Nagarjuna College of Engineering and Technology

(An Autonomous College under VTU)

3.4.4 Number of books and chapters in edited volumes/books published per teacher during the last five years (5)

Sl. No.	Academic year	Number of book chapters published
1	2020-21	1 to 2
2	2019-20	-
3	2018-19	3
4	2017-18	3
5	2016-17	3

3.4.4 Number of books and chapters in edited volumes/books published per teacher during the last five years (5)

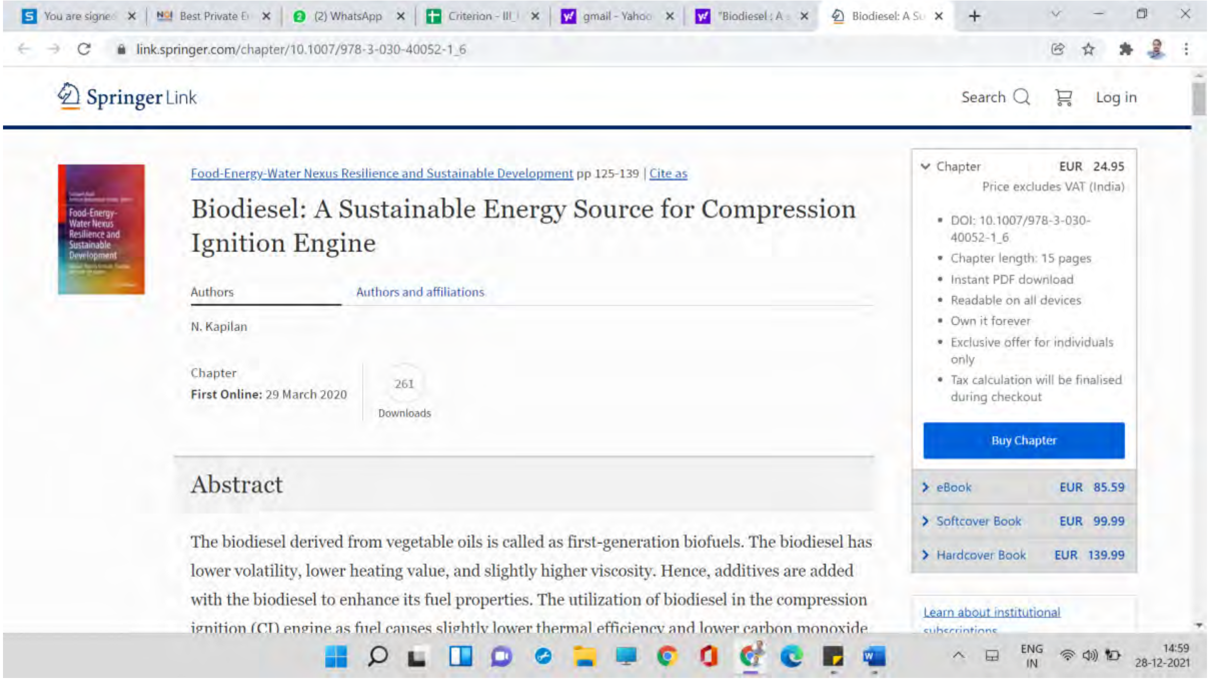
3.4.4.1: Total number of books and chapters in edited volumes / books published, and papers in national/international conference-

Sl. No.	Name of the teacher	Title of the book/chapters published	Title of the paper	Title of the proceedings of the conference	Year of publication	ISBN/ISSN number of the proceeding	Whether at the time of publication Affiliating Institution Was same Yes/NO	Name of the publisher
1	Dr. Vala Ghanshyam Kantilal	Innovations in Technolology, Business and Management	An Empirical Study on Customer-Buyer Behaviour and its impact on customer satisfaction with special reference to maruthi cars at mandovi motors Bengaluru.	International Conference on Innovations in Technolology, Business and Management	2020-2021	978-93-86891-19-8	YES	Enas Publications
2	Dr. Praachi Beriwala	Innovations in Technolology, Business and Management	Women Entrepreneureship actual and national Barriers: A Study in Karnataka with Special reference to Bangalore District	International Conference on Innovations in Technolology, Business and Management	2020-2021	978-93-86891-19-8	YES	Enas Publications
3	Dr. Geethanjali G	Innovations in Technolology, Business and Management	A Study on Challenges & Prospects of Startups with reference to covid-19	International Journal of Advance and Innovative Research	2020-2021	ISSN-2348-2397	YES	Shodh Sarita, An Interantional Bilingual Peer Reviewed Refereed research Journal
4	Dr. Geethanjali G	Proceedings	Accounting For Business	BOOK	2020-2021	ISBN 978-93-90515-10-3	YES	Vision Book House
5	Mr. Arun Kumar S	Advances in Intelligent Systems and Computing	A survey on Graphical Authentication System Resisting Shoulder Surfing Attack	Advances in Artificial Intelligence & Data Engineering	2020-2021	ISSN - 2194-5357	YES	Springer
6	Dr N Kapilan	Role of IoT in Green Energy Systems	Challenges and Issues of IOT Application in Heating Ventilating Air Conditioning Systems		2020-2021	ISBN13: 9781799867098	YES	IGI Global
7	Mr.Amaresh Gunge	Viscoelastic response of polyvinyl alcohol composites	Dynamic mechanical analysis/thermo-mechanical analysis of treated polyvinyl alcohol reinforced natural fiber composites		2020-2021	In- Print	YES	LAMBERT Academic
8	Dr. Anil Kannur	Deep Learning and Its Application in Big Data Information	"Identification of Reasons behind infant Cry Using Acoustic Signal processing and Deep Neural Network for Neonatal Intensive Care Unit"	International journal of information retrieval research	2020-2021	In- Print	YES	IGI Global Publication (SCI Indexing
9	Dr. Anil Kannur	Computational Vision and Bio-inspired Computing	"Identification of Artificial body marks and Skin disorders marks using Artificial Neural Network Approach"	International journal of information retrieval research	2020-2021	In- Print	YES	Scopus Indexing, Springer
10	Mr Sreenivasulu K N	Proceedings	Detection of software intrusion based on machine learning techniques for IOT systems	International Conference on Research in Science, Engineering, Technology and Management (ICRSETM-2020).	2020-2021		YES	Science Direct
11	Dr. Rohith S	Proceedings	An Novel Hand Gesture System for ASL using Kinet Sensor based Images	International Conference On Computing, Communication And Control System (I3CAC) 2021	2020-2021	978-1-63190-306-9 ISSN 2593-7642	YES	EAI Publisher
12	Dr. Sandhya G	Data Science	Data Science book	Book	2020-2021	987-93-91987-04-6	YES	Jayalakshmi Publications
13	Ms. Komal Srivastava, Dr. Ajay Kumar Dwivedi, Dr. Nagesh K.N, Dr. Anand Sharma	Lecture Notes in Electrical Engineering (LNEE), Springer	Dual Band, Dual Port Dielectric Resonator Based MIMO Antenna with Bi-directional Pattern Diversity using PRS	VLSI, Communication and Signal Processing	2020-2021	In-Print	YES	Scopus Indexing, Springer

14	Mr. A.K Singh, Dr. Ajay Kumar Dwivedi, Dr. Nagesh K.N, Dr. V. Singh, Dr. R.S Yadav	Lecture Notes in Electrical Engineering (LNEE), Springer	Modeling and Investigation of Novel Two-Port UWB- MIMO Antenna with Enhanced Isolation	VLSI, Communication and Signal Processing	2020-2021	In-Print	YES	Scopus Indexing, Springer
15	Mr. Mallanna S D, Mr. Sreenivasa T V, Dr. Nagesh K.N, Dr. Ajay Kumar Dwivedi	American Institute of Physics (AIP) Conference Proceedings	Design and Analysis of Four Port MIMO Antenna for n77 and MTI Wireless Edge 5G Applications	International Conference on Industrial Electronics Mechatronics, Electrical & Mechanical Power	2020-2021	In-Print	YES	Scopus Indexed
16	Mr.S. Jadhav and Mr.K. N. Nagesh	Proceedings	Impairment impact on the wireless communication system	2021 2nd Global Conference for Advancement in Technology (GCAT)	2020-2021	978-1-6654-3070-8	YES	IEEE
17	Dr. Rohith S, Dr.Basavaraj G M, Mr.Mahesh M R,Bhavya Sree, Vijay Yellapu	Proceedings of IEEE International Conference on Mobile Networks and Wireless Communications	Image Encryption Scheme Using Combined Key Sequence of Logistic and Tent Map for Medical Applications	2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC)	2020-2021	In-Print	YES	IEEE
18	Dr. Shantakumar B Patil	Proceedings	Computer Peripherals and its Organization A Teacher Handbook		2020-2021	979-8631389250	YES	Independent Publisher
19	Dr. Premjyoti Patil, Dr. Shantakumar Patil	Proceedings	Computer Peripherals and its Organization: A Teacher Handbook	NA	2020-2021	979-8631389250	YES	Independent Publisher
20	Dr. Rohith S	Proceedings of International Conference on Advances in Electronics, Computers	Image Encryption and Decryption Using Key Sequence of Triple Logistic Map for Medical Applications	2020 Third International Conference on Advances in Electronics, Computers and Communications (ICAECC)	2020-2021	978-1-7281-8045-8	YES	IEEE
21	Dr N Kapilan	Food-Energy-Water Nexus: Resilience and	"Biodiesel : A sustainable energy source for compression ignition engine",	Food-Energy-Water Nexus Resilience and Sustainable Development	2020-2021	ISBN 978-3-030-40051-4	YES	Springer
22	Dr N Kapilan	Applications of Nanomaterials in Agriculture, Food Science,	Applications of Nano Materials in Cold Storage	Applications of Nanomaterials in Agriculture, Food Science, and Medicine	2020-2021	9.7818E+12	YES	IGI Global
23	Mr.Sriram	Proceedings	Oppurtunity available for Start Ups in Andaman and Nikobar Islands, with speacial reference to women Entrepreneurs	International Conference on Innovations in Technolology, Business and Management	2020-2021	978-93-86891-19-8	YES	Enas Publications
24	Ms.Rajeshwari	Proceedings	The Role of Information Technology in Shoping of Economic Development in India	International Conference on Innovations in Technolology, Business and Management	2020-2021	978-93-86891-19-8	YES	Enas Publications
25	Dr. Anita Patil, Dr. Srikanta Murthy K, Mr. Bhargava R Mrs. Seema J	Proceedings	Learn Java with Real World Problems	**NA**	2018-2019	9.78939E+12	YES	Akinik Publications
26	Mr. Abraham Rajan	Proceedings	A Systematic And Composed Big Data Entry Restriction Scheme With Isolation-Preserving Policy	2018 International Journal of Advanced Research in Computer Science	2018-2019	0976-5696	YES	International Journal Of Advanced Research In Computer Science
27	Mr.Rohith S, Mr.Chandrashekar	Proceedings	Design of 8-bit Vedic Multiplier Using Urdhva Tiryagbhyam Sutra	2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)	2018-2019	978-1-7281-0631-1	YES	IEEE
28	Mr.S. Jadhav and Mr.K. N. Nagesh	Proceedings	Isolated Node Localization Probablity in Wireless Ad-Hoc Networks	2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAECC)	2018-2019	978-1-5386-3786-9	YES	IEEE
29	Mrs. Ashwini. S	Proceedings	Monitoring and Cotrolling Cold Chin Logistics Using WSN as a part of IOT Technology		2018-2019		YES	

30	Mr. Abraham Rajan	Proceedings	IOT, GPRS, RASPBERRY PI Based Global Industrial Process Monitoring	CiiT International Journal of Programmable Device Circuits and Systems	2018-2019	0974 – 9624	YES	I Scholar
31	Dr. T Y Satheesha	Proceedings	Segmentation of Melanoma Using the EM Algorithm		2018-2019		YES	
32	Dr. T Y Satheesha	Proceedings	Computer-aided Detection of Melanoma Using Geometric Features	National conference on Recent Trends in Electronics and Communication Engineering NCRTECE-18	2017-2018	National	YES	National conference on recent trends in computer science and information technology
33	Dr. Srikanta Murthy K	Proceedings	Efficient Feature Extraction for face recognition with combined method of PCAand GMM	International Conference on Advanced Trends in Computer Science & Information Technology 2018	2017-2018		YES	Science Publishing Corporation
34	Dr. T Y Satheesha	Proceedings	Pre-Processing Technique for detection of Melanoma	3rd National conference on Convergence of Science, Technology and Management NCCSTM-2017	2017-2018		YES	National Conference On Convergence Of Science, Technology And Management
35	Dr. T Y Satheesha	Proceedings	Extraction of melanoma 3D feature from tensor representation	International Conference on Recent Innovations in Engineering Science and Technology 2018	2017-2018	2227-524X	YES	Science Publishing Corporation
36	Mrs. Ashwini. S	Proceedings	Monitoring and Cotrolling Cold Chin Logistics Using WSN as a part of IOT Technology		2017-2018	National	YES	
37	Mr.Basavaraj G M and Mr.Ashok Kusagur	Proceedings	Optical and Streakline flow based crowd estimation for surveillance system(Citation: 2)	2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)	2016-2017	978-1-5090-0775-2	YES	IEEE
38	Dr. Srikanta Murthy K	Proceedings	A Novel Approach for Detection of Breast Cancer at an early stage by Identification of Breast Asymmetry and Microcalcification cancer cells using Digital Image Processing techniques	2nd International Conference for Convergence in Technology (I2CT)	2016-2017	978-1-5090-4308-8	Yes	IEEE
39	Dr. Srikanta Murthy K	Proceedings	A Novel Approach for Detection of Breast Cancer at an Early Stage Using Digital Image Processing Techniques	2017 International Conference on Inventive Systems and Control (ICISC)	2016-2017	978-1-5090-4716-1	Yes	IEEE
40	Mr.Basavaraj G M and Mr.Ashok Kusagur	Proceedings	Vision Based Surveillance System for Detection of Human Fall(Citation :9)	2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)	2016-2017	978-1-5090-3705-6	YES	IEEE
41	Dr. T Y Satheesha	Proceedings	KNNs Classification based on insitu melanoma	International Journal of Advanced Technology & Engineering Research (IJATER)	2016-2017		YES	International Journal of Advanced Technology & Engineering Research (IJATER)
42	Dr. T Y Satheesha	Proceedings	Skin Lesion Analysis System for Melanoma Detection with an Effective Hair Segmentation Method	IEEE Journal of Translational Engineering in Health and Medical Sciences	2016-2017	2168-2372	YES	National conference on recent trends in computer science and information technology
43	ASHWINI, ROHITH S, SUNITA K A,	Proceedings	Implementation of High Speed and Low Power 5T-TSPC D Flip-flop and Its Application	2016 International Conference on Communication and Signal Processing (ICCSP)	2016-2017	978-1-4673-8549-7	YES	IEEE

[Biodiesel: A Sustainable Energy Source for Compression Ignition Engine | SpringerLink](#)



link.springer.com/chapter/10.1007/978-3-030-40052-1_6

Springer Link

Search Log in

Food-Energy-Water Nexus Resilience and Sustainable Development pp 125-139 | Cite as

Biodiesel: A Sustainable Energy Source for Compression Ignition Engine

Authors | Authors and affiliations

N. Kapilan

Chapter
First Online: 29 March 2020

261 Downloads

Chapter EUR 24.95
Price excludes VAT (India)

- DOI: 10.1007/978-3-030-40052-1_6
- Chapter length: 15 pages
- Instant PDF download
- Readable on all devices
- Own it forever
- Exclusive offer for individuals only
- Tax calculation will be finalised during checkout

Buy Chapter

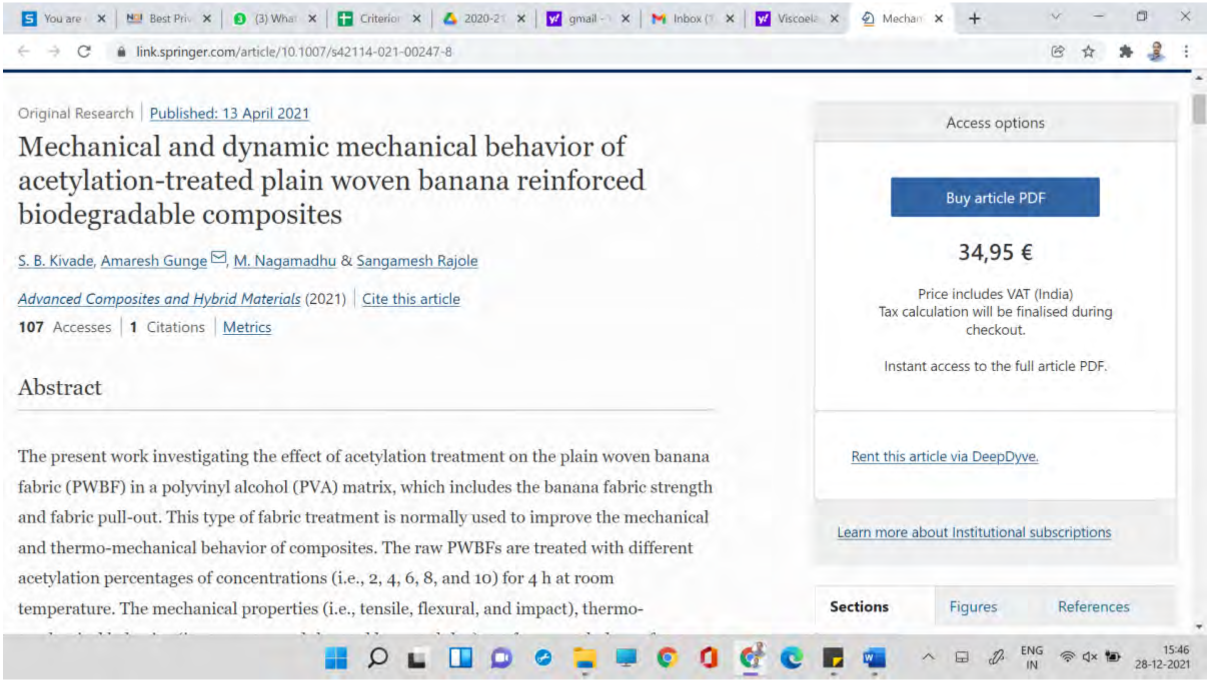
eBook EUR 85.59
Softcover Book EUR 99.99
Hardcover Book EUR 139.99

Learn about institutional subscriptions

Abstract

The biodiesel derived from vegetable oils is called as first-generation biofuels. The biodiesel has lower volatility, lower heating value, and slightly higher viscosity. Hence, additives are added with the biodiesel to enhance its fuel properties. The utilization of biodiesel in the compression ignition (CI) engine as fuel causes slightly lower thermal efficiency and lower carbon monoxide

https://link.springer.com/article/10.1007/s42114-021-00247-8



link.springer.com/article/10.1007/s42114-021-00247-8

Original Research | Published: 13 April 2021

Mechanical and dynamic mechanical behavior of acetylation-treated plain woven banana reinforced biodegradable composites

S. B. Kivade, Amaresh Gunge, M. Nagamadhu & Sangamesh Rajole

Advanced Composites and Hybrid Materials (2021) | Cite this article

107 Accesses | 1 Citations | Metrics

Access options

Buy article PDF

34,95 €

Price includes VAT (India)
Tax calculation will be finalised during checkout.

Instant access to the full article PDF.

Rent this article via DeepDyve.

Learn more about Institutional subscriptions

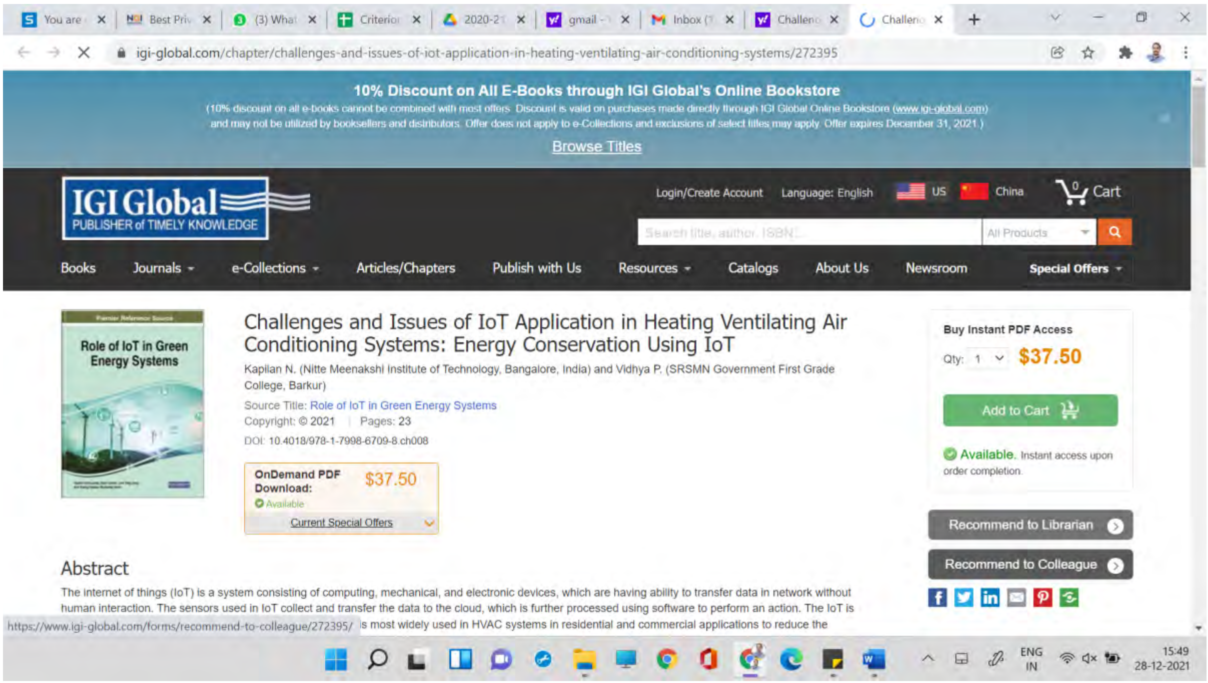
Sections Figures References

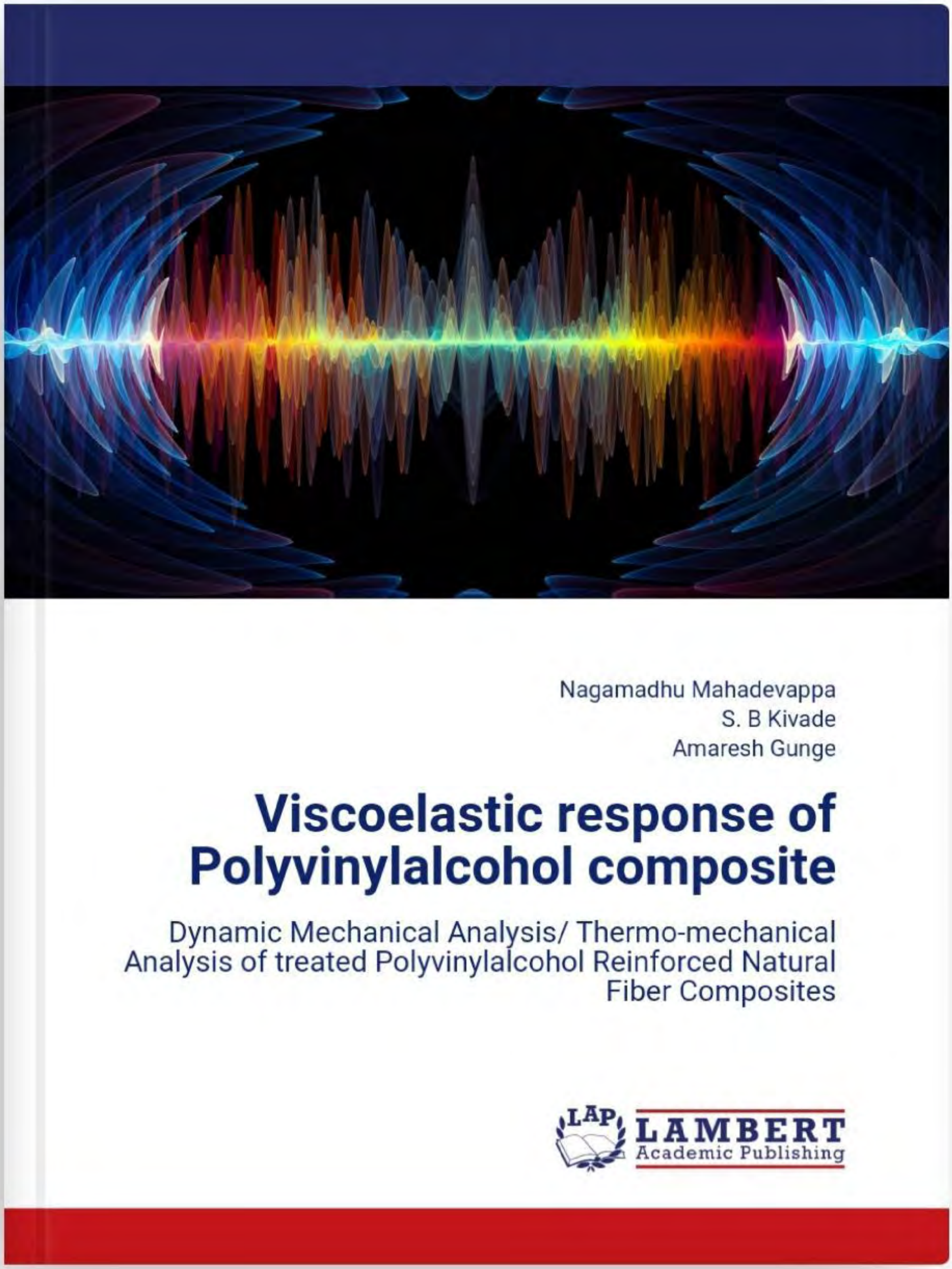
Abstract

The present work investigating the effect of acetylation treatment on the plain woven banana fabric (PWBF) in a polyvinyl alcohol (PVA) matrix, which includes the banana fabric strength and fabric pull-out. This type of fabric treatment is normally used to improve the mechanical and thermo-mechanical behavior of composites. The raw PWBFs are treated with different acetylation percentages of concentrations (i.e., 2, 4, 6, 8, and 10) for 4 h at room temperature. The mechanical properties (i.e., tensile, flexural, and impact), thermo-

https://www.igi-global.com/chapter/applications-of-nano-materials-in-cold-storage/268820

https://www.igi-global.com/chapter/challenges-and-issues-of-iot-application-in-heating-ventilating-air-conditioning-systems/272395




Nagamadhu Mahadevappa
S. B Kivade
Amaresh Gunge
Viscoelastic response of Polyvinylalcohol composite
Dynamic Mechanical Analysis/ Thermo-mechanical Analysis of treated Polyvinylalcohol Reinforced Natural Fiber Composites
LAP LAMBERT Academic Publishing

2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC)

Image Encryption Scheme Using Combined Key Sequence of Logistic and Tent Map for Medical Applications

ROHITH S
ECE Department
NCET, Bengaluru
Karnataka, India
rohithvjp2006@gmail.com

C Bhavya Sree
ECE Department
NCET, Bengaluru
Karnataka, India
cbhavyasree20@gmail.com

Vijay Yellapu
ECE Department
NCET, Bengaluru
Karnataka, India
vijayyellapu08@gmail.com

Basavaraj G M
ECE Department
NCET, Bengaluru
Karnataka, India
gmb2206@gmail.com

Mahesh M R
ECE Department
NCET, Bengaluru
Karnataka, India
mahesathya@gmail.com

***Abstract*—Securing the clinical records of the patient is very much important. Conventional encryption methods might not be robust for the applications of medical images due to their statistical properties. In this paper, the Combination of Logistic and Tent map based on image encryption method for medical image application is proposed. Initially, the Logistic and Tent map key sequence is combined by employing an XOR operation. The resulting sequence is XORed with medical image pixels. Hence the encrypted image is obtained. To evaluate the performance, ten different test images are chosen. The parameters like Entropy, Correlation and Mean Square Error (MSE) are used to analyse the security of the proposed scheme. Further, the proposed scheme is compared with schemes using Tent map and Logistic map alone. The results obtained demonstrate that the proposed encryption scheme provides better performance compared with a scheme using the sequence of Logistic map or Tent map alone.**

Keywords—Combined Key sequence, Medical Image encryption, chaotic map, Logistic Map, Tent Map

I. INTRODUCTION

With the new developments in Personal Computer (PC) and innovation in Internet Technology, the creation, transmission and storage of digital images are more normal. For the diagnosis and therapy of the patient, medical images play a significant role. These medical image provides visual information of the body organs and is achieved through Computed Tomography (CT), Magnetic Resonance Imaging (MRI), X-ray and ultrasound methodology. The majority of those images are going to be exchanged through wired and wireless media or stored in digital devices. The safety of the medical images is mandatory due to private data about the patient health. Henceforth medical images need to be protected from unauthorized users. Encryption schemes are generally used to make sure image protection against unauthorized users during transmission or storage. using an encryption scheme. Encryption may be a process of protecting information to such an extent that it cannot be perused or decoded without applying the proper key. The traditional algorithm, for instance, DES, AES and RSA cryptosystems [1] might not be appropriate for the applications of image encryption. The chaos-based algorithms may be one of the solutions for image encryption [2].

Chaos-based schemes exhibit statistical properties of periodicity with longer orbit, sensitivity to initial conditions and control parameters, simple execution and high encryption rate [3-8]. The chaotic sequences are unpredictable and partially noise like structures. Subsequently, it's utilized in image encryption methods than conventional encryption procedures. Many chaotic maps like logistic map, Henon map, Lozi map, Arnold map, baker map, tent map are used for image encryption applications [3][4][6][7][8][9][10][11][12]. For medical image encryption applications security is a big threat due to the confidentiality of the patient data. In the case of image encryption applications, many scientists suggested using a mixture of different chaos maps, provides better results than a single map.

Many Authors proposed Logistic map based encryptions schemes [6][7][11][12]. In [4] image encryption method based on two logistic map is discussed. They used two logistic separate sequences for encryption of the image. Image encryption based on the varying parameter of the logistic map is discussed in [11]. In [6] Image Encryption using Three different sequences of Logistic Map is discussed. They used three different discrete Logistic sequences were used to generate the key sequence. Image encryption is carried out using the generated key sequence. In [7] image encryption scheme employing a mixture of a Logistic map and chen map was examined. Image encryption based on the permutation-diffusion technique is discussed in [13]. In [9] image encryption by the combination of the logistic map, Henon map and 3D - Baker map was used. The results show the performance of the scheme is best. However, the Histogram of the encrypted image is not uniform. In [10] image encryption scheme for biomedical applications is discussed. The key sequence is derived using two discrete logistic maps. In [8] Image Encryption Scheme based on the combination of Tent map, Sine map and Logistic Map is discussed. After the generation of the sequence, they used simultaneous Confusion and diffusion approach to scramble the image.

In this paper, a mixture of Logistic map and Tent map is employed for the image encryption. The logistic sequence $\{X_i\} \in \{0, 1\}$ is transformed into discrete sequence in the range of 0 to 255. Similarly, the discrete key sequence for Tent map $\{T_i\}$ is also obtained. Finally, using a logical XOR operation, a discrete sequence of logistic map and tent map is

978-0-7381-4637-9/21/$31.00 ©2021 IEEE

combined. The resultant sequence is used to encrypt the medical image.

The rest of the article is arranged as follows. The outline of Logistic Map, Tent Map and the proposed scheme is provided in Section II. In Section III Result and discussion of the proposed scheme is given. In Section IV overall conclusion of the paper is discussed.

II. Proposed Scheme

A. The Logistic Map:

It is a one-dimensional (1D) discrete-map that shows chaotic behaviour. It has a basic equation, however, it shows an unpredicted level of complexity. It is expressed in (1).

$$X_{n+1} = rX_n\,(1- X_n) \tag{1}$$

Where X_n may be an integer sequence that ranges between 0 to 1, X_0 is an initial value, bifurcation parameter 'r' may be a positive integer and its value ranges between 0 to 4. Using equation (1), with chosen value of 'X_0' and bifurcation parameter 'r' the logistic sequence $\{X_i\} = \{X_1, X_2,\ldots..X_t\}$ were obtained. The sequence $\{X_i\}$ is random and ranges 0 to 1[6].

Further, the sequence $\{X_i\}$ is transformed into the discrete sequence in the range of 0 to 255 using equation (2) and used for the encryption process.

$$S_{1,i}= Round\,(X_i*255) \tag{2}$$

B. The Tent Map:

The word tent map is because the structure of the graph resembles the structure of a tent as shown in Fig. 1. It is observed in the figure that two straight lines aligned like a tent. The tent map is simpler than nonlinear systems. It is a kind of equation with a particular control parameter 'µ' value. The tent map function is given by

$$T_{n+1} = \begin{cases} \mu Yn & for\ Yn < \frac{1}{2} \\ \mu(1 - Yn) & for\ Yn \geq \frac{1}{2} \end{cases} \tag{3}$$

Where the control parameter µ ranges between 0 to 2. The initial value Y_0 € [0, 1] which results in a sequence T_n. Where 'µ' values are real and positive.

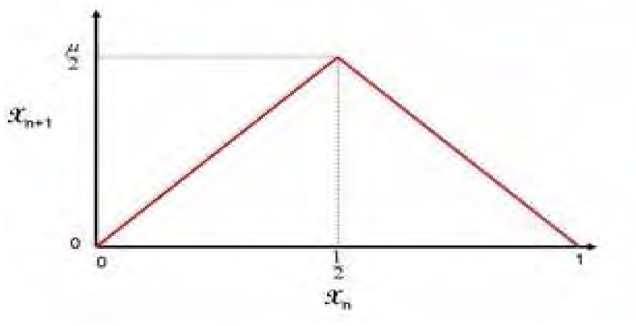


Fig. 1. The plot of the Tent map

Choosing parameter µ with Y_0 as an initial value sequence is obtained using equation (3). The generated sequence $\{T_i\}$ € (0,1) is converted into an integer that ranges from 0 to 255 using equation (4). This sequence is known as $S_{2,i}$.

$$S_{2,i} = Round\,(Y_i*255) \tag{4}$$

Final key sequence $\{S_i\}$ is derived by performing simple XOR operation with $\{S_{1,i}\}$ and $\{S_{2,i}\}$ using equation (5). The resultant sequence is used for image encryption.

$$\{S_i\} = \{S_{1,i}\} \oplus \{S_{2,i}\} \tag{5}$$

C. The Proposed Encryption method

This section discusses the proposed image encryption and decryption method using Logistic and tent map. The detailed algorithm is given below.

Step1: Initially, an 8 bit grayscale 2D image is converted into sequence of 1D image pixels $\{C_i\} = \{C_1, C_2, \ldots C_{M\times N}\}$. M and N represent the total number of rows and columns of the image.

Step2: The 1D sequence of pixels $\{C_i\} = \{C_1, C_2, \ldots C_t\}$ are XORed with the key sequence $\{S_i\}=\{S_1, S_2, S_3\ldots S_t\}$ obtained using equation (5). Resulting sequence is encrypted image pixels $\{E_i\} = \{E_1, E_2, \ldots E_{M\times N}\}$ is given by equation (6)

$$\{E_i\} = \{C_i\} \oplus \{S_i\} \tag{6}$$

Step 3: The resulted 1D sequence of encrypted pixels $\{E_i\}= \{E_1, E_2, \ldots E_t\}$ is transformed into an array of 2D of size M×N. So obtained a 2D array of pixels is named encrypted image E(x, y).

D. The Proposed Decryption method

The stepwise decryption process of the proposed image decryption method is given below.

Step1: Initially, an encrypted image E(x,y) is converted into the 1D array sequence $\{E_i\} =\{E_1, E_2, \ldots E_{M\times N}\}$.

Step2: Decryption process is applied using simple XOR operation between encrypted image pixels $\{E_i\}$ and proposed combined key sequence $\{S_i\}$. The sequence of Decrypted image pixels is obtained using equation (7)

$$\{D_i\} = \{E_i\} \oplus \{S_i\} \tag{7}$$

Step3: The Sequence of Decrypted image pixels $\{D_i\}$ are converted into an array of the 2D image. Hence the resultant image is named as decrypted image D (x,y).

III. Results and Discussion

The image encryption algorithm is implemented through MATLAB tool with version R2014a. The grayscale medical images such as the lungs, teeth, brain, eye and head are chosen as test images to demonstrate the results of the proposed scheme. The parameters like i) Visual analysis ii) Histogram plot iii) Mean Square Error (MSE) iv) Entropy and v) Correlation are used to evaluate the proposed algorithm. Finally, comparisons are made between encryption using the Logistic map sequence $\{S_{1i}\}$ alone, Encryption using tent map sequence $\{S_{2i}\}$ alone and the proposed image encryption scheme.

A. Visual Analysis

Visual analysis is one of the parameters to analyse the encrypted image. About ten different medical images are randomly chosen is depicted in Fig. (2a) (1-10) to test the proposed scheme. The encrypted image using the logistic map, tent map individually tested and proposed is also illustrated in Fig. (2b), Fig. (2c) & Fig. (2d). It is seen that images that have been encrypted do not reveal any information about the original image. Compared to encryption using the Logistic map or the Tent map alone, the performance of the proposed encryption scheme is superior.

To test the sensitivity of the proposed method, a small change in the initial key with 10^{-12} in the decimal value is applied while the generation of the key sequence. The encrypted image and the original image are entirely different after employing the key sequence. It indicates even with a very small change in the initial value of X_0 of the logistic map or Y_0 of the tent map, the decryption of the original image is impossible. It implies the proposed method is very much sensitive to the key sequence.

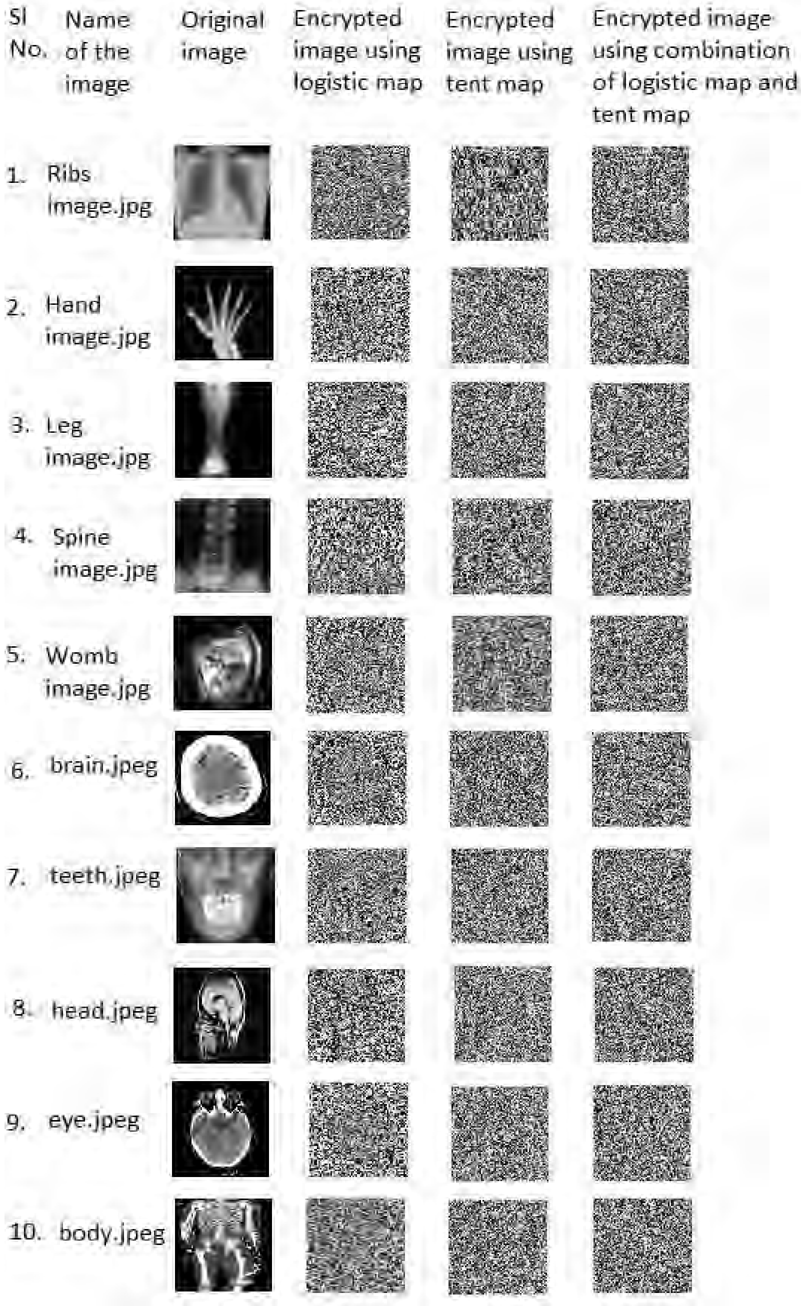


Fig. 2. Visual analysis Fig. 2(a) Original image. Fig. 2(b): Encrypted image using Logistic map {S1i} with X0 =0.1 and r=3.99 Fig. 2(c): Encrypted image using Tent map {S2i} with Y0 =0.5 and μ =1.99. Fig. 2(d): Encrypted image Proposed encryption scheme {Si} with X0 =0.1, r=3.99, Y0=0.5, μ =1.99.

B. Histogram Analysis.

The Histogram plot of the image may be a pictorial representation of the pixel values of the image. It shows the distribution of the grey level values in the range of 0 to 255. In this paper, a plot of the histogram is employed to see the statistical behaviour of the encrypted image. The results of the histogram of the original image and encrypted image for all three cases are shown in Fig 3. Fig. 3(a) depicts the original image histogram plot. Fig. 3(b) depicts Logistic map $\{S_{1,i}\}$ based encrypted image Histogram plot. Fig. 3(c) shows a tent map $\{S_{2,i}\}$ based encrypted image. The histogram plot of the proposed scheme $\{S_i\}$ is shown in Fig. 3(d). It is observed from the results of the original image histogram plot has some peaks and it's irregularly distributed. However, in the case of encrypted images histogram plot is uniformly distributed. In the case of Encrypted image Histogram plot using the proposed method, it is fully flat and pixels are uniformly distributed as compared with logistic map $\{S_{1i}\}$ or tent map-based method $\{S_{2i}\}$ alone. It indicates the proposed method can be robust against statistical attacks.

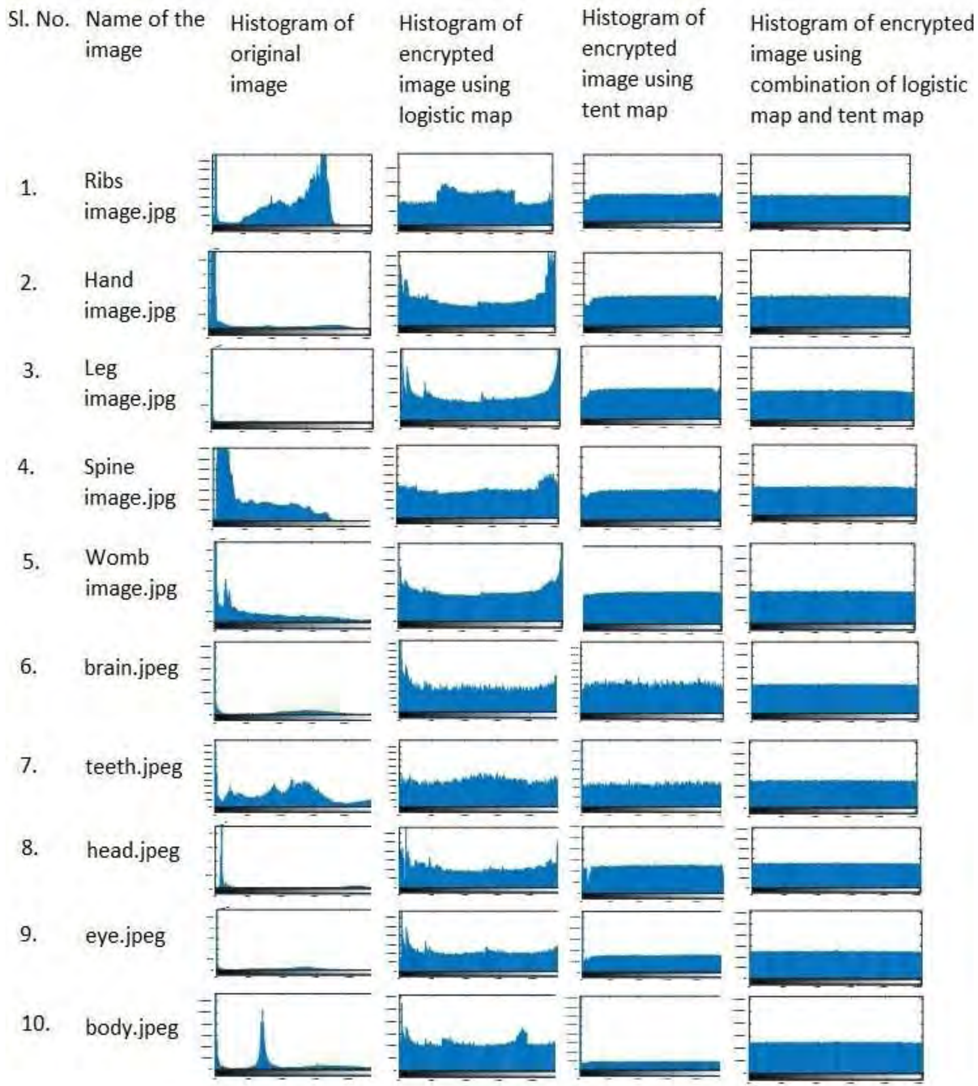


Fig. 3. Histogram analysis Fig. 3(a) Original image Histogram plot. Fig. 3(b): Encrypted image Histogram plot with Logistic map{S1i} sequence X0 =0.1 and r=3.99 Fig. 3(c): Encrypted image Histogram plot with sequence of Tent map {S2i}with Y0 =0.5 and μ =1.99. Fig. 3(d): Encrypted image Histogram plot with sequence of Proposed encryption scheme {Si} with X0 =0.1, r=3.99, Y0=0.5, μ =1.99

C. *Mean Square Error Analysis*

It is a numerical parameter constant used to distinguish original and encrypted images. It is defined in equation (8).

$$MSE = \frac{1}{MN}\sum_{i=0}^{M\times N-1}[C(i,j) - E(i,j)]^2 \qquad (8)$$

Where C (i, j) is a pixel value of the original image

E (i, j) is a pixel value of encrypted image pixel

M and N are the total numbers of rows and columns of the image.

The ideal value of MSE for image encryption should be as high as possible. It indicates that a higher MSE value between the encrypted and original image, more resistance to statistical attacks. In Table1, MSE is computed between original and encrypted images shown in Fig. 2. It is found in Table 1 that, when compared to the logistic map or tent map alone, the MSE value of the proposed encryption technique exhibits better performance.

TABLE I. MSE VALUE BETWEEN ORIGINAL AND ENCRYPTED IMAGE

Sl no	Image	MSE (Logistic map)	MSE (Tent maps)	MSE (Proposed scheme)
01	Ribs image.jpg	7696	8630	8365
02	Hand image.jpg	20574	17911	17387
03	Leg image.jpg	17379	18715	18173
04	Spine image.jpg	16530	13251	12863
05	Womb image.jpeg	25446	14934	14514
06	Brain.jpeg	19414	17232	16565
07	Teeth.jpeg	11720	11318	10877
08	Head.jpeg	19504	17035	16569
09	Eye.jpeg	19922	17610	17135
10	Body.jpeg	15928	14566	14148

D. *The entropy of the image*

The Entropy is used to compute the unpredictability of the image pixels. It determines the degree of unexpected information [2][5]. The information entropy E(C) for source information I is defined in equation (9)

$$\mathrm{H(I)} = \sum_{\mathrm{i=0}}^{\mathrm{M\times N\text{-}1}} \mathrm{P(I_i)} \log_2 \frac{1}{\mathrm{P(I_i)}} \quad (9)$$

Where P (I_i) is the probability of the symbol.

M×N is the total count of the symbols I_i.

The highest entropy value for an 8-bit grayscale image is 8. It occurs when the image pixel values are equiprobable. An ideally larger value of entropy occurs when pixels are uniformly distributed in an image.

The entropy values were computed for the original image and encrypted image by Logistic map, tent map & proposed method for all the ten different medical images and tabulated in Table 2. It is observed that, when compared to encryption using a logistic map $\{S_{1i}\}$ or a tent map $\{S_{2i}\}$ alone, the entropy of the encrypted images found to be high. Hence, the method proposed is very robust against statistical attacks.

TABLE II. ENTROPY VALUES OF THE ORIGINAL AND ENCRYPTED IMAGE

Sl No	**Image Type**	**Entropy (Original image)**	**Entropy (Encrypted image using Sequence {S1,i}of Logistic map)**	**Entropy (Encrypted image using Sequence {S2,i} of Tent map)**	**Entropy (Proposed scheme)**
01	Ribs image.jpg	7.1161	7.9644	7.9985	7.9993
02	Hand image.jpg	5.3034	7.9156	7.9955	7.9996
03	Leg image.jpg	3.7391	7.8789	7.9887	7.9991
04	Spine image.jpg	7.1309	7.9851	7.9987	7.9995
05	Womb image.jpeg	6.7919	7.9726	7.9984	7.9998
06	Brain.jpeg	4.8579	7.9271	7.9904	7.9962
07	Teeth.jpeg	7.4039	7.9874	7.9956	7.9963
08	Head.jpeg	4.5073	7.8958	7.9923	7.9993
09	Eye.jpeg	3.9279	7.8918	7.9910	7.9993
10	Body.jpeg	6.0234	7.9637	7.9972	7.9994

E. The Correlation analysis

The parameter Correlation coefficient [2][5] defined in equation (10)

$$Correlation\ Coefficient = \frac{Cov(x,y)}{\sqrt{A(x)}\sqrt{A(y)}} \quad (10)$$

Covariance of x, y is given in equation (11)

$$Cov(x,y) = \frac{1}{L}\sum_{i=0}^{L}(x_i - B(x)(y_i - B(y) \quad (11)$$

x and y are values of adjacent pixels.

L is total number of pixels

A(x) and A(y) are variance of x_i and y_i respectively in equation (12) and (13).

$$A(x) = \frac{1}{L}\sum_{i=0}^{L}\left(x_i - B(x)\right)^2 \quad (12)$$

$$A(y) = \frac{1}{L}\sum_{i=0}^{L}\left(x_i - B(y)\right)^2 \quad (13)$$

B(x) and B(y) is mean value and given by equation (14) and (15)

$$B(x) = \frac{1}{L}\sum_{i=0}^{L} x_i \quad (14)$$

$$B(y) = \frac{1}{L}\sum_{i=0}^{L} y_i \quad (15)$$

The correlation coefficient is calculated between the original and encrypted image for Logistic map $\{S_{1i}\}$, Tent map $\{S_{2i}\}$, and proposed technique $\{S_i\}$. Ideally coefficient of the correlation value is '0'. From Table 3, it is found that the coefficient of the correlation value is low in the proposed scheme when compared to the scheme using tent map or Logistic map alone. It implies that pixels of the encrypted image in the case of the proposed scheme is highly uncorrelated.

TABLE III. CORRELATION COEFFICIENT

Sl No	**Name of the image**	**Correlation (Logistic map)**	**Correlation (Tent map)**	**Correlation (Proposed Scheme)**
01	Ribs image.jpg	-0.0702	-0.0237	0.0023
02	Hand image.jpg	-0.0509	-0.0212	0.0020
03	Leg image.jpg	-0.0485	-0.0222	0.0035
04	Spine image.jpg	-0.0420	-0.0179	0.0033
05	Womb image.jpeg	-0.0610	-0.0245	0.0018
06	brain.jpeg	-0.0897	-0.0413	0.0119
07	Teeth.jpeg	-0.0702	-0.0224	0.0081
08	Head.jpeg	-0.0609	-0.0263	0.0018
09	Eye.jpeg	-0.0638	-0.0269	0.000784
10	Body.jpeg	-0.0731	-0.0323	0.0011

IV. CONCLUSION

Image Encryption using a Combination of tent map and logistic map for medical applications is proposed in this paper. The sequence {Si} generated was used to encrypt the medical image. Further, results are compared with encryption using a tent map or logistic map-based scheme. The proposed encryption scheme provides better performance with respect to Correlation, Entropy, MSE parameters compared to the scheme of logistic or tent map alone. It is found that the proposed method is simple, robust against statistical attacks.

REFERENCES

[1] W. Stallings, "Cryptography and network security - principles and practice," 7th ed., Pearson Education, 20016, ISBN 978-0-13-111502-6.

[2] S. Lin, and D. Costello, "Error Control Coding," 2nd ed., Pearson Education, 2004, 978-0130426727.

[3] C. Li, T. Xie, Q. Liu, and G. Cheng, "Cryptanalyzing image encryption using chaotic logistic map," An International Journal of Nonlinear Dynamics and Chaos in Engineering Systems, vol. 78, no. 2, pp. 1545–1551, 2014. DOI: 10.1007/s11071-014-1533-8, ISSN 0924-090X

[4] H. Pan, Y. Lei, and C. Jian, "Research on digital image encryption algorithm based on double logistic chaotic map," EURASIP Journal on Image and Video Processing, vol. 142, pp. 1–10, 2018. DOI: https://doi.org/10.1186/s13640-018-0386-3

[5] C. E. Shannon, "A Mathematical Theory of Communication," The Bell System Technical Journal, vol. 27, no. 3, pp. 379–423, July 1948.

[6] S. Rohith, L. Jahnavi, S. C. Bhuvaneshwari, S. Supreeth, and B. K. Sujatha, “Image Encryption and Decryption Using Key Sequence of Triple Logistic Map for Medical Applications,” in Proc. of IEEE ICAECC, REVA University, India, 2020, pp. 1–5.

[7] L. Hong, and C. Li, “A Novel Color Image Encryption Approach Based on Multi-chaotic System,” in Proc. of IEEE International Conference on A-counterfeiting, Security and Identification, 2008, pp. 1–5.

[8] G. Zhang, W. Ding, and L. Li, “Image Encryption Algorithm Based on Tent Delay-Sine Cascade with Logistic Map,” Symmetry, vol. 12, no. 3, pp. 355, 2020. https://doi.org/10.3390/sym12030355

[9] Y. Honglei, and W. Guang-shou, “The compounded chaotic sequence research in image encryption algorithm,” in Proceedings of WRI Global Congress on Intelligent Systems, 19-21 May, 2009, vol. 3, pp. 252–256.

[10] Ismail, M. Samar, “Biomedical image encryption based on double-humped and fractional logistic maps,” in Proc. of IEEE international conference on modern circuits and systems technologies (MOCAST), 2017, pp. 1–4.

[11] L. Liu, and S. Miao, “A new image encryption algorithm based on logistic chaotic map with varying parameter,” Springer Plus, vol. 5, no. 289, pp. 1–12, 2016.

[12] M. Kumar, and P. Gupta, “A new medical image encryption algorithm based on the 1D logistic map associated with pseudo-random numbers,” International Journal of Multimedia Tools Applications, vol. 80, pp. 18941–18967, 2021.

[13] R. Enayatifar, A. H. Abdullah, I. F. Isnin, A. Altameem, M. Lee, “Image encryption using a synchronous permutation-diffusion technique,” Optics and Lasers in Engineering, vol. 90, pp. 146–154, 2017.

Lecture Notes in Electrical Engineering 733

V. Bindhu
João Manuel R. S. Tavares
Alexandros-Apostolos A. Boulogeorgos
Chandrasekar Vuppalapati *Editors*

International Conference on Communication, Computing and Electronics Systems

Proceedings of ICCCES 2020

Springer

Lecture Notes in Electrical Engineering

Volume 733

The book series *Lecture Notes in Electrical Engineering* (LNEE) publishes the latest developments in Electrical Engineering - quickly, informally and in high quality. While original research reported in proceedings and monographs has traditionally formed the core of LNEE, we also encourage authors to submit books devoted to supporting student education and professional training in the various fields and applications areas of electrical engineering. The series cover classical and emerging topics concerning:

- Communication Engineering, Information Theory and Networks
- Electronics Engineering and Microelectronics
- Signal, Image and Speech Processing
- Wireless and Mobile Communication
- Circuits and Systems
- Energy Systems, Power Electronics and Electrical Machines
- Electro-optical Engineering
- Instrumentation Engineering
- Avionics Engineering
- Control Systems
- Internet-of-Things and Cybersecurity
- Biomedical Devices, MEMS and NEMS

For general information about this book series, comments or suggestions, please contact leontina.dicecco@springer.com.

To submit a proposal or request further information, please contact the Publishing Editor in your country:

China

Jasmine Dou, Editor (jasmine.dou@springer.com)

India, Japan, Rest of Asia

Swati Meherishi, Editorial Director (Swati.Meherishi@springer.com)

Southeast Asia, Australia, New Zealand

Ramesh Nath Premnath, Editor (ramesh.premnath@springernature.com)

USA, Canada:

Michael Luby, Senior Editor (michael.luby@springer.com)

All other Countries:

Leontina Di Cecco, Senior Editor (leontina.dicecco@springer.com)

**** This series is indexed by EI Compendex and Scopus databases. ****

More information about this series at http://www.springer.com/series/7818

V. Bindhu · João Manuel R. S. Tavares ·
Alexandros-Apostolos A. Boulogeorgos ·
Chandrasekar Vuppalapati
Editors

International Conference on Communication, Computing and Electronics Systems

Proceedings of ICCCES 2020

Springer

Editors
V. Bindhu
Department of ECE
PPG Institute of Technology
Coimbatore, India

João Manuel R. S. Tavares
Departamento de Engenharia Mecânica
Faculdade de Engenharia
Universidade do Porto
Porto, Portugal

Alexandros-Apostolos A. Boulogeorgos
Aristotle University of Thessaloniki
Thessaloniki, Greece

Chandrasekar Vuppalapati
San Jose State University
San Jose, CA, USA

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-33-4908-7 ISBN 978-981-33-4909-4 (eBook)
https://doi.org/10.1007/978-981-33-4909-4

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

We are honored to dedicate the proceedings of ICCCES 2020 to all the participants, organizers and editors of ICCCES 2020.

Preface

This conference proceedings volume contains the written versions of most of the contributions presented during the ICCCES 2020 Conference. The conference has provided a platform to share and exchange the recent developments in a wide range of topics including computational intelligence, machine learning, signal and image processing, electronic devices and systems, antenna and wave propagation, wireless communication networks and so on. The conference has been a good opportunity for participants coming from various destinations to present and discuss the state-of-the-art topics in their respective research areas.

ICCCES 2020 Conference tends to collect the latest research results and applications on computing, communication and electronics. It includes a selection of 66 papers from 256 papers submitted to the conference from various universities and industries present across the globe. All the accepted papers were subjected to double-blinded peer-reviewing process by 2–4 expert referees. The papers are selected for its high quality and the relevance to the conference.

ICCCES 2020 would like to express our gratitude and appreciation to all the authors for their valuable research contributions to this book. We would like to extend our thanks to Guest Editors **Dr. V. Bindhu, PPG Institute of Technology, India; Dr. João Manuel R. S. Tavares, Rua Doutor Roberto Frias, Porto, Portugal; Dr. Alexandros-Apostolos A. Boulogeorgos, Aristotle University of Thessaloniki, Greece; Dr. Chandrasekar Vuppalapati, San Jose State University, USA;** and all the referees for expressing their constructive comments on all the research papers. In particular, we would like to thank the organizing committee for their tireless hard work. Finally, we would like to thank the Springer publications for producing this volume.

Dr. V. Bindhu
Conference Chair, ICCCES 2020
Head of the Department
Department of ECE, PPG Institute of Technology
Coimbatore, India

João Manuel R. S. Tavares
Porto, Portugal

Alexandros-Apostolos A. Boulogeorgos
Thessaloniki, Greece

Chandrasekar Vuppalapati
San Jose, USA

Contents

About the Editors

V. Bindhu received the B.E. degree in Electronics and Communication Engineering from Bharathiar University, Coimbatore, in 2002, M.E. degree in Applied Electronics from Anna University, Chennai, in 2007, and Ph.D. degree from Anna University, Chennai, in 2014. She has 10 years of teaching experience and 5 years of research experience. Currently, she is Professor at PPG Institute of Technology, Coimbatore. Her area of interest includes signal processing and VLSI design.

João Manuel R. S. Tavares graduated in Mechanical Engineering at the Universidade do Porto, Portugal in 1992. He also earned his M.Sc. degree and Ph. D. degree in Electrical and Computer Engineering from the Universidade do Porto in 1995 and 2001, and attained his Habilitation in Mechanical Engineering in 2015. He is a senior researcher at the Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial (INEGI) and Associate Professor at the Department of Mechanical Engineering (DEMec) of the Faculdade de Engenharia da Universidade do Porto (FEUP).

João Tavares is co-editor of more than 60 books, co-author of more than 50 book chapters, 650 articles in international and national journals and conferences, and 3 international and 3 national patents. He has been a committee member of several international and national journals and conferences, is co-founder and co-editor of the book series "Lecture Notes in Computational Vision and Biomechanics" published by Springer, founder and Editor-in-Chief of the journal "Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization" published by Taylor & Francis, Editor-in-Chief of the journal "Computer Methods in Biomechanics and Biomedical Engineering" published by Taylor & Francis, and co-founder and co-chair of the international conference series: CompIMAGE, ECCOMAS VipIMAGE, ICCEBS and BioDental. Additionally, he has been (co-) supervisor of several MSc and PhD thesis and supervisor of several postdoc projects, and has participated in many scientific projects both as researcher and as scientific coordinator.

His main research areas include computational vision, medical imaging, computational mechanics, scientific visualization, human-computer interaction and new product development.

Alexandros-Apostolos A. Boulogeorgos was born in Trikala, Greece, in 1988. He received the Electrical and Computer Engineering (ECE) diploma degree and Ph.D. degree in Wireless Communications from the Aristotle University of Thessaloniki (AUTh) in 2012 and 2016, respectively. From November 2012, he has been a member of the wireless communications system group (WCSG) of AUTh, while, from November 2017, he has joined the Department of Digital Systems, University of Piraeus. He also serves as an adjunct professor in the department of ECE in the University of Western Macedonia and as a visiting lecturer at the University of Thessaly. Finally, he is a Senior Member of the IEEE as well as the Technical Chamber of Greece, and serves as editor in IEEE Communications Letters.

Chandrasekar Vuppalapati is Software IT Executive with diverse experience in Software Technologies, Enterprise Software Architectures, Cloud Computing, Big Data Business Analytics, Internet Of Things (IoT), and Software Product & Program Management. Chandra held engineering and product leadership roles at GE Healthcare, Cisco Systems, Samsung, Deloitte, St. Jude Medical, Lucent Technologies, and Bell Laboratories Company. Chandra teaches Software Engineering, Mobile Computing, Cloud Technologies, and Web & Data Mining for Masters Program in San Jose State University. Additionally, Chandra held market research, strategy, and technology architecture advisory roles in Cisco Systems and Lam Research and performed Principal Investigator role for Valley School of Nursing where he connected Nursing Educators & Students with Virtual Reality technologies. Chandra has functioned as Chair in numerous technology and advanced computing conferences such as IEEE Oxford, UK, IEEE Big Data Services 2017, San Francisco USA, and Future of Information and Communication Conference 2018, Singapore. Chandra graduated from San Jose State University Masters Program, specializing in Software Engineering, and completed his Master of Business Administration from Santa Clara University, Santa Clara, California, USA.

Security System for Big Data Cloud Computing Using Secure Dynamic Bit Standard

Faiz Akram

Abstract Big data (BD) is a high-volume resource that requests savvy and imaginative types of data handling for improved knowledge revelation, dynamic, and procedure streamlining. CC gives a solid, deficiency open-minded, and adaptable condition to the enormous information appropriated the board frameworks. The Secure Dynamic Bit Standard (SDBS) calculation gives the security through two unique keys, for example, the ace key and meeting key created by cloud service providers (CSP). The SDBS calculation contains the three distinctive key lengths, for example, 128 bits, 256 bits, and 512 bits. The length of the ace key and meeting key is arbitrarily created during the encryption procedure. CSP scrambles the ace key with the meeting key and sends the encoded ace key and meeting key to the information suppliers on demand. The information supplier unscrambles the ace key with the meeting key and encodes the information record with the decoded ace key. This strategy will decrease the most extreme number of unauthenticated and unapproved clients in a huge information cloud.

Keywords Cloud computing (CC) · Cloud service provider (CSP) · Secure Dynamic Bit Standard (SDBS)

1 Introduction

Cloud computing (CC) has numerous applications, for example, empowering access to costly applications at no cost, lessening both foundation and running costs of PCs and programming as there is no requirement for any establishment. Clients can put the information at anyplace [1]. All clients are required to connect with a framework, state the Internet. CC began as an instrument for relational figuring, however now it is generally utilized for getting to programming on the web, online capacity without stressing over foundation cost, and preparing power. Associations can offload their information technology (IT) foundation in the cloud and get entrance. Not just private

F. Akram (✉)
Faculty of Computing and Informatics, Jimma Institute of Technology, Jimma University, Jimma, Ethiopia
e-mail: akram.faiz@ju.edu.et; akram.faiz@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_1

associations are moving to distributed computing, yet the legislature is additionally moving a few pieces of its IT framework to the cloud. Huge information includes the advanced information from a few computerized sources which contain sensors, scanners, numerical displays, recordings and cell phones, digitizers, Internet, messages, and interpersonal organizations which are expanding the information rate.

CC and enormous information are conjoined [2, 3]. BD gives clients the capacity to utilize merchandise figuring to process dispersed questions over numerous datasets and return resultant sets in an opportune way. CC gives a class of dispersed information handling stages. Huge information sources from the cloud and Web are put away in a conveyed flaw lenient database and handled through a programming model for huge datasets with an equal disseminated calculation in a group [4]. The multifaceted nature and assortment of information types are handling the capacity to perform an investigation on huge datasets. CC foundation can fill in as a successful stage to address the information stockpiling required to perform an enormous information examination. CC is associated with another example for the arrangement of registering foundation and enormous information handling strategy for a wide range of assets accessible in the cloud through information examination.

A few cloud-based advancements need to adapt to this new condition since managing BD for simultaneous handling has gotten progressively confounded. MapReduce is a genuine case of BD handling in a cloud domain; it permits the preparation of a lot of datasets put away in equal in the group. Group processing displays great execution in conveyed framework situations, for example, PC force, stockpiling, and system correspondences. Moreover, the capacity of bunch figuring is to give a cordial setting to information development. Database management systems (DBMSs) are viewed as a piece of the currently distributed computing engineering and assume a significant job to guarantee the simple change of utilizations from old venture foundations to new cloud framework designs [5]. The weight for associations to rapidly receive and execute advances, for example, CC, to address the test of large information stockpiling and preparing requests involves unforeseen dangers and results [6].

2 Security in BD-CC

Big data cloud computing (BD-CC) transforms into a valuable and standard plan of action in light of its engaging segments. Notwithstanding the current advantages, the past segments also bring about authentic cloud-specific security issues. The all-inclusive community concern is security in the cloud, and the clients are deferring to trade their business to the cloud. Security issues have been the prevention of the improvement and expansive use of CC [7]. Understanding the security and assurance chances in CC, making rich and powerful arrangements are essential for its thriving, regardless of the way that mists empower clients to avoid fire up costs, lessen working expenses, and unite their speed by rapidly getting administrations and infrastructural assets when required.

In publicizing and business, the greater part of the ventures utilizes large information; however, the key properties of security may not be executed [8]. On the off chance that security penetrates happens to BD, it would result in much more genuine lawful repercussions and reputational harm than at present.

In this new time, numerous organizations are utilizing the innovation to store and examine petabytes of information about their organization, business, and their clients. For making BD secure, procedures, for example, encryption, logging, nectar pot recognition must be fundamental. In numerous associations, the organization of BD for misrepresentation location is alluring and helpful. The test of recognizing and forestalling propelled dangers and noxious interlopers must be explained utilizing large information style examination. The difficulties of security in CC conditions can be classified into an organized level, client verification level, information level, and conventional issues [9].

The difficulties can be sorted under a system-level arrangement with organizing conventions and system security, for example, disseminated hubs, and appropriated information, and inter hub correspondence.

The difficulties can be classified under client verification level arrangements with encryption or unscrambling procedures, validation techniques, for example, authoritative rights for hubs, verification of utilizations and hubs, and logging.

The difficulties can be classified under information level arrangements with information honesty and accessibility, for example, information security and disseminated information.

The difficulties that can be ordered under the general level are customary security apparatuses and utilization of various advancements.

3 Security Attacks

Distributed computing relies for the most part upon the structure of the current system, for example, metropolitan area network (MAN), wide area network (WAN), and local area network (LAN). System-level security assaults might be deliberately made by outside clients or by a malevolent insider staying between the client and the cloud service providers (CSP) and endeavoring to encroach upon the information to or from the cloud. This segment will endeavor to focus on the system-level security assaults and their possible countermeasures to ensure genuine information secrecy and dependability [10, 11].

3.1 Space Name System (DNS)

Attacks on the Internet, since reviewing a framework described by the numbers is troublesome, the aggressors are made to do with names. The Internet Protocol plays a remarkable role over the Internet related to the PC. The names of the DNS automatically change concerning the IP addresses utilizing a disseminated database schema. Web DNS servers are dependent upon various types of attacks, for example, space catching, ARP store hurting, and man-in-the-middle attacks. A discussion of these attacks was found underneath.

3.2 Space Hijacking

Domain capturing alludes to changing the name of an area without the data or consent from the area owner or producer. Area seizing engages intruders to get corporate data and play out the unlawful development, for instance, phishing, where a site is superseded by a similar segment that records private data. Another arrangement is using the Extensible Provisioning Protocol (EPP) that is used by various area vaults. EPP utilizes an approval code gave uniquely to the space registrant as a safety effort to envision unapproved names advancing.

3.3 IP Spoofing Attack

The attacker buildups across unapproved admittance to a PC by proposing the PC to have the intention of undergoing through heavy traffic are known as IP mocking. Various attacks are employed by IP caricaturing, for instance, denial of service assault [12].

3.4 Disavowal of Service (DoS) Attacks

The inspiration driving these assaults is making the physical system and PC assets out of reach. During the DoS attack, the attacker submerges the disaster with a broad number of software packages over short intervals. DoS aids to identify distinguishing the devouring data between the systems. The assailant practices a false IP address as the source IP address to hold the interruption of the DoS. Furthermore, it is conceivable to the aggressor to utilize distinctive traded off machines that need to begin at presently seized to attack the misfortune machine in the interim. This kind of interruption handled by the DoS is known as the distributed DoS [13].

3.5 Transmission Control Protocol Synchronize (TCP SYN) Flooding Attack

While possessing situations like the DoS attack, the hackers employ the TCP SYN packages to pervert the machines. These kinds of attacks will certainly damage the restrictions of the three-course handshake when maintaining the half-open affiliations. A man-in-the-middle (MITM) attack is an overall term for when a culprit positions himself in a discussion between a client and an application by either to snoop or to imitate one of the gatherings, causing it to show up as though an ordinary trade of data is in progress. The bundle filtering is a method adopted to reduce the IP ridiculing that has been executed with the assistance of beginning stage confirmation systems, solid encryption, and a firewall [14, 15].

4 Methodology

SDBS have three-piece levels, for example, 128 bits, 256 bits, and 512 bits. At whatever point the information supplier needs to transfer an information record to the cloud, any of the bit levels is chosen haphazardly and it will get changed over into bytes. Because of byte esteem, CSP will produce the ace key and a meeting key utilizing an irregular generator. The ace key is scrambled by the meeting key, and both the encoded ace key and the meeting key will be sent to the information supplier.

The meeting key will be utilized to unscramble the ace key, and the ace key is utilized to scramble the information document. The encoded information document will be transferred to the huge information cloud server alongside the proof of ownership which is created by CSP. On the off chance that an information client needs to download an information document from the large information cloud server, according to popular demand, the CSP will send the encoded ace key and the meeting key alongside the scrambled information record to the information client after one-time password (OTP) check.

At that point, utilizing this meeting key the ace key will be unscrambled, and utilizing the ace key the information record will be decoded and put away in the arrangement of the information client. SDBS calculation is a novel calculation that has three different guidelines with ten rounds for 256-bit keys, eight rounds for 128-bit keys, and 12 rounds for 512-bit keys. The round contains different activities like substitution, adjustment, and change of the info plaintext into the yield figure text (Fig. 1).

The encryption process of SDBS algorithm with the 128-bit standard is shown in Fig. 2.

The plain content called a state cluster 'S' has four lines of bytes. Each line of a state contains N_b quantities of bytes, where N_b fluctuates for these three guidelines.

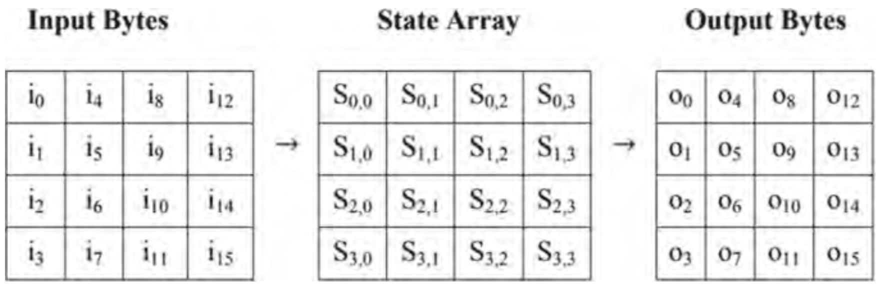


Fig. 1 Input bytes, state array, and output bytes

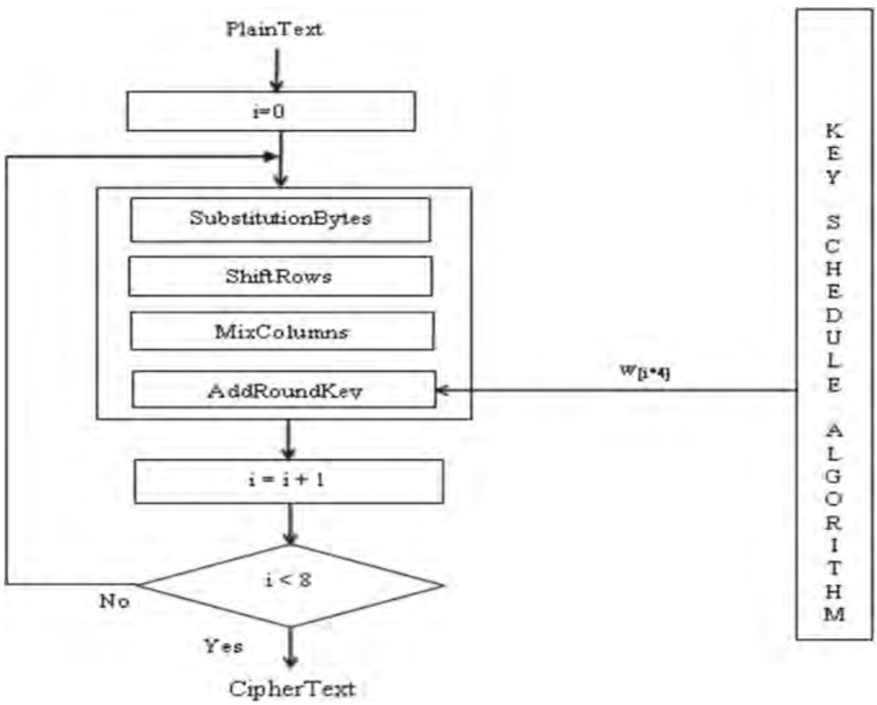


Fig. 2 SDBS encryption process for 128-bits standard

For the 128-bit standard, the estimation of N_b is 4, for the 256-bit standard, the estimation of N_b is 8, and for the 512-bit level, the estimation of N_b is 16.

The variety of information bytes appeared as $i_0, i_1, \ldots, i_{15}$ and the variety of yield bytes is spoken to by $o_0, o_1, \ldots, o_{15}$ as appeared in Fig. 1.

Decoding process: In SDBS unscrambling process, the information supplier must login and afterward select an information record, and the information client needs to download. Before downloading the information document, the information client

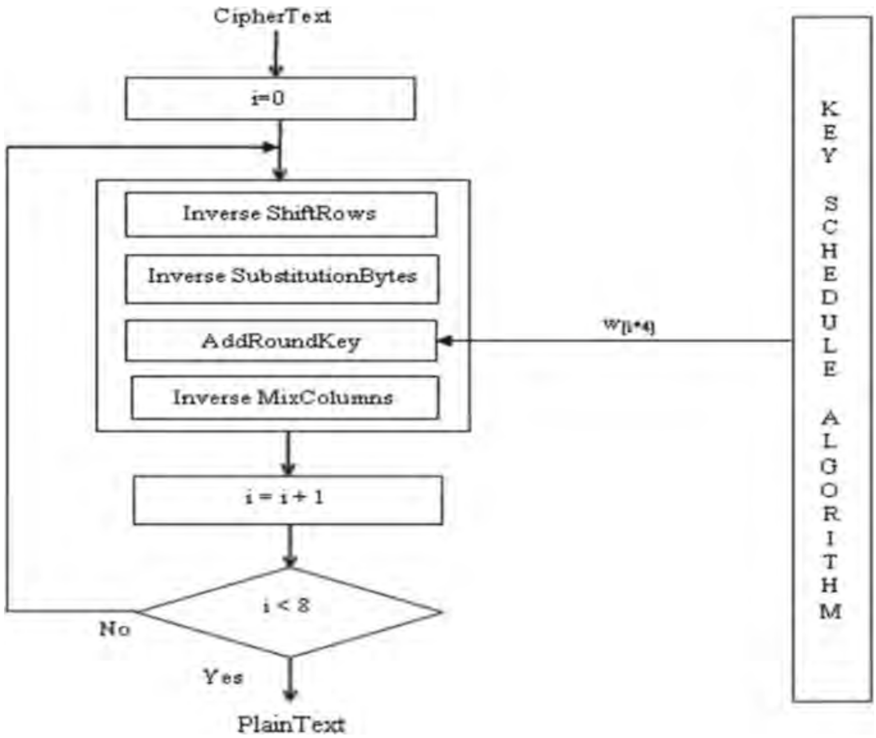


Fig. 3 SDBS decryption process of 128-bits standard

sends the solicitation to the CSP to get the ace key and meeting key. CSP encodes the ace key with the meeting key and sends to the information client.

OTP is additionally sent by the CSP to the information client's email id or versatile no. On the off chance that the entered OTP is substantial, at that point the encoded information document will be downloaded from the cloud, and the decoded ace key with the meeting key is utilized to begin the unscrambling procedure. At long last, the decoded record is put away into the framework (Fig. 3).

5 Results

The proposed method is performed and tried on a workstation with the following hardware specifications like 8 GB RAM, Windows 7 (64-bit) operating system, Intel i7 processor within the cloud storage (dropbox distributed storage) (Table 1).

The exhibition of SDBS calculation could be dissected by two sorts of boundaries which are encryption time and decoding time. The encryption time is measured by the time taken to complete the encryption process along with the record size.

Table 1 Role and operation of multilevel SDBS

Role	Operation
Data Provider (DP)	Encrypt Data File
	Upload Data File
Data User (DU)	Download Data File
	Decrypt Data File
Cloud Service Provider (CSP)	Authenticate—Checking User Name and Password
	Authorize—Checking Credentials of the User
	Key Generation, OTP Generation, PoW Generation
	Block Unauthorized User

And the decryption time is measured by the time taken to complete the decryption process along with the record size. The genuine portrayal of SDBS 128-bits standard encryption time and unscrambling time-dependent on record size is spoken to underneath.

Figure 4 portrays the presentation of encryption time and unscrambling time versus record size in conspire 4. The diagram is completely based on the encryption time and document size. The encryption or unscrambling time of the proposed scheme 4 is less time when contrasted and different methods conspire 1, plot 2, and plan 3. This strategy took 2.56 ms for encoding 1 GB information document, likewise the 24 GB information record took 12.1 ms for encryption. From Fig. 4, 15, 1 GB information record took 2.35 ms for decoding; additionally, the 24 GB information document took 11.72 ms for unscrambling.

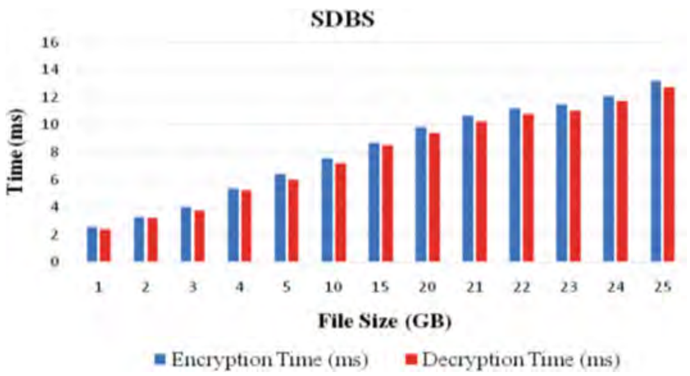


Fig. 4 Performance of SDBS 128-bits standard

6 Conclusion

The proposition gave a point by point prolog to the security framework in big data cloud computing. Big data distributed storage limits and applications are clarified. The conceivable outcomes of various assaults are portrayed in detail. In this article, information moved to the big data cloud has been finished by the information supplier. The proposed framework has high information respectability and information stockpiling without information misfortune. Before the information has been transferred into the capacity region, a high secure calculation called Secure Dynamic Bit Standard is been utilized. Big data investigation report assists by distinguishing each datum supplier and information client use of document size, encryption time or decoding time, and transfer time or download time.

References

1. Nahar AK, Mongia K, Kumari S (2018) Cloud computing and cloud security. Int J Res Adv Eng Technol 4(1):1–8
2. AshwinDhivakar MR, Ravichandran D, Dakha V (2015) Security and data compression in cloud computing using BlobSeer technique. In: National conference on cloud computing and big data, vol 1(12), pp 201–203
3. Ateniese G, Fu K, Green M, Hohenberger S (2006) Improved proxy re-encryption schemes with applications to secure distributed storage. ACM Trans Inf Syst Secur 9(1):1–30
4. Attrapadung N, Herranz J, Laguillaumie F, Libert B, De Panafieu E, Ràfols C (2012) Attribute-based encryption schemes with constant size cipher texts. Theoret Comput Sci 422:15–38
5. Awodele O, Izang AA, Kuyoro SO, Osisanwo FY (2016) Big data and cloud computing issues. Int J Comput Appl 133(12):35–47
6. Bachhav S, Chaudhari C, Shinde N, Kaloge P (2016) Secure multi cloud data sharing using key aggregate cryptosystem for scalable data sharing. Int J Comput Sci Inf Technol 3(1):19–27
7. Balasubramanian N, Balasubramanian A, Venkataramani A (2009) Energy consumption in mobile phones: a measurement study and implications for network applications. In: Proceedings of the 9th ACM SIGCOMM conference on internet measurement conference, vol 1(5), pp 280–293
8. Bavisi S (2018) Computer and information security handbook. Morgan Kaufmann Publication, Elsevier Inc., pp 375–341
9. Bhadauria R, Chaki R, Chaki N, Sanyal S (2011) A Survey on security issues in cloud computing. IEEE Commun Surv Tutor 3(16):1–15
10. Bisong A, Rahman M (2011) An overview of the security concerns in enterprise cloud computing. Int J Netw Secur Appl 3(1):30–45
11. Hemalatha M (2012) Cloud computing for academic environment. Int J InfCommunTechnol Res 97–101
12. Mathew S (2012) Implementation of cloud computing in education—a revolution. Int J Comput Theory Eng 473–475
13. Kaur M, Singh H (2015) A review of cloud computing security issues. Int J AdvEngTechnol (IJAET), pp 397–403
14. Gaikwad BP (2014) A critical review on risk of cloud computing in commercial. Int J Innov Res Comput CommunEng 1–8
15. Ahmed ES, Saeed R (2014) A survey of big data cloud computing security. Int J Comput Sci SoftwEng (IJCSSE), pp 78–85

Distributed DBSCAN Protocol for Energy Saving in IoT Networks

Mazin Kadhum Hameed and Ali Kadhum Idrees

Abstract Sensor networks form a crucial topic in research, as it seems to target a huge variety of uses in which it could be applied, such as health care, smart cities, environment monitoring, military, industrial automation, and smart grids. The clustering algorithms represent an essential factor in conserving power within energy-constrained networks. The selection of a cluster head balances the energy load within the network in a proper way, eventually contributing to the reduction of energy consumed, as well as the enhancement of network lifespan. This article introduced a distributed DBSCAN protocol for saving the energy of sensor devices in IoT networks. This protocol is implemented on each IoT sensor device, and the devices apply the density-based spatial clustering of applications with noise (DBSCAN) algorithm to partition the network into clusters in a distributed way. The efficient periodic cluster head strategy is proposed based on certain criteria like remaining energy, number of neighbors, and the distance for each node in the cluster. The cluster head will be chosen in a periodic and distributed way to consume the power in a balanced way in the IoT sensor devices inside each cluster. The comparison results confirm that our protocol can conserve power and enhance the power conservation of the network better than other approaches.

Keywords Sensor networks · Density-based spatial clustering of applications with noise (DBSCAN) clustering · Wireless sensor network (WSN) · Low-energy adaptive clustering hierarchy (LEACH) · Internet of Things (IoT)

M. K. Hameed
Department of Software, University of Babylon, Babylon, Iraq
e-mail: it.mazen.kadhum@uobabylon.edu.iq

A. K. Idrees (✉)
Department of Computer Science, University of Babylon, Babylon, Iraq
e-mail: ali.idrees@uobabylon.edu.iq

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_2

1 Introduction

Wireless sensor networks (WSNs) have recently gained significant attention for their implications found in various fields including ecosystem monitoring, health care, environment assessment, urban areas applications, control maintenance, and target tracking [1]. The connection of all things that the Internet can monitor or control could be defined as Internet of Things (IoT), which is most preferably achieved through a wireless medium [2, 3]. The network of wireless sensors could be depicted as the set of huge sensor nodes used over a wide area for sensing and accumulating different data from the systems and environment, to be applied in a variety of uses like weather monitors, animal tracking, disaster managing, and bio-medical applications, within IoT [4, 5]. Wireless sensors could be of use with IoT applications for gathering and processing data with the extraction of valuable information to be communicated to the end user, as it can occasionally be unreachable by individuals. Therefore, WSNs are among the integrated parts of IoT applications [6]. Direct communication is made by each node over the BS as the data is distributed [4]. Through the continuity of data transmitted, the furthest node would be more likely to die earlier than others through its energy loss [7]. Consequently, the clustering process applied tends to collect nodes, forming a set of clusters for solving the problem [2]. Its main performance is remarkably improved through clustering several nodes [8]. Besides, the network keeps the demands of the central organization to a minimum, inspiring the local decision-making to enhance the scalability. The clustering procedure tends to collect data through the active network. As well, a suitable cluster head needs to be selected for every data retrieving clusters [9], through sensor nodes to be passed to the BS [10]. The SN eventually creates clusters for the monitoring procedure and the constitution of both cluster member (CM) and cluster head (CH) [7]. The SNs contain battery sources for their performance, which makes them a tool of power starving. The main elements that influence the energy dissipating of WNS are the distance with the sink, the remaining energy of the node, and intra-cluster distance [11]. The remained power, distance among the core nodes in the cluster, and the number of members in each core node are the three factors that the CH choosing is based on in our work. Thus, at the same time, the elected CH must have the highest remained power, maximum members, and finally the lowest distance to all core nodes. The power can be preserved to extend the life of the sensor device if these factors considering, and the simulations show the good results.

The contributions of the proposed work are based on the distributed clustering-based DBSCAN protocol which is for increasing the lifespan of wireless sensors of IoT networks. This protocol is distributed on every IoT sensor device, and the sensor devices are combined with the DBSCAN algorithm to form several groups in the network area of interest. An efficient periodic CH approach is suggested based on several criteria like remaining power, neighbors' number, and the distance for each node in the group. The cluster head will be elected periodically to the consumed power which is balanced in the wireless sensors inside each group. The proposed protocol is evaluated and compared to two existing methods such as I-LEACH [11] and low-energy adaptive clustering hierarchy (LEACH) that are presented in [12] in light of several metrics like the resting energy, network lifetime, and CH count, etc. The comparison of simulation results illustrates that our protocol can preserve energy and increase the network lifetime better than other approaches.

The remaining of the article is structured as follows. The related literature is presented in Sect. 2. The DBSCAN traditional algorithm is explained in more detail in Sect. 3. Section 4 introduces the proposed energy-efficient distributed clustering-based DBSCAN protocol for conserving the power of wireless sensors of IoT networks. Results and discussion are explained in Sect. 5. Section 6 presents the conclusion and the planned work for the future.

2 Related Work

The cluster head (CH) has been selected randomly resulting in a similar likeliness for both nodes with high or low energy to become a CH. Whenever a low energy node is elected to be a CH, it is most likely to die soon, with its eventual effect on how robust the network is. Besides, every round varies in how many CHs it has, as well as their location [13]. As with WSNs, the main focus lies on two essential factors: reducing the consumed power, as well as extending the network lifespan. Taking LEACH protocol to be the basic algorithm, several alterations are made in light of differing applications. LEACH and its related researches have been presented [14], taking into account several significant parameters including the clustering method, data aggregating, scalability, and mobility type. This protocol makes a random selection of CHs without the BS knowing any details on the network's remaining energy [15]. Therefore, the LEACH-C protocol has been suggested for addressing this issue [12].

The PSO-ECHSs are dealt with as the Particle Swarm Optimization (PSO) based CHs are selected with the use of factors such as distance among nodes, remaining energy, and distance to BS [16]. An alternative optimizing method known as grouped gray wolf search optimization has been applied in [17] selecting security-aware CH, to improve the network lifespan choice. The researchers proposed an alternative algorithm which first calculates the ideal cluster number, taking into account its location adaptability and data aggregating rate. Next, a new parameter is presented in light of the residual, initial, and average energy consumption, as well as the node degree for selecting the CH. A third aspect is the proposal of an unevenly self-adaptive clustering algorithm that considers the node degree in solving the "hot spot" problem. At last, a solution is suggested for the "isolated nodes problem" [18].

Jan et al. [19] present a new method known as a mutual authentication approach based on payload, as it consists of two steps selecting nodes optimality which act as CH, being allowed to have communication with its neighbors, and the authentication of every CH of its near nodes for forming clusters. After the former step, a method of authenticating takes place which depends on tokens. The tokens are used in the correlation between the CH and the acknowledge messages it corresponds with. Authenticating the payload cluster contributes for forming clusters from close member nodes and the CH. A comparison between the scheme of each the LEACH-C and Sec LEACH is suggested. The suggested model shows a lack in the use of an encryption method, as well as the improvement of its performance and comparing it to modern clustering models with a random distribution. Purohit and Bhargava introduced the multi-hop routing scheme [20], where the one-hop transmission is transformed into multi-hop way, to reduce the consumed energy by a sensor. This helps in obtaining the effective use of energy. The experiment resulted in the improvement of performance regarding the time for the first node to die (FND). This had a clear influence on improving network energy effectively. The main aim of this proposed idea is inter-cluster communicating; the time needed to receive messages is rather higher, negatively affecting its work. The main use of WSNs is receiving information for doing several performances, mainly in light of the data received. The decrease of messages counts automatically displays the inactive nodes, eventually declining the network's general performance. The work in [11] is introduced an improved method of the LEACH named I-LEACH. It is limited the selection of the cluster head using a certain threshold with concurrently changing the level of power between the nodes. The results explain better performance with the original protocol.

Despite many clustering approaches were proposed for grouping the sensor nodes in the WSNs, but none of them can ensure an optimal energy saving, and this would result in a shorter lifespan of the WSN. This paper suggested a distributed DBSCAN protocol for preserving the power of sensor devices in IoT networks. This protocol is executed at each IoT wireless sensor, and the devices apply the DBSCAN scheme to divide the network into groups in a distributed way. An efficient cluster head strategy is proposed based on certain criteria like remaining energy, number of neighbors, and the distance for each node in the cluster. The cluster head will be selected in a periodic and distributed way to conserve the power in a balanced way in the IoT wireless sensors inside each cluster.

3 The DBSCAN Algorithm

The DBSCAN approach identifies clusters within huge spatial datasets by taking into account the local density of its elements, with the use of a single input parameter. Also, a suitable parameter value is suggested for the user, so little knowledge about the domain itself is needed. The aim of DBSCAN is categorizing them into clusters apart, eventually defining the differing classifications [21].

In traditional DBSCAN, the user requires two parameters to be defined that are the neighborhood range and MinPts refer to the lower required number of points to construct a new cluster. In the beginning, the scheme chooses one-point P randomly and then calculate the distance among this selected point and the rest point in the dataset. The neighborhood condition between the point P and any other point in the dataset if the distance between them is less or equal to e. If the number of points which are in the neighborhood range of point P greater than or equal to MinPts, then the new cluster will be constructed; otherwise, these points as noise are labeled.

This means that noise points can later be within other clusters if they satisfy the condition of the required MinPts in the neighborhood range of the newly selected point. After that, the DBSCAN scheme will check if it is possible to extend this cluster or it chooses another point from the outside of the current cluster. The checking is done by verifying both MinPts and distance conditions if they are satisfied for each point in the range of the cluster. If these conditions are satisfied, then the DBSCAN scheme extends this cluster to each point in the neighborhood range of point P. The extension of the cluster will be stopped, and each point will be labeled as a visited point if the cluster expanded to the required the MinPts. The DBSCAN scheme then chooses another random not visited point from the dataset and repeats the same above scheme. The DBSCAN scheme will be stopped if there is no point labeled as not visited. Algorithm 1 illustrates the DBSCAN scheme with its expanding function. The time requirement of the DBSCAN scheme is O(n2), where n refers to the size of the dataset. The time complexity will be decreased to the O(nlogn) if the spatial indexing is utilized [22, 23].

4 Energy-Efficient Distributed DBSCAN Protocol

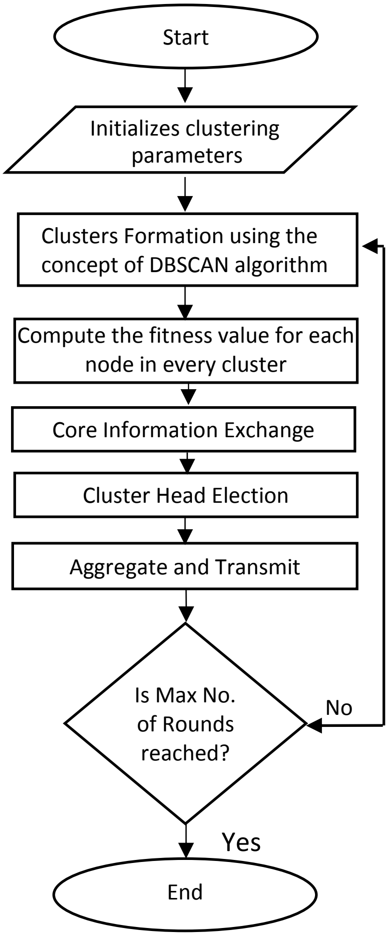
This research aims at proposing a distributed DBSCAN protocol for maximizing the wireless sensor's lifetime. This protocol is distributed at every sensor device deployed in the monitored area. The proposed protocol involving two steps: setup and steady-state. For the sake of simplicity, the proposed distributed DBSCAN protocol is named as DBSCAN protocol in this paper. The DBSCAN protocol is presented in Fig. 1.

In the setup phase, when the sensor devices are deployed in the working area, it is supposed that every sensor device knows its location. According to the DBSCAN algorithm, each sensor implements the algorithm of DBSCAN as follows:

(a) Each sensor node will perform the same test whether it is core point or not, according to the principle of the DBSCAN algorithm, as it scans the surrounding area to find out the number of sensors that are within the sensing range and that must be greater or equal to a specific parameter.
(b) As for the sensor nodes that are within the sensing range for core point, it will be its member.
(c) If this core point does not belong to any cluster, then it forms a new cluster; otherwise, it remains with the same cluster.
(d) The core point sends a message to all its members to be included in the same cluster.
(e) Repeat steps 1–5 until all sensor nodes are passed.

Algorithm 1 explains the distributed DBSCAN algorithm that will be executed in every sensor node *sj*.

Fig. 1 DBSCAN protocol



Algorithm 1. Distributed DBSCAN (sj)
Input: N: number of neighbor nodes, Sr: sensing range, minNodes: minimum number of nodes to create cluster.
Output: sj.rejon: the cluster number for node sj.

```
while  REj ≥ Ethr do
    If  sj  Receive MemberPacket from si  then
        Mark sj  as member to the Core si ;
        Update REj;
  end
        sj.rejon ← 0;
        for each node si in N do  // i ∈ N and i ≠j
        nbrNodes ← nbrNodes + CORE Objective Function (sj, si, Sr);
        if  CORE Objective Function return 1 then
                Send MemberPacket to the sensor node i;
                Update REj;
        end
        if  nbrNodes ≥ minNodes then
            save the information
            if (((sj.rejon = 0) Or (sj.rejon ≠ 0)) and (r==0)) then
             sj.rejon ← sj.rejon +1;
                Call Cluster(sj);
            end
            else if ((sj.rejon = 0) Or (r ≠ 0)) then
                sj.rejon ← sj.rejon +1;
                    Call Cluster1(sj);
             end
              else if ((sj.rejon ≠ 0) Or (r ≠ 0)) then
                    Call Cluster2(sj);
              end
            end
         end for
     end while
returen sj.rejon;
```

CORE objective function return 1 and $r = 0$ if the sensor node i is within the sensing range Sr, and it is not a member in other clusters. Otherwise, CORE objective function returns 0 and $r = 1$. The function cluster put any neighbor node within the sensing range of *sj* in the same cluster and *sj* send MemberPacket to the sensor node *i* to 4 cluster of *sj*. The function cluster1 places any neighbor node within the sensing range of *sj*, and it has not been assigned to any cluster in the same cluster of sensor node *j*. The function cluster2 places any neighbor node within the sensing range of *sj*, and it has not been assigned to any cluster (or it is assigned to the cluster of sensor node *j*) in the same cluster of sensor node *j*. After achieving the functions cluster, cluster1, and cluster2, the remaining energy of the sensor node *j* will be updated due to sending a MemberPacket to the sensor node *i* to inform it that it becomes a member in the same cluster of *sj*.

After the stage of creating clusters, the exchange of information between the core points (nodes) is done inside the single cluster, where each core point sends a message to all the core points inside the cluster; it contains all the necessary information inside like remained power, status, location, number of members, total wireless sensors number in the group (cluster), etc. Every core sensor node inside each cluster will include the information of other core nodes in the same cluster; therefore, every core node in the same cluster will execute Eq. (1) for the information of each member inside the core node. The core node that gives the better value of Eq. (1) will be selected as a cluster head in the current cluster for this round. All the core nodes inside the cluster will achieve the same computation and will give the same results for the winner core node. This will be implemented in a distributed way, and every core node will know if it is a cluster head or not.

$$\mathrm{FitVal}_j = \frac{E_{\mathrm{remaining}}}{E_{\mathrm{initial}}} + \left(1 - \sum_{j \in N} \left| S_j(x, y) - S_i(x, y) \right| \right) + \frac{S_j(\text{ Members})}{\text{Cluster(Members)}} \tag{1}$$

where $E_{\mathrm{remaining}}$ is the remaining power of the wireless node j; E_{initial} is the initial energy value of sensor node j; N is the number of core nodes in the current cluster; S_j (x, y) and S_i (x, y) refer to the positions of core nodes S_j and S_i, respectively. S_j (Members) refers to the number of nodes member of core node j; cluster (Members) refers to the total number of nodes in the cluster.

In the steady phase, after clusters formation and fixing the TDMA schedule, the process of transmitting data may start. With the assumption that all nodes contain data that requires to be sent to the CH, this sending will occur within its allocated time. The cluster head in its turn allocates a TDMA schedule to the actively participating members so that the data transmitting process is managed, and the consumed power is limited to a minimum. Based on the active/idle status, the data is transmitted during the steady-state stage with regards to the timespan assigned for every member node. The power supply of any idle node will be turned off but the CH, which awaits the data from the member nodes. After delivering data through nodes, the CH initiates a data aggregating process followed by transmitting this data to the BS. The algorithm eventually returns to the setup iterating stage to select a different CHs group, followed by the steady stage, and so on.

5 Performance Evaluation, Analysis, and Discussion

This section focuses on evaluating the proposed distributed DBSCAN protocol using different performance metrics like cluster count, remaining energy, dead nodes number, packets number transmitted to cluster head, number of packets sent to cluster sink, and network lifetime. The conduction of the simulation results is performed

Table 1 Stimulation parameters

Symbol	Description	Value
$X1_m$	Distance at X-axis	400 m
$Y1_m$	Distance at Y-axis	400 m
Ns	WSN size	100 nodes
P_{Tx}, P_{Rx}	The initial energy	0.5 J
E_{mp}	Energy consumption for receiving	0.0013/pJ/bit/m^4
E_{fs}	Energy dispersion: free space model	10/pJ/bit/m^2
E_{amp}	Energy dispersion: power amplifier	100/pJ/bit/m^2
E_{DA}	Energy consumption for collection	5/nJ/bit
d_0	Reference distance	87 m
I	Packet size	4000 bits

using a C++ custom simulator for 2500 iterating round so that several plots are obtained. The sensor nodes are deployed in the monitored area randomly. The location of the sink is in the center of the monitored area; with no limitation on energy, normal nodes would have its limitations in terms of energy, memory, and processing capabilities. The suggested protocol is applied for generating the results with regards to the parameters referred to in Table 1.

As for this work, the packet size tends to be relatively larger, namely be 4000 bits. The proposed protocol in this paper is named as DBSCAN. The DBSCAN protocol is applied to the same energy consumption model that is employed in [11]. An obvious result of stimulation is that the DBSCAN outperforms the I-LEACH [11] and LEACH [12] protocols in light of several performance metrics. The number of CHs is found to extend up to 2000 rounds for the DBSCAN, whereas it reaches only 1750 and 850 rounds for I-LEACH and LEACH, respectively. Figure 2 shows the cluster count for the LEACH, I-LEACH, and DBSCAN.

Similarly, the simulation results in Fig. 3a present the fact that at just "500 rounds," the average power of LEACH reaches 0, while I-LEACH goes on to ~"1250 rounds." The DBSCAN continues to 1750. The amount of data packets transmitted to BS within LEACH and I-LEACH reaches maximally (0.5 × 104) and (1.75 × 104), respectively. The rise in rounds leads to the energy depletion of sensor nodes until they terminate.

Figure 4 illustrates the network lifetime through the representation of dead nodes. After 750 rounds, the number of nodes alive levels out at 0 for LEACH, whereas a few nodes remain active till 1500 rounds with I-LEACH. The DBSCAN, on the other hand, has several nodes that remain active till 2000 rounds.

As for the DBSCAN, the value may reach (2×10^4), as is shown in Fig. 5a. A similar increase in data packets transmitted to cluster head (CH) is noticed in Fig. 5b, proving the effectiveness of the suggested protocol. This illustrates the

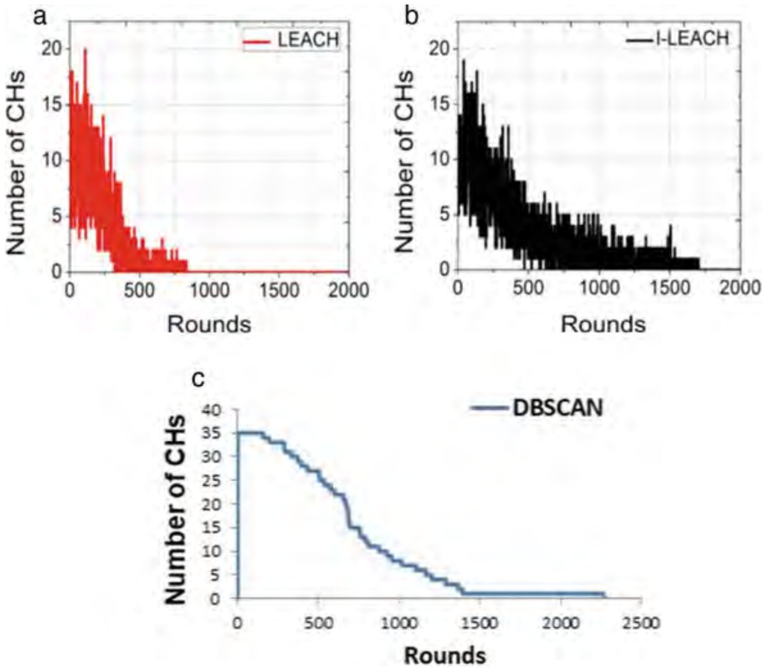


Fig. 2 CH count **a** LEACH, **b** I-LEACH, **c** DBSCAN

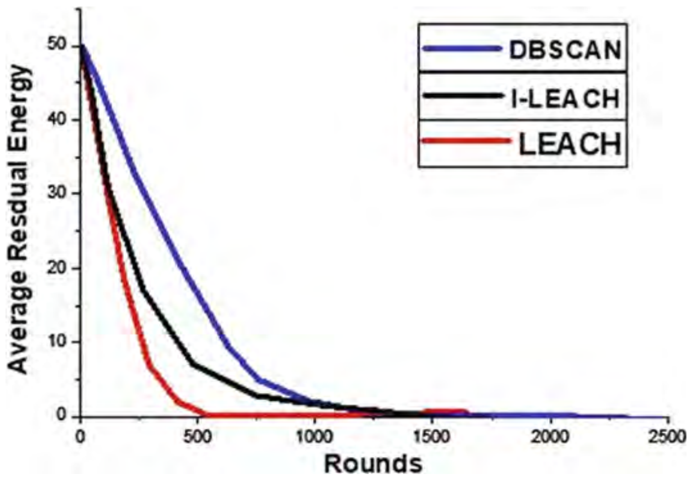


Fig. 3 Network performance **a** average residual energy

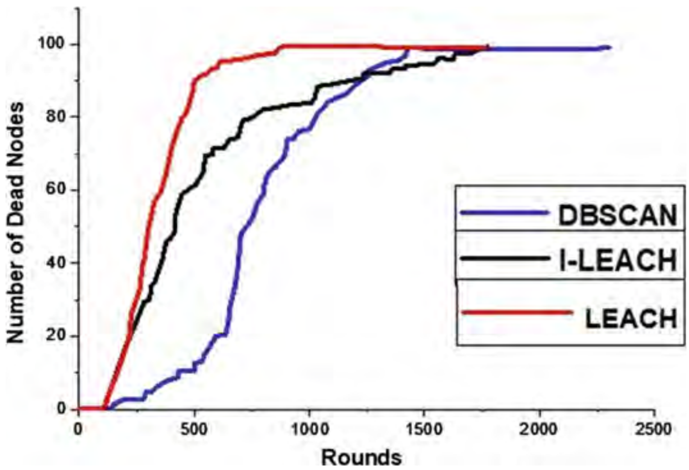


Fig. 4 Lifetime metrics: Dead nodes

whole situation for the maximization of the network lifespan, merely due to assigning various energy levels for differing communicating modes in the network.

The behavior of the algorithm is different in comparison of lifespan metrics with regards to the "first node dead (FND)" and "last node dead (LND)," as illustrated in Table 2. Simulation is performed with regards to three different areas (100, 200, and 400) m^2 with a network that is poorly to richly deployed with sensors. Keeping the initiated energy at (0.5 J) on a poorly deployed area of 100 m^2 improves the network timespan to 1.16, 1.36, and 1.3 for I-LEACH, LEACH, and CPCHSA, respectively. There is a positive relationship between the lifespan and number of nodes (keeping the area and energy constant), as it increases to 1.04, 1.87, and 1.19 times the value for I-LEACH, EECS, and LEACH, respectively. Doubling the area to 100 nodes results in a lifespan of 1.16, 1.3, and 1.47 times, the value for I-LEACH, ModLeach, and LEACH, respectively. At 400 nodes, the lifespan is 1.23, 1.4, and 1.61 times the value for I-LEACH, ModLeach, and LEACH subsequently. With the initiated energy of 1 J for 1000 nodes, the lifespan is 1.4, 1.4, and 3.5 times the value for I-LEACH, LEACH, and EECS. Comparing the DBSCAN to I-LEACH and LEACH over a wider area of 400 m^2 shows the increase of the stability period by (1.23, 2.6) and (1.28, 1.7), and (1.58, 1.7) times for 100, 400, and 1000 nodes, respectively. One can therefore draw the conclusion that the suggested protocol proves a more favorable performance with both smaller and larger areas, regardless of whether the networks were defectively or completely covered with nodes.

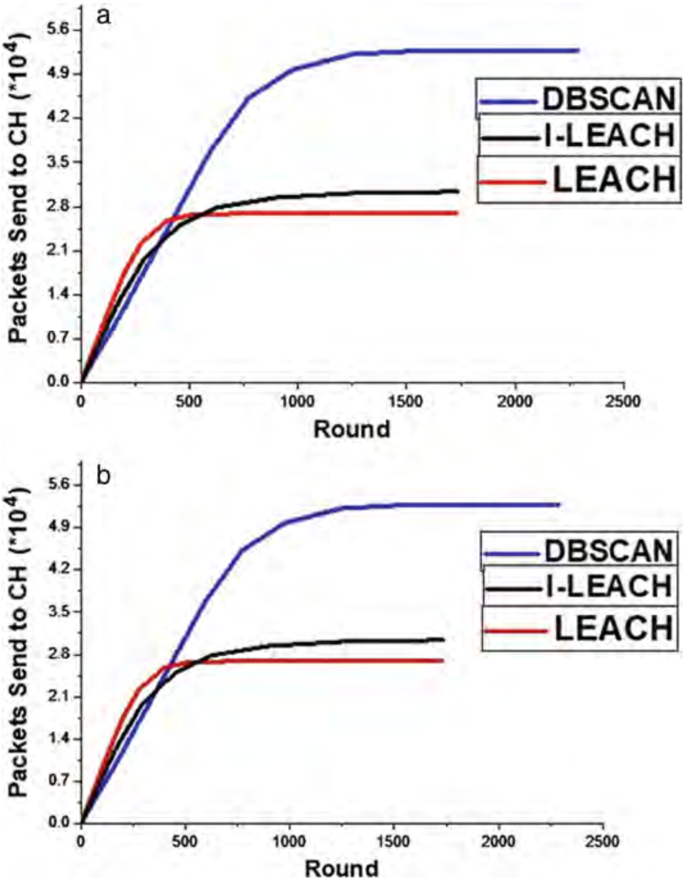


Fig. 5 Sent packets **a** to BS, **b** to CH

6 Conclusion

The clustering algorithms still represent a significant part in the field of wireless sensor network, and it gets a great consideration by many researchers in the current world. This article suggested a distributed density-based spatial clustering of applications with noise protocol for extending the lifetime of wireless sensors of Internet of Things networks. This protocol is distributed on every Internet of Things sensor device, and the sensor devices are cooperated based on the density-based spatial clustering of applications with noise algorithm to form the clusters in the network in a distributed way. The efficient periodic cluster head strategy is introduced based on certain criteria like remaining power, a number of neighbors, and the distance for each node in the cluster. The cluster head will be elected periodically and in a distributed way so as to the consumed power is balanced in the sensors inside each cluster. The

Table 2 Different scenarios for the network lifetime

Nodes	Energy	Area	LEACH		ILEACH		DBSCAN		Other protocols	
			FDN	LDN	FDN	LDN	FD	LD	FDN	LDN
100	0.5	100	980	1450	1050	1700	1150	1972	600 [45]	1500 [45]
		200	780	1150	850	1450	676	1685	200 [46]	1300 [46]
		400	100	800	98	1700	194	2092	–	–
400	0.5	100	1000	1500	1100	1700	1171	1778	820 [47]	950 [47]
		200	850	1300	900	1700	681	2093	190 [46]	1500 [46]
		400	100	1200	100	1600	316	2049	–	–
1000	1	100	2000	2700	2000	2700	2005	4095	–	–
		200	1700	2600	1700	2650	1121	3651	810 [47]	1050 [47]
		400	300	2500	300	2700	842	4284	–	–

proposed protocol is evaluated and compared to two existing methods using several performance metrics like the resting energy, network lifetime, and cluster head count, etc. The comparison results show that the suggested protocol can preserve power and improve the network lifetime better than other approaches.

References

1. Alhussaini R, Idrees AK, Salman MA (2018) Data transmission protocol for reducing the energy consumption in wireless sensor networks. In: New trends in information and communications technology applications, pp 35–49 (2018)
2. Idrees AK, Deschinkel K, Salomon M, Couturier R (2017) Multiround distributed lifetime coverage optimization protocol in wireless sensor networks. J Supercomput 74(5):1949–1972
3. Raj JS (2019) QoS optimization of energy efficient routing in IoT wireless sensor networks. J ISMAC 1(01):12–23
4. Idrees AK, Harb H, Jaber A, Zahwe O, Taam MA (2017) Adaptive distributed energy-saving data gathering technique for wireless sensor networks. In: 2017 IEEE 13th ınternational conference on wireless and mobile computing, networking and communications (WiMob)
5. Duraipandian M, Vinothkanna R (2019) Cloud based Internet of Things for smart connected objects.J ISMAC 1(02):111–119
6. Harb H, Idrees AK, Jaber A, Makhoul A, Zahwe O, Taam MA (2018) Wireless sensor networks: a big data source in ınternet of things. Int J Sens Wireless Commun Control 7(2)
7. Chamam A, Pierre S (2016) A distributed energy-efficient clustering protocol for wireless sensor networks. Comput Electr Eng 36(2):303–312
8. Fotouhi H, Alves M, Zamalloa MZ, Koubaa A (2014) Reliable and fast hand-offs in low-power wireless networks. IEEE Trans Mob Comput 13(11):2620–2633
9. Ranjan NM, Prasad RS (2018) LFNN: Lion fuzzy neural network-based evolutionary model for text classification using context and sense based features. Appl Soft Comput 71:994–1008

10. Geeta D, Nalini N, Biradar RC (2013) Fault tolerance in wireless sensor network using hand-off and dynamic power adjustment approach. J Netw Comput Appl 36(4):1174–1185
11. Behera TM, Samal UC, Mohapatra SK (2018) Energy-efficient modified LEACH protocol for IoT application. IET Wireless Sens Syst 8(5):223–228
12. Heinzelman W, Chandrakasan A, Balakrishnan H (2002) An application-specific protocol architecture for wireless microsensor networks. IEEE Trans Wireless Commun 1(4):660–670
13. Singh SK, Kumar P, Singh JP (2017) A Survey on successors of LEACH protocol. IEEE Access 5:4298–4328
14. Mahapatra RP, Yadav RK (2015) Descendant of LEACH based routing protocols in wireless sensor networks. Procedia Comput Sci 57:1005–1014
15. Heinzelman WR, Chandrakasan A, Balakrishnan H (2000) Energy-efficient communication protocol for wireless microsensor networks. In: Proceedings of the 33rd annual Hawaii International conference on system sciences, Maui, HI, USA, vol 2, p 10
16. Rao PCS, Jana PK, Banka H (2016) A particle swarm optimization based energy efficient cluster head selection algorithm for wireless sensor networks. Wireless Netw 23(7):2005–2020
17. Shankar A, Jaisankar N, Khan MS, Patan R, Balamurugan B (2019) Hybrid model for security-aware cluster head selection in wireless sensor networks. IET Wireless Sensor Syst 9(2):68–76
18. Li J, Jiang X, Lu I-T (2014) Energy balance routing algorithm based on virtual MIMO scheme for wireless sensor networks. J Sens 2014:1–7
19. Jan MA, Nanda P, Usman M, He X (2017) PAWN: a payload-based mutual authentication scheme for wireless sensor networks. Concurr Comput Pract Exp 29(17)
20. Purohit R, Bhargava D (2017) An illustration to secured way of data mining using privacy preserving data mining. J Stat Manag Syst 20(4):637–645
21. Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of 1996 ınternational conference on knowledg discovery and data mining (KDD '96), pp 226–231
22. Han D, Agrawal A, Liao W-K, Choudhary A (2016) A novel scalable DBSCAN algorithm with spark. In: IEEE ınternational parallel and distributed processing symposium workshops (IPDPSW)
23. Idrees AK, Al-Yaseen WL, Taam MA, Zahwe O (2018) Distributed data aggregation based modified k-means technique for energy conservation in periodic wireless sensor networks. In: 2018 IEEE middle east and north africa communications conference (MENACOMM)

Hardware Implementation of Automatic Power Source Controller Cum Power Optimizer

A. Kunaraj, J. Joy Mathavan, and K. G. D. R. Jayasekara

Abstract A device that can switch from one source of power to the other or add on two power sources based on the real-time energy requirement. Power can be generated by various sources. Solar, wind, diesel generator, and electric board main grid supply are the four sources of energy tested here. The number of sources can be more based on user requirements. If the demand exceeds the supply, the device is programmed through AT Mega 2560 microcontroller to add more than one power source to meet the excess demand. With the help of electrical parameters, the power optimizer automatically switches and controls the power source based on the power consumption. If the demand exceeds the total supply, the device is programmed to stop certain devices until the consumption becomes normal.

Keywords Power optimizer · Power source controller · Solar power · CEB grid

1 Introduction

In the developed world, people use innovative ideas to invent new electrical/electronic appliances to do work more efficiently and effectively. The demand for electricity keeps on increasing when everyone starts using electrical appliances. The objective of this research work is to propose an automatic power controller cum power optimizer to switch the power from one source to another or to add up more than one power source based on demand for power. A power generation system should be able to produce enough power using various sources. This device is designed to go for the energy produced by cheap and available sources first, and if the demand exceeds the supply, the device will switch to the next available power source. It can be

A. Kunaraj · J. Joy Mathavan (✉) · K. G. D. R. Jayasekara
Faculty of Technology, University of Jaffna, Jaffna, Sri Lanka
e-mail: joymathavan1991@gmail.com

A. Kunaraj
e-mail: kunaraj12@gmail.com

K. G. D. R. Jayasekara
e-mail: darshanaruma43@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_3

used in situations where the electrical energy consumption is high and fluctuating in an unpredictable manner. Mainly, four power systems can be connected with this proposed device. The power optimizer can measure the power requirement of the consumer and take a suitable real-time decision automatically to process and switch to the next available power source. In most of countries, the energy produced by private companies or by individuals, if it is excess, can be supplied to the main grid line of government, based on the energy requirement. The sources of power generation may be costly or cheap depending on the availability of those sources. If a cheap and available energy resource are used to produce maximum power, a large amount of money could be saved. Solar and wind power are cheap renewable energy sources. In contrast, coal and diesel are non-renewable sources and they harm the environment as well.

Based on the power requirement of the industries and domestic users, there is no proper power optimizing devices for power management. People usually use the main grid line to their electrical energy requirements. When there is a power cut or low voltage supply, they use solar energy, if they have it. If there are only two power sources, it can be switched at least manually. But when the number of power sources increases, it is very difficult to manage the switching between power sources. This project considers a new device to control four power sources manually and automatically. The automatic mode operated based on the power requirement of the industries as well as domestic purposes. In case if the power consumption is high due to the heavy use of electrical appliances, it can be controlled using the wireless control method. The control method can either be switching to the next available power source or turning off certain devices for a limited period.

Solar energy, which is obtained from the sun, can be converted directly to solar energy using photovoltaic solar cells [1]. Ricardo Orduz et al. discussed the development of the PV cell and maximum power point tracking (MPPT) system. Generally, the solar panel combination has one MPPT; but in this work, each PV panel has an individual MPPT module and all the MPPT modules connected [2].

Wind energy is also a readily available and eco-friendly source of energy [3]. The usage of wind energy to produce electrical energy is first implemented in America in 1887 [4]. Various developments in the selection of composite materials for wind turbine blades are still in progress [5]. Sara Mac Alpine et al. discussed mitigating the losses related to non-uniform operating conditions in grid-tied photovoltaic arrays [6]. Elkamouny et al. mentioned in their research work about the recent developments in the technologies associated with capturing solar energy through solar cells [7]. Sanz et al. researched DC-DC converter PV system architecture and they have mentioned that almost all the mismatching losses between modules are eliminated in their outcome, and the energy output has increased [8]. Muthamizhan mentioned that coupled inductor and switched capacitor technologies were used to get high voltage gain when considering DC distribution system with solar power optimizer [9]. Salpe performed research in solar power optimizer to get maximum energy from a photovoltaic panel and send this energy to a DC microgrid. A coupled inductor and switched capacitor are used to increase the voltage gain [10]. Sivakumar explained about boost converter using SIC diodes for PV applications. The converter designed here has two switching cells to distribute the output current [11]. Haoxiang et al. focused on developing a multi-objective optimization algorithm (MO-OPA) for power management in the radio networks. This method helps to reduce power consumption by minimizing the delay in communication [12]. Smys developed ad hoc networks in various traffic conditions [13]. Most of the researches mainly focus either on one or two power sources. One among solar panels or wind or hybrid sources and the other one is the AC grid line. These researchers tried to mitigate the existing issues in the available two power source switching systems. In this current research work, the necessary power is supplied by a design of four power sources that can switch to manage the power requirement.

2 Methodology

The automatic power optimizer cum power source controller can be divided into two parts namely the master module and the slave module. All the slave modules can be controlled by the signals of the master module. The number of slave modules can be increased according to the user requirement. This research work aimed to switch four power sources based on the energy requirement. The master controller mainly focuses on increasing the use of renewable energy sources and if the demand exceeds only, it will switch to the other source. Solar energy is an environmentally comfortable, renewable, limitless [14], and cost-effective [15] source among all other alternative sources used here. Solar energy will become an economical source of energy in the coming years and developing good technology for solar cells, reduction of cost, and efficiency in an application [16, 17]. So, the use of solar power is prioritized here according to the control given to the master controller. Usually, if the solar power supply is not enough to operate appliances which need high power, like refrigerator, cooler or motor, those appliances should be turned off.

But in this device, the power derived from the next available source will be added up with the existing power source and the continuous working of those appliances is ensured. High power electrical equipment is controlled by a slave module which automatically switches off certain devices, in case, if the total power supplied by all the available power sources cannot meet the demand. The slave modules are connected with the master module by the radio signal. If the demand exceeds the supply, the slave module sends an RF signal to the master module, and the master module gives the command to switch up or shut down that particular device. Within 100 m^2, the master module and other slave modules can communicate with each other. The workflow diagram of the system is shown in Fig. 1, and the block diagram of the system is shown in Fig. 2.

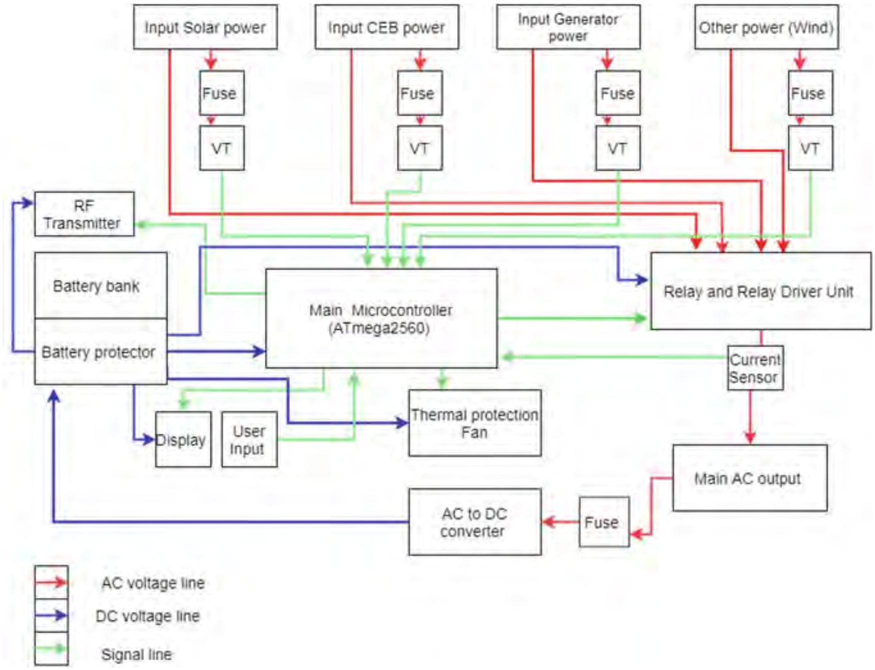


Fig. 1 Conceptual overall workflow diagram

2.1 AT Mega 2560 Microcontroller

AT Mega 2560 microcontroller is selected in this project since a large number of input and output are expected to be controlled by the microcontroller. AT Mega 2560 microcontroller is the big member of AT Mega series, and it has more number of I/O pins. The specifications of it is shown in Table 1. There are 16 analog pin in AT Mega 2560 microcontroller for analog processing, 54 pins for the general digital input–output processing, and 15 pins for the Pulse Width Modulation (PWM) processing.

In this research, work pins of all those categories are used to get inputs and give outputs. The keypad and wireless communication are connected as inputs such as the relay driver circuit, LED indicator, and LCD are connected as output in digital input–output pins. The current sensor is connected as input, and the VT transformer input is connected as output with Pulse Width Modulation (PWM) pin. Generator and solar data input are connected as the input to the system protection (Fan) are connected as output with analog pins.

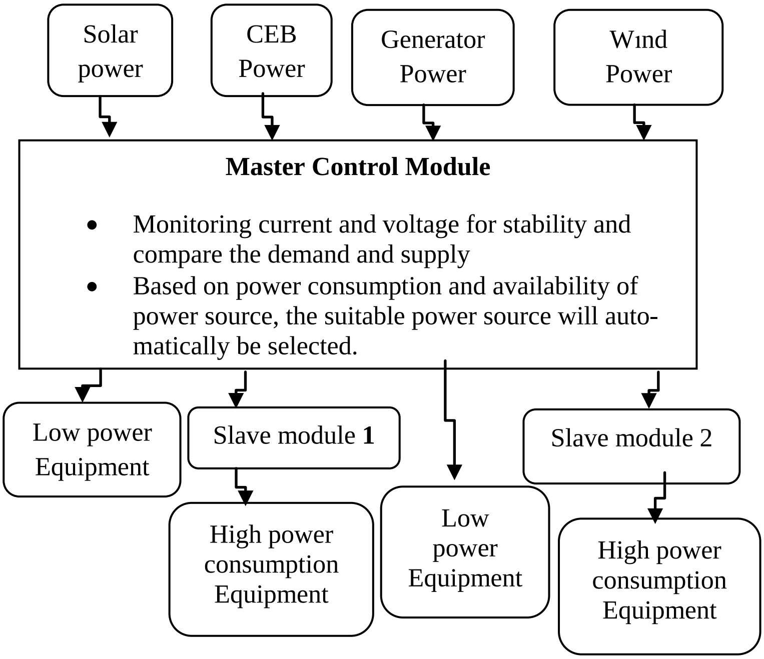


Fig. 2 Power optimizer full system block diagram

Table 1 Specifications of AT Mega 2560 microcontroller

Parameters	Range
Microcontroller	At mega 2560
Operating voltage	5 V
Digital I/O pins	54
Analog pins	16
PWM pins	15
Crystal oscillator	16 MHz
Current rating per I/O pin	20 mA

2.2 Relay Controller Module

The relay module is an important component in this equipment since all the processing is done on the AC power sources through the relay module. The relays are driven by relay driver circuit, and relay driver circuit is controlled by the main microcontroller (Fig. 3).

Based on the power requirement, the relay controller module switches from one source of power to the other or add up with the other and it is shown in Fig. 3. Four power sources namely solar, Ceylon electric board (CEB), generator, and wind are

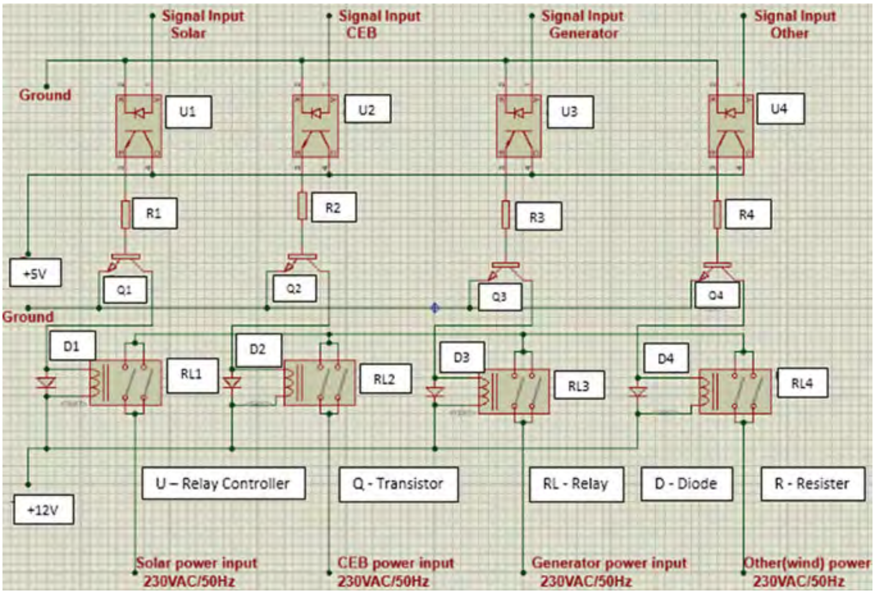


Fig. 3 Relay driver circuit diagram

occupied. The last source is mentioned as "other" in the diagram since any other source of power can be annexed with the existing design. Depend on the power demand, the microcontroller sends the signal to the relay circuit and energizes the relevant power source. When the relay is operated, back EMF would be generated and this signal affects the performance of the microcontroller since the relay is connected directly with the microcontroller. The rectifier diode is connected with the relay driver to send back EMF to the ground and prevent the intervention of back EMF in the performance of the microcontroller.

2.3 AC to DC Converter

Since all the processing inside the master controller module is occurring in DC voltage, the AC voltage needs to be converted to DC voltage by the AC to DC converter as shown in Fig. 4. Firstly, 230 VAC is reduced to 12 VAC, and after rectification and smoothening, 12 VDC output is obtained. 7805 IC is used to regulate the voltage from 12 to 5 VDC since 5 VDC is required to operate the microcontroller. One microfarad capacitor is used to get smoothened 5 VDC output.

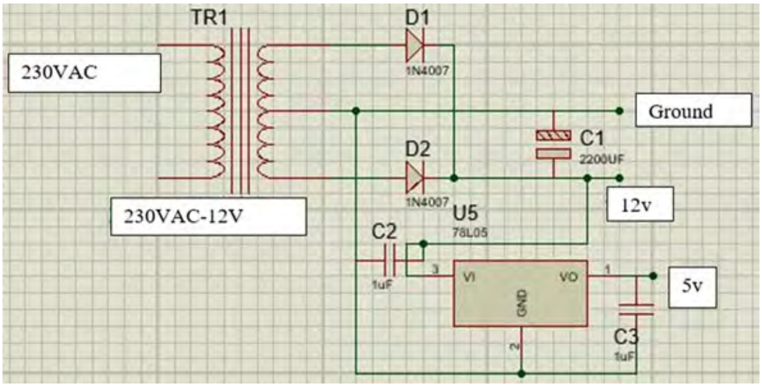


Fig. 4 Circuit diagram of AC to DC converter

2.4 Battery Bank

A battery bank is used as a reserve power source for the internal operation of the master controlling module. If the main power source failed to provide power to the device, the battery bank will provide it. If there is any interruption in the supply power to the master module, the whole system will be shutdown. Since the proper functioning of the device is needed always to switch the power, the battery bank is used as a backup and it is used to store the electricity to provide to the master controlling module. Three Lithium-ion batteries each of 3.7 V and 4000 mAh as shown in Fig. 5 are connected in series to give an approximate of 12 V output.

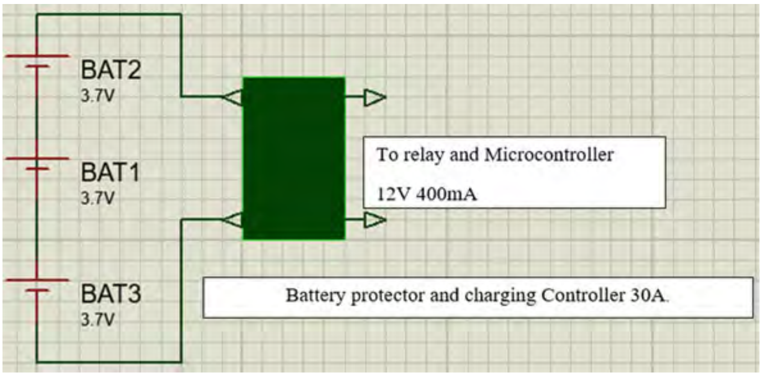


Fig. 5 Circuit diagram of battery bank

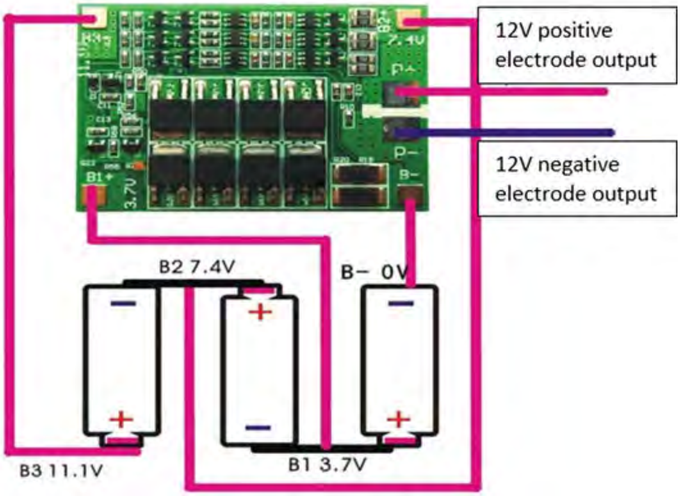


Fig. 6 Charging controller

2.5 Chargıng Controller and Battery Protector

The charging controller is designed in a way as shown in Fig. 6. Usually, the batteries are charged in a way that all batteries are connected in series. In case if one battery is damaged, none of the other batteries would be charged. Therefore, to prevent such happening and to protect the battery and other equipment connected with the battery, the individual charging of each battery is designed. In this way, the battery is charged individually and the lifetime of the battery will also be high.

2.6 Current Sensor

The current sensor is used to detect the electric current in a circuit. This sensor generates a signal proportional to the current. The generated signal may be voltage or current or digital output. ASC712 IC is used as a current sensor.

It has 8 pins and the supply voltage to it is 5 V. One pinout is provided for the analog output. When current flow through the ASC712 IC, the IC output analog voltage change between 0 and 5 V. The variation is mentioned in Fig. 7.

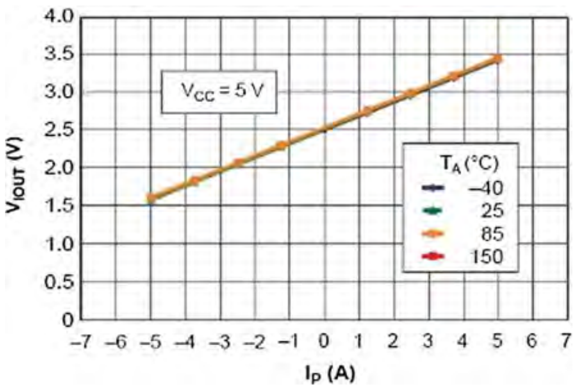


Fig. 7 Variation of current with voltage supply

2.7 Display Module

The display module shown in Fig. 8 is used to display the data needed by the user. 128 * 64 display is used in this project, and it has 20 pins which are also suitable for displaying the graphics. There are many Arduino libraries that can support this display module.

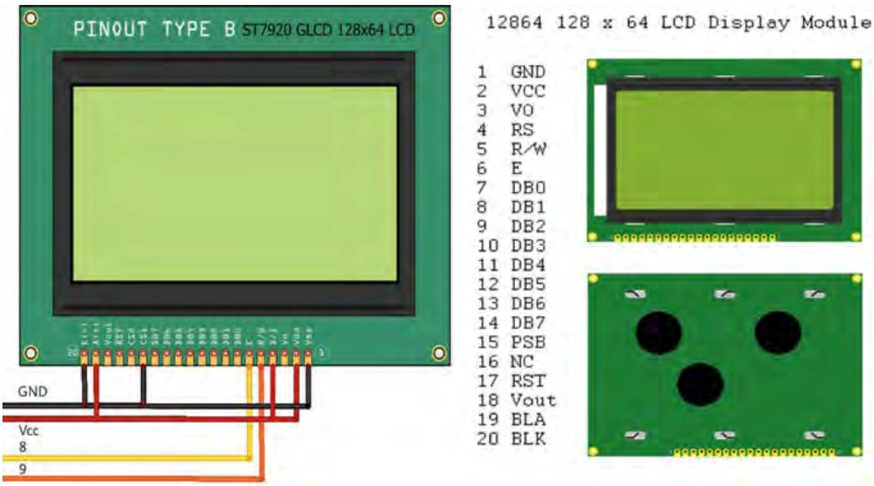


Fig. 8 Display module

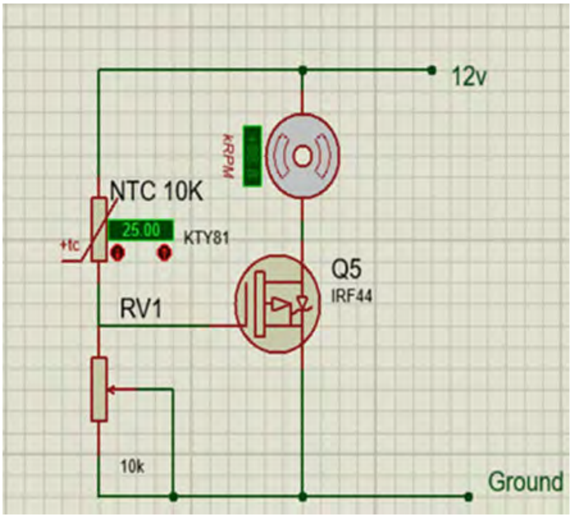


Fig. 9 Cooling system

2.8 Thermal Protection to the System

The master controlling module is a multi-processing unit and VT transformers are installed inside the module. Usually, transformers heat up while operating. In order to prevent the heating up of the transformers, normally an in-built cooling system would be provided in most of the transformers. But VT transformers usually does not have a cooling arrangement. Therefore, a separate cooling system is provided in this research work as shown in Fig. 9. The temperature should be maintained at the proper level by the automatic cooling fan for the efficient functioning of the transformer. When the temperature increases beyond a certain point, the driver circuit of the fan identifies it based on thermistor reading because the resistance of the thermistor changes with temperature. NTC thermistor is used in this project. If the temperature increases, the resistance across the terminals would be reduced. In this condition, IRF44 MOSFET would be biased and current flows through the source to drain. This will switch ON the fan and thereby cool the system. Once the temperature reduces below the prescribed value, the fan automatically switches off.

2.9 Voltage Transformer

The voltage transformer is used to measure the AC voltage. The 230 VAC is converted to 6 VAC, and it is rectified and smoothened by 10 microfarad capacitors to get the regulated voltage of 5.1 V and this is the input voltage signal to the microcontroller. The primary coil side of the transformer is connected with 800 mA fuse in series

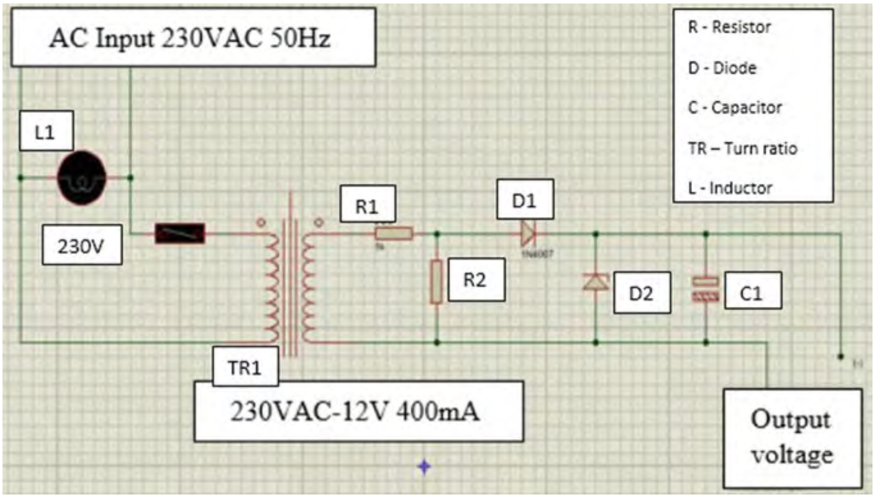


Fig. 10 Circuit diagram of voltage transformer

to protect the VT transformer from overvoltage. The circuit diagram of the VT transformer is shown in Fig. 10.

2.10 Data Access Port

When operating various energy sources, the master controlling module needs the amount of energy that can be supplied by each source. The data access ports are introduced to access the data of each power source. The solar power source produces solar radiation and heat. And, if it is an electricity generator calculate the amount of fuel left and if it is a wind power source determine the speed of wind, etc. Not only the basic data, but also the data like inverter mode and PV voltage for solar source, the pulse of starting motor and generator temperature for generator source, and all the data necessary for the user would be provided by data access port. These data are provided to the master module through a data access port. 5.1 V zener diode is used in this module to limit the input signal up to 5 V since the input data to solar and generator sources should be around 5 V since the operating voltage of the microcontroller is 5 V. If more than 5 V flows through the microcontroller, it will be damaged. The circuit diagram and schematic diagram of the data access port are shown in Figs. 11 and in 12, respectively.

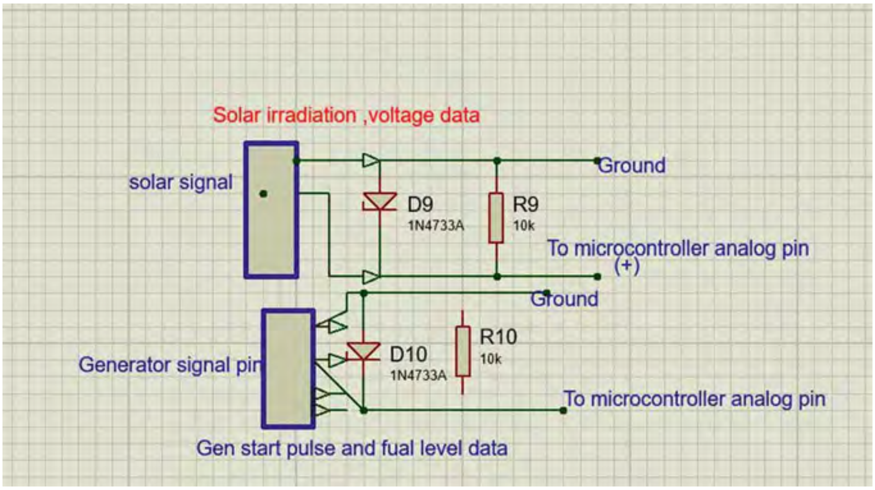


Fig. 11 Circuit diagram of data access port

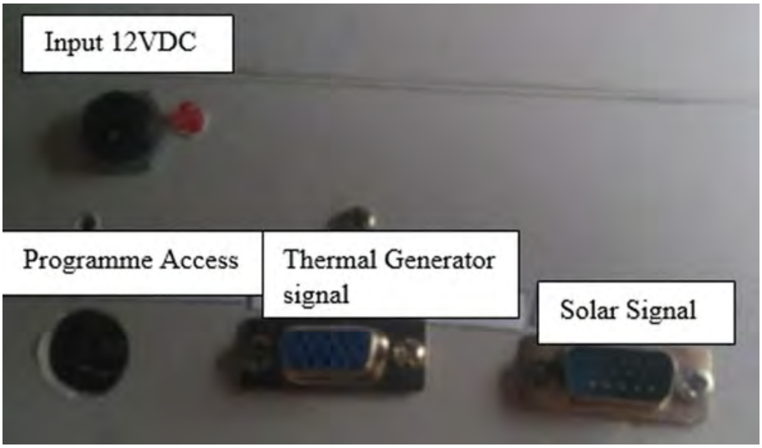


Fig. 12 Diagram of data access ports in the device

2.11 Fuse

Each circuit has a series of fuses connected with it to protect the circuit from the threat of overvoltage or short circuit as shown in Fig. 13. The damage occurs in the circuit of one source, damaging the other circuit is also prevented by the introduction of fuses. The fuses can easily be replaced since the fuse holders are mounted outside the module case. Connectors are used to connect each input and output wires, and it can easily be removed and safely be connected with the master module. Each connection have a separate terminal as shown in Fig. 14.

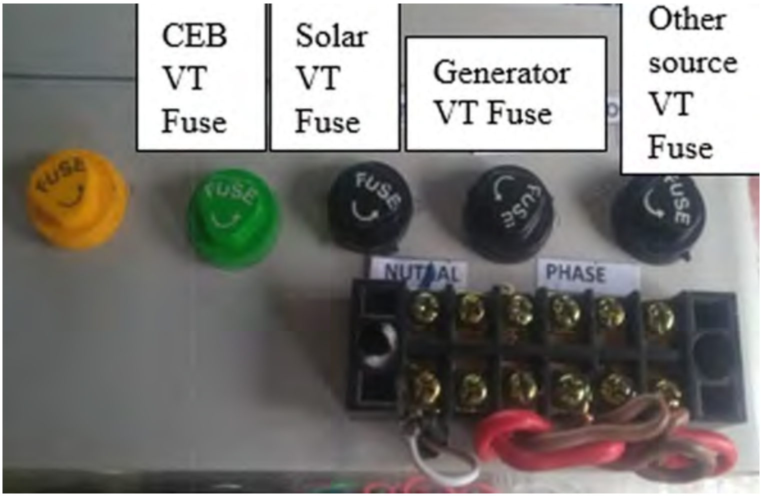


Fig. 13 Arrangement of fuses

Fig. 14 Diagram of connector terminal bar

2.12 Slave Module

The slave module could be located in different locations that are connected with the master module. Heavy equipment that consumes high power is connected with the main output AC line through a slave module as shown in Fig. 15. In case if the combined power supply by all four sources cannot meet the demand, the slave module needs to turn off some equipment connected through it with the main AC line. The power is supplied to the slave module as shown in Fig. 16.

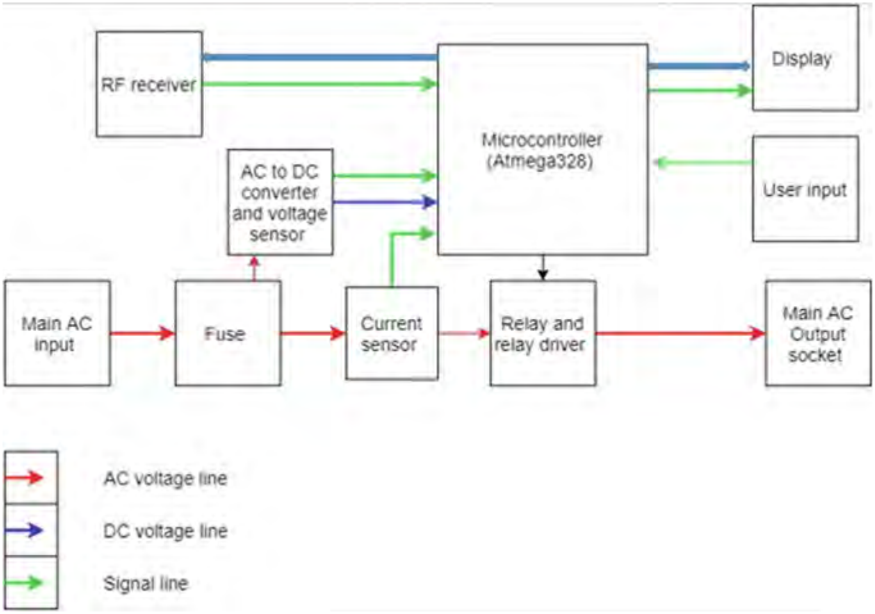


Fig. 15 Block diagram of slave module

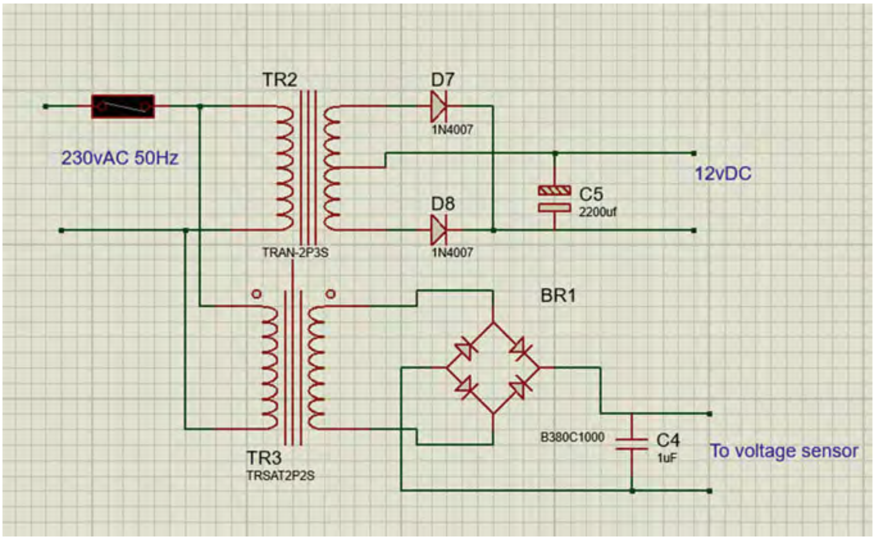


Fig. 16 Power supply for slave module

3 Results and Discussion

Wireless connection is established between the master controlling module and the slave module. A cheap and reliable method of wireless connection are a radio signal. 433 MHz transmission module is used since it can cover a large working area of 100 m and it is comparatively cheap. The transmitter module is in the master controlling module and the receiver module is in each slave module. When four power sources are connected with the master controller module, it checks the availability of the power sources, reads the voltages and currents of all power sources, and also reads the power consumption of the main AC output. The master controller module always tries to connect with the solar power sources at the first attempt.

At first, the master control module checks the availability of solar power. If solar power is available, the solar power source would be selected and connected with the main AC output. If the solar power source is not available, on the next turn, the master control module checks the availability of CEB power. If CEB power is available, the main AC output would be connected with the CEB power source. If the CEB power source is not available, the master controller module on its third run checks the generator fuel level. If the fuel level is sufficient, the master control module sends the signal to the generator to start. So, the generator would be started and generator power can be connected with the main AC output. During peak hours, the solar power is not sufficient most of the time. Therefore, the master controller module always checks the availability of CEB and generator.

In case if the combined supply of all four power sources together is not enough during peak hours of power consumption, the master controller module sends the signal to the slave module to cut-off the power from equipment like air conditioner which consumes high power. Equipment consume less power to keep on working while equipment consumes high power would be shut down until the balance between supply and demand arise. The basic idea of this device is to increase the usage of easily available renewable power sources instead of conventional government power supply. The prevention of environmental pollution and the economic growth of the country is also expected to be addressed through this research work. For example, coal and diesel power which are used in thermoelectric plants are polluting the environment and, it has the threat of extinguishing soon. In most of the countries, coal and petroleum products are imports. Renewable resources like solar, tidal, wind, and hydro energy are environment-friendly and readily available energy sources (Fig. 17).

4 Conclusion and Future Scope

There are two modules developed in this design, one is the master control module and the other one is the slave module. The master control module has control over the power sources and the parameters of all individual power sources. The slave module is used in a way that the heavy equipment which consumes high power is

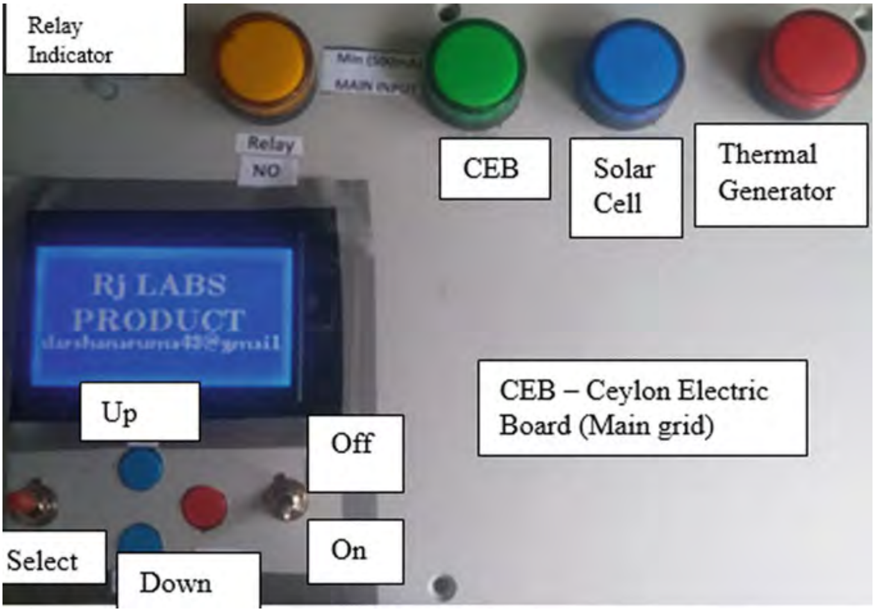


Fig. 17 Real-time working model of the proposed system

connected through the slave module with the main AC grid line. If power production from one source is not enough, the master module programmed to add up more than one power source to provide the necessary power. If the combined supply of all four power sources becomes lower than demand and the power supply from the source is detected by the master controlling module as low, it sends a signal to the salve module to switch off certain devices based on the output AC voltage. Similarly, if the master controlling module detects sufficient power from the sources, then it sends the signal to all the slave modules to switch ON the output AC supplies which are closed.

The working of the power optimizer cum automated power sources controller is based on the availability of the power sources and power demand on the main AC output. As a development of this controller, the addition of the power consumption for every month is expected to be calculated daily based on KW/h to find a graph of peak hours. If it is found, then the master controller can select the power sources according to this graph as a predetermined function. And also it would be very useful in the industrial point of view since the industries run cyclic workloads on their day-to-day functioning. If the heavy-duty hours of the industries are found accurately, this device can be programmed appropriately and maintenance of the device and the system will also be easy.

References

1. Shaikh MRS, Waghmare SB, Labade SS, Fuke PV, Tekale A (2017) A review paper on electricity generation from solar energy. Int J Res Appl Sci Eng Technol 5(IX):1884–1889
2. Orduz R, Solórzano J, Egido MÁ, Román E (2011) Analytical study and evaluation results of power optimizers for distributed power conditioning in photovoltaic arrays. Fundación Tecnalia Research and Innovation, Energy Unit, Edit. 700 Derio 48160, Spain
3. Suresh Babu K, Raju S, Srinivasa Reddy M, Nageswara Rao DN (1887) The material selection for typical wind turbine blades using MADM approach & analysis of blades. In: MCDM 2006, Chania, Greece, June 19–23
4. Kumar A, Dwivedi A, Paliwal V, Patil PP (2014) Free vibration analysis of Al 2024 wind turbine blade designed for Uttarakhand region based on FEA. Procedia Technol 14:336–347
5. Mathavan JJ, Patnaik A (2020) Development and characterization of polyamide fiber composite filled with fly ash for wind turbine blade. In: Emerging trends in mechanical engineering. Springer, Singapore, pp 131–139
6. MacAlpine SM, Erickson RW, Brandemuehl MJ (2013) Characterization of power optimizer potential to increase energy capture in photovoltaic systems operating under nonuniform conditions. IEEE Trans Power Electron 28(6):2936–2945. https://doi.org/10.1109/TPEL.2012.2226476
7. Elkamouny K, Lakssir B, Hamedoun M, Benyoussef A , Mahmoudi H (2017) Simulation, design and test of an efficient power optimizer using DC-DC interleaved isolated boost PV-micro inverter application. In: 2017 14th ınternational multi-conference on systems, signals & devices (SSD), Marrakech, pp 518–525. https://doi.org/10.1109/SSD.2017.8167019
8. Sanz A, Vidaurrazaga I, Pereda A, Alonso R, Román E, Martinez V (2011) Centralized vs distributed (power optimizer) PV system architecture field test results under mismatched operating conditions. In: 2011 37th IEEE photovoltaic specialists conference, Seattle, WA, pp 002435–002440. https://doi.org/10.1109/PVSC.2011.6186440
9. Muthamizhan T (2016) Performance analysis of solar power optimizer for DC distribution system. Int J Adv Res Electr Electron Instrum Eng 5(8):6708–6715
10. Salpe AD (2016) Design, analysis, of efficient solar power optımizer for DC mıcro grid system. Int J Adv Res Sci Eng 5(6):182–189
11. Sivakumar K (2014) Implementatıon of interleaved boost converter using SIC diodes in residential PV pre-regulator application. Int J Adv Eng Technol 6(6):2537–2547
12. Haoxiang W (2019) Multi-objective optimization algorithm for power management in cognitive radio networks. J Ubiquit Comput Commun Technol (UCCT) 1(02):97–109
13. Smys S, Josemin Bala G, Jennifer S (2010) Mobility management in wireless networks using power aware routing. In: 2010 International conference on intelligent and advanced systems, pp. 1–5. IEEE
14. Chu Y, MeisenP (2011) Review and comparison of different solar energy technologies. Report of Global Energy Network Institute (GENI), Diego
15. Choubey PC, Oudhia A, Dewangan R (2012) A Review: solar cell current scenario and future trends. Recent Res Sci Technol 4:99–101
16. Wall A (2014) Advantages and disadvantagesofsolarenergy. Process Industry Forum, 7 Aug 2013. Web, 2 Feb 2014
17. Bagher AM, Vahid MMA, Mohsen M (2015) Types of solar cells and application. Am J Opt Photon 3:94–113

Texture-Based Face Recognition Using Grasshopper Optimization Algorithm and Deep Convolutional Neural Network

Sachinkumar Veerashetty and Nagaraj B. Patil

Abstract Face recognition is an active research area in biometric authentication, which has gained more attention among researchers due to the availability of feasible technologies, including mobile solutions. However, the human facial images are high dimensional, so the dimensionality reduction methods are often adapted for face recognition. However, the facial images are corrupted by the noise and hard to label in the data collection phase. In this study, a new GOA-DCNN model is proposed for face recognition to address those issues. Initially, the face images are collected from two online datasets FEI face and ORL. Next, modified local binary pattern (MLBP) and speeded up robust features (SURF) are used to extract the feature vectors from the collected facial images. The extracted feature values are optimized using grasshopper optimization algorithm (GOA) to decrease the dimensionality of data or to select the optimal feature vectors. At last, deep convolutional neural network (DCNN) was applied to classify the person's facial image. The experimental result proves that the proposed model improved recognition accuracy up to 1.78–8.90% compared to the earlier research works such as improved kernel linear discriminant analysis and probabilistic neural networks (IKLDA + PNN) and convolutional neural network (CNN) with pre-trained VGG-Face.

Keywords Deep convolutional neural network · Face recognition · Grasshopper optimization algorithm · Modified local binary pattern · Speeded up robust features

1 Introduction

In recent years, face recognition plays a vital role in a biometric authentication system that is applied in many applications such as law enforcement, access control, video

S. Veerashetty (✉)
Appa Institute of Engineering and Technology, Kalaburagi, India
e-mail: sveerashetty@gmail.com

N. B. Patil
Government Engineering College, Yaramaras, Raichur, India
e-mail: nagarajbpatil1974@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_4

surveillance, and access control [1]. Due to the rapid growth of electronic equipment techniques, a large number of face images are captured using cell phones and cameras. Hence, the image-based facial recognition becomes essential in real-world applications. The human facial images have high dimensionality that leads to the curse of dimensionality and computational complexity in the face recognition system [2]. Meanwhile, the accuracy of the face recognition system is compromised dramatically in real-world applications by inter and intraclass variations, due to facial expression, occlusion, aging effect, head pose variation, poor illumination, and low and blur resolutions [3—6]. To address these concerns, many dimensionality reduction methods have been developed based on unsupervised, supervised, and semi-supervised conditions. The unsupervised methodologies include principal component analysis [7], sparsity preserving projection, etc. The supervised methods include linear discriminant analysis [8], maximum margin criterion, etc. Still, the existing face recognition techniques are not fulfilling the requirements of real-world applications in the case of large scale data scenes. So, a superior optimization technique with a deep learning classifier (GOA-DCNN) is proposed in this study.

At first, the facial images are collected from two databases FEI face and ORL. Then, image quantization is carried out to improve the quality of the facial images. The image quantization technique reduces the number of colors utilized in the facial images, which is essential to display the images on devices like mobile phones, biometric attendance devices, etc. Besides, MLBP and SURF features are used to extract the feature vectors from the denoised facial images. In LBP, the sign vector is failed to extract the important texture feature vectors, so the rotation invariance, scale, and illumination vectors are calculated in MLBP for extracting all the texture feature vectors. The extracted features are high dimensions in nature, which is optimized by GOA to reduce the curse of dimensionality issues. The optimized features are classified by the DCNN classifier to classify the person's face images.

The paper is arranged in the following manner, a few recent research papers on face recognition is explained in Sect. 2. The proposed model with mathematical expressions is discussed in Sect. 3. Section 4 presents the quantitative and comparative results of proposed and existing models. Finally, the conclusion is described in Sect. 5.

2 Literature Survey

Ouyang et al. [9] presented a hybrid model improved kernel linear discriminant analysis (IKLDA) and probabilistic neural networks (PNNs) for facial recognition. Initially, the IKLDA method was adapted to reduce the dimension of the extracted features to retain the most relevant information about the facial images. Next, the PNN method was applied to solve the difficulties of facial recognition. The developed model (IKLDA and PNN) not only enhanced the recognition accuracy but also improved the overall computing efficacy. In this study, the performance of the developed model was validated on three databases like AR, YALE, and ORL. These databases comprise a wide range of face details, expressions, and scale degrees. The experimental results proved that the developed model achieved better recognition accuracy compared to the existing techniques. In facial image classification, PNN is slower than multi-layer perceptron networks, because it requires more memory space for storing the model.

Faraji and Qi [10] presented a multi-scale approach for facial recognition based on the maximum response filter. Initially, the facial images were scaled using a log function for compressing the brighter image pixels and expanded the darker image pixels. The multi-scale approach used a filter bank to reduce the illumination and enhanced the edges of the image. At last, an improved multi-scale gradient face method was used to capture different properties of the facial images to generate an illumination invariant feature representation. The developed multi-scale approach attained good performance related to other earlier methods of facial recognition. In this literature study, manual intervention is high during testing and training of the data that increase the computational time of the system.

Elmahmudi and Ugail [11] utilized a convolutional neural network (CNN) with a pre-trained VGG face model to extract the features and classify an individual's face. In this study, labeled faces in the wild (LFW) and Brazilian FEI datasets were used to evaluate the performance of the developed model. Simulation outcomes showed that the individual parts of the face such as nose, cheeks, and eyes were achieved a better rate of recognition. The conventional CNN does not encode the position and orientation of the face cues, and also, it is computationally expensive. Besides, Li and Suen [12] developed a new model for facial recognition by extracting the discriminate parts and dynamic subspace of the facial images. These parts represent the discriminative components and provide a recognized protocol to classify the facial images. In this study, the experiment was performed on three online datasets extended Yale B, ORL, and AR to validate the speed, robustness, and accuracy of the developed model. However, face occlusions, and variations are the major concerns in this study to develop a robust face recognition system.

Li et al. [13] implemented recurrent regression neural network (RRNN) for facial recognition. In the RRNN classifier, the encoder-decoder was a first unit, which was used to model sequential reconstruction. The second unit was utilized for constraining the global nature of the sequences, and the final unit was utilized to label the discriminative information. The experimental results proved that the RRNN classifier achieved better recognition results compared to the existing methodologies. However, there is a loss of information while embedding the low-resolution facial images in higher-level layer, where the RRNN classifier contains several a higher-level layers. Tang et al. [14] used LBP to extract the feature values from the collected face images. Then, ten CNN with five dissimilar network structures were employed for extracting the feature vectors for training that enhance the network parameter and classification result utilizing softmax function. At last, a parallel ensemble learning methodology was applied for generating the result of face recognition. As previously mentioned, CNN requires high graphics processing unit system for attaining better performance in face recognition, where it is highly expensive. To highlight these concerns, a new optimization technique with a deep learning classifier is proposed in this study for facial recognition.

3 Methods

The face recognition is an emerging research topic, which attracts more researchers in the field of pattern recognition and computer vision [15, 16]. In multimedia applications, face recognition has great potential, for instance, personal verification, video surveillance, digital entertainment, etc. Therefore, image-based face recognition is necessary for many real-time applications and it becomes a popular research topic in the area of facial recognition [17, 18]. In this research, the proposed GOA-DCNN model contains five steps: image collection from FEI face and ORL databases, image pre-processing (i.e., image quantization), feature extraction using MLBP, and SURF, feature optimization using GOA, and classification using DCNN. The workflow of the proposed model is presented in Fig. 1.

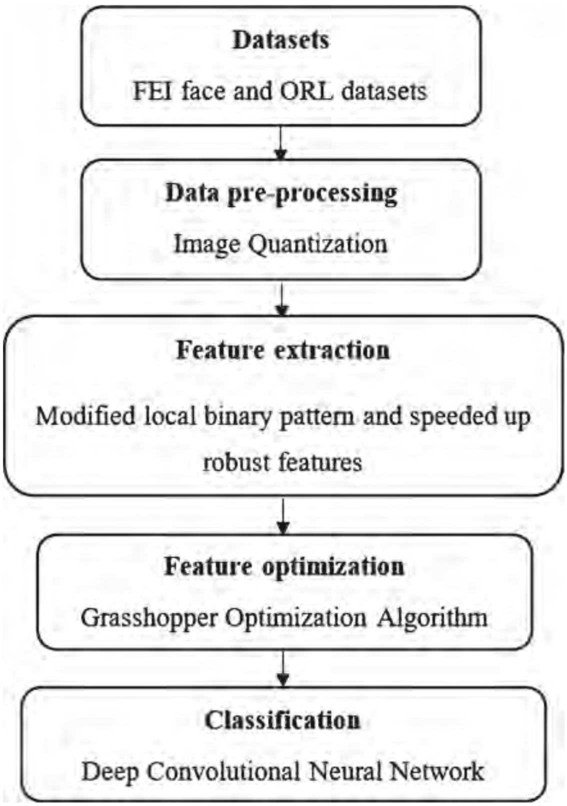


Fig. 1 Block diagram of proposed model

Fig. 2 Graphical illustration of FEI face dataset

3.1 Dataset Description

In this study, the input facial images are collected from two datasets FEI face and ORL. FEI dataset includes different facial images that are collected from the period of June 2005 to March 2006 at the artificial intelligence laboratory of the FEI in Brazil [19]. In the FEI dataset, the facial images are collected from 200 subjects; each subject includes 14 images and a total of 2800 facial images. In this dataset, all the facial images are colorful, those were taken under white homogenous background and the size of every facial image is 640 × 480 pixels. In this dataset, the facial images are collected from staff and students at FEI. Where the age ranges between 19 and 40 years old with a different hairstyle, appearance, and adorns. A graphical illustration of the FEI face dataset is represented in Fig. 2.

Also, the ORL database includes 400 facial images, where each subject contains ten different images. For some individuals, the facial images were captured at different lighting variations, periods, facial expressions (not smiling/eye open/smiling/eye closed), and facial details (no glasses/glasses) [20]. A graphical illustration of the ORL data set is presented in Fig. 3.

3.2 Image Pre-processing

The quantization process is carried out to enhance the visibility level of the facial images after collecting the images. The quantization process includes three steps; initially divide the color components into "*n*" and "*p*" shades, then combine red, green, and blue monochromes into a single channel to build the color features, and finally extracts the set of points with the quantized color "*Q*." Graphically the quantized image is presented in Fig. 4.

Fig. 3 Graphical illustration of ORL dataset

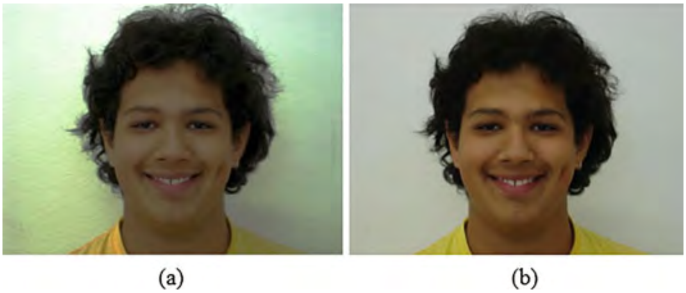


Fig. 4 **a** Input image, **b** quantized image

3.3 Feature Extraction

After denoising the facial images, feature extraction is performed using SURF and MLBP. Compared to other techniques, the selected feature extraction techniques are very simple and efficient in extracting the texture features to achieve better performance in the conditions like illumination condition, lighting variation, facial rotations, etc. Brief explanations about the feature extraction techniques are given below.

3.3.1 Speeded Up Robust Features

The SURF feature is utilized for detecting the blob-like structure when the Hessian matrix determinant is maximized. Consider a point $x = (x, y)$ in a face image I and the hessian matrix $H(x, \sigma)$ at x with scale σ is mathematically denoted in Eq. (1).

$$H(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix} \tag{1}$$

where $L_{xx}(x,\sigma), L_{xy}(x,\sigma),$ and $L_{yy}(x,\sigma)$ are indicated as the convolution of Gaussian 2nd order derivation $\frac{\partial^2}{\partial x^2}$ at point x. The scale space is divided into octaves to detect interest points at different scales, where every octave has a series of intervals. The convolution window scale with parameters interval i and octave o is mathematically indicated in Eq. (2). Meanwhile, the relation between the window size and scale σ is denoted in Eq. (3).

$$L = 3 \times \left(i \times 2^o + 1\right) \tag{2}$$

$$L = \sigma \times 9/1.2 \tag{3}$$

Then, the SURF key point is mathematically denoted in Eq. (4).

$$\mathrm{DoH}(x, L) = \max\left(\sum_{k_i=i-1}^{i+1} \sum_{k_x=x-2^o}^{x+2^o} \sum_{k_y=y-2^o}^{y+2^o} \mathrm{DoH}\left(k_x, k_y, o, k_i\right)\right) \geq \lambda \tag{4}$$

where λ is indicated as a positive threshold and DoH is stated as a Hessian matrix determinant. A bright blob centered at (x, y) with scale $L = 3 \times (i \times 2^o + 1)$ is detected if the trace of the hessian matrix is larger than zero.

3.3.2 Modified Local Binary Pattern

The LBP is a productive and effective methodology in image processing applications like face recognition. LBP is a texture feature descriptor, where the central pixel is indicated as g_a. The vector of image pixel p_x is indexed as $g_0, g_1, g_2 \ldots g_{p_{x-1}}$, where the LBP features are obtained by multiplying binomial factor with every binary value, as stated in the Eqs. (5) and (6).

$$\mathrm{LBP}_{px} = \sum_{p_x=0}^{p_x=1} d\left(g_{p_x} - g_a\right) \tag{5}$$

where

$$d(n) = \begin{cases} 1 & n \geq 0 \\ 0 & n < 0 \end{cases} \tag{6}$$

Initially, the squared neighborhood image pixels are estimated in the conventional texture descriptor systems. There will be a variation in rotation invariance if the sign vector varies, so the combination of sign vector and rotation variance is used to extract all the texture features. Therefore, rotation invariance, illumination, and scale are estimated to extract all the texture features, and the magnitude vector in the difference vector is also considered.

Scale Invariance

It is achieved by eliminating the gray value of g_a from circular symmetric neighborhood gray values $g_{p_x}(p_x = 0, 1, \ldots p_{x-1})$, which is mathematically indicated in Eq. (7).

$$T = t\left(g_a, g_0 - g_a, \ldots.g_{p_{x-1}} - g_a\right) \tag{7}$$

where $t(g_a)$ is stated as the luminance value of the facial image. The scale invariance is achieved by considering circular symmetric neighborhood sets by changing the radius $r = 1, 2, 3$ and 4.

Rotation Invariance

The conventional invariant system utilizes only the sign values for facial image texture investigation. In some cases, the sign vector value varies if there are any variations in the image rotations. So, the sign vector along with the magnitude vector is considered to attain rotation invariance, because the magnitude vector remains the same in all the conditions. The mathematical equations to attain rotation invariance are denoted in Eq. (8).

$$\mathrm{LBP}_{p_{x,r}}^{id} = \min\left(\mathrm{RS}\left(\mathrm{LBP}_{P_{x,r}} j\right), \quad j = 0, 1, \ldots P_{x-1}\right) \tag{8}$$

where $\mathrm{RS} = \left(\mathrm{LBP}_{p_{x,r}} j\right)$ performs a bit-wise circle right on x for j times and $\mathrm{LBP}_{p_{x,r}}^{id}$ is denoted as rotation invariant code. The rotation invariance is also utilized to find the illumination changes in the facial images.

Local Difference Sign and Magnitude Transformation

The local difference vector $\left[E_0, E_1, \ldots E_{p_x} - 1\right]$ is proved to be robust for illumination changes by removing g_a. Compared to the input images, there will be more efficient in pattern matching by eliminating g_a from local difference vector $\left[E_0, E_1, \ldots E_{p_x} - 1\right]$. Hence, E_{p_x} is categorized into two elements, as mentioned in the Eqs. (9) and (10).

$$E_{p_x} = S_{p_x} \times M_{p_x} \ \text{and} \ \begin{Bmatrix} S_{p_x} = \mathrm{Sign}\left(E_{p_x}\right) \\ M_{p_x} = \left|E_{p_x}\right| \end{Bmatrix} \tag{9}$$

$$S_{p_x} = \begin{cases} 1 & E_{p_x} \geq 0 \\ -1 & < 0 \end{cases} \tag{10}$$

where M_{p_x} is indicated as E_{p_x} magnitude and S_{p_x} is signified as E_{p_x} sign. Then, the extracted texture features are optimized by GOA to decrease the dimensionality of the extracted feature values.

3.4 Feature Optimization

GOA is a population-based optimization technique that easily handles the unconstrained optimization problems. GOA imitates the behavior of grasshoppers, where the three components (gravity G_r, social relationship S_i, and horizontal wind movement W_i) affect the flying route of grasshopper. In GOA, the searching process is mathematically denoted in Eq. (11).

$$S_i = \sum_{j=1, j \neq i}^{M} s\left(p_{i,j}\right) \widehat{p}_{ij} \tag{11}$$

where p_{ij} is represented as distance between ith and jth grasshopper that is estimated as $p_{ij} = \left|x_j - x_i\right|$, s is stated as strength of social forces and the unit vector from ith to jth grasshopper is indicated as $\widehat{p}_{ij}$, which is mathematically defined in Eq. (12).

$$\widehat{p}_{ij} = \frac{x_j = x_i}{p_{ij}} \tag{12}$$

The function s is the backbone of the social relationship that represents the grasshopper direction in the swarm, and it is mathematically stated in Eq. (13).

$$s(r) = be^{-\frac{r}{l}} - e^{-r} \tag{13}$$

where l is indicated as an attractive length scale, b is represented as an attraction force, r is denoted as the distance between grasshoppers. In the GOA, two types of forces attraction and repulsion are generated. The repulsion force increases when r is in the range of [0, 2.079] that avoids the collision. Correspondingly, the attraction force increases when r is in the range of [2.079, 4] that efficiently handles the swarm cohesion. Equation (14) states the mathematical expression of grasshopper's interaction.

Table 1 Features selected after applying GOA

Sample image	Extracted features	Selected features
1	42 × 3454	42 × 2592
2	42 × 4951	42 × 3028
3	42 × 3885	42 × 3012
4	42 × 4109	42 × 3482
5	42 × 3984	42 × 3091

$$X_i^k = c\left(\sum_{j=1, j\neq i}^{M} c\frac{\mathrm{up}_k - \mathrm{lp}_k}{s} s\left(x_j^d - x_i^d\right)\frac{x_j - x_i}{p_{ij}}\right) + \widehat{T}_k \tag{14}$$

where $\widehat{T}_k$ is indicated as kth dimension in the target, lp_k and up_k are stated as lower and upper bounds, and c is denoted as a coefficient that is utilized for reducing the comfort, repulsion, and attraction regions. The parameter c moves the grasshoppers closer to the target that is estimated as the best solution. The best solution gets updated if the grasshoppers chase the target. The parameter c is estimated using Eq. (15).

$$c = c_{\max} - l\frac{c_{\max} - c_{\min}}{N} \tag{15}$$

where l is indicated as current iteration, N is stated as maximum iterations, $c_{\min} = 0.00001$ and, $c_{\max} = 1$GOA is effective in solving the optimization issues and also the complexity of the algorithm is very low, due to the simple calculation of the distance between the grasshoppers. Table 1 denotes the feature vectors selected after applying GOA. The flow diagram of the GOA is given in Fig. 5.

3.5 *Classification*

Deep neural network (DNN) is a feed-forward network, where the information is processed by layer-by-layer. In CNN, the layers are stacked where the output of a hidden layer is the input to the succeeding layer. The output of every layer is a function of its internal weights and input. CNN is mathematically defined in Eq. (16).

$$\begin{cases} o_i = X & i = 1 \\ o_i = f_i(z_i) & i > 1 \\ z_i = g_i(o_{i-1}, w_i) \end{cases} \tag{16}$$

where $f_i(.)$ is denoted as the activation function of the ith layer, z_i is stated as weighted operation output of the ith layer, X is indicated as input data, $g_i(.)$ is specified as a weighted operation of the ith layer, w_i is denoted as the weight of the ith layer, and

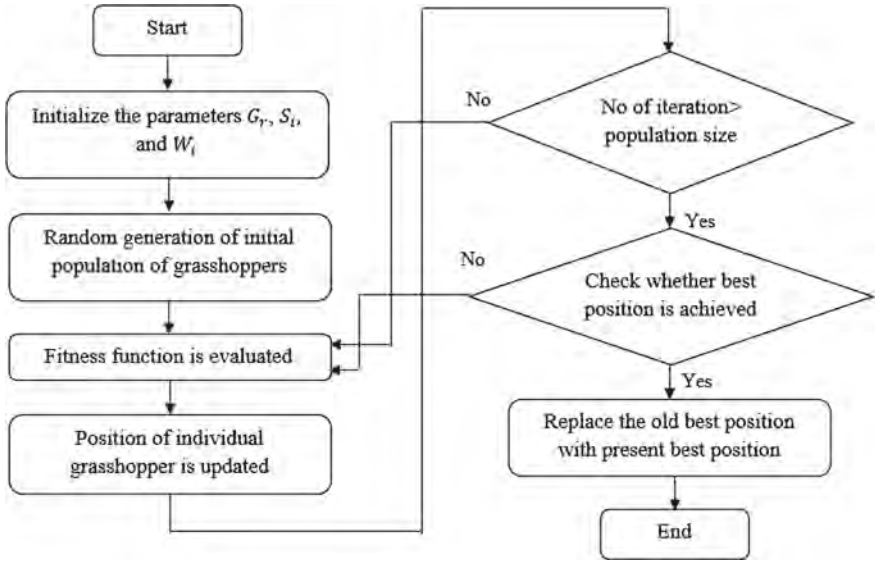


Fig. 5 Flow diagram of GOA

o_i is represented as output of the ith layer. CNN includes three main layers pooling, convolution, and fully connected layers.

The output of the convolutional layer is a convolution operation (Θ) on its weights and inputs and it is known as filters or kernels. There are two kinds of pooling layers average and max-pooling layers that perform a down sampling operation in $r \times c$ window$\left(\mathrm{N}_{r,c}\right)$ for decrease the number of output parameters. In a fully connected layer, the output is a function of the weight multiplied by its input. The CNN layers are mathematically denoted in Eq. (17).

$$\begin{cases} Z_i = w_i \Theta o_{i-1} & \text{if } i\text{th layer is convolutional} \\ z_i = \mathrm{N}_{r,c}\, o_{i-1} & \text{if } i\text{th layer is pool} \\ z_i = w_i o_{i-1} & \text{if } i\text{th the fully connected} \end{cases} \tag{17}$$

The DCNN helps to reduce the error between predicted outputs and training targets. Generally, the minimization of cross-entropy loss is carried out using back propagation and gradient descent in the DCNN classifier. The parameter setting of DCNN is given as follows; convolutional layers are 5, the number of convolution filters in one layer, is 96, the type of activation function is softmax, and pooling window size 5 * 5. The structure of DCNN is given in Fig. 6.

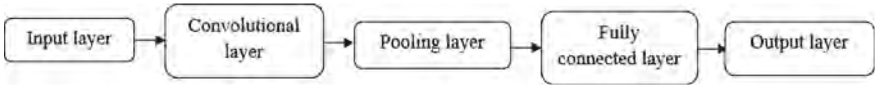


Fig. 6 Structure of DCNN classifier

4 Experimental Investigation

The MATLAB (2019a) environment is used for experimental investigation in a personal computer consist of Intel® Core™ i5-3220 CPU @ 3.30 Hz, 16 GB RAM, and 2 TB hard disc. The performance of the proposed GOA-DCNN model is compared with two existing models IKLDA + PNN [9] and CNN with pre-trained VGG-Face [11] to validate the efficacy of the proposed model in terms of accuracy, precision, f-score, and recall. Though f-measure is used to estimate the test accuracy, and it balances the usage of recall and precision. The mathematical expressions of accuracy, f-score, recall, and precision are presented in the Eqs. (18)–(21).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{FN} + \text{TP} + \text{FP}} \times 100 \tag{18}$$

$$F\text{-score} = \frac{2\text{TP}}{\text{FP} + \text{FN} + 2\text{TP}} \times 100 \tag{19}$$

$$\text{Recall} = \frac{\text{TP}}{\text{FP} + \text{TP}} \times 100 \tag{20}$$

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}} \times 100 \tag{21}$$

where true positive is denoted as TP, true negative is indicated as TN, false positive is stated as FP, and false negative is represented as FN.

4.1 Quantitative Investigation on FEI Face Dataset

In this section, the FEI database is taken for validating the performance of the proposed model. The performance of the proposed method is compared with different classification techniques like multiclass support vector machine (MSVM), K-nearest neighbors (KNN), long short-term memory (LSTM), deep belief network (DBN), and DCNN in terms of f-score, accuracy, recall, and precision. In the FEI database, 80% (2240) of the images are used for training and 20% (560) of the images are used for testing, where the collected FEI database facial images are cropped into the size of 32 * 32. By inspecting Table 2, it is concluded that the GOA with DCNN classifier has achieved 98.90% of recognition accuracy, which is better than the other classifiers MSVM, KNN, LSTM, and DBN. Additionally, GOA with DCNN classifier

Table 2 Performance estimation of the proposed model with dissimilar classifiers on FEI face dataset

Method	Precision (%)	Recall (%)	Accuracy (%)	F-score (%)
GOA-KNN	45.89	77.75	80.03	65
GOA-MSVM	67.80	70	86.52	78.92
GOA-LSTM	78.93	89.60	77.90	80.80
GOA-DBN	89.89	93.34	91.17	92.28
GOA-DCNN	**97.50**	**98.78**	**98.90**	**96.75**

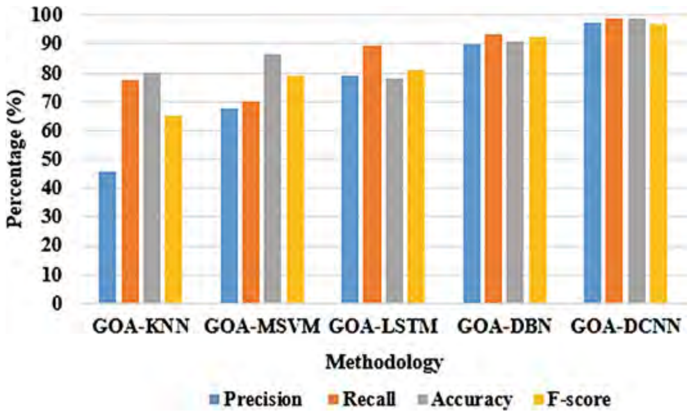


Fig. 7 Graphical depiction of the proposed model with dissimilar classifiers on FEI face dataset

achieved better performance in facial recognition in terms of recall, precision, and f-score. Compared to other classification techniques, GOA with DCNN classifier has faster learning and capacity to manage high dimensional and multivariate data in a dynamic and uncertain environment. Figure 7 presents comparison of performance between proposed and existing models in terms of recall, precision, and f-score on the FEI dataset.

4.2 Quantitative Investigation on ORL Dataset

In Table 3, the performance of the proposed GOA-DCNN model is validated with other classifiers like MSVM, KNN, LSTM, DBN, and DCNN using f-score, accuracy, recall, and precision on the ORL database. The collected ORL dataset images are cropped into the size of 32 * 32, where 80% (320) images are used for training and 20% (40) images are used for testing. From Table 3, it is observed that the proposed GOA-DCNN model has achieved a good performance in face recognition compared to other classification techniques KNN, DBN, MSVM, and LSTM. The

Table 3 Performance estimation of the proposed model with dissimilar classifiers on ORL dataset

Method	Precision (%)	Recall (%)	Accuracy (%)	*F*-score (%)
GOA-KNN	80.74	80.98	87.24	85
GOA-MSVM	88.83	89.30	92.39	90.80
GOA-LSTM	92.64	94.47	91.20	91.09
GOA-DBN	90.09	95.35	94.50	93.30
GOA-DCNN	**97.20**	**97.95**	**99**	**98.61**

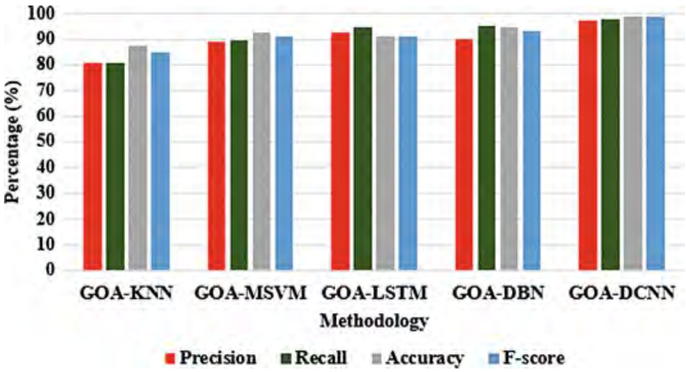


Fig. 8 Graphical depiction of the proposed model with dissimilar classifiers on ORL dataset

proposed GOA-DCNN model has achieved 99% of recognition accuracy which is better compared to other classification techniques. In Table 3, it is observed that the GOA-DCNN classification methodology results in a minimum of 4.50% and a maximum of 11.76% enhanced in recognition accuracy related to other techniques. Respectively, the proposed model GOA-DCNN has attained a good performance in face recognition using f-score, recall, and precision. Figure 8 presents a comparison of performance between proposed and existing models in terms of recall, precision, and f-score on the ORL dataset.

4.3 Comparative Study

Table 4 presents the comparative study of the proposed and existing models. Ouyang et al. [9] developed a hybrid model (IKLDA and PNN) for facial recognition. At first, IKLDA methodology was adapted to lessen the dimension of the extracted feature vectors for retaining the most relevant information about the facial images. Then, the PNN methodology was employed to solve the difficulties of the face recognition system. In this literature, the performance of the developed model was validated on three datasets AR, YALE, and ORL. In the experimental section, the developed

Table 4 Comparative study

Method	Datasets	Accuracy (%)
IKLDA + PNN [9]	ORL	97.22
CNN with pre-trained VGG model [11]	FEI face	90
Proposed model (GOA-DCNN)	ORL	99
	FEI face	98.90

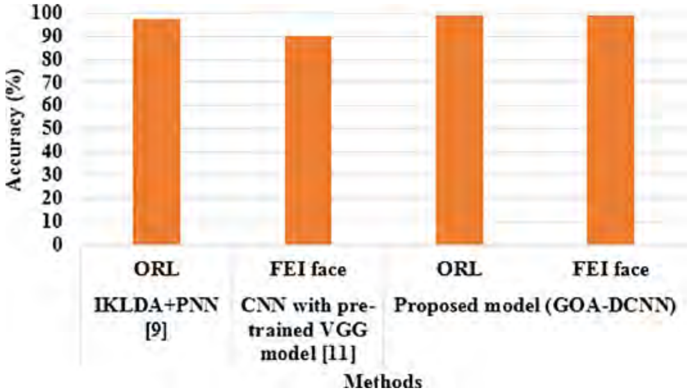


Fig. 9 Graphical comparison of proposed and existing models

model achieved a recognition accuracy of 97.22% in the ORL database. In addition, Elmahmudi and Ugail [11] utilized CNN with a pre-trained VGG model for face recognition. In this paper, the developed model achieved a 90% recognition accuracy in FEI datasets. However, the proposed GOA-DCNN model has achieved better recognition accuracy compared to these existing methods and achieved 1.78%-8.90% of improvement in recognition accuracy. A comparative study is graphically presented in Fig. 9.

5 Conclusion

A new optimization technique with a deep learning classifier (GOA-DCNN) was proposed for facial recognition. The proposed model had three phases feature extraction, optimization, and classification for face recognition. After facial image denoising, MLBP and SURF were applied to extract the features and the extracted feature values are optimized using GOA to decrease the data dimensionality. In the final phase, a DCNN classifier was applied to classify the individual person's facial images. The experimental analysis on ORL and FEI datasets showed that the proposed GOA-DCNN model achieved better performance in face recognition in terms of f-score, accuracy, recall, and precision. Compared to the existing methods, the proposed GOA-DCNN model improved recognition accuracy by up to 1.78–8.90%. In future work, a hybrid optimization technique can be included in the proposed model to further improve the performance of face recognition in the conditions like illumination conditions and light variations.

References

1. Gao G, Yu Y, Yang M, Huang P, Ge Q, Yue D (2020) Multi-scale patch based representation feature learning for low-resolution face recognition. Appl Soft Comput 106183
2. Mi JX, Sun Y, Lu J (2020) Robust supervised sparse representation for face recognition. Cognit Syst Res 62:10–22
3. Zhang G, Porikli F, Sun H, Sun Q, Xia G, Zheng Y (2020) Cost-sensitive joint feature and dictionary learning for face recognition. Neurocomputing 391:177–188
4. Orrù G, Marcialis GL, Roli F (2020) A novel classification-selection approach for the self-updating of template-based face recognition systems. Pattern Recognit 100:107121
5. Shakeel MS, Lam KM (2019) Deep-feature encoding-based discriminative model for age-invariant face recognition. Pattern Recognit 93:442–457
6. Kas M, Ruichek Y, Messoussi R (2018) Mixed neighborhood topology cross decoded patterns for image-based face recognition. Expert Syst Appl 114:119–142
7. Nikan S, Ahmadi M (2018) A modified technique for face recognition under degraded conditions. J Vis Commun Image Rep 55:742–755
8. Gan H (2018) A noise-robust semi-supervised dimensionality reduction method for face recognition. Optik 157:858–865
9. Ouyang A, Liu Y, Pei S, Peng X, He M, Wang Q (2020) A hybrid improved kernel LDA and PNN algorithm for efficient face recognition. Neurocomputing 393:214–222
10. Faraji MR, Qi X (2018) Face recognition under varying illuminations with multi-scale gradient maximum response. Neurocomputing 308:87–100
11. Elmahmudi A, Ugail H (2019) Deep face recognition using imperfect facial data. Future Gener Comput Syst 99:213–225
12. Li H, Suen CY (2016) Robust face recognition based on dynamic rank representation. Pattern Recognit 60:13–24
13. Li Y, Zheng W, Cui Z, Zhang T (2018) Face recognition based on recurrent regression neural network. Neurocomputing 297:50–58
14. Tang J, Su Q, Su B, Fong S, Cao W, Gong X (2020) Parallel ensemble learning of convolutional neural networks and local binary patterns for face recognition. Comput Methods Progr Biomed 105622
15. Dong X, Zhang H, Sun J, Wan W (2017) A two-stage learning approach to face recognition. J Vis Commun Image Rep 43:21–29
16. Roy H, Bhattacharjee D (2018) A novel local wavelet energy mesh pattern (LWEMeP) for heterogeneous face recognition. Image Vis Comput 72:1–13
17. Deng X, Da F, Shao H, Jiang Y (2020) A multi-scale three-dimensional face recognition approach with sparse representation-based classifier and fusion of local covariance descriptors. Comput Electr Eng 85:106700
18. Vijayakumar T (2019) Comparative study of capsule neural network in various applications. J Artif Intell 1(01):19–27
19. Thomaz CE, Giraldi GA (2010) A new ranking method for principal components analysis and its application to face image analysis. Image Vis Comput 28:902–913

20. Jin Z, Yang JY, Hu ZS, Lou Z (2001) Face recognition based on the uncorrelated discriminant transformation. Pattern Recogn 34:1405–1416
21. FEI face dataset. https://fei.edu.br/~cet/facedatabase.html
22. ORL dataset. https://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html

An Interactive Framework to Compare Multi-criteria Optimization Algorithms: Preliminary Results on NSGA-II and MOPSO

David F. Dorado-Sevilla, Diego H. Peluffo-Ordóñez, Leandro L. Lorente-Leyva, Erick P. Herrera-Granda, and Israel D. Herrera-Granda

Abstract A problem of multi-criteria optimization, according to its approach, can mean either minimizing or maximizing a group of at least two objective functions to find the best possible set of solutions. There are several methods of multi-criteria optimization, in which the resulting solutions' quality varies depending on the method used and the complexity of the posed problem. A bibliographical review allowed us to notice that the methods derived from the evolutionary computation deliver good results and are commonly used in research works. Although comparative studies among these optimization methods have been found, the conclusions that these offer to the reader do not allow us to define a general rule that determines when one method is better than another. Therefore, the choice of a well-adapted optimization method can be a difficult task for non-experts in the field. To implement a graphical interface that allows non-expert users in multi-objective optimization is proposed to interact and compare the performance of the NSGA-II and MOPSO algorithms. It is chosen qualitatively from a group of five preselected algorithms as members of evolutionary algorithms and swarm intelligence. Therefore, a comparison methodology has been proposed to allow the user for analyzing the graphical and numerical results, which will observe the behavior of algorithms and determine the best suited one according to their needs.

Keywords Evolutionary computation · Multi-objective optimization · Swarm intelligence

D. F. Dorado-Sevilla
Universidad de Nariño, Pasto, Colombia

D. H. Peluffo-Ordóñez · L. L. Lorente-Leyva (✉) · E. P. Herrera-Granda · I. D. Herrera-Granda
SDAS Research Group, Ibarra, Ecuador
e-mail: leandro.lorente@sdas-group.com

D. H. Peluffo-Ordóñez
e-mail: dpeluffo@yachaytech.edu.ec

D. H. Peluffo-Ordóñez
Yachay Tech University, Urcuquí, Ecuador

Corporación Universitaria Autónoma de Nariño, Pasto, Colombia

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_5

1 Introduction

Most of the optimization problems that people are commonly facing will have more than one objective simultaneously. In this type of problem, it does not allow one single solution that satisfies all the stated objectives, but rather a set of possible solutions. This set could be very extensive, and if obtaining the best results are desired, then the objective functions must be optimized to find the subset that contains the best solutions. The quality of the obtained set of solutions can vary according to the applied method by taking in count that a general rule which allows defining a method A, as better than a method B, does not exist. In this article, the development of an interactive comparative interface of NSGA-II [1] and MOPSO [2] optimization methods is described, which have been selected, after a review of the state of the art, to represent two of the most used optimization branches: the algorithms inspired by evolutionary theories and those inspired by swarm intelligence. Some applications of these algorithms are proposed for performance optimization and adaptive and intelligent routing of wireless networks using energy optimally [3, 4]. The interface developed in MATLAB allows its user to apply the mentioned algorithms to five different test problems with two objectives and bring the necessary information to conclude which method best suits the user's needs.

This paper is organized as follows: Sect. 2 describes multi-criteria optimization and metaheuristics. Section 3 shows the comparison methodology. Section 4 presents the experimental setup, and Sect. 5 depicts the results and discussion. Finally, the conclusion and the future scope are drawn in Sect. 6.

2 Multi-criteria Optimization

The multi-criteria optimization helps to reach a specific goal, looking for a set of solutions that best adapt to the proposed problem criteria. Depending on the characteristics of the problem, optimizing could involve maximize or minimize the objectives. Thus, a multi-criteria optimization problem in terms of minimization is formally defined as [5].

$$\begin{aligned} &\text{Optimize} \quad y = f(x) = (f_1(x), f_2(x), \ldots, f_k(x)) \\ &\text{s.t.} \quad g(x) = (g_1(x), \ldots, g_m(x)) \leq 0 \end{aligned} \tag{1}$$

where

$$\begin{aligned} x &= (x_1, \ldots, x_n) \in X \subseteq R^n \\ y &= (y_1, \ldots, y_n) \in Y \subseteq R^n \end{aligned}$$

The function $f(x)$ depends of k objective functions, and it can represent real numbers, binary numbers, lists, to-do tasks, etc. The decision vector x contains n

decision variables that identify each solution on the problems space X, which is the set of all the possible elements of the problem. The m restrictions for $g(x)$ limit feasible search areas, where the vector is located x. The objective vector y with k objectives belongs to the objective space Y which is the co-domain of the objective functions. The values, found after solving the objective functions with the decision variables, will be known as functionals.

To classify the best solutions of the solution set, the term dominance is used (Vilfredo Pareto, 1896), which mentions that a Pareto optimum solution is found if it reaches equilibrium, where this solution can't be improved without deteriorating another. Formally, since u and v are vectors contained in the decision space $f(u)$ and $f(v)$ then corresponding functionals, it can be said in minimization terms that:

The dominant vector will be which has the minor *functional*. Then,

$$u \prec v(u \text{ dominates a } v) \quad \text{if and only if} \quad f(u) < f(v).$$
$$v \prec u(v \text{ dominates a } u) \quad \text{if and only if} \quad f(v) < f(u).$$

Solutions are not compatible if none of the vectors dominates each other. This is:

$$u \sim v(u\text{y } v \text{ are not comparable}) \quad \text{if and only if} \quad f(u) \neq f(v) \wedge f(v) \neq f(u).$$

The optimization methods try to find, in the decision space, the set called Pareto optimum defined as $X_{\text{true}} = \{x \in X|x$ is not dominated respect a $X\}$, for succeeding, reaching the Pareto front in the objective space defined as $Y_{\text{true}} = F(X_{\text{true}})$ [6].

2.1 *Metaheuristics*

They are algorithms that modify variables trough time, guided by expert knowledge through the feasible area of the decision space in an iterative manner. The best results are obtained by applying improvements to a set of initial solutions, based on the mentioned concept of dominance, to discard the least suitable solutions [7].

2.1.1 NSGA-II (Non-dominated Sorting Genetic Algorithm)

It is a genetic algorithm chosen to represent evolutionary algorithms [8]. It is widely used in the literature for solving multi-criteria optimization problems, as shown in [9, 10]. It is considered one of the best methods for its strategies to maintain elitism and diversity in the search for optimal solutions, using Darwinian natural selection analogy, which establishes that only the fittest individuals survive and reproduce to generate a new generation with improved aspects. In Algorithm 1 is detailed the pseudocode of the algorithm proposed in [1].

Algorithm 1 Algorithm NSGA-II

```
1.Start a population:
2.  Generate an aleatory population P.
3.  Evaluate aptitude.
4.  Assign a level based on Pareto dominance - "arrange".
5.  Generate a P population as follows:
6.    Selection by binary tournament.
7.    Recombination and mutation.
8.  For i = 1 up to generation number Do:
9.    For Father and son population Do:
10.         Assign a level based on Pareto dominance and
arrange.
11.     Generate the set of not dominated fronts.
12.       Sum solutions to the next generation, starting by
        the
        hierarchically first and use the stacking factor
        (crowding) on each front.
13.   End For.
14.    Select points on the lowest front, which are out of
      the stacking factor distance.
15.   Create the next generation:
16.     Select by binary tournament.
17.     Recombination and mutation.
18. End For.
```

Initially, the algorithm randomly creates an initial population of feasible P_0 solutions of N size and then forms a Q_0 population also of N size using binary tournament selection, recombination, and mutation. The next step is to combine the two populations in such a way that from the new population $R_0 = P_0 + Q_0$, using selection, mutation, and recombination, a new P_1 population is born. The process is repeated in the following generations as shown in Fig. 1.

2.1.2 Multi-objective Particle Swarm Optimization (MOPSO)

It is an algorithm representative of swarm intelligence, popular in the literature for solving multi-criteria optimization problems, due to its good performance as can be seen [11, 12]. Besides, it is a collective intelligence algorithm with similar search behavior to the flight of starlings. These birds move in flocks and coordinates the direction and speed of their flights, so that a subgroup of the population, in response to an external stimulation, transmits clearly and immediately the state of their movement to the rest of the group. Each individual maintains a maximum susceptibility to any change in the flight of their neighbors, who react quickly to external stimulation and transmit the information to the whole group [13]. The pseudocode proposed in [2] is shown in the Algorithm 2.

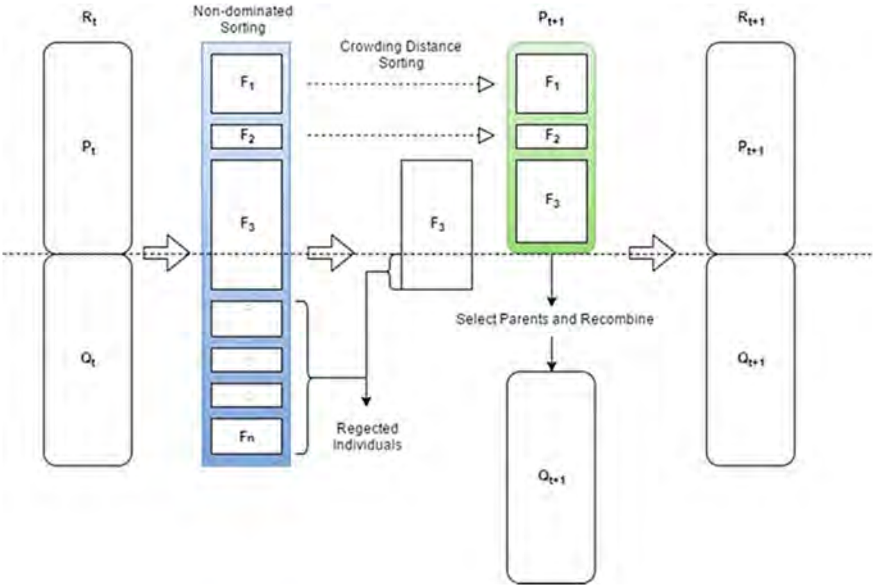


Fig. 1 NSGA-II algorithm search process

Algorithm 2 Algorithm MOPSO

```
Initialization phase:
For i = 1 to N Do:
      Apply to position_i an aleatory value in the rank [X_min,X_max]
    Assign zeros to velocity_i.
    Evaluate functional_i.
    Best_position_i equal to position_i.
    Best_functional_i equal to funcional_i.
End For
Evaluate dominance.
 Save not dominated individuals on a repository and assign a quadrille index.
Search phase:
For i_t = 1 to the maximum number of iterations Do
    For i = 0 to N Do.
        Chose the global best from repository.
        Calculate velocity_i.
        Calculate position_i.
        Keep individuals within the search space and evaluate functional_i.
        Apply the mutation factor.
         If position_i dominates the best_position_i in its memory then: Pbest_i = position_i.
      End For.
      Evaluate dominance.
      Update repository and evaluate dominance.
              Eliminate the dominated individuals from repository.
      Decrease the factor w.
End For
```

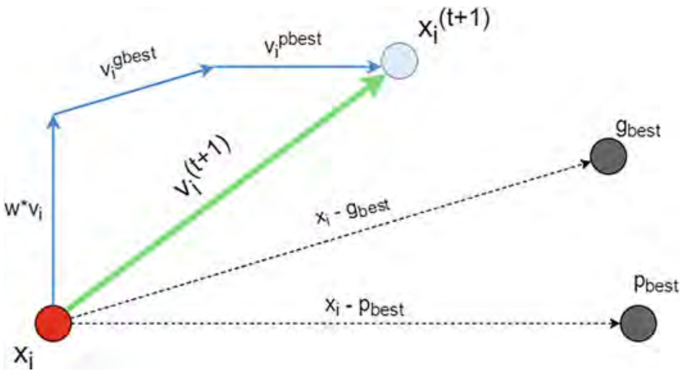


Fig. 2 Individual's change of position

The MOPSO performs the search for optimal solutions imitating the behavior of a flock in search of food. The position of each individual is obtained from the following equations.

$$v_{id}(t+1) = w * v_{id}(t) + c_1 * r_1 * [pbest_{id} - x_{id}(t)] \\ + c_2 * r_2 * [(gbest_{id} - x_{id}(t)] \tag{2}$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \tag{3}$$

where v_{id} is the speed value of individual i in the d dimension; $c1$ is the cognitive learning value; $c2$ is the global learning factor; $r1$ and $r2$ are random values uniformly distributed in the range [0.1]; x_{id} is the position of individual i in the d dimension; *pbestid* is the value in the d dimension of the individual with the best position found by individual i; and *gbestd* is the value in the d dimension of the individual in the population with the best position. The value w is important for the convergence of the algorithm. It is suggested that $c1$ and $c2$ take values in the range [1.5, 2] and w in the range [0.1, 0.5] [14]. The change of position is shown in Fig. 2.

3 Comparison Methodology

It is necessary to establish guidelines that allow understanding of how the two methods of optimization perform against certain objective functions. To evaluate its performance, four metrics are used to measure the convergence to the optimum Pareto front [15, 16].

*3.1 Error Ratio (*E*)*

This measure determines the portion of individuals in the set of solutions found by the algorithm Y_{known} that belongs to the Pareto optimal solution Y_{true}, where a value of $E = 0$ is ideal. Formally, it is defined as follows:

$$E \triangleq \frac{\sum_{i=1}^{N} e_i}{N} \tag{4}$$

$$e_i = \begin{cases} 0, & \text{if a vector of } Y_{\text{known}} \text{ is in } Y_{\text{true}} \\ 1, & \text{otherwise} \end{cases} \tag{5}$$

*3.2 Generational Distance (*DG*)*

This measure determines the solutions which are found by the Pareto optimal algorithm. Mathematically, it is defined as:

$$\text{DG} = \sqrt{\frac{\sum_{i=1}^{N} d_i^2}{N}} \tag{6}$$

where d_i is the Euclidean distance between each objective vector that belongs to the solution set found and its closest corresponding member in the real optimal Pareto front.

*3.3 Spacing (*S*)*

Verifies the dispersion of the elements of the Pareto set X found by the algorithm. Knowing the individuals at the extremes of the set, this measure proposes to use the variance of the distance between neighboring vectors of the current X set.

$$S \triangleq \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\overline{d} - d_i\right)^2} \tag{7}$$

For two objective functions, $d_i = \min_j\left(\left|f_1^i(x) - f_1^j(x)\right| + \left|f_2^i(x) - f_2^j(x)\right|\right)$ is the Euclidean distance between consecutive solutions of Y_{known} $i, j = 1, 2, \ldots, n$, where n is the number of individuals in the set.

4 Experimental Setup

To test the performance of the optimization algorithms, in [17] the test functions, proposed by Zitzler, Deb, and Thiele, are used. The functions ZDT1, ZDT2, ZDT3, ZDT4, and ZDT6 allow analyzing the behavior of the algorithms when optimizing five different Pareto fronts. The optimal fronts of the five functions are given for $g(x) = 1$.

4.1 ZDT1 Function

This function has a convex and continuous front. With $n = 30$ as the number of decision variables and x_i in the [0, 1] rank.

$$f_1(x) = x_1 \tag{8.}$$

$$g(x) = 1 + \frac{9}{n-1}\sum_{i=2}^{n} x_i,$$

$$h(f_1, g) = 1 - \sqrt{\frac{f_1}{g}},$$

$$f_2 = g(x) * h(f_1(x), g(x)) \tag{9}$$

4.2 ZDT2 Function

This function has a convex and continuous Pareto front. With $n = 30$ as the number of decision variables and x_i in the [0, 1] rank.

$$f_1(x) = x_1, \tag{10}$$

$$g(x) = 1 + \frac{9}{n-1}\sum_{i=2}^{n} x_i,$$

$$h(f_1, g) = 1 - \left(\frac{f_1}{g}\right)^2,$$

$$f_2 = g(x) * h(f_1(x), g(x)) \tag{11}$$

4.3 ZDT3 Function

This function has a discontinuous Pareto front segmented into five parts. With $n = 30$ as the number of decision variables and x_i in the [0, 1] rank.

$$f_1(x) = x_1 \tag{12.}$$

$$g(x) = 1 + \frac{9}{n-1}\sum_{i=2}^{n} x_i,$$

$$h(f_1, g) = 1 - \sqrt{\frac{f_1}{g}} - \frac{f_1}{g}\sin(10\pi f_1),$$

$$f_2 = g(x) * h(f_1(x), g(x)) \tag{13}$$

4.4 ZDT4 Function

This is a multi-modal function that has several convex and continuous Pareto fronts. With $n = 10$ as the number of decision variables, x_i in the [0, 1] rank and x_i in the [−5, 5] rank for $i = 2, \ldots, n$.

$$f_1(x) = x_1 \tag{14}$$

$$g(x) = 1 + 10(n-1) + \sum_{i=2}^{n}\big(x_i^2 - 10\cos(4\pi x_i)\big),$$

$$h(f_1, g) = 1 - \sqrt{\frac{f_1}{g}},$$

$$f_2 = g(x) * h(f_1(x), g(x)) \tag{15}$$

4.5 ZDT6 Function

This function has a non-convex and continuous Pareto front. With $n = 10$ as the number of decision variables and x_i in the [0, 1] rank.

$$f_1(x) = 1 - \exp(-4x_1) * \sin^6(6\pi x_1) \tag{16}$$

$$g(x) = 1 + 9\left[\frac{\sum_{i=2}^{n} x_i}{9}\right]^{0.125}$$

$$h(f_1, g) = 1 - \left(\frac{f_1}{g}\right)^2$$

$$f_2(x) = g(x) * h(f_1(x), g(x)) \tag{17}$$

4.6 Parameters

Tables 1 and 2 show the parameters used in the execution and simulation of NSGA-II and MOPSO algorithms.

Table 1 Parameters for execution of the NSGA-II algorithm

Parameters	
Population size	100
m	30
Number of iterations	200
Range, decision variables	[0 1]
Crossover rate	0.8
Mutation rate	0.033
Number of mutants	20

Table 2 Parameters for execution of the MOPSO algorithm

Parameters	
Population size	100
Decision variables	30
Number of iterations	200
Range, decision variables	[0 1]
w	0.5
Wdamp	0.99
Mutation rate	0.01
$c1$	1
$c2$	2

5 Results and Discussion

A graphic interface in Fig. 3 using MATLAB was developed, in which the user is allowed to apply the NSGA-II and MOPSO algorithms to the five ZDT test functions mentioned above, to obtain numerical results of the proposed performance measures. In addition, this interface shows iteratively how each algorithm tracks the best possible solutions in the search space. In order to execute the interface, it is necessary to introduce certain evaluation parameters that guide the search of each algorithm. These parameters are loaded for each test function automatically. In the following Tables 1 and 2, the parameters loaded in the interface for the two algorithms, and their respective test functions are shown.

Since the NSGA-II algorithm is based on a population for the solutions search, the N size of this population must be defined. A stop parameter is needed to stop the search, in this case, a maximum of *MaxIt* iterations. To create the population, define the number of *parents* to generate a group of *descendants*, where Pc is the crossing rate. The number of mutants is defined as $\mathrm{nm} = \mathrm{round}(\mathrm{pm1} * N)$, where Pm1 is the mutation rate. Table 3 shows the parameters used in the NSGA-II to evaluate the five defined test functions.

Like the previous algorithm in the MOPSO, you must define the N size of the individuals that will take flight in search of optimal solutions and a *MaxIt* stop parameter. As for the search procedure, the change of individual's position is fundamental, the parameters w, $c1$, $c2$ defined in Eq. (2) must be defined. To generate diversity,

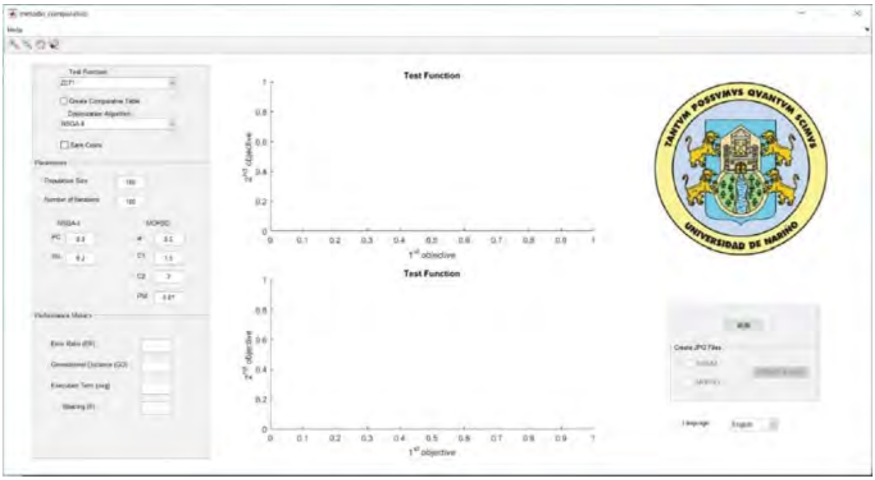


Fig. 3 Interactive comparator developed in a graphical interface using MATLAB. More information and MATLAB scripts available at: https://sites.google.com/site/degreethesisdiegopeluffo/interactive-comparator

Table 3 Evaluation parameters NSGA-II

Parameters	ZDT1	ZDT2	ZDT3	ZDT4	ZDT6
N	100	100	100	100	100
MaxIt	500	500	500	500	500
Pc	0.67	0.63	0.63	0.67	0.67
Pm1	0.33	0.33	0.33	0.33	0.33

Table 4 Evaluation parameters MOPSO

Parameters	ZDT1	ZDT2	ZDT3	ZDT4	ZDT6
N	100	100	100	100	100
MaxIt	100	100	100	100	200
W	0.5	0.5	0.5	0.5	0.5
$c1$	1.5	1.5	1.5	1.5	1.5
$c2$	2	2	2.5	2	2.5
Pm2	0.1	0.1	0.5	0.1	0.5

the algorithm simulates turbulence in flight using a mutation operator. In each iteration, all individuals are assigned a mutation probability (Pm2). Table 4 shows the parameters used to evaluate the five defined test functions.

To allow the user to conclude the results easily, the "Create Comparative Table" function is created in the interface, which executes automatically each algorithm ten times in a row and creates an excel file that contains a table with the numerical results of the performance measures for each execution.

5.1 ZDT1 Results

In Table 5, the results obtained with the ZDT1 execution are presented, where it can be evidenced that the performance of the MOPSO when optimizing a problem with continuous and convex front is better than the NSGA-II, where the *E* and *DG* metrics are very close to the real Pareto front. It is also noted that the swarm intelligence algorithm is much faster. It is also shown that according to the *S* metric, the NSGA-II has a better dispersion than the MOPSO.

Figures 4 and 5 show the Pareto front of the ZDT1 function and the solutions distribution.

In the previous figures, it is shown that the analysis made in the execution of the ZDT1 function with the developed interface, where a user will be able to make in the interface, the analysis for the rest of the presented problems (ZDT2, ZDT3, ZDT4, and ZDT6). And obtain in this way, the behavior of each algorithm is used against

Table 5 ZDT1 results

Execution	E_NSGA	E_MOPSO	DG_NSGA	DG_MOPSO	Time NSGA	Time MOPSO	S_NSGA	S_NSGA
1	1.000	0.000	0.012	0.000	318.006	44.153	0.007	0.035
2	0.990	0.000	0.013	0.001	384.542	42.192	0.006	0.023
3	0.970	0.000	0.016	0.001	466.853	40.592	0.009	0.023
4	1.000	0.000	0.013	0.001	517.460	40.701	0.009	0.018
5	1.000	0.000	0.021	0.001	574.547	41.070	0.016	0.022
6	1.000	0.000	0.013	0.000	645.206	42.438	0.007	0.020
7	1.000	0.000	0.018	0.001	712.545	41.705	0.013	0.021
8	0.990	0.000	0.013	0.000	777.596	40.710	0.007	0.021
9	0.970	0.000	0.013	0.001	893.973	42.343	0.007	0.018
10	1.000	0.000	0.015	0.001	1196.094	39.570	0.008	0.019
Average	**0.992**	**0.000**	**0.015**	**0.001**	**648.682**	**41.547**	**0.009**	**0.022**

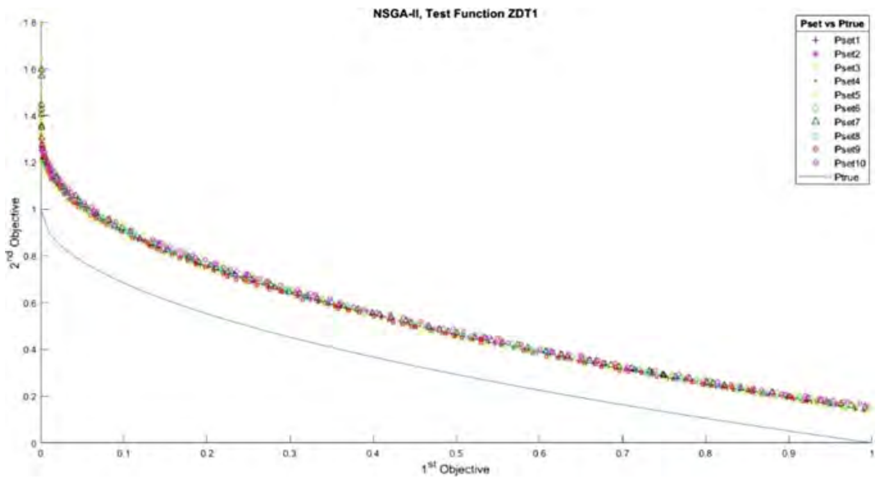


Fig. 4 NSGA-II solutions with ZDT1 optimization versus continuous convex Pareto front

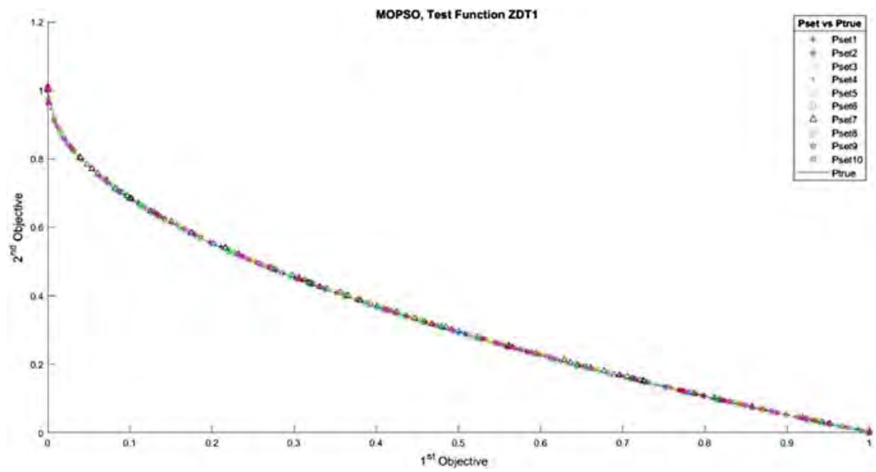


Fig. 5 MOPSO solutions with ZDT1 optimization versus continuous convex Pareto front

all the given conditions. Determine the algorithm that has the best performance and obtains the best results.

6 Conclusion and Future Scope

The results thrown by an optimization algorithm can reach different quality levels, depending on the variation of the evaluation parameters. Therefore, the comparative studies of multi-criteria optimization methods available in the literature, limit the reader performance analysis of the algorithms, by basing their experiments on fixed evaluation parameters.

The development of the interactive comparative interface offers the possibility of easily carrying out an optimization process in an intuitive way. This interactive interface allows the user to choose the optimization algorithm and the test function that he wants to optimize according to the Pareto front of interest and establish a man–machine communication, through several inputs defined as parameters of evaluation, which can be modified to obtain a dynamic graphic and numerical response.

Using the mapping of the objective functions, the set of solutions found in the target space at the end of each iteration can be observed, allowing them to observe the search procedure in a dynamic way. The user can accurately measure the performance of the algorithms by evaluating the numerical results of the performance measures, which are on the final set of solutions found. For the above, a not necessarily expert user will have a greater understanding of the optimization process and will choose the appropriate method more easily, according to their needs. As future work, it is proposed to expand the number of optimization algorithms and add new test functions such as three objective functions and so on.

Acknowledgements The authors are greatly grateful for the support given by the SDAS Research Group (https://sdas-group.com/).

References

1. Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 6(2):182–197
2. Coello Coello C, Lechuga M (2002) MOPSO: a proposal for multiple objective particle swarm optimization. In: Proceedings of the 2002 Congress on evolutionary computation, CEC'02, pp 1051–1056
3. Rahimunnisa K (2019) Hybridized genetic-simulated annealing algorithm for performance optimization in wireless Adhoc network. J Soft Comput Paradigm 1(01):1–13
4. Shakya S, Pulchowk LN (2020) Intelligent and adaptive multi-objective optimization in WANET using bio inspired algorithms. J Soft Comput Paradigm 2(01):13–23
5. Deb K, Agrawal S, Pratap A, Meyarivan T (2000) A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-II. In: International conference on parallel problem solving from nature. Springer, pp 849–858
6. Veldhuizen DAV, Lamont GB (2000) Multiobjective evolutionary algorithms: analyzing the state-of-the-art. Evolut Comput 8(2):125–147
7. Melián B, Pérez JAM, Vega JMM (2003) Metaheurísticas: Una visión global. Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial 7(19)
8. Kannan S, Baskar S, McCalley JD, Murugan P (2009) Application of NSGA-II algorithm to generation expansion planning. IEEE Trans Power Syst 24(1):454–461
9. Kwong WY, Zhang PY, Romero D, Moran J, Morgenroth M, Amon C (2014) Multi-objective wind farm layout optimization considering energy generation and noise propagation with Nsga-II. J Mech Des 136(9):091010
10. Lorente-Leyva LL et al (2019) Optimization of the master production scheduling in a textile ındustry using genetic algorithm. In: Pérez García H, Sánchez González L, Castejón Limas M,

Quintián Pardo H, Corchado Rodríguez E (eds) HAIS 2019. LNCS 11734, Springer, Cham, pp 674–685
11. Robles-Rodriguez C, Bideaux C, Guillouet S, Gorret N, Roux G, Molina-Jouve C, Aceves-Lara CA (2016) Multi-objective particle swarm optimization (MOPSO) of lipid accumulation in fed-batch cultures. In: 2016 24th Mediterranean conference on control and automation (MED). IEEE, pp 979–984
12. Borhanazad H, Mekhilef S, Ganapathy VG, Modiri-Delshad M, Mirtaheri A (2014) Optimization of micro-grid system using MOPSO. Renew Energy 71:295–306
13. Marro J (2011) Los estorninos de san lorenzo, o cómo mejorar la eficacia del grupo. Revista Española De Física 25(2):62–64
14. Parsopoulos KE, Vrahatis MN (2002) Recent approaches to global optimization problems through particle swarm optimization. Nat Comput 1:235–306
15. Van Veldhuizen DA, Lamont GB (1999) Multiobjective evolutionary algorithm test suites. In: Proceedings of the 1999 ACM symposium on applied computing. ACM, pp 351–357
16. Eberhart R, Kennedy J (1995) A new optimizer using particle swarm theory. In: Micro machine and human science. In: Proceedings of the Sixth International Symposium on MHS'95. IEEE, pp 39–43
17. Zitzler E, Deb K, Thiele L (2000) Comparison of multi-objective evolutionary algorithms: empirical results. Evolut Comput 8:173

A Dynamic Programming Approach for Power Curtailment Decision Making on PV Systems

Yasmany Fernández-Fernández, Leandro L. Lorente-Leyva, Diego H. Peluffo-Ordóñez, and Elia N. Cabrera Álvarez

Abstract The new grid codes for large-scale photovoltaic power plants require power curtailment despite the variation of solar irradiance. This power curtailment is been developed considering one reference of active power. However, this value is chosen according to the demand, but it is not considering other aspects as solar irradiance or cloudiness. Therefore, this article presents a novel approach to tackle this issue. For this, stochastic dynamic programming is considered to optimize the decision of the power reference every hour considering the solar irradiance and cloudiness during different stages of the day. The results obtained are compared with the performance of the photovoltaic power plant, and it is a referential approach that uses the maximum power point tracking algorithms for the construction of referential power intervals over longer time intervals.

Keywords Maximum power point tracker (MPPT) · Photovoltaic power plant (PVPP) · LS-PVPP

Y. Fernández-Fernández
Universidad Politécnica Estatal del Carchi, Tulcán, Ecuador
e-mail: yasmany.fernandez@upec.edu.ec

Y. Fernández-Fernández · L. L. Lorente-Leyva (✉) · D. H. Peluffo-Ordóñez
SDAS Research Group, Ibarra, Ecuador
e-mail: leandro.lorente@sdas-group.com

D. H. Peluffo-Ordóñez
Corporación Universitaria Autónoma de Nariño, Pasto, Colombia

Yachay Tech University, Urcuquí, Ecuador
e-mail: dpeluffo@yachaytech.edu.ec

E. N. C. Álvarez
Universidad de Cienfuegos, Cienfuegos, Cuba
e-mail: elita@ucf.edu.cu

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_6

1 Introduction

The photovoltaic power plants of the LS-PVPP type operation represent a problem for operators. Normally, there is a variability in the behavior of solar irradiance mainly during the day. Using photovoltaic inverters, it has been possible to integrate voltage and frequency support and active and reactive power control [1].

One of the issues to solve is to find an adequate value of the active power that the LS-PVPP must supply to comply with the grid codes despite the variation of solar irradiance. The reduction of the active power to a fixed power is called by the grid operators as "power curtailment." Currently, this curtailment is only performed due to demand and grid behavior. The decision of this value is performed by the grid operator, and the LS-PVPP must supply this power at any moment of the day. However, this decision is not developed considering solar irradiance variability, temperature, costs or any other factor that could affect directly to the operation of the LS-PVPP.

For the management of active power in renewable energy, some optimization techniques have been used [2, 3]. One of these is the stochastic dynamic programming as it helps to add uncertainty scenarios due to the variability of the input energy [4]. For instance, Papavasiliou et al. [5] use this technique to optimize the curtailment of a wind power plant according to the grid response, the technical requirements and the variation of wind power. The challenges of the stochastic dynamic programming addressed by this research are (i) the appropriate selection of the weighing scenarios and (ii) the computational intractability of the resulting problem.

In the photovoltaic field, this optimization technique has not been used for power curtailment as it is a new requirement asked by the grid codes as the case of Puerto Rico [6]. Thus, the aim of this work is to find the optimal power point that the LS-PVPP has to supply considering solar irradiance, cloudiness and in hourly basis using a stochastic dynamic programming approach.

The rest of the paper is structured as follows: Sect. 2 presents an explanation of stochastic programming. Section 3 shows the formulation of the problem considering the uncertainty of cloudiness and solar irradiance during different stages of the day. In Sect. 4, a scenario is tested for the given algorithm and the results are presented. Finally, the conclusion and the future scope are described in Sect. 5.

2 Background and Problem Statement

Linear models represent the basis for formulating linear problems, in general, a basic linear model has the following form:

$$\begin{aligned} &\min_x \mathbf{c}^T\mathbf{x} \\ &\text{st} \quad \mathbf{Ax} = \mathbf{b} \\ &\qquad \mathbf{x} > \mathbf{o_n}, \end{aligned} \tag{1}$$

where $\mathbf{x}$, $\mathbf{c}$, $\mathbf{b} \in \mathrm{R}^{\mathbf{n}}$, $\mathbf{A} \in \mathrm{R}^{\mathbf{nxn}}$, and $0_{\mathbf{n}}$ is a n-dimensional all-zeros vector.

To make a forecast, it is necessary to have quantitative information on-demand behavior over time, with analysis using classic statistical techniques such as ARIMA, Holt-Winters, among others, the most widely used to predict their behavior [7, 8]. Through this representation, thousands of people doing research have been able to represent their problems through a standard mathematical model. The modeling of dynamic programming problems has a considerable focus on complexity with respect to linear problems [9].

The challenge of dynamic programming is decision making. Two general procedures in the deterministic sampling or stochastic sampling approaches are important for considering the problem to solve. The first approach involves the representation of the stochastic process through a decision tree from which an associated equivalent deterministic problem is obtained that is solved by an optimization technique that may or may not lead to a discretization of the problem with a deterministic approach. The second approach assumes the complete tree of the problem which cannot be enumerated, so it is necessary to approximate successive sampling that according to the bibliography can be done with two approaches of exterior and interior sampling [7].

Another approach is presented in [10], which combines clustering algorithms with dynamic programming for designing a demand response strategy applied to residential electric water heaters. In [11], authors develop a multi-objective optimization algorithm to manage energy in radio networks and reduce consumption by minimizing communication delays. Other researchers [12] use some methods to optimize the performance of a wireless network to retain the energy level of the devices.

In [8], authors use a real-time dynamic economic load dispatch integrated with renewable energy curtailment to detect a minimum amount of supply–demand mismatch in advance, managing reliably for a considerable time horizon. One last experience using dynamic models is presented in [13] where the authors propose an improved multistage converter topology intended for single-phase solar rooftop applications with battery storage.

Stochastic dynamic programming refers to the existence of a probability to obtain the results of a near state.

By expanding the relationships represented in Fig. 1 to include all states and possible decisions in all stages, one gets what is often called the decision tree, which, if not very large, provides a useful way to summarize the different possibilities. Let us define $f_n(S_n, x_n)$ and $f^*_{n+1}(S_{n+1}, x_{n+1})$, respectively, as the instantaneous and optimal objective function value in terms of the state S and the independent variable x. Due to the probabilistic scheme, the relationship between $f_n(S_n, x_n)$ and $f^*_{n+1}(S_{n+1}, x_{n+1})$ is more complicated than the deterministic case, and it will depend on the general form of the objective function.

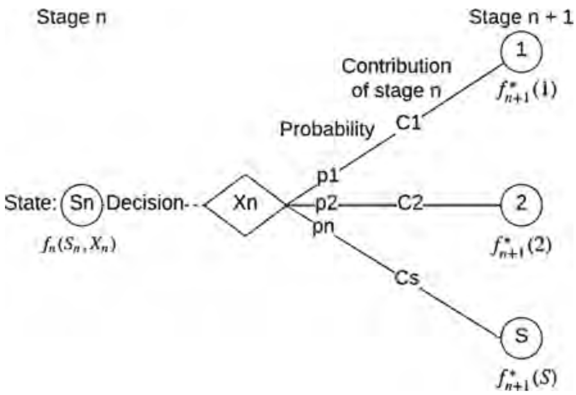


Fig. 1 Probabilistic dynamic problem scheme

3 Formulation of the Proposed Mathematical Model

PV is used as a basic unit for large-scale photovoltaic power plants. This generator controls the active power at every instant according to the solar irradiance. Commonly, the PV generator follows the maximum power possible to get at the given conditions of solar irradiance and ambient temperature. The control is called the maximum power point tracker (MPPT). Different algorithms have been developed to track this point at any instant [14].

Because grid code requirements are not necessary to track the maximum power point at each instant, a referential power can be used. To optimize this power reference considering solar irradiance and cloudiness, the approach of this model is based on a decision tree of stochastic dynamic programming.

For this model, the day will be divided into several parts that are the stages of the model (i). Each stage i represents a constant time in which the system calculates the MPPT, if the MPPT is calculated every 5 min, then $i \in \{1, \ldots, n\}$ corresponds to 5, …, nt minutes where the system collects MPPT data, as shown in Eq. (2).

$$i = n \to nt \text{ minutes} \tag{2}$$

The range of active power values that can be signed in the stage i can be given as follows:

$$\delta_i \left[\theta_i, \theta_{i+1}\right] \tag{3}$$

where θ_i the maximum power point in stage i and θ_{i+1} is the maximum power point in stage $i + 1$. Considering Eq. (3), the optimal power function in stage i (ρ_i) can be defined depending on the range of active power and also on the optimized value of active power for the next stage (θ_{i+1}).

$$\rho_i(\delta_i)\rho(\delta_i, \theta_{i+1}) \tag{4}$$

where δ_i, θ_i is the contribution to the objective function of stages i, $i + 1$, …, n. If the system is in the state δ_i in stage i, the immediate decision is θ_i and from now on optimal decisions are taken in the form:

$$\rho_i^*(\delta_i) = \max_{\rho} \rho(\delta_i, \theta_i)$$
$$\text{s.t.} \quad \theta_i \geq 0 \tag{5}$$

Then, the average accumulated contribution of power θ_i up to stage $i = n$ is expressed as follows:

$$C(\theta_i) = \frac{1}{n}\sum_{i=1}^{n} \theta_i \tag{6}$$

The decision of the maximum value possible between the current power δ_i and power θ_{i+1} can be expressed as follows:

$$\rho_i^*(\delta_i, \theta_i) = \sup\{[\delta_i, \theta_{i+1}]\} \tag{7}$$

So, the optimal decision for each stage i is written in the form:

$$\rho_i^{**}(\delta_i) = \max_{\rho*}\{kb_i(C(\theta_i) + \rho_i^*(\delta_i, \theta_i))\} \tag{8}$$

where kb_i is the clarity index in stage i. This index is calculated by daily accumulated values in two temporary resolutions, hourly and daily. The time series are grouped in ranges of $kb = 0.2$ assuming that the behavior of the fluctuations of the radiation for each range is very similar. Table 1 shows this classification based on the value of the clarity index [15].

The result of this decision is a recursive function, Eq. (8), which considers the value of the accumulated referential power that follows each stage. This model is considering the possible maximum power of the solar irradiance at each stage and the optimal possible power considering different sky conditions. The recursive form of this function is illustrated in Fig. 2.

Table 1 Clarity index for different sky types

kb values	Type of sky
$kb \leq 0.2$	Completely covered
$0.2 \geq kb \leq 0.4$	Mostly covered
$0.4 \geq kb \leq 0.6$	Partially covered
$0.6 \geq kb \leq 0.67$	Mostly clear
$kb \geq 0.67$	Completely clear

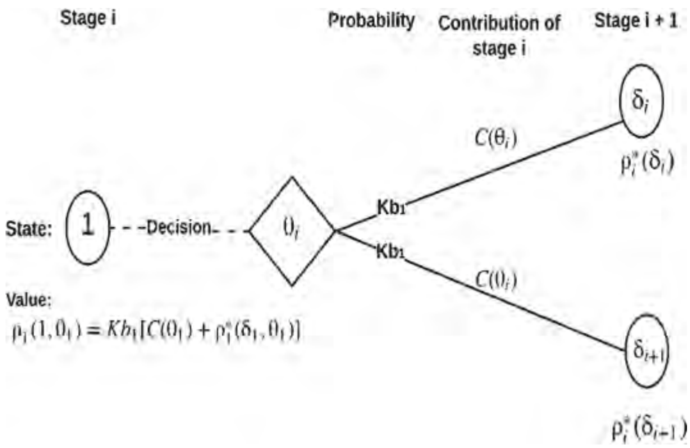


Fig. 2 Representation of the problem through a stochastic dynamic programming model

4 Results and Discussion

To study the model proposed, real data from a photovoltaic power plant located in the north of Ecuador is considered. Figure 3 shows the maximum power supplied by the PVPP at different solar irradiance taken each 5 min for a single day in 2018.

The proposed model and the data from the PVPP are introduced in the software "Wolfram Mathematica 11.2." To simulate the model, the day (from 6h30 to 18h30)

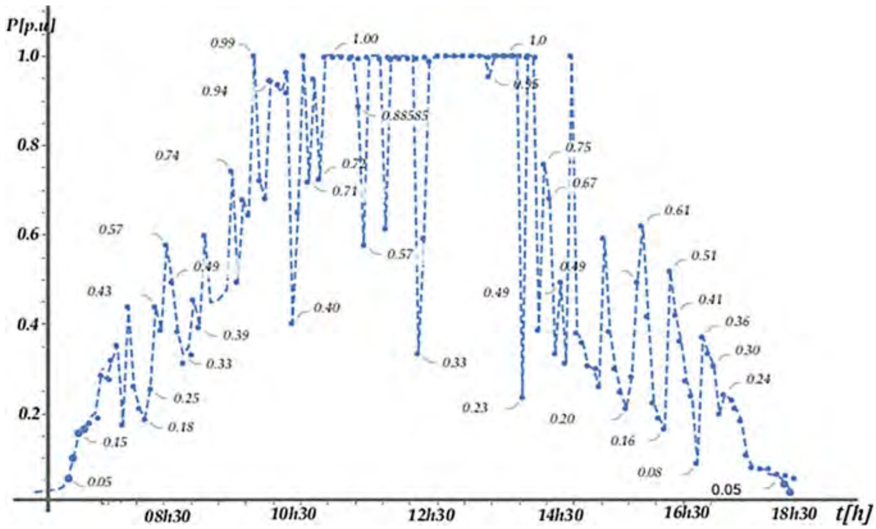


Fig. 3 Active power supplied by a photovoltaic power plant located at the north of Ecuador

is divided into six parts of 2 h each. Figure 3 shows the representation of the data of a common day used in the proposed simulation. Each value represents the maximum power point determined every 5 min by a search algorithm for the maximum power point installed in an inverter.

The initial simulation parameters for a stage are set as follows: $i = 24$ such that for each i there is a 5 min interval where the AC–DC inverter algorithm determines the maximum power point data that is used to estimate the referential power curtailment $\rho_i^{**}(\delta_i)$. For a time interval of 5 s * 24 states, the analysis is made on what should be the referential working power for the maximum use of the weather conditions, setting a clarity index of kb = 0.78 for the next two hours (120′). In summary, the fundamental parameters used to obtain a referential power for each state are $\delta_1 = \theta_1$ (initial condition) where θ_1 is the first MPP captured by the investor, the number of states to use for prediction (in this case δ_{24}) such that for each i between 1 and 24 to obtain a referential power value that constitutes the working power $\rho_i^{**}(\delta_{i+1})$ in the next time interval.

For instance, the maximum power that the PVPP generates from 8:30 to 10:30 is plotted in Fig. 4. Applying the optimization tool created and considering the cloudiness, the new active power reference for these two hours is 0.5 p.u as it is illustrated in Fig. 4.

The following diagram in the algorithm in Table 2 shows in detail how the dynamic algorithm works to perform the power curtailment in such a way that once the overall optimal solution is known, any partial solution that involves only a part of the stages is also an optimal solution [16].

Considering this, the results for the complete day are illustrated in Fig. 5. The active power instead of being following the maximum power point follows a new

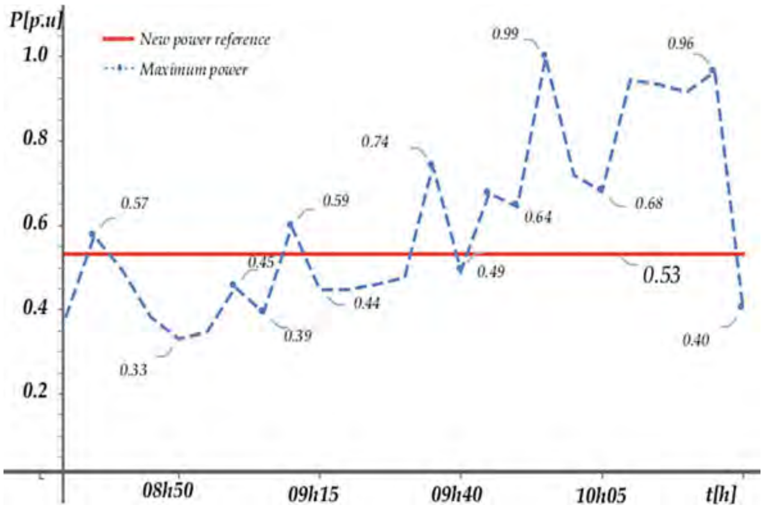


Fig. 4 Optimized active power from 8h30 to 10h30

Table 2 Algorithm for calculating the power curtailment

Algorithm	
Init in: Stage 1	• First time interval: $t = 5', i = 1$ • Initial decision: State $\delta_1 = \theta_1$ (Initial MPP value) • Save the historical power value by an accumulative mean: $C(\theta_1) = \theta_1$ • Find the optimal initial decision for next stage: $\boldsymbol{\rho_1^{**}(\delta_1, \theta_1) = Kb_1\left[C(\theta_1) + \rho_1^{*}(\delta_1, \theta_2)\right] = \delta_2}$
Stage 2	• Second time interval: $t = 10', i = 2$ • State 2 analysis: $\boldsymbol{\delta_2 = \rho_1^{**}(\delta_1, \theta_1)}$ (Optimal Decision in Last Stage) • Save the historical power value by an accumulative mean: $C(\theta_2) = \frac{(\theta_1+\theta_2)}{2}$ • Optimal decision for next stage: $\boldsymbol{\rho_2^{**}(\delta_2, \theta_2) = Kb_1\left[C(\theta_2) + \rho_1^{*}(\delta_1, \theta_3)\right] = \delta_3}$
⋮	⋮
End: Stage n	• General time interval: $\boldsymbol{t = 5' \cdot i, i = n}$ • State n analysis: $\boldsymbol{\delta_n = \rho_{n-1}^{**}(\delta_{n-1}, \theta_{n-1})}$ (Optimal Decision in Last Stage) • Save the historical power value by an accumulative mean: $C(\theta_n) = \frac{1}{n}\sum_{i=1}^{n}\theta_i$ • Optimal Decision for Last Stage: $\boldsymbol{\rho_n^{**}(\delta_n, \theta_n) = Kb_1\left[C(\theta_n) + \rho_1^{*}(\delta_n, \theta_{n+1})\right] = \delta_{n+1}}$

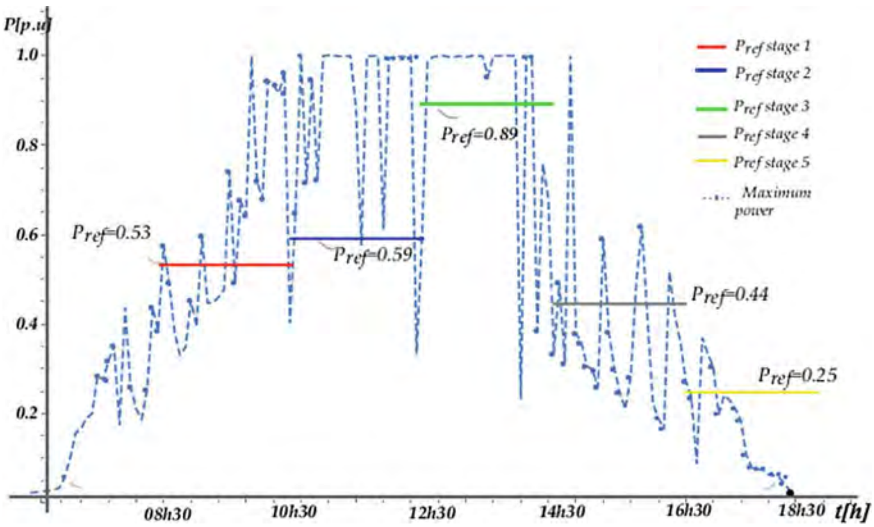


Fig. 5 Optimized active power generated for one day versus the maximum possible power

reference. This reference has been calculating by the model which considers the behavior on each part of the day and the cloudiness expected. As can be seen, when the PVPP uses this reference, the active power reduces its intermittent behaviors during the day.

The conditions that have been considered for decision making have been analyzed mainly considering that the state of the subsequent stages does not fully determine a decision policy of the current state, for this reason, it is considered of vital importance, the use of the clarity index (*Kb*) to conclusively determine the next state, which represents the optimal decision policy for each stage.

5 Conclusion and Future Scope

An optimization approach is to calculate the active power reference when power curtailment is employed. For this solution, stochastic programming has been handled. The optimization has been used considering one day data of a real photovoltaic power plant. The results determine the intermittent performance of the PVPP, which is reduced in each part of the day.

In the future scope, the optimization tool is considered, not only for solar irradiance and cloudiness but also for the economic aspects.

Acknowledgements The authors are greatly grateful for the support given by the SDAS Research Group. https://sdas-group.com/.

References

1. Cabrera-Tobar A, Bullich-Massagué E, Aragüés-Peñalba M, Gomis-Bellmunt O (2016) Review of advanced grid requirements for the integration of large scale photovoltaic power plants in the transmission system. Renew Sustain Energy Rev 62:971–987. https://doi.org/10.1016/j.rser.2016.05.044
2. Morais H, Kádár P, Faria P, Vale ZA, Khodr HM (2010) Optimal scheduling of a renewable micro-grid in an isolated load area using mixed-integer linear programming. Renew Energy 35(1):151–156. https://doi.org/10.1016/j.renene.2009.02.031
3. Fernández YF, Tobar AC, Peluffo-Ordóñez DH, Manosalvas TS, Miranda R (2019) Optimization-based algorithms applied in photovoltaic systems. RISTI Revista Iberica De Sistemas E Tecnologias De Informacao 2019(E22):242–255
4. Kaewpasuk S, Intiyot B, Jeenanunta C (2017) Stochastic unit commitment model for power system with renewable energy. In: 2017 International Electrical Engineering Congress (iEECON), pp 1–4 (2017). https://doi.org/10.1109/IEECON.2017.8075781
5. Papavasiliou A, Oren SS, O'Neill RP (2011) Reserve requirements for wind power integration: a scenario-based stochastic programming framework. IEEE Trans Power Syst 26(4):2197–2206. https://doi.org/10.1109/TPWRS.2011.2121095
6. Gevorgian V, Booth S (2013) Review of PREPA technical requirements for interconnecting wind and solar generation. Technical report, NREL, USA. https://doi.org/10.2172/1260328

7. Diniz AL, Maceira MEP (2013) Multi-lag benders decomposition for power generation planning with nonanticipativity constraints on the dispatch of LNG thermal plants. World Sci Ser Finance, Stochast Program 443–464. https://doi.org/10.1142/9789814407519_0016
8. Sasaki Y, Tsurumi T, Yorino N, Zoka Y, Rehiara AB (2019) Real-time dynamic economic load dispatch integrated with renewable energy curtailment. J Int Council Electr Eng 9(1):85–92. https://doi.org/10.1080/22348972.2019.1686861
9. Powell WB (2011) Approximate dynamic programming: solving the curses of dimensionality, 2nd edn. Wiley
10. Alvarez MAZ, Agbossou A, Cardenas A, Kelouwani S, Boulon L (2019) Demand response strategy applied to residential electric water heaters using dynamic programming and k-means clustering. IEEE Trans Sustain Energy 11(1):524–533. https://doi.org/10.1109/TSTE.2019.2897288.25
11. Haoxiang H (2019) Multi-objective optimization algorithm for power management in cognitive radio networks. J Ubiquit Comput Commun Technol 2:97–109. https://doi.org/10.36548/jucct.2019.2.004
12. Smys S, Raj JS (2019) Performance optimization of wireless Adhoc networks with authentication. J Ubiquit Comput Commun Technol 2:64–75. https://doi.org/10.36548/jucct.2019.2.001
13. Valsala KD, Premkumar K, Beevi AB (2019) Development of battery intervention power supply for solar roof top installations. Environ Progr Sustain Energy 38(2):570–583 (2019). https://doi.org/10.1002/ep.12958
14. Eltawil MA, Zhao Z (2013) MPPT techniques for photovoltaic applications. Renew Sustain Energy Rev 25:793–813. https://doi.org/10.1016/j.rser.2013.05.022
15. Dai Q, Fang X (2014) A simple model to predict solar radiation under clear sky conditions. Adv Space Res 53(8):1239–1245. https://doi.org/10.1016/j.asr.2014.01.025
16. Sniedovich M (1978) Dynamic programming and principles of optimality. J Math Anal Appl 65:586–606. https://doi.org/10.1016/0022-247X(78)90166-X

Arabic Braille Numeral Recognition Using Convolutional Neural Networks

Shurouq Alufaisan, Wafa Albur, Shaikha Alsedrah, and Ghazanfar Latif

Abstract Braille is a system that is designed to assist visually impaired individuals to acquire information. It consists of raised dots arranged in a cell of three rows and two columns. Visually impaired individuals rely on the sense of touch to read and write. However, it is difficult to memorize the arrangement of dots that compose a character. This research aims to design an application that recognizes and detects Arabic braille numerals and convert it to plain text and speech by implementing convolutional neural network variation Residual Network (ResNet). A new dataset was collected by capturing Arabic braille numerals using smartphone cameras. The recognition accuracy for Arabic braille numerals achieved 98%, taking into accountability different light and distance conditions.

Keywords Braille recognition · Deep learning · Arabic braille numerals classification · Convolutional neural network · Residual network

1 Introduction

Currently, everywhere in the world operate with the data as it is the most valuable part of our society, simply known as information. Collecting data successfully is the first step to operate effectively and make decisions efficiently. Sharing knowledge can be interpreted as communication, and it is one way of acquiring the needed information. In addition, reading is an important factor to learn and to obtain the information needed to prosper in our society. However, for visually impaired and deaf individuals, it is impossible to acquire information with plain texts, and there is no means of sharing information through communication with sighted individuals [1]. Therefore, a system, known as the braille system, was designed for visually impaired groups of individuals to access and receive information. With such a system, it is now possible to make decisions, operate upon events effectively, and communicate with sighted individuals.

S. Alufaisan · W. Albur · S. Alsedrah · G. Latif (✉)
College of Computer Engineering and Sciences, Prince Mohammad Bin Fahd University, Al Khobar, Saudi Arabia
e-mail: glatif@pmu.edu.sa

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_7

Braille system can be used as a mean of communication to share information. It is used for sighted individuals who wish to communicate with visually impaired individuals through written informational communication. Visually impaired individuals may prosper in our society and can play a significant role in our world. Thus, it is important to have a mean of communication between sight and visually impaired individuals to share information and to learn from each other.

Braille is a system that depends on the sense of touch. It is a system that enables visually impaired individuals to write and read with the help of the touch sense. It uses a series of raised dots that are used to read with the personal sense of touch. The language consists of six dots arranged in a rectangular shape. The dots can be arranged in any of the six position to compose a word, letter, number, or a special character. Braille system can be used to write different languages such as Arabic or English. In addition, the system can be used to write musical compositions and mathematical notations. Reading braille texts, for both Arabic and English, are read by moving the index finger from left to right.

Reading braille language with no previous experience can be difficult at first. To be able to read it effectively, the sense of touch must be trained first while making all the other senses unused. Also, users must memorize the positions of the dots and make sure what each composition of dots mean. Of course, not all individuals were born visually impaired. Some might go through events in life that make a person loses his/her sight, such as a chemical accident. In addition, there are situations, where a person might want to learn the language to teach it or to communicate with a visually impaired individual. Being a beginner in learning, this system can be hard at first. Ideally, if there exists a technology that can make reading and writing braille language easier, then it would be beneficial for the users to learn faster and to acquire the information that is needed efficiently. Such a system would have to translate the braille language into text and speech to be familiar with the system.

Therefore, the objective of this research is to develop an Arabic braille numeral system detection and recognition based on a deep learning algorithm. The proposed system will have the ability to recognize Arabic braille numeral images captured by cameras and to be processed by CNN variation, residual network [2, 3]. The goal of this work is to assist a person in reading and learning Arabic braille numeral. The ultimate goal is to have a pi camera that is integrated with the application, to scan and recognize printed Arabic braille numeral to translate it and display it to the user in plain text and speech [4–6]. This aims to improve the learning process of the Arabic braille system to acquire the required information for visually impaired users which lead to improve their daily life activities [7]. In addition, to help those who wish to communicate with visually impaired individuals.

2 Literature Review

Classifying and recognizing Arabic braille are significant to assist visually impaired individuals to learn and to obtain the necessary knowledge. In addition, to assist who wishes to communicate with a visually impaired individual. Extensive research has been conducted in the area of classifying and recognizing Arabic braille scripts. In [8], the authors proposed the use of find contour and artificial neural network for braille character recognition. Their method consists of preprocessing the image to prepare it for the process of the finding contour to get the black dots on the image for different datasets consisting of tilted images. The authors used segmentation to read the area of the braille cell. Later on, the artificial neural network was conducted as the final step to have the system learn by feeding it data input to obtain the desired data output value. The method achieved 99% of titled images of $-1°$ to $1°$. The authors showed that the level of accuracy decreased when the image is tilted more than $1°$.

In [9], the author's dataset consisted of braille documents with the color of green and yellow that included dots in one side of the document. The author's method was to use image processing techniques to recognize Arabic braille characters. The authors preprocessing steps include: converting the image to grayscale, filtering the image, applying local thresholding for green braille documents, applying adaptive thresholding for yellow braille documents, segmentation, and extracting features. The authors described that Arabic braille characters were successfully recognized. Afterward, the authors proposed to convert Arabic braille character to binary strings which are converted into ASCII code to obtain the correct Arabic translation. The method achieved 98.04–100% for green braille document and 97.08–99.65% for yellow braille document. In [10], the authors proposed performing image preprocessing techniques to prepare the image for feature extraction. In the feature extraction step, the authors computed centroids of dots in the image to extract the relevant information from the image. Afterward, the authors aligned the coordination by applying many operations to rotate the centroids to align the page and the braille dots. Finally, the authors were able to recognize braille cells by grouping the dots to acquire a combination of letters and words. The method achieved between 94 and 99% braille cell recognition accuracy. In [11], the authors proposed converting the scanned document of braille page to gray color. After converting the image, the authors used the threshold technique to obtain three classed of regions. The authors used the three classes to initially identify braille dots. The possibility of valid dot identification was used in braille cell recognition. The method achieved 99% accuracy for skewed, reversed, or workout braille document.

In [12], the authors proposed classifying Arabic braille character using fuzzy classification, character feature extraction, and character recognition. The authors proposed system was developed with the use of segmenting Arabic braille characters. The authors use of fuzzy classification was inspired by the Fuzzy C-Mean (FCM) and fuzzy KNN classification algorithms. The method achieved up to 83% accuracy of classification and recognition of braille character. In [13], the authors suggest using a text-matching method as a way to recognize images. This system implies that starting with observing the interaction between words, then the use of several matching patterns between the phrases and ending up with matching entire sentences. This paper tried two methods to see what gives the highest accuracy, one of which was using paper citation matching, where the authors used a large academic dataset along with their citations and abstracts. The dataset sized 838,908 instances in total, containing 279,636 positive instance pairs and 559,272 negative instance pairs. The negative pairs were selected randomly, where they do not have citations along with them. Moreover, one out of three models that the authors trained outed the other two models. The three models being MP-IND, MP-COS, and MP-DOT, where MP-DOT was the model that gave the highest accuracy with a percentage of 88.73. In [14] seen that camera-enabled smartphone was used as the main method to capture braille characters. The paper suggested an algorithm that manipulates the images of braille documents that interpret the document's highlights and convert them to their equivalent English characters. Taking into account the lighting conditions while capturing the images, the authors obtained an accuracy of over 80% by developing an application that works under Android platform. The authors also stated that under the right conditions while capturing braille texts, the accuracy can go up to 84%.

In [15], the authors proposed a system for recognition double-sided Amharic braille documents that use the identification of three methods. Those methods are recto, verso, and overlapping dots. The system that the authors suggested in this paper works by simply converting braille texts into codes, and those codes are later on being translated into texts again. On top of that, adding the concept of reflection to reverse wrongly scanned braille documents automatically. While the dataset was collected from Addis Ababa University's Kennedy Library that contains good and bad scanned braille documents, the system was evaluated to give a high accuracy of 99.3% for identification and accuracy of 95.6% for translation. The authors in [16] used optical braille recognition based on semantic segmentation network along with auxiliary learning strategies. Using the OBR framework along with BraUNet and morphological post-processing procedures, the authors also used corresponding pixel-level annotations of braille characters with 64 other classes as an input in the system for both training and testing. The results of the methods they applied on DSBI dataset, type recto braille character along with BraUNet gave the best results with an accuracy of 0.9966%. While the regular number of classes for braille classes is 64 the authors in [17] took another approach and added 7 more classes to add up to 71 classes of characters to corresponding to the braille dataset that consists of 37 characters. Moreover, a collective dataset of 26,724 labeled braille images now has 37 braille symbols that correspond to the 71 classes. Using a novel method that pairs ratio character segmentation, RCSA's algorithm was used aside with CNN to translate a line of braille into its English counterpart. By linking the CNN model to two recognition techniques: character and word recognition the system proposed in this paper were able to reach an accuracy of 98.73% on the test set. In [18], the authors used Convolutional Neural Network (CNN) techniques to develop a system that can identify Cyrillic braille characters. After scanning the braille documents, image preprocessing techniques were used to make the recognition process easier. Then, character segmentation was performed to improve recognition accuracy. Subsequently, a modified backpropagation algorithm was used to train neural networks. The system has achieved 95.7% training accuracy and 95% testing accuracy. The authors also concluded that the use of the artificial neural network is very helpful in identifying characters due to the ease of programming the network architecture to train and test with any image sizes as an input. In [19], the authors proposed a module that uses associative memories to recognize single-sided braille documents and then convert it to audio. Their module consisted of two stages, preprocessing and recognition. In the preprocessing stage, different operations are performed on the scanned braille papers such as grayscale conversion and dilation to prepare the images for the next stage. Then, Modify Multi-Connect Architecture (MMCA) and Modify Bidirectional Associative Memory (MBAM) algorithms are used to recognize the characters. The authors compared results of MMCA and MBAM algorithms, where (MMCA) achieved an average accuracy of 98.26% for characters recognition while (MBAM) achieved 91.87%. Afterward, the proposed module converts the recognized text into audio.

In [20], the authors proposed developed an Optical Braille Translator (OBT) system that identifies single-sided Sinhala braille characters and translates it into texts. The systems features were developed based on image processing techniques in MATLAB environment. Their methodology was to apply preprocessing techniques like grayscale conversion, image rescaling, and angle correction functions on both handwritten and computer-generated braille documents. Then, segmentations are used to recognize the braille character cells. Afterward, extracted characters are resized into a 21×16 matrix binary images. Braille characters are regenerated using an algorithm to improve accuracy. The developed system was able to reach an accuracy of over 99%. In [21], the authors proposed a system equipped with a scanner and a webcam. The braille characters are captured by the webcam or the scanner. Then, grayscale conversion, filtering, segmentation, and normalization are done on the captured images using MATLAB IDE. After the image processing stage, Artificial Neural Networks (ANN) and feature extraction are used to recognize the patterns of braille characters as well as obtaining a training model as a dataset. Furthermore, the recognized characters are then converted to audio. The system results showed that the characters captured with webcam resulted in 57.69% accuracy while the characters captured using the scanner resulted in an average of 93.26% accuracy. In [14], the authors proposed a method to process braille documents and convert them into English text and speech. Their method performs Hough's circle detection algorithm on the phone captured braille images to identify the dots. Then, maximum length and width are calculated for all dots to generate a new image of equaled size dots. After that, the authors used image segmentation to divide the rows and columns into cells. Later on, each cell in a row is read to recognize its pattern by the position of its dots. Finally, each cell pattern is mapped to its matched English character. The output text is converted into speech using a text-to-speech engine. In ideal lighting and alignment conditions, the method achieved more than 80% accuracy.

3 Methodology

The overall aim of this research is to design an application that uses deep learning techniques that recognize Arabic braille numerals with high accuracy. Captured images of Arabic braille numerals are used to train the CNN variation Residual Network (ResNet). The model was modified to achieve high accuracy of recognizing Arabic braille numerals.

Training a deep network is a challenge, and it has been proven that the depth of a network degrades the network performance [22, 23]. To address this problem, ResNet is chosen as the building block of our network due to its methods of training a deep network efficiently. Adding more layers to the network leads to the vanishing gradient issue which shows high training error. Thus, the authors in [24] suggested adding skip connections that skip one or more layers. The authors in [24] proved that when the model uses skip connection, the training efficiency improved since the gradients can travel to many layers with the help of skip connections. The proposed model is implemented using python libraries that will accept images and classify them to its respective classes. The application will be able to detect and recognize Arabic braille numerals using a pi camera that will be integrated with raspberry pi 4. Once the model is trained, captured images of Arabic braille numerals will be classified to its respective classes, provide a correct translation of it and then convert it to speech audio representing the numeral that was processed.

3.1 Data Description

Arabic braille numerals were printed on a single side A4 embossed paper with color blue and white documents. Braille dots are arranged in a cell with three rows and two columns. Dot height is approximately 0.02 inches, and the horizontal and vertical spacing between the dots are 0.1 inches. The blank space between the cells is 1.3 inches.

As Table 1 shows, 5000 images of Arabic braille numerals were collected. Smartphone cameras are used to generate the dataset. The images are captured under different lights, such as natural and industrial lighting. The images range from several colors, white, yellow, and gray. Images were captured on all possible angles and heights.

3.2 Preprocessing

In this stage, images are being prepared to make it easier for the model to train and to recognize braille numerals.

Table 1 Arabic braille numerals sample dataset

#	Name in Arabic	Name in English	Sample image	Number of images
0	صفر	Sifer		500
1	واحد	Wahed		500
2	اثنين	Ethnein		500

(continued)

Table 1 (continued)

#	Name in Arabic	Name in English	Sample image	Number of images
3	ثلاثة	Thalatha		500
4	اربعة	Arba-a		500
5	خمسة	Khamsa		500

(continued)

Table 1 (continued)

#	Name in Arabic	Name in English	Sample image	Number of images
6	ستة	Sitta		500
7	سبعة	Sab-a		500
8	ثمانية	Thamanya		500

(continued)

Table 1 (continued)

#	Name in Arabic	Name in English	Sample image	Number of images
9	تسعة	Tis-a		500

3.2.1 Grayscaling

Computers identify RGB images as 3D arrays. Meanwhile, grayscale images are identified as 2D arrays, which make the preprocessing stage more efficient. Thus, captured RGB images of braille numerals are converted to grayscale.

3.2.2 Resizing

Training images of larger size will add computational power to the network as well as time complexity. Therefore, images are resized to 256, along the y and x-axis, before inputting them into ResNet. With this approach, the model will be able to train faster with less pixels, while preserving important features.

3.2.3 Converting to Array

In deep learning, the model is trained with images in a NumPy array form. Thus, grayscaled images will be converted to an array using the NumPy library. Later on, the arrays will be fed to the network to start the training.

3.3 Deep Residual Learning

The depth of a network is significant in any neural network model [25]. However, the deeper the network the harder it is to train the model. It is noticed that as the network gets deeper, there will be a degradation problem which will lead to a decrease in the validation set. In residual block, the layers would be fed into the next layer and the layer after it, while having the advantage to skip the number of layers known as skip connection. The diagram below (Fig. 1) illustrates the skip connection.

The diagram on the right describes a deep network that shows stacked layers one after the other. However, the diagram on the left describes a deep network with stacked layers as before but the original input is now added to the output of the block. This operation is called a skip connection. Skipping the training of a few layers can be beneficial in solving the degradation problem that affects the accuracy negatively due to having more layers than needed for training.

In [17], the authors suggested using the building block, which is defined as:

$$y = F(x, \{W_i\} + x \tag{1}$$

The above equation is the shortcut connection which does not hold extra parameters and computation complexity. The identity mapping is responsible for adding the output from the previous layer to the next layer. Input and output vectors are represented as x and y of the considered layers. $F(x, \{W_i\}$ is the function that is used

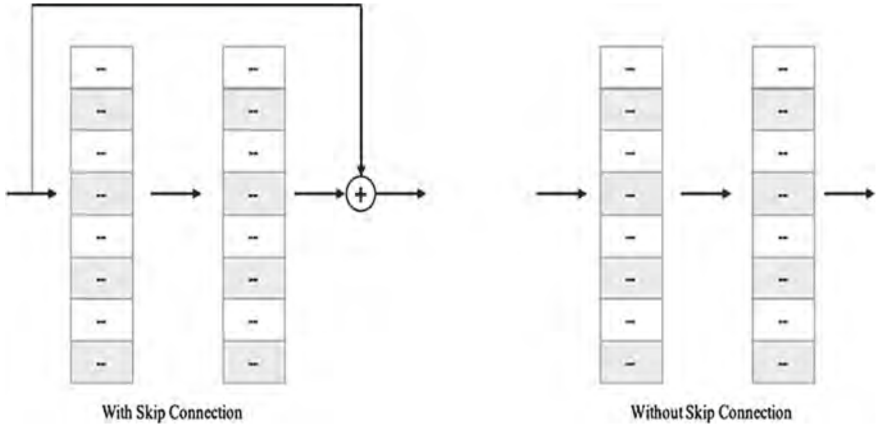


Fig. 1 Skip connection concept

for the residual mapping that will be learned. It is a requirement for the input and output to have equal dimensions to use the shortcut connection equation. Also, in [17], the authors discussed a solution in case when the dimensions are not equal that is linear projection. If the model found that the dimensions are not equal, then it will have an operation which will multiply a linear projection W to the identity mapping to match the dimensions. The equation is defined as:

$$y = F(x, \{W_i\} + W_s x \tag{2}$$

while identity mapping is used to solve the degradation problem multiplying is W_s used when there is a need to match the dimensions [17].

4 Results and Discussion

In this section, the performance of CNN was examined for training and recognizing Arabic braille numerals. A modified ResNet architecture model was used to implement our system. The dataset has been divided based on their corresponding labels to guarantee that each image will be classified based on their labels. The model was trained for 120 epochs on the braille numerals dataset, which means training for 10 classes for 120 cycles. 20% of the dataset images were used for validation and testing. This means that 3000 images were used for training, 1000 for validation, and 1000 for testing. The results using the ResNet model were extremely high, the model achieved a validation accuracy of 98%, as long with high model performance. Figure 2 shows the validation accuracy and test accuracy.

Table 2 shows a comparison between the results of the first and last epoch. The validation has increased from 0.40 to 0.98 as the training has progressed.

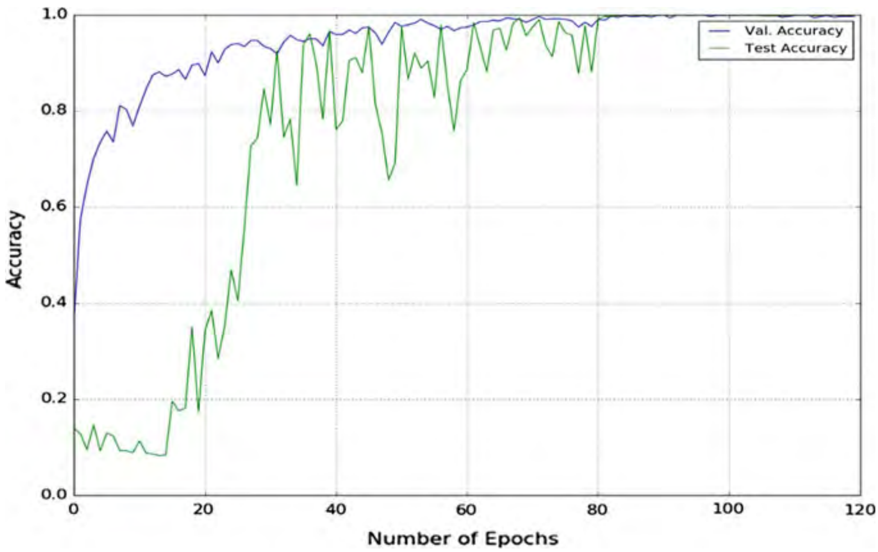


Fig. 2 Model validation and testing accuracy curve

Table 2 First and last epoch results

Epoch no.	Accuracy	
	Test	Val
First epoch	0.14	0.40
Last epoch	0.99	0.98

From Fig. 2 can be observed that the test accuracy has dramatically increased after epoch no. 25.

5 Conclusion

In this article, Arabic braille recognition using deep learning techniques was proposed. Braille is a system that is designed to assist visually impaired individuals to acquire information. Visually impaired individuals rely on the sense of touch to read and write. The proposed system can help teachers, parents, and people who have lost their vision recently. The system uses a modified ResNet model, which overcomes the degradation problem. A validation accuracy of 98% and a test accuracy of 99% were able to achieve. Deep learning techniques have proved their efficiency in recognizing braille characters.

References

1. Latif G, Mohammad N, AlKhalaf R, AlKhalaf R, Alghazo J, Khan M (2020) An automatic arabic sign language recognition system based on deep CNN: an assistive system for the deaf and hard of hearing. Int J Comput Dig Syst 9(4):715–724
2. Alghmgham DA, Latif G, Alghazo J, Alzubaidi L (2019) Autonomous traffic sign (ATSR) detection and recognition using deep CNN. Procedia Comput Sci 163:266–274
3. Shaikh E, Mohiuddin I, Manzoor A, Latif G, Mohammad N (2019) Automated grading for handwritten answer sheets using convolutional neural networks. In: 2019 2nd International conference on new trends in computing sciences (ICTCS), pp 1–6. IEEE, Oct 2019
4. Mahmoud AA, Alawadh INA, Latif G, Alghazo J (2020) Smart nursery for smart cities: ınfant sound classification based on novel features and support vector classifier. In: 2020 7th ınternational conference on electrical and electronics engineering (ICEEE), pp 47–52. IEEE, Apr 2020
5. Al-Hmouz A, Latif G, Alghazo J, Al-Hmouz R (2020) Enhanced numeral recognition for handwritten multi-language numerals using fuzzy set-based decision mechanism. Int J Mach Learn Comput 10(1)
6. Alzubaidi L, Latif G, Alghazo JM, Zikria M (2019) Cloud-based ınteractive hands free e-learning environment for students with disabilities
7. Latif G, Alghazo JM, Maheswar R, Jayarajan P, Sampathkumar A (2020) Impact of IoT-based smart cities on human daily life. In: Integration of WSN and IoT for smart cities. Springer, Cham, pp 103–114
8. Subur J, Sardjono TA, Mardiyanto R (2016) Braille character recognition using find contour and artificial neural network. J Electr Electron Eng 14(1)
9. Authman ZI, Jebr ZF (2012) Arabic Braille scripts recognition and translation using image processing techniques. J Educ Pure Sci 2(3):18–26
10. Mousa A, Hiary H, Alomari R, Alnemer L (2013) Download limit exceeded, citeseerx.ist.psu.edu, Nov 2013. Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.403.6856&rep=rep1&type=pdf.
11. Al-Salman A, AlOhali Y, AlKanhal M, AlRajih A (2007) An Arabic optical braille recognition system
12. Al Nassiri A, Abdulla S (2018) A fuzzy classification and recognition system for Arabic Braille segmented characters. Int J Appl Eng Res 13(6):3662–3669
13. Pang L, Lan Y, Guo J, Xu J, Wan S, Cheng X (2016) Text matching as image recognition. arXiv preprint arXiv:1602.06359
14. Venugopal-Wairagade GA (2016) Braille recognition using a camera-enabled smartphone. Int J Eng Manuf 6(4):32–39
15. Seid H, Assabie Y (2017) Recognition of double sided Amharic Braille documents recognition of double sided Amharic Braille documents view project NLP tools for Ethiopian languages view project recognition of double sided Amharic Braille documents. Int J Image Graph Sig Process 4:1–9
16. Li R, Liu H, Wang X, Xu J, Qian Y (2020) Optical Braille recognition based on semantic segmentation network with auxiliary learning strategy. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp 554–555
17. Hsu B-M (2020) Braille Recognition for reducing asymmetric communication between the blind and non-blind. Symmetry 12(7):1069
18. Smelyakov K, Chupryna A, Yeremenko D, Sakhon A, Polezhai V (2018) Braille character recognition based on neural networks. In: 2018 IEEE second ınternational conference on data stream mining & processing (DSMP), pp 509–513. IEEE, Aug 2018
19. Khaled S, Safana H, Abbas (2017) Braille character recognition using associative memories. Int J Eng Res Adv Technol (IJERAT) 1:2454–6135
20. Perera TDSH, Wanniarachchi WKIL (2018) Optical Braille translator for Sinhala Braille system: paper communication tool between vision ımpaired and sighted persons. Int J Multimed Its Appl (IJMA) 10

21. Ramiati, Aulia S, Lifwarda, Nindya Satriani SN (2020) Recognition of ımage pattern to ıdentification of braille characters to be audio signals for blind communication tools. İn: IOP conference series: materials science and engineering, vol 846, p 012008, May 2020
22. Chen Z, Xie Z, Zhang W, Xu X (2017) ResNet and model fusion for automatic spoofing detection. In: INTERSPEECH, Aug 2017, pp 102–106
23. Latif G, Iskandar DNFA, Alghazo J, Butt MM (2020) Brain MR ımage classification for glioma tumor detection using deep convolutional neural network features. Curr Med Imag
24. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: IEEE conference on computer vision and pattern recognition, 2016, pp 770–778
25. Latif G, Alghazo J, Maheswar R, Vijayakumar V, Butt M (2020) Deep learning based intelligence cognitive vision drone for automatic plant diseases identification and spraying. J Intell Fuzzy Syst (Preprint) 1–12

In-Silico Modeling of Sleep Stage Classification System Using Covariance Normalization

Nishant Gautam, G. B. Mrudula, and C. Santhosh Kumar

Abstract With surge among count of sleeping disorders across the globe and among every strata of society and the non-availability of sleep medicine tools in the backward regions of the third world nations, the need of automated systems arises. This paper introduces an inexpensive, computerized binary sleep stage classification system classifying the rapid eye movement (REM) and non-rapid eye movement (NREM) stages with the usage of electrocardiogram (ECG) and respiratory effort signals (RES). To avail a baseline classification of the sleep stages, support vector machine (SVM) was employed as the backend classifier that uses the heart rate variability (HRV) and respiratory rate variability (RRV) features derived from ECG and RES, respectively. The baseline system developed using linear SVM kernal performed better with performance accuracy of 73.83% , sensitivity of 84.37% and specificity of 30% in totality. The statistical features extracted from the data contain patient-specific variations that are irrelevant for sleep stage classification. As an effort to minimize these variations, covariance normalization (CVN) was performed, and a system is obtained with an absolute overall classification accuracy of 81.30%.

Keywords Electrocardiogram · Respiratory effort signals · Respiratory rate variability · Heart rate variability · Covariance normalization

N. Gautam (✉) · G. B. Mrudula · C. Santhosh Kumar
Machine Intelligence Research Laboratory, Department of Electronics and Communication Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: g.nishant12329@gmail.com

G. B. Mrudula
e-mail: gb_mrudula@cb.students.amrita.edu

C. Santhosh Kumar
e-mail: cs_kumar@cb.amrita.edu

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_8

1 Introduction

Sleep [1], the ingrained natural, easily reversible periodic state of all the living beings marked by the state in between wakefulness and loss of consciousness of one's surrounding, is accompanied by the typical body posture (lying down, closed eyes), occurrence of dreams, changes in brain's electrical activity, reduced muscle activity and other physiological signaling. Sleep comprises recurrent stages of NREM and REM within the sleep cycle and is considered to be highly important for the restoration and recovery of vital bodily and cerebral functioning. The electrophysiological variables associated to the sleep study are electroencephalography (EEG), electrocardiogram (ECG), electrooculogram (EOG), chin electromyogram (Chin-EMG), respitory effort signals (RES) and oxygen saturation (SpO_2) [2]. Sleep is greatly affected by the circadian cycle managed by cerebral neurons, which collectively builds up the sleep cycle of human body.

Traditionally, sleep cycle is sorted out into two stages, non-rapid eye movement (NREM) and rapid eye movement (REM). Both of these differentiated using not just neuro-physiological but also psycho-physiological characteristics. NREM and REM stages of sleep alternate in a cyclic manner, with a total of 4–6 cycles, each cycle lasting for 90–110 min. The NREM sleep often referred as quiet sleep, which continues as REM sleep also known as active sleep is subdivided to foour stages. Initial stage of NREM sleep is described as a transition time from wakefulness to sleep. As NREM sleep advances, the heartbeat gradually slows down, muscle activity decreases, the consciousness of the individual fades out completely, and the sleep cycle enters into the REM stage. The REM stage is a sleep period with intense brain activity, where muscles are inactive and unresponsive characterized by rapid movements of eye in various directions. Sleep disorders like sleep apnea are very common to occur in this stage due to the lose of muscle tone. During REM sleep, there is an overall increase in breathing rate, and there is a more pronounced variation in cardiovascular activity, with overall increases in blood pressure and heart rate and also ST segment elevation. Additionally, changes in blood flow that cause erections to occur in males or swelling of the clitoris in females are characteristic of REM sleep. The underlying reason for these considerable cardiac and physiological variations in REM sleep is currently unknown and may be a by-product of REM-related changes in nervous system activity or related to dream content. In the contemporary world, large population suffers from the sleeping disorders [3] irrespective of age and gender, like sleep apnea, insomnia, restless leg syndrome (RLS) and others. Limitations of the traditional non-invasive sleep diagnostic tool polysomnography (PSG) [4] are that, to avail the electrophysiological data, a complex array of electrical sensors is attached to the body making patient uncomfortable. Also, the procedure is time consuming and requires an expert neurotechnician to interpret the results and proffer the diagnosis which makes the sleep medicine an expensive and arduous task. Hence, the non-familiar environment and the possibility of interpretation error makes the data erroneous that lead to having a great potential of misdiagnosis. Therefore, the need of development of an automated, portable, cheap sleep state classification system arises.

The foundation of this cardiorespiratory system resides on H.R Variability and R.R Variability features [5, 6]. Till date several researches have been carried out in the sleep stage classification domain in which the features were derived from varied sleep stages and substages within them. Hence, our research work eliminates the particular stage and the varied dependencies of feature values. HRV, aka cardiac beat by beat variation and is highly necessary to identify the autonomic vitality defined by the parasympathetic superiority within NREM stage also with a shift to sympathetic ascendency within REM stage. RRV is also an important criterion for different humane parameters and system. RRV varies greatly with the sleep stages matching up with the defined depths of sleep and posseses greater utility when a person is awake. Hence, the proposed system employs only ECG and RES signals, making the system effective and accessible [7] to use in backward localities having only primary healthcare furnished with the simple data acquisition systems [8]. Sleep impoverishment is associated with alteration in HRV and RRV, which makes these features a promising identifier for sleep disorders and sleep stages.

2 Methodology

2.1 Dataset

PSG signals were acquired from the Department of Neurology at Amrita Institute of Medical Science (AIMS), Kochi. The dataset comprised of full single night PSG recordings of total 32 subjects aging from 17 to 75 years. The sleep study data comprised of physiological parameters like both side occipital EEG, SpO_2, ECG, RES, nasal intake along with left and right side EOG. The final dataset constructed comprised of 247 samples, out of which the division of data for testing and training was achieved. The data was split into 70:30 ratio, hence 173 samples were kept for training, and 74 samples were maintained for testing purpose.

2.2 Feature Extraction and Feature Fusion

Signals were processed in advance, earlier than extracting needed features for the study. A 50 Hz notch filter was used to filter out the power line interference from ECG. Artifacts from effort signals were removed using a filter having a cut-off frequency of 1 Hz. R-R [9] and breath to breath interval was drawn out from effort signals and cardiogram signal. HRV and RRV features in both time domain and frequency domain were extracted from the interval data. Feature extracted was fused together through feature fusion [10]. Hence, a combined featured data comprising fused time and frequency domain features was also constructed for the study.

- **R-R Interval**
  Time elapsed time between two successive R waves within the QRS complex on the electrocardiogram has unit of measurement in seconds. General (normal) value of RR interval is 0.6–1.2 s. And any sort of variability besides the mentioned normal value reflects the disorder within the subject.
- **B-B Interval**
  Breath to breath interval [11] was monitored within the breath cycle of the test subjects to analyze the patient's degree of breathing while she/he is asleep. All breath to breath intervals were automatically analyzed from flow signal, displayed and manually corrected for artifacts. Variations among the B-B interval during sleep cycle can indicate toward the morbidity within that patient.

Time Domain Features (TD) Refers to variation of amplitude of signal with time. Here in this domain features extracted encompass mean, standard deviation and root mean square successive difference (RMSSD) of R-R and B-B intervals, mean heart and breathing rates, percentage of the mean number of times an hour in which the change in successive normal sinus (NN) intervals exceeds 50 ms. Above them all, the data from respiratory inductance plethysmography (RIP) was involved for evaluating pulmonary ventilation by taking the movement of chest wall and abdomen into the equation. This feature known as rib cage percentage (RC%) is defined as the contribution (in percentage) of the rib cage to the inspiratory tidal volume, given by the equation:

$$\mathrm{RC\%} = \frac{\text{Thoracic Movement}}{\text{Thoracic Movement} + \text{Abdomen Movement}} \tag{1}$$

Frequency Domain Features (FD) The frequency domain refers to the analytic space in which mathematical functions or signals are conveyed in terms of frequency, rather than time. Features of both HRV and RRV were considered for this respective domain, and therefore, three frequency bands given in Tables 1 and 2 were considered and employed. Frequency domain features encompass absolute power, maximum frequency, peak power, ratio among low frequency and high frequency.

Table 1 HRV frequency bands

Band name	Abbreviation	Wavelength (Hz)
Very low frequency	VLF	0.0003–0.04
Low frequency	LF	0.04–0.15
High frequency	HF	0.15–0.4

Table 2 RRV frequency bands

Band name	Abbreviation	Wavelength (Hz)
Very low frequency	VLF	0.0003–0.05
Low frequency	LF	0.05–0.2
High frequency	HF	0.2–0.25

2.3 Baseline System

Baseline system which is to act as reference point of this study was created using the features extracted from the ECG and RES data of all subjects. The database then later on separated into the training and testing set, and the SVM [12]-based backend classifiers were used for the training the classifier and to develop the SVM-based sleep stage classification model. Representation of the baseline/reference system is given in Fig. 1.

Polysomnography (PSG) Polysomnography, a diagnostic tool to analyze patients having a potential sleeping disorder, records the multiple parameters that are associated with the sleeping disorders such as heart rate, EEG, skeletal muscle activity (EMG), oxygen saturation, eye movement (EOG) and also the respiratory effort indicators. Test is performed within sleep lab in the presence of an expert technologist, who gives actual diagnosis after a step of sleep stage scoring which is performed by epoch method through assessment of sleep parameters.

Here, in our research work instead of using all the extracted parameters, only ECG and RES signals were employed and processed.

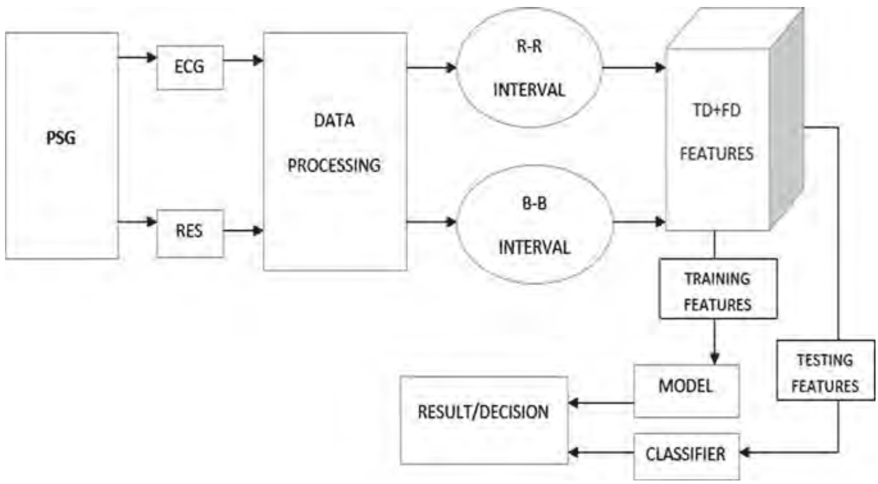


Fig. 1 Baseline system

Electrocardiogram (ECG) An electrocardiogram records the heart's electrophysiological data abbreviated as EKG and ECG. ECG shows graph of voltage vs time of cardiac bio-potential. Traditionally, it is of 12 leads out of which ten were placed on patient's limbs and chest. ECG is made up of frequency peaks, intervals and segments named PR interval, PR segment, QRS complex, ST segment, etc., each having a unique significance and importance when vitality of the heart is in question.

Respiratory Effort Signal (RES) RES was measured using sensor bands placed around thoracic and abdominal regions, and respiratory airflow was measured using thermocouple sensors placed at the nasal and buccal offices. It involves the procedure of respiratory inductance plethysmography which measures the pulmonary ventilation by evaluating chest and abdominal movements. Although number of complex respiratory parameters such as tidal volume (vt), minute ventilation, respiratory rate can be obtained, but here for our research purpose, RC% (rib cage percentage ratio) was considered.

Data Processing Raw data was processed by employing different filters (LPF and notch filters). Artifacts were removed from the ECG and RES signals, and the time domain and frequency domain features were extracted (refer Sect. 2.2). Post data processing features extraction procedure was employed that gives rise to our final dataset which further was divided into testing and training data (refer Sect. 2.1).

Model and Classifier SVM backend classifier was employed over the final dataset, and implementation of four different kernels (linear, polynomial, rbf and sigmoid) was performed over the datasets (test and training). Finally, best performance was noted down and taken as the result.

2.4 Covariance Normalization

A feature normalization [13] approach safeguards the inherent correspondence among the feature vectors. In real-time data acquisition, there is a presence of specific attenuation to the physiological signals of each subject. This normalization contains the ability to minimize the noisy factor established due to subject-specific variations. Stepwise procedure of covariance normalization (CVN) [14] is given in Fig. 2.

Covariance Matrix Covariance matrix is a square matrix providing the covariance between each pair of elements of a given feature vector. Covariance gives the joint variability of two random variables. It measures the direct relationship between two quantities.

$$\mathrm{Cov}(X, Y) = \frac{\sum (Xi - \bar{X})(Yi - \bar{Y})}{N} \tag{2}$$

In true physical sense, if covariance is positive, the quantities related move together in a proportional fashion, on the other hand, if the covariance is negative, the quantities are inversely proportional to each other. Covariance matrix is positive semidefinite

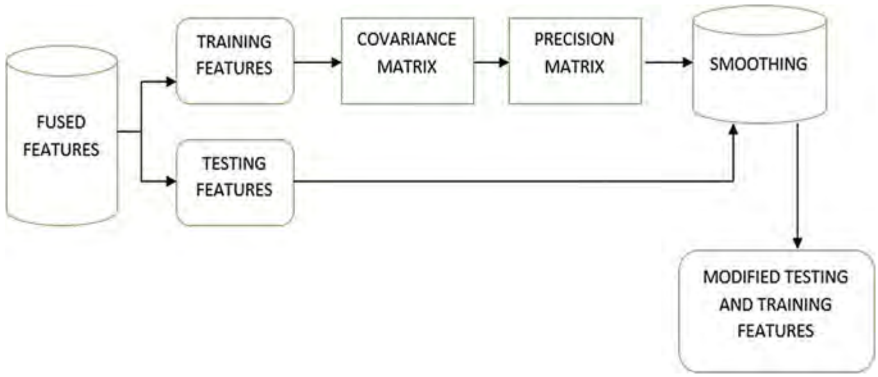


Fig. 2 CVN block diagram

matrix having the eigenvalue positive and the scalar strictly positive for every non-zero column vector z of n real numbers. Here, the covariance was employed over the training data resulting in the matrix formulation.

Precision Matrix Commonly known as inverse covariance matrix or concentration matrix. In true physical sense, it is identified as the converse of covariance matrix. Reason to obtain the concentration matrix is its association with the multivariate normal distribution and partial correlation.

Data Smoothing Estimates obtained after covariance are considered as noisy, which are smoothened out by computing smoothening matrix. This allows important patterns to stand out while leaving noise out. Smoothing can be used to help predict trends. Here, data is compiled, and it can be manipulated to remove or reduce any fallacy or noise.

Here, the smoothing of the data is performed using the equation:

$$S = \lambda P + (1 - \lambda) I \tag{3}$$

within the equation S is smoothing matrix, P is upper triangular matrix extracted from cholesky factorization, lambda is the smoothing factor, and identity matrix is represented with a symbol I. The main motive to smooth the data is to identify simplified alterations in order to help predict different trends and patterns. The smoothened training features were obtained and were fed to the classifier to develop the CVN-based sleep stage classification system.

3 Results and Discussion

Here, in the results, the best performance out of the four kernels employed was taken into the account.

3.1 Support Vector Machines (SVMs)

Support vector machine is a widely used supervised machine learning model useful for classification approaches and works effectively for two stage classification problems. After feeding an SVM model sets of labeled training data for each category, they are able to categorize or classify the dataset efficiently into binary classes. A support vector machine takes the data values and outputs the hyperplane (which in two dimensions is simply a line) that best separates the labeled data. Hyperplane is a hypothetical decision boundary that separates the data points into the feature space: Anything that falls to one side will classify as the class 0 and other will be class 1. LIBSVM [12] package was employed over the Python platform and thereby usage of four kernels (shown below). And the accuracy of each one was noted.

- **Linear Kernel**.
  Simplest of them all, this function is the inner product of x&y added with a constant value (c). This function separates the data into two classes using a linear function given by:
  $$P(x, y) = (x^T * y) + c \tag{4}$$
- **Polynomial Kernel.**
  It is a non-stationary function, suitable for the situations where the data is pre-normalized. Represented as:
  $$P(x, y) = (ax^T * y) + c \tag{5}$$
- **RBF (Radial Basis Function Kernel).**
  It is nonlinear function that separates the data into classes, used when the boundaries are hypothesized and have a curvy fashion.
- **Sigmoid Kernel.** Also known as hyperbolic tangent kernel. This function placed itself under the neural networks learning field, where it is used as an activation function for the perceptrons.
  $$P(x, y) = \tan h(ax^T * y + c) \tag{6}$$

Baseline System System's performance for TD, FD and TD-FD features is tabulated in Table 3. The baseline performance was seen best for linear kernel-SVM.

Table 3 Baseline system performance (%)

Input features	Accuracy	Sensitivity	Specificity
TD	73.83	78.30	30
FD	63.5	84.37	32.55
TD and FD	57	77.61	22.5

Table 4 CVN system performance (%)

Input features	Accuracy	Sensitivity	Specificity
TD	80.373	81.63	66.6
FD	81.30	81.80	75
TD and FD	81.30	80.50	100

Covariance Normalization System The performance accuracy of the normalized system for all the features is tabulated in Table 4. The performance of the system was seen best for linear kernel-SVM.

4 Conclusion

Within this work, an approach of covariance normalization was explored to reduce the disparites associated to a particular subject, embedded within features in order to ameliorate the execution of the sleep stage classification system. Reference system was created using an SVM classifier. The statistical features extracted from the heart's electrophysiology and effort signals were fed to the baseline/reference system as an input. Reference system showed a performance of 73.83% for time domain features, 63.5% for frequency domain features and 57.0% for both time domain and frequency domain features combined. Further, the patient- and stage-specific variations of the features were reduced using a feature normalization technique called covariance normalization. The performance of the CVN-SVM system with TD features as input had a performance improvement of 6.54% absolute, CVN-SVM system with FD features as input, had performance improvement of 17.8% absolute and CVN-SVM system with both TD and FD features as input had a performance improvement of 24.3% absolute. This cardiorespiratory system could be helpful to the patients suffering from sleeping disorders like sleep apnea, restless leg syndrome and others. CVN approach favored the aim and needs of the study and performed better than the baseline system, with the best performance of 81.30%

Acknowledgements My deepest and unfeigned appreciation to my friends and faculties for the support and guidance throughout the research work. Also, I sincerely express my humble gratitude to Haritha H and Sreekumar K T of Machine Intelligence Research Laboratory (MIRL) for providing essential technical support and the needed motivation all along.

References

1. Borb A, Achermann P (1999) Sleep homeostasis and models of sleep regulation. J Biol Rhyth 14(6):559–570
2. Redline S et al (1998) Methods for obtaining and analyzing unattended polysomnography data for a multicenter study. Sleep 21(7):759–767

3. Sateia M (2014) International classification of sleep disorders-third edition. Chest 146(5):1387–1394
4. Douglas N, Thomas S, Jan M (1992) Clinical value of polysomnography. Lancet 339(8789):347–350
5. Stein PK, Pu Y (2012) Heart rate variability, sleep and sleep disorders. Sleep Med Rev 16(1):47–66
6. Gutierrez G, Williams J, Alrehaili GA, McLean A, Pirouz R, Amdur R, Jain V, Ahari J, Bawa A, Kimbro S (2016) Respiratory rate variability in sleeping adults without obstructive sleep apnea. Physiol Rep 4(17)
7. Joby PP (2019) Exploring devops: challenges and benefits. J. Inf. Technol. Digital World 01(01):27–37
8. Shakya D (2020) Analysis of artificial intelligence based image classification techniques. J Innov Image Process 2(1):44–54
9. Manikandan MS, Soman K (2012) A novel method for detecting r-peaks in electrocardiogram (ECG) signal. Biomed Signal Process Control 7(2):118–128
10. Yang J, Yang J, Zhang D, Lu J (2003) Feature fusion: parallel strategy versus serial strategy. Pattern Recogn 36(6):1369–1381
11. Kowallik P, Jacobi I, Jirmann A, Meesmann M, Schmidt M, Wirtz H (2001) Breath-to-breath variability correlates with apnea-hypopnea index in obstructive sleep apnea. Chest 119(2):451–459
12. Chang C-C, Lin C-J (2011) Libsvm. ACM Trans Intell Syst Technol 2(3):1–27
13. Kumar CS, Ramachandran KI, Kumar AA. Vital sign normalisation for improving performance of multi-parameter patient monitors. Electron Lett 51(25)
14. Mrudula GB, Kumar CS, Haritha H, Anand Kumar A, Gopinath S (2018) Feature normalization for improving the performance of sleep apnea detection system. In: IEEE India Council international conference (INDICON), India, Dec 2018

Comparative Analysis of LSTM, One-Class SVM, and PCA to Monitor Real-Time Malware Threats Using System Call Sequences and Virtual Machine Introspection

Jayesh Soni, Suresh K. Peddoju, Nagarajan Prabakar, and Himanshu Upadhyay

Abstract System call analysis is based on a behavior-oriented anomaly detection technique, which is well accepted due to its consistent performance. This study compares two popular algorithms long short-term memory (LSTM) sequence to sequence (Seq-Seq), and one-class support vector machines (OCSVM) for anomalous system call sequences detection. The proposed framework monitors running processes to recognize compromised virtual machines in hypervisor-based systems. The evaluated results show the comparative analysis and effectiveness of feature extraction strategies and anomaly detection algorithms based on their high detection accuracy and with a low loss. This study demonstrates a comparative analysis of detecting anomalous behavior in any process using OCSVM and LSTM Seq-Seq algorithms. A bag-of-2-g with PCA feature extraction strategy and LSTM Seq-Seq with a sequence length of five provides higher detection accuracy of 97.2%.

Keywords Virtual machine (VM) · Long short-term memory (LSTM) · Sequence to sequence (seq-seq) · One-class support vector machines (OCSVM) · Virtual machine introspection (VMI)

1 Introduction

Nowadays, virtualization environments are playing a vital role in different application domains. The major advantage of virtualization is the most utilization of resources. Due to virtualization, there is vast resource sharing, and the scope of protecting resources has improved. Advancements in virtualization methods evolved in extracting virtual machine (VM) [1, 1] information and detected numerous malicious activities are a challenging task. At the present-day market, more sophisticated

J. Soni (✉) · N. Prabakar
School of Computing and Information Sciences, Florida International University, Miami, USA
e-mail: jsoni@fiu.edu

S. K. Peddoju · H. Upadhyay
Applied Research Center, Florida International University, Miami, USA

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_9

and vigorous malware is spreading where no detection methods are not in a position to identify malicious activities happening on VM. Thus, most of the businesses are facing cyber threats along with data breaching issue. To overcome these problems, hypervisor-enabled technologies are getting popular in detecting malicious activities and the performances of these techniques are far better than guest-based systems. Virtual machine introspection (VMI) is a technique to extract memory insights into VM to analyze cyber threats and detect malicious activities in it [3–5]. VMI is a mechanism to manage run-time activities of VM at the hypervisor level and memory forensics is a suitable method to study VM activities at the memory level.

The main goal of intruders is to run malicious code on the VM to modify the system files and footprints of resources. There are many more security techniques are available in the market to avoid attacks, such as anti-viruses and security patches. Despite these methods, there are various unknown attacks on the VM by anti-virus software and updating security patches. Prospective malware assaults are identified with the study of characteristics of executing software, which is also known as behavioral analysis [6, 7]. Analysis of system call [8, 9] is coming under the same category with which we can identify the malicious behavior of the software.

The anomaly detection methods and feature extraction techniques are compared and implemented using system call sequences. This study examines and compares OCSVM and LSTM Seq-Seq algorithms in detecting anomalous behavior in any process. The detection system generalizes and performs well in detecting previously unseen anomalous data under different machine configurations and typical load balancing of work. Further, feature engineering is done by studying various methods, and appropriate system call patterns were identified.

The rest of the paper consists of eight sections. Section 2 discusses latest research occurring on the current approach. The procedure for feature extraction and feature engineering for system calls traces using VMI application is described in Sect. 3. In Sect. 4, an overview of detection algorithms is explained. Section 5 describes the experimental setup. Experimental results and discussions are presented in Sect. 6. Section 7 concludes the proposed work, and finally, Sect. 8 proposes the future scope of the research.

2 Related Work

Host-based malware detection frameworks are devised to detect malware on a host system by capturing and studying system-related access. These host-based frameworks utilize signatures of existing malware and try to detect with detection algorithms. The limitation of these frameworks is to detect unknown malware footprints. Apart from this, there are some techniques that are based on the behavior of executing the program in its regular mode. There are different data processing techniques [10, 10] that are evolved in recent times to address these issues and automatically learn and train models based on its characteristics such as “good behavior” and “bad behavior” with benign systems [12, 12]. Behavioral-based malware detection methods are used to address the limitations of static analysis based on traditional detectors. Static analysis techniques are outperforming in their detection based on static features, which can be quickly puzzled [14]. Authors in papers [15, 16] discuss different behavior-related analysis methods. Anomaly detection is placing a significant role in analyzing system call data with the help of statistical and signature detectors. They have also considered several ways of extraction features from the document classification mentioned [17, 18].

Neural networks in combination with anomaly detection are a popular and good idea to detect malicious activities [19–24]. A recent study proves that deep learning methods are extensively used in detecting anomalies, and especially LSTM plays a vital role [25–28]. However, their work was a feature-based supervised classifier, which requires much work to create labels. As a result, there is a chance of labels attack, which inherently creates limitations to fail to detect any unknown attacks [29]. Furthermore, this method requires careful and specific feature engineering techniques which will provide meaningful representations in terms of features for the supervised classification problems [30, 31]. To train their model, only one binary label per sequence was provided, unlike our proposed method, which is trained to predict the next few system calls and that effectively captures contextual information. Few other papers [32–34] discussed the different machine learning algorithms to train models and detect anomalies in different fields.

In this paper, a comparative study is made with different anomaly detection algorithms to study and identify anomalies in extracted data. The normal behavior of the system is trained, and test data is compared with it to see unusual and abnormalities within it. This training and test data comparison is made with a well-defined model with the help of deep learning algorithms. This procedure identifies unknown attacks happening on VM by extracting data with the VMI technique.

3 System Call Traces with VMI

The proposed approach analyzes system calls extracted by using VMI-based techniques [35, 35] to detect malicious activities happening on VM. The main objective of this technique is to extract system call data of active program under execution and inspects the abnormal behavior. This approach alternates existing defenses such as firewalls, antivirus, and many more in identifying compromised VMs.

The framework shown in Fig. 1 extracts system call traces with the VMI method. Guest VM is monitored by the host (monitoring) system with the help of VMI application programming interfaces such as introspector and security agent modules. The introspector module will collect system call information via the security agent. Then system call traces are analyzed with appropriate detection algorithms to identify and detect malicious behavior of processes. Custom test vectors are created for performance evaluation. System call trace sequences are obtained from Guest VM running Windows 10 operating system.

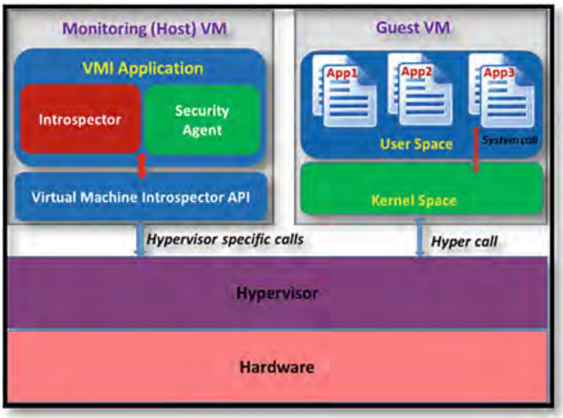


Fig. 1 Virtual machine introspection

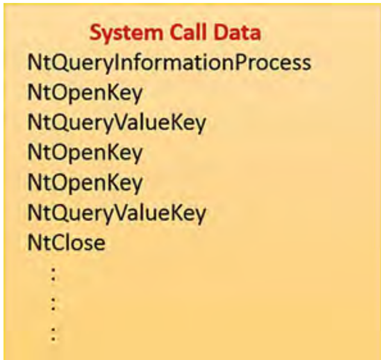


Fig. 2 System call trace example

A hypervisor is used to extract the system call sequences using a custom VMI application, which includes introspector and agent. This is a service application that receives and logs system call sequences at the process level in recent versions of Microsoft Windows. Figure 2 shows a representation of system call traces, depicting the system calls that occurred during a process. For this study, the VMI application has recorded up to 1,000,000 calls made by each process.

3.1 Feature Engineering for System Call Sequences

In this study, feature engineering is done on system call traces data by using a bag-of-words model. In natural language processing, a document is featured using bag-of-words representation by creating a vector of frequencies of each word. Similarly, the same approach is used by featuring system call sequences as a vector of unique n-gram system call frequencies. An n-gram is a vector of length n of system call sequence

Table 1 Bag-of-2-g representation of system call trace

2-g	Frequency
NtQueryInformationProcess, NtOpenKey	1
NtOpenKey, NtQueryValueKey	2
NtQueryValueKey, NtOpenKey	1
NtOpenKey, NtOpenKey	1
NtQueryValueKey, NtClose	1

occurring continuously in a process under execution. Bag-of-2-g highlighting of the system call trace of Fig. 2 and also is depicted in Table 1. This VMI approach will monitor 450 distinct system calls of windows application.

3.2 Feature Selection

Since features generated through 2-g are sparser, the PCA is utilized to select features with high variance. Orthogonal transformation is used in this method to extract sets of highly correlated variables. Principal components (PCs) are ordered so that the first PC has the most substantial highest variance. Based on PCA, only the primary few components are selected. As a result, the dimension of the dataset is reduced. It is a statistical approach that transforms a set of inter-correlated variables into low dimensions, and such transformed dimensions have a more significant amount of the variability.

4 Detection Algorithms

The anomaly detection learning algorithms are studied and analyzed in the proposed work that operates well for large-scale data. The algorithms exhibit well with less run-time computational complexity during detection. It is an unsupervised learning context, where no label is required in training data for model creation. This context gives an overview of LSTM Seq-Seq and OCSVM.

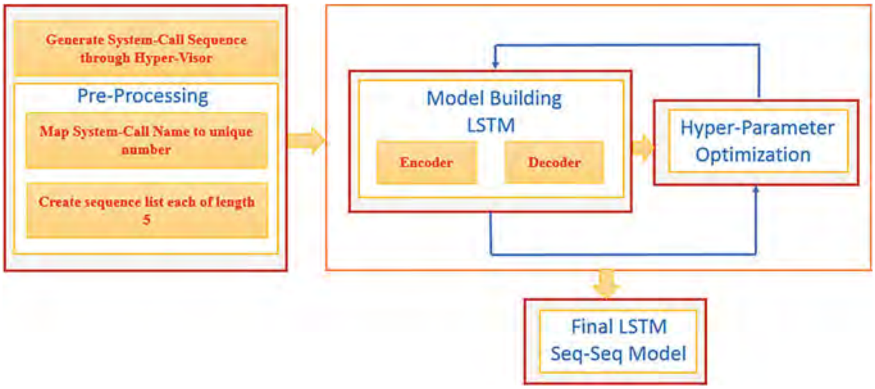


Fig. 3 Preprocessing and LSTM Seq-Seq model building mechanism

4.1 Long Short-Term Memory (LSTM) Sequence to Sequence (Seq-Seq)

4.1.1 Overview of Our Approach

The LSTM Seq-Seq model is a model based on system calls considering its sequence. Figure 3 depicts our proposed method. There are two parts to our approach, the first is the preprocess data, and the second is the model using LSTM by tuning hyperparameter for efficient optimizations.

4.1.2 Prediction Model

Consider the system call sequences generated by hypervisor during program execution as sentences. Natural language processing (NLP)-based sequence-sequence architecture by feeding the first few system call sequences as the input, which in turn generates the next sequence of system call as output. The sequence to sequence architecture is an encoder–decoder framework based on RNN [37], and it can be viewed as a sentence-to-sentence mapping. The idea of encoder and decoder works in the way that humans think. The same level of processing can be used in system call sequences, consideration is based on the first part as a source sequence in a Q&A system, and the second part is based on the answer.

To transform each document into its numerical vector, this creates its vocabulary. Similarly, a unique lexicon for each system call for a windows machine is employed. Let us assume that there are m individual system calls. A set of all system calls are defined by T, and these are extracted by hypervisor while executing a program. Convert each system call to a numeric vector. The system call sequence s can be represented as $s = (s_1, s_2, s_3, \ldots, s_n)$, and the target system call sequence as $t = (t_1, t_2, t_3, \ldots, t_m)$, where $s_i, t_i \in T$.

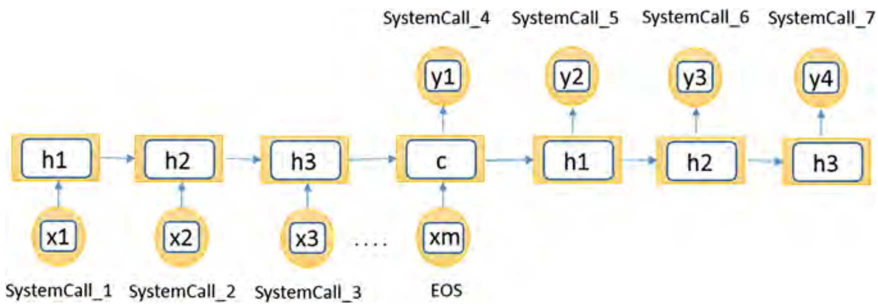


Fig. 4 LSTM encoder–decoder architecture

Given a system call sequence, the encoder produces two states (hidden and output) by completing a forward propagation operation. The formula is as follows:

$$ht = (\mathrm{c}^{HX} x_t + \mathrm{c}^{HH} h_{t-1}) \tag{1}$$

$$y_t = \mathrm{c}^{YH} h_t \tag{2}$$

h_t is encoded information vector c as shown in Fig. 4.

The weight update equations for an LSTM cell are as follows:

$$y_t = \sigma(T_{xi} x_t + T_{hi} k_{t-1} + b_i) \tag{3}$$

$$f_t = \sigma\left(T_{xf} x_t + T_{hf} k_{t-1} + b_f\right) \tag{4}$$

$$o_t = \sigma(T_{xo} x_t + T_{ho} k_{t-1} + b_o) \tag{5}$$

$$g_t = \tanh(T_{xc} x_t + T_{hc} k_{t-1} + b_c) \tag{6}$$

$$c_t = f_t c_{t-1} + i_t g_t \tag{7}$$

$$k_t = o_t \tanh(c_t) \tag{8}$$

In above equations, **tan h** is the hyperbolic tangent function, $\boldsymbol{\sigma}$ is a sigmoid function, and $\mathbf{i_t}$, $\mathbf{f_t}$, $\mathbf{o_t}$, and $\mathbf{c_t}$ are the input gate, forget gate, output gate, and memory cell activation vectors, respectively.

Using context vector c and initial hidden state, the conditional probability of the decoder is shown in Eq. 9.

$$p(y_1, \ldots, y_{T'}|x_1, \ldots, x_1) = \prod_{t=1}^{T'} p(y_t|c, y_1, \ldots, y_{t-1}) \tag{9}$$

Furthermore, perform the attention mechanism by which the conditional probability changes to the below equation.

$$p(y_i|y_1, \ldots, y_{i-1}, X) = g\left(y_{i-1}, s_{i,} c_i\right) \tag{10}$$

where c_i is the context vector calculated during training:

$$c_i = \sum_{j=1}^{T_x} a_{ij} h_j \tag{11}$$

where a_{ij} coefficient of hidden state h_j at time step j. a_{ij} is calculated as follows:

$$a_{ij} = \frac{\exp\left(e_{ij}\right)}{\sum_{k=1}^{T_x} \exp(e_{ik})} \tag{12}$$

where e_{ij} is the relationship learned during the training phase between input and output.

The system call sequence analysis is performed by varying sequence lengths. Finally, 1,000,000 training sequences are collected.

4.2 One-Class Support Vector Machine (OCSVM)

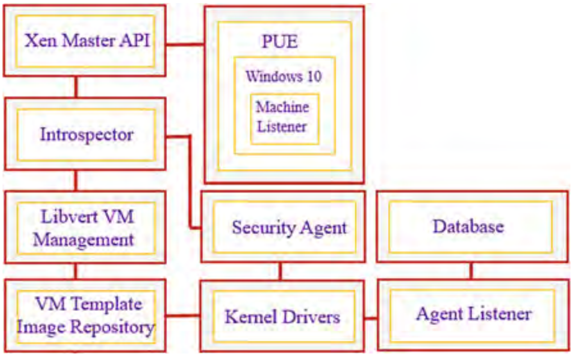
Notational, let us say dataset $\{x_1, x_2, \ldots, x_{\mathrm{i}}, \ldots, x_{\mathrm{N}}\}$, with each $X_i \in R^D$ is the one class. The motive of the OCSVM [38] is to detect a hyperplane that separates data from the origin with a high margin. The hyperplane is in high-dimensional space for nonlinear problems with nonlinear transformation Φ(.). The following quadratic function is used:

$$f(x) = \mathrm{sgn}((w.\emptyset(x)) - \rho) \tag{13}$$

where w is a perpendicular vector to the maximizing hyperplane, and ρ is the distance. A set of variables $\xi i \geq 0$ is introduced to detect outliers. Further, the following decision functions are used.

The OCSVM is trained with and without principal component analysis using system call sequences.

Fig. 5 Framework of proposed evaluation method



5 Experimental Setup

System call data is extracted through a virtual machine, with the help of the Xen hypervisor and libvirt library. Virtual memory introspection is done through DRAKVUF. There are two divisions named introspector and security agent developed using GO language, for data extraction. The virtual machine under inspection is called the system under test (SUT). The recall profile of Google is used for extracting the data. It is a JSON file which contains the memory mappings-related resources. LibVMI library services introspection requests. Various operations of VMs are handled using the libvirt library. An application is designed for performing virtual machine operations and further extracts data. The application is written in the Microsoft Visual Studio.NET framework comprised of user-defined API calls for introspector communication and other related function calls. With the help of an agent, the extracted data is stored into the database. Furthermore, extensive analysis using deep learning algorithms is performed on this data to gain relevant insights. Figure 5 shows the experimental setup.

6 Experimental Results and Discussion

In this subsection, the results of detection algorithms were discussed.

6.1 LSTM Seq-Seq Model Results

In the training of our proposed LSTM Seq-Seq model, the tuning of hyperparameters is critical. The hyperparameters such as sequence length, number of nodes, dropout rate, number of epochs, and batch size are incorporated.

Accuracy as a metric for comparing both the models is defined as follows:

$$\text{Accuracy} = \frac{\text{\# correctly predicted sequences}}{\text{Total sequences of system calls}} \tag{14}$$

The LSTM Seq-Seq neural network in Python using Keras is implemented as the framework. The model with different sequence length values is tuned properly. The dropout rate is 0.2, batch size as 128, and trained with 100 epochs. Different sequence length applied is {3, 5, 10}. The LSTM model is trained with 1,000,000 records and optimized using Adam optimizer. Figures 6 and 7 depict that the sequence length is five and also the low loss and high accuracy, respectively.

Optimizing the number of nodes is critical since it directly affects training time. Utilization of the #Nodes: {16, 32, 64, 128, and 256} is for keeping the other parameters constant.

Figures 8 and 9 depict that 64 nodes perform well in comparison with the varying number of nodes and further increasing the nodes results in model overfits. This model is able to detect the next sequence of system call with high accuracy of 97.2%. On checking against malicious sequences, the model is correctly able to detect malicious behavior.

Fig. 6 Loss versus epoch

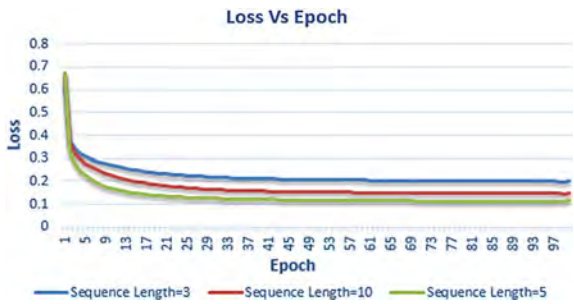


Fig. 7 Epoch versus accuracy

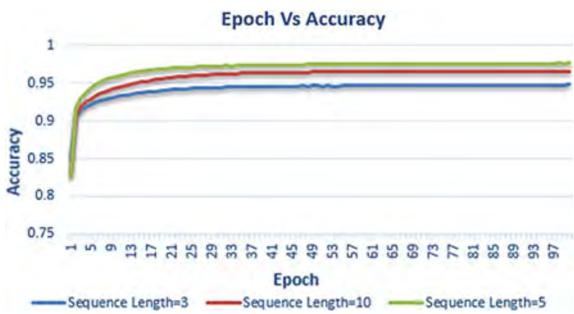


Fig. 8 Training versus testing accuracy

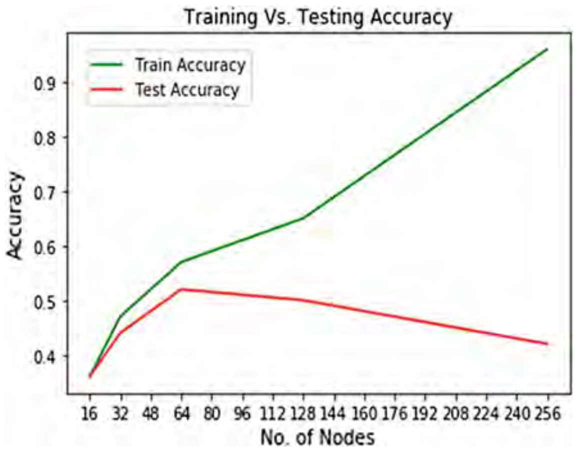
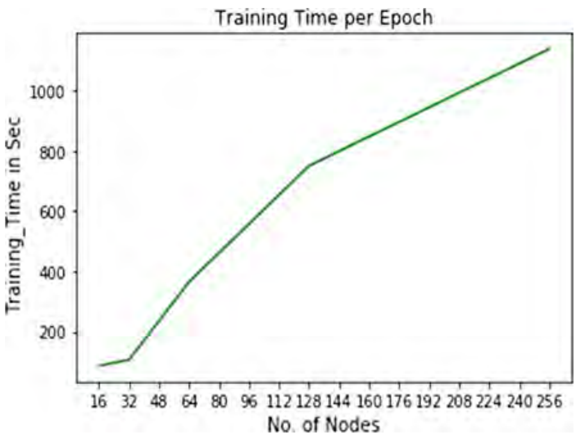


Fig. 9 Training time per epoch w.r.t number of nodes



6.2 *OCSVM Model Results*

For evaluating the OCSVM technique, the similarity score is a metric. The similarity score is a measure which checks how much the training data matches with the testing data. A high similarity score with benign sequences and low similarity scores with malicious sequences is a good sign for an accurate model.

Figures 10 and 11 depict the results of OCSVM trained with PCA on benign sequences. The similarity score with another benign sequence is 90.7% as seen in Fig. 10, whereas the similarity score with malicious sequences is 2.56% as found in Fig. 11.

Furthermore, the OCSVM is trained with raw features without PCA components. The similarity score with another benign sequence is 87.2% as seen in Fig. 12, whereas the similarity score with malicious sequences is 85.56% as found in Fig. 13.

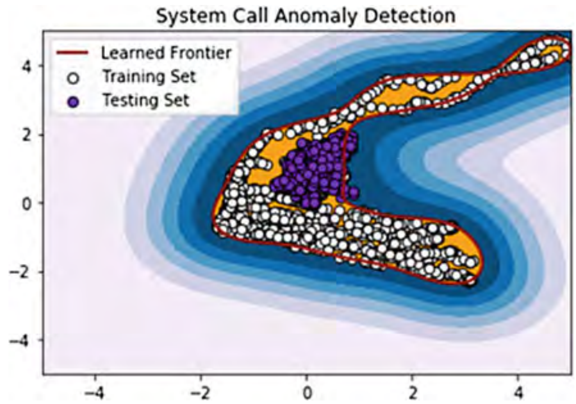


Fig. 10 OCSVM with PCA on benign sequences

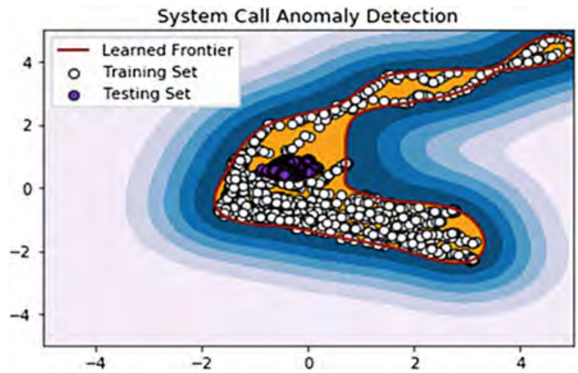


Fig. 11 OCSVM with PCA on malicious sequences

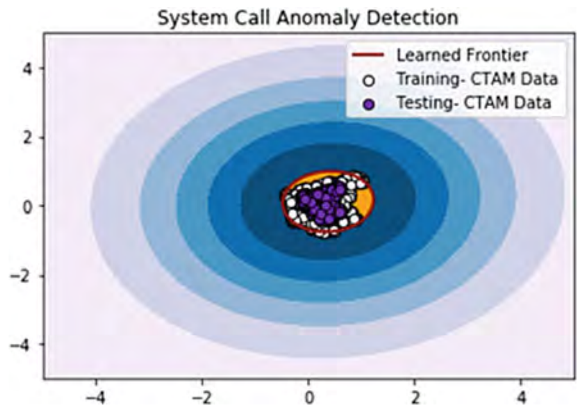


Fig. 12 OCSVM without PCA on benign sequences

Fig. 13 OCSVM without PCA on benign sequences

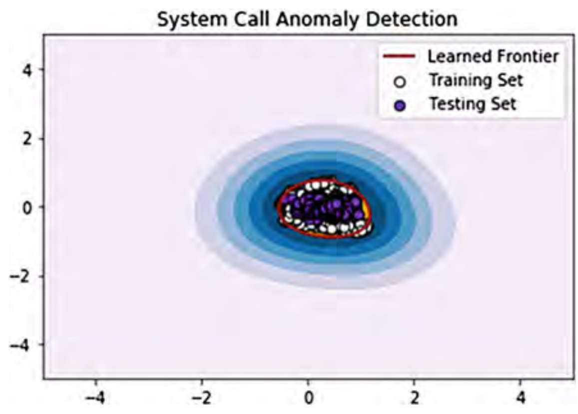


Table 2 Comparative results of the detection algorithms

Model	Accuracy (%)
LSTM Seq-Seq	97.2
OCSVM with PCA	91.6
OCSVM without PCA	87.4

Table 2 lists the testing results of LSTM Seq-Seq, OCSVM with PCA, and OCSVM without PCA on the test set and observed that the LSTM Seq-Seq has the highest accuracy. Moreover, OCSVM with PCA improves the accuracy of a small amount compared with OCSVM without PCA, which indicates that the training on PCA components features is better than the raw features on anomaly detection.

7 Conclusion

The comparative analyses of anomaly detection algorithms, namely LSTM Seq-Seq and one-class SVM with and without PCA using system call sequences data structure are discussed in the proposed work. The introspection technique is used to extract relevant data structures. Subsequently, the filtering and ordering techniques are applied to discard the redundant system calls and obtain valid sequences. To experiment, 1,000,000 samples as datasets are collected. Through the experiment, it is achieved that the LSTM Seq-Seq has the best performance for anomaly detection. It performed with an accuracy of 97.2% as compared to OCSVM. Furthermore, OCSVM with PCA gives better results than OCSVM without PCA.

8 Future Scope

The LSTM Seq-Seq and one-class SVM are applied with and without PCA for anomaly detection using system call sequences. Further, extend this work by applying new natural language processing-based algorithms like transformers, BERT, and variants of LSTM such as bi-directional LSTM and GRU. Furthermore for preprocessing, apart from PCA apply the LDA and autoencoder for feature extraction.

Acknowledgements This work is supported by TRMC, USA.

References

1. Peddoju SK, Upadhyay H, Lagos L (2020) File integrity monitoring tools: issues, challenges, and solutions. Concurr Comput Pract Exper e5825. https://doi.org/https://doi.org/10.1002/cpe.5825
2. Suresh Kumar P, Ramachandram S (2019) Fuzzy-based integration of security and trust in distributed computing. In: Soft computing for problem solving. Advances in intelligent systems and computing, vol 816. Springer, Singapore
3. Ligh MH, Case A, Levy J, Walters A (2014) The art of memory forensics
4. Xen Project (2013) available at https://www.xenproject.org/
5. Hizver J, Chiueh T (2014) Real-time deep virtual machine introspection and its applications.ACM SIGPLAN Notices 49(7)
6. Egele M, Scholte T, Kirda E, Kruegel C (2008) A survey on automated dynamic malware-analysis techniques and tools. ACM Comput Surv
7. Suresh Kumar P, Ramachandram S (2017)Fuzzy based integration of security and trust in distributed computing. In: Proceedings of Springer 7th international conference soft computing for problem solving (SocProS'2017). Indian Institute of Technology, Bhubaneswar, December 2017
8. Forrest S, Hofmeyr S, Somayaji A, Longstaff T (1996) A sense of self for UNIX processes. In: IEEE Security and Privacy
9. Joshi N, Choksi DB (2014) Implementation of process forensic for system calls. Int J Adv Res Eng Technol (IJARET) 5(6):77–82. ISSN 0976–6480(Print), ISSN 0976–6499
10. Lee W, Stolfo S, Mok K (1999) A data mining framework for building intrusion detection models. In: IEEE symposium on security and privacy, pp 120–132
11. Mahoney M, Chan P (2001) Detecting novel attacks by identifying anomalous network packet headers. Technical Report CS-2001-2
12. Suresh Kumar P, Pranavi S (2017) Performance analysis of machine learning algorithms on diabetes dataset using big data analytics. In: Proceedings of IEEE 2017 international conference on infocom technologies and unmanned systems (ICTUS'2017), Dubai, United Arab Emirates (UAE), December 2017, pp 580–585
13. Rishika Reddy A, Suresh Kumar P (2016) Predictive big data analytics in healthcare. In: Proceedings of IEEE 2016 second international conference on computational intelligence & communication technology (CICT), Ghaziabad, pp 623–626
14. Moser A, Kruegel C, Kirda E (2007) Limits of static analysis for malware detection. In: Annual computer security applications conference, ACSAC
15. Ye N, Li X, Chen Q, Emran SM, Xu M (2001) Probabilistic techniques for intrusion detection based on computer audit data. In: IEEE transactions on systems, man, and cybernetics
16. Zhang Q, Reeves D (2007) Metaaware: identifying metamorphic malware. In: Annual computer security applications conference.

17. Kang D-K, Fuller D, Honavar V (2005) Learning classifiers for misuse and anomaly detection using a bag of system calls representation. In: Annual information assurance workshop
18. Xiao H, Stibor T (2011) A supervised topic transition model for detecting malicious system call sequences. In: The workshop on knowledge discovery, modeling, and simulation
19. Debar H, Becker M, Siboni D (1992) A neural network component for an intrusion detection system. In: Proceedings of 1992 IEEE computer society symposium on research in security and privacy, pp 240–250. IEEE
20. Ryan J, Lin M-J, Miikkulainen R (1998) Intrusion detection with neural networks. Adv Neural Inform Proces Syst 943–949
21. Mukkamala S, Janoski G, Sung A (2002) Intrusion detection using neural networks and support vector machines. In: Proceedings of the 2002 international joint conference on neural networks, 2002. IJCNN'02, vol 2, pp 1702–1707. IEEE
22. Wang G, Hao J, Ma J, Huang L (2010) A new approach to intrusion detection using artificial neural networks and fuzzy clustering. Expert Syst Appl 37(9):6225–6232
23. Creech G, Jiankun Hu (2014) A semantic approach to host-based intrusion detection systems using contiguous and discontiguous system call patterns. IEEE Trans Comput 63(4):807–819
24. Suresh Kumar P, Upadhyay H, Bansali S (2019) Health monitoring with low power IoT devices using anomaly detection algorithm. In: IEEE conference SLICE-2019, Rome, Italy
25. Staudemeyer RC (2015) Applying long short-term memory recurrent neural networks to intrusion detection. South African Comput J 56(1):136–154
26. Staudemeyer RC, Omlin CW (2013) Evaluating performance of long short-term memory recurrent neural networks on intrusion detection data. In: Proceedings of the South African institute for computer scientists and information technologists conference, pp 218–224. ACM
27. Soni J, Prabakar N, Upadhyay H (2019) Deep learning approach to detect malicious attacks at system level. In: WiSec'19: Proceedings of 12th ACM conference on security & privacy in wireless and mobile networks, May 15–17, 2019, Miami, FL, USA, 2p
28. Soni J, Prabakar N (2018) Effective machine learning approach to detect groups of fake reviewers. In: Proceedings of the 14th international conference on data science (ICDATA'18), Las Vegas, NV
29. Soni J, Prabakar N, Upadhyay H (2019) Feature extraction through deepwalk on weighted graph. In: Proceedings of the 15th international conference on data science (ICDATA'19), Las Vegas, NV
30. Soni J, Prabakar N, Kim J-H (2017) Prediction of component failures of telepresence robot with temporal data. In: 30th Florida conference on recent advances in robotics
31. Thejas GS, Soni J, Chandna K, Iyengar SS, Sunitha NR, Prabakar N (2019) Learning-based model to fight against fake like clicks on Instagram posts. In: SoutheastCon 2019. Huntsville, Alabama, USA. In Press
32. Tejaswi U, Suresh Kumar P (2016) Diagnosing diabetes using data mining techniques. Int J Sci Res Publ 7(6):705–709
33. Chaudhary A, Peddoju SK, Peddoju SK (2020) Cloud based wireless infrastructure for health monitoring. Virt Mob Healthcare 34–55
34. Peddoju SK, Kavitha K, Sharma SC (2019) Big data analytics for childhood pneumonia monitoring. In: IGI global edited book, 2019, pp 1–17
35. Peddoju SK, Upadhyay H (2020) Evaluation of IoT data visualization tools and techniques. Data visualization. Springer, Berlin
36. Peddoju SK, Upadhyay H, Soni J, Prabakar N (2020) Natural language processing based anomalous system call sequences detection with virtual memory introspection. Int J Adv Comput Sci Appl (IJACSA) 11(5). http://dx.doi.org/https://doi.org/10.14569/IJACSA.2020.0110559
37. Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. Comput Sci
38. Schölkopf B, Platt JC, Shawe-Taylor J, Smola AJ, Williamson RC (2001)Estimating the support of a high-dimensional distribution. Neural Comput 13(7):1443–1471

SBP: Preparation of Schema Elements to Schema Matching

Aola Yousfi, Moulay Hafid El Yazidi, and Ahmed Zellou

Abstract Schema pre-matching is very critical for having schema elements fully ready for schema matching. Ideally, words are first extracted from the schema elements' labels, and then the semantically corresponding elements are generated accordingly. Searching for the sense of words based on their vertical and horizontal contexts, and before performing schema matching is very crucial to obtain high matching accuracy and as a result increase the amount of accurate matches and reduce the number of inaccurate matches and missed matches. Nonetheless, this problem is much more challenging than it seems. This is because complete and precise information about the meaning behind each element is often unavailable. This paper presents SBP, a Sense-Based Pre-matching approach designed for hierarchical data representations. SBP consists of two main components. First, the words sets generator generates, from each element, a set of words that fully describes its meaning. Second, the words qualifier identifies the senses of words based on both their vertical and horizontal contexts. Experimental results on real-world domains show high matching accuracy obtained when using the SBP approach.

Keywords Schema matching · Schema pre-matching · Semantic similarity · Vertical context · Horizontal context · Matching accuracy

---

A. Yousfi (✉) · M. H. El Yazidi · A. Zellou
Software Project Management Research Team, ENSIAS,
Mohammed V University, Rabat, Morocco
e-mail: aola.yousfi@gmail.com

M. H. El Yazidi
e-mail: my-hafid.elyazidi@um5.ac.ma

A. Zellou
e-mail: ahmed.zellou@um5.ac.ma

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_10

1 Introduction

Schema matching is very critical for applications that manipulate data across schemas of distinct data sources, examples of areas where this kind of applications are used include mainly data integration on the World Wide Web, data warehousing, e-commerce, scientific collaboration and bioinformatics. Schema matching not only requires a semantic comparison between elements from different schemas, but also needs an identification of the full meaning of each schema element before proceeding with the matching, which is called schema pre-matching or schema pre-processing. This paper presents two interesting observations for schema pre-processing. First, the labels of the schema elements are ambiguous. They often include acronyms (that correspond to the acronyms and abbreviations database entries), abbreviations (that also correspond to the acronyms & abbreviations database entries) and words (that correspond to the lexical dictionary entries) separated by underscores (e.g. *academic_conf_name*) or juxtaposed against each other (e.g. *academicConfName*). Second, the meanings, also known as senses, of words often change in different contexts.

Let S_1 and S_2 bet two schemas, and let e_1 and e_2 be two semantically corresponding elements (also called matches according to [1]) from S_1 and S_2, respectively. If e_1 and e_2 happens to use the same exact naming convention, then schema matching would be straightforward and pre-matching would not be a topic of discussion. Nevertheless, since there is no universal naming standard, schema pre-matching is very critical when matching new schemas. Therefore, plenty of schema matching systems have been introduced throughout the years (see [2–5] for recent surveys) to search for the matches in different schemas with the idea that semantically corresponding elements may likely be spelled differently. Although the state of the art schema matching systems obtain accurate results, they also obtain inaccurate matches and miss some accurate matches. As a result, these systems will remain completely dependent on human assistance in order to correct the output of the matching systems.

When matching a huge number of schemas, it is way much better to try to get as many accurate matches as possible right from the beginning. Hence, capturing the correct and complete meaning of schema elements prior to generating the semantically corresponding elements is very important to increase the total number of accurate matches and decrease the total number of inaccurate matches and missed matches. The problem of searching for the meanings of schema elements is not easy though.

Next, this paper presents an example that shows the importance of identifying words' senses prior to performing any schema matching operation.

Example 1.1 Let S_1 (see Listing 10.1) and S_2 (see Listing 10.2) be two snippets of two XML schemas describing the domain of organising conferences.

If the matching approach ignores schema elements contexts, then it will end up matching S_1.conference.rented_products.chairs to S_2.conference.organizing_committee.chair (which are not semantically similar). Nonetheless, if the matching approach takes into consideration schema elements contexts {conference, rented_products} for S_1.chairs and {conference, organizing_committee} for S_2.chair, then it will end up not matching S_1.conference.rented_products.chairs to S_2.conference.organizing_committee.chair (simply because rented_products and organizing_committee refer to two completely different real-world entities, which implies that S_1.chair and S_2.chair are in fact homonyms).

Listing 10.1 S_1

```
<?xml version="1.0"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
<xs:element name="conference">
 <xs:complexType>
  <xs:element name="title" type="xs:string"/>
  <xs:element name="date" type="xs:date"/>
  <xs:element name="address" type="xs:string"/>
  <xs:element name="rented_products">
   <xs:complexType>
    <xs:element name="chairs" type="xs:integer"/>
    <xs:element name="tables" type="xs:integer"/>
   </xs:complexType>
   </xs:element>
 </xs:complexType>
</xs:element>
</xs:schema>
```

Listing 10.2 S_2

```
<?xml version="1.0"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
<xs:element name="conference">
 <xs:complexType>
  <xs:element name="name" type="xs:string"/>
  <xs:element name="location" type="xs:string"/>
  <xs:element name="date" type="xs:date"/>
  <xs:element name="organizing_committee">
   <xs:complexType>
    <xs:element name="chair" type="xs:string"/>
    <xs:element name="program_committee" type="xs:string"/>
    <xs:element name="steering_committee" type="xs:string"/>
    <xs:element name="publicity_committee" type="xs:string"/>
   </xs:complexType>
  </xs:element>
 </xs:complexType>
</xs:element>
</xs:schema>
```

This paper introduces SBP, a Sense-Based Pre-matching approach designed for hierarchical data representations. SBP main idea is to prepare schema elements for the schema matching step. It suggests to generate, from each schema element, a words set that fully describes its sense. SBP deals with the problem of imprecise and sometimes unavailable information about the meanings of words as follows: it exploits schema elements' labels, and uses the horizontal and vertical contexts (see Definitions 3.3 and 3.4) of schema elements.

In summary, this paper makes the following concrete contributions:

- It introduces a new approach to capture the complete and correct meaning of schema elements.
- It proposes a novel algorithm that generates, from each schema element, a set of words that fully describes its meaning.

- It proposes a novel equation that determines the accurate sense of a word according to its horizontal and vertical contexts.
- It evaluates SBP on eight real-world domains and show that it significantly improves the matching accuracy: increases the total number of accurate matches, and decreases the total number of missed and inaccurate matches.

The remaining of this paper is structured as follows. Section 2 discusses related work. Section 3 defines the problem of schema pre-matching. Section 4 describes the architecture of SBP. Section 5 presents the experimental results. Section 6 concludes this paper and discusses future research directions.

2 Related Work

To the best of our knowledge, schema pre-matching has received very little attention from the research community compared to schema matching (see [2, 6–9] for surveys on schema matching and [10–17] for current matching tools). Moreover, current schema matching tools proceed directly with the matching as they start searching for semantically corresponding elements between schemas right from the beginning, and do not propose a solution to the pre-matching problem. In what follows, the paper will describe such state of the art matching systems.

PORSCHE [10] (Performance Oriented SCHEma mediation) is an automatic schema matching tool. It uses external thesauri and tree mining techniques, along with string-based information and language-based information. PORSCHE proceeds in three main steps. First, it transforms schemas into rooted ordered labeled trees. Second, it combines all schema trees into one single integrated schema. Third, PORSCHE captures the semantically corresponding elements between the integrated schema and the schema trees.

AgreementMakerLight (AML) [12] derives from AgreementMaker [18]. AML consists of two main modules: ontology loading module and ontology matching module. The ontology loading module loads the ontology files and then uses dictionaries to generate ontology objects. Note that the ontology loading module allows the virtual integration of new matching algorithms. The ontology matching module then aligns the ontology objects generated by ontology loading module.

COMA++ (COmbining Match Algorithms) [19, 20] is an updated version of COMA [19]. It transforms schemas into rooted directed acyclic graphs by means of external dictionaries as well as structure-based information, language-based information and instance-based information.

This section showed that none of the state of the art schema matching tools that are most relevant to our work takes into consideration the different senses a word might have in different contexts, and proposes a pre-processing strategy accordingly. Also, none of the state of the art schema matching tools assigns words sets to schema elements, and identifies the sense of words according to both their vertical and horizontal contexts.

3 Problem Statement

This section presents the definitions used throughout this paper, and states clearly the problem statement.

Definition 3.1 (*Schema element*) Let S be a schema. A schema element e refers to an element from S. Note that a schema element refers to both simple type elements and complex type elements. For example, in S_1 (see Listing 10.1), conference is a schema element and title is also a schema element.

Remark 1 In hierarchical data structures, this paper refers to the leaves by simple type elements, and the inner nodes as well as the root by complex type elements.

Definition 3.2 (*Words Set*) Let S be a schema and e an element from S. A words set θ generated from e is defined as follows: $\theta = \{w_1, w_2, \ldots, w_{|\theta|}\}$, where $\forall i \in [\![1, |\theta|]\!]$, w_i is a word from the lexical dictionary or the acronyms & abbreviations database. For example, the words set generated from chair in S_2 (see Listing 10.2) is defined as follows: $\theta_{\text{chair}} = \{conference, organizing, committee, chair\}$ (see Sect. 4 for more details on the words sets generation process).

Definition 3.3 (*Vertical context*) Let S be a schema and e be an element from S. The vertical context of e refers to all the complex type elements e is contained in. For example, in S_1 (see Listing 10.1), the vertical context of *chairs* is defined as follows: $\theta_{\text{VC}} = \{conference, rented, products\}$.

Definition 3.4 (*Horizontal context*) Let S be a schema and e be an element from S. The horizontal context of e refers to the elements surrounding e in S. For example, in S_1 (see Listing 10.1), the horizontal context of *chairs* is defined as follows: $\theta_{\text{HC}} = \{title, date, address, tables\}$.

Definition 3.5 (*Problem Statement*) Given a schema S, let ce be a complex type element from S and se be a simple type element contained in ce. Our main objective is to exploit the relationship between se and ce, apply some extracting techniques, and use external resources (a hierarchical lexical dictionary along with an acronyms and abbreviations database) to find out a words set that fully describes the meaning of se.

Table 1 lists the notations used throughout this paper.

The next section describes SBP, the solution to the problem presented in Definition 3.5.

Table 1 Summary of symbol notations

Notation	Description
$\mathbb{S}$, S, e	All input schemas, a schema from $\mathbb{S}$, an element from S
θ', $\|\theta'\|$, Θ', $\mathbb{T}'$	Words set generated from e, cardinality of θ', the sets of words generated from S, the sets of words generated from $\mathbb{S}$
θ, $\|\theta\|$, Θ, $\mathbb{T}$	Words set generated from e (with the senses of words identified), cardinality of θ, the sets of words generated from S (with the senses of words identified), the sets of words generated from $\mathbb{S}$ (with the senses of words identified)
w	Word
$\theta_{\mathbb{LD}}$	Set of words that corresponds to the lexical dictionary entries
θ_{acr}, θ_{abbr}	Set of words whose acronyms correspond to the acronyms & abbreviations database entries, set of words whose abbreviations correspond to the acronyms and abbreviations database entries
θ_{VC}, θ_{HC}	Vertical context, horizontal context
$\mathbb{LD}$, $DB_{\text{acr\&abbr}}$	Lexical dictionary, acronyms and abbreviations database

4 SBP: The Sense-Based Pre-matching

This section describes the solution to the schema pre-matching problem. The vast majority of current schema matching systems consider mainly hierarchical data structures, such as XML schemas. Hence, this paper focuses on the problem of preparing such schemas for the actual matching, leaving other data representations as future work.

The architecture of the Sense-Based Pre-matching approach (SBP) (see Fig. 1) includes two key components: words sets generator and words qualifier. Let $\mathbb{S}$ be a set of input schemas to match, and let $S \in \mathbb{S}$ be a schema, the *words sets generator* first applies fuzzy string matching techniques, and exploits a hierarchical lexical dictionary $\mathbb{LD}$ along with an acronyms & abbreviations database $DB_{\text{acr\&abbr}}$ so as to generate, from every schema element $e \in S$, a words set $\theta' \in \Theta' \in \mathbb{T}'$ that describes its meaning (Θ' refers to all the words sets generated from S, and $\mathbb{T}'$ refers to all the sets of words generated from $\mathbb{S}$). Then, for every θ', the *words qualifier* employs the horizontal context of e to determine the sense of words, therefore updating the words set from θ' to $\theta \in \Theta \in \mathbb{T}$ (Θ refers to all the words sets generated from S, which in addition to the information already given by Θ', Θ states the senses of words in terms of their context; and $\mathbb{T}$ refers to all the words sets generated from $\mathbb{S}$ which specifies the senses of words).

The rest of this section describes thoroughly the words sets generator in Sect. 4.1 and the words qualifier in Sect. 4.2.

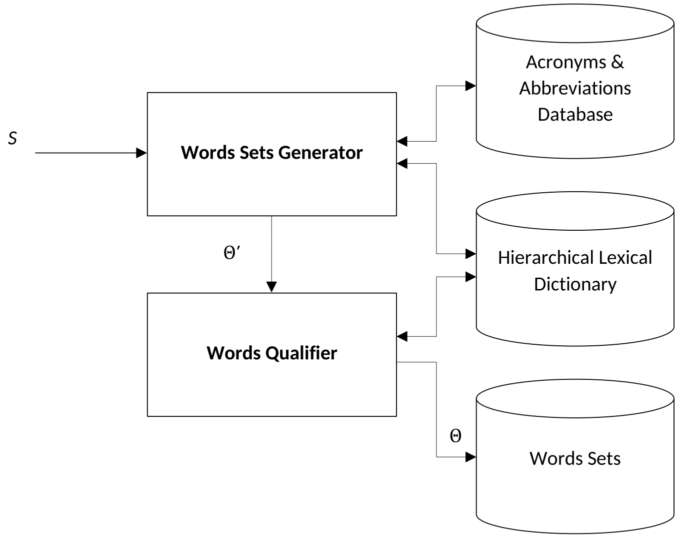


Fig. 1 The SBP architecture

4.1 Words Sets Generator

The words sets generator takes as input schemas $S \in \mathbb{S}$ and delivers as output words sets $\Theta' \in \mathbb{T}'$ that do not specify the senses of words just yet. Below, the paper will describe the three steps (summarized in Algorithm 1) the words sets generator goes through before it generates the sets of words.

Step 1: Extract, from each schema element, plain words that correspond to the lexical dictionary entries.

Simple type elements describe the data stored inside data sources. Nonetheless, the simple type elements' labels on their own do not fully describe the definition of the data. Hence, it was decided to use the simple type elements along with the complex type elements they are included in. But, here again a new challenge was faced: the schema elements labels are neither explicit nor complete. Therefore, given a schema element e (e can be a simple type element or a complex type element), the words sets generator uses fuzzy string matching techniques to extract from e words $\theta'_{\mathbb{LD}}$ that correspond to the hierarchical lexical dictionary entries [see formula (1)].

$$e \xrightarrow{\text{convert into}} \theta'_{\mathbb{LD}} \tag{1}$$

Then, the words sets generator assigns $\theta'_{\mathbb{LD}}$ to θ' as presented in formula (2).

$$\theta' \leftarrow \theta' \cup \theta'_{\mathbb{LD}} \tag{2}$$

Step 2: Substitute acronyms and abbreviations with their corresponding full forms.
The words sets generator exploits $DB_{\text{acr\&abbr}}$ to replace acronyms acr in e, if any, with their full forms θ'_{acr} (see formula (3)).

$$\text{acr} \xrightarrow{\text{convert into}} \theta'_{\text{acr}} \tag{3}$$

Then, it assigns the full forms' set θ'_{acr} to θ' as displayed in formula (4).

$$\theta' \leftarrow \theta' \cup \theta'_{\text{acr}} \tag{4}$$

Similarly, the words sets generator uses an acronyms & abbreviations database to replace abbreviations abbr in e, if any, with their full forms θ'_{abbr}, as presented in formula (5).

$$\text{abbr} \xrightarrow{\text{convert into}} \theta'_{\text{abbr}} \tag{5}$$

Then, it assigns the full forms' set θ'_{abbr} to θ' as shown in formula (6).

$$\theta' \leftarrow \theta' \cup \theta'_{\text{abbr}} \tag{6}$$

Step 3: Generate words sets.
Finally, the words sets generator enriches the sets assigned to the simple type elements in order to gain new insights [see formula (7)]: let *se* be a simple type element, the words sets generator converts *se* into a union of its words set and the sets of its vertical context θ'_{VC}.

$$\theta'_{se} \leftarrow \theta'_{se} \cup \theta'_{\text{VC}} \tag{7}$$

4.2 *Words Qualifier*

Given a schema element e, let $\theta' = \{w_1, w_2, \ldots, w_{|\theta'|}\} \in \Theta' \in \mathbb{T}'$ be the words set generated from e, $\forall\, i \in [\![1, |\theta'|]\!]$, $\exists j \geq 1$, such that w_i has j different senses. The words qualifier exploits the horizontal context $\theta'_{\text{HC}} = \{w_{\text{HC}_1}, w_{\text{HC}_2}, \ldots, w_{\text{HC}_{|\theta'_{\text{HC}}|}}\}$ of e, as shown in Eq. (8), so as to identify the appropriate sense of w_i in the given context.
$\forall\, i \in [\![1, |\theta'|]\!]$, $\exists j \geq 1$,

Algorithm 1 WordsSetsGenerator($\mathbb{S}$)

Input:
$\mathbb{S}$ *: Input schemas*
Output:
$\mathbb{T}'$ *: Words sets*

for each S in $\mathbb{S}$ **do**
  for each se in S **do**
    if $\exists\ w \in se$ and $w \in \mathbb{LD}$ **then**
      $\theta'_{se} \leftarrow \theta'_{se} \cup w$
    end if
    if $\exists$ abbreviation $abbr \in se$ and $abbr \in DB_{acr\&abbr}$ **then**
      Replace *abbr* with its full form
      Add its expanded form to θ'_{se}
    end if
    if $\exists$ acronym $acr \in se$ and $acr \in DB_{acr\&abbr}$ **then**
      Replace *acr* with its full form
      Add its expanded form to θ'_{se}
    end if
  end for
  for each θ'_{se} in Θ' **do**
    $\theta'_{se} \leftarrow \theta'_{se} \cup \theta'_{VC}$
  end for
end for
return $\mathbb{T}'$

$$\begin{aligned}\text{score}_j(w_i)_{1\le j\le|\text{senses}(w_i)|} = {} & |[\text{Definition}_j(w_i) \cup \text{Synonyms}_j(w_i) \cup \text{Examples}_j(w_i)] \\ & \cap [(\bigcup_{\substack{k=1\\ \theta'\setminus w_i}}^{|\theta'|} \bigcup_{q=1}^{|\text{senses}(w_k)|} (\text{Definition}_q(w_k) \cup \text{Synonyms}_q(w_k) \\ & \cup \text{Examples}_q(w_k))) \\ & \cup (\bigcup_{o=1}^{|\theta'_{\text{HC}}|} \bigcup_{p=1}^{|\text{senses}(w_{\text{HC}_o})|} (\text{Definition}_p(w_{\text{HC}_o}) \\ & \cup \text{Synonyms}_p(w_{\text{HC}_o}) \cup \text{Examples}_p(w_{\text{HC}_o})))]| \qquad (8)\end{aligned}$$

where

- $|\text{senses}(w_i)|$ refers to the total number of senses of w_i in $\mathbb{LD}$.
- $\text{Definition}_j(w_i)$ is the definition of the jth sense of w_i in $\mathbb{LD}$.
- $\text{Synonyms}_j(w_i)$ is the set of synonyms of the jth sense of w_i in $\mathbb{LD}$.
- $\text{Examples}_j(w_i)$ are the examples assigned to the jth sense of w_i in $\mathbb{LD}$.
- w_k denotes a word from $\theta' \setminus w_i$.
- Given a word $w_k \in \theta' \setminus w_i$, $|\text{senses}(w_k)|$ refers to the total number of senses of w_k in $\mathbb{LD}$.
- $\text{Definition}_q(w_k)$ is the definition of the qth sense of w_k in $\mathbb{LD}$.

- $\text{Synonyms}_q(w_k)$ is the set of synonyms of the qth sense of w_k in $\mathbb{LD}$.
- $\text{Examples}_q(w_k)$ are the examples assigned to the qth sense of w_k in $\mathbb{LD}$.
- Given a word $w_{\text{HC}_o} \in \theta'_{\text{HC}}$, $|\text{senses}(w_{\text{HC}_o})|$ refers to the total number of w_{HC_o} senses in $\mathbb{LD}$.
- $\text{Definition}_p(w_{\text{HC}_o})$ is the definition of the pth sense of w_{HC_o} in $\mathbb{LD}$.
- $\text{Synonyms}_p(w_{\text{HC}_o})$ is the set of synonyms of the pth sense of w_{HC_o} in $\mathbb{LD}$.
- $\text{Examples}_p(w_{\text{HC}_o})$ are the examples assigned to the pth sense of w_{HC_o} in $\mathbb{LD}$.

Then, the words qualifier assigns the sense j of w_i with the largest $score_j(w_i)$ to w_i:

$$\text{sense}(w_i) \leftarrow \text{sense}_j(w_i), \text{ such that } \text{score}_j(w_i) = \max(\text{score}_j(w_i))_{1 \le j \le |\text{senses}(w_i)|} \tag{9}$$

As a consequence, the following results are obtained such that the sense of every single word in θ' is stated clearly.

$$\theta' = \{w_1, w_2, \ldots, w_{|\theta'|}\} \xrightarrow{\text{convert into}} \theta = \{w_1\#n\#s, w_2\#n\#s, \ldots, w_{|\theta|}\#n\#s\} \tag{10}$$

where

- $\#n$ is short for noun.
- $\#s$ is the sth sense of w_i (where $i \in [\![1, |\theta|]\!]$) in $\mathbb{LD}$, such that $s \in \mathbb{N}$. $w_i\#n\#1$ is the first meaning of w_i in $\mathbb{LD}$, $w_i\#n\#2$ is the second meaning of w_i in $\mathbb{LD}$, and so on.

Algorithm 2 summarizes this.

Algorithm 2 WordsQualifier($\mathbb{T}'$)

Input:
$\mathbb{T}'$ *: Words sets before sense identification*
Output:
$\mathbb{T}$ *: Words sets with the senses identified*

for each Θ' in $\mathbb{T}'$ **do**
  for each θ' in Θ' **do**
    for each w in θ' **do**
      Calculate the score of w according to Eq. (8)
      Keep only the sense with the maximal *score*
    end for
  end for
end for
return $\mathbb{T}$

---

5 Experimental Results

The paper runs extensive experiments to assess SBP using real implementation. It mainly focuses on evaluating two key issues. First, it examines the accuracy of the generated words sets by comparing them against the reference sets. The reference words sets were found manually by a group of forty-five Ph.D. candidates from our university who also specified the exact sense of words based on the words contexts. Second, it verifies the ability of SBP to determine correct matches by applying it to some current schema matching tools.

5.1 Experimental Settings

Datasets: This paper experimented SBP on eight datasets (see Table 2) from TEL (Travel, Entertainment and Living) which are publically available on the Web.[1] The travel domain groups its dataset into three distinct sub-domains: *Car Rentals*, *Hotels* and *Airfares*. The Entertainment domain groups its dataset into three distinct sub-domains: *Music Records*, *Movies* and *Books*. And, the Living domain groups its datasets into two distinct sub-domains: *Automobiles* and *Jobs*.

Implementation: This paper first implements SBP using WordNet[2] [21] as our hierarchical lexical dictionary, and evaluates the accuracy of the sets of words it generates. Then, it uses SBP with COMA++ [19, 22] (COMA++$_{\mathrm{SBP}}$), PORSCHE [10] (PORSCHE$_{\mathrm{SBP}}$) and AML [9, 12, 18] (AML$_{\mathrm{SBP}}$), and compare the results to those obtained by COMA++, PORSCHE and AML.

Measures: This paper first exploits the metrics [19] defined in (11)–(14) to evaluate the accuracy of the words sets generated by SBP.

$$\mathrm{Precision}_{\mathrm{Sets}} = \frac{\text{Accurate Sets}}{\text{Accurate Sets} + \text{Inaccurate Sets}} \tag{11}$$

(11) identifies the percentage of accurate sets among all sets returned by SBP.

$$\mathrm{Recall}_{\mathrm{Sets}} = \frac{\text{Accurate Sets}}{\text{Missed Sets} + \text{Accurate Sets}} \tag{12}$$

(12) determines the percentage of accurate sets returned by SBP among all reference sets.

$$F\text{-Measure}_{\mathrm{Sets}} = \frac{2 \times \mathrm{Precision}_{\mathrm{Sets}} \times \mathrm{Recall}_{\mathrm{Sets}}}{\mathrm{Precision}_{\mathrm{Sets}} + \mathrm{Recall}_{\mathrm{Sets}}} \tag{13}$$

[1] http://metaquerier.cs.uiuc.edu/repository/datasets/tel-8/browsable.html.
[2] http://wordnetweb.princeton.edu/perl/webwn.

Table 2 Evaluation datasets

Domain	Total number of schemas
Car rentals	25
Hotels	39
Airfares	47
Music records	65
Movies	73
Books	65
Automobiles	84
Jobs	49

(13) is the harmonic mean of $\text{Precision}_{\text{Sets}}$ and $\text{Recall}_{\text{Sets}}$.

$$\text{Overall}_{\text{Sets}} = \text{Recall}_{\text{Sets}} \times \left(2 - \frac{1}{\text{Precision}_{\text{Sets}}}\right) \tag{14}$$

(14) determines the manual post-effort required to remove inaccurate sets and add missed sets. Unlike $\text{Precision}_{\text{Sets}}$ and $\text{Recall}_{\text{Sets}}$, $\text{Overall}_{\text{Sets}}$ may have negative values if $\text{Precision}_{\text{Sets}} < 0.5$. Note that if $\text{Overall}_{\text{Sets}} < 0$ then most of the pre-matching work is going to be performed manually.

The ideal case scenario is when $\text{Precision}_{\text{Sets}} = \text{Recall}_{\text{Sets}} = F\text{-Measure}_{\text{Sets}} = \text{Overall}_{\text{Sets}} = 1$.

The paper then employs the metrics [19] defined in (15)–(18) to evaluate the quality of the matches generated by $\text{COMA++}_{\text{SBP}}$, $\text{PORSCHE}_{\text{SBP}}$ and AML_{SBP}.

$$\text{Precision}_{\text{Matches}} = \frac{\text{Accurate Matches}}{\text{Accurate Matches} + \text{Inaccurate Matches}} \tag{15}$$

(15) determines the percentage of accurate matches among all matches generated by the matching tool.

$$\text{Recall}_{\text{Matches}} = \frac{\text{Accurate Matches}}{\text{Missed Matches} + \text{Accurate Matches}} \tag{16}$$

(16) identifies the percentage of accurate matches generated by the matching tool among all reference matches (i.e. matches found manually).

$$F\text{-Measure}_{\text{Matches}} = \frac{2 \times \text{Precision}_{\text{Matches}} \times \text{Recall}_{\text{Matches}}}{\text{Precision}_{\text{Matches}} + \text{Recall}_{\text{Matches}}} \tag{17}$$

(17) is the harmonic mean of $\text{Precision}_{\text{Matches}}$ and $\text{Recall}_{\text{Matches}}$.

$$\text{Overall}_{\text{Matches}} = \text{Recall}_{\text{Matches}} \times (2 - \frac{1}{\text{Precision}_{\text{Matches}}}) \tag{18}$$

(18) identifies the amount of manual post-effort necessary to remove inaccurate matches and add missed matches. Different from $\text{Precision}_{\text{Matches}}$ and $\text{Recall}_{\text{Matches}}$, $\text{Overall}_{\text{Matches}}$ might have a negative value if $\text{Precision}_{\text{Matches}} < 0.5$. Note that if $\text{Overall}_{\text{Matches}} < 0$ then almost all the matching work will be performed manually. Ideally, $\text{Precision}_{\text{Matches}} = \text{Recall}_{\text{Matches}} = F\text{-Measure}_{\text{Matches}} = \text{Overall}_{\text{Matches}} = 1$.

5.2 Results and Discussions

Figure 2 displays the $\text{Precision}_{\text{Sets}}$, $\text{Recall}_{\text{Sets}}$, $\text{Overall}_{\text{Sets}}$ and $F\text{-Measure}_{\text{Sets}}$ obtained by SBP.

The findings (graph in Fig. 2) indicate that SBP reaches a high accuracy. The sets of words generated by SBP are very similar to the reference words sets. This is very promising.

The full matching results in terms of $\text{Precision}_{\text{Matches}}$, $\text{Recall}_{\text{Matches}}$, $\text{Overall}_{\text{Matches}}$ and $F\text{-Measure}_{\text{Matches}}$ are shown in Fig. 3.

The findings (graphs in Fig. 3) indicate quite similar results obtained by AML_{SBP}, $\text{PORSCHE}_{\text{SBP}}$ and $\text{COMA++}_{\text{SBP}}$ this is because SBP produces accurate and complete definitions for every schema element which helps improve the accuracy of the matches. The results also indicate that AML_{SBP}, $\text{PORSCHE}_{\text{SBP}}$ and $\text{COMA++}_{\text{SBP}}$ outperform AML, PORSCHE and COMA++, respectively, in terms of $\text{Precision}_{\text{Matches}}$, $\text{Recall}_{\text{Matches}}$, $F\text{-Measure}_{\text{Matches}}$ and $\text{Overall}_{\text{Matches}}$.

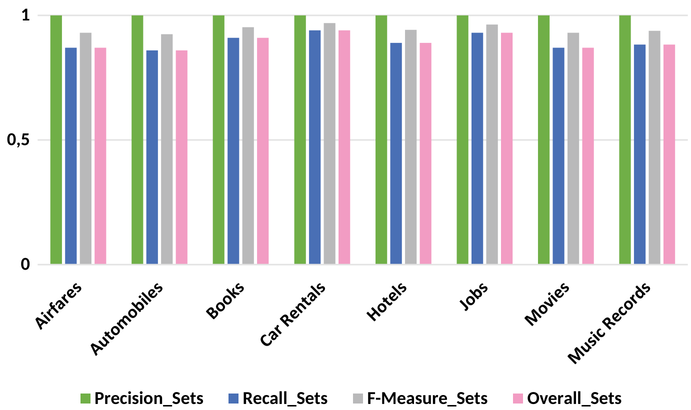


Fig. 2 The results obtained by SBP

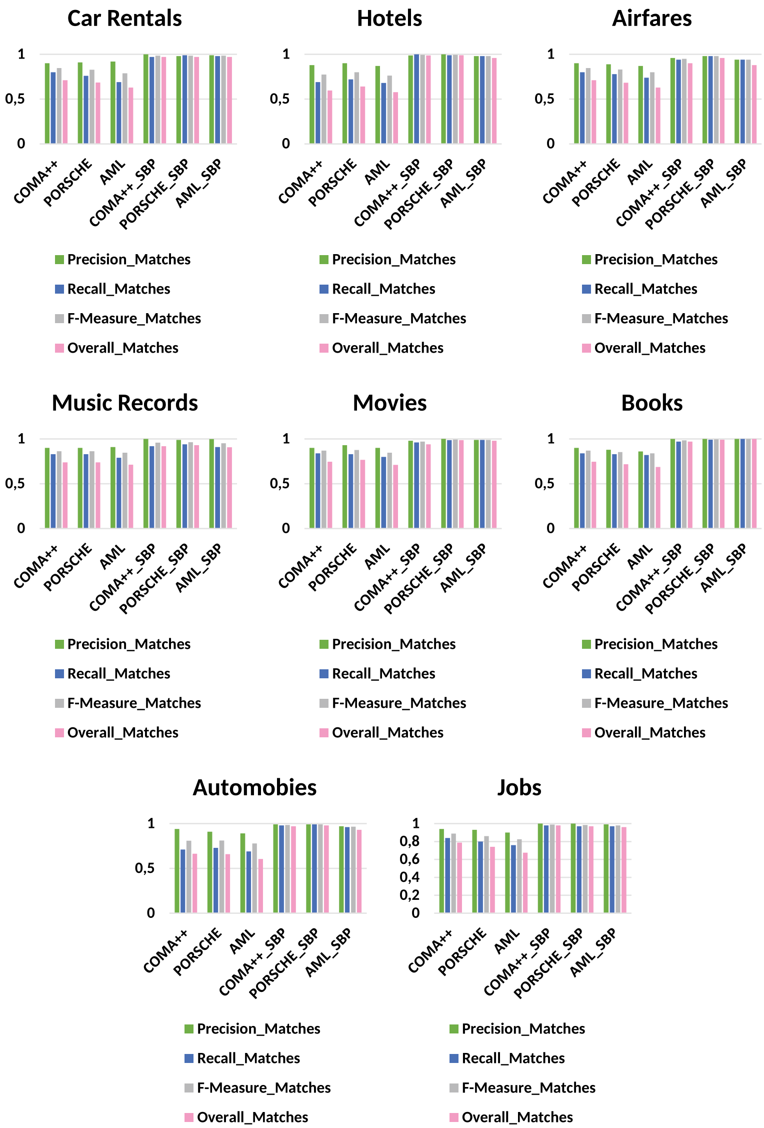


Fig. 3 The results obtained by AML_{SBP}, $\text{PORSCHE}_{\text{SBP}}$, $\text{COMA++}_{\text{SBP}}$, AML, PORSCHE and COMA++

6 Conclusions and Future Work

This paper has demonstrated that schema pre-matching is very critical for obtaining high matching accuracy. The state of the art matching tools do not take into consideration the fact that the sense of words often changes according to the context. Hence, this paper introduces an unprecedented a Sense-Based Pre-matching approach (SBP) that overcomes those limitations.

Given a schema S, our key idea is to prepare schema elements e for the actual matching. SBP captures the full meaning of e using both its vertical and horizontal contexts in S. It generates from e a words set θ that fully describes its meaning. This way, elements from different schemas are ready for the schema matching. This paper evaluated SBP on three state of the art matching tools over eight real-world domains. The results show that the matching tools applying SBP reach a superior matching accuracy. Future research work includes the following.

Consider other data structures. This paper focused mainly on hierarchical data structures and left other data structures for future work.

Study the impact of SBP on data source selection and ordering. Prior to answering the query, the system selects a subset of data sources that contain the complete or just a piece of the answer to the query (process is called source selection); next, the system orders the sources in a decreasing order of their coverage (given a query, a source coverage refers to the amount of answers contained in the source) (process is called source ordering). Thus, a future research direction would be to study the impact of SBP on source selection and ordering.

Work on schema matching. This paper focused on schema pre-matching; a future direction would be to come up with a holistic matching approach.

References

1. Zhang CJ, Chen L, Jagadish HV, Zhang M, Tong Y (2018) Reducing uncertainty of schema matching via crowdsourcing with accuracy rates. CoRR abs/1809.04017
2. Sutanta E, Wardoyo R, Mustofa K, Winarko E (2016) Survey: models and prototypes of schema matching. Int J Electr Comput Eng 2088-8708) 6(3)
3. Shvaiko P, Euzenat J (2005) A survey of schema-based matching approaches, pp 146–171
4. Yousfi, A, Yazidi MHE, Zellou A (2020) xmatcher: matching extensible markup language schemas using semantic-based techniques. Int J Adv Comput Sci Appl 11(8)
5. Bernstein PA, Madhavan J, Rahm E (2011) Generic schema matching, ten years later. PVLDB 4(11):695–701
6. Shvaiko P, Euzenat J (2013) Ontology matching: state of the art and future challenges. IEEE Trans Knowl Data Eng 25(1):158–176
7. Otero-Cerdeira L, Rodríguez-Martínez FJ, Gómez-Rodríguez A (2015) Ontology matching: a literature review. Expert Syst Appl 42(2):949–971
8. Ardjani F, Bouchiha D, Malki M (2015) Ontology-alignment techniques: survey and analysis. Int J Mod Educ Comput Sci 7(11):67

9. Faria D, Pesquita C, Balasubramani BS, Tervo T, Carriço D, Garrilha R, Couto EM, Cruz IF (2018) Results of AML participation in OAEI 2018. In: Proceedings of the 13th international workshop on ontology matching co-located with the 17th international semantic web conference, OM@ISWC 2018, Monterey, CA, USA, 8 Oct 2018, pp. 125–131
10. Saleem K, Bellahsene Z, Hunt E (2008) PORSCHE: performance oriented schema mediation. Inf Syst 33(7–8):637–657
11. Yazidi MHE, Zellou A, Idri A (2015) Fgav (fuzzy global as views). AIP Conf Proc 1644(1):236–243
12. Faria D, Pesquita C, Santos E, Palmonari M, Cruz IF, Couto FM (2013) The AgreementMaker light ontology matching system. On the move to meaningful internet systems: OTM 2013 conferences—confederated international conferences: CoopIS, DOA-trusted cloud, and ODBASE 2013, Graz, Austria, 9–13 Sept 2013. Proceedings, pp 527–541
13. Yousfi A, Yazidi MHE, Zellou A (2018) Assessing the performance of a new semantic similarity measure designed for schema matching for mediation systems. In: Computational collective intelligence—10th International conference, ICCCI 2018, Bristol, UK, 5–7 Sept 2018. Proceedings, Part I, pp 64–74
14. Yazidi MHE, Zellou A, Idri A (2013) FMAMS: fuzzy mapping approach for mediation systems. IJAEC 4(3):34–46
15. El Yazidi MH, Zellou A, Idri A (2012) Towards a fuzzy mapping for mediation systems. In: 2012 IEEE international conference on complex systems (ICCS), pp 1–4
16. Bourennani E, Bourque M (2019) A content-based schema matching tool. World Comput Sci Inf Technol J 9(5)
17. Ding G, Sun S, Wang G (2020) Schema matching based on SQL statements. Distrib Parallel Databases 38(1):193–226
18. Cruz IF, Antonelli FP, Stroe C (2009) AgreementMaker: efficient matching for large real-world schemas and ontologies. PVLDB 2(2):1586–1589
19. Do HH, Rahm E (2002) COMA—A system for flexible combination of schema matching approaches. In: Proceedings of 28th international conference on very large data bases, VLDB 2002, Hong Kong, 20–23 Aug 2002, pp 610–621
20. Massmann S, Engmann D, Rahm E (2006) COMA++: results for the ontology alignment contest OAEI 2006. In: Proceedings of the 1st international workshop on ontology matching (OM-2006) collocated with the 5th international semantic web conference (ISWC-2006), Athens, Georgia, USA, 5 Nov 2006
21. Miller GA (1995) Wordnet: a lexical database for english. Commun ACM 38(11):39–41
22. Aumueller D, Do HH, Massmann S, Rahm E (2005) Schema and ontology matching with COMA++. In: Proceedings of the ACM SIGMOD international conference on management of data, Baltimore, Maryland, USA, 14-16 June 2005, pp 906–908

Acceptance of Biometric Authentication Security Technology on Mobile Devices

William Ratjeana Malatji, Tranos Zuva, and Rene Van Eck

Abstract The development of mobile devices is quick and changes our daily personal and business lives. Every mobile user wants to be sure about individual data security, and for this reason, biometrics come into existence for mobile devices. Many studies were conducted on the acceptance of biometric authentication technology, but only a few of these studies focused on mobile devices-based biometry and the current study based on the mobile technology. To observe the reliability of the broadcast services, it is essential to offer better security for the biometry mobile phones. The limitations of this study were addressed by proposing a new mobile biometric technology acceptance model (MBTAM) that contains perceived humanness (PH), perceived interactivity (PI), and perceived social presence (PSP). The combined model for this quantitative study was tested on 302 mobile users through the distribution of the survey questionnaire, and examined by using the statistical package for social science (SPS). The results indicate that only one variable of the proposed model is not supported, which calls for further research. Furthermore, the functional elements of the research model become more prominent on the customer's intention to practice the mobile biometric device than the social elements. The research contributes to academic by suggesting new constructs that join together MBTAM to evaluate the possibility of mobile users to accept biometric authentication technology.

Keywords Mobile biometric technology acceptance model (MBTAM) · Perceived humanness (PH) · Perceived interactivity (PI) · Perceived social presence (PSP) · Statistical package for social sciences (SPSS)

W. R. Malatji (✉) · T. Zuva · R. Van Eck
Department of ICT, Vaal University of Technology, Andries Potgieter Blvd, Johannesburg 1911, South Africa
e-mail: villywr@gmail.com

T. Zuva
e-mail: tranosz@vut.ac.za

R. Van Eck
e-mail: rene@vut.ac.za

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_11

1 Introduction

In a technological era, mobile devices are most increasingly used for basic communications as well as a tool for managing individual issues and processing data obtained from anywhere at any time [1]. Over recent years, information access from mobile devices has become mainstream both in business and personal environments. The world is turning out to be more connected and every mobile user wants to be sure about individual data security [2].

Mobile device services assist as the base for business transactions but the traditional way of providing the security privileges is represented in terms of a mixture of alphanumeric and symbols. This ancestral process leads the users to avoid using mobile devices for reaching business data [3]. With the increase of its functionality including mobile banking, internet access, remote work, e-commerce, and entertainment, more confidential data is stored on these devices. For these reasons, biometrics comes in existence for mobiles [2].

To intensify the reliability of Wi-Fi services over mobile phones, a new trending and advanced technology have emerged that is biometric technology for mobile devices to promote the security levels [4]. Biometric technology refers to any technique that reliably uses measurable physiological or behavioral characteristics of distinguishing one individual from another [5].

Many studies were carried out on the acceptance of biometric devices and applications, users' attitudes towards such devices, and measurements of impact on performance. However, only a few of the studies focused on the factors that affect the acceptance of biometric devices [6]. Many studies have insisted on an investigation behind the biometric technology and stated the issues which are faced with user acceptance [7]. The acceptance of biometrics for other technologies still needs to be investigated deeply [8].

There were very few studies that measured the acceptance of biometric authentication technology on mobile devices. Therefore, this study efforts to regulate the reception of biometric corroborate technology on mobile devices.

The layout of the article is arranged in the following manner. Section 2 describes the related works of the proposed system. Section 3 denotes the significance of the study. Section 4 describes the methodology for the proposed system. Section 5 illustrates the results, and Sect. 6 reviews the discussions. Section 7 proposes the future scope of the research, and finally, Sect. 8 concludes the research work.

2 Literature Review

In literature, there have been many research studies on the acceptance of biometric devices and applications, users' attitudes towards such devices, and measurements of impact on performance. However, only a few of those studies focused on the factors that affect the acceptance of biometric devices [6]. Besides, each one of those examinations analyzed the adequacy of biometric procedures, however, do not contemplate the purposes for such acceptability. According to [9], many studies discussed the acceptance of technology, and the studies focused on technical issues such as algorithms, accuracy performance, etc.

The survey was carried out with 1206 respondents with the age of 18 years and above to find out the level of the acceptance of biometric technology (specifically facial recognition) from the Australian public [10]. This was achieved by asking how acceptable they thought it was if this technology was to be used in certain circumstances. It was found that 95% of respondents supported that the security can be used by airport staff as a way of passenger identification on police watch-lists. A similar report suggested with accuracy 92% of respondents have confirmed the security procedures chosen by the police for identifying the culprits in the criminal cases are of the video footage gathered through the security cameras. Among the survey report, quarter of the respondents weighed that this technology is not preferable for acceptance. One part of the respondents was bothering about the reflections of social media across these technologies (for example, Twitter, Facebook, and so on). It was found that 50% of the respondents declared this was an unacceptable technology to be applied [11].

Researchers conducted a review predicted on the physiological and behavioral biometric methods for user acceptance [12]. Later observed that these methods are rated very feeble in general except for fingerprint, voice, and hand geometry. All the above-mentioned studies have not been conducted based on mobile biometric devices.

According to [13], few studies have been conducted on mobile biometric devices and the good including the bad side of such devices were also discussed. Research conducted on both the pros and cons of the particular technique where there is no clear idea stated for the factors affecting the usage of biometric authentication technologies through mobile devices. The outcome of these factors is affecting the workplace, education, government sectors, and so on. Due to this report, there exists a phenomenon of technology for user acceptance [13].

Investigators studied modern mobile supporters towards their PDAs [14]. The particular biometric strategies were presented as elective confirmation measures to make sure about their mobile phones and observed that respondents reflected all techniques positively. The impediment of the investigation made by Clarke et al., Deane et al. and Furnell et al. [14–16] was that there was no attempt to comprehend the level of association concerning the members for biometrics on phones.

A portion of the effective determinants of biometric has been analyzed by Giesing [17] assessed the issues projected by the user and the social factors of biometric discovery. This examination leads to the new technology development towards the acceptance model designed by Davis [18].

3 Significance of This Study

By identifying the user acceptance issues from the research question, this research will at point consider how to address such issues to escalate the user acceptance of mobile biometric technology based on security. New devices are coming with biometric authentication security technology; however, few studies have tested the user acceptance of such technology on mobile devices. This examination will emphatically supplement the clients' consciousness of the biometric security reformation on cell phones. The findings of this study will assist decision-makers to be aware of the issues that affect users' decisions to welcome and utilize a specific system so that they would be capable of considering them during the development stage. It is hoped that this research would be beneficial to future researchers by providing them with helpful information about biometric authentication technology on mobile devices and some of their research questions may be answered by this study.

4 Methodology

4.1 Participants of the Study

Participants for this study were South African citizens in Vanderbijlpark. Three hundred and five (305) questionnaires were distributed to the target population. Only 302 responses were returned out of 305. The results of the demographic characteristics of the respondents are shown in Table 1.

Table 1 Questionnaire, source, and number of items

Constructs	Number of items	Source-citations
Perceived Ease of Use (PEOU)	4	Emily, Johnson and Carmen (2019)
Perceived Usefulness (PU)	4	Emily, Johnson and Carmen (2019)
Subjective Social Norm (SSN)	4	Barbara, Belanger and Schaupa (2017)
Perceived Humanness (PH)	3	Lankton, Knight and Tripp (2015)
Perceived Interactivity (PI)	3	Gao, Rau and Salvendy (2009)
Perceived Social Presence (PSP)	3	Lankton, Knight and Tripp (2015)
Intention to Use	2	Weng, Yang, Ho and (2019)
Actual Use of Mobile biometric device (AUMBD)	3	Asiimwe and Orebro (2015)
Trust	4	Cheng, Sun, Bilgihan and Okumus (2019)
Reliability	1	Tuunainen, Pitkanen and Hovi (2009)

4.2 Research Instruments

In this quantitative study, a simple random sampling technique was used to choose the participants. The items of this study in the survey questionnaire were constructed from the review of the related works that is appropriate to the research model. A five-point Likert—scale type measurement from one "strongly agree" to five "strongly disagree" was used in this study. After developing the questionnaire, it has been circulated to 30 participants (10% of the sample size) to ensure good clarity of questions, good length of instruments, and content completeness. The questionnaire is further sub-divided into two parts namely the first section and second section. The former part includes the details of the question linked to internet usage, technology expertise, demographics, and awareness of internet scams. The latter part consists of enquires about the estimation of the value of mobile biometrics (appropriate use of varied biometrics). Table 1 shows the questionnaire, source, and number of items. The questionnaire of this study was created based on the research framework derived from Ho et al. [19] shown in Fig. 1.

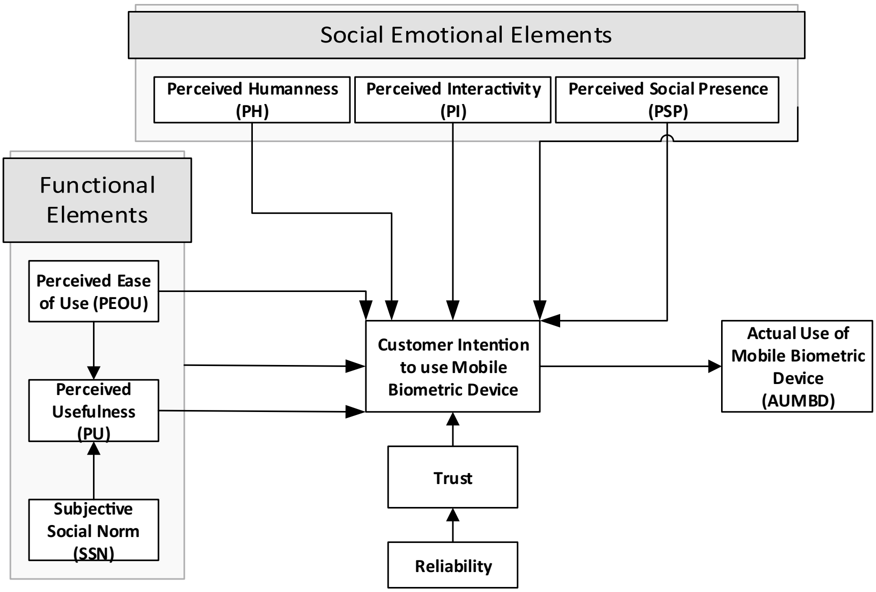


Fig. 1 Proposed mobile biometric technology acceptance model. *Source* Ho et al. [19]

5 Results

5.1 Demographic Characteristics

The data that is presented in Table 2 of this study provides the demographic characteristics of the respondents on age, gender, race, employment status, and the level of the study. The results indicated that 186 respondents were male and 116 were female, which shows that the number of male respondents is larger when compared to the number of female respondents. The greater number of the respondents is between 20 and 30 years of age with 69.9%, while the smallest is between 51 and 60 years of age with 2.3%. Considering the nature of mobile devices, this imbalance is understandable, because most mobile users are usually the youth [20]. Of the different races that participated in the study, the results indicated that 198 respondents were black, 87 were white, and 17 were other races. Regarding the participant's employment status, the results show that 8. 6% were self-employed, 25.5% were employed, and 1.0% retired, while 62.3% were students, and 2.6% other. It was further indicated in the results that on the level of the study, the majority of the respondents were undergraduate students with 38.7%, and the lowest was primary with only 3% (Tables 3, 4, 5 and 6).

5.2 Statistical Analysis

The displayed research model in Fig. 1 was evaluated by employing the statistical package for social sciences. The primary solution for factor analysis of this study revealed that the model was appropriate for factor analysis. The assumptions were tested, and it was found that the data contained no outliers, and the level of close to normality was excellent. The produced results indicated that the dependent variables do not violate the presupposition of linearity. Moreover, the results indicated that there is no presence of homoscedasticity and there is no multicollinearity. This shows that the statistical inferences made regarding the data may be reliable. In this study, items reliability test was performed and it was found that the reliability analysis of all variables was fairly high, which showed that the internal consistency among variables was robust and greater. Furthermore, items validity test was performed and the results indicated the satisfactory level of the construct validity of items.

5.3 Regression Analysis

The objective of this work is to measure user acceptance of biometric authentication technology on mobile devices, the analysis will focus on the main variables of acceptance in our acceptance model. The key variables of the customer's for the purpose

Table 2 Respondents demographic informations

Variable	Frequency	Percent (%)
Gender		
Male	186	61.6
Female	116	38.4
Age		
19 and Below	14	4.6
20–30	211	69.9
31–40	53	17.5
41–50	17	5.6
51–60	7	2.3
61 and Above	0	0
Race		
Black	198	65.6
White	87	28.8
Other	17	5.6
Employment status		
Self-employed	26	8.6
Employed	72	25.5
Retired	3	1.0
A student	188	62.3
Other	8	2.6
Level of study		
Primary	1	0.3
Secondary	11	3.6
Undergraduate	117	38.7
Postgraduate	97	32.1
Other	76	25.2
Do you own a mobile device		
Yes	294	97.4
No	3	1.0
Owned it before	5	1.7
Have you used biometric authentication security before		
Yes	215	71.2
No	87	28.8
Would you prefer to use a mobile biometric device		
Yes	256	84.8
No	15	5.0

(continued)

Table 2 (continued)

Variable	Frequency	Percent (%)
Not sure	31	10.3
I have accessed the internet using a mobile biometric device before		
Yes	141	46.7
No	161	53.3
Total	**302**	**100**

Table 3 Regression results of PU, PEOU, SSN, trust, PH, PI, PSP and intention to use

Model		Unstandardized coefficients		Standardized coefficients	*t*	Sig.
		B	Std. Error	Beta		
1	(Constant)	0.026	0.166		0.154	0.877
	PU	0.350	0.066	0.324	5.278	0.000
	POEU	0.195	0.063	0.162	3.054	0.011
	SSN	−0.054	0.038	−0.070	−1.418	0.157
	Trust	0.350	0.066	0.311	5.103	0.000
	PH	0.132	0.060	0.126	2.196	0.029
	PI	0.196	0.064	0.166	3.070	0.002
	PSP	0.211	0.052	0.229	4.030	0.000

[a]Dependent variable: intention to use

Table 4 Regression results of intention to use and AUMBD

Model		Unstandardized coefficients		Standardized coefficients	*t*	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.114	0.141		14.988	0.000
	Intention to use	0.224	0.072	0.177	3.107	0.002

[a]Dependent variable: actual use

Table 5 Regression results of PEOU, SSN, and PU

Model		Unstandardized coefficients		Standardized coefficients	*t*	Sig.
		B	Std. Error	Beta		
1	(Constant)	0.427	0.132		3.228	0.001
	POEU	0.589	0.048	0.576	12.340	0.000
	SSN	0.128	0.032	0.180	3.961	0.000

[a]Dependent variable: PU

Table 6 Regression results of reliability and trust

Model		Unstandardized coefficients		Standardized coefficients	*t*	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.207	0.182		6.634	0.000
	Reliability	0.466	0.052	0.468	8.971	0.000

[a]Dependent variable: trust

of using the mobile biometric devices (Intention to use) are PEOU ($\beta = 0.162$; $p < 0.05$), PU ($\beta = 0.324$; $p < 0.01$), PH ($\beta = 0.126$, $p < 0.05$), PI ($\beta = 0.166$; $p < 0.05$), PSP ($\beta = 0.229$; $p < 0.01$) and trust ($\beta = 0.311$; $p < 0.01$). The results indicate that trust and PU are the most important variables in explaining customer's intention to utilize the mobile biometric devices (Intention to use). Intention to use on its own is a key variable to AUMBD with ($\beta = 0.177$; $p < 0.05$). It is indicated in the results that PEOU is the most important variable that explains PU ($\beta = 0.576$; $p < 0.01$) succeeded by SSN ($\beta = 0.180$; $p < 0.01$). Moreover, reliability is the most important variable that explains trust with ($\beta = 0.468$; $p < 0.01$). The sum of functional elements of our model indicates that PEOU, PU, and SSN altogether, strongly explain intention to use with ($\beta = 0.390$; $p < 0.01$) (Fig. 2).

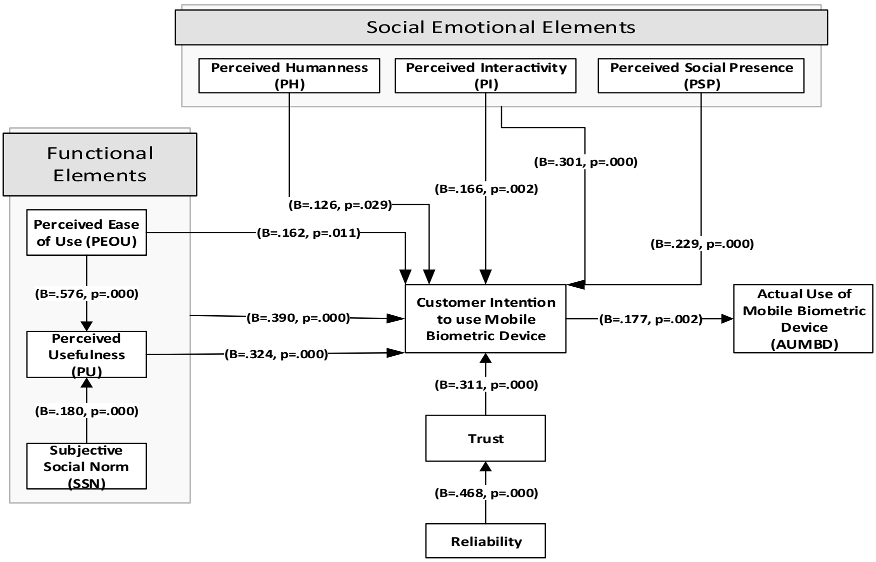


Fig. 2 Proposed Model for this study

6 Discussion

The overall mobile biometric acceptance model that is proposed in this study is validated. Starting with the functional elements (PEOU, PU, and SSN) of the model, the results indicated that PEOU has a positive influence on customers' intention to use mobile biometric devices (intention to use), and these results were supported by Suki and Suki [21]. This is an indication that when PEOU increases also intend to use increases. The results show that PU obtained impacts the positive plan to accept the usage of mobile, and these results are also in line with [21]. However, SSN on its own was not supported in this study. These same results were found on Chao [22]'s study on "factors determining the behavioral intention to use mobile learning: an application and extension of the ATAUT model." Based on the obtained results, it is concluded that PEOU and PU can be kept and used in future research to measure the acceptance of biometric authentication security technology on mobile devices. Although SSN is not supported, the variable on its own influence PU, moreover, the sum of all functional elements indicates a very strong influence on intention to use. Therefore, the conclusion cannot yet be made on whether the variable must be removed or not.

The social elements (PH, PI, and PSP) of the proposed model are all supported. It was indicated by the results that PI, PH, and PSP have a positive intention to use, and the results of these three variables are supported by Lankton [23]. Therefore, it is concluded that these variables can be kept and used in future research to estimate the user acceptance of biometric authentication technology on mobile devices [22]. Trust on its own is strongly influenced by reliability. Reliability is the most important variable that explains trust, and these results are in line with [7]. Intention to use on its own has a positive influence on AUMBD which is supported by Suki and Suki [21]. Based on these results, the conclusion can be made that trust, reliability, intention to use, and AUMBD can be kept and used in future research to measure the acceptance of biometric authentication technology on mobile devices.

7 Limitations and Suggestions for Further Research

This study focused on the two limitations as follows. Firstly, the study focused on the acceptance of biometric authentication technology on mobile devices only. Further research must be carried out on the acceptance of biometric authentication on other existing technologies except for mobile devices. The second important limitation of this study concerns gender and age of the respondents. The majority of the respondents for this study were male, and the highest age group of respondents was between 20 and 30. This brings about an issue of unbalanced results. Generally, both males and females in different age groups nowadays are using mobile devices. The conceptual framework used in this study should also be tested on the acceptance of biometric authentication technology on other existing technologies.

8 Conclusion

This study aimed to measure the acceptance of biometric authentication technology on mobile devices. The model that was used in this study proved to be valid, suitable, and supported. The researcher suggested that further research must be done especially using the variables that were supported in the model. The results and findings of this research showed that the majority of respondents acknowledged or are willing to accept biometric authentication technology to be used as security on mobile devices. However, further research needs to be conducted in this area.

Acknowledgements The author like to appreciate Professor Tranos Zuva and Doctor Rene Van Eck from the Vaal University of Technology for their support and supervision in this research. This is the revised version of the earlier paper. The author also likes to thank the IMITECH-2020's two anonymous reviewers for their valuable suggestions and comments throughout the research.

References

1. Wang H, Liu J (2009) Mobile phone-based health care technology. Recent Patents Biomed Eng 2(1):15–21
2. Kadena E, Ruiz L (2018) Adoption of biometrics in mobile devices. Obuda University, Doctoral school on safety and security sciences, Budapest Hungary. technologies to support teachers and improve practice
3. Bao P, Pierce J, Whittaker S, Zhai S (2011) Smartphone use by non-mobile business users. In: MobileHCI, Stockholm, Sweden. Attitudes and Practices: Computers & Security, vol 24, no 7, pp 519–527
4. Clarke N, Furnell S (2005) Authentication of users on mobile telephones
5. Kaur G, Kaur D (2013) Multimodal biometrics at feature level fusion using texture features. Int J Biometr Bioinform 7(1):58–73
6. James T, Pirim T, Boswell K, Reithel B, Barkhi R (2017) Determining the ıntention to use biometric devices: an application and extension of the technology acceptance model. J Organ End User Comput 18(3)
7. Chau A, Jamieson R (2004) Biometrics acceptance-perception of use of biometrics. Assoc Inform Syst
8. Uzoka FE, Ndzinge T (2009) Empirical analysis of biometric technology adoption and acceptance in Botswana. J Syst Softw 82:1550–1564
9. Chau A, Jamison R (2004) Biometric acceptance-perception of use of biometrics. In: ACIS 2004, Proceedings
10. Newspoll (2012) Rite aid deployed facial system in hundreds of Australia public. J Organ End User Comput 4(10):110–115
11. Unisys.Unisys security index report australia: facial recognition.https://www.unisyssecurityindex.com/system/resources/uploads/101/original/Australian2012.pdf
12. Miltgen L, Popovic C, Oliveira T (2013) Determinants of end-user acceptance of biometrics: ıntegrating the Big 3 of technology acceptance with privacy context. Decis Support Syst 56:103–114
13. Vrana R (2018) Acceptance of mobile technologies and m-learning in higher education learning: an explorative study at the faculty of humanities and social science at the University of Zagreb. Department for İnformation and Communication Science

14. Clarke NL, Furnell S, Rodwell PM, Reynolds PL (2002) Acceptance of authentication methods for mobile telephony devices 21(3):220–228
15. Deane F, Barrelle K, Henderson R, Mahar D (1995) Perceived acceptability of biometric security systems. Comput Secur 14(3):225–231
16. Furnell SM, Dowland PS, Illingworth HM, Reynolds PL (2000) Authentication and supervision: a survey of user attitudes. Comput Secur 19(6):529–539
17. Giesing I (2020) User perceptions related to ıdentification through biometrics within electronic business. University of Pretoria. https://upetd.up.ac.za/thesis/available/etd-01092004-141637/. Accessed 17 Feb 2020
18. Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quart 13:319–339
19. Ho G, Stephens G, Jamieson R (2003) Biometric authentication adoption ıssues. In: Presented at the proceedings of the 14th Australasian conference on ınformation systems, Perth, Western Australia, 26–28th November 2003
20. Hosokawa R, Katsura T (2018) Association between mobile technology use and child adjustment in early elementary school age: Plos One J 13(7)
21. Suki NM, Suki NM (2011) Exploring the relationship between perceived usefulness, perceived ease of use, perceived enjoyment, attitude and subscribers' intention towards using 3G mobile services. J Inf Technol Manage (2011)
22. Chao C (2019) Factors determining the behavioural intention to use mobile learning: an application and extension of the UTAUT model. Front Psychol 10
23. Lankton M, McKnight DH, Tripp J (2015) Technology, humanness and trust: rethinking trust in technology. J Assoc İnform Syst 16(10) (2015)

Impact of Agile Methodology Use on Project Success in South African Banking Sector

Makoena Moloto, Anneke Harmse, and Tranos Zuva

Abstract Agile methodology has become the most used software development methodology in different software communities. Besides the extensive benefits that agile methods offer, it also posseses several critical issues. Most studies have reported that, when comparing structured methodologies with agile methodologies, agile improves software quality, whereas other studies have contradicted and questioned the claimed benefits. In contrast, others argued that it does not have enough empirical evidence. This study aims at investigating the impact that agile method implementation on project success in the South African banking sector. The survey was conducted through questionnaires. The theoretical model for analyzing the impacts of agile method implementation on project success was proposed. Reliability was tested using Cronbach's alpha. The data were quantitatively analyzed by using correlation and regression approaches that use statistical package for social sciences. The results indicated that agile methodology implementation positively influences the project success. Furthermore, this study has indicated that reduced upfront planning, iterative delivery, environmental feedback and technical feedback as individual agile method use components positively impact project success in the South African banking sector.

Keywords Agile methodologies · Agile impact · Agile benefits · Project quality · Project success · Project performance

1 Introduction

The success of the project is usually measured by the perceived benefits to the user and the significant positive impact the project brings to the organization together with the return of investment [33]. Lately, industry experts have been keen to enhance performance at both the organization level and project level by adopting agile practices because agile enables organizations to stay more flexible [36]. Organizations

M. Moloto (✉) · A. Harmse · T. Zuva
Faculty of Applied and Computer Sciences, Vaal University of Technology, Vanderbjilpark, South Africa
e-mail: koenasm@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_12

following a plan-driven approach are primarily unable to accommodate changes during software development; they face poor communication between customers and developers, poor project management and high project cost [16]. They are seeking ways to stay competitive in the market share, to comply with relevant standards and deliver software on time and therefore adopt and use agile methodologies in their projects [7]. Agile seems to be the rational approach to deal with the instability of the market and to meet dynamic business needs [36].

Software development industries around the globe started changing their development process to agile since its manifesto was introduced over a decade ago [16]. Agile has formed in a manner that software development communities put people first to ensure transparency, trust and teamwork between stakeholders and to solve problems that the traditional software development methodologies could not solve [35]. It was created to ensure projects do not fail anymore because of people's problems, as demonstrated by most studies [6]. People are creative and that is why agile depends on them over a process and organizational maturity [6].

Agile methodology has been adopted in many software development projects due to its ability to better cope with frequent changes in requirements maturity [8]. It can improve the management of the development process and developer or user relationships [5]. Teams move to agile to improve quality, to reduce defects rates, to comply with standards, to deliver systems in time, under budget and to have a better interaction with customers [7]. Asnawi et al. [3] indicated that even if the use of agile is still emerging in Malaysia and its use is still forming, benefits are showing in the adoption of the methods. Agile ensures the delivery of quality software products on time [15].

1.1 Problem Statement

Regardless of how widely agile methodologies have been adopted and used as a means of solving the issues plan-driven methodologies could not solve, with a number of authors pointing the benefits and advantages it brings, very less empirical studies have been done to support the argument on project success [24] Many authors have written about the benefits agile brings to software engineering, many contradicted and questioned the claimed benefits. In contrast, others argued there is not enough empirical evidence of the claimed benefits [25].

Agile has been significantly adopted, progressed and used in the information system industries. Still, there are no neutral academic studies to its advantages. All the benefits claimed need more empirical studies [17]. The claims that are made by agile communities do not have enough scientific evidence supporting them [20]. Owen and Koskela [24] stated that because projects tend to be complicated by nature, it is essential to verify that when comparing agile to other methodologies, it improves projects for better and ensure the claimed benefits are true.

It is for these reasons that this study aimed to evaluate the impact of agile methodology use on project success in the South African banking sector.

Research questions. How does agile methodology use impact project success in the South African banking sector?

Aims and objectives. This study aimed to measure the impact of agile methodologies uses on project success in the South African banking sector.

The layout of this paper is as follows:

1. Introduction; to give an introduction to our research.
2. Related work; finding out what research has been done and if there is enough evidence on the impacts of agile methodology use on project success.
3. Proposed work; broadly outlining the research plan, research procedure, the research design and methods used in this paper.
4. Results analysis; explaining the findings of the research in detail and finally, the conclusion summarizing the research study.

2 Related Work

In the last years, agile development has been the most adopted process that has gained strong momentum [15]. A need for quality, efficient, reliable and useable software is continuously growing across the globe. As a result, a need for geographically accommodative software development methodologies rises [30]. Waterfall methodology exposes significant project risk because it relies on extensive and comprehensive planning; it also does not favor customers in changing their needs due to its sequential software development life cycle [9]. Fast software delivery, stable development process, ambiguous and/or evolving requirements are the popular motivations for adopting agile [37].

Agile is aiming at implementing the projects responsive to change by emphasizing less on the development of the rigorous project requirements and plans, but more on revising the requirements after every cycle [19].

The idea that agile is undisciplined and risky is a myth. Agile principles like stand up meetings keep communication and everyone engaged and informed of what is happening [28]. Daily stand-ups meetings improve communication between the team members, and this leads to trust between them and therefore improves the return of investment performance. They share knowledge and become self-organized [10]. It practices better and efficient communication methods between developers and customers [14]. When using agile, communication strategies improve, and stakeholders become more engaging. Hence, issues and faults get addressed in good times, and costs are reduced [2].

Agile iterations help communicate the progress of the project quickly and transparently than traditional methodologies. Software changes are then picked up as soon as possible because customers have a close relationship with the team members and progress feedback is given after every iteration [33]. To build high-quality projects on time, agile encourages customers to meet with developers regularly to validate and verify the requirements [29]. Velumani [36] agrees that agile methodologies helped many organizations to deal with the volatility of the market and to meet dynamic business needs after employing it in the past decade. When [15] compared agile techniques with traditional approaches, 89.6% of participants stated that with agile productivity was higher, 84.8% of participants stated that the quality is higher, and 86.5% stakeholders were highly satisfied. Agile helps teams become disciplined. They become better in communication, how to set goals and how to define acceptance criteria. Coordination and prioritization of activities can be enhanced through an integrated environment where testing can occur anytime, and system architecture should make parts of the system as independent as possible to avoid code conflicts across all project teams. Project organization structure hierarchy may also be introduced to enhance coordination [19].

People who successfully follow agile processes acquire their benefits such as delivery time improvements, quality improvements and development costs reduction [22]. Sidky and Arthur [29] suggested that it is important for organizations to analyze their past projects and define success factors. They need to ensure their organizational culture fits well with agile. Sidky and Arthur [29] further emphasize that people must know and understand agile principles through training and swallow it into their work culture. Vijayasarathy and Turk [37] surveyed early adopters in agile software development and mentioned that organizations that implemented agile methodologies had given positive feedback. Increased productivity with less defects, maintainable and extensible code with high test coverage and reduced time and costs are amongst the positive feedback given by the organizations which participated in a survey [37]. On a personal level, they mentioned that morale has improved with better collaboration and their customers are very satisfied, which is the key benefit of using agile development methodology. Agile adoption should not only be on an organization level but also a project level and be implemented within teams, not at a different time [28].

Agile software is deployed continuously, and the teams get to test the system regularly and continuously as they test every phase and therefore detect the issues at an early stage before they become more severe. Fewer defects mean better quality, which means more customer satisfaction [2]. Unlike traditional methods, faults are detected early and not at the end of the whole software development emphasized [11]. High-quality software is built because of the trust developers have with customers who also improves the initial states of the methodology. The owner's needs are met through onsite interaction from customers. Costs are reduced because changes are implemented early and the software is delivered incrementally [2].

Different projects in software development industries are developed and prototyped using agile methodology due to its ability to satisfy the changes in requirements, unlike the traditional development methods, which are not efficient in that regards [10]. Agile sees coding as the core of software development; it has, however, been criticized that code alone can lead to the loss of information due to lack of detailed documentation and models, especially in large complex systems. Those in favor argue that code is the base and the pillar, the only deliverable that matters in software creation and evolution, rather than analysis, designs and documentation [34]. Unlike traditional methods which spare a huge amount of time planning and documenting, and additional great effort of rework, defects are usually discovered early while they are embedded because products are verified during creation, which then becomes less costly, less time consuming and less effort to identify and remove them [7]. Plan-driven methodologies waste time documenting bulk requirements that might have to be reworked as the customers change their needs [9].

Generally, agile or scrum provides better results in software development. In the long term, it also increases quality and productivity [15]. When [35] made his research on The Adoption of Agile Software Development Methodologies by Organizations in South Africa, he mentioned that project visibility improved, team management regarding changing priorities improved, market time was faster, software quality and team morale also improved. Organizations switching from the traditional waterfall process model can increase the morale of the team and decrease large-scale software application risks [9]. An open-minded team, together with a strong scrum master, also helps to have a successful agile software development project [15].

When [22] researched adopting agile practices when developing medical device software, they mentioned that the project was completed seven per cent faster than if it were to be implemented using plan-driven approaches and stakeholders confirmed it would have overrun by 14% with budget overrunning by approximately seven per cent. A number of the benefits, including but not limited to productivity, project visibility and software quality, have been reported on adoption and use of agile methodologies [17]. The plan-driven methodologies processes do not allow changes to be adopted during implementation, therefore, end up with possibilities of obsolete product implemented, with costs increases and wasted resources because the initial requirements are no longer desired [9].

Agile teaches people to do just enough documentation needed. For people who have been on traditional methods for a while, where comprehensive and detailed documentation is the key, through training and educations, they will have a smooth, agile adoption [18]. Incorporating agile in software development can benefit organizations [9]. "Many of us have encountered people and teams who claim to be 'agile,' but who are just making this claim to avoid planning, documentation and design. If agile methods were simply about avoiding those aspects of software and systems delivery, then agile methods would be entirely inappropriate for safety–critical systems development. That is a misrepresentation of the intent and the practice of agile methods" [7]. Agile results in a high project satisfaction because it is adaptive and open to continuous customer feedback while accommodating changes due to its nature to have repeated requirements gathering, production deployable working source code, implementation of functionality and system testing [9].

Because of the freedom given by most agile methods, most people do follow the manifesto but not adhere to the specific agile methodology [15]. When using agile, projects become efficiently delivered because they are delivered in short iterations [13].

2.1 Agile Project Success

Project cost estimations using traditional software development methodologies have always been difficult. However, delivering the project on time, under budget, within scope has always been the critical aspect of a successful project [13]. Serrador and Pinto [27] used two dimensions to define project success, Project efficiency: defined by whether the scope goals of the project are met and whether the project is delivered on time within budget and Stakeholder success: defined as the best judges of the overall success measuring whether the project expectations are met and therefore satisfying the stakeholders. Project managers always find making software successful and increasing the customer satisfaction challenging, especially in an environment where changes come first not taking consideration of scope, time, cost and quality and not thinking of the positive or negative impact they might have to the project [38].

For a project to be declared successful, the scope has to be delivered on time under budget and be of good quality [23]. Projects that are not delivered on time may have sales lost and higher development costs, which may also result in customers not being satisfied emphasized [32]. It is crucial for organizations to understand the critical success factors of the projects to improve project management strategies, cost benefits, profitability and productivity. By so doing, they will positively impact social change and benefits management, employees and customers [23]. Planning and managing scope, time, cost, quality, risk, human resources, stakeholders and procurement management are the primary keys of project management with time and cost being the main issues [38].

Because systems are delivered over budget and costly, methodologies that can manage system development are needed [2]. The inappropriate choice of a project management methodology may contribute to the project being unsuccessful [1]. Agile helps the project manager and business development manager become more prepared for the ever-changing requirements of the clients. It also equips them with the understating of scope, cost and budget estimates, to avoid using the patterns without an understanding [13]. Requirements, scope and timeline change profoundly affect cost estimation in most cases. Agile companies, unlike companies using traditional methods, have their cost estimation at the beginning of every iteration, which helps them have accurate estimates easily. For the project to succeed, an organization must have agile engineering practices continuously integrated to deliver quality scope under budget, finalize all requirements before starting development to avoid bugs and glitches, ensure there are automated tests and builds in place, engage and motivate the team to achieve higher quality project deliverables, have stronger technical practices combined and have time-boxed iterations [28]**.** Managers must avoid having failing projects due to the methodologies which do not accommodate changes and always try and maximize profits and minimizes costs [2].

Regardless of software development difficulties in estimating costs, projects must be delivered under budget, within time, and the requirements must be met [13]. When using agile, projects become efficiently delivered because they are delivered in short iterations [13]. Agile put people first over the process. Creative and innovative team members help project success [2].

3 Proposed Work

The quantitative research method is used in this study. It is considered conclusive in its purpose, and it can be used to measure statistical results that are interpreted objectively [21]. This study used inferential statistic to analyze the impact of agile methodology use on project success in the South African banking sector. A research model in Fig. 2 for effectively analyzing the impacts of agile method use on project success in the South African banking sector was proposed and its effectiveness measured. Our model was derived from the theoretical model for analyzing the impacts of agile method adoption on project success by Tripp [33]. As seen in Fig. 2. The Model consists of two main variables, agile method use and project success. Agile method use has multiple components, and they are discussed below:

Reduced upfront planning determines how much time spent is reduced by the team before beginning the work. The reduced upfront planning is to directly impact the project performance by reducing time to initial feedback and waste from planning tasks too far in advance [33]. Iterative delivery determines how much time has taken the team to deliver the functional work in an iteration. It reduces ambiguity in requirements due to the feedback received from the users, which is from the working system rather than the abstract documents [33]. Environmental feedback determines the mechanisms utilized by the team to obtain feedback from the customers and stakeholders [33]. Technical feedback determines to what level the mechanisms have been utilized by the team to ensure that the system is functioning properly [33]. Project success is dependent on agile method use, and its components are discussed below.

Project performance determines whether the project results have matched the defined goals of scope schedule and budget. It is measured using the project budget outcome, project time outcome scale and project scope outcome [33]. Product quality determines whether the system is useful, reliable, complete, effective, suitable and accurate as perceived [33]. Project impacts determine to what extent or how positively has the project impacted the organization as perceived. The perceived benefits of the system are measured with two dimensions: Customer satisfaction and organizational impacts [33].

3.1 Sample Size and Profile Participants

The criteria for participating in our study were that a team is using agile software development methodology operating in the application domains of a South African banking sector and it had delivered software to a customer at least once. The respondents varied from a small team with less than 10 team members to large teams with several hundred team members. One hundred and fifty responses were collected, and 41 were incomplete and discarded. The remaining 109 complete responses were used for analysis. Our participants comprised mainly but not exclusively of IT stakeholders, IT management and team members (software developers, quality assurance personnel, project managers, business analysts, system analysts and architects).

3.2 Data Collection

The questionnaire was the key data collection instrument in this study. Data were quantitatively collected in South African banks. The data collection was administered using primary data. A non-probability purposive sampling method has been used for data collection. This sampling method was used because not all project teams in the South African banking sector are using the agile method. It was appropriate to make use of a purposive sample in this survey because sampling from outside the population, which is not using the agile development method, would be inappropriate. For us to gather as much data as possible, Snowball sampling is also used and asked the participants to indicate any other potential participants who are working in the banking sector on different projects and are using an agile methodology. For our respondents to take part in the survey, the questionnaire is placed online using google forms.

3.3 Data Analysis

The statistical package is used for Social Sciences (SPSS) V25 software to analyze our data. Figure 1 shows the methods used to evaluate our data. As shown in Fig. 1, the common weaknesses of the survey study (validity and reliability) were mitigated by using Cronbach's alpha coefficient and Kaiser–Meyer–Olkin (KMO). According to [12], Cronbach's alpha is not a statistical test. It measures the relationship between many items in a group, how closely related they are and therefore measure consistency. Cronbach's alpha coefficient ranges between 1 and 0. The closer the value approaches 1, the higher the reliability, and the closer it approaches 0, the lower the reliability [31]. George and Mallery [12] provided the following Cronbach's alpha rules of thumb: _ > 0.9—Excellent, _ > 0.8—Good, _ > 0.7—Acceptable, _ > 0.6—Questionable, _ > 0.5—Poor and _< 0.5—Unacceptable. To measure how adequate our sampling is, KMO is used. "KMO test reflects the sum of partial correlations relative to the sum of correlations" [4]. A measure of >0.9 is marvelous, >0.8 is meritorious, >0.7 is middling, >0.6 is mediocre, >0.5 is miserable, and <0.5 is unacceptable [12].

To test how strong the relationship is between our variables, correlation analysis has been used. Correlation analysis tests how strong the relationship is between two continuous variables given as a coefficient between -1 and 1. A score ranging -1

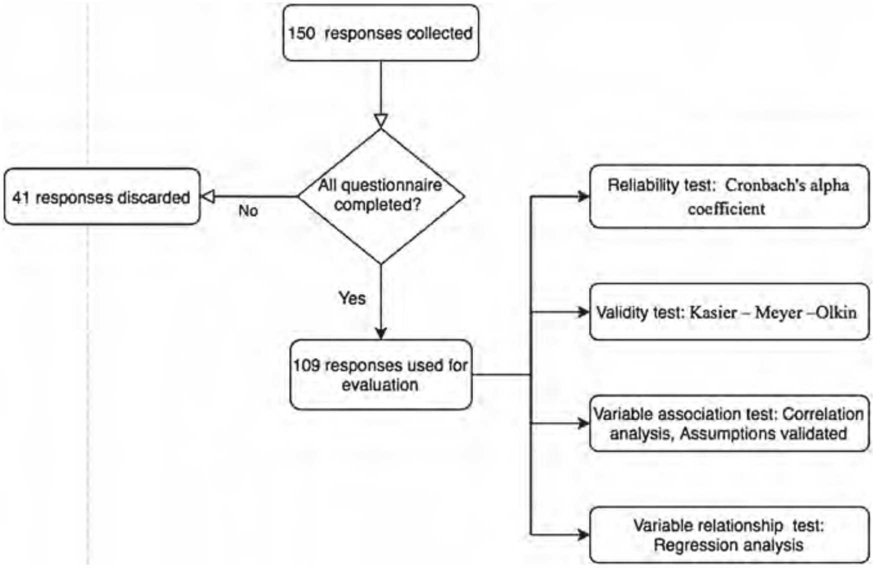


Fig. 1 Evaluation method

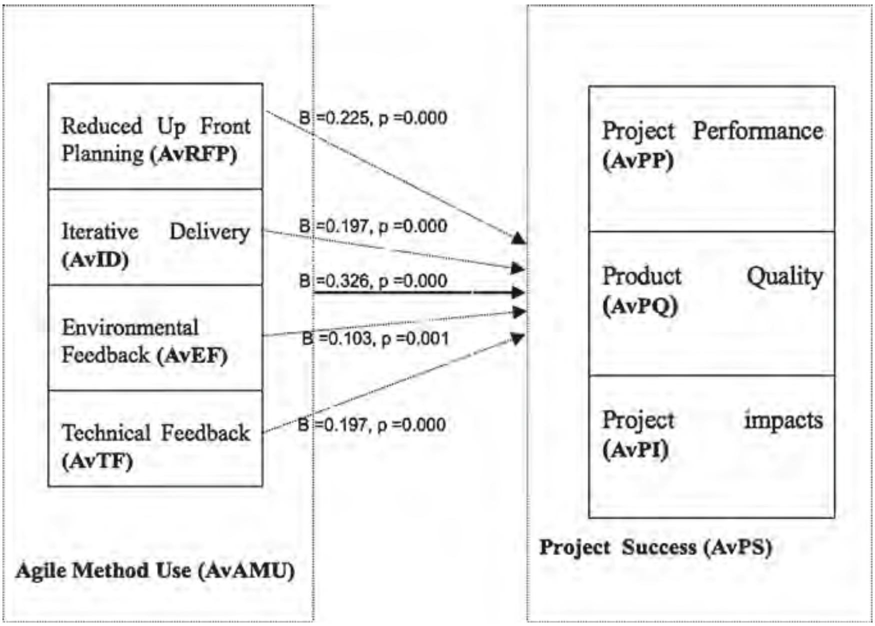


Fig. 2 Theoretical model for analyzing the impacts of agile method adoption on project success [33]

1 is perfect, −0.9 to −0.7 0.7 to 0.9 is strong, −0.6 to −0.4 0.4 to 0.6 moderate and −0.3 to −0.1 0.1 to 0.3 weak [12].

Regression analysis is also used to indicate how significant the relationship is between our independent and dependent variables. Regression indicates the effects of the relative strength of different independent on a dependent variable [26].

4 Results Analysis

The reliability and validity were measured, and the following results were recorded as seen in Tables 1 and 2.

A score 0.801 is indicated in our results. As [12] indicated, our results show high reliability of our variables.

KMO was used to test the validity of our study and indicated a value sampling adequacy of 0.687. According to [12], this is acceptable. These variables have proven to be statistically significant at the 1% level.

4.1 Correlation

To test how strong the relationship is between our variables, a correlation analysis has been performed. The results are shown in Table 3.

From Table 3, it can be seen that there is a positive relationship between the individual agile method use variables and project success. The correlation showed that reduced upfront planning (Pearson correlation 0.467, Sig. 0.000), iterative delivery (Pearson correlation 0.356, Sig. 0.000), environmental feedback (Pearson correlation 0.327, Sig. 0.000), technical feedback (Pearson correlation 0.403, Sig. 0.000), agile

Table 1 Cronbach's alpha

Reliability statistics	
Cronbach's alpha	No. of items
0.801	9

Table 2 Kaiser–Meyer–Olkin

KMO and Bartlett's test		
Kaiser–Meyer–Olkin measure of sampling adequacy		0.687
Bartlett's test of sphericity	Approx. Chi-Square	2728.856
	Df	1225
	Sig.	0.000

Table 3 Correlation analysis for agile method use variables and project success

Correlations						
		AvRFP	AvID	AvEF	AvTF	AvPS
AVPS	Pearson correlation	0.467**	0.356**	0.327**	0.403**	1
	Sig. (2-tailed)	0.000	0.000	0.001	0.000	
	N	109	109	109	109	109

**Correlation is significant at the 0.01 level (2-tailed)

method use (Pearson correlation 0.533, Sig. 0.000) have a positive relationship with project success.

4.2 Regression

The direct effects of the agile method use on project success were tested. Regression analysis was used to determine the actual contribution of agile method use on project success, and the results are shown in Tables 4 and 5.

In Table 4, it is indicated that the influence of the individual agile method use variables on project success is positive. The following variables were noted; reduced upfront planning influences project success with a B coefficient values of 0.225 and *p*-value of 0.000, iterative delivery influences project success with B coefficient value of 0.197 and *p*-value of 0.000, environmental feedback influences project success with a B coefficient value of 0.103 and *p*-value of 0.001, and technical feedback influences project success with a B coefficient value of 0.197 and *p*-value of 0.000. These variables have proven to be statistically significant at the 1% level. The equation for the use of agile method variables to project success is as follows:

$$AvPS = 0.225AvRFP + 0.197AvID + 0.103AvEF + 0.197AvTF + 2.5, \tag{1}$$

where AvPS is the project success depends on individual agile method use components. AvRFP is reduced upfront planning determining the time reduced by the team, AvID is iterative delivery determining the time taken by the team to deliver a functional code, AvEF is environmental feedback determining the feedback mechanism used by the team, and AvTF is technical feedback determining the feedback provided by the team mediating technology.

Using regression analysis, the direct effects and the actual contribution of agile method use as a whole on project success were tested. The results are seen in Table 5.

The results indicated that the influence of the extent of agile method use as a whole on project success is positive. Table 5 indicates that the agile method use positively

Table 4 AVPS dependent on AvAMU components (AvRFP, AvID, AvEF, AvTF)

Coefficients[a]

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.	95.0% confidence interval for B	
		B	Std. error	Beta			Lower bound	Upper bound
1	(Constant)	2.423	0.117		20.796	0.000	2.192	2.654
	AvRFP	0.225	0.041	0.467	5.461	0.000	0.143	0.306
	AvID	0.197	0.050	0.356	3.940	0.000	0.098	0.296
	AvEF	0.103	0.029	0.327	3.584	0.001	0.046	0.159
	AvTF	0.197	0.043	0.403	4.561	0.000	0.111	0.283

[a]Dependent variable: AVPS

Table 5 AVPS dependent on AvAMU

Coefficients[a]

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.	95.0% confidence interval for B	
		B	Std. error	Beta			Lower bound	Upper bound
1	(Constant)	2.459	0.110		22.325	0.000	2.241	2.678
	AvAMU	0.326	0.050	0.533	6.510	0.000	0.227	0.425

[a]Dependent variable: AVPS

influences project success with a B coefficient value of 0.326 and p-value of 0.000 proving that agile method use is statistically significant at the 1% level. The equation of project success in relation to agile methodology use is as follows:

$$AvPS = 0.326AvAMU + 2.5, \tag{2}$$

where AvPS is the project success, AvAMU is an agile method use theorized to likely directly be impacts project success.

This is our main equation for calculating the influence of project success in relation to agile method use. The results discussed above are summarized in Fig. 2 illustrating the influence of individual components of agile methodology use on project success and the influence of agile methodology use as a whole on project success.

5 Discussions of Key Findings

This study aimed to examine the impact that agile method use has on project success in the South African banking sector. This study was motivated by many studies that highlighted that the claims that were made about agile that it improves the project and organization for better have not been proven [24]. A model Fig. 2 was proposed and used to test our theory.

There have been numerous claims made by agile practitioners that agile improves team efficiency, team performance, software quality and organizational benefits [7]. In this study, agile method use had four components, reduced upfront planning aimed at determining the time reduced by the team before beginning the work. Iterative delivery aimed at determining the team taken to deliver the functional work. Environmental feedback aimed at determining the mechanisms utilized by the team to obtain feedback from the customers and stakeholders, and technical feedback aimed at determining the level at which the team used the mechanisms to ensure that the system is functioning properly [33]. These variables were theorized to directly impact project success and proven true.

Project success was measured using project performance, product quality and project impacts. Project performance aimed at determining whether the project results have matched the defined goals of scope schedule and budget. Product quality aimed at determining whether the system is useful, reliable, complete, effective, suitable and accurate as perceived, while project impacts aimed at determining the extent or how positively has the project impacted the organization as perceived [33].

In accordance with [33], our results indicated that the agile method uses positively impact project success. Furthermore, this study has indicated that reduced upfront planning, iterative deliver, environmental feedback and technical feedback as individual agile method use components positively impact project success in the South African banking sector.

6 Conclusion

The results of this research indicated that the agile method use positively impacts project success. In agile method use, the following variables: Reduced upfront planning, iterative delivery, environmental feedback and technical feedback influence project success. In future, it is necessary to understand the effect that the individual agile method use variables (reduced upfront planning, iterative delivery, environmental feedback and technical feedback) have on the individual project success components (project performance, product quality and project impacts). How project success impacts the organizational net benefits indicates a significant need for future research as well.

References

1. Ahimbisibwe A (2015) Critical success factors for outsourced software development projects from a Vendor's perspective: a structural equation modelling analysis of traditional plan-based and agile methodologies. Victoria University of Wellington
2. Altameem E, Mohammad AI (2015) Impact of agile methodology on software development. In: Computer and Information Science. Canadian Center of Science and Education, pp 9–14. doi: https://doi.org/10.5539/cis.v8n2p9
3. Asnawi AL, Gravell AM, Wills GB (2012) Emergence of agile methods: perceptions from software practitioners in Malaysia. Doi: https://doi.org/10.1109/AgileIndia.2012.14
4. Balasundaram N (2009) Factor Analysis: nature, mechanism and uses in social and management science research. J Cost Manage Acc XXXVII(2):15–25
5. Ceschi M et al (2005) Project management in plan-based and agile companies. University of Bolzano-Bozen
6. Chevers DA, Whyte CC (2015) The adoption and benefits of Agile software development methods in Jamaica. In: 2015 Americas conference on information systems, AMCIS 2015, pp 1–9. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84963616582&partnerID=40&md5=5823bedf6722a72069d185562a3b1e7a
7. Douglass BP, Ekas L (2012) Adopting agile methods for safety-critical systems development. Longman, Britain
8. Fergal M et al (2016) Introducing agile practices into MDevSPICE. Int J Adv Life Sci 8(1):133–142
9. Fergis K (2012) The impact of an agile methodology on software development costs. University of Pennsylvania Technical Papers, pp 1–16. Available at: https://repository.upenn.edu/cis_reports%0A
10. Gaurav Kumar PKB (2012) Impact of agile methodology on software development. Int J Comput Technol Electron Eng (IJCTEE) 2(4):46–50. https://doi.org/10.5539/cis.v8n2p9
11. Gaurav Kumar PKB (2012b) Impact of agile methodology on software development. Comput Inform Sci 8(2). doi: https://doi.org/10.5539/cis.v8n2p9
12. George D, Mallery P (2003) SPSS for windows step by step: a simple guide and reference. 11.0 update, 4th edn
13. Haider A (2017) Impact of agile methodologies on cost estimation techniques in software industry of Pakistan. IndustrEng Manage 6(03). doi: https://doi.org/10.4172/2169-0316.1000218
14. Hneif M, Ow SH (2009) Review of agile methodologies in software development 1. Int J Res Rev Appl Sci 1(1):2076–2734. ISSN:2076-734X, EISSN:2076-7366

15. Kapitsaki GM, Christou M (2015) Learning from the current status of agile adoption. In: Filipe J, Maciaszek L (eds) ENASE 2014, CCIS 551, pp 18–32. Springer International Publishing Switzerland 2015, pp 18–32. doi: https://doi.org/10.1007/978-3-319-27218-4
16. Kautsar E et al (2013) Challenges in adopting agile practices: perceptions of software practitioners in Indonesia. In: The 5th international conference on internet (ICONI) 2013. Malaysia, pp 1–9
17. Lagerberg L et al (2013) The impact of agile principles and practices on large-scale software development projects: a multiple-case study of two projects at Ericsson. Int Symp Empirical Softw Eng Meas 10:348–356. https://doi.org/10.1109/ESEM.2013.53
18. Mahanti A (2006) Challenges in enterprise adoption of agile methods—a survey. J Comput Inform Technol 03:197–206. https://doi.org/10.2498/cit.2006.03.03
19. Masood ZA, Farooq S (2017) The benefits and key challenges of agile project management under recent research opportunities. Int Res J Manage Sci 5(1):20–28. Available at: https://www.researchgate.net/publication/316239082
20. Mcbreen P (2002) Questioning extreme programming. Canada
21. McCrindle C (2008) Choosing and using quantitative research methods and tools. FIL-IDF world dairy summit and exhibition. Available at: https://www.up.ac.za/media/shared/624/choosing-and-using-quantitative-research-methods-and-tools.zp119932.pdf
22. Mchugh M, Fergal M, Coady G (2015) Adopting agile practices when developing medical device software adopting agile practices when developing medical device software. Ireland: creative commons attribution-noncommercial-share alike 3.0 License
23. Nguyen DS (2016) Success factors that influence agile software.pdf. Am Sci Res J EngTechnol Sci (ASRJETS) 17(1):172–222
24. Owen RL, Koskela L (2006) Agile construction project management. In: 6th international postgraduate research conference in the built and human environment, 6(7), pp 22–33. Available at: https://www.irbnet.de/daten/iconda/CIB9021.pdf
25. Penn DM (2016) Agile and conventional methodologies: an empirical investigation of their impact on software quality parameters. IOSR J Econ Finance 3(1):56. https://doi.org/10.3929/ethz-b-000238666
26. Sarstedt M, Mooi E (2014) A concise guide to market research—Chapter 7 regression analysis. Springer Texts in Business and Economics, Berlin. https://doi.org/10.1007/978-3-642-53965-7
27. Serrador P, Pinto JK (2015) Does agile work?—a quantitative analysis of agile project success. Int J Project Manage APM and IPMA 33(5):1040–1051. https://doi.org/10.1016/j.ijproman.2015.01.006
28. Shiner K, Pitt A (2015) Implementing-agile-in-financial-services-intelliware-development.pdf. Toronto, Ontario, Canada
29. Sidky A, Arthur JD (2007) A structured approach to adopting agile practices: the agile adoption framework a structured approach to adopting agile practices: the agile adoption framework. Innov Syst Softw Eng 3(3):1–12
30. Sinha S (2017) Exploratory study on the influence of agile on project management of outsourced software projects in India. Dublin
31. Taber KS (2018) The use of Cronbach's alpha when developing and reporting research instruments in science education. Res Sci Educ Res Sci Educ 48(6):1273–1296. https://doi.org/10.1007/s11165-016-9602-2
32. Totten J (2017) Critical success factors for agile project management in non-software related product development teams, Dissertation. Western Michigan University. Available at: https://scholarworks.wmich.edu/dissertations/3178
33. Tripp JF (2012) The impacts of agile development methodology use on project success: a contingency view, Michigan
34. Turk D, France R, Rumpe B (2002) Limitations of agile software processes. In: Extreme programming and flexible processes in software engineering, Alghero, Italy, pp 43–46. www.se-rwth.de/publications
35. Vanker C (2015) The adoption of agile software development methodologies by organisations in South Africa. doi: https://doi.org/10.13140/RG.2.1.1831.7683

36. Velumani M (2017) Adoption of agile enterprise architecture in large organisation: a case study
37. VijayasarathyLR, Turk D (2008) Agile software development: a survey of early adopters. J InfTechnol Manage XIX(2):1–8. https://www.aom-iaom.org/jitm_pdfs/jitm_08/article3.pdf
38. Zafar I, Nazir AK, Abbas M (2017) The impact of agile methodology (DSDM) on software project management. In: Circulation in computer science: international conference on engineering, computing & information technology (ICECIT 2017). Islamabad, Pakistan, pp. 1–6. www.ccsarchive.org

Automated Industrial Sound Power Alert System

R. Vishva, P. Harish Annamalai, K. Raja Raman, B. Vijay, J. Rolant Gini, and M. E. Harikumar

Abstract Noise pollution is a threat to health and well-being. Normal environmental noise is around 40–60 dB. But if the noise level increases above 80 dB, it can affect our psychomotor performance. To address this issue, the noise decibel level should be calculated and preventive measures should be taken. This work proposes an Automated sound power alert system at an industrial level that displays the decibel value of noise around heavy machinery and if it exceeds the threshold value of 80 dB, it notifies the authority by delivering an automated text message. It follows message queuing telemetry transport protocol which forms a basis for communication between microcontroller and Adafruit web. Data is transmitted to the web feed using the Message Queuing Telemetry Transport (MQTT) Publish pattern. The web feed is continuously monitored by the IF This Then That (IFTTT) for triggering the alert message. It is also equipped with liquid crystal display, Light-emitting diodes, and Piezo Buzzer that notifies workers around the machinery about the noise level as well as provides continuous monitoring of the same.

Keywords Message queuing telemetry transport (MQTT) protocol · IF this then that (IFTTT) · Liquid crystal display (LCD) · Light-emitting diode (LED)

R. Vishva · P. Harish Annamalai (✉) · K. Raja Raman · B. Vijay · J. Rolant Gini · M. E. Harikumar
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: harishanna336@gmail.com

R. Vishva
e-mail: vishvar0610@gmail.com

K. Raja Raman
e-mail: 2000rajaraman@gmail.com

B. Vijay
e-mail: vijaybaskar391999@gmail.com

J. Rolant Gini
e-mail: j_rolantgini@cb.amrita.edu

M. E. Harikumar
e-mail: me_harikumar@cb.amrita.edu

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_13

1 Introduction

The word noise is derived from the Latin word “Nausea” implying “unwanted sound” or sound that is loud, unpleasant, or unexpected. It can also be defined as a wrong sound, in the wrong place, and at the wrong time; is more severe and widespread than ever before which is known as noise pollution or sound power. It will continue to increase because of population growth and urbanization [1]. The population in the world continues to increase every day and India is said to become the most populated during 2019 and 2050. This will lead to an increase in noise pollution. Current guidelines used are insufficient to control noise pollution. It has been an underestimated threat that can cause short-term and long-term problems [2]. Continuous exposure to noise leads to hearing impairment and health problems like arterial constriction which leads to high blood pressure, sleep disturbance, etc. Emotional stress triggered by noise was suggested to play a role in the respiratory problems in children [3–5]. In India, the prevalence of hearing loss was estimated to be 63 million (6.3%). Preventive measures of noise pollution include the use of protective equipment like ear muffs and earplugs. Industries must also design machines that produce less noise and use noise abatement equipment such as acoustic louvers, Industrial exhaust silencers [6]. This kind of noiseless industry machinery design requires more cost and it may not be possible to implement in all the industries. The government of India has also implemented Noise Pollution (Regulation and Control) rules which are not regularly implemented and inspected [7]. Many people are not aware of the consequences of noise pollution. A study conducted in Orissa, a state in India found that, though people experienced noise-induced symptoms such as headaches, bad temper, hearing problems, loss of concentration, and sleep disturbance, they were unaware of the ill-effects of noise on health. If every person present in such an environment is aware of the consequences involved and alerted when the environment becomes dangerous then this would lead to the prevention of health effects caused by an increase in noise levels [8]. The current decibel meters are accurate but the user should stay near the heavy machinery for quite an amount of time to take readings. This could lead to hearing impairment and the health problems of the testing person [9].

The proposed Automated Sound Power alert system in this paper uses a condenser microphone to obtain noise input from industrial types of machinery. It is then processed by the microcontroller and the decibel value is displayed on the Liquid Crystal Display (LCD) screen. Moreover, the decibel values that are calculated by the microcontroller are periodically sent to an Adafruit server which acts as an intermediate medium for communication between the microcontroller and IFTTT. The IFTTT reads data from the Adafruit server and triggers the alert message if the value is greater than the threshold value. By employing the proposed system in industries, workers can be made aware of the noise decibel level around their workplace, and also hearing impairment and health problems can be avoided.

The full paper is organized in the following manner. Section 2 describes the design and working of the proposed system. Section 3 represents the results and discussions. Section 4 concludes the research work.

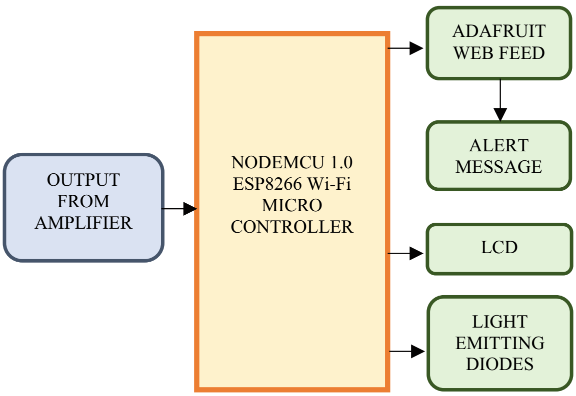
2 Design and Working

In this work, the input signal is obtained from the Condenser microphone, an analog sensor that converts sound pressure into an electrical signal. It is then processed by a preamplifier which converts the weak electrical signal into an output signal strong enough for further processing. The output analog signal is converted to a digital signal with discrete values ranging from 0 to 1024 by the Analog to Digital Converter (ADC) in the microcontroller ESP8266. The resulting digital values are then processed to estimate the noise level by the microcontroller ESP8266 to provide the current situation. LCD and Light-Emitting Diode (LED) are used to display the current decibel level. If the resultant noise level is over the threshold value then the piezo buzzer will alert the workers around. Being a Wi-Fi microchip, the ESP8266 also sends the resultant data to an Adafruit feed. IFTTT, a freeware web-based service is used to read data from the Adafruit server, which helps in triggering the alert message.

The system design is illustrated in Fig. 1, in the form of a block diagram. The Automated sound power alert system has two different ways to alert the workers. It displays the decibel value directly on the LCD screen and the other is the number of LED lights that glow in correspondence to the decibel value. Each LED light that glows denotes a 10 dB noise level in the system. The more the number of light glows, the higher the decibel value. The latter way is very useful to find the decibel value without the user stay near the heavy machinery to find out the value.

The condenser microphone is to be placed between 5 and 10 cm distance from the noise source (Industrial machinery) to be measured for getting good input. The

Fig. 1 Block diagram of the system

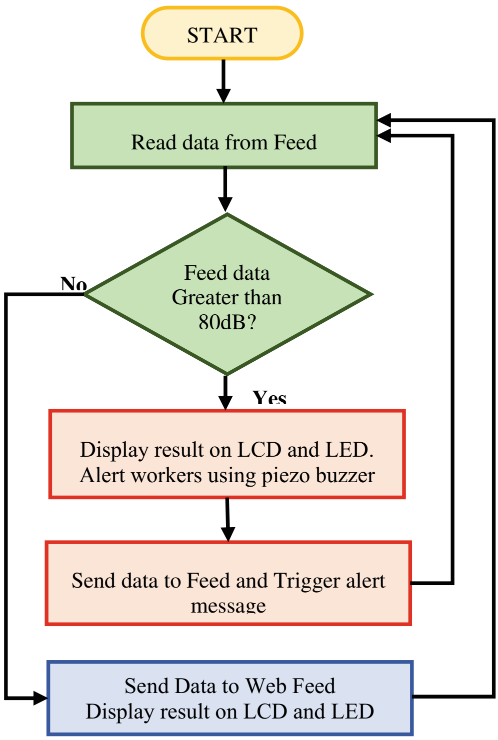


preamplifier amplifies the signal, which is then processed by the microcontroller and the result is displayed. The microcontrollers can be powered using micro USB 2.0 cables. Using the MQTT library, the data has been transmitted periodically from the microcontroller to the web feed. Adafruit is a free web-based service where only 30 data points can be sent to the server for a minute. Hence, the Automated sound power alert system sends the value to the feed every two seconds. The IFTTT reads data from the updated feed from the Adafruit web server and triggers the alert message when the value is over the threshold value. The basic control flow of the Automated sound power alert system is illustrated in Fig. 2.

The 16X2 LCD used in this experiment requires 5 V DC voltage to display output. Since the General Purpose Input–Output pins (GPIO) in ESP8266 microcontroller provide only 3.3 V. The V_{in} pin of the ESP8266 microcontroller gives 5 V DC. So, an IC7408, commonly known as AND gate is used to get 5 V from V_{in} pin to power up the LCD. From Fig. 3, the circuit connection for LCD through IC7408 is understood.

Inputs obtained from the condenser microphone may not be strong enough for processing. It needs to be amplified to convert it into a digital signal. The LM358, a preamplifier shown in Fig. 4b, is used to amplify the weak input signal. By using Preamplifier, it is possible to obtain a higher gain value and improved sound quality. The LEDs are connected to GPIO pins of the ESP8266 Wi-Fi Microcontroller. From

Fig. 2 Control flow diagram of the proposed system



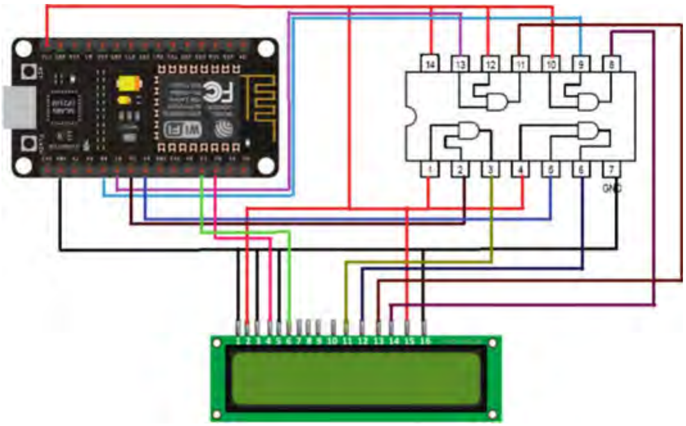


Fig. 3 LCD circuit connection of the proposed system

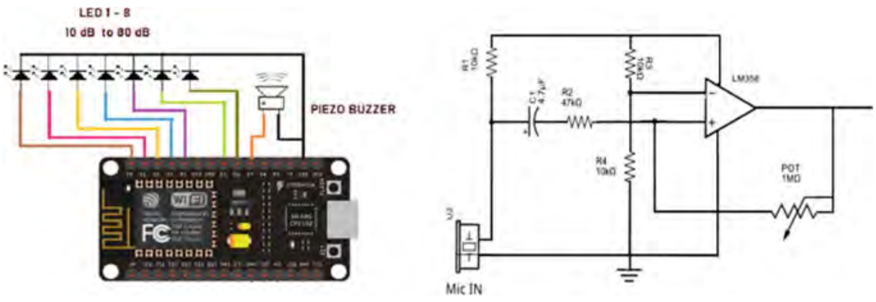


Fig. 4 **a** LED circuit connection of the proposed system. **b** Preamplifier circuit connection

Fig. 4a, the circuit connection for LEDs and From Fig. 4b, preamplifier circuit connection is understood. The components used and their purpose in this Automated sound power alert system are given in Table 1.

Table 1 Components used in the automated sound power alert system

Components	Purpose
Condenser microphone	The most common instrumentation microphone—used to get input signals
LM358 (preamplifier)	To amplify the weak signal to a strong output signal for processing [10]
ESP8266 Wi-Fi microchip	The processing unit for the sound power alert system [11]
IC7408 and LCD	Display the decibel value for the user
LEDs and Piezo Buzzer	To indicate the decibel range for the user

2.1 Software Used

Arduino IDE is an open-source Arduino software [12] is used for control code and uploaded to the ESP8266 Microcontrollers. Adafruit IO, which supports MQTT protocol [13], acts as a feed for the data transmitted by the ESP8266 microcontroller. IFTTT, a free web-based service [14] is used to monitor the changes at Adafruit web feed and trigger an alert message if the feed value is over the threshold value.

3 Results and Discussion

The Automated Sound power alert system is tested for various noise inputs. For experimental trails, inputs are given through a Bluetooth speaker. Input for the speaker consists of various songs that are played at different volumes and also inputs are taken from a very noisy traffic environment. The outputs are displayed on LCD and LEDs. The interpretation of decibel values from the LEDs is understood from Table 2.

The working model of the sound power alert system can be understood from Fig. 5. As soon as the condenser microphone receives noise from various sources, it converts it into an electrical signal and the preamplifier amplifies it. The processed result is displayed in the LCD and LEDs. The number of LEDs that glow, in this case, 5 corresponds to the range 50–59 dB which is also shown by the LCD.

The data which is displayed in the LCD, in this case, 57 dB is sent to the Adafruit web feed. The same has been shown in Fig. 6a. IFTTT provides many web-based services. These services are known as applets. For the proposed automated sound power alert system, the Adafruit applet with the android SMS applet is combined to trigger alert messages. Since ESP8266 doesn't have an inbuilt messaging service this is setup is required. The IFTTT setup interface has also been shown in Fig. 6b. For Adafruit feed values that are over the threshold value, the IFTTT triggers an alert message.

The input to the Bluetooth speaker is given at maximum volume. From Fig. 7a, it is understood that the calculated decibel value is over the threshold value of 80 dB

Table 2 Decibel range correspondence

Decibel range	Number of LED glows
10–19 dB	1
20–29 dB	2
30–39 dB	3
40–49 dB	4
50–59 dB	5
60–69 dB	6
70–79 dB	7
Greater than 80 dB	7 + 1 (Piezo Buzzer)

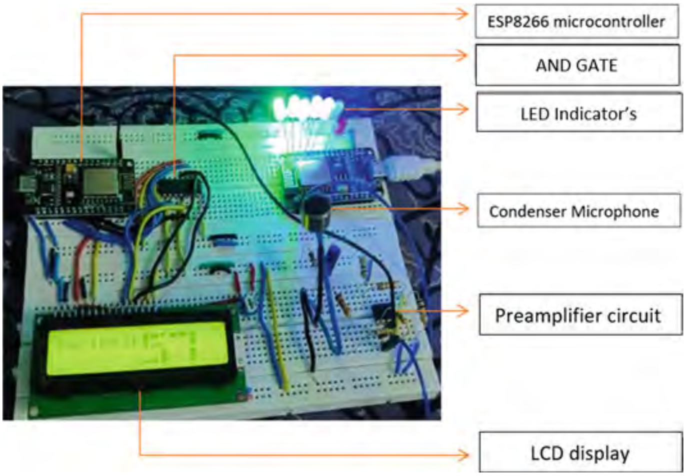


Fig. 5 Working model of the automated sound power alert system

Fig. 6 **a** Adafruit Web Feed displaying transmitted data from ESP8266 Wi-Fi Microcontroller. **b** IFTTT setup interface for Adafruit and Android SMS

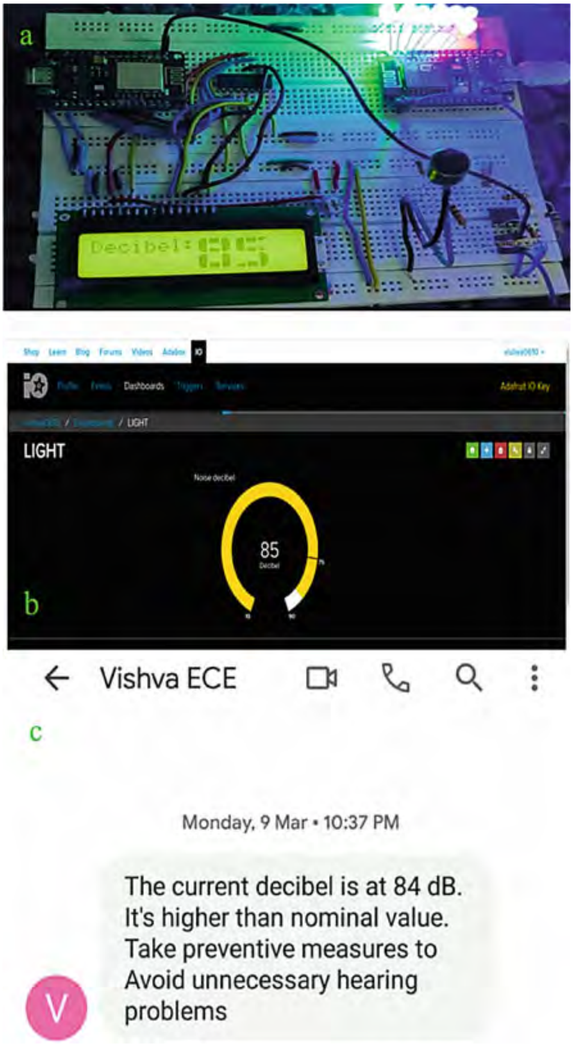


Fig. 7 **a** Sound power alert system result for the noisy case. **b** Adafruit Web Feed displaying transmitted data. **c** The triggered alert message by the IFTTT

and the same has been displayed in the LCD. The MQTT publish protocol transmits the resulting value which is displayed on the LCD to the Adafruit web feed. The transmitted value is shown on the Adafruit feed dashboard as shown in Fig. 7b. The IFTTT, which connects the Adafruit and the android SMS applets monitor the Adafruit feed, finds the value from feed to be greater than the threshold value of 80 dB. so, it triggers an alert message as programmed. The alerted text message is as displayed in Fig. 7c.

The inputs given to the condenser microphone should be monotonous. Calibration of the condenser microphone is a very important process. Since Adafruit feed can support only 30 data points per minute, data is transmitted to the Feed with a significant delay of 2 s. Care must be taken so that the two ESP8266 controllers aren't out of sync. This is the drawback of the proposed Automated sound power alert system. The use of a Wi-Fi microcontroller with a greater number of General Purpose Input–Output (GPIO) pins to support both LCD and LEDs at the same time could solve the above case. This can be upgraded and made to work collectively with other finished products. For example, the Police of Mumbai city, India has equipped traffic signals with noise decibel meters, which increases the waiting time for the signal to turn green if the decibel value crosses 85 dB [15, 16]. This reduces people from honking unnecessarily while waiting for the traffic signal. This can also be extended for domestic use with minor changes.

4 Conclusion

The proposed Automated sound power alert system is lightweight, low cost, and user friendly for workers to use. It is easy to use and it does not require a noise assessor to be present all the time. Health risks like stress and hearing loss can be minimized to an extent with the use of these systems. It helps industries in managing Occupational Health and Safety claims regarding the noise of machineries. The proposed Automated sound power alert system would create a healthy work environment that improves productivity at an affordable cost. This would increase the Gross Domestic Product and economy of a country.

References

1. Singh N, Davar SC (2004) Noise pollution-sources, effects, and control. J Hum Ecol 16:181–187
2. Smys S, Raj JS (2019) Virtual reality simulation as therapy for posttraumatic stress disorder (PTSD). J Electronic 1(1):24–34
3. Hahad O, Kroller-Schon S, Daiber A, Munzel T (2019) The cardiovascular effects of noise. Deutsches Arzteblatt International 116:245–250
4. Clark WW, Bohne B (1984) The effects of noise on hearing and the ear. Medical Times 112:17–22
5. Petric D (2020) Detrimental health effects of noise pollution
6. Mohanapriya SP, Sumesh EP, Karthika R (2014) Environmental sound recognition using Gaussian mixture model and neural network classifier. In: International conference on green computing communication and electrical engineering (ICGCCEE), Coimbatore, pp 1–5
7. High Court warns of action for non-implementation of noise pollution rules. https://economictimes.indiatimes.com/news/politics-and-nation/high-court-warns-of-action-for-non-implementation-of-noise-pollution-rules/articleshow/47800925.cms
8. Balazikova M, Salaj L, Wysoczanska B (2019) Analysis of human factor reliability in workplace with noise load. In: 2019 international council on technologies of environmental protection (ICTEP), Stary Smokovec, Slovakia, pp 25–29
9. Davis AH (1938) An objective noise-meter for the measurement of moderate and loud, steady and impulsive noises. J Inst Electr Eng 83(500):249–260

10. Neri B, Pellegrini B, Saletti R (1991) Ultra-low-noise preamplifier for low-frequency noise measurements in electron devices. In: IEEE transactions on instrumentation and measurement 40:2–6
11. Raj JS, Vijitha Ananthi J (2019) Automation using IoT in greenhouse environment. J Inform Technol 1(1):38–47
12. Ramon MC (2014) Arduino IDE and wiring language. In: Intel® Galileo and Intel® Galileo Gen 2. Apress, Berkeley, CA
13. Jayan AP, Balasubramani A, Kaikottil A, Harini N (2019) An enhanced scheme for authentication using OTP and QR code for MQTT protocol. Int J Recent Technol Eng 7:70–75
14. Ovadia S (2014) Automate the internet with "If This Then That" (IFTTT). Behav Soc Sci Libr 4:208–211
15. Mumbai Police Play a Trick on Honking Drivers. https://www.nytimes.com/2020/02/04/world/asia/mumbai-horn-honking.html
16. The more you honk, the longer you wait! Signal to stay red if decibel level high. https://timesofindia.indiatimes.com/city/mumbai/more-you-honk-the-longer-you-wait-signal-to-stay-red-if-decibel-level-high/articleshow/73819813.cms

Predicting Early Readmission of Diabetic Patients: Toward Interpretable Models

Mir Moynuddin Ahmed Shibly, Tahmina Akter Tisha, and Md. Mahfuzul Islam Mazumder

Abstract Hospital readmission among diabetic patients is a common phenomenon throughout the world. Predicting such patients with a high risk of readmission at the time of discharge even before can help us to provide better health care to them. It can minimize the cost associated with readmission too. This study aims at creating a decision support system that can find diabetic patients who are prone to early readmission. To do that, several data mining techniques have been used. Two regular classifiers using the decision tree and random forest have been developed. After that, two rule-based classifiers using Repeated Incremental Pruning to Produce Error Reduction (RIPPER) and PART algorithms have been developed to provide better interpretability and understandability of the support system. Between two regular classifiers, the random forest has shown a better performance with 89.5% accuracy and 89.5% recall. And, between rule-based classifiers, PART has demonstrated promising performance with 86.6% accuracy and 84.6% recall. Using these classifiers, smart and improved health care can be ensured.

Keywords Rule-based classifiers · RIPPER · PART · Decision support system · Dataset imbalance · Synthetic Minority Oversampling Technique (SMOTE)

1 Introduction

Hospital readmission is an important issue in the health care system. It can be a matter of inconvenience for patients, doctors, and other stakeholders. It can also unbalance the overall cost management of the health care system. Around 17 billion dollars is spent yearly on hospital readmission in the USA [1]. Follow-up treatment

M. M. A. Shibly (✉) · T. A. Tisha · Md. M. I. Mazumder
Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh
e-mail: shiblygnr@gmail.com

T. A. Tisha
e-mail: tahminatish001@gmail.com

Md. M. I. Mazumder
e-mail: 2016-3-60-048@std.ewubd.edu

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_14

of the critical patients and taking appropriate measures can reduce the rate of readmission. The readmission can happen because of various reasons like heart failure, pneumonia, acute myocardial infarction, etc., [2]. But among the patients who have stayed in a hospital for various reasons, diabetic patients are more vulnerable to hospital readmission [3]. The reason for diabetic patients to stay in the hospital does not necessarily have to be a diabetes-related illness. Readmission often occurs for those patients who have diabetes as comorbidity but admitted for other reasons. A study shows that the readmission of diabetic patients can be managed by better follow-up treatment [4]. To ensure better treatment after discharge, the first step is to identify those who are at risk of getting readmitted early. A statistical decision support system based on data mining techniques that can detect those risky patients early can be helpful to mitigate the risks of mortality. Early predicting of them using this system can also be financially helpful. In the current world where ensuring proper health care for people is the most challenging task; such support systems can make life easier for all stakeholders. If a patient who is likely to be readmitted early after discharge, then he/she can be monitored with high importance to prevent readmission. Customized health care plan can be designed for the targeted patients. Such precautions can result in to provide better health care services in terms of reducing risk of getting readmitted and improving cost management associated with it.

For being an important and challenging issue in the health care system, many methods have been used to accomplish the task of early prediction of readmission of diabetic patients. Different researchers have applied different techniques to provide classification-based predictive models. Some have used machine learning algorithms [5] like support vector machine [6], random forest [7], neural networks [8], etc. The typical machine learning models can act as a black box, i.e., how the model decides something based on input is unknown to the users. Figure 1a illustrates an example of such a machine learning model. Usually, the neural network-based models are treated as the black box system. If the users cannot perceive how a model makes a decision, then there might be a chance of growing untrust among them [9]. In health care-related problems, the predictions must have to be interpretable, understandable, and transparent. Few methods interpret the results of a machine learning model like Local Interpretable Model-agnostic Explanations (LIME) [10].

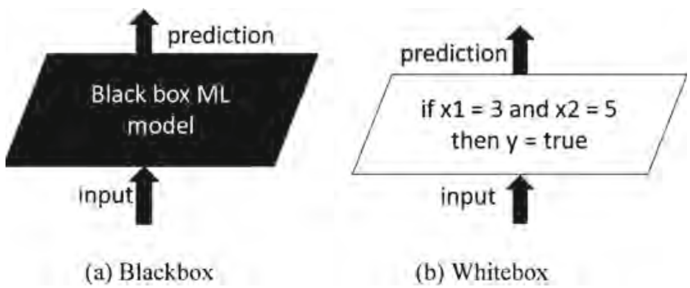


Fig. 1 Machine learning models

In this domain of the health care system, rule-based classifiers can be utilized to design an improved predictive support system. As their results are easier to understand than the other "black box" systems, rule-based classifiers can also provide better interpretability of the system to the stakeholders. Figure 1b illustrates a rule-based predictive system. There is a gap of knowledge of how the rule-based classifiers can perform in early readmission prediction of diabetic patients. There have not been many works to predict early readmission using rule-based interpretable classifiers.

This study designs interpretable rule-based classifiers to predict the early readmission possibility of diabetic patients. The objectives of this study are:

1. To predict early readmission of diabetic patients using traditional classifiers.
2. To develop interpretable rule-based predictive models using Repeated Incremental Pruning to Produce Error Reduction (RIPPER) and PART algorithms.
3. To compare the results of traditional and rule-based classifiers.

This article is arranged in the following manner, Sect. 2 presents the related works to this study, and Sect. 3 describes the methodologies and datasets. Section 4 presents the results of this study with appropriate comparative analysis, and Sect. 5 discusses the results. And, finally, Section 6 draws some concluding remarks.

2 Related Works

Cui et al. [6] have conducted experiments on data of diabetic patient hospital readmission. The aim has been to reduce the readmission rate so that better health care can be provided by mitigating associated costs incur for hospital readmission. With the samples from the hospital readmission dataset [11] based on a criterion, a dataset has been constructed by the authors. They have used synthetic minority oversampling technique to reduce the imbalance between the class that has early hospital readmission risk and the class that has less chance of getting readmitted. They have used a hybrid feature selection method combining the filter method and wrapper method. For the binary classification task, they have proposed a Support Vector Machine-based (SVM) method. They have used a genetic algorithm to optimize the parameters of SVM. They have also employed k-fold cross-validation in the training stage. The study has achieved 81.02% accuracy in the testing phase.

In another study [7], an intelligent decision support system has been developed using the random forest algorithm and Bayesian network. They have experimented with a series of classification algorithms, and the best performing classifiers have been selected. Before training, they have created a dataset by taking samples from the hospital readmission dataset [11] based on diabetes medication. They have achieved up to 82.97% accuracy with their works. Additionally, they also have identified the optimal medications for 28 comorbidity combinations to reduce the chance of readmission after discharge by increasing monitoring for risky patients.

Another study has been conducted by Bhuvan et al. [12]. They have used naive Bayes, Bayes network, random forest, and neural networks for the same classification problem. They have obtained a maximum of 65.4% accuracy. Additionally, they have calculated the costs that can be saved by employing their models. Salian et al. [13] have used traditional classifiers decision tree, logistic regression, SVM, k-nearest neighbors to accomplish a similar task. They have achieved a minimum misclassification rate of 28% using a decision tree. Additionally, they also have generated some rules from the decision tree. Similar studies have been conducted by Alajmani et al. [14], Alloghani et al. [5], Shameer et al. [15], etc. Different studies have utilized different techniques but to achieve a common goal. All of these studies are aimed to find predictive systems that can identify the high-risk patients that are more likely to get readmitted to the hospital 30 days after discharge. Early identification of such patients can help the stakeholders by following up with better treatment and by minimizing cost. After analyzing these related works, it has been seen that there have not been enough significant works using rule-based classifiers to predict early readmission.

3 Methodology

The objective of this study is to predict the risk of a diabetic patient being readmitted to the hospital 30 days after discharge. This work designs a decision support system to predict whether a diabetic patient is needed to be under constant monitoring or not after discharge. In this section of the article, the methods and materials that have been used to create such a support system are described. The complete flow of the works has been demonstrated in Fig. 2.

3.1 Dataset

In this study, "Diabetes 130-US hospitals for years 1999–2008 dataset" [11] containing more than 100,000 instances from the UCI machine learning repository has been used. This dataset has been constructed from the health facts database (Cerner Corporation, Kansas City, MO) based on five criteria. There are 50 features like age, race, weight, the medical speciality of a patient, different diagnoses and medications, etc., in this dataset. Each instance of the dataset is labeled with a class having one of the three outcomes—not readmitted, readmitted in less than 30 days, and readmitted after 30 days. Eleven of the features of the dataset are numeric, and the rest are nominal.

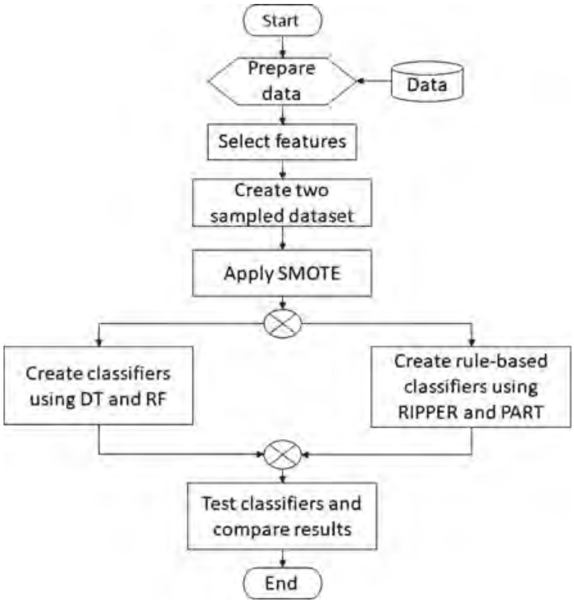


Fig. 2 Flow diagram of the study

3.2 Data Preparation and Feature Selection

There are few nominal features in the dataset that contain null values. The features that have less than or equal to 2% missing values are replaced with the mode of the corresponding feature. On the other hand, the feature "medical specialty" has 53% missing values. They are replaced with a new feature value named "unknown." Another feature named "weight" which has 97% missing entries is removed from the dataset. Three other attributes—"encounter ID," "patient no", and "payer code" are also removed. Three categorical variables "diag_1," 'diag_2," and "diag_3" contain international statistical classification of diseases and related health problems, i.e., ICD9 codes of primary, secondary, and tertiary diagnosis of a patient. These features are mapped to nine major diagnoses as suggested in the original paper of this dataset [11]. In this study, the focus group is the patients with readmission status "<30." The ">30" status has been merged with the "NO" group to make it a binary classification problem. In this way, a complete categorical dataset with 45 features with the target feature having binary classes is prepared. The dataset has been needed to be binarized further in some experiments for the compatibility issue of used tools. Additionally, for feature selection, the Gini index—a filter model has been used. Gini index calculates the discriminative power of each feature [16]. Using this, 24 features with lower scores are selected for classification.

3.3 Data Sampling

The initial experiments of this study have been conducted on the complete dataset. But our scope is only diabetic patients. Two sampled datasets have been prepared based on two criteria. With the data points that have a primary diagnosis as diabetes, i.e., diag_1 = Diabetes and diabetes_med = Yes, two sampled datasets are created as suggested in these papers [6, 7]. In the original dataset and two sampled datasets, there is an imbalance between two target classes. This imbalance in datasets may result in poor performance of the classifiers. And, the classifiers can perform poorly to detect the patients who have a higher chance to get readmitted despite performing well overall. To tackle this imbalance, the Synthetic Minority Oversampling Technique (SMOTE) [17] has been used. In this method, each instance of the minority classes' k-nearest neighbors from the same class is calculated. After that, a portion of the neighbors is selected randomly based on the number of oversampling instances needed. Then, for each instance neighbor pair, a synthetic data point is generated and added to the training set.

3.4 Classifiers

After preparing data and feature selection, two classifiers have been developed using decision tree and random forest algorithms. The decision tree algorithm continuously divides the work area by plotting lines into sub-areas until a specific class emerges. In the decision tree algorithm, the process of classification is created with a set of hierarchical decisions on the attributes structured with a tree [16]. The decision at a node is a condition on one or more attributes in the training set. The condition divides the training data into two or more parts. The aim is to separate the training data into a smaller portion of the target classes in the best possible way. In this manner, the tree grows until it meets a stopping criterion. A stopping criterion can be where all the training examples in a leaf node belong to a single class. This algorithm finds the best split using some quantifications. In this study, the decision tree with a Gini index has been used for classification tasks.

Another classification algorithm that has been used is the random forest algorithm. It is an ensemble learning method which is widely used in regression and classification problems. A random forest is nothing but a combination of many decision trees. In this method, k decision trees are created using training data to form a forest, and for the test data, the majority voting technique is followed. The majority class predicted by individual decision trees is the predicted class. The term forest came from the idea that to create each decision tree, a random set of attributes is selected. The trees are not developed with all features. Another way to create a random forest to use a bootstrap aggregating technique [18]. A portion of the training data points is selected with replacement, and for each decision tree, a newly sampled training dataset is used. Random forest resolves the problem of overfitting in decision tree classifier [19]. In this study, random forest with random attribute selection has been used.

3.5 Rule-Based Classifiers

As the objective of this study is stated, in this stage, three rule-based classifiers have been developed. The rule-based classifiers have better interpretability than the typical classifiers. How a regular trained model predicts a class in a real-life environment can be less understandable by the common stakeholders who use the system. But if there is some rule upfront supporting a classifier, then the working mechanism behind a decision can be perceived by the patients, doctors, and the other stakeholders. Rules can be generated from a decision tree. But the number of rules generated by a decision tree is enormous. That is why algorithms specially designed for rule learning are used. Three algorithms have been used to generate rules, namely Apriori algorithm, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), and PART algorithm.

3.5.1 Apriori

Apriori is an association pattern mining algorithm that generates frequent itemsets present in a dataset. This algorithm generates a smaller candidate itemsets first. Then, the support of each itemset is calculated. By joining the itemsets that satisfy minimum support, the next smaller candidate itemsets are generated rather than generating all itemsets. From these candidate itemsets, the downward closure property violating itemsets are removed. After that, the supports of the remaining candidates are calculated, and the satisfying itemsets are added to the set of frequent itemsets. This process continues until all the frequent itemsets based on minimum support are found. After mining frequent itemsets, a suitable rule generation framework can be employed which works based on conditional probability. Only the rules that have greater or equal to minimum confidence are added to the ruleset. In this study, a modified rule generation framework has been used. Rules are being looked at that can predict the classes. First, the frequent itemsets that have either of the two classes are filtered. After that, only the rules of types $\{x_1x_2 \ldots x_n\} \rightarrow \{readmitted = '{<}30'\}$ and $\{x_1x_2 \ldots x_n\} \rightarrow \{readmitted = 'NO'\}$ are considered for association rules, where $x_1x_2 \ldots x_n$ are individual items in a frequent itemset. If these rules satisfy the minimum confidence, then they are the desired rules mined.

3.5.2 Repeated Incremental Pruning to Produce Error Reduction

RIPPER is a sequential covering algorithm. It adds three improvements over the Incremental Reduced Error Pruning (IREP) algorithm. IREP starts pruning the created rule right after it was developed. In basic IREP, the training dataset is divided into two sets naming “growing set” and “pruning set.” The growing set has two-third of the training data, and the pruning set has the other one-third. For the growing set, using a basic sequential covering algorithm, the best rule is generated. After that, the worth of the rule is calculated on the pruning set based on a measurement metric. Then, a clause from the end of the rule is omitted, and the worth of the reduced rule is computed. If the worth does not decline, then the pruning process continues. After a rule is pruned, the training instances covered by the rule are removed, and the next rule generation starts. In this way, the classification rules are generated [20]. RIPPER offers few modifications in the IREP procedure. The first improvement is a better metric to calculate the worth of a rule [21]. If the total number of positive and negative class in the pruning set is P and N, and p and n are the number of positively and negatively identified instances, respectively, then the worth of a rule is defined by $(p + (N - n))/(P + N)$. The modification made by RIPPER is to use $(p - n)/(p + n)$ as the metric.

Another limitation of IREP is that it stops adding new rules when the last rule has more than 50% error on the pruning set. The developer of RIPPER has said that this stopping criterion is too early and has suggested a new stopping criterion based on description length. Description length of a rule is the number of bits needed to encode a rule, and the number of examples covered by the class-specific rule in the training dataset, which belongs to a different class, i.e., error made by the rule. After generating a rule, its description length is computed, and if that is b bits larger than the smallest description length obtained so far, then it is not added to the ruleset. The last improvement is rule optimization using minimum description length. After creating all the rules, each of them is analyzed, and two variants of it are created—one is the extended version, and the other is generated from scratch. Among the three, a rule is selected according to the minimum description length. In this study, the RIPPER algorithm has been used to create a rule-based classifier.

3.5.3 PART

The last algorithm to create a rule-based classifier is PART. It is developed by Frank et al. [22]. As mentioned earlier, RIPPER globally optimizes the rulesets after rule generation using complex methods. In contrast, PART does not need this global optimization step to generate accurate rules. It uses a typical separate-and-conquer method to build a rule. It follows a sequential covering algorithm. But it differs in how a rule is created. To create each rule, it creates a partial decision tree on current instances. After building the partial tree, one of the leaves that have the highest coverage of training instances is turned into a rule. In this method, the global optimization step is not needed. Even without this optimization step, the algorithm can perform well. This method has been used to build the last classifier to predict early hospital readmission among diabetic patients.

4 Results

This study aims to develop a system that predicts early hospital readmission of diabetic patients. After data preparation, decision tree and random forest classifiers have been developed on the original dataset including 44 categorical attributes with a target feature having binary classes. Then, 24 features have been filtered using a Gini index. Then, two sampled datasets are created based on primary diagnosis and diabetes medication. The same experiments have been carried out on these two datasets too. In the final phase of the study, three rule-based predictive models have been developed using Apriori, RIPPER, and PART algorithms. To implement the decision tree and random forest classifiers, the scikit-learn library from Python has been used. And, for the rule-based models, Weka—a data mining tool developed at the University of Waikato, New Zealand has been used. For all the classifiers, the datasets have been split into training and testing set having 80 and 20% data, respectively. In this section of the article, the outcome of the study based on some evaluation metrics, and the prominent rules generated by predictive models have been presented. All the performances that are reported in this section are test performances.

4.1 Evaluation Metrics

Four evaluation metrics have been used throughout the whole working process of this study—precision, recall, f1-score, and accuracy. Precision is the portion of the correct predictions made by the classifier with respect to total predicted classes. And, recall is the portion of the correct predictions made by the classifier regarding the total existing accurate classes. F1-score is the harmonic mean of precision and recall, i.e., both precision and recall are given equal importance while evaluating the performance of a classifier. And, accuracy is nothing but the portion of the classifier being right.

$$\text{precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{recall} = \frac{TP}{P} \tag{2}$$

$$f1\_\text{score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{3}$$

$$\text{accuracy} = \frac{TP + TN}{P + N} \tag{4}$$

Here, P, N TP, TN, FP, and FN are the number of actual positive instances, negative instances, true positive, true negative, false positive, and false negative, respectively.

4.2 Results Analysis

The initial experimentations on the full dataset show that the overall performance of decision tree (DT) and random forest (RF) is good. The accuracies of classifiers are 88.46 and 89.24%, respectively. But we are more interested to detect the patients that have a higher chance of readmission than the patients that have no chance of readmission. Both classifiers have performed poorly to find out the patients that have been readmitted within 30 days after discharge. But after adding synthetic data using SMOTE, the class-wise performance has been improved. Table 1 shows the comparison of the classifiers' class-wise performance. Another aspect is to consider the impact of feature selection. It has been observed that the classifiers with 45 features and with 24 features have similar performance. This justifies the employing feature selection strategy before training. A comparison with and without feature selection has been demonstrated in Table 2.

The scope of this study is focused on diabetic patients. To provide a better decision support system for this domain-specific patient, the original dataset has been sampled into two sub-samples. In the original paper of the dataset [11], it has been said that the probability of hospital readmission depends on the primary diagnosis of the patients

Table 1 Class-wise performance of the classifiers on the full dataset with and without SMOTE

Sampling technique	Algorithm	Classes	Precision	Recall	Support
Without sampling	DT	<30	0.30	**0.05**	2189
		NO	0.90	0.98	18,165
	RF	<30	0.48	**0.01**	2189
		NO	0.89	1.00	18,165
SMOTE	DT	<30	0.86	**0.90**	17,930
		NO	0.90	0.85	18,234
	RF	<30	0.98	**0.91**	17,930
		NO	0.92	0.99	18,234

Table 2 Impact of feature selection with Gini index

Algorithm	Without feature selection			With feature selection		
	Precision	Recall	Accuracy	Precision	Recall	Accuracy
DT	0.884	0.883	**0.883**	0.883	0.882	0.882
RF	0.952	0.950	0.950	0.952	0.950	**0.951**

Table 3 Performance of decision tree and random forest on two sampled datasets

Dataset	Algorithms	Precision	Recall	F1-score	Accuracy
Sampled based on the primary diagnosis	DT	0.792	0.793	0.792	0.792
	RF	**0.896**	**0.895**	**0.895**	**0.895**
Sampled based on diabetes medication	DT	0.795	0.797	0.796	0.797
	RF	**0.906**	**0.906**	**0.906**	**0.906**

(feature name: diag_1). The ICD9 codes ranging from 250 to 251 denote that a patient is diagnosed with diabetes. The data points that have diag_1 value in this range are considered for the experiment in this phase of the study. 8772 data points fall under this criterion. In the next phase, the data points that have diabetes medication equal to "Yes" has been considered as the samples. 78,363 data points fulfill the diabetes medication criterion. 24 features selected in the previous phase of the study have been used in these experiments. Since synthetic minority oversampling technique has been proven to work better to tackle both imbalance and overfitting problems, it has been used in the experiments too. In Table 3, the performances of the classifiers created on these two sampled datasets have been shown. For both datasets, random forest classifiers have outperformed the decision tree classifiers. The decision tree classifiers have an accuracy of 79.2 and 79.7% in two datasets, respectively, while the random forest classifiers have an accuracy of 89.5 and 90.6%, respectively. And, the class-wise recalls of the classifiers have also been observed. Since SMOTE has been applied on the dataset before training, unlike the initial experiments, the recalls of individual classes have been better.

The results show that classifiers to predict hospital readmission of diabetic patients can be developed using decision tree and random forest. But this type of classifier lacks interpretability. To provide a better decision support system for the doctors, patients, and other stakeholders, rule-based classifiers have been developed. The predictive rules have been generated using Apriori, RIPPER, and PART algorithms. While creating rules from frequent itemsets that are mined using Apriori, only the itemsets having either of two target classes are selected. For mining frequent itemsets, minimum support of 0.2 is used. And for rules have been generated with a minimum confidence of 0.7. In Table 5, five rules generated using the Apriori algorithm are demonstrated. In this study, two rule-based classifiers have also been developed to provide the interpretability of the models. The algorithms used for rule-based prediction are RIPPER and PART. These two algorithms have produced 90 and 635 rules, respectively. Table 4 shows a few of the top rules generated by two algorithms.

The rules generated by three predictive models are easy to understand. The patients and doctors will understand the reason behind a decision that has been made by just looking at them. The rules are also self-explanatory. If the first rule generated by the Apriori algorithm (Table 5) is looked at, then it will eventually be apparent that a patient who has not been under any procedure ($num\_procedures = (-2, 0]$), who has not any emergency ($number\_emergency = (-10, 0]$), and has not any inpatient visit ($number\_inpatient = (-5, 0]$) during the stay in hospital is not likely to be readmitted.

Table 4 Rules generated by RIPPER and PART

No	RIPPER	PART
1	(medical_specialty = unknown) and (number_inpatient = (0, 5]) and (num_lab_procedures = (40, 60]) and (race = Caucasian) and (A1Cresult = None) and (diag_3 = Circulatory) and (diag_2 = Circulatory) → readmitted = <30	number_inpatient = (10, 15] AND medical_specialty = unknown AND A1Cresult = None → readmitted = <30
2	(medical_specialty = unknown) and (num_lab_procedures = (40, 60]) and (race = Caucasian) and (A1Cresult = None) and (diag_3 = Other) and (admission_type_id = Emergency) → readmitted = <30	glipizide = Up → readmitted = NO
3	(number_inpatient = (0, 5]) and (medical_specialty = unknown) and (age = [40–50)) and (diag_2 = Neoplasms) → readmitted = <30	number_inpatient = (5, 10] AND admission_source_id = AS1 AND race = Caucasian AND medical_specialty = unknown → readmitted = <30
4	(number_inpatient = (0, 5]) and (medical_specialty = InternalMedicine) and (number_emergency = (0, 10]) and (age = [40–50)) → readmitted = <30	number_inpatient = (5, 10] AND admission_source_id = AS7 AND discharge_disposition_id = D1 AND race = Caucasian AND diag_3 = Other AND num_medications = (10, 20] → readmitted = <30
5	(metformin = Steady) and (num_lab_procedures = (20, 40]) and (admission_source_id = AS7) and (number_diagnoses = (5, 10]) and (number_emergency = (-10, 0]) → readmitted = NO	glipizide = Steady AND number_emergency = (-10, 0] AND admission_source_id = AS7 AND medical_specialty = unknown AND admission_type_id = Emergency AND number_diagnoses = (5, 10] AND discharge_disposition_id = D1 → readmitted = NO

Table 5 Class association rules generated by Apriori with minconf = 0.70

No.	Rules	Confidence
1	num_procedures = (−2, 0] number_emergency = (−10, 0] number_inpatient = (−5, 0] → readmitted = NO	0.72
2	discharge_disposition_id = D1 number_inpatient = (−5, 0] time_in_hospital = (0, 5] → readmitted = NO	0.71
3	diag_1 = Diabetes num_procedures = (−2, 0] number_outpatient = (−4, 5] repaglinide = No → readmitted = NO	0.71
4	glipizide = No medical_specialty = unknown metformin = No number_outpatient = (−4, 5] race = Caucasian repaglinide = No A1Cresult = None → readmitted = <30	0.70
5	admission_source_id = AS7 admission_type_id = Emergency glipizide = No medical_specialty = unknown metformin = No number_outpatient = (−4, 5] race = Caucasian → readmitted = <30	0.70

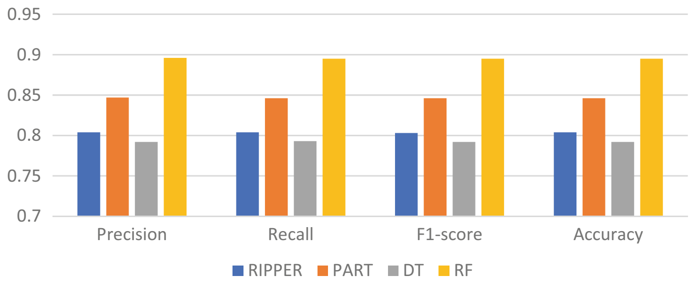


Fig. 3 Performance comparison of RIPPER, PART, decision tree, and random forest

Similarly, the first rule generated by the RIPPER algorithm suggests that if a patient has a large number of laboratory procedures and is diagnosed with a circulatory system during the stay in hospital is likely to be readmitted. Moreover, comparing the length of the generated rules by RIPPER and PART, it has been observed that PART has generated shorter and concise rules. However, RIPPER has generated fewer numbers of rules that PART. It is to be noted that for Apriori, RIPPER, and PART algorithms, only the sampled dataset based on primary diagnosis has been used.

In Fig. 3, there is a performance comparison among a regular classifier, and two rule-based classifiers have been demonstrated. The chart depicts that the rule-based classifiers have outperformed the decision tree-based classifier. Though they could not beat the random forest ensemble classifier, in terms of interpretability of the rules and overall performance, RIPPER and PART algorithms have shown good performances. Between two rule-based predictive models, the PART algorithm has outperformed the RIPPER algorithm having 84.6% accuracy, while the other one has 80.6% accuracy. The overall performance of this study is better than the other existing works too. In Table 6, a comparison with a few other existing works has been shown.

Table 6 Performance comparison with other works

Works	Test accuracy (%)
Cui et al. [6]	81.02
Ossai et al. [7]	82.97
Rubin et al. [23]	70
The proposed method (RF)	**89.5**
The proposed method (PART)	**84.6**

5 Discussion

Predicting early hospital readmission of patients can be helpful to reduce hassles to treat them properly. It can also minimize the costs associated with readmission. For diabetic patients, a decision support system that can filter out the patients susceptible to readmission is important too. This study is an effort to develop such a system with better understandability and interpretability. As per the objective of this study, classifiers have been created using various algorithms and have analyzed their performances. It has been proved that both typical and rule-based classifiers can produce good performing models. Few issues are to be discussed based on the findings from various experiments. One of the major challenges during experimentations was the high imbalance in the dataset on which has been worked on. To manage this problem, multiple measures have been taken, and the synthetic minority oversampling technique has been proven to work better than other techniques. Generating synthetic data points and adding them to the original dataset may cause the classifiers to behave differently in the real-world environment. Collecting more data for the minority class can be helpful to develop more sustainable classifiers. The experiments have been carried out to feature selection too. The empirical data shows that with feature selection, the classifiers can be robust. Among typical classifiers, the random forest algorithm has yielded better performance and has beaten currently existing works in this domain.

The main research outcome of this study lies in the rule-based predictive models. There have not been many works with those methods in this domain-specific classification tasks. Such models can help the stakeholders in a unique understandable way. By looking at the generated rules, the doctors and patients can easily grasp the way the models work. Although, the large number of rules generated by them can be overwhelming for them. Further works on reducing the rules and exploring ways to simplify them can add an extra dimension to this type of decision support system. Among the rule-based predictive models, the PART algorithm has the best performance. Though it has generated more rules than the RIPPER algorithm, the rules are short and simple. There are some limitations to this study too. To reduce the complexity of the classifiers, the three classes problem has been treated as a binary classification problem. The ">30" class has been merged with the "NO" class as interested in predicting early readmission. Although this issue could easily be resolved by conducting experiments for all these three classes, but that would add dis-ambiguity to the interpretation of the developed systems. Not utilizing more traditional algorithms is also a limitation of us. Only the decision tree and random forest algorithm have been applied. The other classification algorithms, like naive Bayes, SVM, logistic regression, etc., could be used for classification purposes, and their comparative analysis could add extra weights to the study. Despite the limitations, the outcome of this study has been good, and with the concepts presented by it, hospital readmission of diabetic patients can easily be predicted.

6 Conclusion

Ensuring better health care for risky patients even after the discharge is important. High-risk diabetic patients should be monitored carefully at the time of discharge. It should be ensured that they do not get readmitted to the hospital soon. Lacking in the follow-up treatment may increase the mortality rate and cost of health care too. To prevent these, an automated classifier that predicts the hospital readmission among diabetic patients can be useful. With that view, in this study, the unique idea of rule-based classifiers has been introduced. These classifiers can predict the 30 days readmission of a patient in an understandable way. This study has developed a few such systems, and it has been shown that with empirical data, they can surely work as a good decision support system.

References

1. Swain MJ, Kharrazi H (2015) Feasibility of 30-day hospital readmission prediction modeling based on health information exchange data. Int J Med Inform. https://doi.org/10.1016/j.ijmedinf.2015.09.003
2. Dharmarajan K, Hsieh AF, Lin Z, Bueno H, Ross JS, Horwitz LI, Barreto-Filho JA, Kim N, Bernheim SM, Suter LG, Drye EE, Krumholz HM (2013) Diagnoses and timing of 30-day readmissions after hospitalization for heart failure, acute myocardial infarction, or pneumonia. JAMA J Am Med Assoc. https://doi.org/10.1001/jama.2012.216476
3. Rubin DJ (2018) Correction to: hospital readmission of patients with diabetes. Curr Diab Rep. https://doi.org/10.1007/s11892-018-0989-1
4. Lutz R: Patients with diabetes often readmitted for hypo- and hyperglycemia. https://www.hcplive.com/view/patients-with-diabetes-often-readmitted-for-hypo-and-hyperglycemia
5. Alloghani M, Aljaaf A, Hussain A, Baker T, Mustafina J, Al-Jumeily D, Khalaf M (2019) Implementation of machine learning algorithms to create diabetic patient re-admission profiles. BMC Med Inform Decis Mak. https://doi.org/10.1186/s12911-019-0990-x
6. Cui S, Wang D, Wang Y, Yu PW, Jin Y (2018) An improved support vector machine-based diabetic readmission prediction. Comput Methods Programs Biomed. https://doi.org/10.1016/j.cmpb.2018.10.012
7. Ossai CI, Wickramasinghe N (2020) Intelligent therapeutic decision support for 30 days readmission of diabetic patients with different comorbidities. J Biomed Inform. https://doi.org/10.1016/j.jbi.2020.103486
8. Hammoudeh A, Al-Naymat G, Ghannam I, Obied N (2018) Predicting hospital readmission among diabetics using deep learning. Procedia Comput Sci. https://doi.org/https://doi.org/10.1016/j.procs.2018.10.138
9. Ribeiro MT, Singh S, Guestrin C (2016) Model-agnostic interpretability of machine learning
10. Ribeiro MT, Singh S, Guestrin C (2016) Why should I trust you? Explaining the predictions of any classifier. In: Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining. https://doi.org/https://doi.org/10.1145/2939672.2939778
11. Strack B, Deshazo JP, Gennings C, Olmo JL, Ventura S, Cios KJ, Clore JN (2014) Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records. Biomed Res Int. https://doi.org/https://doi.org/10.1155/2014/781670
12. Bhuvan MS, Kumar A, Zafar A, Kishore V (2016) Identifying diabetic patients with high risk of readmission
13. Harisekaran SSDG (2015) Big data analytics predicting risk of readmissions of diabetic patients. Int J Sci Res

14. Alajmani S, Elazhary H (2019) Hospital readmission prediction using machine learning techniques. Int J Adv Comput Sci Appl 10. https://doi.org/https://doi.org/10.14569/IJACSA.2019.0100425
15. Shameer K, Johnson KW, Yahi A, Miotto R, Li LI, Ricks D, Jebakaran J, Kovatch P, Sengupta PP, Gelijns A, Moskovitz A, Darrow B, Reich DL, Kasarskis A, Tatonetti NP, Pinney S, Dudley JT (2017) Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: a case-study using mount sinai heart failure cohort. In: Pacific Symposium on Biocomputing (2017). https://doi.org/https://doi.org/10.1142/9789813207813_0027
16. Aggarwal CC (2015) Data mining. Springer International Publishing, Cham. https://doi.org/https://doi.org/10.1007/978-3-319-14142-8
17. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority oversampling technique. J Artif Intell Res. https://doi.org/10.1613/jair.953
18. Han J, Kamber M, Pei J (2012). Data mining: concepts and techniques. https://doi.org/10.1016/C2009-0-61819-5
19. Vinet L, Zhedanov A (2010) A “missing” family of classical orthogonal polynomials. Math Intell. https://doi.org/10.1088/1751-8113/44/8/085201
20. Fürnkranz J, Widmer G (1994) Incremental reduced error pruning. In: Machine learning proceedings 1994. https://doi.org/https://doi.org/10.1016/b978-1-55860-335-6.50017-9
21. Cohen WW (1995) Fast effective rule induction. In: Machine learning proceedings 1995. https://doi.org/https://doi.org/10.1016/b978-1-55860-377-6.50023-2
22. Frank E, Witten IH (1998) Generating accurate rule sets without global optimization. In: Proceedings of fifteenth international conference on machine learning. https://doi.org/1-55860-556-8
23. Rubin DJ, Handorf EA, Golden SH, Nelson DB, McDonnell ME, Zhao H (2016) Developement and validation of a novel tool to predict hospital readmission resk among patients with diabetes. Endocr Pract. https://doi.org/10.4158/E161391.OR

Design and Analysis of Mobile-Based Tourism Security Application: Concepts, Artifacts and Challenges

Deepanjal Shrestha, Tan Wenan, Bikram Adhikari, Deepmala Shrestha, Adesh Khadka, and Seung Ryul Jeong

Abstract Tourism security has become a matter of great concern over the years for tourists around the world, and this attribute plays an important role in the selection of a destination. The application of technology can greatly enhance the safety and security environment of a tourist in a particular destination. This work proposes the analysis and design of a mobile-based application targeted to provide security and safety to a tourist. The design is based on requirements gathered from the tourist and tourism business personals through online and offline interviews which are grouped and coded into user requirements. Unified Modeling Language (UML) basic notation and tools are used to analyze and design the system in a comprehensive manner. Pokhara city of Nepal is chosen as the destination to deploy destination city to deploy and test the system. Further, the challenges and limitations of the work are discussed in light of privacy and system implementations. This work has a great significance in the

D. Shrestha · T. Wenan
School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China
e-mail: deepanjal@hotmail.com

T. Wenan
e-mail: wtan@foxmail.com

T. Wenan
School of Computer and Information Engineering, Shanghai Polytechnic University, Shanghai, China

B. Adhikari
Genese Solution, Kathmandu, Nepal
e-mail: bikram@genesesolution.com

D. Shrestha · D. Shrestha
School of Business, Pokhara University, Pokhara, Nepal
e-mail: deepmala@pusob.edu.np

A. Khadka
Ministry of Education, Science and Technology, Singhadarbar, Kathmandu, Nepal
e-mail: adesh.khadka@gmail.com

S. R. Jeong (✉)
Graduate School of Business IT, Kookmin University, Seoul, South Korea
e-mail: srjeong@kookmin.ac.kr

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_15

application of technology for personal security especially in the tourism area. It is one of the first kind of work done in the tourism sector of Nepal which adds a great value to the tourism industry. The work also contributes as a knowledge domain for the digital and mobile technology implementations in tourism destinations.

Keywords Unified Modeling Language (UML) · Global Positioning System (GPS)

1 Introduction

Tourism business around the world has accounted for 10.4% of global GDP and created 319 million jobs as per the report of World Travel and Tourism Council in 2018. The report also states domestic tourism had the strongest growth with 71.2% of the total spending in 2018 [1]. A lot of factors account for the growth of tourism business around the world which includes tourism environment, policies, infrastructure, cultural and natural resources [1], safety and security, and health and hygiene. Tourism safety has become a greater concern for the tourist around the world, and a lot of attention is paid to travel safety by the tourist. Tourism safety can be managed and handled properly with the use of technology. The IoT devices, social sites, Global Positioning System (GPS) technologies, image satellites, CCTVS, etc., are very powerful technological tools that are capable of working with security and disaster-related situations [1]. Mobile devices are the handiest and preferred devices that can be used as a personnel means of communication to get connected with individuals during sensitive situations. Mobile in tourism security is the best tool that can provide easy communication and act as an effective medium to avoid disaster and risk. The role of mobile devices and applications can be extremely vital for underdeveloped countries like Nepal which are in the middle of the technological boom and has concerns over its tourism business. Further, the data from Nepal Tourism Ministry depicts that in 2018, 619 cases related to tourism safety were reported which mostly include cases related to personal incidents. Cases with women, aged tourists, and minors were also seen in existence. Data published in 2017 by The Himalayan Times reported that 62 tourists have gone missing in the last five years which was alarmingly big for safe tourist destination image [2]. In light of these incidents, it can be inferred that the tourism security system is a must and one of the most important systems required. Thus, this research work is undertaken to analyze, design, and develop a mobile application for tourism safety in Nepal. The requirements are gathered from tourists, tourism stakeholders, and associated literature in an iterative cycle to come up with a mobile-based software requirement specification document. UML tools and techniques are used to design and develop the system. Easy to use interface with a priority of action and sequences during risky situations are taken as major considerations. The application design is of hybrid nature that works with mobile devices, tablets, PCs and computer systems having Internet access. The system is targeted for Pokhara city of Nepal, which is the tourist capital and the second-biggest city of Nepal.

2 Literature Review

Tourism security and technological implementations have taken the attention of a wide range of researchers around the world. Many studies have been done in the past that talk about design issues, social impacts, business impacts, future trends AI role in tourism security, and disaster management. Andriy Volkoviy and Llya Tikhov in their paper, Use of mobiles in security and safety systems explored all the different possibilities that can be implemented with mobile to provide safety to the users. They listed some good mobile applications like, Rescue, iMap Weather Radio, Earthquake Alert, Real-Time Warning, and Tsunami in the paper with their features and shortcomings [3]. Similarly, studies have been done on safety forecasting and early warning [4] safety of food services in the hospitality industry [5] safety and security in different geographies [6] and mobile application as guides integrated with security features [7] and using social site data in understanding tourist pattern for better management [8]. The book written by Michael and Annika extensively talks about the use of mobile applications frequently on vacation, where they talk about the varied uses of mobile applications in tourism destinations [9]. Studies are also carried on the role of mobile applications in tourism [10] and mobile application in women safety [11–14].

Table 1 represents a survey of some recent research work in the field of tourism, women's safety, and mobile technology. Further, the literature work regarding tourism risk and safety of Nepal has portrayed Nepal as a risky country with 84th rank out of 176 countries [1]. Various studies have been conducted on disaster management, risk mitigation, and adventure tourism in Nepal but very less data is available for technology and tourism. The data of Table 2 depicts the technological development of Nepal with technological breakthroughs (mobile, internet, and social site) and the current status quo [15]. The technological development has seen few published works of few notable authors Shrestha and Jeong, Bidur et al. Tan et al. [16] who talk about the role of technology in tourism but there is a gap and absence of literature in the case of mobile technology and tourism security.

3 Research Framework

The research framework plays a vital role in guiding the overall study of the research work. In our study, UML is used as a basic tool for the analysis and design of a mobile application for the tourism industry. Figure 1 depicts, five classes of the user (tourist, tourism business personnel, tourist officials, security agencies) who provide input to the system requirement. The requirements are gathered using an iterative approach which results in identifying concepts, use cases and systems artifacts which are further elaborated to come up with the development of the conceptual model, use case diagram, and UI model for the system. Some specific algorithms of tourism security are also proposed based on user requirements in Nepal as shown in Fig. 1.

Table 1 Survey of mobile technology and application in tourism safety

Paper details (year/publisher)	Research details	References
2012, Journal of Information and Security	Use of mobiles in security and safety systems	[3]
2015, IEEE Digital Explore	An android app for the safety of women	[11]
2015, Journal of Computer Science and IT	Android application for women security	[13]
2015, Journal of Computer Applications,	Mobile-based intelligent safety system for women	[14]
2015, KSII Transactions	Exploring the attractive factors of app icons	[17]
2017, KSII Transactions	An app visualization design for car accident prevention	[18]
2018, CISTI Conference	Technology usage for safety and security	[6]
2019, ICCPCT (Conference)	Mobile application for women safety and RT database	[12]
2019, Journal of Physics	Mobile apps in tourism communication	[10]
2019, Third World Conference	Twitter mining sentiment analysis in tourism industry	[8]
2019, IEEE Access	Safety forecasting and early warning	[4]

4 System Analysis and Design (Concepts and Artifacts)

The section below represents different concepts, artifacts, and processes identified and used for the development of mobile-based security system application.

4.1 User Requirements and Functional Description

User interview and narrative description of users which include tourist, tourism business personnel, tourist officials, and security agencies are recorded. The requirements are analyzed to identity artifacts, objects, processes, and functions. The functional requirements of the system are shown in Table 3. The functional requirements are coded and categorized as evident and hidden depending on their state in the system. Altogether ten major categories of functional requirements are identified numbered from R0 to R9, with each major category further extended to specific functional requirements. The extended functional requirements are coded and grouped based on major functional types, also identified as evident and hidden.

Table 2 Data representing technological development in Nepal [2, 15]

ICT Development	Year	Details
E-readiness and digital implementations	2016	135 as per world bank
Government development index rank	2017	1.85 as per maturity
Mobile communications	2013–2022 (forecasted)	25–136% (forecasted)
Mobile Internet penetration	2012–2017	21–57% by 2022
Internet penetration	2012–2017	21–57%
Active Internet users	2018	250 new Internet users per hour
Social site users	January 2018	9.3 M Facebook, 604 YouTube
Web sites status	2017	56,286 registered Web sites in Nepal
4G subscribers	2017	2,144,887
Broadband services	2020	90% of population gets connected
E-commerce	2018	Recorded a growth rate of 62% CAGR
Online travel	2018	Recorded a growth rate of 15% CAGR
Online media	2018	Recorded a growth rate of 44% CAGR

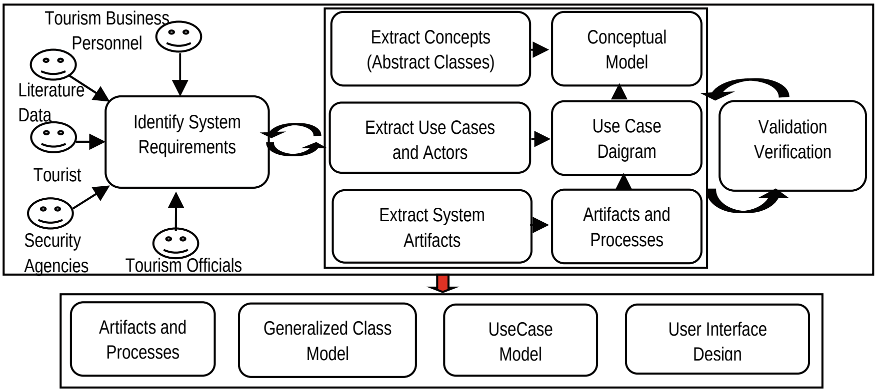


Fig. 1 Research framework of the study with detailed components

Table 3 User requirements for mobile security application

Reference	Function description	Category
R0	User registration requirements	Evident
R0.1	Get information details of the user who wants to use this application	Evident
R0.2	Check information with the database to find out errors, redundancy, format, etc.	Hidden
R0.3	Register information of the user after validation and provide a unique ID	Evident
R1	Destination-based information requirements	Evident
R1.1	Legal issues in sensitive places like airport for drugs, cash, gold, banned items, etc.	Evident
R1.2	Administrative issues: Visa regulation, fee	Evident
R2	Warning and alert message requirements	Evident
R2.1	On travel for taxi fare, bus fare, or other travel vehicles	Evident
R2.2	Sensitive situations like accidents, disaster, road blocks, rallies, etc.	Hidden
R3	Information requirements	Evident
R3.1	Alerts on destinations with regard to health and hygiene	Hidden
R3.2	Alerts on routes, services, prices, support, digital communication	Evident
R3.3	Information on special regulations for religious, cultural, sensitive places	Evident
R3.4	For change in tourism destinations, activities, and critical updates	Hidden
R4	Connection and information sharing requirements	Evident
R4.1	Police services, rescue agencies, local governance, etc.	Evident
R4.2	Tourism information centers and agencies	Evident
R4.3	Emergency rescue teams and agencies	Evident
R4.4	Nearby tourists or peers in existence	Evident
R5	Risk sensing and mitigation requirements	Evident
R5.1	Sense any incoming risk situation and provide solutions	Hidden
R5.2	Track the location and provide rescue	Hidden
R5.3	Provide backup plans for any risk situation	Evident
R6	Personal safety requirements	Evident
R6.1	Track and individual to ensure personal safety and inform for suspicion	Evident
R6.2	Provide safety to personal data including data and information	Hidden
R6.3	Prioritize solo females, aged, or very young tourist	Evident
R7	User application requirements	Evident
R7.1	Provide easy interface to trigger emergency situation	Evident
R7.2	Provide real-time sensing and monitoring feature	Hidden
R7.3	Provide information on safe routes or alternative routes short routes	Evident
R7.4	Provide information on safe food and hygiene locations	Evident
R7.5	Provide analysis of the risk of a visiting a place and provide instructions	Evident
R7.6	Provide features to report a complain	Evident

(continued)

Table 3 (continued)

Reference	Function description	Category
R7.7	Provide features for historical data of incidents and mishappenings	Evident
R7.8	Provide alarms, alert messages for any risk or disaster information update	Hidden
R7.9	Provide information recording and location-sensing feature for emergency	Hidden

4.2 *Conceptual Design*

Concepts define the entities of the system at an abstract level that can be later refined to form software class and objects which interact with one another to provide a system functionality [19].

4.2.1 Concept Extraction Process

The process of use case and concept extraction is subject to user requirements which are traced from interview data. For example user requirement R0 that states that a user must be a registered user to use the security application. The user gets registered by providing the necessary data and information as requested by the application. In this case, required data here is a user name, country, mobile, or passport number, whereas the process is to validate and register users by the system. This requirement is related with a major use case Register_User. Further, Register_User has further dependencies to complete this process which include get info, check info, process registration, and invoke OTP. The conceptual class person (tourist/general user) invokes this use case and provides all the necessary information and gets himself stored as a user object in the system. Further in Register_User, use case process needs other conceptual classes like policy information from governing bodies, communication roles from devices classes, and other objects to validate and complete this process. Thus, user requirements and use cases serve as the base to extract the concepts. A similar process when mapping requirements can be applied to design concepts and map them as conceptual models in the early requirement and detailed design phase. Figure 2 shows the conceptual class modeled in a generalized relationship with few extended as inherited classes.

4.2.2 Generalized Conceptual Model

This section depicts the process of identification of concepts in a tabular form extracted from user requirements and other design artifacts of the system. Table 4 depicts a class person that extends the types of users which include tourist type (inbound, domestic), tourism personnel, and government officials. All identified

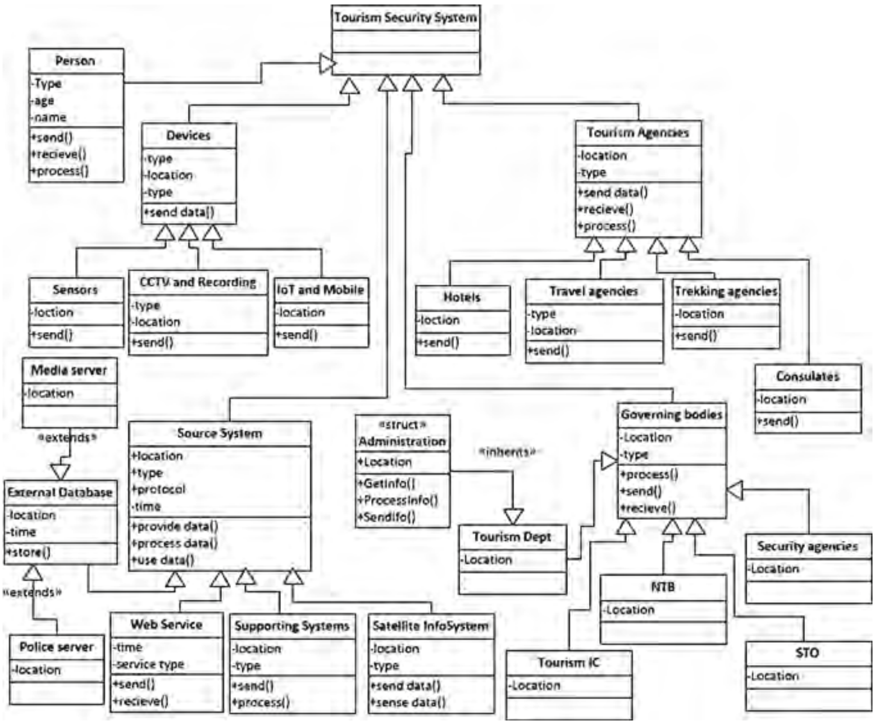


Fig. 2 Conceptual model of security system is depicted with a generalized relationship

Table 4 Representing base class person with other attributes and extended classes

Base class	Extended classes	Properties	Functions	Multiplicity
Person (Users)	Local citizen	Name, age, address	Post data, use data, get information	*….1
	Domestic tourist	Name, age, address	Post data, use data, get information	*….1
	Inbound tourist	Name, age, address	Post data, use data, get information	*….1
	Tourism personnel’s (Media, tourism labor)	Name, age, address	Post data, use data, get information	*….1
	Government officials	Name, age, address	Post data, use data, get information	*….1

concepts have properties, functionalities, relationships, and multiplicities associated with them. Similarly, the generalized systems represent other important base concept classes which include the source systems, governing bodies, tourism agencies, and devices class at a higher level. Each base concept class is further extended into many generalized classes that further includes the specialized classes needed by the system to form the real artifacts of the system. The objects of the classes form the dynamic part of the system that is refined from the conceptual model and use the case model in the later phase of the design process in an iterative manner. The generalized conceptual model is shown in Fig. 2 that represents the conceptual classes of the mobile-based security system.

Further, the class concept of external database and source system are important concepts that are related to information provider and consumer objects which include media servers, police servers, etc. The device class in the diagram is a dedicated information providers related to Sense_Data; use case and includes sensors, mobile devices like smartphones, iPad, CCTVS, recording system, and devices. These devices are capable of performing functions that include sending data, receive instructions, post data, and can create information instances. The class tourism agencies act as an important actor during rescue operations so it is taken as a default security agency for the tourist and is related to many use cases which include Create_Circle, Invoke_Panic, and Connect_Agencies. The embassies and consulates are more concerned with the well-being of their citizens. Similarly, the class governing body is represented with associated classes Department of Tourism, Security Agencies, Nepal Tourism Board (NTB), State Tourism Office (STO), and Tourism Information Centers (TIC). These classes are related to security policymaking, management, and implementation of it.

4.3 Use Case Model

Use cases are the stories that illustrate and imply the requirements of a system in an informal way [19]. Use cases are the narrative descriptions that help in identifying the system requirements that can be further elaborated to form a sequence diagram, class diagram, and collaboration diagram representing the static and dynamic view of the system. The section below represents various high level and extended use cases of the system.

4.3.1 The High-Level Use Cases

The high-level use cases are created at the initial level of system analysis and design. İt is more abstract and gives a good understanding of the system processes that include one or many functions to complete a particular task of the system [19]. Table 5 represents 13 major uses cases of the mobile security application where each major use case is related to the number of extended use cases that capture the complete

Table 5 Representing major tourist (user) initiated use cases and extended use cases

Major use cases	Extended use cases	Actors
Register_User	Get info, check info, process registration, invoke OTP	Tourist, user
Create_Circle	Get contact list, add contact, add message, define level	Tourist, user
Emergency_Rescue	Send info, display security agencies, contact rescue agency, monitor real time	Tourist, user
Real-time_Monitoring	Capture location, sample info, check status, issue alerts, mark tourist	Tourist, user
Invoke_Assistance	Ask Info, check database, compute info, provide info	Tourist, user
Plan_Trip	Collect travel info, check database, create plan, display plan, provide instructions	Tourist, user
Route_Navigate	Get location, compute alternate paths, get related data, provide suggestion	Tourist, user
Security_Update	Check database, capture changes, display info	Tourist, user
Share_Location	Capture location, send data, send message	Tourist, user
Connect_Peers	Inform peers, inform circle, display peers, monitor real time	Tourist, user
File_Complaint	Open form, post data, upload info, submit data	Tourist, user
Offline_Assistance	Invoke backup, search request, provide info	Tourist, user
Invoke_Panic	Get location, alter agencies, make rescue, update info	Tourist, user

functionality of the system. As seen, the major use case Create_Profile requires Get_info, Check_info, Process_Registration, and Invoke_OTP use cases to complete its functional process. A person is an actor that initiates this process and closes it. Similarly, we can relate, understand and deduce other uses cases of the system domain. The actors for these use cases are a person that represents domestic, inbound, and tourism personnel. The other uses cases include Create_Circle, Emergency_ Rescue, RT_ Monitoring, Invoke_ Assistance, and Plan_Trip, etc., as shown in Table 5.

4.3.2 Use Case Diagram

Use case diagrams are the pictorial representation of the use cases that show the system processes in a diagrammatic form [19]. Figure 3 represents complete system use cases including their actors and system boundary. The use case diagram represents 13 major use cases that are triggered by tourist and user and seven major use cases that are triggered by the external server system. The tourist use cases are connected with black lines, while the users with red lines and systems with blue lines as shown below. The mobile application, in this case, serves as the system boundary and is a part of a larger system, the security management system. Further, in Table 6, the sequence of events is explained between the actor and system to show the implementation of various use cases in the system. The sequence of events can

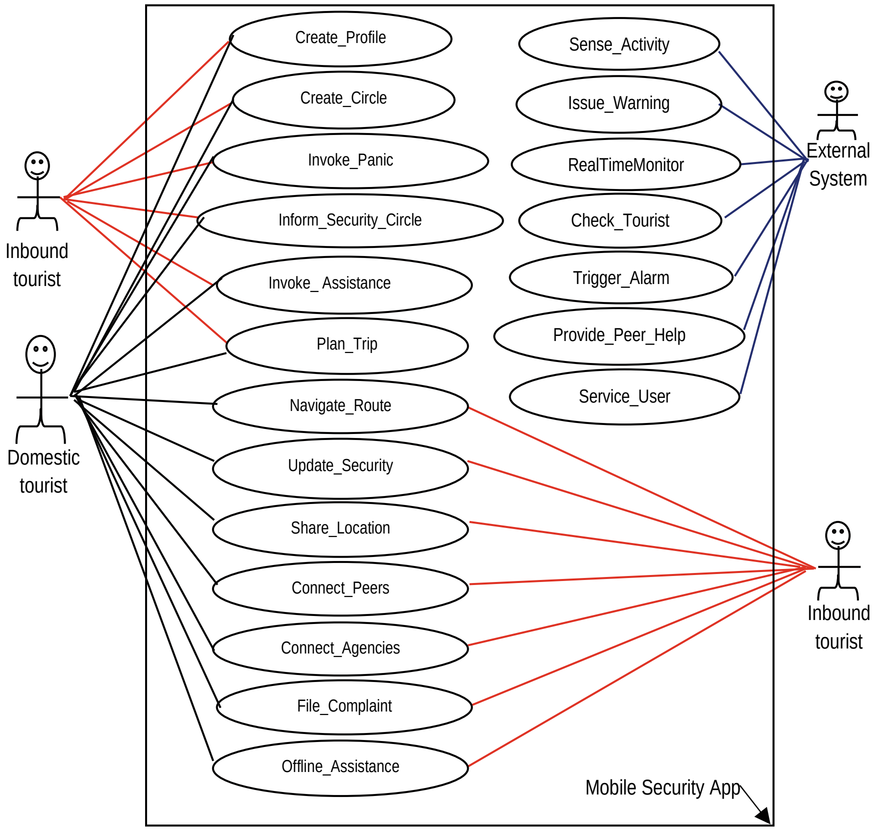


Fig. 3 Use case diagram of the mobile security system application with related actors

Table 6 Typical course of events for use case: Invoke_Application

Actor action		System response	
1	User initializes application with a touch	2	System prompts for user input
3	User inputs preferred details	4	System validates input types (options)
5		4a	Logs in as register user and waits for the next action
6	User acts with information	4b	Logs in as guest user and waits for the next action
		4c	Initiates panic module and distribute information of the user
Dependency		**GetUser_info, Check_Input_Type, Validate_Input_Type, Trigger_Action**	

later be extended to come up with sequence diagrams and collaboration diagrams that show the behavioral design of the system. Table 6 shows a typical course of events for Invoke_Application that has three optional paths that are shown as 4a., 4b. 4c in course of events. Table 6 also shows the dependency of these primary use cases on other use cases which include GetUser_info, Check_Input_Type, Validate_Input_Type, and Trig-ger_Action to complete a particular system process.

4.4 The Radius R-algorithm

The radius R algorithm is a security recommender system that works for providing travel route and blockade/accident scenario checkup suggestions to the tourist-based on their current location and destinations as shown in Fig. 4. The algorithm is specially designed for the security scenario of Nepal considering its important parameters. The system is named as radius R algorithm because it takes radius as its main parameter to compute paths and trace agencies for making recommendations from the specified or current node. In case, the tourist meets an accident at node (n), then also system considers radius R to make a scan to locate health agencies, security agencies, and rescue teams so that appropriate recommendations can be made. The radius is set to 2 km in case of metropolitan and sub-metropolitan areas while the radius is set to 4 km in case of rural areas because the metro and sub-metro are stuffed with health, security, and rescue facilities closely which is not present in the case of rural areas. On the first scan, if the defined radius area finds no health, security, or rescue agency, the algorithm increases the radius by 50% and makes the next cycle of scan again, repeating further.

This process keeps increasing the radius by 50% each time until agencies or nodes are spotted. The same logic applies to recommend routes and paths also. Besides this,

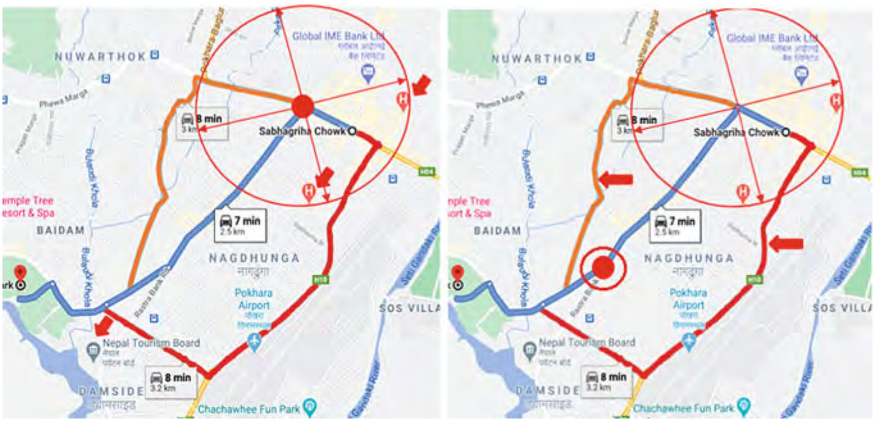


Fig. 4 Representing different conditions for route finding and recommendations

the algorithm synchronizes its data for day and time with the stored calendar of Nepal to see the holiday period and working hours of a concerned agency which is further computed to make accurate recommendations as explained in Fig. 5.

Table 7 shows the comparison of attributes of R algorithm with some of the other algorithms, and it is seen that this algorithm has some features that are not present in other algorithms like the backtracking of the path and recommending agencies based on radius. Further, this algorithm is improvised to get synchronized with the Nepal government officials calendar data to make realistic recommendations. The R algorithm is applicable for smart transportation, smart city management, security, and disaster management applications as it combines the features of searching the travel paths, travel points, and security agencies located in the circular periphery with a backtracking system. The adjustment of an algorithm to keep increasing the radius with scan cycles to locate a valid information cluster is a strong feature. The other algorithms that were specifically targeted for the Nepalese tourism security system included the connect peers and rescue algorithm along with other general algorithms.

1. Capture user mobile location Start → SU_{loc}
2. Ask User Destination location → DU_{loc}
3. Check SU_{loc} = [Metro, Sub-Metro, Rural], Set SU_{loc} as Type = SU_{loct}
4. Set (R = 3 Km for Metro and Sub Metro and R = Rx2 for Rural Type)
5. Define → R_{radius} = R (Based on SU_{loct})
6. Mark SU_{loct} → Primary location (*PL*)
7. Set PL ← as primary location, scan travel points (Tp) to find paths
8. Compute paths scores for [C1, C2, C3.... Cn] based on [distances, time and path type]
9. Sort path in order of computed score [SC_{P1}, SC_{P2}.... SC_{Pn}]
10. Check path for accidents and blockade
11. If path → [Accident = False or Blockade = False]

    Recommend → Paths [SC_{P1}, SC_{P2}, SC_{P3}]
12. Else

    Cycle = 1, increase = 2

    Scan "R" for Cycle 1 = $S_{can}C_{ycle}$1

    If HA = ø, SA = ø and RA = ø

        R = R + R/Increase, Repeat Step 12 until If HA ≠ ø, SA≠ ø and RA ≠ ø,

        Increase = increase+1

        Mark health agencies → HA, Security Agencies → SA, Rescue Agencies → RA

        Create list HA ← [HA1, HA2...Han], SA ← [SA1, SA2...San], RA ← [RA1, RA2,..RAn]

        Compute date and time, Dt and Tt

        Check if Dt → Holiday = [True], Tt → No Office Hours = [True]

        Scan List [HA, SA, RA] = Open = True

        Recommend HA → [HA1, HA2...HAn] ,SA ← [SA1, SA2..San] , R ← [Ra1, RA2..RAn] True

    Else

        Recommend HA → [HA1, HA2...HAn] ,SA ← [SA1, SA2..San] , R ← [Ra1, RA2..RAn]
13. Stop

Fig. 5 Radius "R" algorithm structure

Table 7 Comparison of R algorithm with some existing systems and algorithms

Attributes	Captures location of the user	Suggest shortest path and time	Suggest places based on sensing (Mishappening, roadblocks)	Suggest location based on circular area (radius) and backtracking	Works with Nepal government data for holidays, office hours for recommendations	References
R algorithm	Yes	Yes	Yes	Yes	Yes	–
safety forecasting and early warning	Yes	Yes	Yes	No	No	[4]
Intelligent safety system	Yes	Yes	Yes	No	No	[8]
Distributed societal security system	Yes	No	Yes	No	No	[11]
Google map	Yes	Yes	No	No	Yes	[20]
Dijkstra's algorithm	No	Yes	No	No	No	[20]
Any angle path planning	No	Yes	No	No	No	[20]
Mobile/IoT-based Algo	Yes	Yes	Yes	No	No	[21]
Mobile in security and safety	Yes	Yes	Yes	No	No	[22]

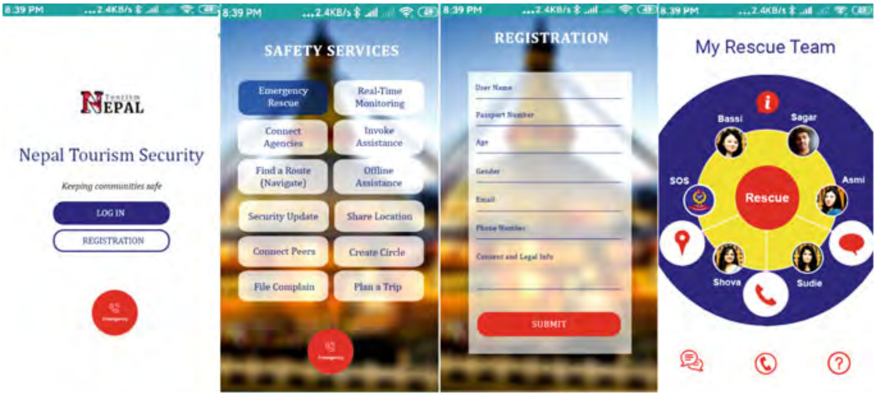


Fig. 6 Snapshots of screens including welcome, services, rescue, and registration

4.5 User Interface Design

The user interface design serves as the heart of the whole system process, and several designs were reviewed before coming up with the interface of the application. The more sensitive requirements were the given priority in user interface design and icons. The GUI principles especially for visually targeted web applications were considered with more attention to color and text size. Few snapshots of the user interface are depicted in Fig. 6.

5 System Design Verification and Validation

The system analysis and design are verified and validated using design walkthrough and backtrack tracing in the case of different models developed [19]. It can be seen from Table 8 that class user/tourist is related to use case Create_Profile which has tourist/user as an actor, and it can be traced back to user requirement R0, which consists of sub-functions R0.1, R0.2, and R0.3, respectively. Similarly, class source systems can be related to use case Realtime_Monitor, Update_Security, which is further related to actor tourist/user and server system and can be traced back to requirements R1, R2, R3, and R4, respectively. This shows that analysis and design are in coherence with the user requirements and follow a connection between design artifacts to guarantee a good design.

Table 8 Verification and validation matrix

Class	Use case	Actor	Traced to requirement
User/tourist	Create profile	Tourist/user	R0
Source system	Realtime_Monitor, Update_Security, Navigate_route, Sense_Activity	Tourist/user, Server system	R1, R2, R3, R4
Device systems	Realtime_Monitor, Update_Security, Navigate_route, Sense_Activity	Tourist/user, server system	R1, R2, R3, R4
User/tourist	Plan_Trip, Navigate_Route	Tourist/user, server system	R1, R3, R6, R7
User/tourist	Plan_Trip, Navigate_Route	Tourist/user, server system	R1, R3, R6, R7

6 System Security and Testing

Digital systems and mobile systems are prone to attack from hackers, malicious bots, viruses' worms, Trojan horses, etc. To provide security to the application, current security measures are used in the system that includes an SSL certificate, strong password policy, mobile or web-based OTP, and registration based on either passport number or mobile phone to authenticate. The guest login is limited to view the only privilege. The password is assigned 15 days of life, and the user has to go with a two-factor authentication after the period expires. To protect the personal data of users, the database is designed to show no policy and is centrally stored in a secure server in a two-layered system that can only be accessed by authorized government officials through verification and validation of their identity. The testing of the system is done with user-generated data and in a controlled environment. Almost all the modules and paths are tested in the application and data results are satisfactory. The work lacks testing with real-world data and real-world implementation. Performance, load balancing, and recommendation preciseness need to be tested together for 100% accuracy.

7 Challenges and Limitations of the System

The design and implementation of the mobile-based tourism security application have different challenges associated with it. Infrastructure is one of the biggest challenges for the robust implementation of the system. The security application is dependent on different sources and systems to extract data related to the security environment. The real-time collection, processing, and distribution of data require a huge digital and communication infrastructure that must be capable of processing complex data and algorithms. User data privacy and security are another major concern, as the data of individuals have a high chance of being misused, and it needs a strong legal and system policy to ensure user data safety. The optimal use of the application by users as a trusted system can be difficult due to the privacy concerns and nature of the tourist as they are from different countries, backgrounds and have issues with language, icons, and information models used. In the context of Nepal, IT education, system support, and governing policies can add more challenges to the system implementation. The designed application is good for tourism recommendation with security as the leading priority but still suffers from issues of cyber-attacks and misuse of personal data from officials having access to it. It is also dependent on external sources for data and information.

8 Conclusion

The above study concludes that security applications are one of the most vital needs for the tourism industry to guarantee safety as well as create image of the destination. A secured tourism destination has more chances of getting tourists from all sections and age groups compared to unsafe destinations. Technology can be aggressively used to implement security in the tourism sector at various levels. An integrated tourism security system with mobile application extension for smartphones can enable safety in the hands of tourists visiting a place. A good and well-analyzed mobile application can not only work as a good security system for tourist but can be enhanced as a personal security system too. Though privacy and security of personal data remains a big question for the implementation aspect, the overall realized advantage minimizes the privacy concern. A well-thought plan with stiff regulations must be imposed before implementing this kind of system. Further the design of software systems using UML tools and techniques greatly helped in identifying the base components, processes, and functions of the system. The UML approach made an analysis and design process more effective with designated notations and symbols. The designs carried with extensive research and GUI principles made it look more appealing and attractive. User-provided excellent feedback in the testing phase for the UI design, which was considered as a merit achievement. The specifically designed algorithms serves the local customized needs of the tourist in Nepal. It is concluded that the system is well designed with engineering principles, in time and has been well documented. This work serves as a good source of knowledge and reference for future systems.

Acknowledgements We would like to thank the Ministry of Tourism, Culture and Civil Aviation of Nepal for providing necessary support and data for this work. Further, this work is supported in part by the National Natural Science Foundation of China under Grants No. 61672022 and No. U1904186.

References

1. WTTC, Travel and Tourism Economic Impact 2019, World Travel & Tourism, 65 Southwark Street, London SE10HR, United Kingdom. https://www.slovenia.info/uploads/dokumenti/raziskave/raziskave/world2019.pdf
2. Aryal NP et al (2019) Nepal tourism statistics 2018. Ministry of Culture, Tourism and Civil Aviation, Kathmandu, Nepal. https://tourism.gov.np/files/statistics/19.pdf
3. Volkoviy A, Tikhov L (2012) Use of mobiles in security and safety systems, information & security. Int J 28(1):146–153
4. Yin J, Bi Y, Zheng X, Tsaur R (2019) Safety forecasting and early warning of highly aggregated tourist crowds in China. IEEE Access 7:119026–119040. https://doi.org/10.1109/ACCESS.2019.2936245
5. DusenkoSV (2018) Digital technology in ensuring the safety of food services in the hospitality ındustry,2018. IEEE ınternational conference (IT&QM&IS), St. Petersburg, pp 618–619, doi: https://doi.org/10.1109/ITMQIS.2018.8524926
6. Au-Yong-Oliveira M, Moreira F, Martins J, Branco F, Goncalves R (2018) Technology usage as a way to increase safety and security in different geographies: testimonials on the use of technology in Rio de Janeiro, Brazil,2018. In: 13th Iberian conference on ınformation systems and technologies (CISTI), Caceres, 2018, pp 1–7. doi: https://doi.org/10.23919/CISTI.2018.8399266
7. Alrehili M, et al (2018) Tourism mobile application to guide madinah visitors. In: 2018 1st ınternational conference on computer applications & ınformation security (ICCAIS), Riyadh, pp 1–4. doi: https://doi.org/10.1109/CAIS.2018.8442023
8. Gupta G, Gupta P (2019) Twitter mining for sentiment analysis in tourism ındustry. In: 2019 third world conference on smart trends in systems security (WorldS4), UK, pp 302–306. doi: https://doi.org/10.1109/WorldS4.2019.8903940
9. Beier M, Aebli A (2016) Who uses mobile apps frequently on vacation? Evidence from Tourism in Switzerland. ICT in Tourism 2016. Springer. https://doi.org/https://doi.org/10.1007/978-3-319-28231-2_40
10. Abdul R et al (2019) Mobile apps in tourism communication: the strengths and weaknesses on tourism trips. In: The 2nd joint ınternational conference on emerging computing technology and sports 2019, Bandung, Indonesia. https://doi.org/https://doi.org/10.1088/1742-6596/1529/4/042056
11. Yarabothu RS, BramarambikaThota A. An android app for the safety of women. In: 12th IEEE India ınternational conference E3-C3, At: JamiaMilliaIslamia, New Delhi, Indıa. https://doi.org/10.1109/INDICON.2015.7443652
12. Prashanth DS, Patel G, Bharathi B (2017) Research and development of a mobile based women safety application with real-time. Int Conf ICCPCT 2017:1–5. https://doi.org/10.1109/ICCPCT.2017.8074261
13. Harini R, Hemashree P (2019) Android application for women security. J Comput Sci Inform Technol 8(10):54–59. ISSN 2320-088X
14. Paradkar A, Sharma D (2015) All in one ıntelligent safety system for women security. Int J Comput Appl 130(11):33–40
15. Frost & Sullivan and Government of Nepal, Ministry of Communication & Information Technology, 2018 Digital Framework Nepal. https://mocit.gov.np/application/resources/admin/uploads/source/EConsultation/Final%20Book.pdf
16. Tan W, Shrestha D, Jeong SR (2019) Digital tourism development and sustainability model for Nepal. In: 2019 IEEE CSCWD, Portugal, 2019, pp 182–187. doi: https://doi.org/10.1109/CSCWD.2019.8791852
17. Ho C, Hou K (2015) Exploring the attractive factors of app icons. KSII Trans Internet Inf Syst 9(6):2251–2270. https://doi.org/10.3837/tiis.2015.06.016
18. Jeong Y, Jeong E, Lee B (2017) An app visualization design based on ıot self-diagnosis micro control unit for car accident prevention. KSII Trans Internet InfSyst 11(2). doi: https://doi.org/10.3837/tiis.2017.02.020

19. Larman C (2005) Applying UML and patterns, 3rd edn. Pearson Education. ISBN-13: 978-0131489066, ISBN-10: 0131489062
20. Wikimedia Foundation, Inc., Pathfinding, Creative Commons Attribution-ShareAlike License July 2020. https://en.wikipedia.org/wiki/Pathfinding
21. Bin C et al (2019) A travel route recommendation system based on smart phones and IoT environment. Wirel Commun Mob Comput. https://doi.org/10.1155/2019/7038259
22. Volkoviy A, Tikhov I (2012) Use of mobile application in security and safety systems. Inf Secur Int J. https://doi.org/10.11610/isij.2812

An Efficient Approach to Software Fault Prediction

Md. Anuvob Pradan, Mahia Binte Mizan, Moon Howlader, and Shamim Ripon

Abstract The use of machine learning concepts in the software engineering field is now ubiquitous to predict software defects. Most of the Software Defect Prediction (SDP) datasets are highly imbalanced and filled with multiple irrelevant features that cause negative effects on the results. The goal of this paper is to create an approach to predict the software faults efficiently from imbalanced and multi-featured SDP datasets. Two highly imbalanced and multi-featured NASA MDP datasets have been used in this experiment. Initially, data cleaning has been performed with the help of *Z*-score technique to eliminate noisy instances. To balance the datasets, Synthetic Minority Oversampling Technique (SMOTE) oversampling technique has been used. Furthermore, to select the relevant features, three well-known feature selection techniques, as well as their ensembles, have been applied. Finally, to measure the performance, four well-known classification algorithms are implemented and evaluated the results by their accuracy, TPR and TNR.

Keywords Synthetic Minority Oversampling Technique (SMOTE) · Software Defect Prediction (SDP) · Machine Learning (ML) · K-Nearest Neighbor (KNN) · Logistic regression · Decision tree · Random forest

1 Introduction

Softwares are now vital for both technical and scientific purposes. As the need for good quality software increases, so does the necessity for quality assurance activities

Md. A. Pradan (✉) · M. B. Mizan · M. Howlader · S. Ripon
Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh
e-mail: mdanuvobprodan@gmail.com

M. B. Mizan
e-mail: m.mizan0129@gmail.com

M. Howlader
e-mail: ahmoon67@gmail.com

S. Ripon
e-mail: dshr@ewubd.edu

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_16

such as testing, verification and validation, fault tolerance, and fault prediction. Fault prediction algorithms are often used by organizations with a low budget and time constraints for detecting fault-prone modules. To improve the detection rate and performance of software fault detection, Machine Learning (ML) is considered to be the most popular method [1]. However, applying ML raises the concern of imbalanced datasets. As the Software Defect Prediction (SDP) datasets are highly skewed and filled with multiple irrelevant features, it results in biased trained models, which return faulty modules as non-faulty. This was shown in earlier works, e.g., Mohsin et al. [2]. When these models are used on various other software the cumulative effect of low true positive rate causes numerous errors to be generated.

In this paper, a novel approach to predict software faults efficiently is proposed. The proposed approach consists of a data cleaning method to reduce noisy instances, SMOTE oversampling technique to generate pseudo instances, and four feature selection algorithms, for dimensionality reduction, have been included.

In an imbalanced dataset, the results get biased toward the majority class; this, in turn, produces models with poor predictive performance, specifically for the minority class. For datasets, such as the SDP dataset where defect cases are the minority and less likely to happen than non-defect cases, the minority class is more crucial. Therefore, classes with fewer instances are more prone to classification errors than those with a higher number of instances. Oversampling techniques try to solve this problem by adding instances to the minority class either by duplicating or adding fake data. It can further be rectified by performing data cleaning first, to remove noisy instances or outliers. An outlier is a data point that differs so much from other values that it is suspected of being obtained by a separate mechanism [3].

The SDP datasets contain a large number of features, but every feature in the dataset does not influence the classification results. The initial dataset may contain irrelevant, duplicate, and useless data that do not affect the learning outcome; instead, these attributes can consume extra processing time, and for that, the quality of the result might decline [4]. By removing these irrelevant and unimportant features, the classification performance can be improved [5]. Three well-known filter-based feature selection techniques, namely chi-square, feature importance, and relief, and their ensemble have been applied in this work. Our proposed approach works by minimizing the effect of a skewed distribution of classes in the training data to increase the value of true positive rate and classification performance.

The rest of the paper is organized as follows. After reviewing related work in Sect. 2, a brief overview of the dataset used in this study is discussed in Sect. 3. Then, the proposed methodology is given in Sect. 4, while in Sect. 5 provides the evaluation results through experiments and comparative analysis. Finally, the conclusion and the future scope are presented in Sect. 6.

2 Related Work

Dataset balancing concept is being used effectively for the imbalanced dataset. Kamei et al. evaluated the effect of over and undersampling on imbalance fault-prone module detection [6]. Ramezankhani et al. [7] applied SMOTE oversampling on imbalanced diabetes dataset to study the impact. Pelayo et al. [8] focus on the SMOTE oversampling to determine to improve the recognition of defect-prone modules. Huda et al. [9] used different oversampling technique to balance software defect prediction dataset. Zohu et al. [10] utilize the combined sampling technique which includes SMOTE oversampling and undersampling to balance the dataset. The feature selection has been performing effectively on software fault prediction since the beginning until now. Many approaches to feature selection have been used to bring efficiency to the software fault prediction sector. Agarwal et al. [11] propose a feature selection-based Linear Twin Support Vector Machine (LSTSVM) model to predict software modules that are defect prone. Xu et al. [12] in their paper investigate the impact of 32 feature selection techniques on the defect prediction performance. Shivaji et al. [13] have investigated multiple feature selection techniques that are generally applicable to classification-based bug prediction methods. With the use of five feature selection techniques, they discarded less important features. The findings and limitations of some previous works are summarized in Table 1.

3 Dataset Overview

The datasets used in the experiments herein are collected from the NASA MDP repository software engineering databases [25]. These datasets vary in the number of instances, their degree of complexity, and Imbalance Ratio (IR). In Fig. 1, a short overview of the two datasets used in this paper is shown. The imbalance ratio varies from 12.2% (highly imbalanced) to 1% (only slightly imbalanced). The diversity of the dataset is also considered in the number of instances, while CM1 has 344 instances, PC2 dataset contains 1585 instances.

Only the class attribute is categorical, the rest of the attributes contain numerical data. Both datasets contain 38 attributes, and all attributes are common.

4 Proposed Model for Fault Detection

Numerous fault detection methods have been studied over the years, and increasing the detection accuracy remains to be an important criterion for developers. However, when the dataset is imbalanced, and a single class makes up a significant part of the dataset, merely using accuracy as a metric is not sufficient. If data size differs greatly between major and minor classes, then detection rate of minor classes is heavily

Table 1 Comparative analysis of the state of the art

References	Findings	Limitations
[14]	Authors have applied five feature selection techniques on software defect prediction datasets to reduce the dimension	Data balancing was not discussed or performed
[15]	Authors have addressed different topics in which feature selection plays a crucial role	
[16]	Seven feature selection techniques were applied to select the most relevant features from software defect prediction dataset	
[12]	Two versions of NASA dataset were used in their study. One was noisy, and the other one was clean dataset. And 32 feature selection method was used	
[17]	Four filter feature ranking and fourteen filter subset selection methods were evaluated in this study	
[18]	Selection of attribute with log filtering was applied for feature selection, and naive Bayes classifier was used for defect prediction	No data balancing was applied, and only one classifier was considered to predict the software defects
[19]	Authors have used bat-based search algorithm for feature selection and resampling to balance the dataset	Only one classification algorithm was used for the defect prediction purpose
[10]	Their model adopts chi-square test for attribute selection, combination oversampling and undersampling to balance the dataset and J48 classifier for classification	Barely one feature selection method was used to select relevant features, and one classification algorithm was used for classification
[20]	Authors developed a hybrid sampling technique to balance dataset then compared their technique with SMOTE and virtual oversampling	The dimensionality of the dataset was not reduced
[21]	SMOTE oversampling was used to balance the dataset, and six classifiers were used for the classification	
[22]	Ensemble classifier was used for classification, and SMOTE oversampling was used to tackle imbalance distribution	
[23]	A fuzzy-based oversampling algorithm was used to handle the imbalanced data	
[9]	Authors built an ensemble model using different oversampling techniques, which consider the class imbalance problem	PROMISE software engineering dataset was used, and feature selection was not applied

(continued)

Table 1 (continued)

References	Findings	Limitations
[24]	A resampling technique with three types of ensemble learners was introduced to balance the dataset	

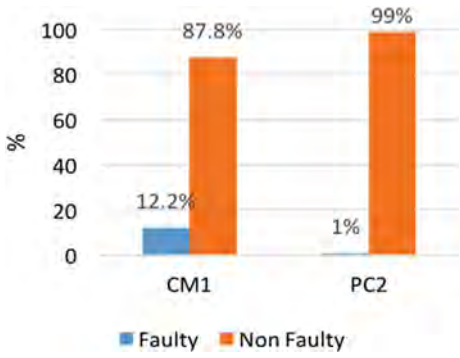


Fig. 1 Dataset overview [25]

affected; therefore, in this study, the skewness of the dataset is focused to remove and improve the recognition of the minority class or true positive rate.

To reduce the effect of imbalanced data in fault detection, the proposed model is implemented with original data according to the flowchart shown in Fig. 2. Techniques used such as the *Z*-score technique for the data cleaning step, SMOTE for data balancing, chi-square, relief, feature importance, and ensemble for feature selection step contributes to the elimination of outliers, minimization of the effect of asymmetry between classes and reduction of multi-dimensionality of the dataset, respectively. All the selected learning algorithms cover distinct types of methods in the training of imbalanced dataset.

All steps along with the techniques used have been briefly discussed below.

4.1 Data Cleaning

To reduce the possibility of noise generation, data cleaning has been applied in this model. It improves the quality of the training data by removing noise and correcting inconsistencies in the data.

First, the instances of the majority class that contain outliers have been identified, using the *Z*-score technique. The *Z*-scores [26] can quantify the unusualness of observation or in other words determine extreme points numerically. Instances that have a higher absolute value than the normalization threshold of 3 or −3 is recognized. This threshold value is commonly used for large datasets to detect and eliminate noisy

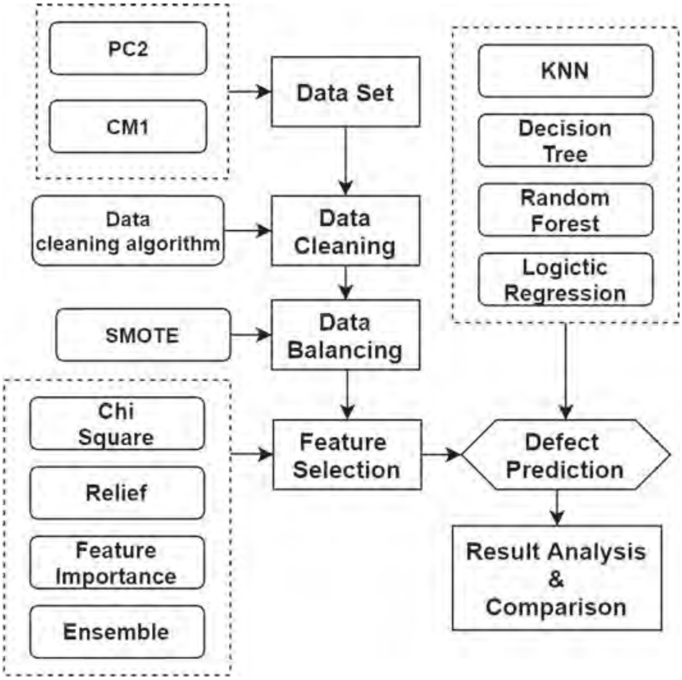


Fig. 2 Process mapping

data, where 99% of the values have a Z-score between 3 or -3; this means that they lie 3 standard deviations above and below the mean.

(a) Parameters for calculating Z-score

$Z = Z$-Score value

$x =$ each value from the dataset

$\mu =$ mean or average

$\sigma =$ standard deviation

$n =$ total instances in the majority class

threshold = 3.

(b) Formulas for calculation Z-score

$$\mu = \frac{\sum_{i=1}^{n} x^i}{n} \tag{1}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(x^{i} - \mu\right)^{2}}{n}} \tag{2}$$

$$Z = \frac{x^{i} - \mu}{\sigma} \tag{3}$$

To data cleaning, only the instances of the majority class have been considered, as the number of instances of the minority class is significantly lower. Any reduction from the minority class may lead to the generation of a high percentage of artificial data when oversampling will be performed. The pseudo-code of the proposed algorithm is illustrated in Algorithm 1.

```
1.  Algorithm Begin
2.  Separate the majority instances form the dataset
3.  n← total instances in majority class
4.  outlier_count←{0}
5.  for i=1 to total_feature_in_dataset:
        average[i]←average of all x in feature i
        S.D[i]←standard deviation of each feature i
                End for
6.  for i=1 to total_feature_in_dataset:
        for j=1 to n:
          Z←(x_j-average[i])/S.D[i]
          If(abs(Z)>=threshold):
          outlier_count[j]+ +
        end for
              end for
7.  for i=1 to n:
        if outlier_count[i]=0:
          remove i^th instance form the dataset.
                  End for
8.  End of algorithm
```

In Algorithm 1, first, the instances of majority class are separated from the dataset, and the count value of the total number of majority instances is stored in variable *n*. An array is defined for outlier count, and it is initialized with the default value, zero. The index number of an array denotes the instance number. Therefore, outlier_count[*i*] indicates the total number of outliers in the *i*th instance.

Next, the average value and the standard deviation are calculated for each feature and stored in two different arrays named average[] and S.D[]. Here, the index number of the arrays denotes the feature number. Hence, average[*i*] means the average value of the *i*th feature and S.D[*i*] denoted the standard deviation value of the *i*th feature.

Then, for each feature, the *Z*-score value of every instance is calculated, and the absolute value of *Z*-score is compared with the threshold value. If the *Z*-score value is above or equal to the threshold value for jth instance, then the outlier_count[j] is increased by one. And finally, the instances, which have the outlier count greater than or equal to zero, are eliminated from the dataset.

4.2 Data Balancing

SMOTE oversampling generates synthetic instances in minority class to balance the dataset. This technique involves two key steps. In the first step, it searches k samples that are closest in distance to the minority class samples. The usual value of k is 5 [7]. In the final step, new synthetic instances are generated by the following procedure:

The difference between the minority class sample variable (x_i) and its nearest neighbor (x_j) is calculated. Then, the distance between x_j and x_i is multiplied by a random value between 0 and 1 finally added to variables value of minority sample.

$$\boldsymbol{x_{new}} = \boldsymbol{x_i} + \left(\boldsymbol{x_j} - \boldsymbol{x_i}\right) \times \boldsymbol{\delta} \tag{4}$$

Here, $\boldsymbol{x_{new}}$ is the generated synthetic instance, and $\boldsymbol{\delta}$ is the random value between 0 and 1.

4.3 Feature Selection Techniques

In the next step, feature selection is carried out. The applied feature selection algorithms used in the study are briefly discussed below.

4.3.1 Chi-Square

The chi-square [10] feature selection tests the independence of two events. The occurrence of a feature or attributes is considered as the first event, and the second is the occurrence of the class. It is a nonparametric test, where the hypothesis of no association between two or more groups is tested.

$$x^2 = \frac{(O_i - E_i)^2}{E_i} \tag{6}$$

In this formula, O_i represents the occurrence of attributes, and E_i represents the occurrence of the class. These two factors are involved in the calculation for a rank.

4.3.2 Feature Importance

Feature importance refers to techniques that assign a score to each input feature based on their usefulness to predict the target variable, and based on that score, more privilege is given toward the output variable of the features with higher scores than those with lower scores. It works by using the importance score to select features that are to be kept (highest scores) and those that are to be removed (lowest scores). It decreases the dimensionality of the dataset and thus simplifies the problem that is to be modeled. To compute the relative importance for all the attributes individually, an ensemble of decision trees is used by this technique [14].

4.3.3 Relief

The relief algorithm inspired by instance-based learning [27, 28] was formulated by Kira and Rendell [29, 30]. It works by finding the conditional dependencies between attributes of a dataset. During preprocessing, the relief algorithm is viewed as a feature subset selection algorithm that selects only those features that are statistically relevant to the target class. Relief requires linear time for the number of training instances and the number of supplied features irrespective of the target concept to be learned.

To help with the selection of relevant attributes, relief tends to find the sample size and a threshold of relevancy from the given training data. A total number of triplets is collected by relief. These are its near miss instance and near hit instance. To select near miss and near hit, Euclidean distance is used. Relief also calls a routine for updating weight vectors of features for every triplet and finds an average feature weight vector that is relevant to all the features to the target concept. The features that contain the average weight which is above the given threshold are selected by relief [5].

4.3.4 Ensemble

An ensemble of feature ranking techniques is an approach, where multiple feature ranking lists are obtained from corresponding feature ranking techniques [31], and all the obtained lists, one single ranking list is generated. This process takes place in two steps. First, a set of ranking lists is obtained with help of various feature selection algorithms. Second, a function is generated from a combination of functions which transform the ranking lists in step 1 into one individual list. The second step is more important as it includes the combining method.

This can be demonstrated formally. Consider a dataset D consisting of Y attributes with X records, the initial step would be to obtain a set of X ranking lists $\{F_1, F_2, \ldots F_n\}$, and the next step would be to work out a combination method T; let f_j^i indicate the rank of feature i from ranking list j, such that the set of rankings of feature i is given by $S_i = \{f_i^1, f_i^2, \ldots f_i^n\}$. Using the combination method T, the new score obtained by feature I is

$$f_i^{'} = T\left(f_i^1, f_i^2, \ldots f_i^n\right) \tag{6}$$

4.4 Classification Algorithms

For the classification process, four widely used classifiers have been considered for evaluating the efficiency of the feature selection methods. These are K-Nearest Neighbor (KNN) [32], logistic regression [33], decision tree [34], and random forest [35]. Table 2 gives a short description of the classification algorithms based on their characteristics and parameter settings.

5 Results and Analysis

In this study, the performance of data balancing and feature selection algorithms in fault detection procedures is examined. After data cleaning, majority class instances of CM1 are reduced to 266 from 302, whereas for PC2 dataset the majority instances become 1242 from 1569. The parameters are selected by default in all the implemented algorithms except for SMOTE and feature selection algorithms. In SMOTE, the percentage parameter for CM1 was determined to be 532%, and for PC2, it was 7662%. The percentage parameter says how many synthetic instances are created based on the number of the class with less instances. Moreover, the parameter numToSelect of the feature selection algorithms was set to 20. This resulted in 20 most significant features to be selected from each dataset.

All the experiments have been carried out by applying ten-fold cross-validation. This is to decrease the variability of the performance results due to the random generation of train and test sets. Weka 3.8.3 [36] was used on a desktop PC with 3.6 GHz Intel Core i7-4790 processor and 16 GB RAM to perform the experiments.

Table 2 Machine learning algorithms with their description and parameter settings

Classifiers	Description	Parameter settings
KNN	An instance-based classification algorithm	$K = 5$, NNSearch = LinearNNSearch
Decision tree	Tree-based classification technique	Confidence factor = 0.25
Random forest	Ensemble-based learning method for classification	nEstimators = 100
Logistic regression	Function-based classification technique	Kernel = PolyKernel; lambda = 0.01

Table 3 Performance results based on SMOTE after feature selection (20 features)

Dataset	Classifiers	Performance					
		Accuracy		TPR		TNR	
		Before balance (%)	After balance (%)	Before balance	After balance	Before balance	After balance
CM1	KNN	83.721	88.329	0.286	0.943	0.887	0.823
	Decision tree	84.593	85.687	0.262	0.879	0.927	0.835
	Random forest	86.046	91.713	0.024	0.925	0.977	0.910
	Logistic regression	86.628	79.473	0.214	0.785	0.957	0.805
PC2	KNN	97.918	98.591	0.063	0.998	0.989	0.974
	Decision tree	98.864	98.873	0.00	0.990	0.999	0.987
	Random forest	98.927	99.355	0.00	**0.**990	0.999	0.990
	Logistic regression	98.864	94.645	0.188	0.965	0.997	0.928

In Tables 3 and 4, the experimental results for the evaluated classification algorithms are tabulated in terms of accuracy, TPR, and TNR. These metrics are calculated according to Eqs. (7)–(9). To calculate these matrices, four important counts are collected from the confusion matrix: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Accuracy refers to the percentage of correctly classified defects.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \tag{7}$$

True Positive Rate (TPR) or sensitivity is the rate of actual positives that are correctly identified.

$$\text{TPR} = \frac{TP}{(TP + FN)} \tag{8}$$

True Negative Rate (TPR) or specificity is the rate of actual negatives that are correctly identified.

$$\text{TNR} = \frac{TN}{(FP + TN)} \tag{9}$$

In the classifiers, defects are considered as positive class, and non-defects are considered as negative class.

Table 4 Performance results based on feature selection (20 features) after applying SMOTE

Dataset	Classifiers	Performance					
		Accuracy		TPR		TNR	
		Before feature selection (%)	After feature selection (%)	Before feature selection	After feature selection	Before feature selection	After feature selection
CM1	KNN	89.266	88.323	0.947	0.932	0.838	0.835
	Decision tree	88.89	88.135	0.868	0.894	0.910	0.868
	Random forest	93.408	93.032	0.947	0.943	0.921	0.917
	Logistic regression	78.907	81.167	0.785	0.826	0.793	0.797
PC2	KNN	98.671	97.918	0.998	0.998	0.976	0.970
	Decision tree	98.752	98.832	0.990	0.986	0.986	0.991
	Random forest	99.396	99.519	0.997	0.997	0.991	0.994
	Logistic regression	95.330	95.491	0.969	0.977	0.938	0.933

Due to the limited scope of the paper, some results have been omitted from the presentation.

5.1 *Comparison With no Balancing*

As explained in Sect. 1, the initial datasets reflect a skewed class distribution. Every machine learning algorithm encounters a significant challenge when working with these higher number of defective classes and its high imbalanced instances. For the performance evaluation of our proposed approach, experiments have been carried out, with and without data balancing, among two preprocessed sets with feature importance feature selection.

Table 3 summarizes the performance of the SMOTE algorithm based on the four classifiers. It is indicated that the TPR of CM1 and PC2 dataset increased significantly. For both CM1 and PC2, the random forest algorithm is the most successful algorithm with a TPR increase of approx. 0.9 while decision tree and KNN being the second most efficient. This is followed by other algorithms applied to both the datasets. Similarly, it can be observed from Fig. 3 that after performing SMOTE, the TPR increase of approx. 0.99 can be obtained for PC2 dataset while applying a different feature selection algorithm. However, the accuracy of PC2 and CM1 shows a small change before and after SMOTE is applied. The variation in TNR is also very slight

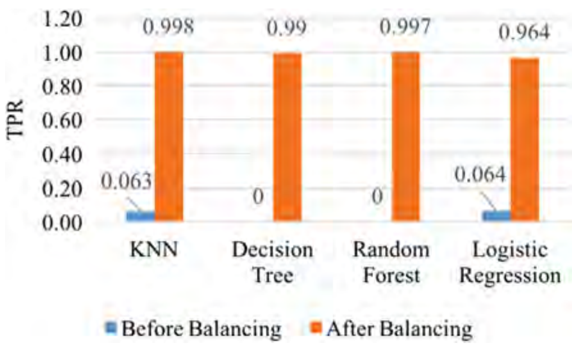


Fig. 3 Comparison of performance based on the selected attributes by ensemble feature selection of PC2 dataset before and after applying the SMOTE technique

and therefore negligible. This does not affect the goal of this paper as the focus was to predict faulty software correctly. Identifying all data points as non-faulty in the software fault detection problem is not helpful and instead it is important that we concentrate on identifying the positive class. The metric our knowledge tells us to maximize, TPR or the ability of a model to find all the relevant cases within a dataset.

5.2 *Comparison with no Feature Selection*

A comparison between before and after applying feature selection algorithms has been made to evaluate the performance of the proposed SDP. As a result of the selection of relevant features by the chosen attribute selection methods, the values of the metrics have increased slightly.

Table 4 presents the accuracy, True Positive Rate (TPR), and True Negative Rate (TNR) of the classifiers before and after the relief algorithm has been applied.

The results show that the performance of the classifiers is nearly similar before and after feature selection. This is also further proven in Fig. 4, where identical experiments have been carried out but with ensemble feature selection.

5.3 *Comparison with Other Feature Selection Methods*

To further evaluate our approach, a comparison between all the feature selection techniques are executed based on the CM1 and PC2 datasets and the random forest classification technique.

Although it is a common practice to apply the concept of ensemble learning to classification, other fields of machine learning, such as feature selection can also be improved by using it [37]. Machine learning methods alone can no longer deal with datasets efficiently as they increase in dimensionality and get more complex.

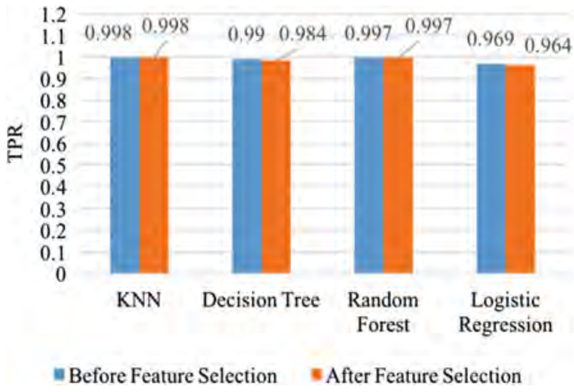


Fig. 4 Comparison of performance based on balanced PC2 dataset before and after applying ensemble feature selection

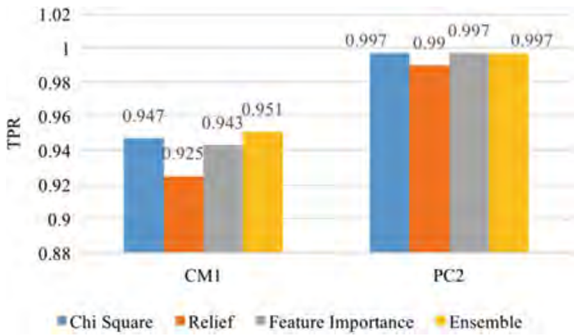


Fig. 5 Comparison performance of different feature selection techniques based on balanced datasets

As shown in Fig. 5, ensemble feature selection achieves the highest TPR on both CM1 and PC2. Therefore, in terms of performance and efficiency, it has superiority over other feature selection algorithms.

6 Conclusion and Future Scope

In automated software fault detection, balanced data which results in high accuracy, high TPR, and a high TNR is an important requirement. However, in real-life situations, most datasets are highly imbalanced and complex. The effect of low TPR and TNR can grow if not detected early in the software cycles. Hence, this issue must be addressed. In addition to balancing data, dimensionality reduction by choosing useful features can improve the efficiency of the classifiers and decrease training time.

This article shows the effect of oversampling and feature selection techniques on the results of four different machine learning algorithms when dealing with high-dimensional and imbalanced data. First, data cleaning is applied to minimize noise generation. Second, an oversampling technique, SMOTE is applied to both CM1 and PC2. This is followed by the application of four different feature selection techniques on each algorithm. The experimental results reveal that even though feature selection on its own has some effect albeit minimal but after combining oversampling and feature selection, the performance of the classifiers has improved significantly. Such results confirm the strength and usefulness of the proposed approach. Our approach outperforms other approaches that take only feature selection into account. Most importantly, because of initial data cleaning in the proposed method, can be balanced and apply any feature selection on noise-free datasets.

An apparent limitation of this study is that it examines only two datasets, but for better analysis of the performance in terms of consistency, the suggested method should be evaluated with other datasets. Moreover, only a small group of filter-based feature selection techniques and SMOTE oversampling method was considered. Filter-based feature selection methods ignore the dependencies between the features, and they also dismiss interaction with the classifiers.

This study considers only a small group of feature selection techniques and an oversampling method. In future work, investigating filter and wrapper-based feature selection techniques along with a combination of undersampling and oversampling methods might prove important. Also, instead of considering only a fixed number of features (in our case, 20), conducting experiments with a varying number of features could be beneficial.

References

1. Kaur R, Sharma ES (2018) Various techniques to detect and predict faults in software system: survey. Int J Futur Revolut Comput Sci Commun Eng (IJFRSCE) 4(2):330–336
2. Ali MM, Huda S, Abawajy J, Alyahya S, Al-Dossari H, Yearwood J (2017) A parallel framework for software defect detection and metric selection on cloud computing. Cluster Comput 20(3):2267–2281. https://doi.org/10.1007/s10586-017-0892-6
3. Ben-Gal I (2005) Outlier detection
4. Chandrashekar G, Sahin F (2014) A survey on feature selection methods. Comput Electr Eng 40(1):16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024
5. Vege SH (2012) Ensemble of feature selection techniques for high dimensional data
6. Kamei Y, Monden A, Matsumoto S, Kakimoto T, Matsumoto K (2007) The effects of over and under sampling on fault-prone module detection
7. Ramezankhani A, Pournik O, Shahrabi J, Azizi F, Hadaegh F, Khalili D (2016) The impact of oversampling with SMOTE on the performance of 3 classifiers in prediction of type 2 diabetes. Med Decis Mak 36(1):137–144. https://doi.org/10.1177/0272989X14560647
8. Pelayo L, Dick S (2007) Applying novel resampling strategies to software defect prediction. In: Annual conference of the North American fuzzy information processing society—NAFIPS, pp 69–72. doi: https://doi.org/10.1109/NAFIPS.2007.383813
9. Huda S et al. (2018) An ensemble oversampling model for class imbalance problem in software defect prediction. In: IEEE Access 6:24184–24195. doi: https://doi.org/10.1109/ACCESS.2018.2817572
10. Zhou L, Li R, Zhang S, Wang H (2018) Imbalanced data processing model for software defect prediction. Wirel Pers Commun 102(2):937–950. https://doi.org/10.1007/s11277-017-5117-z

11. Agarwal S, Tomar D (2014) A feature selection based model for software defect prediction. Int J Adv Sci Technol 65:39–58. https://doi.org/10.14257/ijast.2014.65.04
12. Xu Z, Liu J, Yang Z, An G, Jia X (2016) The impact of feature selection on defect prediction performance: an empirical comparison. In: Proceedings of ISSRE '96: 7th international symposium on software reliability engineering, ISSRE, pp 309–320. doi: https://doi.org/10.1109/ISSRE.2016.13
13. Shivaji S, James Whitehead E, Akella R, Kim S (2013) Reducing features to improve code change-based bug prediction. IEEE Trans Softw Eng 39(4):552–569. doi: https://doi.org/10.1109/TSE.2012.43
14. Tasnim Cynthia S, Rasul MG, Ripon S (2019) Effect of feature selection in software fault detection BT—multi-disciplinary trends in artificial intelligence, 2019, pp 52–63
15. Bolón-Canedo V, Sánchez-Maroño N, Alonso-Betanzos A (2016) Feature selection for high-dimensional data. Prog Artif Intell 5(2):65–75. https://doi.org/10.1007/s13748-015-0080-y
16. Jakhar AK, Rajnish K (2018) Software fault prediction with data mining techniques by using feature selection based models. Int J Electr Eng Inform 10(3):447–465. doi: https://doi.org/10.15676/ijeei.2018.10.3.3
17. Balogun AO, Basri S, Abdulkadir SJ, Hashim AS (2019) Performance analysis of feature selection methods in software defect prediction: a search method approach. Appl Sci 9(13). doi: https://doi.org/10.3390/app9132764
18. Sharmin S, SAL: an effective method for software defect prediction, pp 184–189
19. Ibrahim DR (2017) Software defect prediction using feature selection and random forest algorithm, pp 252–257. doi: https://doi.org/10.1109/ICTCS.2017.39
20. Zengin A (2016) HSDD : a hybrid sampling strategy for class imbalance in defect prediction data sets
21. Karatas G (2020) Increasing the performance of machine learning-based IDSs on an imbalanced and up-to-date dataset. IEEE Access 8:32150–32162. https://doi.org/10.1109/ACCESS.2020.2973219
22. Alsawalqah H, Faris H, Aljarah I, Alnemer L. Hybrid SMOTE-ensemble approach Adv Intell Syst Comput 1. doi: https://doi.org/10.1007/978-3-319-57141-6
23. Liu S, Zhang J, Wang Y, Xiang Y (2016) Fuzzy-based feature and instance recovery, pp 605–615. doi: https://doi.org/10.1007/978-3-662-49381-6
24. Abdou AS (2018) Early prediction of software defect using ensemble learning : a comparative early prediction of software defect using ensemble learning : a comparative study. doi: https://doi.org/10.5120/ijca2018917185
25. Gray D, Bowes D, Davey N, Sun Y, Christianson B (2012) Reflections on the NASA MDP data sets. IET Softw 6(6):549–558
26. Kannan KS, Manoj K, Arumugam S (2015) Labeling methods for identifying outliers. Int J Stat Syst
27. Aha DW, Kibler D, Albert MK (1991) Instance-based learning algorithms. Mach Learn 6(1):37–66
28. Callan JP, Fawcett T, Rissland EL (1991) CABOT: an adaptive approach to case-based search. IJCAI 1991(12):803–808
29. Kira K, Rendell LA (1992) A practical approach to feature selection. In: Sleeman D et al (eds) Morgan Kaufmann, San Francisco (CA), pp 249–256
30. Kira K, Rendell LA (1992) The feature selection problem: Traditional methods and a new algorithm. AAAI 2:129–134
31. Rahman A, Verma B (2013) Ensemble classifier generation using non-uniform layered clustering and Genetic Algorithm. Knowl-Based Syst 43:30–42. https://doi.org/10.1016/j.knosys.2013.01.002
32. Yihua Liao VR, Vemuri (2002) Use of K-nearest neighbor classifier for intrusion detection 21(5):439–448
33. Peng C-YJ, Lee KL, Ingersoll GM (2002) An introduction to logistic regression analysis and reporting. J Educ Res 96(1):3–14. https://doi.org/10.1080/00220670209598786

34. Quinlan JR (1986) Induction of decision trees. Mach Learn 1(1):81–106. https://doi.org/10.1007/BF00116251
35. Fawagreh K, Gaber MM, Elyan E (2014) Random forests: from early developments to recent advancements. Syst Sci Control Eng 2(1):602–609. https://doi.org/10.1080/21642583.2014.956265
36. Witten IH, Frank E (2002) Data mining: practical machine learning tools and techniques with java implementations. SIGMOD Rec 31(1):76–77. https://doi.org/10.1145/507338.507355
37. Pes B (2020) Ensemble feature selection for high-dimensional data: a stability analysis across multiple domains. Neural Comput Appl 32(10):5951–5973. https://doi.org/10.1007/s00521-019-04082-3

Indoor Navigation Assistant for Visually Impaired (INAVI)

U. B. Mahadevaswamy, D. Aashritha, Nikhil S. Joshi, K. N. Naina Gowda, and M. N. Syed Asif

Abstract Navigating in unknown places can be challenging for people who are deprived of the benefit of sight. The focus of advancements in navigation systems is extended onto helping the visually impaired understand the structure of their surroundings whilst they are traveling. The main idea behind the proposed method is to eliminate the dependency of the visually impaired on unreliable sources in an unfamiliar locality. Most of the public buildings these days, such as college and office buildings are equipped with their own Wi-Fi network which is used by the proposed indoor navigation assistant to direct the user while navigating inside the building. Using the Received Signal Strength Indicator (RSSI) values that are taken from each access point, trilateration is performed for localization, and the speech output guides the person by informing about the current location and thus making it possible for the visually impaired to move inside a building without any human assistance.

Keywords Wireless Fidelity (Wi-Fi) · Received Signal Strength Indicator (RSSI) · NodeMCU

U. B. Mahadevaswamy · D. Aashritha (✉) · N. S. Joshi · K. N. Naina Gowda · M. N. Syed Asif
Department of Electronics and Communication, JSS Science and Technology University, Mysuru, India
e-mail: aashutrisha@gmail.com

U. B. Mahadevaswamy
e-mail: mahadevaswamy@sjce.ac.in

N. S. Joshi
e-mail: nikhilsj98@gmail.com

K. N. Naina Gowda
e-mail: naiiinagowda@gmail.com

M. N. Syed Asif
e-mail: syedasifm.n@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_17

1 Introduction

Humans have been exploring and navigating around the world over centuries. Navigation systems, over the years, have evolved from maps and compasses to global positioning systems with detailed user interfaces, portraying the locations and routes with constantly improving accuracy. The realistic experience offered by current navigation systems adds on to the assistance provided by the same for finding places and directions to the same.

Even though these systems are very useful on road, they fall behind in providing the same facilities inside the four walls of a building. These systems are designed and modeled in such a way that they guide the user outdoors in unfamiliar places. However, this modeling has not been extended to indoors. The need for guidance inside unfamiliar buildings is tantamount to the requirement of directions while traveling around in a new city.

Assistance in moving around the corridors of a building is of most importance to the visually impaired people. The people deprived of the benefit of sight struggle to move around in an unfamiliar environment, especially without any guidance from others. An indoor navigation assistant can make their movement inside the building a lot comfortable by guiding them through the building's ways.

An indoor navigation system is a concept which is designed to assist the indoor movement to the user. The idea behind the indoor navigation assistant is to locate the user and guide through the ways of a building to reach the desired destination.

2 Related Work

Numerous research and implementations have been carried out on developing an efficient indoor positioning system that can be used worldwide. Several journal papers and research publications were considered for a better understanding of the field of this venture. The closest approaches have been studied in detail, and some of them are elaborated further.

In [1], Sunmin Lee, Jinah Kim, and Nammee Moon have proposed a smart watch which basically acts as a Wi-Fi-based indoor location recognition. The problem of position recognition due to the similar signal strength is solved by using both the Received Signal Strength Indication (RSSI) and Basic Service Set Identifier (BSSID).

The authors of [2] present an algorithm that a mobile device can utilize, as GPS-like reference nodes, either in range location-aware compatible mobile devices or pre-installed low-cost infrastructure-less location-aware beacon nodes.

The model proposed in [3, 4] uses neural networks to train the distance model. The authors propose an RSSI real-time correction method which is based on Bluetooth gateway.

In [5], the authors have built a Long Short-Term Memory (LSTM) recurrent neural network. This network makes regression between fingerprints and the locations in order to track the moving target.

Furthermore, indoor positioning has been an area of constant development with technologies ranging from IR to BLE included. In that view [6–9] discuss indoor localization using Bluetooth Low Energy (BLE) along with other additional protocols and techniques. The BLE beacons are deployed in different locations, and the different RSSI techniques are used for identifying the position of a user and mobile device.

A new technique of self-localization using infrared sensors is proposed [10]. The authors have come up with a configuration which consists of an IR LED array equipped with unique ID encoding capabilities which are based on a combination of different frequencies, and the repeated use of each ID encoding LED is done to address the issues of limited frequencies.

In recent times, Wi-Fi technology is preferred over other methods. The methods in [11, 12] provide some insight into the usage of Wi-Fi for indoor localization. Due to its wide deployment, Wi-Fi is expected to become a prominent tool for indoor positioning.

Along with this, Wi-Fi fingerprints can be further explored to identify human activities and locations. These conspicuous benefits of using Wi-Fi finger printing are well discussed in [13, 14]. Furthermore, [15, 16] provide details regarding the recent researches carried out in indoor positioning and the results of the same.

The existing indoor navigation methods use the maps and blueprints of the campus or buildings to guide the visitors. Some work has been done in developing various methods to determine the location of the visitors. An eclectic range of techniques have been discussed and proposed, explaining different methods of determining the position of the visitor by using the RSSI from the BLE beacons and Wi-Fi signals. Most of the works are limited to finding the position of the visitor.

Even though the ideas and experiments carried out so far have provided results to some level, they have either been conducted in a limited space area or have their own limitations with respect to the technology and approach. Thus, these previous works provide a good platform for the development of the existing methods.

3 Proposed Work

The system proposed intends to cover upon the area of indoor navigation. Further, specifically, under indoor navigation, the approach focuses on developing indoor navigation assistance for the visually impaired people.

In recent times, it can be seen that wireless technology has taken over the market completely. The indoor navigation assistant takes advantage of the available Wi-Fi network of a building to carry out the localization of the user, thus providing a novel facility by involving existing technology.

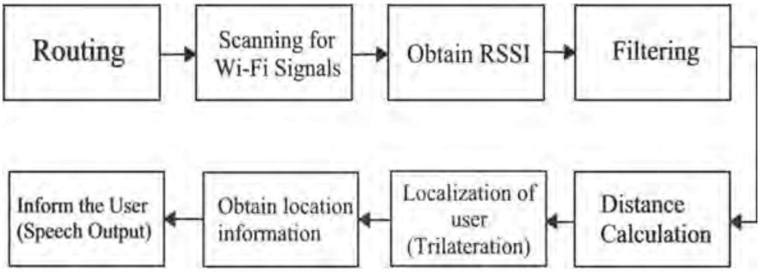


Fig. 1 Proposed block diagram

Among the various algorithms used for localization, the most popular method is the Received Signal Strength Indicator (RSSI)-based algorithm. Received Signal Strength Indication (RSSI) is an indicator of the power level that the receiver sensors receive from access points. It is measured in terms of decibels from 0 (zero) to −120 (minus 120).

The implemented method utilizes Wi-Fi technology and trilateration techniques to determine the position of the user. Each user is provided with an ESP8266 device that is programmed to scan for the Wi-Fi signals and obtain their respective RSSI strength. Three Wi-Fi signals with the best RSSI are used to perform trilateration and obtain the distance of the user from the access points, thus obtaining the location of the user.

Once the location is obtained, the pre-fed database of the architecture of the building is searched, and the physical location details of the user are obtained. After this, the next important step is to inform the user about it. This is then achieved with the help of the speaker connected for audio output. In a new indoor environment, the visually impaired persons can survive independently without any human assistance. Figure 1 shows the block diagram which represents the working of the implemented system. The functionality of each block is as follows:

1. **Scanning for Wi-Fi signals**: The NodeMCU ESP8266 device is used to scan for the Wi-Fi signals from nearby access points.
2. **Obtain RSSI**: The Received Signal Strength (RSSI) of individual signals is obtained. The nearest three access points are chosen based on the strongest RSSI.
3. **Filtering**: The signals from the routers might undergo multi-path fading causing fluctuations in the RSSI. These fluctuations are controlled by the implementation of a Kalman filter.
4. **Distance calculation**: Once the RSSI is obtained, the distance of the device from the access points is calculated.
5. **Obtain (x, y) coordinates**: Coordinates of the three chosen access points are obtained from the pre-defined database.
6. **Localization of user**: Trilateration algorithm is used, and the device's location is determined.
7. **Obtain location information**: With the determined location, the information regarding that location is obtained from the pre-defined database.

8. **Inform the user**: The information fetched from the database is communicated to the user; thus, the user is aware of his\her surroundings using speech output.

The proposed method utilizes the Wi-Fi technology and trilateration techniques in order to determine the position of the user. Each user is provided with a NodeMCU ESP8266 device that is programmed to scan for the Wi-Fi signals. The device obtains the Received Signal Strength Indication (RSSI) from at least three Wi-Fi access points. RSSI is basically the measurement of power that is present in a received radio signal.

In case, if the device receives signals from more than three access points, then the top three signals with higher RSSI values are considered. To avoid the fluctuations in RSSI from affecting the calculations, a Kalman filter is used before the RSSIs which are considered for trilateration. The response time is affected by RSSI if the device fails to receive any Wi-Fi signals, thereby continuing its search for the signals in a loop. Using the RSSI, the distance of the device from the access point is determined using the formula:

$$\mathrm{RSSI(dbm)} = -10\,\mathrm{m}\ \log(d) + R \tag{1}$$

where

R = signal strength of received signal in dBm.

d = distance between the device and the access point.

m = propagation constant or path loss (for free space: $m = 2$).

The NodeMCU ESP8266 device computes the distance "d" from the mentioned formula using three different signal strengths. To determine the position, trilateration algorithm is used. Trilateration is a technique used to determine the coordinates of a point using the distance of that point from three known locations. Using three distances, trilateration can pinpoint the exact location.

In Fig. 2, the signals coverage from the access points forms three circles that intersect each other at different points. When the device is in the intersection area, using the distances calculated before the location of the device can be determined by solving the system of linear equations given below. The following five steps are conducted to calculate (x, y) coordinates:

Step 1: The three respective equations for each of the three circles are as follows:

$$(a - x_1)^2 + (b - y_1)^2 = d_1^2 \tag{2}$$

$$(a - x_2)^2 + (b - y_2)^2 = d_2^2 \tag{3}$$

$$(a - x_3)^2 + (b - y_3)^2 = d_3^2 \tag{4}$$

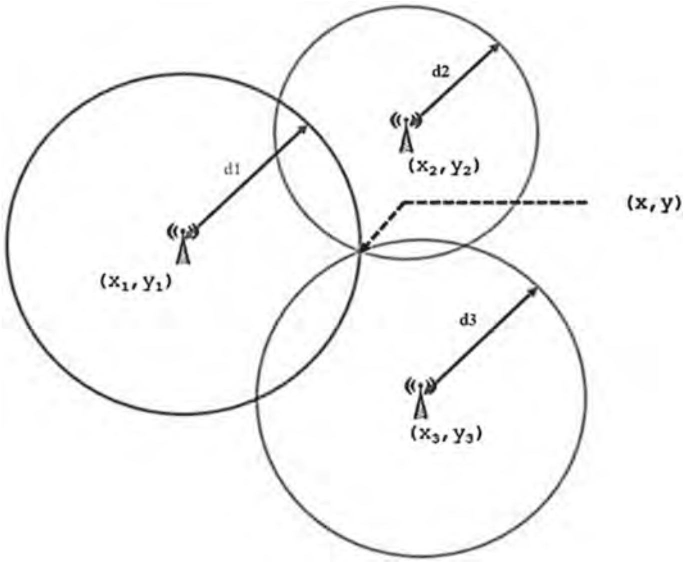


Fig. 2 Position coordinates detection

where

$a = x$ coordinate of the user.

$b = y$ coordinate of the user.

$x_1 = x$ coordinate of the first access point.

$x_2 = x$ coordinate of the second access point.

$x_3 = x$ coordinate of the third access point.

$y_1 = y$ coordinate of the first access point.

$y_2 = y$ coordinate of the second access point.

$y_3 = y$ coordinate of the third access point.

$d_1 =$ distance between the user and the first access point.

$d_2 =$ distance between the user and the second access point.

$d_3 =$ distance between the user and the third access point.

Step 2: Expanding out the squares in each of the above three equations

$$a^2 - 2x_1a + x_1^2 + b^2 - 2y_1b + y_1^2 = d_1^2 \tag{5}$$

$$a^2 - 2x_2a + x_2^2 + b^2 - 2y_2b + y_2^2 = d_2^2 \tag{6}$$

$$a^2 - 2x_3a + x_3^2 + b^2 - 2y_3b + y_3^2 = d_3^2 \tag{7}$$

Step 3: Subtracting Eq. 6 from Eq. 5:

$$(-2x_1 + 2x_2)a + (-2y_1 + 2y_2)b = d_1^2 - d_2^2 - x_1^2 + x_2^2 - y_1^2 + y_2^2 \tag{8}$$

Similarly, subtracting Eq. 7 from Eq. 6:

$$(-2x_2 + 2x_3)a + (-2y_2 + 2y_3)b = d_2^2 - d_3^2 - x_2^2 + x_3^2 - y_2^2 + y_3^2 \tag{9}$$

Step 4: Rewriting Eqs. 8 and 9 from step 3 using M, N, O, P, Q, R values:

$$Ma + Nb = O \tag{10}$$

$$Pa + Qb = R \tag{11}$$

Step 5: The solution of this system is:

$$a = \frac{OQ - RN}{QM - NP} \tag{12}$$

$$b = \frac{OP - MR}{NP - MQ} \tag{13}$$

The coordinates of the checkpoints including important laboratories, seminar halls, staff rooms, and staircases in the floor plans are determined by the implemented system and stored in the database by the campus admin. After the coordinates of the user are acquired, they are compared to that of the checkpoints in the database, thus obtaining the location of the user. The next important step is to convey this location information to the user.

The ESP8266 audio library comes in hand while generating speech output from the text with the support of ESP8266SAM library. Using a single transistor amplifier circuit with a speaker at the end, the speech output can be obtained. The transistor emitter is grounded, and the transistor base is driven by the ESP8266-I2SOUT (Rx) pin. The collector terminal and the USB 5 V supply are connected to negative and positive terminals of the speaker as shown in Fig. 3.

Even the 3 V from ESP8266 can provide the power but the volume is comparatively low. Since the ESP8266 pins cannot provide the necessary current, the amplifier is essential. Without that, there is a possibility of the device being damaged. Using an object of the audio output I2SNoDAC class, the speech output is produced from the text in the program. The location names are previously stored in the program which is converted into speech when the user is near that location.

```
                          2N3904 (NPN)
                          +---------+
                          |         |     +-|
                          | E  B  C |    / S|
                          +-|--|--|-+    | P|
                            |  |  +-------+ E|
                            |  |         | A|
ESP8266-GND ----------------+  |  +-------+ K|
                               |  |      | E|
ESP8266-I2SOUT (Rx) -----/\/\/\--+  |      \ R|
                                  |       +-|
USB 5V ---------------------------+
```

Fig. 3 Connection between the ESP8266 module and the speaker

4 Result Analysis

The proposed system was implemented and was tested for its functionality inside the college building. Multiple trials were carried out to check the working of user localization. The testing was carried out by navigating from a source location to a destination location inside the building.

For testing, the system is designed to guide the user while navigating within the three floors of the college building. The architectural structures of each of these floors are depicted using Figs. 4, 5, and 6.

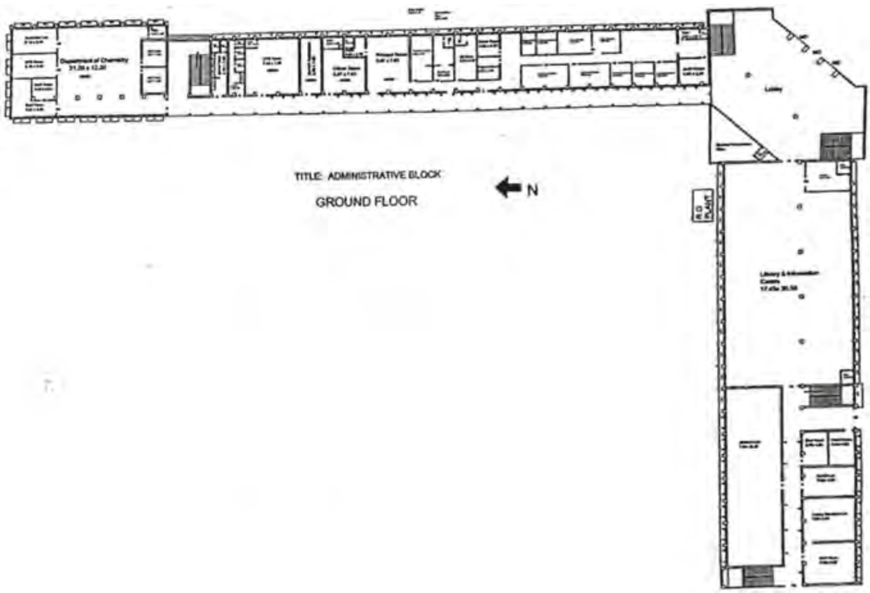


Fig. 4 Schematic of the ground floor

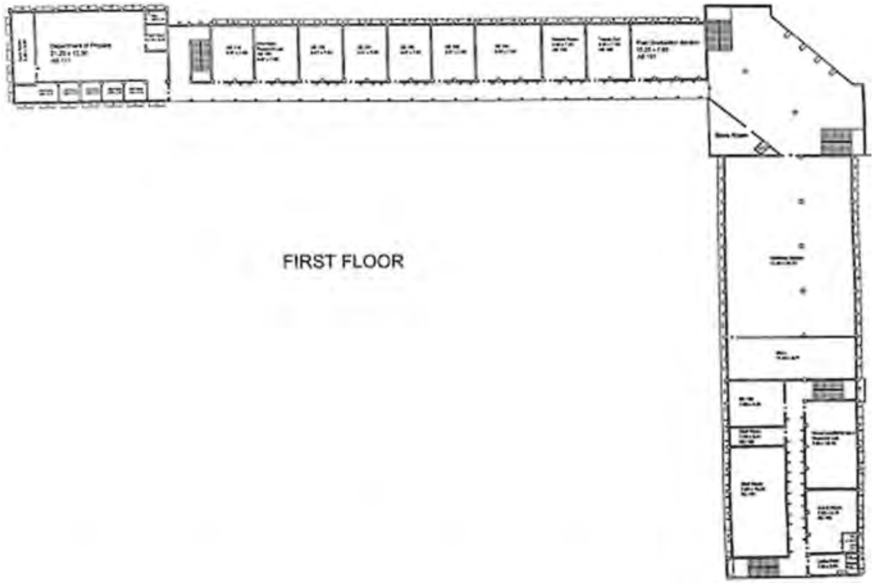


Fig. 5 Schematic of the first floor

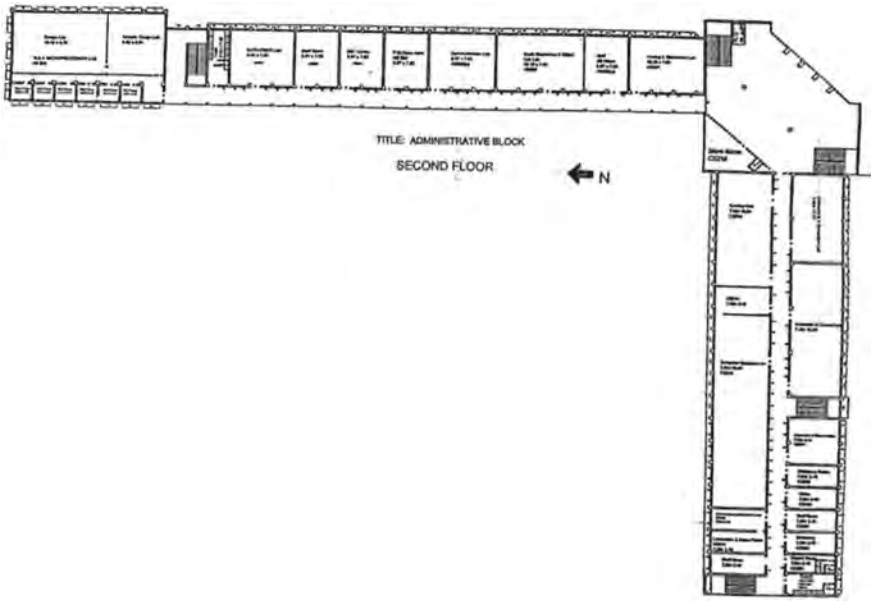


Fig. 6 Schematic of the second floor

With the inclusion of knowledge of staircases interconnecting all three floors, the system is equipped to work in the tortuous structure of the building. The system was tested for its accuracy in delivering correct location information, and the time taken to intimate the user was also noted over multiple trials.

Once the NodeMCU determines the user's location, it informs the user with a speech output using the ESP8266SAM library. The user becomes aware of his/her location once he/she reaches the checkpoints. Figure 7 shows a snippet of guiding directions that are conveyed to the user. However, in real time, the user is directed by similar speech commands.

The NodeMCU ESP8266 device functions quite effectively while dealing with this amount of data. The delay obtained while using the system is not very significant and is found to be under tolerable limits. It was observed that in real time, the device was successful in identifying the checkpoints inside the building (rooms or halls inside the building) as per the floor plans. Table 1 provides the time taken by the device to inform the user about his/her location over multiple trials.

Furthermore, the system's efficiency is analyzed by the accuracy with which the user's location is determined and conveyed to the user. At every stage of database

```
You are near Cisco Lab

Take ten steps to reach Phillips lab
You are near Phillips Lab

Take ten steps to reach Robotics lab
You are near Robotics Lab

Take twenty steps to reach Communication lab
You are near Communication Lab

You have reached your destination.
```

Fig. 7 Output commands from INAVI device

Table 1 Time taken to inform user

Trial no.	Database strength (No. of checkpoints)	Response time (seconds)
1	10	2.193
2	20	2.197
3	30	2.198
4	40	2.198
5	48	2.199
Average response time (seconds)		2.197

strength, 25 trials were done, and the trials, where the location was correctly identified, were considered a success. The trials wherein the location identification was either delayed or erroneous amounted to the error.

The error in measurement is calculated using equation

$$\%\text{Error} = \frac{|\text{Actual value observed} - \text{Expected value}|}{\text{Expected value}} \times 100 \tag{14}$$

$$\%\ \text{Accuracy} = 100 - \%\ \text{Error} \tag{15}$$

Using Eqs. 14 and 15, the accuracy at each database strength level was calculated, and in the final stage when all the checkpoints were included in the database, the system showed a location recognition accuracy of 96%. Even though the effect of varying database strength is negligible on the response time, an improvement in the accuracy can be seen. With the increase in the number of checkpoints, it was observed that an average localization accuracy of 1.5 m was achieved.

Wi-Fi signals are known to undergo fading, causing fluctuations in the RSSI. Therefore, to achieve accurate localization of the user not only the RSSI was used in the proposed method, but also the BSSIDs were utilized. The plot of the accuracy of the device, when tested over varying database strength, is shown in Fig. 8.

The implemented system was compared with that of the other systems implemented, and the findings are indicated in Table 2. The key difference when it comes to the system implementation and the complexity of the previously implemented systems and the drawbacks found in them are observed.

The system is found to be efficacious in dynamically tracking the user and identifying the user's position inside the building and guide him/her with further directions with voice output. This way it can keep the visually impaired user informed about

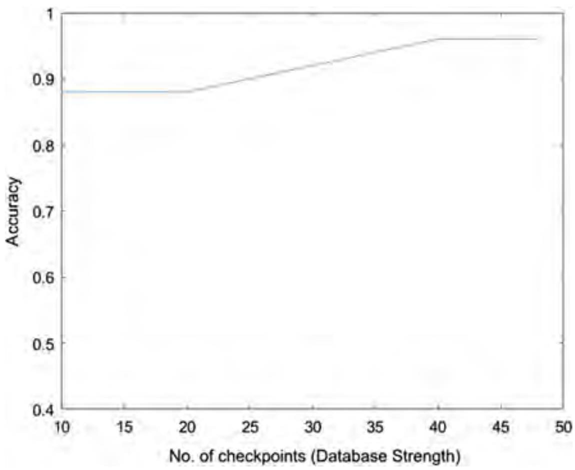


Fig. 8 Plot of accuracy with respect to the database strength

Table 2 Comparison with existing work

S. no.	Approach	Area of implementation	Key difference	System complexity	Result interpretation	Accuracy
1	RSSI random forest BSSID [1]	The system was implemented in open space (no walls in the surrounding)	Loss of signal strength and interference from other sources was not considered	Highly complex system due to the implementation of server and client approach for computations	Execution speed and accuracy results of the experiment are presented	Localization accuracy of 97.5%
2	RSSI DALIS [2]	The testing was conducted in an open space indoor environment with reference nodes deployed at three corners of area 4 × 4 m^2	The device was incapable of picking up low-level Wi-Fi signals	Simple setup but results in lagged consistency due to non uniform radiation patterns	The proposed method's architecture, scenario, localization, and accuracy for <3 m are compared with other methods	Localization accuracy is 97%
3	RSSI PSO-BPNN distance model BLE [3]	The implemented system did not consider the complex electromagnetic locations and the corners of the room	Noise affects the connection between the beacon and the smartphone	The involvement of BLE technology complicates the process of obtaining continuous positions with high precision	Comparison of positioning RMSE, MAE, and maximum error is done between the traditional and proposed method	Positioning error of 1.61 m No mention of accuracy percentage
4	Long short-term memory recurrent neuron network [5]	MATLAB and Spyder simulations with localization areas limited to 4 × 4m^2	Only simulation results	Complex requires GPUs for fingerprints preparation	A comparison between LSTM, RNN, and BP is presented for training and testing time. Mean error and centralized time for different lengths in different trajectories are also presented	The mean error of different neural networks in different trajectories is plotted

(continued)

Table 2 (continued)

S. no.	Approach	Area of implementation	Key difference	System complexity	Result interpretation	Accuracy
5	INAVI	The system was implemented in a three story building	The system accesses the low-level Wi-Fi signals. Interference factors are also considered	Simpler, as computations are carried out only using NodeMCU ESP8266	Time is taken to inform the user, and the accuracy with respect to varying database strength is presented	Localization accuracy of 96% (an average accuracy of 1.5 m)

his/her position near the important landmarks of the building, thereby eliminating the dependency of the user on any other person.

5 Novelty

The proposed indoor navigation assistant has been developed with the main objective of helping the visually impaired inside the building. The proposed work not only locates the visitor but also intimates him/her with the information regarding his/her surroundings. Unlike the existing applications, the proposed idea does not demand the user to have the map of the building since it uses the pre-defined database with important locations of the building. The method presented here employs the trilateration technique instead of triangulation and thus does not require the orientation of the device and access points. The proposed idea uses minimal hardware that includes ESP8266 and a speaker prominently. Most of the existing works limit themselves to indoor localization whereas the idea proposed here is an extension of the former, and the visitor is informed about the important locations such as rooms, cabins, and halls as he/she wanders inside the building. The proposed system can be developed with the deployment of undemanding technologies whilst providing efficacious results and serving the visually impaired.

6 Conclusion and Future Scope

The proposed system is designed and implemented such that it tracks the user dynamically and updates the user about his/her current position. The system can inform the user about his location once he reaches near the designated checkpoints (rooms or halls inside the building). This will make it possible for the user to be independent while navigating in an unfamiliar environment. This leads to eradicating the need of human assistance to the visually impaired person.

However, the system has the limitation that the device has to receive at least three signals since three coordinates are a requirement for trilateration. Therefore, the system might be incompetent in coverage of fewer areas. This can be remedied by relocating the Wi-Fi routers to avoid areas without reception. The system performance can be enhanced by the additional feature of speech interaction, where the user can enter the destination location using speech input. A more dynamic approach in tracking and assisting the user can be accomplished with the inclusion of guidance to prevent obstacles.

The system implemented is such that the speech output is available in the English language as a default. But if developed further, then the system can be customized to provide the speech output to the user in the local languages as preferred by the user.

Moreover, the response time of the system can be improved by using a processor with high processing speed so that the delay in localization can be minimized. The further improvised systems can be implemented using Bluetooth low-energy devices, which will further enable a reduction in power consumption when implemented on a larger scale.

References

1. Lee S, Kim J, Moon N (2019) Random Forest and WiFi fingerprint-based indoor location recognition system using smart watch. Human-centric Comput Inf Sci 9(6)
2. Awad F, Al-Sadi A, Al-Quran F, Alsmady A (2018) Distributed and adaptive location identification system for mobile devices. EURASIP J Adv Signal Process 61
3. Li G, Geng E, Ye Z, Xu Y, Lin J, Pang Y (2018) Indoor Positioning Algorithm based on the improved RSSI distance model. Sens J 18(9):2820
4. Robesaat J, Zhang P, Abdelaal M, Theel O (2017) An improved BLE indoor localization with Kalman-based fusion: an experimental study. Sens J 17(5):951
5. Xu B, Zhu X, Zhu H (2019) An efficient indoor localization method based on the long short-term memory recurrent neuron network. IEEE Access 7:123912–123921. IEEE
6. Mekki K, Bajic E, Meyer F (2019) Indoor positioning system for IoT device based on BLE technology and MQTT protocol. In: IEEE 5th world forum on internet of things (WF-IoT), pp 787–792. IEEE, Limerick, Ireland
7. Heyn R, Kuhn M, Schulten H, Dumphart G, Zwyssig J, Trsch F, Wittneben A (2019) User tracking for access control with bluetooth low energy. In: IEEE 89th vehicular technology conference (VTC2019-Spring), pp 1–7. IEEE, Kuala Lumpur, Malaysia
8. Qureshi UM, Umair Z, Hancke GP (2019) Indoor localization using wireless fidelity (WiFi) and bluetooth low energy (BLE) signals. In: IEEE 28th international symposium on industrial electronics (ISIE), pp 2232–2237. IEEE, Vancouver, Canada
9. Sawaby AM, Noureldin HM, Mohamed MS, Omar MO, Shaaban NS, Ahmed NN, El Hadidy SM, Hussein RS, Hassan AH, Mostafa H (2019) A smart indoor navigation system over BLE. In: IEEE 8th international conference on modern circuits and systems technologies (MOCAST), pp 1–4. IEEE, Thessaloniki, Greece
10. Wang J, Takahashi Y (2018) Indoor mobile robot self-localization based on a low-cost light system with a novel emitter arrangement. ROBOMECH J 5(17)
11. Yang C, Shao HR (2015) WiFi-based indoor positioning. IEEE Commun Mag 53(3):150–157
12. Ebner F, Fetzer T, Deinzer F, Grzegorzek M (2019) On Wi-Fi model optimizations for smartphone-based indoor localization. Int J Geo-Inf 6(8)

13. Wang F, Feng J, Zhaoi Y, Zhang X, Zhang S, Han J (2019) Joint activity recognition and indoor localization with WiFi fingerprints. IEEE Access 7
14. Molina B, Olivares E, Palau CE, Esteve M (2018) A multimodal fingerprint-based indoor positioning system for airports. IEEE Access
15. Liu Q, Qiu J, Chen Y (2016) Research and development of indoor positioning China communications. 2016(2z):67–79
16. Yazti DZ, Laoudias C, Georgiou K, Chatzimilioudis G (2017) Internet-based indoor navigation services. IEEE Internet Comput 21(4). IEEE Computer Society

On the Evaluation of Effectiveness of eLearning and Blended Learning

Sarka Hubackova

Abstract Currently, the development of information and communication technology is influenced by versatile information and increases the potential to analyze the information through digital technologies. Most of the learners utilize ICT as a tool to get enhanced vision in their carrier. The use of multimedia in teaching and learning is examined to explore research. The effectiveness of teaching is dealt with interactive teaching and the use of eLearning. In order to obtain the relationship between the students and teachers, ICT will be a better platform. This research work presents an analysis of issues in educational process and its effectiveness and provides possible solutions for eLearning to increase the effectiveness in face-to-face teaching learning process.

Keywords Effectiveness · Assessment · eLearning · Teaching · Learning · Foreign languages · Teaching methodologies

1 Introduction

In the present situation, everyone uses smart gadgets like mobile phones, tabs for accessing various applications through Internet. All these applications are data oriented and include various data processing features. The technology development brings multimedia into handheld device and influences the society through its communication modules and reduces the communication gap through its fast processing characteristics. This provides various new opportunities and opens up various ways in education systems. Effective and efficient education system are a complex process, and it requires various factors which are practically difficult to measure. Various research works are evolved to evaluate the eLearning and teaching practices, and among them, Kirkpatrick methodology of evaluation is much familiar.

It is a simple and common model, although it was created as early as in 1959. Later, it began to be used for evaluation of eLearning. Among contemporary authors

S. Hubackova (✉)
Department of Applied Linguistics, Faculty of Informatics and Management, University of Hradec Kralove, Hradec Kralove, Czech Republic
e-mail: sarka.hubackova@uhk.cz

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_18

dealing with the measurement and evaluation of effectiveness, the example of Khan or Syverson should be shown. Original view on the aspects of eLearning (CAPEODL) was created by Khan [1, 2].

In education, different types of effects can be seen such as training effectiveness, individual effectiveness, pedagogical effectiveness, education system effectiveness, education process effectiveness and school effectiveness.

2 Methods

Initially, a method of the literature review of available sources exploring the issue of modern teaching methods was used. Research work utilizes databases such as science direct, Scopus, Web of science and Springer. The research issues are identified by evaluating the research works. A questionnaire is framed to obtain the relationship between the students and ICT. Also to obtain the teaching supported by ICT in student's view. The research process is started during 2018–2019, and all the questions are provided with multiple choices.

3 Effectiveness of eLearning

In the following part of our paper, the effectiveness and its relation to eLearning have conversed, respectively. The basic assumption of the course is a kind of technical, and any student who might take part in an eLearning teaching process can receive their appropriate e-material and be excellent [3]. This presumption stands complex of pedagogical assumptions.

Unlike face-to-face teaching, eLearning is more turned to an individual. The academic result of it is that a tutor should have a certain and if possible, a clear idea about the language knowledge of the students attending the seminar group [4].

Nowadays, eLearning is widely adopted, and smartphones supports the teaching and learning process. Smartphones are easily available, and multimedia teaching through smartphones gains more attraction in the recent days [5].

A tutor can add to a student's interests but cannot rely on the attractiveness of the new method. Also, an attractive and newly treated eLearning material are very important here; it supports student's attention and concentration, and in this way, it becomes a good assumption of eLearning effectiveness.

An important role in effectiveness may be played also by a suitable time extent in which eLearning is used. If this use should be effective, then it cannot be too long. So, if the contents of eLearning material and new information should be more complicated, then it is better to divide the instruction into two or three shorter intervals. Their effectiveness is higher than the efficiency of a one long time stage.

eLearning offers a good possibility of checking the quality of a student's cooperation. The grade of effectiveness depends here not only on the intensity of student's work, but can also be influenced by a teacher's attitude. It is necessary to mention in this connection a certain tutor's self-criticism and his willingness to revise the contents of the given task, its extent, the sequence of its items that the student must observe. eLearning makes it possible to inform the tutor about all of it. Such information is usually also a picture of the immediate effectiveness of the tutor's activity. Any change in the procedure or improvisation is here difficult. It is therefore important to think of a possibility of change during the tutor's home preparation. The tutor has to count on the possibility to change the method and use blended learning.

The necessity of making short inputs with closed content directed if possible, to a simple problem has its close connection with a time limitation of eLearning. In accordance with its complexity, the tutor decides if the relevant foreign language is used and assigns the tasks or explains the instruction in the students' mother tongue. Sometimes, it is advantageous to explain the instruction in the foreign language but to repeat its succinct contents in the mother tongue.

The effectiveness of the eLearning method is higher in the case when the task with its all details is assigned in a way that excludes further students' questions on how to do this or that. It is valid here in principle: The more distinct the assignment of a task is, the higher the measure of the effectiveness of the used method will be. The specification of tasks has also its psychological importance: A student takes for superfluous any help of another person, and he embarks upon his task with interest and force.

It is not suitable to use eLearning partially for presenting of information only on the one hand and for assigning of tasks on the other hand. Any one-sidedness here has a calamitous psychological influence, and it reduces the measure of effectiveness.

Both the variants have just mentioned can be accompanied by a very short and simple text. It does not necessarily have the nature of an examination. Sometimes, the tutor wants to check that the student understands an instruction or a task well. In another case, the assigned task applied to the information or instruction explained earlier. The test is not only a depiction of the short-time effectiveness of tutor's attitude, but it can show a student some faults in his work. This short test opens the possibility to check the cooperation of a seminar group and its members with a tutor. The awareness that he has such a possibility usually forces students to higher effort. But in our experience, the test assigned in this way is not the suitable source material for obtaining a credit. A face-to-face way for a much more suitable form is taken in this case.

4 Effectiveness of the Process of Teaching

The concept of effectiveness of teaching as a didactical category is not unambiguously delimitated. Contemporary literature discussing this pedagogical field proves this fact quite clearly. The terms effectiveness, efficiency as synonyms were used. In this sense, the effectiveness is seen as a measurable trace of a teaching process, a trace having a certain duration in the cultural consciousness of an individual who completed a teaching process. [4, 5].

The sub-terms used in our explanation are tried to explain with some occasional examples from different fields of teaching. The similar trace that has been in mind when mention a teaching process can be left in an individual's cultural awareness also by some other process. One of them is, for example, reading: a contact with individuals from other national fields or cultures, the experience of life or contact with educated members of the same nation often belonging to different generations. A special place is then taking here by different forms of self-study connected both with direct school teaching and with further lifelong education. The student's does not have school environment on the mind, but the educational occasions that can take their places in a gallery, a museum, on archaeological workplaces, in a workshop, in a concert hall or a laboratory are considered.

It is not possible to reflect on the full content and extent of a teaching process in this connection. There are, namely—almost at the same time—also two other processes: An individual, who accepts the content of a teaching process, makes with its perception concerning usually knowledge almost immediately a selection of both its content and its extent. But the facts that an individual keeps unwittingly or consciously in his consciousness are touched also by a quite different process, by a process of forgetting. Of course, a teaching process does not concern new information only, but also skills and teaching methods. And the process of forgetting treats the named entities in a quite different way than the pieces of information. Those plain facts are mentioned for two reasons: On one hand, wanted to point out that the effectiveness of a teaching process does not concern the mediation of pieces of information. On the other hand, wanted to emphasize that a face-to-face teaching must be taken into account and the more permanent effectiveness in which revision and strengthening of learned facts, their connection with other facts, etc. The duration of effectiveness which has been already mentioned in our definition represents a very complicated problem. To simplify it, immediate, short-term and long-term efficiency are distinguished. The share of one of them reveals most markedly in the situations, where it concerns the trace of face-to-face teaching. In the short-term and long-term effectiveness, the share of the effectiveness of teaching is always very high.

Also, some other factors come into effect: organization of learning and teaching of the relevant subject, its position in the educational plan of the relevant school purposefulness of the sequence of its contents, logical relations of separate basic elements of learning materials, used methods and books or teaching aids. Even the connection of effectiveness with the aim of the relevant subject in a certain span of time might shows itself as a very tight one. A teacher himself sometimes makes a self-examination about the study material and also in need of having a decision over the material. The important factors to be considered are with usefulness and applicability of the acquired facts.

The teaching in a foreign language lesson devoted for acquiring new vocabulary might have a very high immediate effectiveness. The teacher and the method can guarantee also it is extending to an acceptable level of short-time effectiveness and the focus on the new words from different points of view. The corresponding tutor chooses and explains the most important new words in connection with their frequency, meanings and grammar or orthography. And a certain kind of selection that have been mentioned. However, one can guarantee a prolongation of the immediate effectiveness into a short time one also by means of the textbook utilized. School textbooks usually remember the taught new words consistently and deliberately in a form of different exercises; they occasionally repeat at least the basic words of the new vocabulary. The teacher depends on the textbook in this case and does not pay any specific attention to newly educated vocabulary.

The basic relation mentioned here is to figure out the valid time. The intermediate effectiveness is a significant condition of the short-time efficiency, and this one is a condition of the long-time effectiveness.

The false students' opinion of the usefulness of the material should study usually is a big obstacle of the effectiveness of teaching. The tutor should judge the material that should be taught in the lesson from this point of view in his home preparation and prepare a suitable motivation. The purpose of the material and its place in the system of information and its indispensable place in the cultural consciousness of an individual.

The relation of methods and effectiveness leads to another speculation. One example is that a foreign language teaching students' mutual communication and conversation with the teacher play an important role. However, organic parts of a lesson can represent also some other activities, as, for instance, explanations concerning orthography or grammar rules that do not proceed as a communication nevertheless they can contribute to lessons' effectiveness evaluated in its relation to its scientific aim. Even such a modern method as eLearning does not necessarily contribute to higher effectiveness if its excessive use limits the process of authentic students' communication.

Motivation, often beyond the advancements and older students at universities always lead to a higher level of effectiveness. An interesting text attractive not only by its content but also by the tasks connected with its analysis can play a certain motivation in foreign language teaching. Even perfectly silent reading might be a very effective process. Teachers know well the situations when one can feel in the very quiet atmosphere of class that the students work with interest and very hard.

There is also a short of possibility of effectiveness' measuring in our definition. The effectiveness is usually measured by means of a test. A teacher does not measure his effectiveness in most cases but interested only in the extent or content of the taught material that the student masters. The harmful influence of frequent testing in this connection is to be mentioned. Students feel very soon what kind of information their teacher will check because he takes it for important. And this fact results in students' attitude to such information and their quite different attitude to the rest of the taught material. This student's attitude reflects the effect that even a very conscientious and scrupulous teacher's work leaves behind itself a very vague trace.

5 Findings

Contemporary pedagogy takes advantage of the most modern methodologies. A standard connectivity between eLearning and teaching is termed as blended learning [6, 9]. Through this blended learning, the satisfaction level of students with the teaching process could be furnished. Students responses for the survey is depicted as a graph below, and the feedback will help to find out the students rank and course complexity. In the survey, 78 students are allowed to participate, and out of 78, 72 students submitted the survey. The question framed was more suitable for face-to-face learning or learning supported by an online course. The students are asked about which method they most admire and most effective. Students considered blended learning to be the most effective method which is described in Fig. 1.

The importance of a teacher for effective teaching is also asked, and the role of the teacher is also considered by the students to be very significant which is explained in Fig. 2.

Figure 3—During the self-study, the following eLearning materials are used most often:

Dictaphone	1%
TV	4%
DVD/video	4%
Textbook	15%
Smartphone	18%
Internet	27%
Multimedia courses	31%

Fig. 1 Method of teaching

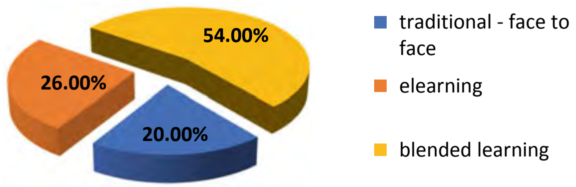
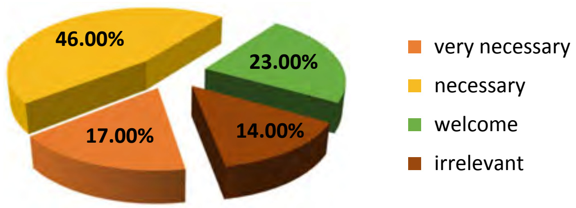


Fig. 2 Role of a teacher



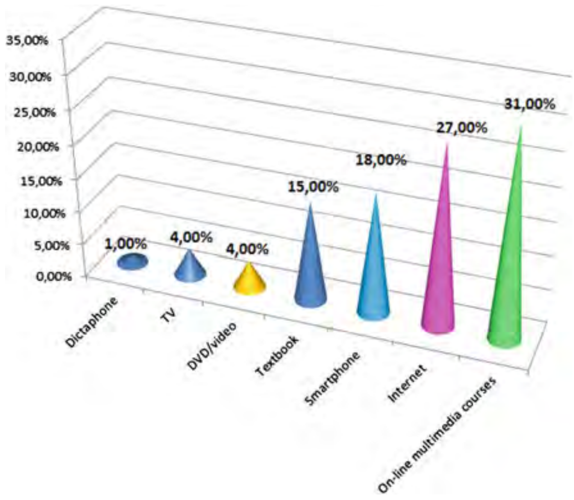


Fig. 3 Most often used eLearning materials

6 Conclusion

The success of eLearning depends on many factors. The quality of the course as a whole, on the virtual environment in which the education takes place, on the readiness of students to work in a virtual learning environment, their ability to orient in the environment and in the use of all the tools that information and communication technologies offers, the personality of the tutor and the ability to run the course well and responsibly. And a very important role in the effective use of ICT in the education process is played by the student's attitude to individual work in the virtual learning space.

The eLearning represents potential possibilities for measuring both the method and the eLearning materials as Khan points out. But the contemporary pedagogic research heads towards another direction. The scholars do not talk about the effectiveness of a teaching process, but about the effectiveness of schools.

Acknowledgements This article was supported by the project titled SPEV (2104/2020) at the Faculty of Informatics and Management of the University of Hradec Kralove, Czech Republic.

References

1. Khan BH, Granato LA (2017) Program evaluation in eLearning. https://asianvu.com/digitallibrary/elearning/elearning_program_evaluation_by_khan_and_Granato.pdf. Last accessed 23 Sept 2017
2. Syverson MA, Slatin J (2017) Evaluating learning in virtual environments.https://www.learningrecord.org/caeti.html. Last accessed 18 Sept 2017

3. Poulova P (2003) Využití eLearningu ve vysokoškolské výuce z pohledu studentů a vyučujících. In: E-learning—Sborník příspěvků ze semináře a soutěže eLearning 2003, Gaudeamus Hradec Králové
4. Průcha J (2002) Moderní pedagogika. Praha
5. Skalková J (2007) Obecná didaktika. Praha
6. Frydrychova Klímova B, Poulova P (2011) Tutor as an important eLearning support. Proc Comput Sci 3:1485–1489. https://www.sciencedirect.com/science/article/pii/S1877050911000378. Last accessed 22 Feb 2011
7. Frydrychova Klimova B (2009) Blended learning. In: Research, reflections and innovations in integrating ICT in education, Lisboa
8. Hoffmann L (1987) Kommunikationsmittel Fachsprache: Eine Einfuehrung. 3. Aufl. Akademie Verlag, Berlin
9. Pikhart M (2014) New horizons of intercultural communication: applied linguistics approach. Proc Soc Behav Sci 152:954–957
10. Hubackova S (2010) Foreign language teaching with WebCT support. Proc Soc Behav Sci 3:112–115

Design and Performance Analysis of GaAs-Based P-i-N Photovoltaic Using AlGaAs as Window Layer

Rocky Chakma, S. S. Mahtab, M. J. Alam, and Rupa Akter

Abstract In this study, the simulation is based on the P-i-N solar Photovoltaic. P-i-N-based reference solar cell founded on GaAs for enhanced performance was utilized with AlGaAs as the window layer. Different layers have been optimized according to the best performance. Impact of using Anti Reflecting coating has also been studied. It has been found that with Anti Reflecting coating, P-i-N solar cells efficiency has increased significantly. A horizontal band section has been placed in the intrinsic substance of the P-i-N model and because of this, high η (efficiency) is build up with high short circuit current (Isc); high open-circuit voltage (Voc) as extra short circuit current density (Jsc) comes from the intrinsic section. 40.9133% is the highest η in the research.

Keywords P-i-N · GaP · Photovoltaic · Inorganic solar cell · GaAs

1 Introduction

Producing electrical power from light power is a one-step alteration by PV arrangement whose justification is dependent on quantum theory. Photons observed from light, whose power relies on frequency and shade of the light. Energy from photons is enough to electrify electrons and in the advanced power levels, they are freer to

R. Chakma
Department of EEE, USTC, Chittagong, Bangladesh
e-mail: rocky.cht@gmail.com

S. S. Mahtab
Department EEE, Feni University, Feni, Bangladesh
e-mail: mahtabshahzad@gmail.com

M. J. Alam (✉)
Department of EEE, Feni University, Feni, Bangladesh
e-mail: alameee1993@gmail.com

R. Akter
Department of EEE, Mymensingh Engineering College, Mymensingh, Bangladesh
e-mail: rupa.mec.eee.bd@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_19

be in motion [1]. In a p-i-n solar unit, an intrinsic section is inserted between p- and n-type coating of a GaAs p-n solar unit. For this case, Jsc is higher than the common p-n cell because of the supplementary carrier donation of the intrinsic section. As a result, efficiency increases. In quantum well solar units, the difference among the only bandgap and solar band's event power is provided by supplementary power like photons from diverse power can be concentrated well. So, it has an essential likeness. Here, mainly the concern is the production of various light-produced quasi-Fermi stages [2]. Among the two entrances, the transfer of the carriers is significantly unlike. In quantum well solar unit application, transport is achieved when the carriers at each sectional energy level escape due to light absorption. The runaway moment ought to be faster than recombination moment so that the accumulation of η can be obtained in the greatest number. Well and truly the quantum-well infrared photodetectors illustrate the expediency of the scape process [3], which have peak assemblage from intra-sub band process. Sectional band approaches have additional superiority in that consecutive sectional energy levels can have dissimilar energies, therefore authorizing a good quantity of productive band gaps and high performance [4].

2 Structure Analysis

The individual solar cell has a complex construction with various coatings over and above the p- and n-coating. The function of these coatings is to decrease face and reverse shell recombination along with superficial mirroring. A rough copy of a P-i-N indicated cell is shown in Fig. 1 that is screening every coating such as windows, anti-reflective coating, p^+ , P-i-N, and n^+) to get great η.

The V-I characteristic of the only p- and n-coating is presented. In this paper, the V-I characteristic of a solar cell configuration is derived from the continuity equation for both electrons and holes, with the correct boundary conditions for this construction. Inspired by [5] and the V-I quality of n^+-n-p junction is resultant.

2.1 Proposed Structure

In Table 1, various resources are used in suggestion cells.

For the opposition of reflective coating, R(E) is very much condensed also the reflection damage may become short around 2% [6]. In this research, the model has been used as a reflectivity equal to zero for every wavelength. The I-coating which is in among p- with n-coating is preferably un-doped. Having a grounding doping, N_i is greatly lesser than the injecting of p- and n-coating. Deficiency section expands in P-i-N connection addicted to the mild doped section. Within the p-i-n connection, depleted thicknesses of both p- and n-coating are understood as little [7]. Throughout a front outside field, greatly doped p^+-coating above the p-coating condenses the exterior recombination rate. P-layer's electrons get together a probable

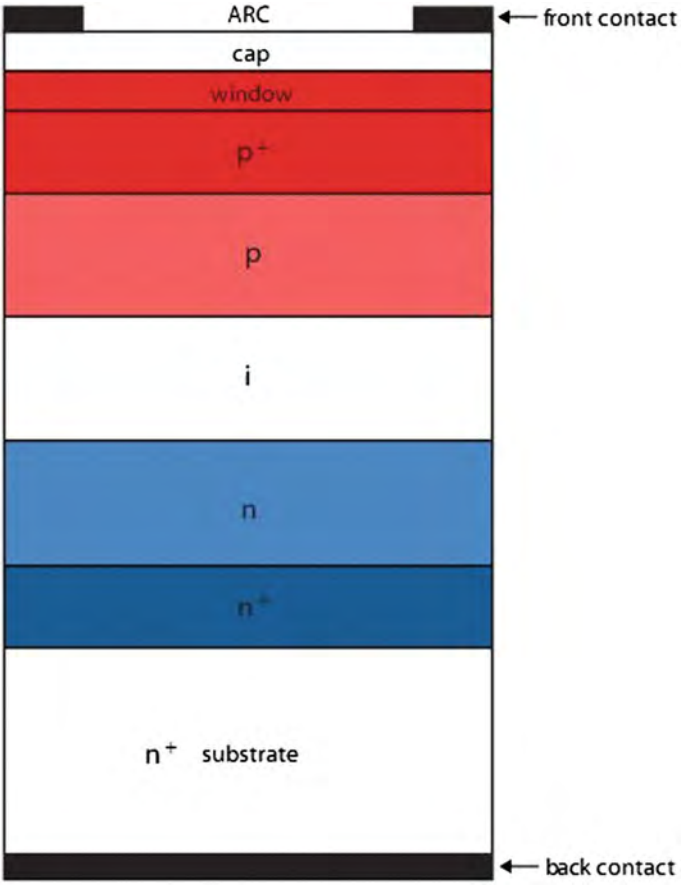
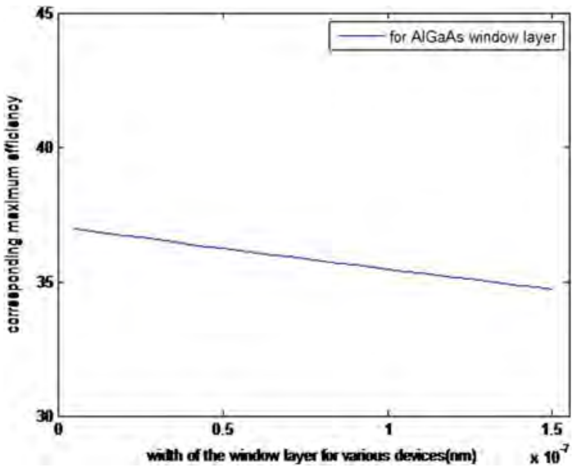


Fig. 1 P-i-N reference cell's construction

Table 1 Materials used in a suggestion cell

Coatings	Materials
Window	AlGaAs
p^+	GaAs
P	GaAs
I	GaAs
N	GaAs
n^+	GaAs

Fig. 2 Various efficiency for various thickness of $Al_xGa_{1-x}As$ ($x = 0.804$) window layer



barrier reliant on top of the proportion of doping intentness both in p$^+$-and p-coating while going as of p-coating hooked on the p$^+$ coating. It is known as front exterior field [8]. As the P$^+$ coating is located beside the front exterior where there is greatest photon flux, the carriers produced within p$^+$-coating supply much to the photocurrent [5]. In the boundary of un-depleted p-coating at z = 0, whole electron current (Fig. 2) is the addition of the donations. It is from the P$^+$ coating, p-coating along with depleted section among the P$^+$, and p-coating.

$$j_{n,\mathrm{total}}(0) = j_{n,p}(0) + j_{n,p} + (0) j_{n,\mathrm{dept}}(0)$$

Inserting a window coating, above the p$^+$-coating; which is prepared by advanced band gap substance, the valuable exterior recombination rate is more condensed. By placing supplementary greatly doped n$^+$ coating below the n-coating provides exterior field. It reduces the valuable rear exterior recombination rate similarly as the front exterior.

2.2 AlGaAs Seeing that Window Coating

Surplus carrier's recombination takes place not only inside the size of a semiconductor but also inside the exterior of it. By cutting short the periodicity at the exterior, it performs as boundary among the semiconductor and an additional substance. So, the surface recombination rate is different (mainly superior); which can be cut by passivating (or window coating) with the intention that it can avoid minority carrier as of going to the exterior, rather than the size of the semiconductor. Exterior recombination rate is robustly reliant on its roughness, pollution, ambient gases. In other material-based window layer such as GaAs, exterior recombination rate is excessive (of the order of cm/s). It can be cut capable of 10-cm/s by deposing a thin coating of AIGaAs [9].

3 Computational Methodology

3.1 P-i-N Solar Cell

3.1.1 State's Band Construction with Valuable Density

Al_xGa_{1-x} As and GaAs come together having the same lattice constants in zinc blende arrangement (variance is just 0.12%). Which indicates As that can be developed on GaAs with no damage can now stay away from the configuration of dislocations that might raise the recombination [10].

$$E_{\mathrm{GaAs}}(T) = \left[1.519 - \frac{5.405 \times 10^{-4} T^2}{T + 204}\right] \mathrm{eV}$$

For Al_xGa_{1-x} As smallest conduction band's location the Γ, X, and L bands are reliant on x, and As is a straight bandgap semiconductor for the condition of $x <$ 0.45. It is shown away in bellow [11]

$$E_{\mathrm{Al}_x}\mathrm{Ga}_{1-x}\mathrm{As} = E_{\mathrm{GaAs}} + 1.247x$$

Al_xGa_{1-x} is for an indirect bandgap semiconductor for the condition of $x \geq 0.45$ with a bandgap written as

$$E_{\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}} = 1.911 + 0.005x + 0.245x^2$$

Gallium arsenide's permittivity having temperature reliance is specified as follows:

$$\in = \left[12.79\left(1 + 1.0 \times 10^{-4} T\right)\right] \in_0$$

This equation for the extension of the depletion region in the moderately injected region which is utilized from [12] and manipulated as an estimation for w_{a}

$$w_a = \sqrt{\frac{2 \in k_B T}{q^2 N_a}} \arctan\left(\frac{q V_{pp} +}{k_B T} \sqrt{\frac{N_a^+}{2 N_a}}\right)$$

A fixed voltage of the n^+-n joint has the form [13]. On or after [12] the depletion region's length is in use hooked on the lightly doped n-section and utilized at the same time as an approximation for,

$$w_b = \sqrt{\frac{2\epsilon k_B T}{q^2 N_d}} \arctan\left(\frac{qV_{nn}+}{k_B T}\sqrt{\frac{N_d^+}{2N_d}}\right)$$

4 Numerical Results and Discussion

4.1 The Window Layer's Consequence of Thickness

For a variety of thickness of window coating, the highest η and Jsc have been estimated with the constant thickness of other coatings and further factors (Table 2).

From the data table, it is clear that increasing the width of the window layer causes to decrease efficiency because it increases recombination. As $Al_xGa_{1-x}As$ (where $x = 0.804$) has an indirect bandgap, non-radiative recombination (i.e., Auger recombination) dominant here where radiative recombination is concealed because of the requirement of photons in this method. The usefulness of neither a pure initial substance nor a fresh fabrication procedure can improve this auger recombination. Only its lower thickness can cut the auger recombination rate in the interior of the cell. So, a lower thickness device gives better performance (Fig. 3).

This graph represents how efficiency is decreasing with the increase of windows layer thickness. At 5 nm thickness of windows layer, the peak efficiency can be calculated which is approximately 36.9481 (Fig. 4).

From the graph, the highest density can be obtained at 5 nm windows coating which is nearly 450 A/m^2 whereas the cut off voltage 0.9 v is estimated (Fig. 5).

This graph elucidates the decrease of quantum efficiency with the increase of wavelength at different windows layers. From the graph, it is more clear that 5 nm windows layer is much more effective. Although the other thickness shows nearly the same values but can be ignored. At 400 nm wavelength, the peak quantum efficiency

Table 2 Changeable thickness of window coating and simulation result

Device no	Window layer's thickness in nm	AlGaAs seeing that a window coating		
		Volt	A/m^2	Max efficiency
1	5	1.0054	445.1377	36.9481
2	10	1.0053	444.2197	36.8683
3	50	1.0046	436.9647	36.2371
4	100	1.0038	428.1129	35.4670
5	150	1.0029	419.4953	34.7172

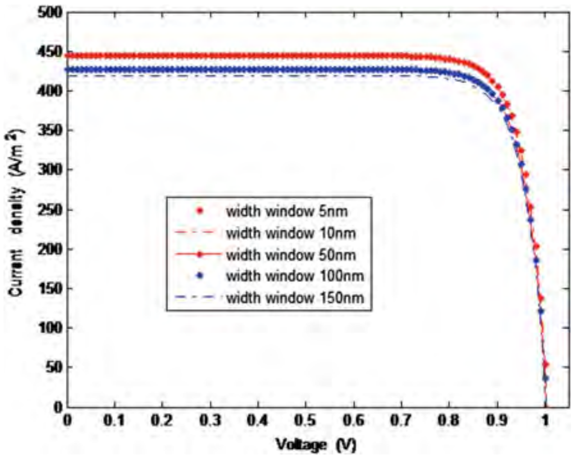


Fig. 3 V-I density arc for diverse width of the AlGaAs window coating

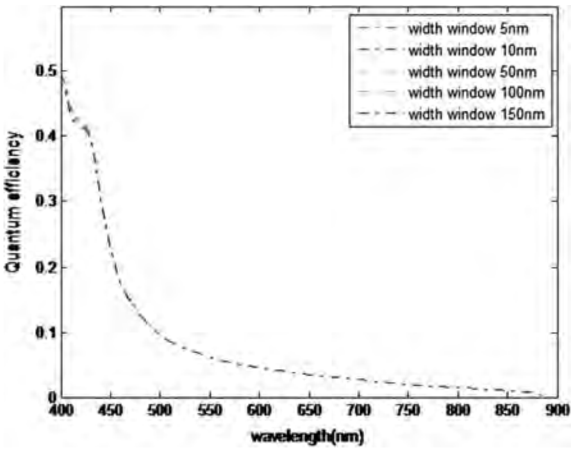


Fig. 4 λ versus QE in favor of diverse thickness of AlGaAs window coating

can be found. No light absorbed after the 900 nm wavelength because of the bellow bandgap of the active layer.

4.2 *P^+ Coating's Consequences of Thickness*

For a variety of thickness of p^+ coating, the highest η and Jsc have been estimated with the constant thickness of other coatings and further factors. To see the difference, AlGaAs window coating devices information is contained in Table 3.

A dark injected p^+-layer is found below the window layer, which additionally diminishes the front-surface recombination. The data table confirms that η and thickness of the p^+ layer are in addition linked. Thickness and η are proportional to each

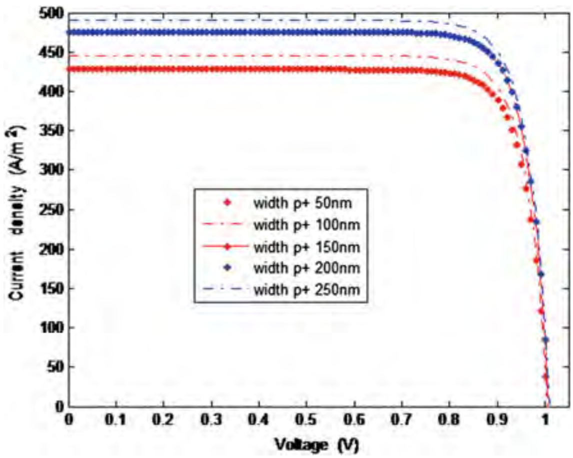


Fig. 5 V-I density arc of the p^+ coating in favor of a range of thickness (via AlGaAs while a window coating)

Table 3 AlGaAs window coating devices information

Device no	p^+ layer's thickness in nm	AlGaAs seeing that a window coating		
		Volt	A/m^2	Max efficiency
1	50	1.0038	428.4171	35.4935
2	100	1.0054	445.1377	36.9481
3	150	1.0068	461.0628	38.3385
4	200	1.0080	476.1407	39.6652
5	250	1.0092	490.3254	40.9133

other as a consequence of the absorption of more photon can be done in this coating. The data table also confirms that AlGaAs window coating has an identical mechanism presentation (Fig. 6).

The data from the graph directly represents that efficiency is also associated with the thickness of the p^+ layer which is also reflected in the data table. Enhancing the width of the p^+ layer also enhances the efficiency of the PV solar cell. This is happened because of the absorption of more photons by this layer. Considering the optimizing thickness, 250 nm p^+ layer has been calculated for the value of 40.9133% (Fig. 7).

This graph better designates how quantum efficiency enhances with the enhancing of the thickness of the p^+ layer considering optimized wavelength. The photocurrent is increased with the increase of quantum efficiency; As photocurrent is precisely associated with the quantum efficiency.

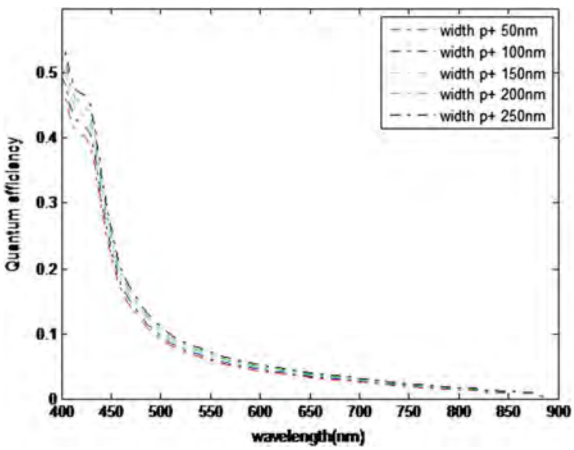


Fig. 6 λ versus QE in favor of diverse thickness of p+ coating (via AlGaAs while a window coating)

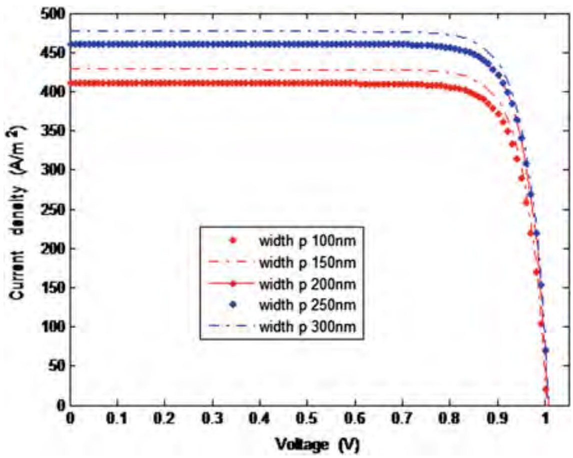


Fig. 7 V-I density arc of the p-coating in favor of a range of thickness (via AlGaAs while a window coating)

4.3 *P-Coating's Consequences of Thickness*

For a variety of thickness of p-coating, the highest η and Jsc have been estimated with the constant thickness of other coatings and further factors. To see the difference, AlGaAs window coating devices knowledge is carried in Table 4.

Various curves are obtained from a simulation that is given below (Fig. 8).

It can be seen that the current density increases when the thickness of p-coating increases. Considering the pv cell thickness, 300 nm P-coating has been optimized for 476.6671 A/m^2 current density (Fig. 9).

As the quantum efficiency directly associated with the thickness of P-coating, the quantum efficiency enhances with the enhancing of P-coating at an optimized

Table 4 For a changeable thickness of p-coating and simulation results

Device no	p layer's thickness in nm	AlGaAs seeing that a window coating		
		Volt	A/m^2	Max efficiency
1	100	1.0022	410.9044	33.9701
2	150	1.0038	428.3638	35.4890
3	200	1.0054	445.1377	36.9481
4	250	1.0068	461.2356	38.3536
5	300	1.0081	476.6671	39.7113

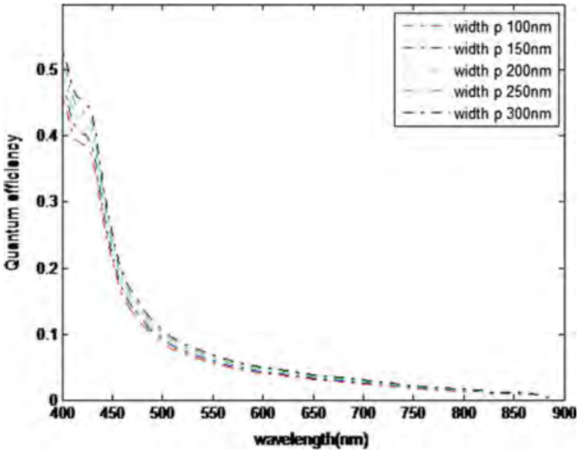


Fig. 8 λ versus QE in favor of diverse thickness of p-coating (via AlGaAs while a window coating)

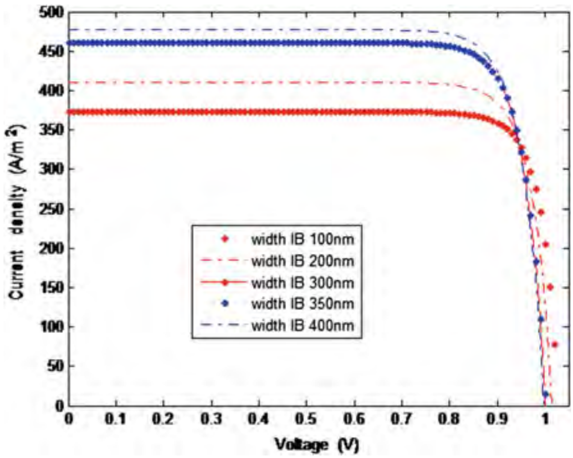


Fig. 9 V-I density arc of the intrinsic coating in favor of a range of thickness (via AlGaAs while a window coating)

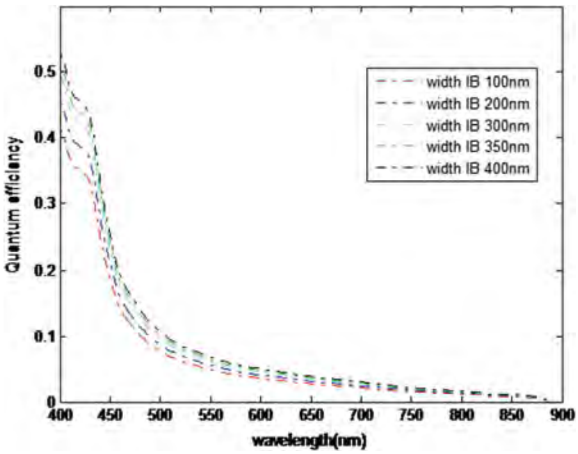


Fig. 10 λ versus QE in favor of diverse thickness of p-coating (via AlGaAs while a window coating)

wavelength which is reflected in the graph. It also can be noticed from the graph that the peak quantum efficiency can be obtained at 300 nm P-coating layer.

4.4 Intrinsic Coating's Consequences of Thickness

For a variety of thickness of the intrinsic coating, the highest η and Jsc have been estimated with the constant thickness of other coatings and further factors. The device information of the AlGaAs window coating is contained in Table 5.

Various curves are obtained from the simulation that is given below.

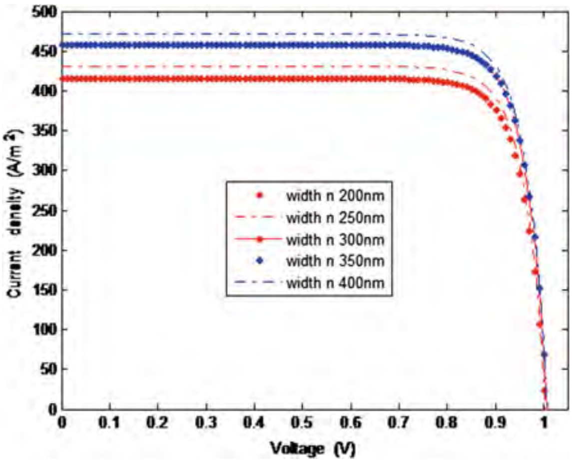
As an intrinsic layer increase the photon absorption, the short circuit current is increased with the increase of intrinsic layer and this is well demonstrated in the graph. 400 nm intrinsic layer has been optimized for the short circuit current of 477.2973 A/m^2.

Figure 11 confirms that QE is proportional to the thickness of the intrinsic coating. As a result, Isc too increases which is given in Fig. 10. The thickness is proportional to the number of photons soaked up, and power is also relative to the quantity of

Table 5 AlGaAs window coating

Device no	Intrinsic layer's thickness in nm	AlGaAs seeing that a window coating		
		Volt	A/m^2	Max efficiency
1	100	1.0281	373.4749	32.3951
2	200	1.0150	410.5894	34.6451
3	300	1.0054	445.1377	36.9481
4	350	1.0014	461.5055	38.0810
5	400	0.9979	477.2973	39.1655

Fig. 11 V-I density arc of the n coating in favor of a range of thickness (via AlGaAs while a window coating)



photon soaked up so η is related to the thickness of the intrinsic coating. But if the thickness is sufficiently increased, power saturates.

4.5 N-Coating's Consequences of Thickness

For a variety of thickness of n-coating, the highest η and Jsc have been estimated with the constant thickness of other coatings and further factors. To see the difference, AlGaAs window coating devices knowledge is carried in Table 6 (Fig. 12).

From the graph, it can be noticed that the maximum short circuit current (Jsc) can be obtained at 400 nm depositing n layer whereas the minimum Jsc is at 200 nm.

The data table and arc confirm that thickness is proportional to η and Isc, because it soaks further photon.

Table 6 Changeable thickness of n-coating and simulation results

Device no.	Intrinsic coating's thickness in nm	AlGaAs seeing that a window coating		
		Volt	A/m^2	Max efficiency
1	200	1.0024	415.9974	34.4099
2	250	1.0039	430.9711	35.7150
3	300	1.0054	445.1377	36.9481
4	350	1.0067	458.5358	38.1171
5	400	1.0079	471.2188	39.2340

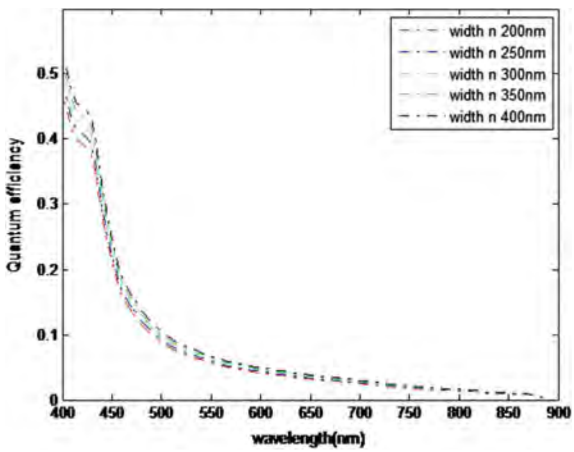


Fig. 12 λ versus QE in favor of diverse thickness of n-coating (via AlGaAs while a window coating)

4.6 n^+ Coating's Consequences of Thickness

To decrease the back surface recombination, a dark injected n^+-layer is accommodated beneath the n layer. For a variety of thickness of n^+ coating, the highest η and Jsc have been estimated with the constant thickness of other coatings and further factors. To see the difference, AlGaAs window coating devices knowledge is carried in Table 7.

The arc from Fig. 13 confirms that the highest Isc can be originated when the thickness of the n^+ coating is 150 nm, and the minimum can be found when the thickness of the n^+ coating is 25 nm. So, η increases proportionally with the increase of the depth of the n^+ coating. Figure 14 also confirms that maximum quantum efficiency can be originated when the thickness of the n^+ coating is 150 nm at an optimized wavelength.

Table 7 Changeable thickness of n^+ coating and simulation result

Device No.	Intrinsic coating's thickness in nm	AlGaAs seeing that a window coating		
		Volt	A/m^2	Max efficiency
1	25	1.0000	426.2351	35.2831
2	50	1.0022	433.4167	35.9173
3	75	1.0039	439.5884	36.4606
4	100	1.0054	445.1377	36.9481
5	150	1.0076	455.0994	37.8234

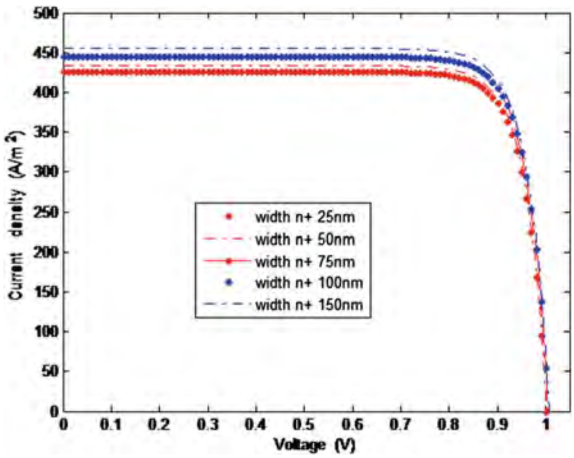


Fig. 13 V-I density arc of the n$^+$ coating in favor of a range of thickness (via AlGaAs while a window coating)

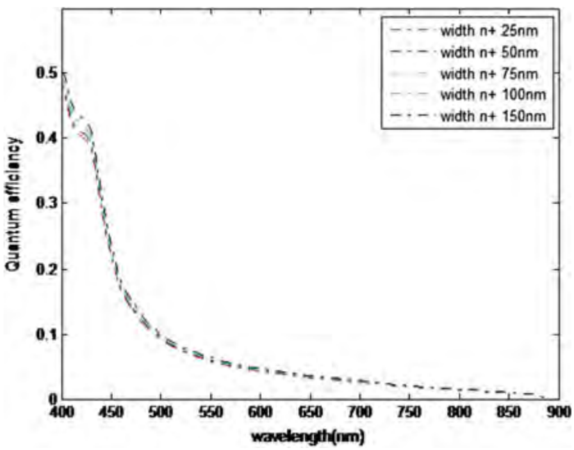


Fig. 14 λ versus QE in favor of diverse thickness of n$^+$ coating (via AlGaAs while a window coating)

5 Conclusion

In our research, the AlGaAs formed P-i-N indicated cell designed for leading presentation have been studied. To understand it, the P-i-N indicated cell's theoretical learning, as well as presentation, is vastly vital. Design of p-i-n indication cell is optimized by a simulation representation for great η from which preparation and presentation are put together. The indication cell's representation describes the consequence of a window coating. It also describes densely injected and coating for obtaining a small efficacious exterior alliance velocity with an anti-reflective coating to minimize the mirroring dissipation. Via this thickness, great QE is obtained. For the representation of P-i-N cell with the fundamental substance that is accommodated in a flat band area, a representation of great η is built up and high Isc and Voc is acquired because further Jsc comes from the intrinsic area. For AlGaAs substances, simulation is done with various factors particularly the thickness of diverse coatings. For AlGaAs, the greatest η is 40.913% for this research.

Conflict of Interest In this paper, all the authors have worked and contributed equally. The serial of author list included is basically according to seniority not by contribution. They all contributed equally and will be designated as the first author of this paper.

References

1. International Energy Outlook (2009) Energy Information Administration Office of Integrated Analysis and Forecasting U.S. Department of Energy, Retrieved online on 2009-06-13. http://www.jointsolarpanel.nl/fileadmin/jointsolarpanel/user/documents/seminar2004/stanleyzc04.pdf
2. Twidell J, Weir T (2006) Renewable energy resources. Taylor and Francis, Milton Park
3. Nelson J (2007) The physics of solar cells. Imperial College Press, London
4. Greenpeace and European Photovoltaic Industry Association (2008) Solar Generation V-2008, Solar electricity for one billion people and two million jobs by 2020
5. Dai XM, Tang YH (1996) A simple general analytical solution for the quantum efficiency of front-surface-field solar cells. Solar Eenergy Mater Solar Cel ls 43:363
6. Tobin SP, Vernon SM, Bajgar C, Geoffroy LM, Keavney CJ, Sanfacon MM, Haven VE (1988) Device processing and analysis of high efficiency GaAs cells. Solar Cel ls 24:103
7. Luque A, Martí A (2001) A metallic intermediate band high efficiency solar cell. Prog Photovolt: Res Appl 9:73
8. Martí A, Stanley CR, Luque A (2006) Intermediate band solar cells (IBSC) using nanotechnology, chapter 17 in Nanostructured Materials for Solar Energy Conversion. Elsevier B. V., Amsterdam
9. Martí A, Cuadra L, Luque A (2002) Quasi-drift diffusion model for the quantum dot intermediate band solar cell. IEEE Trans Electron Devices 49:1632
10. Streetman BG, Banerjee SK (2006) Solid state electronic devices. Pearson, London
11. Adachi S (1993) Properties of aluminium gal lium arsenide. INSPEC, 1993
12. Hauser JR, Littlejohn MA (1968) Approximations for accumulation and inversion space-charge layers in semiconductors. Solid-State Eelectron 11:667
13. Demoulin PD, Lundstrom MS, Schartz RJ (1987) Back-surface field design for n + p GaAs cells. Solar Cel ls 20:229

Impact Analysis of Anti-Reflection Coating on P-i-N Solar Device

S. S. Mahtab, Rupa Akter, Tauhidul Mahbub, Ahamed Raihan, Rapsan Amin Anonto, and M. J. Alam

Abstract In this article, p-i-n solar have been simulated with SCAPS-1D solar photovoltaic device simulator to observe the effect of anti-reflecting coating over the p-i-n solar device. The device was constructed using AlGaAs. GaAs was used as P^+ , P, i, n^+, n-layer with changing doping in it. It is observed in the proposed design that there is a significant impact of this layer in p-i-n PV Device. In the p^+-p-i-n-n^+ device, 36.948% efficiency has been observed with anti-reflective coating. But without anti-reflective coating, it went down only at 24.3850%. Almost 12% variation with a layer. In the device of p^+-p-i-n, 33.5399% efficiency has been observed with anti-reflective coating. In the device of p-i-n, 33.4215% efficiency has been observed with anti-reflective coating. But without anti-reflective coating, it went down only at 22.0723%. Almost 11% variation observed with a layer. So, it can be said that anti-reflecting coating reduces the loss and increases the efficiency of a p-i-n PV Device.

S. S. Mahtab (✉)
Department of EEE, Feni University, Feni, Chittagong Division, Bangladesh
e-mail: mahtabshahzad@gmail.com

R. Akter
Department of EEE, Mymensingh Engineering College, Mymensingh, Bangladesh
e-mail: rupa.mec.eee.bd@gmail.com

T. Mahbub
Department of EEE, Independent University of Bangladesh, Dhaka, Bangladesh
e-mail: tony.mahbub@gmail.com

A. Raihan
Department of CSE, Jahangirnagar University, Savar, Dhaka, Bangladesh
e-mail: araihan13@gmail.com

R. A. Anonto
Department of EEE, American International University Bangladesh, Feni, Bangladesh
e-mail: rapsanaminanonto@gmail.com

M. J. Alam
Department of EEE, Feni University, Feni, Bangladesh
e-mail: alameee1993@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_20

Keywords P-i-N PV device · GaAs · 3–5 material · Photovoltaic · ARC · Anti-reflection coating

1 Introduction

The global average electricity consumption in 2006 was approximately 1610^{13} W [1]. The cumulative incident of solar radiation at ocean level is approximately 210^{17} W per day, which is more than 12,500 times the centre global energy utilization in 2006 [2]. It demonstrates the amazing potential solar power for providing the planet with electricity [3]. Photovoltaic devices are one way to harvest the sun's power [4]. The big field of delivering this enormous amount of energy is a fossil fuel with scarce resources [5]. A few of these capabilities are accessible only in a very few areas around the world; security also poses a challenge to the stability and sovereignty of national and regional entities, as well as global security [6, 7]. The protection issue and cost of disposing of radioactive material and hazardous waste renders the use of nuclear and chemical technology a dubious option. The fossil fuel is also expensive for transmission in remote zones. Accordingly, the interest in renewable energy has risen the last few decades along with the world's energy requirements, the alarms of global warming, and oil prices [8, 9]. A PV device is a solid-state semiconductor system that, when stimulated by photons, produces electricity from DC (direct current) [10].

As photons come into contact with the device's atomic structure, they dislodge electrons from the atoms. Which leaves a vacuum that makes other free electrons usable. If the device fabricates a PN junction, the dislodged photons flow towards the junction's P side. The effect of this electron transfer is a flow of electrical current which can be redirected through electrical contacts from the exterior of the membrane to generate energy [11, 12]. PV device's conversion η is measured as the ratio of input power(radiant energy) to power output (electric energy). In a p-i-n solar device, an intrinsic region is inserted between p- and n-type layer of a GaAs p-n solar device [13, 14]. In this case, J_{sc} is greater than the conventional p-n solar device. Sue to the additional carrier contribution of the intrinsic region. As a result, efficiency increases. In quantum well PV devices, variation among the appearance of photo energy levels to sustain photons from different energy sources is the incidence energy of the solar spectrum and single strip width. There is a fundamental similarity in this case, as seen is the main issue is the generation of multiple light-generated quasi-Fermi rates. However, the carriers' transport between the two approaches is considerably different. In quantum well solar device approaches, transportation is achieved by letting the carriers escape through light absorption at each localized energy level. The time of escape should be faster than the time of recombination to maintain a high collection η [15]. Viability of the escape mechanism is well demonstrated by infrared photodetectors with a large range of intra-sub-band detectors. Localized band solutions have another benefit in that successive localized energy rates may have different energies, thus allowing for a large number of successful band gaps and high efficiencies [9, 12, 16, 17].

The reflectivity, R(E), is significantly reduced by using an anti-reflective coating, and losses of reflection can be as small as possible [18]. In this article, the solar devices with anti-reflective coatings and without anti-reflective coatings are modeled on a reflectivity equal to zero for all wavelengths [10]. The impact of anti-reflective coatings has been checked in PIN PV Device [19].

2 Device Structure

Practical solar cells have a perplexing formation with different layers besides to the p-and n-layer got in the facile p-n photovoltaic solar cell [20]. The motive of these layers is to assuage back and front periphery recombination and periphery reflection. A plan of a p-i-n relevance cell is represented in Fig. 1 indicating all the layers (window, p^+, p, i, n, and n^+) comprised to gain a high efficient PV solar cell.

Aside from p- also n-surface, it is been found the easiest p-n PV device, practical PV devices have a complex design with many surfaces [21]. These layers aim to reduce the recombination of the front and back surface and the reflections on the

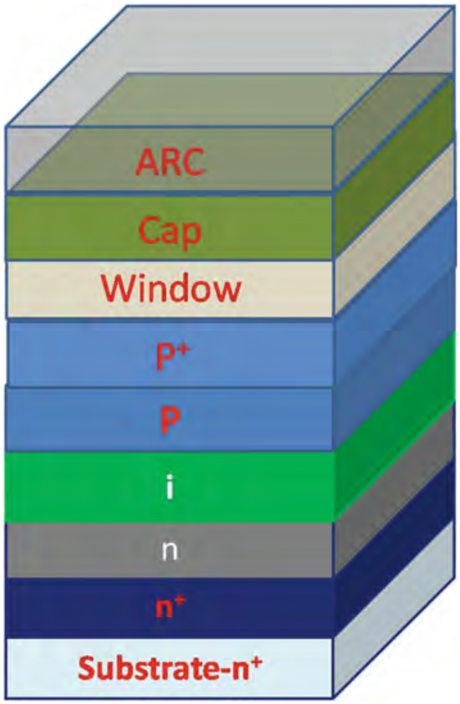


Fig. 1 Reference device p-i-n's structure

Table 1 Reference device's materials

Layers	Materials
Window	AlGaAs
p^+	Gallium Arsenide
P	Gallium Arsenide
İ	Gallium Arsenide
N	Gallium Arsenide
n^+	Gallium Arsenide

surface. A reference device with p-i-n is given in Table 1, for all the levels (anti-reflective surface(ARC), window, p^+, p, i, n, and n^+) included obtaining a great η PV device. It includes of an anti-reflective surface, a cap-, window-, p^+, -, p-, i-, n- and n^+ surface mounted above the surface.

Various materials on the table have been used below in the reference device.

3 Simulation Results

In the simulation model, the below method is considered for calculating the system parameters.

Efficient band gap E_g. The made-in capacity determined by exercising the dominance carrier quantities on the interfaces n and p, the useful state solidities in the valence band N_{valen} and the conduction band N_{cond}. It is possible to write the dynamic equation for the valence band and the conduction band conducting unit geometry as follows:

$$E_{\text{cond}} = -\chi - q\varphi \tag{1}$$

$$E_{\text{valen}} = -\chi - E_g - q\varphi \tag{2}$$

The Concentration of minority carriers can also be estimated using

$$n_l = N_c \exp\left(\frac{F_n - E_{\text{cond}}}{kT}\right) \tag{3}$$

where, in n-side F_n is the Fermi equilibrium stage.

The charge carrier conductance was determined exercising the bi-polar formulas of drift-diffusion as given bellow: for electron,

$$J_n = q\mu_e n_f \frac{\partial E_{\text{cond}}}{\partial x} + qD_n \frac{\partial n_f}{\partial x} \tag{4}$$

And also for holes,

$$J_p = q\mu_h p_f \frac{\partial E_{\text{valen}}}{\partial x} + qD_p \frac{\partial p_f}{\partial x} \tag{5}$$

For the measurement of carrier trapping, the Shockley-Read-Hall (SRH) recombination principle was used [9]. The efficiency of the unit was theoretically determined by using the formula

$$\text{FF} = \frac{P\text{max}}{I_{\text{sc}} \times V_{\text{oc}}} \tag{6}$$

$$\eta = \frac{I_{\text{sc}} \times V_{\text{oc}} \times \text{FF}}{P_{\text{light}}} \tag{7}$$

where FF is considered as fill factor, P_{max} is known as maximum power, V_{oc} is as open-circuit voltage, Isc is known as short circuit current. The standards were accepted from the curve of I-V characteristics. The plight is the input power of 100 mw/cm^2 beneath 1 illumination with sunlight intensity at 1.5 AM.

The photocurrent produced at the short circuit by a solar device under illumination is dependent on the light from the incident. To relate the density of the photocurrent, J_{sc}, to the spectrum of incidents, the quantum efficiency of the device is needed (QE). $QE(E)$ is the probability of one electron being transmitted to the external circuit by an incident photon of power E. The photocurrent can then be written in $QE(E)$

$$J_{\text{light}} = q \int F(E)QE(E)dE \tag{8}$$

The dark current density J_{dark} (V) varies for an ideal diode as

$$J_{\text{dark}}(\text{V}) = J_0\left(e^{qV/k_BT} - 1\right) \tag{9}$$

The sign convention in photovoltaics for current and voltage is such that the photocurrent is positive. That is the reverse of the standard for electronic devices commonly used. The net current density in the device is with this sign convention

$$J(V) = J_{\text{SC}} - J_{\text{dark}}(V) \tag{10}$$

Total current can be express as

$$J = J_{\text{light}} - J_{\text{dark}} - \frac{V}{AR_{\text{sh}}} \tag{11}$$

where,

V is the terminal voltage and A is the area of the device and R_{sh} is the shunt resistance.

And voltage can be calculated as

$$\text{Voltage} = V - J\,R_s A \tag{12}$$

where R_s is the series resistance.

3.1 SCAPS-1D as the Simulation Tool

For this work, SCAPS-1D has been used as a simulation tool.A solar cell capacitance simulator (SCAPS) is a one-dimensional solar cell simulation programme developed at the Department of Electronics and Information Systems (ELIS) of the University of Gent, Belgium [22]. Several researchers have contributed to its development: Alex Niemegeers, Marc Burgelman, Koen Decock, Stefaan Degrave, Johan Verschraegen. The original programme is developed for cell structures of the CuInSe2 and the CdTe family [5]. Recent developments make the programme now also, applicable to crystalline solar cells (Si and GaAs family) and amorphous cells (a-Si and micromorphous Si) [9]

4 Results and Discussion

The maximum efficiency and current density of the short circuit were calculated for devices with different thickness of the window layer when other parameters, like the thickness of the other surface, were kept constant. To observe the effect of the anti-reflective coating simulation performed for with or without ARC. Both data are included in Table 2 for comparison between AlGaAs window layer system and GaP window layer system.

Table 2 Simulation results for anti-reflective coating, p^+ and n^+ layer

Anti-reflective coating (ARC)	Device layers	AlGaAs as the window layer		
		V_{oc} V	J_{sc} A/m^2	Max efficiency (%)
1	p^+-p-i-n-n^+	1.0054	445.1377	36.9481
0	p^+-p-i-n-n^+	0.9891	300.0228	24.3850
1	p^+-p-i-n	0.9880	407.1663	33.5399
1	p-i-n	0.9877	405.8243	33.4215
0	p-i-n	0.9734	273.5256	22.0723

ARC = 1 means present of anti-reflective coating, ARC = 0 means absent of anti-reflective coating

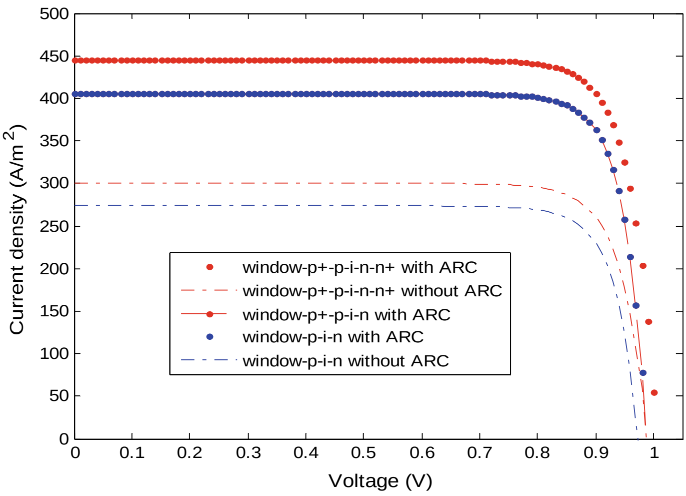


Fig. 2 V-I signifier of solar p-i-n devices with different extra layers

From this table, it is quite clear that there is a significant impact of this layer in p-i-n PV device. In the device of p^+-p-i-n-n^+, observed 36.948% efficiency with an anti-reflective coating. But without anti-reflective coating, it went down only to 24.3850% with 12% variation. In the device of p^+-p-i-n, 33.5399% efficiency observed with anti-reflective coating. The p-i-n device has 33.4215% efficiency with anti-reflective coating. But without anti-reflective coating, it shows 22.0723%, almost 11% varies with a layer.

Various curves are obtained from the simulation that is given below.

It can be observed from the curve using anti-reflective coating efficiency increase. This is due to minimize the reflection loss by ARC. It can also be observed from Figs. 2 and 3, that short circuit current and quantum efficiency is more for p^+-p-i-n-n^+ solar device than simple p-i-n PV device because both p^+ and n^+ layer reduce front and back surface recombination, respectively, so efficiency increase.

5 Conclusion

The p-i-n solar with SCAPS-1D solar photovoltaic device simulator is simulated here to observe the effect anti-reflecting coating over p-i-n solar device. The device was constructed using AlGaAs. GaAs was used as P^+, P, i, n^+, n-layer with dope changing. In p^+-p-i-n-n^+ an efficiency of 36.948% is observed in the anti-reflective coating and got 24.3850% without anti-reflective coating with 12% variation. In p^+-p-i-n observed 33.5399% efficiency with anti-reflective coating. In the p-i-n device, 33.4215% efficiency observed with an anti-reflective coating and 22.0723% without

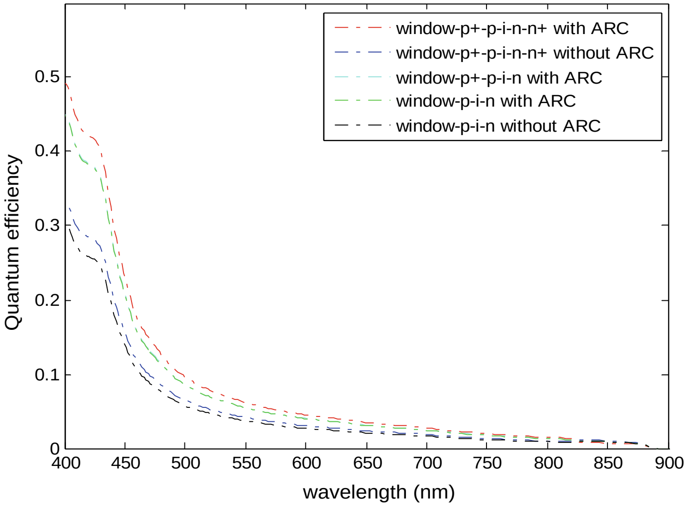


Fig. 3 Comparative analysis of p-i-n solar device's quantum performance with various additional layers

anti-reflective coating, almost 11% varies with a layer. So, it can be said that anti-reflecting coating reduces the loss and increases the η of a p-i-n PV device.

6 Limitation and Future Scope of This Study

Due to the lackings of fabrication instrument, this result can only be simulated. It is needed to fabricate in real life and check the validity if it matches with the simulation result or not. It is needed to check more materials for future increases in efficiency in term stability in solar energy. Lots more other 3–5 semiconductor materials are available like InSb and others which may give more efficiency with stability. Looking forward to checking them.

References

1. International Energy Outlook (2009) Energy Information Administration Office of Integrated Analysis and Forecasting U.S. Department of Energy. Retrieved online on 2009–06. http://www.jointsolarpanel.nl/fileadmin/jointsolarpanel/user/documents/seminar2004/stanleyzc04.pdf
2. Nelson J (2007) The physics of solar devices. Imperial College Press, London
3. Klausmeier-Brown ME, DeMoulin PD, Chuang HL, Lundstrom MS, Melloch MR, Tobin SP (1988) Influence of bandgap narrowing effects in p^+-GaAs on solar device performance. Confer Record Twentieth IEEE 1:503

4. Luque A, Hegedus S (2003) Handbook of photovoltaic science and engineering. Wiley, Hoboken
5. Kumakura K, Makimoto T, Kobayashi N (2000) Activation energy and electrical activity of Mg in Mg-doped $InxGa_{1-x}N$ ($x < 0.2$). Jpn J Appl Phys 39:L337
6. Emon IS, Ahmed SS, Milu SA, Mahtab SS (2019) Sentiment analysis of bengali online reviews written with english letter using machine learning approaches. In: Proceedings of the 6th international conference on networking, systems and security (NSysS '19). Association for Computing Machinery, New York, NY, USA, pp 109–115. doi: https://doi.org/10.1145/3362966.3362977
7. Milu SA et al (2020) Sentiment analysis of bengali reviews for data and knowledge engineering: a Bengali language processing approach. In: Bindhu V, Chen J, Tavares J (eds) International conference on communication, computing and electronics systems. Lecture Notes in Electrical Engineering, vol 637. Springer, Singapore. https://doi.org/10.1007/978-981-15-2612-1_8
8. Twidell J, Weir T (2006) Renewable energy resources. Taylor and Francis, Milton Park
9. Chakma R, Mahtab SS (2019) Navigation and tracking of AGV in ware house via wireless sensor network. In: 2019 IEEE 3rd international electrical and energy conference (CIEEC), Beijing, China, 2019, pp 1686–1690. https://doi.org/10.1109/cieec47146.2019.cieec-2019589
10. Trondheim, Modeling of intermediate band solar Devices. Department of Physics, Faculty of Natural Sciences and Technology, Norwegian University of Science and Technology, Norway
11. Smith LL, Davis RF, Kim MJ, Carpenter RW, Huang Y (1997) Microstructure, electrical properties, and thermal stability of Au-based Ohmic contacts to p-GaN. J Mater Res 12:2249
12. Khan MFS, Mahtab SS (2019) PLC based energy-efficient home automation system with smart task scheduling. In: 2019 IEEE sustainable power and energy conference (iSPEC), Beijing, China, 2019, pp 35–38. https://doi.org/10.1109/ispec48194.2019.8975223
13. Greenpeace and European Photovoltaic Industry Association (2008) Solar Generation V-2008, Solar electricity for one billion people and two million jobs by 2020
14. Vassilevski KV, Rastegaeva MG, Babanin AI, Nikitina IP, Dmitriev VA (1997) Ti/Ni Ohmic contacts to n-type GaN. Mater Sci Eng, B B43:292
15. Wen TC, Lee WI, Sheu JK, Chi GC (2001) Characterization of p-type InxGa1-xN grown by metalorganic chemical vapor deposition. Solid State Elect 45:427
16. Mahtab SS, Alam MJ (2020) Numerical modeling and simulation of high-efficiency thin Cu(In, Ga)Se photovoltaic by WxAMPS. In: Bindhu V, Chen J, Tavares J (eds) In: International conference on communication, computing and electronics systems. Lecture Notes in Electrical Engineering, vol 637. Springer, Singapore. https://doi.org/10.1007/978-981-15-2612-1_40
17. Mahtab SS, Monsur A, Ahmed SS, Chakma R, Alam MJ (2018) Design and optimization of perovskite solar cell with thin ZnO insulator layer as electron transport. In: 2018 international conference on advancement in electrical and electronic engineering (ICAEEE), Gazipur, Bangladesh, pp 1–4. https://doi.org/10.1109/icaeee.2018.8643012
18. Ahmed SS, Milu SA, Emon IS, Mahtab SS (2020) Opinion mining of bengali review written with english character using machine learning approaches. In: Bindhu V, Chen J, Tavares J (eds) International conference on communication, computing and electronics systems. Lecture Notes in Electrical Engineering, vol 637. Springer, Singapore. https://doi.org/10.1007/978-981-15-2612-1_5
19. Strite S, Morkoç H (1992) GaN, AlN, and InN: a review. J Vac Sci Technol B 10:1237
20. Kumakura K, Makimoto T, Kobayashi N (2000) High hole concentrations in Mgdoped InGaN grown by MOVPE. J Cryst Growth 221:267
21. Mori T, Kozawa T, Ohwaki T, Taga Y, Nagai S, Yamasaki S, Asami S, Shibata N, Koike M (1996) Schottky barriers and contact resistances on p-type GaN. Appl Phys Lett 69:3537
22. Yamasaki S, Asami A, Shibata N, Koike M, Manabe K, Tanaka T, Amano H, Akasaki I (1995) P-type conduction in Mg-doped Ga0.91In0.09 N grown by metalorganic vapor phase epitaxy. Appl Phys Lett 66:1112

Microstrip Antenna with SVASTU Slot for UWB Applications

Shivani Chourasia, Sudhir Kumar Sharma, and Pankaj Goswami

Abstract The svastu slot stuffed with a rectangle microstrip antenna (MA) through a microstrip line and examines its realization. It is also called as a printed antenna. It comprises of the four-sided mark with svastu form which is assembled with one postern of the substrate ($30 \times 35 \times 1.6$ mm^3) and restricted base plane (16×8 mm^2) on the other postern. The FR-4 sealant substrate is used, and MA is delineated through the HFSS simulation technique which helps in providing support to give rise high frequency of 10.5 GHz which is working with the rate of occurrences of 4.1–14.6 GHz with VSWR < 2, and whereas the features procreate the delineate MA which is advisable for multiple UWB applications.

Keywords High frequency structure simulator (HFSS) · Ultra-wideband (USB) · Svastu slot

1 Introduction

MA has manifold benefits like lightweight, popularly priced, lean profile, informal to outline area whereas it can cover entirely larger technologies such as aircraft, satellite as well as wireless conversation [1, 2]. The major drawback of this MA is precarious transmission capacity. Whereas several efforts have taken to improve and raise the transmission capacity of the MA, although there have been some procedures to implement the transmission capacity with the help of parasitic patches one of the two same different layers. The similar layer in case parasitical patches are showing on some dissimilar layer then the solidity of the antenna raises or if the particular is nonce on a similar layer, then the lateral magnitude of the antenna raises. In harmony

S. Chourasia (✉) · S. K. Sharma
Department of Electronics and Communication Engineering, Jaipur National University, Jaipur 116067, India
e-mail: shivanicp01@gmail.com

P. Goswami
Department of Electronics and Communication Engineering, Teerthankar Mahaveer University, Moradabad, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_21

to overcome this obstacle to evolve excellent layer patch transmission capacity MA [3].

UWB, a wireless communication (WC) automation, which holds the attention of massive transmission capacity of a minimum 20% of center rate of occurrence government correspondence commission, designated a transmission capacity of 7.5 GHz, i.e., in distinction to 3.1–10.6 GHz during UWB tolls [4, 5].

UWB engender by short span throbbing in the magnitude of picoseconds so radical wide cord will give very elevated data quota up to multiple hundred Mbps and aphoristic period throbbing also eschew multipath evanescent. UWB articulation systems have newly acknowledged great thinking in the cellular universe [6]. This a broadly used high tech in scanning system and frontier sensing tools. This transmission capacity satiates the system needs for S-DMB, WLAN as well as CMMB. UWB antenna is having less return loss (S11 < -10 dB), high emission productiveness over UWB against 3.1–10.6 GHz [7].

One of the transcendent demanding roles is to expand the UWB antenna for adaptable tools as well as the technology which gives a very trifling portion for antenna settling. In a compact four-sided MA is delineating where U pattern slot has been taken to develop the impedance transmission capacity. Also, it is a requirement of an antenna that has several bands as well as it exhilarates exigency for UWB tools.

The svastu slot brimming four-sided MA is delineating prosperously and outcome after facsimile fulfill transmission capacity needs for UWB tools. In this depiction, for a rate of occurrences range of 4.1–14.6 GHz, using this range the VSWR < 2, and S11 < -10 dB is realized. Its E- as well as H-sphere radiation molds are balanced over this rate of occurrences range [8, 9].

2 Microstrip Antenna

In the time of 1970s, different techniques have become excellent from over all the existence where the idea of MA is distinguished in the year 1953 and all the legitimate techniques like patents have been distributed in the time of 1955. It typically examines and comprises a lean strip that has been fixed on the head of the ground plane [10] which is depicted in Fig. 1.

The strip just as the ground sphere is separated by a dielectric membrane (intimate to as the substrate). There are different and various assorted substrates that can be taken into consideration for the outline of MA. There are distinct substrates that utilize the layout of MA, and their dialectic constraints are for the most part in the scope of $2.2 \leq \in r \leq 12$. The most beneficial is that these generally satisfy the receiving wire performance and are large substrates whose dielectric steady is in the optional finish of the range which is flexible in nature. Since they provide extremely bigger transmission limits and approximately headed fields for outflow into space, however, at the danger of bigger part [11, 12].

The transmission strip can be in any shape like might be square, four-sided, lean strip (dipole), round shape, egg molded, trilateral, or some other structure [13]. In

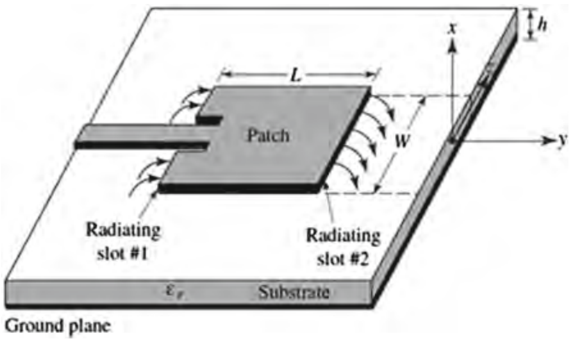


Fig. 1 Block diagram of microstrip antenna

the event that the more important demanding of MA, they are utilized in numerous advances where they are intensely required, for example, airplane, rocket, satellite, rocket instruments, where size, weight, cost, execution, simplicity of establishment, just as streamlined profile are limitation in the current situation, there are enormous quantities of government and business devices correspondingly portable radio just as remote innovation that have same necessity. To coordinate a similar degree of the requests, MA is comprehensively utilized and demanded [14–16].

3 Antenna Design

MA is delineated with the help of ANSOFT HFSS that is the highest rate of occurrence formative simulator. This is for forming MA with very limited and restricted components that have been used four-sided patched and the thickness of 11.964 mm as well as the length of 16 mm (Table 1).

The substrate is formed of FR-4 epoxy stuff having a diameter of 30 mm, length 35 mm, and compactness 1.6 mm. FR-4 epoxy is a dispensable nonconductor that is inexpensive as well as accessible. Which is having a stooge situation is 2.5 mm aside from situation which overlays an integrated ultra-wideband (Fig. 2).

Table 1 SVASTU dimension

S. No.	Parameter	Dimension	Material
1	Substrate	$W_{\text{sub}} = 30\,\text{mm} L_{\text{sub}} = 35\,\text{mm} H_{\text{sub}} = 1.6\,\text{mm}$	FR-4
2	Rectangular	$L_{\text{p}} = 16\,\text{mm} W_{\text{p}} = 11.964\,\text{mm}$	Copper
3	Ground plan	$W_{\text{g}} = 16\,\text{mm} L_{\text{g}} = 8\,\text{mm}$	Copper
4	Svastu slot	$6 * 1.5\,\text{mm}^2$ straight slot $L_{\text{S}} = 2.2\,\text{mm}$ $W_{\text{s}} = 1.5\,\text{mm}$	-
5	Feed line	$W_{\text{f}} = 3.01\,\text{mm} L_{\text{f}} = 8\,\text{mm}$	Copper

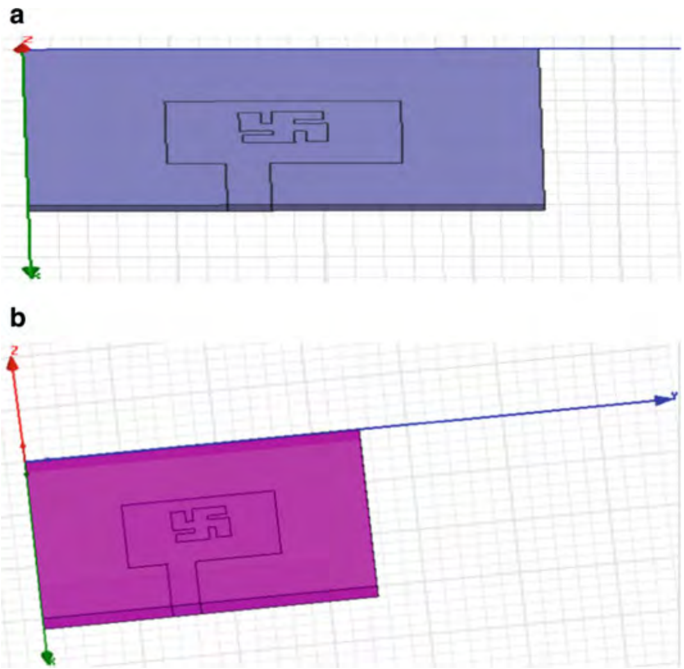


Fig. 2 **a** Svastu slot loaded MA. **b** Svastu slot loaded MA

4 Simulation Result

The outcome which is elucidating here is imitated on HFSS software. Circumstances that are used for overseeing substantial is inclination as skilled E furthermore for transmission box usual air reputation are fit which is elucidated in Fig. 3 which demonstrate reoccur loss vs. rate of occurrences plot.

Reoccur loss provides a quantity of power that has been reproduced separately input port. Whereas the UWB antenna is reoccurred detriment below −10 dB is treated to be precisely economical. Further, antenna layout reoccur detriment is lesser as compare −10 dB in the rate of occurrence limits 4.1–14.6 GHz.

The E-sphere is illustrated as the sphere which accommodates the electric field trajectory and pioneers the uttermost transmission meantime; the H-sphere is the sphere accommodating the magnetic field trajectory as well as the administration of altitude emission.

The x–z sphere upgrading sphere which is having some certain smooth angle φ is the principle E-sphere. Meantime for the x–y sphere, azimuth sphere, which is having a few specific altitude angle θ, is principle H-sphere and in Fig. 4 has elucidated 2-D *E*-sphere Radiation pattern (RP) at a unique managing rate of occurrence.

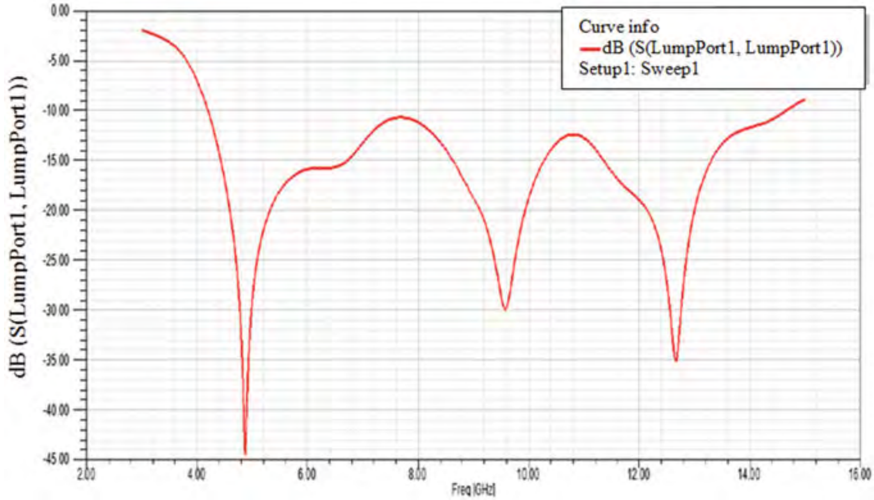


Fig. 3 Variation of Frequency vs Return loss (S11)

For the *x*–*y* sphere is a smooth plane with some specific levitation angle θ is criterion H-sphere which has been elucidated in Fig. 5, 2-D H-plane emission design at a dissimilar operating rate of occurrence (Fig. 6).

Figure 7 displays the voltage standing wave ratio (VSWR) plot as compared to the rate of occurrence in GHz. VSWR is directly the ratio of crest amplitude of eminence twirl to the littlest amplitude of eminence twirl. VSWR below 2 is thought hardy for an antenna. For this delineation, VSWR is lesser than 2 take-ups away 4.1–14.6 GHz.

5 Conclusion

A rectangular microstrip antenna is advisable for UWB tools that is delineation. It shows a transmission capacity of 112.31% (4.1–14.6 GHz, centralize at 9.35 GHz). The suggested outlet of this antenna could be used for a multiple of UWB tools in addition to greater speed data transfers, wireless connectivity in the middle of UWB-enabled equipment, and diversification of medical tools.

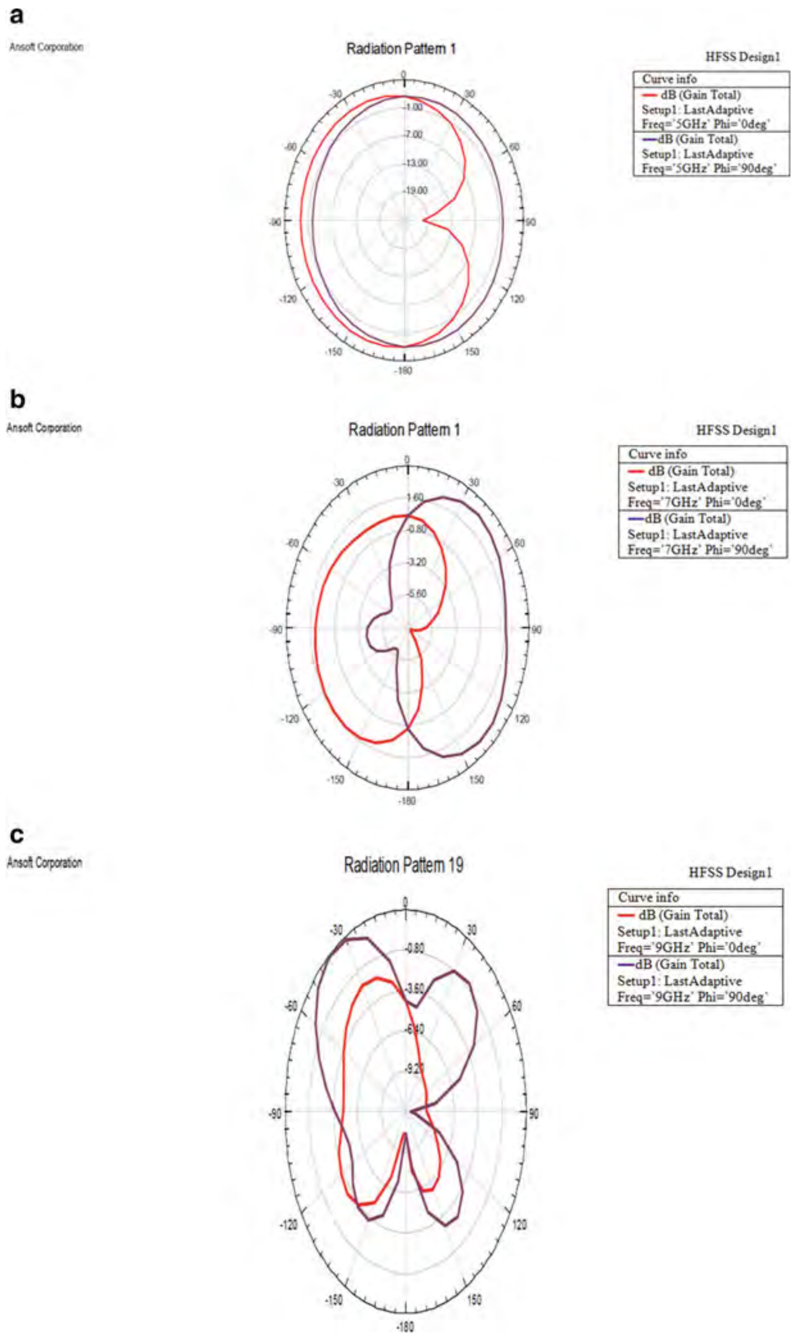
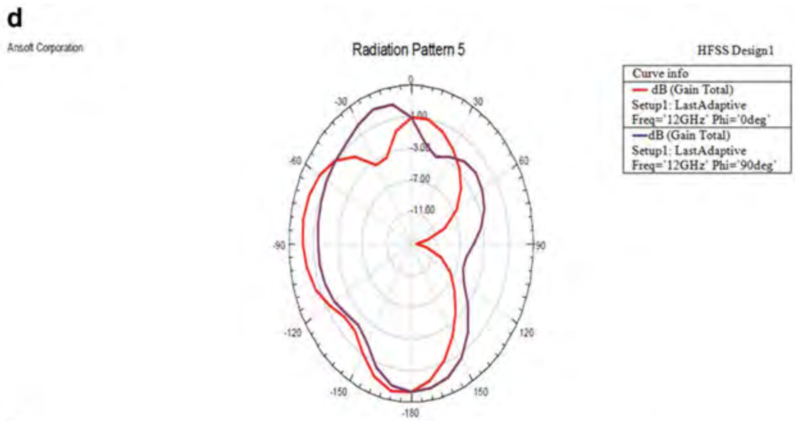


Fig. 4 **a** 2-D E-plane RP at 5 GHz. **b** 2-D E-plane RD at 7 GHz. **c** 2-D E-plane RD at 9 GHz. **d** 2-D E-plane RD at 12 GHz. **e** 2-D E-plane RD at 12 GHz



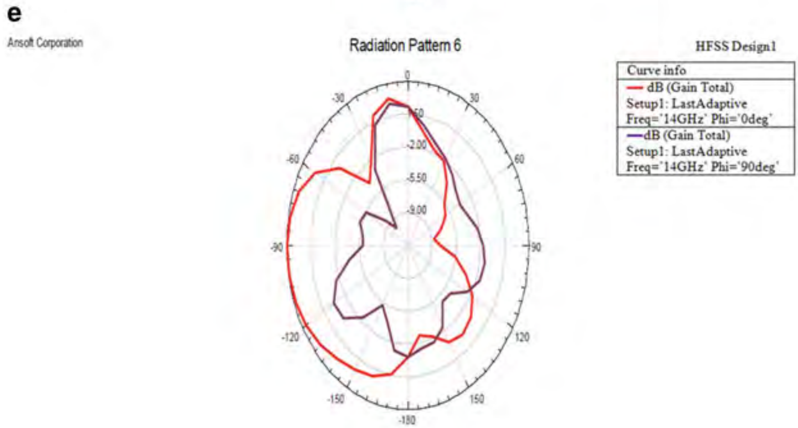


Fig. 4 (continued)

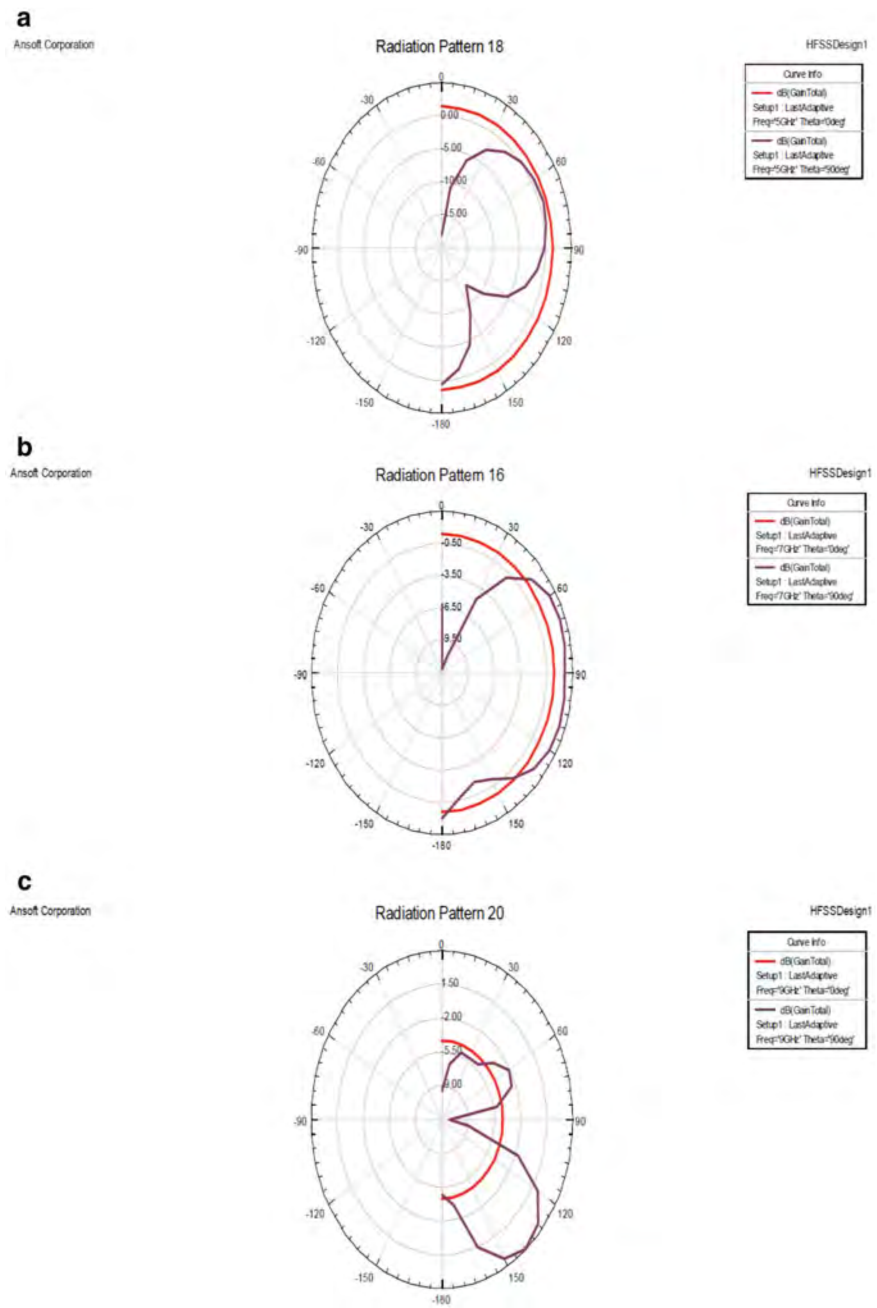
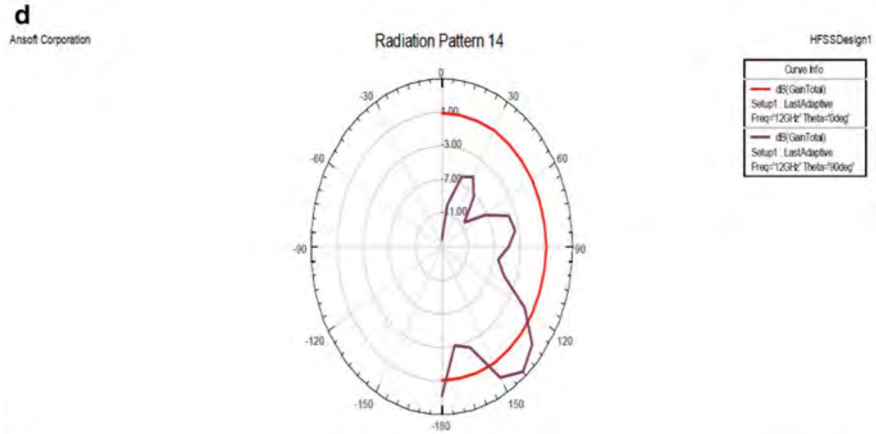


Fig. 5 **a** 2-D H-plane RD at 5 GHz. **b** 2-D H-plane RD at 7 GHz. **c** 2-D H-plane RD at 9 GHz. **d** 2-D H-plane RD at 12 GHz. **e** 2-D H-plane RD at 14 GHz



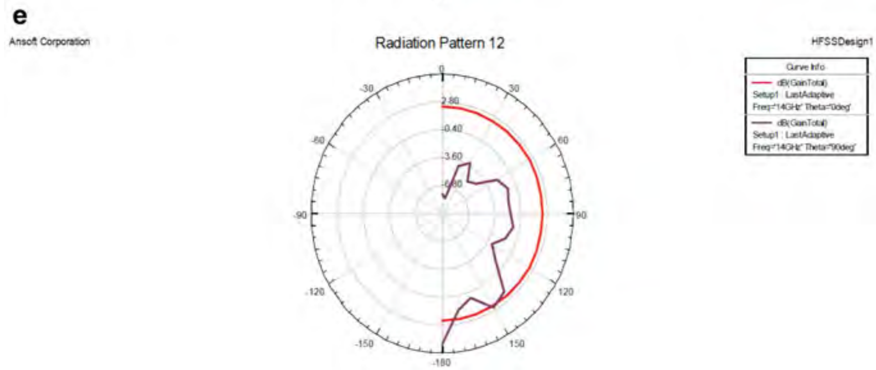


Fig. 5 (continued)

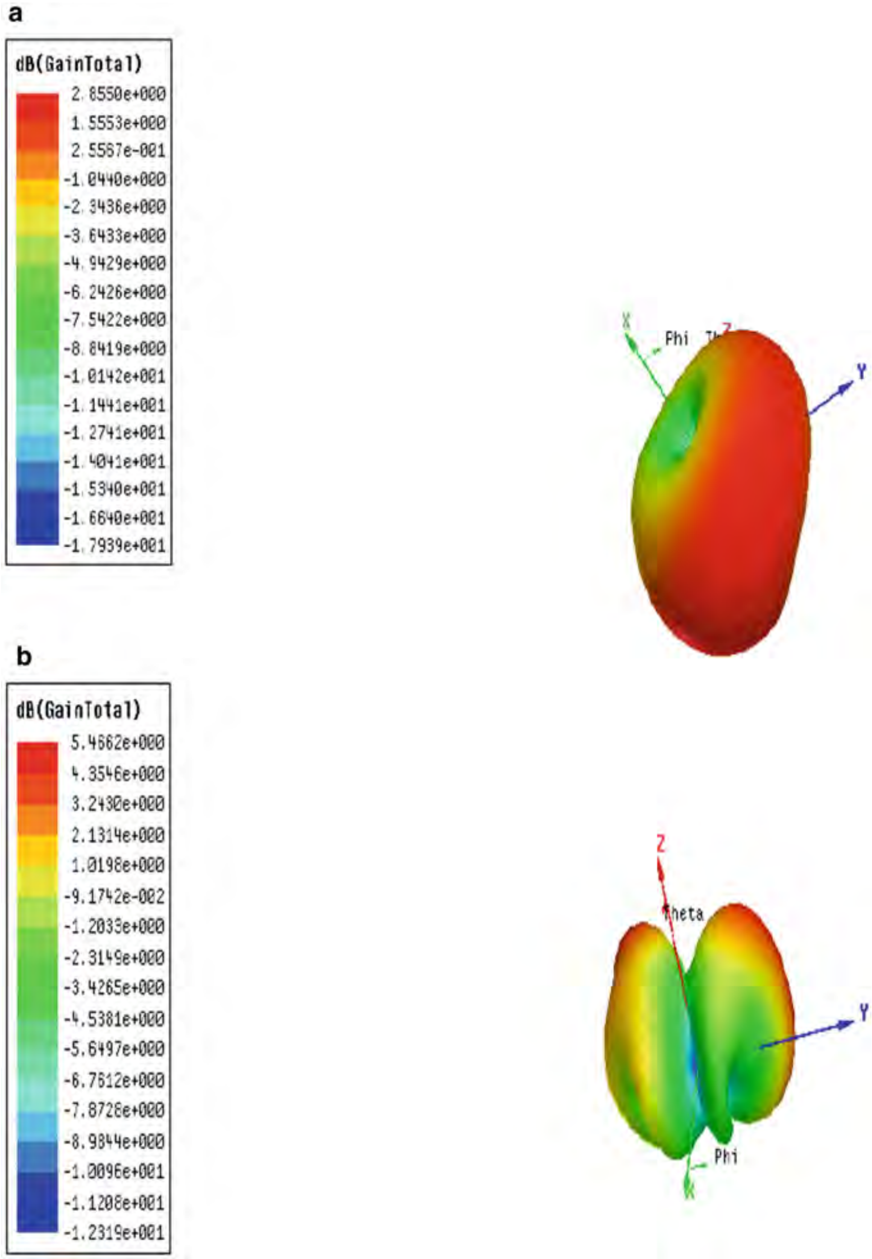


Fig. 6 **a**3-D RD at 5 GHz. **b** 3-D RD at 10 GHz

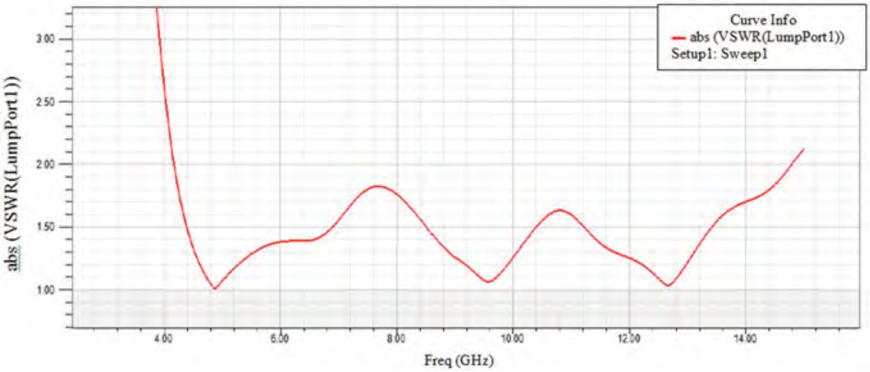


Fig. 7 Variation of Frequency vs VSWR

References

1. Sharma A, VishwakarmaRK (2014) Microstrip antenna with Swastik slot for UWB applications.In: 2014 IEEE students' conference on electrical, electronics and computer science, pp 1–5. IEEE
2. Prakasam V, Sandeep P (2017) Mode patterns in rectangular waveguide. Int J Trend Res Dev (IJTRD)
3. Luo J, Chi L, Li C, Sun B (2018) Side edge frame printed eight port dual band antenna array for 5G smart phone applications. IEEE Trans Antennas Propag 66(12):7412–7417
4. Sun L, Fang H, Li Y, Zhang Z (2018) Compact 5G mobile phone antennas with tightly arranged orthogonal mode pairs. IEEE Trans Antennas Propag 66(11):6364–6369
5. Porchan NO, Al- YIA, Ali AH (2019) Eight element dual polarized MIMO slot antenna system for 5G smart phone applications. IEEE Access 7:15612–15622
6. Ayyappan M, Chadran J (2016) Design and analysis of circular micro-strip antenna at 2.4 GHz with Fr-4 substrate. Int J Adv Res Electr Electron InstrumEng 5(4)
7. Prakasam V, Sandeep P (2018) Design and analysis of 2×2 circular micro-strip patch antenna array for 2.4 GHZ wireless communication applications. Int J InnovEng Manage Res 7(12)
8. Naik KK (2018) Asymmetric CPW-fed SRR patch antenna for WLAN/WiMAX applications. Int J Electron Commun (AEÜ) 93:103–108
9. Checkatla AR, Ashtankar S (2019) Compact microstrip antenna for 5G mobile phone application. Int J Appl Eng Res 14(2):108–111
10. Fallahpour M, Zoughi R (2018) Antenna miniaturization techniques: a review of topology- and material-based methods. Antennas Propag Magaz IEEE 60(1):38–50
11. Mohammad HA, Mohammad U, Chughtai T, Nasir J (2018) Cross polarized 2×2 UWB-MIMO antenna system for 5G wireless applications. PIER-M 76(5):157–166
12. Wang Y, Zhu L, Wang H, Yang G (2018) Design of compact wideband meandering loop antenna with a monopole feed for wireless applications. PIER Lett 73:1–8
13. Chen T, Li ZN (2019) Shared surface dual band antenna for 5G application. IEEE Trans Antennas Propag 1–1
14. Khan R, Abdullah A, Soh PJ (2018) User influence on mobile terminal antennas: a review of challenges and potential solutions for 5G antenna. IEEE Access 6:77695–77715
15. Haun C, Kuster Li N, Ofli E (2007) The effect of hand phantom on mobile phone antenna OTA performance. EuCAP Edinburgh
16. Jain S, Tomar PS, Tomar GS (2012) Design & analysis of proximity fed circular disk patch antenna. Int J Emerg Technol Adv Eng 2(10):126–131

Design and Implementation of a Smart Helmet System for Underground Miner's Safety

S. M. Minhajul Alam, Arnob Barua, Ahamed Raihan, M. J. Alam, Rocky Chakma, S. S. Mahtab, and Chitra Biswas

Abstract In industrial applications, safety is determined as a crucial theme. With time, many solutions are used to save labors from traumatic events or constructional failures. A wearable helmet is exhibited in this study, warns leaking toxic gas in a large amount. This study is also aid for petro-chemical to save the lives as well as the safety of workers. This prototype provides real-time monitoring of harmful gases, temperature, humidity, and worker's heart-rate. To overcome the hazardous situations, this system provides an emergency alarm for the monitoring station. With the aid of unified sensors, the helmet is adroit to execute nearby observing actions of the labor and rapidly forwards alarm regarding the un-avoidable collision. This system is very user-friendly and cost-effective.

Keywords Internet of things (IoT) · Radio frequency (RF) · GPS/GPRS

S. M. Minhajul Alam · A. Barua · R. Chakma · C. Biswas
Department of EEE, USTC, Chittagong, Bangladesh
e-mail: alameee1993@gmail.com

A. Barua
e-mail: arnob303@gmail.com

R. Chakma
e-mail: rocky.cht@gmail.com

C. Biswas
e-mail: chitra.biswas86@gmail.com

A. Raihan
Department of CSE, Jahangirnagar University, Savar, Dhaka, Bangladesh
e-mail: araihan13@gmail.com

M. J. Alam (✉) · S. S. Mahtab
Department of EEE, Feni University, Trunk Road, Feni 3900, Bangladesh
e-mail: alameee1993@gmail.com

S. S. Mahtab
e-mail: mahtabshahzad@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_22

1 Introduction

In this system, a safety helmet is made for the worker, rescuer, and miner. The natural condition of the current situation like humidity, temperature, heart-rate, and harmful gas detection can be detected by radio frequency (RF) communication. The position of a worker can be found using GPS/GPRS. The helmet's sensor senses the data from weather and the human body [6]. Then, this data is sent to the monitoring module by RF transmitter. After that, the monitoring section can receive all the data from the RF receiver and display.

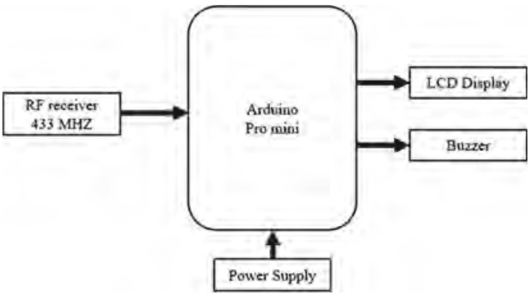
The demand for coal as energy resources is always important and significant. Nevertheless, thousands of people have forfeited their breath in mining accidents, all over the world. In their article, Jing change, Qinggui Cao and Yonjige Yang listed 100 of the major mining accident which had taken place from 2001 to 2010 [1]. As most of the coal mines in the North East region of the country are still in a primitive state, the mining accidents here are also very frequent. The main reason is these accidents occur due to the presence of methane and carbon monoxide gas in theses mines. These gasses are colorless, odorless and are undetectable by human sensors [2]. Such kind of accident can be controlled in the prophecy of the explosion process by executing microcontrollers and sensors and to develop a siren scheme before demanding climatic level. A steady supervising is important which further recommends some efficient and solid sensing method. To sense the existence poisonous gas, several techniques can be accommodated, using semiconductor type gas is much more effective among the techniques. The advantage of these sensors is, can be mounted in the coal mine location [3]. Apart from this, some disadvantages also can be calculated at the time of mining. The damage of the sensor apparatus from an accident often took place. Using robot can be another fruitful technique [4]. In no doubt, these kind of robots are good to use but for a country like India where industrialists are not much concerned about the safety of the workers, a robot cannot be imagined. However, considering cost-effectiveness and some other factors, safety helmet technology can be the further solution for the coal mine workers, whereas a smart safety helmet with updated sensor array will be presented to sense data and also a wireless modem will be available to transmit the data [5].

2 Literature Review

A helmet is an embodiment of defensive equipment on the head to shield it from wounds. In 900 BC, Assyrian soldiers used a helmet which is considered the oldest one, he wore bronze or thick leather helmets to shield the head from the unsharpened body and arrow strikes and sword blows in war. In civilian life, helmets are also used for enjoyment activities, transportation, and ports. On the other hand, different types of the innovative helmet can be seen in the modern world [7]. Among these, Baseball batting helmets is one which has a bolster safety over the ear and it saves the mouth and jaw from the wound. For safety from wind and rain, motorcycle helmets generally have flip-down face screens, and also a projecting visor is randomly used to shield the eyes from glare [8]. Advance technology can be seen in modern firefighter's helmets where include communication systems, masks, and other accessories which save the face and back of the head from accident, electricity, and fires. There is another helmet named mixed martial arts helmets where a pair of ear pads are included to restrict serious injuries to the players, those who do not commonly consume such force to the ears [9].

To upgrade safety, comfort, and current innovations for the target market, Brembo helmets are devised, these helmets are in the new varieties of automatic fit belt fastening of the helmet and the configuration of the visors. Due to hot weather and humid condition, forest workers will remove the head shields. But, it is essential for the workers to wear helmet during work which is claimed by occupational safety and health administration (OSHA) in consequence to face a compliance difficulty [10]. To identify which features, provide to forest labors' thermal embarrassment, this research assessed subjects' psychophysical and physiological reactions at the time of tasks estimating the assignment of forest labor in a large amount of temperature condition are observed throughout the summer in southeastern united states. In this study, three more helmets assessments were practiced: (a) a standard helmet, (b) an actively ventilated helmet, and (c) a passively ventilated helmet. It was observed that none of the physiological variables are examined for the body loaded with helmets [11]. Also, it is observed that dry bulb temperature and wet bulb temperature are differed in the tested helmets. psychophysical outcomes appears as air circulation will commits as a prominent suitable helmet and its fit, and weight are significant features in helmet pattern which is much important to factories and industries. Protective helmets for utilizing in humid and hot atmospheres must be altered to be convenient and motivating forest labors to wear and obey with OSHA regulations. 2001 Elsevier Science B.V. All rights set aside. In consequence, the head substantially regulates heat transfer and performs a crucial role in governing comprehensive individual comfort. Thermal comfort has suited progressively vital in designing different kinds of helmets (safety, firefighter, bicycle, etc.) as it can importantly ameliorate users' safety and health. Attaining thermal comfort implies considering multiplex airflow, moisture movement, and heat transfer [12]. Beginning evaluation of thermal properties and comfort of helmets disseminate the subsequent similarities between industrial safety helmets and cricket: the purposed of the these helmets is to save the user head [1] in case of shortest impact with a thing; both cricket players and industrial labors [2] are displayed to an utmost environmental situation for a prolonged spell of time; most of the users [3] noticed that most of the convention helmets have more weight, and it produces uncomfort to wear due to poor ventilation. International cricket environments have moderate airflow and to analyze the thermal distributions for helmets is accepted as <0.8 ms À1 which is tested under laboratory. Considerably, the helmet design used for cricket and motorcycle are much similar in construction and the appearance will control the air flow in the helmets [13]. Existing research models confines that position of vents increases the airflow and reduces the heat. Usage of dissimilar polymeric materials such as acrylonitrile butadiene styrene, Polycarbonates and polyethylene will be a suitable choice for industry helmets.

3 Hardware Block Diagram of Helmet



Two basic elements of system block diagrams are considered for this project. In addition, the receiver and transmitter block diagram are suggested for this approach.

3.1 Block Diagram of Transmitter

An electronic circuit with a switched-mode power supply (SMPS) transmuted power operating switching appliance that is turned on and off at high frequencies, and elements stored inside such as capacitors or inductors to transmit power during the switching apparatus is in its non-conduction condition. Switching power supplies have excessive efficiency and are extensively utilized in a difference of electronic appliance, incorporating computers, and other sensitive appliance demanding steady and effective power supply. A switched-mode power supply is also familiar as a switching-mode power supply (Figs. 1 and 2).

MQ2 gas sensor performs on 5 V DC and takes out approximately 800mW. It also can recognize LPG, alcohol, propane smoke, hydrogen, methane, and carbon monoxide concentrations everywhere from 200 to 10,000 ppm. When semiconductor particles such as tin dioxide are warmed in the air at overpriced temperature, oxygen is sucked up on the exterior. In scrubbed air, donor electrons in tin dioxide are engaged about oxygen which is sucked up on the exterior of the sensing substances. This obstructs electric current pass. In the appearance of condensing gases, the exterior density of adsorbed oxygen attenuates as it behaves with the condensing gases. Electrons are subsequently

delivered into the tin dioxide, authorizing current to pass openly direct to the sensor. In this project, a prominent network technology (GSM) is applied to utilize for

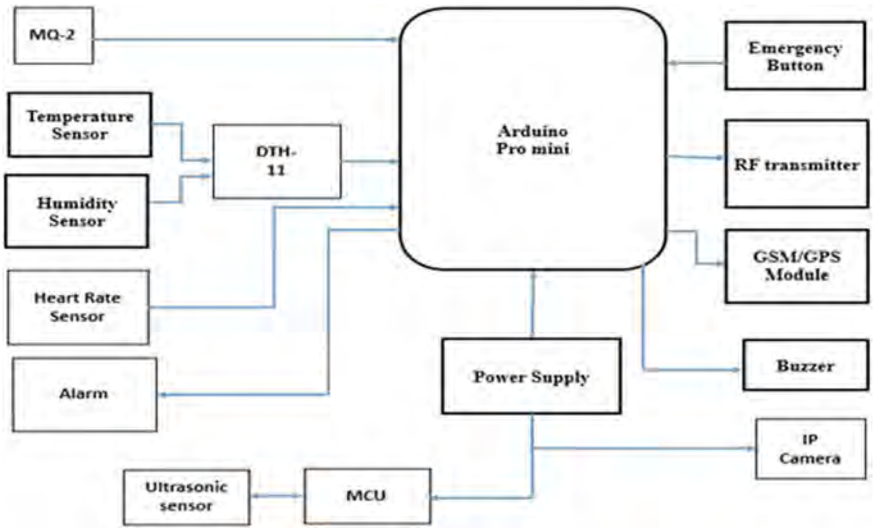


Fig. 1 Transmitter

Fig. 2 Receiver

transmission of SMS from sender to receiver. SMS sending and receiving is generally manipulated for omnipresent access of device and permitting breach control at here and there. The security alert network accords self-executing security observation. The system is efficient adequate to command user via SMS from a particular cell number to wide awake the circumstances of the industry machine as per to the user's demands and requirements. The second feature is that of security alert which is attained in a method that on the observation of intrusion, the system permits self-executing creation of SMS, therefore, warning the user against security risk [2].

4 Hardware Desıgn

Here, a gas sensor, humidity, and temperature sensor, Arduino, and a buzzer have been used for emergency purpose. Our system includes an emergency sensing switch. The gas sensor is operated to recognize the weather bad conditions, the achievement is fed to the Arduino. The gas sensor and switch both are placed accurately in the helmet. Based on the gas sensors, humidity, and temperature sensor concentration gas sensors, and the humidity and temperature sensor provides an analog resistive result. Arduino unit, which controls all the functions of other blocks in this system. Indeed, Arduino accepts or reads data comes from the sensors and governs all the tasks of the entire system by utilizing these data. Arduino secures data from these sensors and it represents a digital data wireless tie in with to the output receiver edge (Figs. 3, 4 and 5).

4.1 System Result

When the worker presses the emergency button, the system will send the message to the monitoring section. A monitor will have a message and the worker's location (Latitude and Longitude) value, Google map includes this message. Then, the rescue team will find out and help the worker.

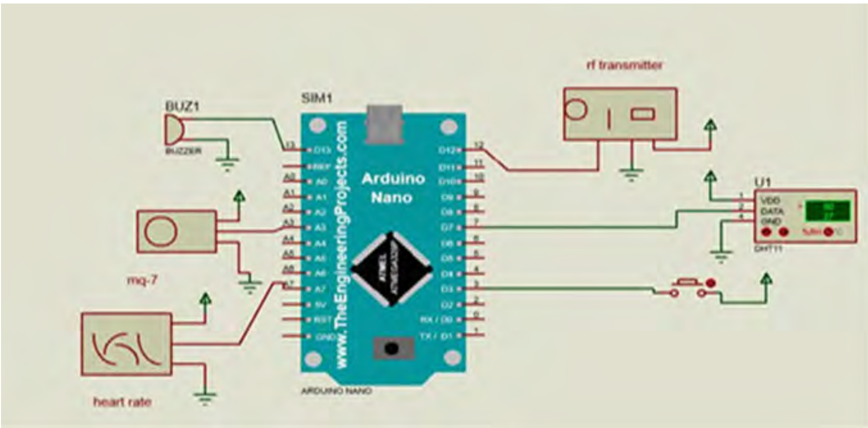


Fig. 3 Circuit diagram for sending

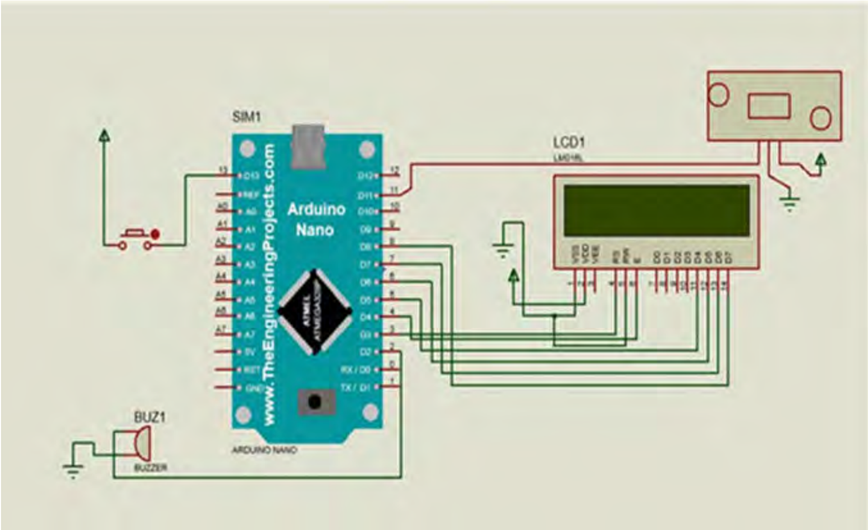


Fig. 4 Circuit diagram for receiving

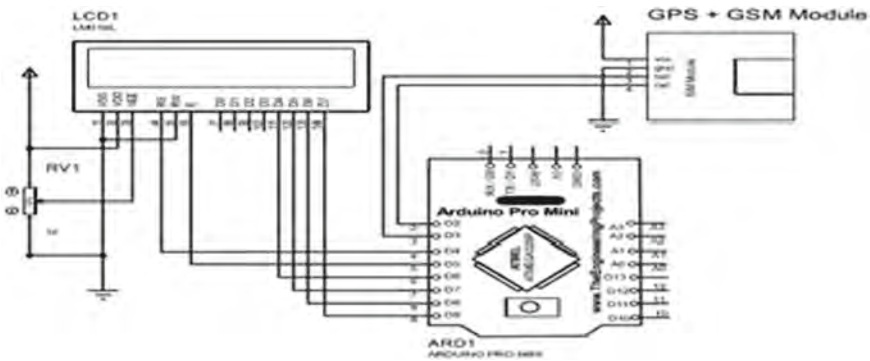


Fig. 5 Arduino accepts or reads data from Sensor

5 Result Analysis

This project gives an overview of the project introduction, overview, features, aims and objectives, scope, problem definition, background, and operation environment of the system. Here, the design (Fig. 6) and the circuit (Fig. 8) used to operate the system is shown below (Figs. 7 and 9).

Fig. 6 System send the message

6 Conclusion and Future Scope

A smart helmet band have been sufficiently devised in this project utilizing both smart technology GPS and GSM. The project led mandatory of put on a helmet to commence the flaming of a motor vehicle and interval of driving. If any unanticipated transformation in acceleration takes place afterwards, an accelerometer will supervise the transformation and a direct message with the position of the driver will be posted to the predetermined number utilizing GSM constituent. However, GPS/GSM constituent can't be able to address labor's particular position longitude and latitude value. This is a condition where a few solutions are discovered to the issue of expanded death percentage. In the future, the proposed system can connect the sensor's values to the online server with the internet of things technology.

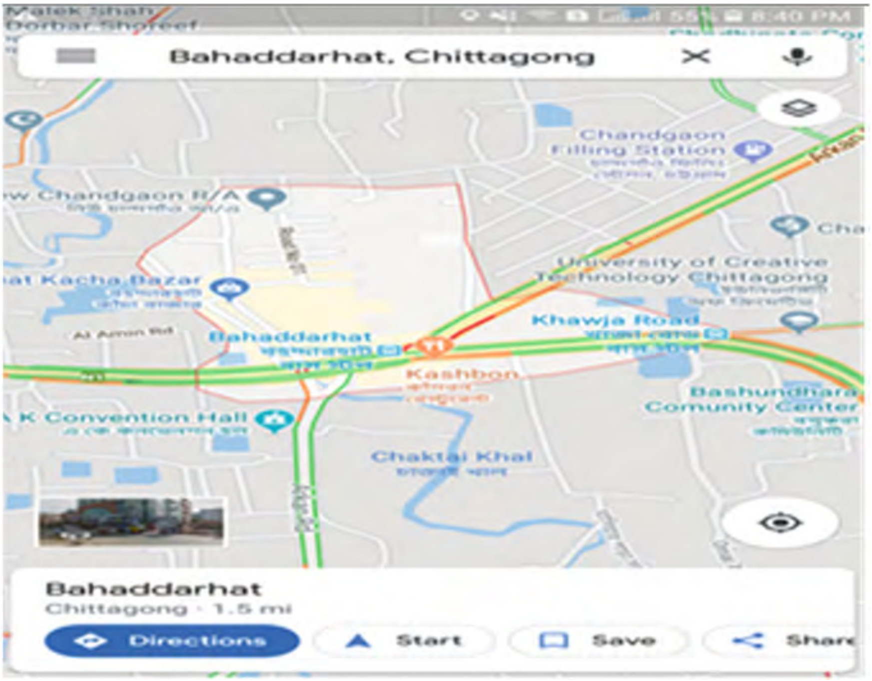


Fig. 7 Google map

Fig. 8 Sensor value

Fig. 9 Physical overview of the project

References

1. Khan MFS et al (2019) PLC based energy-efficient home automation system with smart task scheduling. In: IEEE sustainable power and energy conference (iSPEC), Beijing, China, pp 35–38
2. Chakma R et al (2019) Navigation and tracking of AGV in ware house via Wireless Sensor Network. 2019 IEEE 3rd international electrical and energy conference (CIEEC), Beijing, China, pp 1686–1690
3. Mahtab SS, Alam MJ, Khan AM, Uddm Z, Mamun AA, Uddin MM (2017)Optimization of InSb QWFET layer structure for high-speed and low power nano electronics applications. In: 2017 4th international conference on advances in electrical engineering (ICAEE), Dhaka, pp 707–711. IEEE. doi: https://doi.org/10.1109/ICAEE.2017.8255447
4. Mahtab SS, Monsur A, Ahmed SS, Chakma R, Alam MJ (2018) Design and optimization of perovskite solar cell with thin ZnO insulator layer as electron transport. In: 2018 international conference on advancement in electrical and electronic engineering (ICAEEE), Gazipur, Bangladesh, pp 1–4. IEEE. doi: https://doi.org/10.1109/ICAEEE.2018.8643012
5. Mahtab SS, Hossain MA (2019) Efficient and stable perovskite solar cell with TiO_2 thin insulator layer as electron transport. In: 2019 international conference on robotics, electrical and signal processing techniques (ICREST), Dhaka, Bangladesh, 2019, pp 54–58. IEEE. doi: https://doi.org/10.1109/ICREST.2019.8644093
6. Chakma R, Mahtab SS (2019) Navigation and tracking of AGV in ware house via wireless sensor network. In: 2019 IEEE 3rd international electrical and energy conference (CIEEC), Beijing, China, pp 1686–1690. IEEE. doi: https://doi.org/10.1109/CIEEC47146.2019.CIEEC-2019589
7. Khan MFS, Mahtab SS (2019) PLC based energy-efficient home automation system with smart task scheduling. In: 2019 IEEE sustainable power and energy conference (iSPEC), Beijing,

China, pp 35–38. IEEE. doi: https://doi.org/10.1109/iSPEC48194.2019.8975223

8. Emon IS, Ahmed SS, Milu SA, Mahtab SS (2019) Sentiment analysis of Bengali online reviews written with english letter using machine learning approaches. In: Proceedings of the 6th international conference on networking, systems and security (NSysS '19). Association for Computing Machinery, New York, NY, USA, pp 109–115. doi: https://doi.org/https://doi.org/10.1145/3362966.3362977
9. Milu SA et al (2020) Sentiment analysis of Bengali reviews for data and knowledge engineering: a Bengali language processing approach. In: Bindhu V, Chen J, Tavares J (eds) International conference on communication, computing and electronics systems. Lecture Notes in Electrical Engineering, vol 637. Springer, Singapore. https://doi.org/https://doi.org/10.1007/978-981-15-2612-1_8
10. Ahmed SS, Milu SA, Emon IS, Mahtab SS (2020) Opinion mining of Bengali review written with English character using machine learning approaches. In: Bindhu V, Chen J, Tavares J (eds) International conference on communication, computing and electronics systems. Lecture Notes in Electrical Engineering, vol 637. Springer, Singapore. https://doi.org/https://doi.org/10.1007/978-981-15-2612-1_5
11. Buurat J, Mahtab SS, Milu SA, Emon IS (2020) An automated Bengali text summarization technique using Lexicon based approach. In: International conference on innovations in computer science and engineering (ICICSE-2020). Springer "Lecture Notes in Networks and Systems"
12. Buurat J, Mahtab SS, Milu SA, Emon IS (2020) A pronoun replacement based special tagging system for Bengali language processing (BLP) In: International conference on innovations in computer science and engineering (ICICSE-2020). Springer "Lecture Notes in Networks and Systems"
13. Hoque F, Chakma R, Mahtab SS, Akter R, Ahmed SS. Design and developing real time interactive IIUC bus tracking system. J Innov Comput Sci Eng (JICSE) with ISSN 2278-0947 in vol 9(2), vol 10(1)

A Brief Review on Instance Selection Based on Condensed Nearest Neighbors for Data Classification Tasks

Yasmany Fernández-Fernández, Diego H. Peluffo-Ordóñez, Ana C. Umaquinga-Criollo, Leandro L. Lorente-Leyva, and Elia N. Cabrera-Alvarez

Abstract The condensed nearest neighbor (CNN) classifier is one of the techniques used and known to perform recognition tasks. It has also proven to be one of the most interesting algorithms in the field of data mining despite its simplicity. However, CNN suffers from several drawbacks, such as high storage requirements and low noise tolerance. One of the characteristics of CNN is that it focuses on the selection of prototypes, which consists of reducing the set of training data. One of the goals of CNN seeks to achieve the reduction of information in such a way that the reduced information can represent large amounts of data to exercise decision-making on them. This paper mentions some of the most recent contributions to CNN-based unsupervised algorithms in a review that builds on the mathematical principles of condensed methods.

Keywords Prototypes · Nearest neighbor algorithms · Classification

Y. Fernández-Fernández (✉)
Universidad Politécnica Estatal del Carchi, Tulcán, Ecuador
e-mail: yasmany.fernandez@upec.edu.ec

Y. Fernández-Fernández · D. H. Peluffo-Ordóñez · A. C. Umaquinga-Criollo · L. L. Lorente-Leyva
SDAS Research Group, Ibarra, Ecuador
e-mail: dpeluffo@yachaytech.edu.ec

A. C. Umaquinga-Criollo
e-mail: acumaquinga@utn.edu.ec

L. L. Lorente-Leyva
e-mail: leandro.lorente@sdas-group.com

D. H. Peluffo-Ordóñez
Yachay Tech University, Urcuquí, Ecuador

Corporación Universitaria Autónoma de Nariño, Pasto, Colombia

A. C. Umaquinga-Criollo
Universidad Técnica del Norte, Ibarra, Ecuador

E. N. Cabrera-Alvarez
Universidad de Cienfuegos, Cienfuegos, Cuba
e-mail: elita@ucf.edu.cu

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_23

1 Introduction

Instance selection methods represent an important approach in different areas of data science. In [1], some important elements are considered in topics related to instance selection. There are two important processes, namely training set selection and prototype selection.

The selection of a subset of data for another very large data set is summarized in the concept of "data condensation" [2]. This form of data reduction differs from the others and is integrated as one of the families of the instance selection methods. Mainly, data condensation approaches are studied based on the classification processes, particularly the k-nearest neighbor (KNN) methods which refer to obtain a consistent minimum set that classifies the entire original set. Figure 1 shows a simple representation of the KNN.

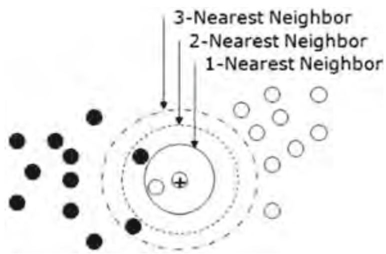
One of the first pioneering methods in the analysis in the data structure for the selection of instances was CNN [4]. The methods of condensation of data that are not related to the classification process are also known as methods of condensation of generic data, such condensation is performed through the so-called vector quantization (VQ), and example of this is the self-organization map and other ways of organizing the data as shown in Fig. 2.

1.1 Vector Quantization

Vector quantization (VQ) is a classic method that consists of approximating a continuous probability density function $p\ (x)$ of the vector input variable x by using a finite number of book-encoded vectors mi, $i = 1, 2, \ldots, k$; once these book-encoders have been chosen, the approximation of x implies finding the reference vector closest to x. An optimal location type of m minimizes to E where E is the rth power of the reconstruction error [5]:

$$E = \int \|x - m_c\|^r p(x)\mathrm{d}x \quad (1)$$

Fig. 1 K-nearest neighbor representation [3]



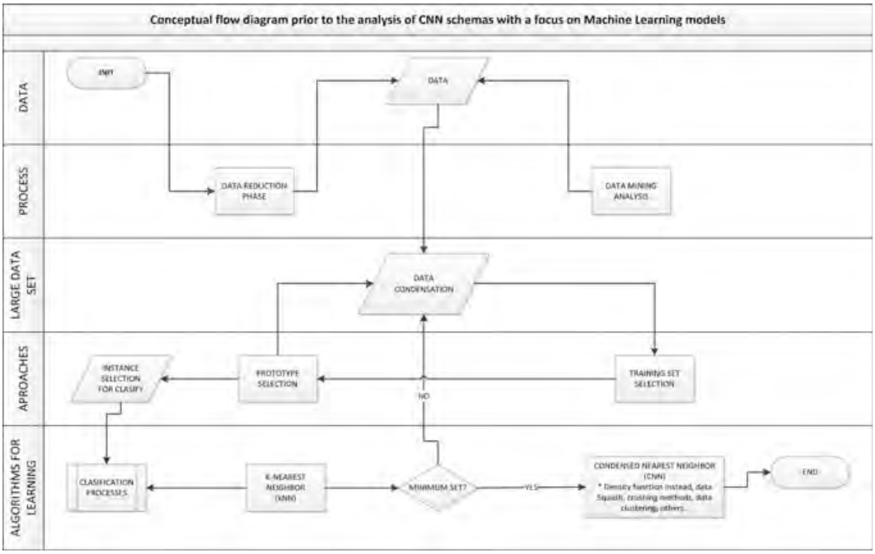


Fig. 2 CNN decision diagram for data reduction task

where dx represents the differential volume in the space x and the index $c = c\,(x)$ of the best match between the book-encoders (winner) is a function of the input vector:

$$\|x - m_c\| = \underbrace{\min}_{i}\{\|x - m_i\|\} \tag{2}$$

In general, a closed solution for the optimal location of m is not possible, so iterative approximation schemes can be used.

1.2 Condensed Methods

Generic data condensation methods are based on techniques that consider density; they consider the density function instead of minimizing the quantification error; that is, for a specific input set, the condensed output set [6] is established.

Other methods such as data squash or data clustering are used for sample selection. A crushing method seeks the compression of the data in such a way that a statistical analysis performed on the compressed data obtains the same result as with the original data. Clustering-based algorithms [7, 8] divide data into samples like each other and different from examples of data belonging to other groups [1].

Figure 3 is represented according to a distance function where the quality of the cluster could be measured according to the dimension of its diameter which is the maximum between two samples belonging to the same group.

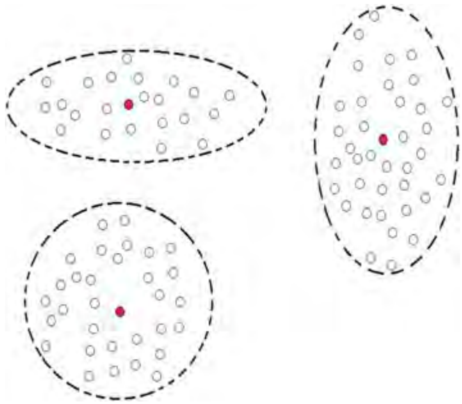


Fig. 3 Three clusters obtained from a set of two-dimensional data

1.3 Machine Learning and Feature Selection

In machine learning, a process known as feature selection consists of the selection of characteristics, attributes or selection of variable subsets for use in model building. In [2], two feature selection strategies are mentioned, the first based on feature ranking and the other based on best subset selection. In the case of the methods based on feature ranking, some statistical metrics are used, some of the simple complexity uses the correlation coefficient instead of other more complex used methods such as the Gini index, and this index can be used to quantify inequalities in variable distributions. Other feature ranking methods mentioned in the literature [9] are the bivariate and multivariate methods; these methods calculate the distance between the actual joint distributions of the characteristics of two or more variables and answer the question of what the joint distribution would be if these variables were independent, further. The joint distribution represents the probability distribution of existing case studies. Among the multivariate analysis, methods are the stepwise linear regression [10, 11] which has been used in cluster tasks and sample selection [12]; other slightly more complex algorithms include the use of machine learning and advanced statistics, for example, partial least squares regression [13] and sensitivity analysis [14]. Also, in performance analysis of virtual clusters [15] and architecture in wireless networks [16].

The second strategy based on subset selection has its focus on the selection of a subset for the selection of characteristics or attributes that have a significant effect on the prediction of a variable. The classic methods of data reduction and sample selection [17] mention its importance given the analysis of large amounts of data for each sample and the time consumed which may cause an over-adjustment of the model of training.

In all the approaches seen so far in a very simple way, the importance of selecting a suitable sample has been evidenced to reduce computational cost and time among other aspects. From now on, the various efforts made to obtain results using the

CNN method [18] with the prototype approach that facilitates the machine learning approach [19] will be more rigorously required.

The rest of this paper is structured as follows: Sect. 2 presents the theoretical background and overview referring to the main problem by the CNN method. Section 3 describes more practically by introducing the idea of the use of metrics in unsupervised learning and its relationship with CNN. Finally, the conclusions are presented in Sect. 4.

2 Theoretical Background and Overview

In practical problems, one of the most important elements to handle is the elimination of noise, redundancies, useless instances and therefore the selection of prototypes, constituting the first step for any practical application.

2.1 Problem Definition

It is desired to isolate the smallest set of instances that could predict the class with the same or greater precision than the original set [20]:

Lemma 2.1.1. *Let $X_p$be an instance where $X_p = \left(X_{p1}, X_{p2}, \ldots, X_{pm}, X_{pc}\right)$, with $X_p \in c$given by $X_{pc}$and a $X_{pi} \in R^m$being the value of the ith feature of the p_{th}sample. A training set TR, and also the N instances $X_p$and a validation set TS with t instances X_p, is obtained. $S \subset TR$ is the subset of the selected samples that resulted from applying an instance selection algorithm.*

Summarizing Lemma 2.1.1.*, the objective of an instance selection method is to obtain a subset $S \subset T$ such that S does not contain unnecessary instances* [21]*:*

$$\mathrm{Acc}(S) \cong \mathrm{Acc}(X) \tag{3}$$

where Acc(X) is the qualifier of the training set X.

2.2 Prototype-Based Approach on Unsupervised Learning

Models based on prototype analysis represent several appealing concepts such as the explicit representation of observations, data or typical representatives that exhibit some relation to psychology and neuroscience.

In Sect. 1, the relationship between condensation methods and vector quantization was approached in a very simple way, and this subsection discusses how to prototype selection matches the instance selection approach with a competitive perspective in unsupervised learning [18].

The vector quantization mathematical statement is formulated in terms of a function that represents costs and generally guides the computation of prototype vectors. A prototype-based representation [22] of a given set of P is defined in Lemma 2.2.1.

Lemma 2.2.1. *Assign the representation of a set of P feature vectors* $\{x^{\mu} \in \mathbb{R}^{n}\},\ \mu = 1, 2, ..., P$ *that represent a particular input values.*

A popular approach considers the assignment of any data point to the closest prototype, the so-called winner in the set $W = \{w^1, w^2, \ldots, w^K\}$ *in terms of a predefined distance measure.*

Using the Euclidean metric in feature space with:

$$d^2(x, y) = (x - y)^2 \text{ for } x, y \in \mathbb{R}^N \tag{4}$$

Having the quantization error [3] as the corresponding cost function:

$$H_{\mathrm{VQ}} = \sum_{i=1}^{P} \frac{1}{2} d^2\left(w^*(x^{\mu}), x^{\mu}\right) \tag{5}$$

where $w^*(x^{\mu}) \in \mathbb{R}^N$ denote the closest prototype using a Euclidean metric $x^{\mu} \in \mathbb{R}^n$:

$$d\left(w^*(x^{\mu}), x^{\mu}\right) \leq d\left(w^j, x^{\mu}\right) \text{ for all } j = 1, 2, \ldots, K \tag{6}$$

The quantization error quantifies the fidelity with which the set of prototypes represent data.

2.3 The Condensed Nearest Neighbor Rule (CNN Rule)

An in-depth study on the pillars that support the CNN method [23] and that will be specified below:

Let $\left(X_1^{'}, Y_1^{'}\right) \ldots \left(X_m^{'}, Y_m^{'}\right)$ be a sequence that depends somehow on the data D_n, and let g_n be the 1-nearest neighbor rule with $\left(X_1^{'}, Y_1^{'}\right) \ldots \left(X_m^{'}, Y_m^{'}\right)$ where m is previously set. One way to find the data is to find the subset of the size m data, for the remained minimal $n - m$ data is confirmed by the error with the I-NN rule (this is known as Hart's rule).

If:

$$\hat{L}_n = \left(\frac{1}{n}\right) \sum_{i=1}^{n} I_{\{g_n(X_i) \neq Y_i\}} \tag{7}$$

And:

$$L_n = P\{g_n(X) \neq Y|D_n\} \tag{8}$$

Then, we have the following:

Lemma 2.3.1. *$\forall\, \varepsilon > 0$,*

$$P\left\{|L_n - \hat{L}_n| \geq \varepsilon\right\} \leq 8\mathrm{e}^{-\frac{n\varepsilon^2}{32}}\left(\frac{ne}{d+1}\right)^{(d+1)m(m-1)} \tag{9}$$

where $\hat{L}_n$is about the estimate error probability.

Observe that:

$$\hat{L}_n = \left(\frac{1}{n}\right)\sum_{i=1}^{n} I_{\left\{(X_j,Y_j)\notin\bigcup\limits_{i=1}^{m} B_i\times\left\{Y_i'\right\}\right\}} \tag{10}$$

where B_i is the Voronoi cell of X_i' corresponding to $X_1' \ldots X_m'$, where $B_i \subset R^d$ is the closer partition to X_i' than to any other X_j':

$$L_n = P\left\{(X, Y) \notin \bigcup_{i=1}^{m} B_i \times \left\{Y_i'\right\}|D_n\right\} \tag{11}$$

Using simple upper bound:

$$|L_n - \hat{L}_n| \leq \underbrace{\mathrm{Sup}}_{A\in A_m}|v_n(A) - v(A)| \tag{12}$$

where v denotes the measure of (X, Y), v_n is some measure and A_m refer a set of all subsets of $R^d \times \{0, 1\}$ of the form $\bigcup_{i=1}^{m} B_i \times \{y_i\}$ where $B_1, \ldots, B_m$ are Voronoi's cells corresponding to $x_1, \ldots, x_m, x_i \in R^d$, $y_i \in \{0, 1\}$.

Using the Vapnik–Chervonenkis inequality [24]:

$$s(A_m, n) \leq s(A, n)^m \tag{13}$$

Such that A is the class of sets $B_1 \times \{y_1\}$ and each set in A intercepts in at most $m - 1$ hyperplanes. Then:

$$s(A, n) \leq \underbrace{Sup}_{n_0,n_1:n_0+n_1=n}\left(\prod_{j=0}^{1}\left(\frac{n_j e}{d+1}\right)\right)^{(d+1)(k-1)} \leq \left(\frac{n_j e}{d+1}\right)^{(d+1)(k-1)} \tag{14}$$

where n_j denotes the points $R^d \times \{j\}$ and the result follows from the Vapnik–Chervonenkis.

Other condensate rules based on CNN were also presented in [25, 26].

Table 1 New approaches based on traditional CNN methods

Method	Short description	References
Extended nearest neighbor	Used for pattern recognition	[10]
The fast-condensed nearest neighbor algorithm	Reuse Voronoi's concepts	[18]
Hierarchy extreme learning machine, for instance, selection	Fuzzy c-means utilizes condensed nearest neighbor (CNN) to make a preliminary selection of training samples	[7]
A modified firefly algorithm for image classification	Used in image classification task	[8]
Nonparametrically regression algorithm with instance selection	Provide flexible forms of prediction	[11]

An approach to the CNN algorithm [27, 28] can be as follows:

Algorithm 1
1. $\boldsymbol{T} \leftarrow \emptyset$
2. Do
3. $\forall \boldsymbol{x} \in \boldsymbol{X}(\boldsymbol{inrandomorder})$
4. Find $\boldsymbol{x}' \in \boldsymbol{T}$ such that
$\boldsymbol{x} - \boldsymbol{x}' = \underbrace{\boldsymbol{min}}_{\boldsymbol{x}^{\boldsymbol{w}} \in \boldsymbol{T}} \boldsymbol{x} - \boldsymbol{x}^{\boldsymbol{w}}$
5. If $\boldsymbol{Class}(\boldsymbol{x}) \neq \boldsymbol{Class}(\boldsymbol{x}^{\boldsymbol{w}})$ insert x to T
6. While T does not change

Several investigations have been carried out to interpret, extend and enhance the traditional CNN algorithm [29, 30]. In Table 1, some novel variants of implementation and application of the CNN method are shown.

3 Results and Discussion

A small review of the process of selecting instances has shown the high potential of sample selection techniques. Its application is valid in all areas and sub-areas of the modern world. The prototyping approach given by machine learning contributes too many investigations to reduce the computational cost of processes and the tasks of classifying huge amounts of data. Stopping in the analysis of the condensed nearest neighbor (CNN) algorithm [31], it represents a cognitive and theoretical element that means the basis of other evolutionary models.

The CNN algorithms use one nearest neighbor rule to iteratively decide if a sample should be removed or not [4].

3.1 Metric Considerations and Visual Scheme for the CNN

Many unsupervised algorithms perform unsupervised learning of distance metrics using information from the data itself or from the dimension where they are represented. In the selection of instances, the measurement of the distance between instances or the metric used is of crucial importance.

To formalize, denote the vectors x and y to those that represent the attributes of two instances x and y (classes are excluded).

A widely used metric is the Minkowski metric, which is defined as:

$$d = \sqrt[p]{\sum_{j=1}^{m} d_j^p} \tag{15}$$

where d_j is defined for continuous attributes such as $d_j = \left|x_j - y_j\right|$.

For some values of p, the Minkowski distance corresponds to a special metric as reflected in Table 2.

There are other important metrics such as the Mahalanobis distance based on the location of multivariate outliers to indicate an unusual combination between one or more variables.

A simple definition to this problem [10] is defined by:

$$d(\text{Mahalanobis}) = \left[(x_B - x_A)^{\mathrm{T}} * C^{-1} * (x_B - x_A)\right]^{0.5} \tag{19}$$

where:

x_A and x_B are a pair of objects and C is the sample covariance matrix.

The following figure shows some examples of sample selection using the Euclidean and the Mahalanobis distance using the CNN algorithm and comparing some values for the n-neighbors:

Figure 4 shows the importance of the selection and use of metrics at the time of clustering, as indicated by the classic methods of selection of instances, and the

Table 2 Minkowski metrics for different p values

Minkowski variant metric	The p value	Metric
Manhattan distance	1	$d = \sum_{j=1}^{m} d_j^p$ (16)
Euclidean distance	2	$d = \sqrt{\sum_{j=1}^{m} d_j^p}$ (17)
Chebyshev distance	∞	$d = \overbrace{\underbrace{\text{Max}}_{j=1}}^{n} \left\|d_j\right\|$ (18)

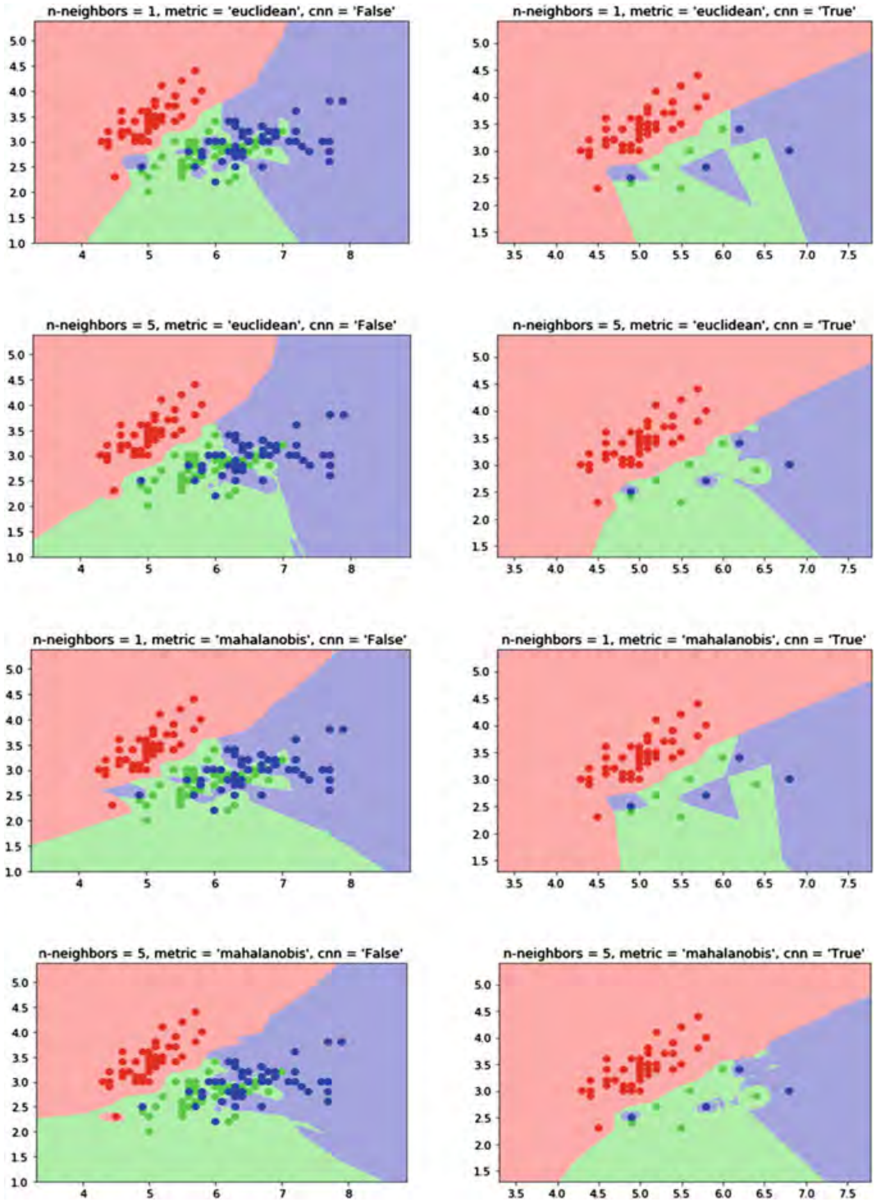


Fig. 4 Sample selection considering the Euclidean and Mahalanobis distance

fact of resorting to a sample that is sufficiently representative of a large population constitutes a difficult job. In this case, the example presented in Fig. 4 shows how the red, green and blue points are selected reflecting their color in a determined area according to the Euclidean and Mahalanobis metrics but using the CNN algorithm (squares on the right in Fig. 4) or simply using the aforementioned metrics (left squares in Fig. 4). As can be seen, using the CNN algorithm in combination with one of the two metrics achieves a clearer and more precise level of the reduced instances.

4 Conclusion

The beginning of the history of instance selection algorithms can be placed in the CNN algorithm (condensed nearest neighbor rule) whose contribution is due to Hart in 1968. The algorithm in its simplest state leaves in S a subset of T such that each element of T is closer to an element of S of the same class than to an element of S of a different class. From this idea, various variants have been formulated with an elegant mathematical profile that has allowed the reduction of computational costs in various modern problems given its simplicity.

Finally, the aim of this work has been to show some theoretical elements about the importance of the sample selection process and the condensed nearest neighbor method collected in the effort of several authors who have tried to theorize in complex aspects of the real world to give solutions to problems of today's world.

Acknowledgements The authors are greatly grateful for the support given by the SDAS Research Group. https://sdas-group.com/.

References

1. García S, Luengo J, Herrera F (2015) Data preprocessing in data mining. Springer International Publishing, Cham
2. Nisbet R, Elder J, Miner G (2009) Handbook of statistical analysis and data mining applications. Elsevier
3. Liu B (2011) Web data mining: exploring hyperlinks, contents, and usage data, 2nd edn. Springer, Heidelberg, New York
4. Hart P (1968) The condensed nearest neighbor rule (Corresp.). IEEE Trans Inf Theory 14:515–516. https://doi.org/10.1109/TIT.1968.1054155
5. Kohonen T (1990) The self-organizing map. Proc IEEE 78:1464–1480. https://doi.org/10.1109/5.58325
6. Girolami M, He C (2003) Probability density estimation from optimally condensed data samples. IEEE Trans Pattern Anal Mach Intell 25:1253–1264
7. Tang B, He H, Zhang S (2020) MCENN: a variant of extended nearest neighbor method for pattern recognition. Pattern Recogn Lett S0167865520300143. https://doi.org/10.1016/j.patrec.2020.01.015
8. Dey N (2020) Applications of firefly algorithm and its variants: case studies and new developments. Springer, Singapore

9. Chen Y, Liu Y, Ning J, Nie L, Zhu H, Chu H (2017) A composite likelihood method for bivariate meta-analysis in diagnostic systematic reviews. Stat Methods Med Res 26:914–930. https://doi.org/10.1177/0962280214562146
10. Stephanie (2017) Mahalanobis distance: simple definition, examples. In: statistics how to. https://www.statisticshowto.com/mahalanobis-distance/. Accessed 19 July 2020
11. Gong C, Wang P, Su Z (2020) An interactive nonparametric evidential regression algorithm with instance selection. Soft Comput. https://doi.org/10.1007/s00500-020-04667-4
12. Silhavy P, Silhavy R, Prokopova Z (2017) Evaluation of data clustering for stepwise linear regression on use case points estimation. Adv Intell Syst Comput 575:491–496. https://doi.org/10.1007/978-3-319-57141-6_52
13. Biancolillo A, Næs T (2019) The sequential and orthogonalized PLS regression for multiblock regression. In: Data handling in science and technology. Elsevier, pp 157–177
14. Barraza N, Moro S, Ferreyra M, de la Peña A (2019) Mutual information and sensitivity analysis for feature selection in customer targeting: a comparative study. J Inf Sci 45:53–67. https://doi.org/10.1177/0165551518770967
15. Smys S, Bala GJ (2012) Performance analysis of virtual clusters in personal communication networks. Cluster Comput 15:211–222. https://doi.org/10.1007/s10586-012-0209-8
16. Jyothirmai P, Raj J, Smys S (2017) Secured self organizing network architecture in wireless personal networks. Wireless Pers Commun 96:5603–5620. https://doi.org/10.1007/s11277-017-4436-4
17. Xu X, Li S, Liang T, Sun T (2020) Sample selection-based hierarchical extreme learning machine. Neurocomputing 377:95–102. https://doi.org/10.1016/j.neucom.2019.10.013
18. Ros F, Guillaume S (2020) Sampling techniques for supervised or unsupervised tasks. Springer International Publishing, Cham
19. Cerruela-García G, de Haro-García A, Toledano JP-P, García-Pedrajas N (2019) Improving the combination of results in the ensembles of prototype selectors. Neural Netw 118:175–191. https://doi.org/10.1016/j.neunet.2019.06.013
20. Brighton H, Mellish C (2002) Advances in instance selection for instance-based learning algorithms. Data Min Knowl Disc 6:153–172. https://doi.org/10.1023/A:1014043630878
21. Garcia S, Derrac J, Cano JR, Herrera F (2012) Prototype selection for nearest neighbor classification: taxonomy and empirical study. IEEE Trans Pattern Anal Mach Intell 34:417–435. https://doi.org/10.1109/TPAMI.2011.142
22. Biehl M, Hammer B, Villmann T (2016) Prototype-based models in machine learning: prototype-based models in machine learning. WIREs Cogn Sci 7:92–111. https://doi.org/10.1002/wcs.1378
23. Devroye L, Györfi L, Lugosi G (1996) A probabilistic theory of pattern recognition. Springer, New York, NY
24. Blumer A, Ehrenfeucht A, Haussler D, Warmuth MK (1989) Learnability and the Vapnik-Chervonenkis dimension. J ACM 36:929–965. https://doi.org/10.1145/76359.76371
25. Gates W (1972) The reduced nearest neighbor rule
26. Fukunaga K, Mantock JM (1984) Nonparametric Data Reduction. IEEE Trans Pattern Anal Mach Intell PAMI-6:115–118. https://doi.org/10.1109/TPAMI.1984.4767485
27. Ullmann J (1974) Automatic selection of reference data for use in a nearest-neighbor method of pattern classification (Corresp.). IEEE Trans Inform Theory 20:541–543. https://doi.org/10.1109/TIT.1974.1055252
28. Ritter G, Woodruff H, Lowry S, Isenhour T (1975) An algorithm for a selective nearest neighbor decision rule (Corresp.). IEEE Trans Inform Theory 21:665–669. https://doi.org/10.1109/TIT.1975.1055464
29. TOMEK I (1976) Two modifications of CNN. IEEE Trans Syst, Man, Cybern SMC-6:769–772. https://doi.org/10.1109/TSMC.1976.4309452
30. Swonger CW (1972) Sample set condensation for a condensed nearest neighbor decision rule for pattern recognition 511–519
31. Gowda K, Krishna G (1979) The condensed nearest neighbor rule using the concept of mutual nearest neighborhood (Corresp.). IEEE Trans Inform Theory 25:488–490. https://doi.org/10.1109/TIT.1979.1056066

Virtual Group Movie Recommendation System Using Social Network Information

Tranos Zuva and Keneilwe Zuva

Abstract Recommendation systems (RS) are software tools and methods designed to give recommendations to support customers in different decisions in terms of what items to buy, music to listen to, news to read, and so forth. Most recommender systems recommend items in terms of individual user likings and group recommender systems recommend items taking into consideration the likings and personalities of group members. To generate effective recommendations for a group, the system must satisfy, to the greatest extent possible, the individual interests of the group members. With the social networks, it is possible to recommend to a virtual group thus this study endeavors to develop a virtual group recommender system prototype using a model-based matrix factorization algorithm of collaborative filtering technique then popularity vote for virtual group. A publicly available dataset was used in this study. The results of the prototype showed the proposed collaborative filtering algorithm for prediction of user rating preferences demonstrated a good mean average error (MAE) of 0.70 and root mean square error (RMSE) of 0.89. Virtual groups of social networks user were then formed using the popularity vote algorithm and the results were plausible. This type of recommendation to a virtual group also enables members of the group to have something to talk about on the social network.

Keywords Mean absolute error (MAE) · Root mean squared error (RMSE) · Recommendation system (RS) · Collaborative-based filtering (CF) · Content-based filtering (CBF)

1 Introduction

With the rapid development of the Internet, more and more online services inevitably suffer from information overload, which makes it very hard for users to find the

T. Zuva (✉)
ICT Department, Vaal University of Technology, Vanderbijlpark 1900, South Africa
e-mail: tranosz@vut.ac.za

K. Zuva
Computer Science Department, University of Botswana, Gaborone, Botswana

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_24

information they need. Recommendation systems have proved to be effective means of dealing with the knowledge overload for online users and have become one of the most important and popular resources in electronic commerce [1]. Recommendation systems help users to identify items that match a user's needs and preferences from a generally long list of possibly interesting items. In literature, several recommendation techniques have been proposed [1].

Content-based filtering (CBF) endeavors to recommend almost identical items to the ones the consumer has preferred before; this is done through considering the items' features [2]. The usual approach is to use the same feature space to represent both the consumers and the items. The similarity scores are then computed between the user's profile and items' profiles. The similarity scores are then used for the recommendation of items to the user concerned. The algorithm performs perfectly well for users that do not have a lot of historical data that can be used during the recommendation period.

Collaborative-based filtering (CF) algorithm is an algorithm that is used for prediction of ratings of an item of interest of a user using many related users of similar tastes (collaborating). Collaborative filtering recommendation systems use a database about user preferences to predict additional items a new user might have interest in [3]. Collaborative filtering, seen as one of the most popular algorithms in developing recommendation systems applications, predicts the unknown preferences ratings using the established preferences ratings of a group of consumers [1].

Hybrid recommendation systems as the name entails are based on the combination of recommender systems algorithms such as content-based and collaborative-based filtering techniques. A hybrid system combines collaborative and content-based filtering techniques and tries to use the advantages of CBF to ameliorate the disadvantages of CF. For example, CF experiences problems with new items, i.e., they have problems in recommending items that have no ratings. Content-based techniques can recommend new items, this is so because they predict using the descriptions (features) of the items that are usually available. The creation of new hybrid recommender systems can be done by combining two or more basic RS techniques in several ways [1, 4].

Social network sites are Web-based services that permit individuals to:

(a) Build a public or semi-public profile within a bounded system,
(b) Formulate a list of other users with whom they share a connection, and
(c) Display and browse their list of ties and those created inside the system by others.

The nature and classification of these connections can vary from one site to another [1]. With very little time and effort, user registered with a social networking site creates his profile which contains some basic details. A social network user can do things such as adding new friends, uploading images and/or audio/videos, setting status messages, making comments, joining various groups of people who share similar interests, and joining forums for discussion but not limited only to the ones listed.

Although most recommendation systems recommend items according to an individual consumer's preferences, group recommendation systems propose items that take into account the group members' preferences and personalities [5]. To produce appropriate recommendations for a group, the program must meet the individual needs of the members of the group as much as possible [1].

2 Background

A variety of recommendation generation techniques, including content-based filtering (CBF), collaborative filtering (CB), and hybrid recommendation systems, have been proposed. These techniques assist users in finding items of their choices such as services, products, or information. Through aggregating and reviewing recommendations from other users, feedback from different authorities, and user characteristics, most of these recommender systems on the market suggest digital goods, books, Web sites, music, movies, and TV shows, to name only a few. Collaborative filtering (CF) makes suggestions to consumers according to other consumers' ratings on items, putting more weights on those from similar consumers [6]. It is considered as one of the most successful recommendation techniques.

Most of the recommendation systems on the market recommend personal items to individuals rather than to a group of people to participate or use in a group [7]. In some recommendation domains, it is necessary and suitable to recommend items to individuals such as shopping and asset investment. In these domains, personal interests and behavior are very important in personal recommender systems. In other domains where a group of individuals require an item such as a movie(s), trip(s), book club(s), and restaurant(s) for use as a group require aggregating individual consumers' likings into a group's preference properly. Choosing the aggregation algorithm is the most daunting task in group recommender systems.

From this information, it is clear that content-based filtering has some limitations. These limitations can be countered by using the strengths of collaborative filtering to solve the weaknesses of content-based filtering. The study investigates existing filtering techniques, hence applying the collaborative filtering technique to recommend movies for a group of participants.

This research, therefore, attempts to make use of the advantages of the collaborative filtering technique to overcome the limitations of the content-based filtering technique to recommend movies to individuals, and eventually making movie suggestions to diversified virtual groups using information gathered from the social network, Facebook.

Recommendation systems are available in many Web applications to help the user in making their choices. They improve sales and are of benefit to businesses, but whether they benefit customers/users by providing relevant products is still questionable [8]. The relevance of this proposal is to assist diversified groups of users by engaging them in the process of movie selection, acquiring information from their social network profiles, making individual movie recommendations, and ultimately making movie suggestions for their respective diversified groups to meet an acceptable level of member satisfaction.

Movie recommendation systems have been beneficial to viewers for years. The virtual group movie recommendation system will be of paramount importance and change user's perception by providing the necessary information, not just to a single user, but also to diversified groups of viewers who will be watching movies together, in form of virtual groups. The system will carry out a series of calculations to reach an acceptable level of satisfaction for all the virtual group members. The objective is that with time, and as technology evolves, the proposed virtual group movie recommendation system for will be updated by either adding or removing features so that it may adapt to technological changes.

The utilization of recommendation systems, specifically movie recommendations systems, has been studied with various recommendation systems approaches employed. However, this study is expected to make a major contribution to how recommendation systems recommend movies to users. The study further focuses on how individual user ratings can be considered when making group movie recommendations. In addition, the study attempts to form virtual groups based on the similarities of users in terms of their ratings. The aim is that with these formed virtual groups, a movie will be recommended to each group.

3 Methodology

In this study, a model-based matrix factorization algorithm of collaborative filtering technique was used. In this perspective, the algorithm was deployed because of its accuracy in making predictions and because of its ability to improve prediction performance [9]. In addition, this algorithm has been proved to be a better option to address the issues of data sparsity, over-fitting, and convergence speed [10].

The general approach of the virtual group movie recommendation system followed these steps:

1. Prediction of movie ratings using matrix factorization
2. The standard ranking of movies above a pre-set threshold value (3.0)
3. Recommendations of three movies to individuals
4. Plurality check
5. Generation of virtual groups
6. Standard ranking to generate a group recommendation list
7. Recommendation of movies to a virtual group, together with a list of group members.

The approach followed by the virtual group movie recommendation system is illustrated in Fig. 1.

As illustrated in Fig. 1, the recommendation system predicts ratings for all unrated movies using matrix factorization algorithm of collaborative filtering technique. The

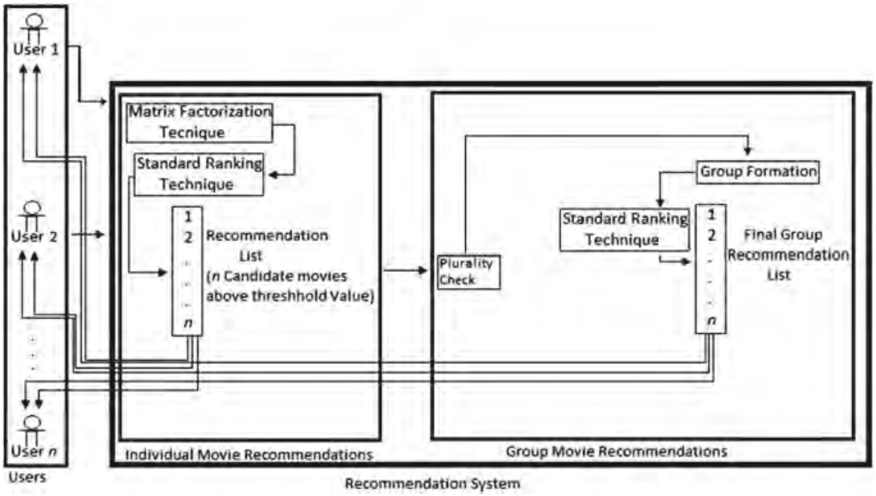


Fig. 1 Virtual group movie recommendation system architecture derived from [1]

system then uses a standard ranking technique to determine top n predicted movies, after which the top n predicted movies that are above the pre-set threshold are sent to each user. Based on common movies recommendations with a predicted rating above the pre-set threshold, the systems go through the plurality check, generates virtual groups, undergoes standard ranking technique, recommends top n movies to the generated virtual groups, and sends movie recommendations to each group member. The list is sent together with the list of group members.

Prediction of movie ratings using matrix factorization that approximates matrix X by the product of two smaller matrices W and H, i.e., $X \approx WH^{\mathrm{T}}$. In terms of recommendation systems, the matrix X is the partially observed rating matrix, $W \in \mathbb{R}^{U \times K}$ is a matrix where each row i is a vector containing the K features describing the item i. Let $w_{uk} \in W$ and $h_{ik} \in H$, then the user u predicted rating of item i is calculated using Eq. 1.

$$\hat{r}_{ui} = \sum_{k=1}^{K} w_{uk} h_{ik} = \left(WH^{\mathrm{T}}\right)_{u,i} \tag{1}$$

where the model parameters are W and H, these can be learnt by optimizing a given criterion using stochastic gradient descent.

The standard ranking of movies above the pre-set threshold value is a commonly used approach for ranking the items in recommender systems. In this approach, the predicted rating of movies was ranked from highest to lowest. In our case, all the predicted movies were ranked in descending order above the pre-set threshold value, which was 3, which means that the highly predicted item comes first in the list and the lowest predicted item is at the bottom of the list. This is to guarantee that the suggested products are correct. This was achieved by utilizing Eq. 2 the same way it was used by [1].

$$\text{rank}_{\text{standard}}(i) = R * (u, i)^{-1} \tag{2}$$

where $R*(u, i)$ is the predicted rating. The power of -1 indicates that the items with the highest predicted are recommended to the user. This approach increases the accuracy in the recommendation system [11].

After movie rating predictions were computed using matrix factorization technique, and the standard ranking of movies to ensure accuracy was carried out, each user was sent a list of three recommended movies starting with the highest to the lowest predicted rating (descending). Figure 2 demonstrates an example of the steps carried out in generating the final movie recommendation list for each user.

Generation of virtual groups was the step. After the individual recommendations were made for each user, the recommendation system identifies four movies with the highest predicted rating (above threshold 3) to users. Four virtual groups were then formed out of these users based on the predictions above the threshold.

The plurality voting aggregation strategy was used in this study. The assumption is that the movie that is recommended to more people has the advantage of being watched by these people as a virtual group. The implication is that the virtual group movie recommendation system went through the plurality check to pick the four movies that have the highest number of votings (recommendations). All these movies were picked among the movies the users have not seen, and that have a prediction above the threshold. Table 1 shows an example of how the strategy was applied.

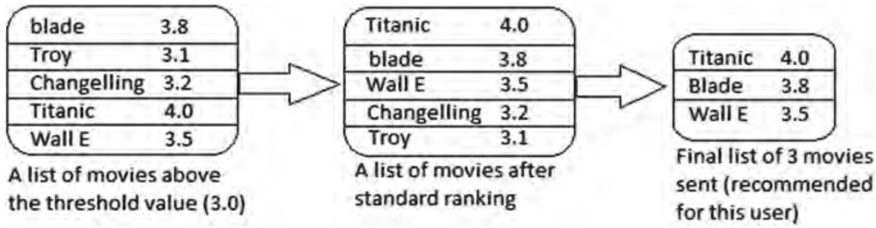


Fig. 2 Individual movie recommendation list derived from[12]

Table 1 Plurality voting strategy example derived from [1]

Movie	User 1	User 2	User 3	User 4	User 5	User 6	Group average
Blade	4.0		4.8		5.0		4.6
Troy			4.5	3.3			3.9
Changeling	3.5	3.2		3.1	3.0		3.2
Titanic		4.5	3.5		3.2		3.7
Wall E				3.0		5.0	4.0

In the case of the example stated in Table 1, it is evident that the movie Blade has the highest group average of 4.6. But the plurality voting strategy would recommend the movie Changeling even though it has a group average of only 3.2. This would be done because the movie Changeling has the greatest number of votings (recommendations). It is worth noting that all the movies that go through this stage of plurality check are all predicted to be above the threshold.

On completion of the plurality check, the recommendation system undergoes the standard ranking process for group recommendations the same way it does with individual recommendations. Just like with individual recommendations list, this is done to increase accuracy. The list is displayed to virtual group members in descending order, starting with the movie with the highest group average to the lowest.

When both the plurality check and the standard ranking processes are complete, the recommendation list reaches its final stage where it sends the final list to each virtual group member as shown in Table 2. The final group movie recommendation list is sent to all the virtual group members together with the names/user ID's of all the group members as shown in Table 3.

Another essential part of this investigation was to evaluate the developed prototype to determine its effectiveness and accuracy in generating predictions and making recommendations. The evaluation metrics may be selected depending on the goal that the researcher wishes to achieve. To measure the predictive performance of the system, to obtain the error of the system during the implementation, the mean absolute error (MAE) and root mean squared error (RMSE) were calculated. These are the two most common metrics used to measure accuracy for continuous variables.

Mean absolute error (MAE): It was used to measures the average magnitude of the errors in a set of predictions, without considering their direction. As MAE measures accuracy for continuous variables, it was used to determine the average over the test

Table 2 Movies arranged in popularity

Movie	User 1	User 2	User 3	User 4	User 5	User 6	Group average	Popularity
Changeling	3.5	3.2		3.1	3.0		3.2	4
Blade	4.0		4.8		5.0		4.6	3
Titanic		4.5	3.5		3.2		3.7	3
Wall E				3.0		5.0	4.0	2
Troy			4.5	3.3			3.9	2

Table 3 Recommendations and virtual groups

Recommended movies	Virtual groups
Changeling	User 1, user 2, user 4, and user 5
Blade	User 1, user 3, and user 5
Titanic	User 2, user 3, and user 5
Wall E	User 4 and user 6
Troy	User 3 and user 4

sample of the absolute differences between prediction and actual observation where all individual differences have equal weight. Equation 3 illustrates how MAE was calculated.

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n}\left|\mathcal{Y}_j - \hat{\mathcal{Y}}_j\right| \tag{3}$$

where $\hat{\mathcal{Y}}$ is the predicted rating, $\mathcal{Y}$ is the actual rating, and n is the number of occurrences/instances (amount of ratings).

The RMSE was also used to measure the accuracy of the prototype. Equation 4 was used for calculating the RMSE.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\mathcal{Y}_j - \hat{\mathcal{Y}}_j\right)^2} \tag{4}$$

where $\hat{\mathcal{Y}}$ is the predicted rating, $\mathcal{Y}$ is the actual rating, and n = number of occurrences/instances (the amount of ratings).

Both the MAE and RMSE can range from 0 to ∞ where ∞ is the maximum error depending on the rating scale of the measured application. They are negatively oriented scores, which mean lower values are better.

In this investigation, both the MAE and the RMSE were used together to diagnose the variation in the errors in a set of predictions. The RMSE always gives the larger or equal to the MAE; the greater the difference between them, the greater the variance in the individual errors in the sample. If the RMSE = MAE, then all the errors are of the same magnitude.

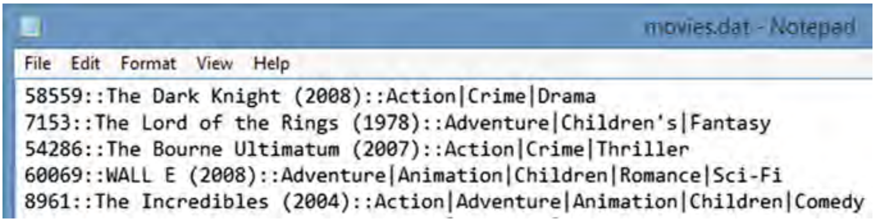
A publicly available dataset based on group recommender systems enhanced by social elements is constructed by Lara Quijano from the Group of Artificial Intelligence Applications (GIGA) obtained from (https://gaia.fdi.ucm.es/research/happymovie/download). The dataset consists of a sample of 58 users and 50 movies selected from the MovieLens dataset. Datasets were in form of two separate files, movies dataset and ratings dataset, in notepad files.

The movies file contains fifty movies that users had to rate movies he or she may have seen. This set of movies consists of a sample of all the different genres that existed and movie types so that with fifty movies a general idea of what types of movies a given user liked could be formed. The dataset had movie_id, movie_name, and genre attributes. Table 4 gives detailed descriptions of the attributes in the movies dataset while Fig. 3 shows an overview of the movies dataset Microsoft Excel file. Only the first ten elements of the movies dataset and its attributes are displayed.

The rating file had ratings ranging from 0.5 ratings as the minimum possible rating and 5.0 as the highest possible rating. The file had user_id, movie_id, and rating attributes. Table 5 gives detailed descriptions of the attributes in the rating dataset. Figure 4 shows an overview of the ratings dataset after the data. Only the first ten elements of the dataset are displayed here.

Table 4 Movies dataset attribute description table

Attribute	Description
MovieLens id number	A unique number identifying each movie in the dataset. Every single movie has a special movie_id for identification. movie_ids used here are similar to those used by grouplens in movielens
MovieLens name	This attribute holds the title of the movie. Similar to movie id, the movie names used in the dataset are identical to movie names used by grouplens in movielens
Year	This attribute holds a specific year a movie was released
Genre	A movie genre attribute refers to a motion-picture category based on similarities either in the narrative elements or in emotional response to the film. The genres contained in the dataset are Action, Adventure, Animation, Children, Comedy, Crime, Documentary, Drama, Fantasy, FilmNoir, Horror, Musical, Mystery, Romance, SciFi, Thriller, War, and Western


movies.dat - Notepad

File Edit Format View Help

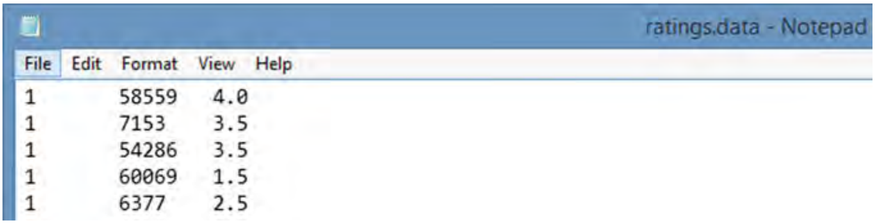
```
58559::The Dark Knight (2008)::Action|Crime|Drama
7153::The Lord of the Rings (1978)::Adventure|Children's|Fantasy
54286::The Bourne Ultimatum (2007)::Action|Crime|Thriller
60069::WALL E (2008)::Adventure|Animation|Children|Romance|Sci-Fi
8961::The Incredibles (2004)::Action|Adventure|Animation|Children|Comedy
```

Fig. 3 Overview of movies dataset notepad file

Table 5 Ratings dataset attribute description table

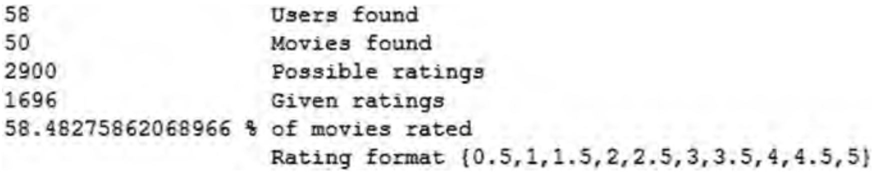
Attribute	Description
user_id	A unique number identifying the user in the dataset. Every single movie has a special user_id for identification. The dataset consists of 58 users. As a result, user_ids begin at 1–58
movie_id	A unique number identifying movies in the dataset. Every single movie has a special movie_id for identification. movie_ids used here are similar to those used by grouplens in movielens
rating	Rating attribute holds the ratings than 58 users gave to the 50 movies that presented them ratings range from 0.5 which is awarded for movies the user least liked to 5.0 for movies the user considered flawless in every department and left a long-lasting impression on them

When conducting this investigation, the author understood that data quality is important and without accurate, good-quality data, a significant amount of time, effort, and resources would be wasted trying to develop a recommendation system. As a result, it was ensured that only the most accurate and relevant data was entered and used in the datasets. Data scrubbing also referred to as data cleansing may be



1	58559	4.0
1	7153	3.5
1	54286	3.5
1	60069	1.5
1	6377	2.5

Fig. 4 Overview of ratings notepad dataset



```
58                      Users found
50                      Movies found
2900                    Possible ratings
1696                    Given ratings
58.48275862068966 %     of movies rated
                        Rating format {0.5,1,1.5,2,2.5,3,3.5,4,4.5,5}
```

Fig. 5 Results after preprocessing data

described as the identification of errors within a dataset, and the removal or correction of those errors. This process involves ensuring that your data is correct, consistent, and usable by identifying and removing or correcting any errors or corruptions in the data. After an establishment of which attributes to use, the irrelevant data was then left out and the movies and ratings datasets were loaded into the data frame, using python, in Jupiter notebook integrated development environment. To obtain meaningful data from the datasets, movies and ratings datasets were merged into one data frame using movie_id attribute.

After the data was cleaned, the author then preprocessed the data to obtain information from it as the raw data is non-comprehensive. By so doing, the author added a visual aspect to the data, making it easier and quicker to understand. This process included opening movies dataset (movies.dat) and ratings dataset (ratings.data), removing attributes that are not relevant to the study, removing or correcting errors from movies.dat and ratings.data and saving the changes on both movies.dat and ratings.data files.

After finalizing the preprocessing, the datasets contained 58 users and 50 movies with 1696 given ratings out of possible 2900 ratings. This implies that 58.5% ratings were given and the ratings expected were in a format 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 5.5, and 5.0. Figure 5 gives a general overview of the data after preprocessing.

4 Result Evaluation of the Prototype

Another crucial stage of the prototype was to evaluate its accuracy by comparing the predicted ratings directly with the actual ratings given by the users. To fulfill

```
Evaluating MAE, RMSE of algorithm SVD.
------------
------------
Mean MAE : 0.7027
Mean RMSE: 0.8996
------------
------------
```

Fig. 6 MAE and RMSE results of the prototype

this purpose, mean absolute error (MAE) and root mean squared error (RMSE) were deployed with the results shown in Fig. 6 [13]. A movie recommendation system using movielens dataset is proposed and 0.709531 MAE and 0.905520 RMSE using FunkSVD are achieved while achieving 0.717344 MAE and 0.9200979 RMSE using item-based collaborative filtering [14]. A movie recommendation system via K-means PSO-FCM technique is proposed that achieved 0.7547 as the MAE [1]. A system that achieved 0.82 MAE and 1.08 RMSE using fast maximum margin matrix factorization, 0.80 MAE and 1.05 RMSE using Iterative SVD, and 0.72 MAE and 0.95 RMSE using repeated matrix is proposed. Our prototype showed better performance over all these with 0.7027 MAE and 0.8996 RMSE as Fig. 6 illustrates.

The proposed collaborative filtering virtual group movie recommendation system using social network information has demonstrated a good 0.70 MAE and 0.89 RMSE. The algorithm has been explored and evaluated comprehensively. The findings depicted that the prototype fulfills its objectives and performs better than other recommendation systems considered for comparison.

5 Conclusion

In this article, a model-based matrix factorization algorithm of collaborative filtering technique was used to predict user preference scores for movies. The popularity vote ranking algorithm was then used for virtual group recommendation. The accuracy of the collaborative filtering technique was measured using mean absolute error and root mean squared error. The results were acceptable in comparison with other researchers thus gave us confidence in the results for the virtual group. The virtual group recommender system is believed to improve the interaction of users on social networks through discussion of the movies they would have watched at the same time. Our system might have to request users if they agree to let their contact details availed to people for those who have the same preference as theirs. The privacy issues will need to be addressed in future work.

References

1. Ricci F, Rokach I, Shapira B (2010)Introduction to recommender systems handbook.In: Nature S (ed) Recommender systems handbook. Bozen-Bolzano
2. Ricci F, Rokach L, Shapira B (2011) Introduction to recommender systems handbook. Springer, Boston
3. Ricci F, Rokach L, Shapira B (2015)Recommender systems handbook. In: Nature S (ed) Introduction and challenges. Springer International Publishing, Bolzano, pp 1–34
4. Lops P, Gemmis M, Semeraro G (2010)Recommender systems handbook. s.l.In: Nature S (ed) Content-based recommender systems: state of the art and trends. Springer International Publishing AG, pp 73–105
5. Breese J, Kadie CC (1998) Empirical analysis of predictive algorithms for collaborative filtering. Morgan Kaufmann Publishers Inc., Madison, Wisconsin
6. Zhongqi L (2015) Content-based collaborative filtering for news topic recommendation. AAAI Press ©2015, Austin, Texas
7. Ricci F, Rokach L, Shapira B (2010)Introduction to recommender systems handbook.In: Nature S (ed) Recommender systems handbook. Bozen-Bolzano, pp 1–35
8. Boyd DM, Ellison NB (2007) Social network sites: definition, history, and scholarship. J. Comput. Med. Commun. 1:210–230
9. Jameson A, Smyth B (2017)Recommendation to groups.In: Nature S (ed) The adaptive web. 2017 Springer International Publishing AG, Dublin, pp 596–627
10. Baltrunas L, Makcinskas T, Ricci F (2010) Group recommendations with rank aggregation and collaborative filtering. ACM, New York, NY, USA
11. Karunanithi N, Alspector J (1996)Feature-based and clique-based user models for movie selection. In: Proceedings of the fifth international conference, vol II, pp 130–136
12. Porcel C, Tejeda-Lorente A, Martínez MA, Herrera-Viedma E (2012) A hybrid recommender system for the selective dissemination of research resources in a Technology Transfer Office.Inf Sci 2:1–19
13. Al-Barznjl K, Atanassov A (2017) Collaborative filtering techniques for generating recommendations on big data. s.n., Sofia, Bulgaria
14. Kim JK, Kim KH, Oh YH, Ryu YU (2010) A group recommendation system for online communities. International Journal of information management. J Inf Prof 30:1–13

Hindi to English: Transformer-Based Neural Machine Translation

Kavit Gangar, Hardik Ruparel, and Shreyas Lele

Abstract Machine Translation (MT) is one of the most prominent tasks in Natural Language Processing (NLP) which involves the automatic conversion of texts from one natural language to another while preserving its meaning and fluency. Although the research in machine translation has been going on since multiple decades, the newer approach of integrating deep learning techniques in natural language processing has led to significant improvements in the translation quality. This paper has developed a Neural Machine Translation (NMT) system by training the Transformer model to translate texts from Indian Language Hindi to English. Hindi being a low resource language has made it difficult for neural networks to understand the language thereby leading to a slow growth in the development of neural machine translators. Thus, to address this gap, back-translation is implemented to augment the training data and for creating the vocabulary, it has been experimented with both word and subword level tokenization using Byte Pair Encoding (BPE) thereby ending up training the Transformer in 10 different configurations. This led us to achieve a state-of-the-art BLEU score of 24.53 on the test set of IIT Bombay English-Hindi Corpus in one of the configurations.

Keywords Neural machine translation · Transformer · Byte pair encoding · Back-translation

K. Gangar (✉) · H. Ruparel · S. Lele
Veermata Jijabai Technological Institute, Mumbai, India
e-mail: kavitgangar34@gmail.com

H. Ruparel
e-mail: hardikruparel14@gmail.com

S. Lele
e-mail: shreyaslele2398@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_25

1 Introduction

Machine translation is one of the oldest tasks taken up by computer scientists and the development in this field has been going on for more than 60 years. The research in this field has made remarkable progress to develop translator systems to convert source language to target language while maintaining the contextuality and fluency. In earlier times, the translation was handled by statically replacing words with the words from the target language. This dictionary look-up led technique led to inarticulate translation and hence was made obsolete by Rule-Based Machine Translation (RBMT) [1]. RBMT is a system based on linguistic information about the source and target languages derived from dictionaries and grammar including semantics and syntactic regularities of each language [2]. With the absence of flexibility and scalability to incorporate new words and semantics and the requirement of human expertise to define numerous rules, rule-based machine translation systems could only achieve accuracy on a subset of languages. To overcome the issues of the RBMT system, a new approach called Statistical Machine Translation (SMT) was introduced. Instead of having rules determine the target sequence, SMT approaches leverage probability and statistics to determine the output sequence. This approach made it feasible to cover all types of language within the source and target language and to add new pairs. Most of these systems are based on Bayesian prediction and have phrases and sentences as the basic units of translation. The main issue faced by this approach is the requirement of colossal amounts of data, which is a huge problem for low resource languages.

Due to these prevailing issues, there is a demand to explore alternate methods for creating a smarter and more efficient translation system. The development of various deep learning techniques and the promising results shown by the combination of these techniques with NLP created a new approach called NMT. NMT's advantage lies in two facts that are its simplistic architecture and its ability to capture long dependencies in the sentence, thereby indicating its huge potential in emerging as a new trend of the mainstream [3]. Conceptually speaking, NMT is a simple Deep Neural Network (DNN) that reads the entire source sentence and produces an output translation one word at a time. The reason why NMT systems are appealing is that they require minimal domain knowledge which makes it well-suited for any problem that can be formulated as mapping an input sequence to an output sequence. Also, the inherent nature of the neural networks to generalize any input implies that NMT systems will generalize to novel word phrases that are not present in the training set.

Moreover, almost all the languages in the world are continuously evolving with new words getting added, older words getting amended and new slangs getting introduced very frequently. Even though NMT systems generalizes the input data well, they still lack the ability to translate the rare words due to their fixed modest-size vocabulary which forces the NMT system to use *unk* symbol for representing out-of-vocabulary (OOV) words [4]. To tackle this issue, a subword tokenization technique called Byte Pair Encoding (BPE) was introduced. BPE divides the words such that the frequent sequence of letters is combined thereby forming a root word and affix. This approach alone handles the OOV words by merging the root word and the different combinations of affixes thereby creating the rare word [5].

This paper presents the experimental setup and the state-of-the-art results obtained after training the Transformer model on the IIT Bombay CFILT English-Hindi dataset of 1.5 million parallel records. The paper is organized as follows: Sect. 2 describes the motivation behind our work. Section 3, describes the implemented model. In Sect. 4 the details of the experimental setup is presented for training the model. In Sect. 5a comparative analysis of the results obtained by training the model is displayed in different configurations. Finally, Sect. 6 concludes the paper.

2 Motivation

With the power of deep learning, Neural Machine Translation has arisen as the most powerful approach to perform the translation task.

In [6] a model called Transformer was introduced which uses the encoder-decoder approach where the encoder first converts the original input sequence into its latent representation in the form of hidden state vectors. By using latent representation, can decode the predicted output sequence. Here transformed helps to achieve the parllelization to encode the symbol data position from the sequence.

In [7], explains about the improvement of machine translation for the monolingual series in the training data set. Dummy source sentence helps to detect the monolingual training instances and provides the better results in accuracy.

In [8], investigates the open vocabulary translation approach based on the byte pair encoding. This byte pair encoding is helps to compress the word segmentation and make it as open vocabulary with fixed sized by using neural network models.

Motivated with the results obtained by the transformer for machine translation on various languages, a translation system that translates a Hindi sentence to English is created. Since a low resource Hindi language is used for which the amount of good quality parallel corpus is limited, back-translation is applied to increase the quantity of training data. To overcome the problem caused by out of vocabulary words, BPE is used.

3 NMT Model Architecture

3.1 Structure

The Transformer model is the first NMT architecture that completely relies on the self-attention mechanism to calculate the representation of its input and output data without using recurrent neural networks (RNNs) or convolutional neural networks (CNNs) [9]. The Transformer model consists of an encoding unit and a decoding

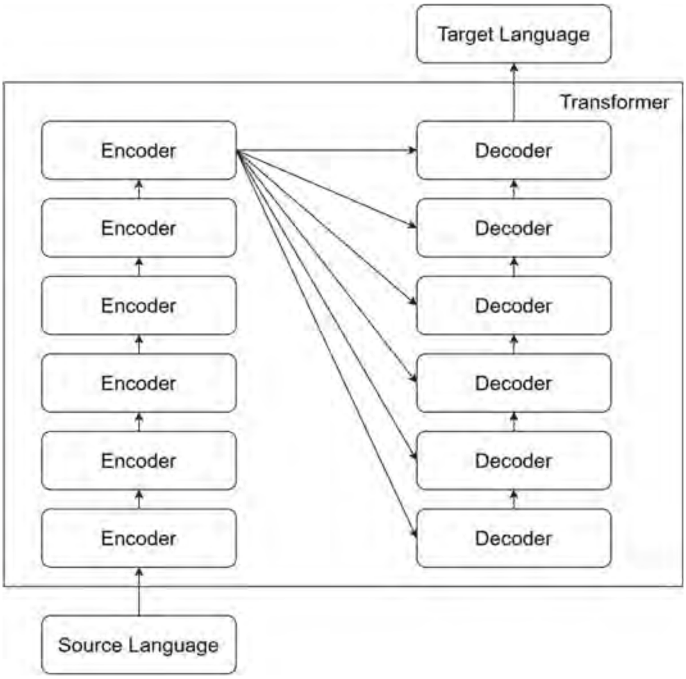


Fig. 1 Transformer structure—Bird's-eye view

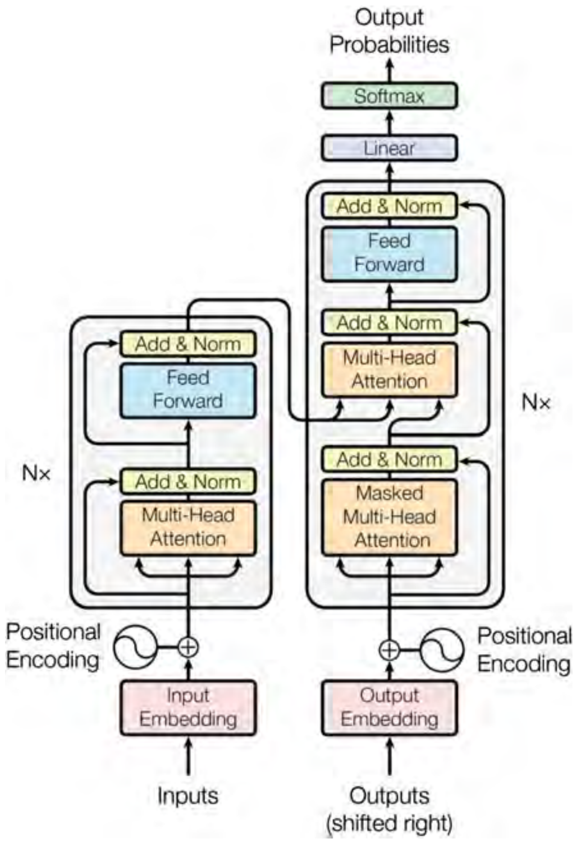
unit wherein each of these components consists of a stack of 6 layers of encoders and decoders respectively. (see Fig. 1).

Each encoder layer consists of two sublayers. The first sublayer is the multi-head self-attention layer and the second sublayer is a position-wise fully connected feed forward network [6]. Each decoder layer in the decoding component consists of 3 sublayers. The function of the first two sublayers is the same as that in the encoder. However, the third sublayer performs multi-head attention mechanism over the output of the encoder stack (see Fig. 2).

3.2 Working

Before passing the input data to the encoder stack, the input data is first projected into an embedded vector space. To capture the notation, distance between different words and the order of the words in the input sequence, a positional embedding vector is added to the input vector. This intermediate vector is then fed to the first layer of the encoding component of the transformer. A multi-head self-attention is then computed on this intermediate embedded vector space. Multi-headed mechanism improves the performance of the attention-layer in two ways. First, it helps expand the model's

Fig. 2 The transformer architecture [6]



ability to focus on the words in different positions. Second, it gives the attention layer multiple representation subspaces, concatenates them and then projects linearly onto a space with initial dimensions [10]. The output of the self-attention layer is then passed onto a dense feed forward network which consists of two linear functions with RELU in between them. The output of this feed forward network is then passed on to another encoder layer stacked on top of it. All the encoder layers in the encoding component have the same functionality. Finally, the output of the encoding unit is then passed as an input to the decoding unit.

The decoder has similar functionality as that of the encoder. The output of the top encoder is converted into a set of attention vectors which is then used by each decoder in the decoding component. This helps the decoder focus on the appropriate position in the input sequence. The decoder predicts words one word at a time from left to right. Upon prediction of each word, it is again fed into the bottom decoder after converting it into an embedded vector space and adding a positional embedding vector. The decoder's self-attention mechanism works in a slightly different way than the encoder. In the decoder, the self-attention layer is only allowed to look at the words in earlier positions. This is done by masking the words at future positions to *-inf*. Each decoder layer in the decoding component performs the same function. The output of the last decoder layer is then fed into a linear layer and softmax layer. The linear layer outputs a vector having a size equal to the size of the target language vocabulary. Each position in this output vector determines the score of the unique word. This vector of scores is then converted into probabilities by the softmax layer and the position with the highest probability is chosen, and the word associated with it is produced as the output for the particular time step.

4 Experimental Setup

4.1 Dataset

The fundamental requirement for assembling a machine translation system is the availability of parallel corpora of the source and the target language. In this paper, the transformer model is trained on the Hindi-English parallel corpus by the Center for Indian Language Technology (CFILT), IIT Bombay [11]. The training data consists of approximately 1.5 million texts from multiple disciplines while the development and the test set contains data from the news domain. Table 1 provides the details about the number of sentences and the number of unique tokens in English and Hindi that are present in our chosen dataset.

4.2 Data Preprocessing

Data preprocessing is an essential data mining technique that helps clean the data by removing the noise and outliers which can then directly be used by the model for training and testing purposes. Our preprocessing pipeline consists of 3 main steps viz. Data Cleaning, Removal of duplicates and Vocabulary creation. Each step is explained in detail below:

Data Cleaning (Step 1) In this step, the special characters, punctuation and noise characters are removed from both the English and Hindi text corpus. After the elim-

Table 1 Metadata of the dataset

Dataset	# Of sentences	Unique Hindi tokens	Unique English tokens
IITB train	1,267,502	421,050	242,910
IITB dev	483	2479	2405
IITB test	2478	8428	9293

ination of all the noise characters, the empty lines are removed. The resulting text corpus was then converted into lower case and was then fed into the next step to remove the duplicates.

Removal of Duplicates (Step 2) The cleaned and noise-free text corpus obtained as a result of the above step was then used to remove the duplicate records. This resulted in the creation of our training universe containing approximately 1.2 million unique parallel text corpus which was used for creating the vocabulary.

Vocabulary Creation (Step 3) Vocabulary creation is one of the most fundamental step in Neural Machine Translation. The coverage of the vocabulary is a major factor that drives the overall accuracy and the quality of the translation. If the vocabulary under-represents the training data universe, then predicted translation will contain many *unk* tokens thereby reducing the BLEU score drastically. Thus creating a modest-size vocabulary that optimally represents the data is a challenging task. For creating the vocabulary for both Hindi and English language, two approaches are implemented: word level tokenization and subword level tokenization. In the word level tokenization, it first extracted 50,000 most frequently used words from the training set and then added it to the vocabulary. While in the subword level tokenization, Byte Pair Encoding (BPE) is used for creating 50,000 subword tokens which were added in the vocabulary. The evaluation of the performance of our model on both word and subword level tokenization is presented in Sect. 5.

4.3 Back-Translation

Hindi, being a low resource language as compared to its counterpart European languages has made the availability of data quite difficult. Many institutions around the world are creating larger and a more complete text corpus for the low resource languages. To tackle the lack of availability of Hindi-English parallel corpus, back-translation technique is used. Back-translation technique is used for augmenting the training data which leads to increasing the output accuracy of the translation. There is a plethora of monolingual English data available on the internet which can be used to generate text corpus of a low resource language. To generate the additional Hindi-English parallel text corpus, an English to Hindi machine translation system is first trained on our training data and then translated the 3 million WMT14 English monolingual data to generate the corresponding predicted Hindi text corpus.

To observe the effect of back-translation, it first divided the 3 million back-translated parallel records in 4 batches. Then cumulatively added the back-translated records to the original training data in each of these batches. The first batch contains the 0.5 million back-translated records along with the 1.2 million original training data. In the same way, an additional 1 million, 1 million and 0.5 million are added in the second, third and fourth batch respectively. Table 2 summarizes the training data universe for each batch.

Table 2 Training data universe: Batch-wise summary

Batch number	# Of back-translated records added (million)	Total records (million)
Batch 1	0.5	1.7
Batch 2	1.5	2.7
Batch 3	2.5	3.7
Batch 4	3	4.2

4.4 Training Details

After the data preparation and segregation into batches, our transformer model is trained using Opennmt-tf toolkit [12]. For training the model, the NVIDIA Tesla K80 GPU provided by Google Colab [13] is used. For our transformer model, the default 6 layers setting in both encoder and decoder each of which contains 512 hidden units are used. Further the proposed work used the default embedding size of 512 along with 8 headed attention. The batch size is configured to be equal to 64 records and the effective batch size which is defined as the number of training examples consumed in one step to be equal to 384. The model parameters are optimized using the LazyAdam optimizer. The model was trained on 10 different configurations 5 each for word and subword level tokenization till convergence or till 70,000 steps at max (hard stop). The GPU run-time provided on Google Colab resulted in a training duration of approximately 20–24 h for each configuration.

5 Results

The quality of translation of our model are trained on the test set using the Bilingual Evaluation Understudy (BLEU) score [14] and the Rank-based Intuitive Bilingual Evaluation (RIBES) score [15]. For depicting the performance of subword level tokenization, where it divides the test set into 2 subsets. The first set (Set-1) consists of sentences whose words are present in the vocabulary generated from word level tokenization. This set consists of 1694 sentences. The second set (Set-2) is the complete test set consisting of 2478 sentences.

In Table 3, after adding the first batch of 0.5 M parallel back-translated records with the original training data, the BLEU score increased by 3.79 and with the subsequent addition of other 2 batches the BLEU and the RIBES score reached a maximum of 24.79 and 0.741 respectively. However, when the 4th batch of 0.5 M back-translated records was added with the previous batches, the scores decreased by a small margin indicating convergence with respect to the addition of back-translated data.

Similar to the results obtained with word level tokenization, in Table 4, after adding the first batch of back-translated records the BLEU score increases by 4.78 and with

Table 3 Results of word level tokenization (Set-1)

Model ID	Model	BLEU	RIBES
1	Transformer	18.76	0.699708
2	Transformer with Batch 1	22.55	0.730440
3	Transformer with Batch 2	23.95	0.735804
4	Transformer with Batch 3	24.79	0.741369
5	Transformer with Batch 4	24.68	0.740567

Table 4 Results of subword level tokenization (Set-1)

Model ID	Model	BLEU	RIBES
6	Transformer	19.10	0.695566
7	Transformer with Batch 1	23.98	0.733614
8	Transformer with Batch 2	25.44	0.737078
9	Transformer with Batch 3	25.87	0.739192
10	Transformer with Batch 4	25.74	0.742397

Table 5 Results for the transformer with Batch 4 model on Set-2

Model ID	Tokenization	BLEU	RIBES
5	Word level	21.22	0.728683
10	Subword level	24.53	0.735781

the subsequent addition of other 2 batches the BLEU score reached a maximum of 25.87. After adding the 4th batch, the BLEU scored decreased by 0.13 however the RIBES score increased by 0.003.

When compared with word level tokenization, subword level tokenization achieves a better BLEU score which can be attributed to the fact that it has the advantage of not having an out-of-vocabulary case and also to learn better embeddings for rare words since rare words can enhance the learning from its subwords that occur in other words. This fact is further strengthened in Table 5 which shows the BLEU and RIBES score for the Transformer with Batch4 model using word and subword level tokenization on Set-2. The decrease in the BLEU and RIBES score as compared to Tables 3 and 4 is due to the fact that the Set-2 consists of additional sentences as compared to Set-1 which contain rare words that are not included in the vocabulary for word level tokenization. When subword level tokenization is used, the model performs reasonably well even in the presence of rare words which is not the case for word level tokenization.

6 Conclusion

The transformer model has displayed promising results for neural machine translation involving low resource languages as well. It is observed that after adding the back-translated records the performance was certainly improved, however when the amount of generated data increases beyond a certain level, there is no further improvement in the performance. Using a combination of the transformer model, back-translation technique and a subword tokenization method like BPE, a BLEU score of 24.53 which is the state-of-the-art on this dataset to the best of our knowledge has been achieved. The future research directions on the proposed work will try to incorporate state-of-the-art Natural Language Processing models like BERT [16] into NMT to further improve the quality of translation.

References

1. Saini S, Sahula V (2018) Neural machine translation for English to Hindi. In: Fourth international conference on information retrieval and knowledge management (CAMP). IEEE, pp 1–6
2. Rule-based machine translation. https://en.wikipedia.org/wiki/Rule-based_machine_translation
3. Yang S, Wang Y, Chu X (2020) A survey of deep learning techniques for neural machine translation
4. Luong M, Sutskever I, Le Q, Vinyals O, Zaremba W (2014) Addressing the rare word problem in neural machine translation
5. Tacorda A, Ignacio M, Oco N, Roxas R (2017) Controlling byte pair encoding for neural machine translation. In: International conference on Asian language processing (IALP) 2017. IEEE, pp 168–171
6. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, Kaiser Ł (2017) Attention is all you need. In: Advances in neural information processing systems, pp 5998–6008
7. Sennrich R, Haddow B, Birch A (2015) Improving neural machine translation models with monolingual data
8. Sennrich R, Haddow B, Birch A (2015) Neural machine translation of rare words with subword units
9. Goyal V, Sharma D (2019) LTRC-MT simple & effective Hindi-English neural machine translation systems at WAT 2019. In: Proceedings of the 6th workshop on Asian translation 2019, pp. 137–140
10. The illustrated transformer. http://jalammar.github.io/illustrated-transformer
11. Kunchukuttan A, Mehta P, Bhattacharyya P (2017) The IIT Bombay English-Hindi parallel corpus
12. Klein G, Kim Y, Deng Y, Senellart J, Rush A (2017) OpenNMT: open-source toolkit for neural machine translation. In: Proceedings of ACL, system demonstrations 2017. Association for Computational Linguistics, pp 67–72
13. Google Colab. http://colab.research.google.com/
14. Papineni K, Roukos S, Ward T, Zhu W (2002) BLEU: a method for automatic evaluation of machine translation. In: Proceedings of the 40th annual meeting. Association for Computational Linguistics 2002, pp 311–318

15. Isozaki H, Hirao T, Duh K, Sudoh K, Tsukada H (2010) Automatic evaluation of translation quality for distant language pairs. In: Conference on empirical methods in natural language processing, vol 2010, pp 944–952
16. Devlin J, Chang M, Lee K, Toutanova K (2018) Bert: pre-training of deep bidirectional transformers for language understanding

CNN-Based Alphabet Identification and Sorting Robotic Arm

Saleh Al-Faraj, Mustafa Al-Bahrani, Saeed Al-Ghamdi, Marwan Rafie, Abul Bashar, and Ghazanfar Latif

Abstract Automated identification of objects and sorting them based on specified criteria is a crucial problem which is encountered by various manufacturing companies. To this end, they employ automated robots which need to perform these tasks with high accuracy. In this paper, a CNN-based machine learning system is proposed, designed, and implemented to accurately identify objects marked with English alphabets and sort them in a correct order based on the input given by the user. It consists of a hardware module which incorporates a robotic arm controlled by a Raspberry Pi microcontroller. The software module is based on a CNN-based image identifier model trained on an indigenous dataset consisting of 3898 images of English alphabets rotated at random angles. The experimental results demonstrate training and validation accuracies of 99.06% and 98.79%, respectively, based on the model trained over 120 epochs. Furthermore, our system was able to successfully sort and arrange the alphabets with the desired accuracy.

Keywords Alphabet sorting · Robotic arm · Image recognition · Convolutional neural networks · Robotic automation

S. Al-Faraj · M. Al-Bahrani · A. Bashar (✉) · G. Latif
Computer Engineering Department, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia
e-mail: abashar@pmu.edu.sa

S. Al-Faraj
e-mail: 201501220@pmu.edu.sa

M. Al-Bahrani
e-mail: 201601053@pmu.edu.sa

G. Latif
e-mail: glatif@pmu.edu.sa

S. Al-Ghamdi · M. Rafie
Computer Science Department, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia
e-mail: 201601001@pmu.edu.sa

M. Rafie
e-mail: 201600035@pmu.edu.sa

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_26

1 Introduction

Generally speaking, sorting is such a common problem observed in various applications and researchers have tried to find fast and efficient ways to sort. Warehouse systems utilize the need for robots to manage their inventory and sort goods using automated systems. There have been many solutions toward automated sorting problem which will be presented in the next section. Alphabetical sorting is one problem that has not been given due attention and is a challenge the industry faces today. The main idea of this paper is to propose an automated robotic arm for identifying and sorting objects which have English alphabets written on them. Essentially, in our approach, a robotic arm is set up, which has the means to read images through a camera and identify the alphabet with the aid of computer vision algorithms. To make the system more interactive, the user inputs a word and based on this the robotic arm is supposed to pick the alphabets in the correct order and build the word with the correct spelling (i.e., sort the alphabets correctly). Robotic arm in general is one of the best options suitable for arranging and sorting problems. The proposed system utilizes a Raspberry Pi microcontroller to control the robotic arm movement, while the alphabet identification is performed with the help of a CNN-based machine learning model which is trained on our generated dataset.

The rest of this article is organized as follows. Section 2 discusses the background of this domain, while Sect. 3 presents the architectural details of the proposed system. Section 4 provides the results achieved and the related discussion. Section 5 concludes the paper with the main contributions and future recommendations.

2 Background

To design our system, the review of the literature is done for the automated sorting problem and found papers where the approach is either arranging objects or recognizing handwritten alphabets using a model which is pretrained dataset (such as the MNIST dataset) [1–4]. Digital image processing becomes important field with the emergence of new machine learning and deep learning techniques with the availability of high computational resources for the recognition, classification, and segmentation in different areas such as medical diagnosis from images [5], sign language recognition for disabled people [6, 7], traffic signs recognition [8], image enhancement [9, 10] and assisted living for visually impaired persons [11, 12]. Unable to find any research work that would arrange alphabets at a rotated angle which is scattered in a region of interest, it is believed that this is a new take on arranging scattered alphabets that are rotated at some random angle.

Table 1 provides a summary of the related work in this domain, where each approach is compared to our approach in terms of the key idea, their benefits, and drawbacks. This was also a discussion that wanted to know how beneficial our research work can be used as a real-world application. It is understood that some

Table 1 Comparison of different existing techniques with our proposed solution

References	Main idea	Benefits	Drawbacks
[13]	Recognizes alphabets based on their color. The arm detects the position of the alphabets from the image. The arm sorts the alphabets according to their color	Sorting colored objects from the panel and put them in the right boxes	There are many different items with the same color
[14]	Robotic arm sorts alphabets based on speech recognition. Alphabets are recognized based on RGB color and shape from the camera images	It avoids mistake and can carry heavyweight	The sound may face a problem since there are different accents diversity
[15]	A webcam is used to captures colored object cubes. Based on the color of the cubes, the robotics arm will place them into different cups	It works well for the manufacturer that uses color in organizing products	Based on the color of the image so leads to miss recognition if colors change
[16]	The camera image is processed using GNU Octave to determine the color and the shape	It is low cost and can sort colored objects	The objects around it can lead to wrong shape detection measurements
[17]	A handwritten optical character recognition (OCR) from the camera images to control the robotic arm	It can be used to scan handwritten notes and similar texts	Possibility for the wrong prediction if it is written by different persons
Proposed alphabet sorting robotic arm	The camera will take a picture of the workspace, and the alphabets are recognized using CNN that will send the position to the Robotics arm and arrange them accordingly	Using CNN as classifier gives high accuracy	It will only work for English Alphabets

industries have an arranging issue, where some packages would be labeled, and they would need to arrange them according to some criteria. One example application would be a library, as people come and borrow books. Once they return it, all these books would be added in some cart, shuffled, with labels that indicate which shelf should be placed alphabetically and it is a hassle for a librarian to sift through these books and find the "correct" book with the specified labeled to be added to the shelf. This is one possible application, however, there could be many other applications to use our proposed solution.

3 Proposed System Design

The proposed system has two major modules, namely the hardware and the software. The hardware module consists of the robotic arm, Raspberry Pi, and the servo controller. The software module is composed of the OCR system and the CNN-based image identifier and alphabet classifier. Their details are now presented below.

3.1 New Alphabets Dataset

A new data is prepared which has 3898 images A to Z capital alphabets as shown in Table 2. Firstly, the data processing packages were imported and then the images were captured from the images dataset. The sample of captured images is shown in Fig. 1.

Here, it can be seen that the alphabets are randomly rotated a certain angle. Next, they imported the machine learning packages, loaded the processed data, and split them into testing data and training data. The test data had 20% of the total number of images while the training data had 80% of the images. Later, a predefined model with four layers is created. The next step was to train the CNN model.

The model was then validated and found to be fit for testing. After that, came up with the trained and validate model which have been used to recognize the randomly arranged alphabets. The above-mentioned steps of alphabets image recognition are depicted in Fig. 2.

Table 2 Newly build alphabets dataset

A	190	F	128	K	155	P	127	U	131	Z	157
B	157	G	152	L	155	Q	153	V	149	Total	
C	158	H	155	M	155	R	137	W	152	3898	
D	126	I	156	N	157	S	140	X	149		
E	156	J	149	O	155	T	145	Y	154		

Fig. 1 Sample images of the newly built dataset

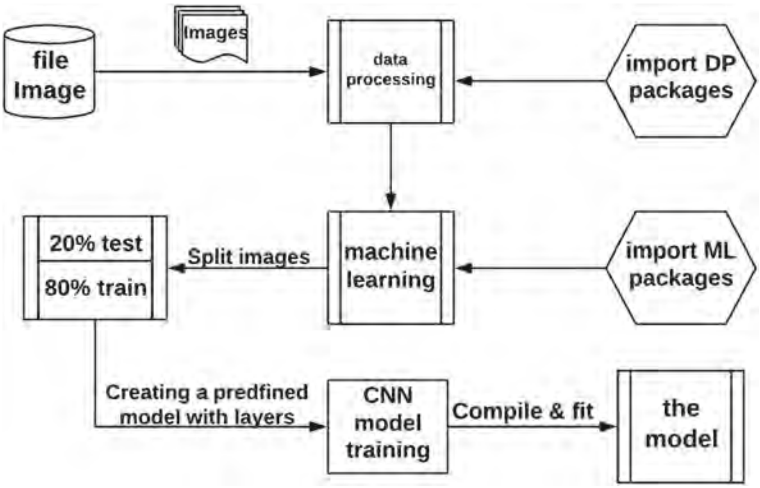


Fig. 2 Process of alphabets image recognition

3.2 Alphabets Image Processing

Now, the process of the alphabet OCR is described which is shown in Fig. 3. The camera takes the picture of the workspace which is saved as an image in a local file. Then, the alphabets in the image are recognized using optical character recognition (OCR). OCR is a technology that translates an image to be recognized by a machine. It distinguishes printed or handwritten text within digital images. Also, it is mostly used to scan documents. The software module first imports the packages and reads

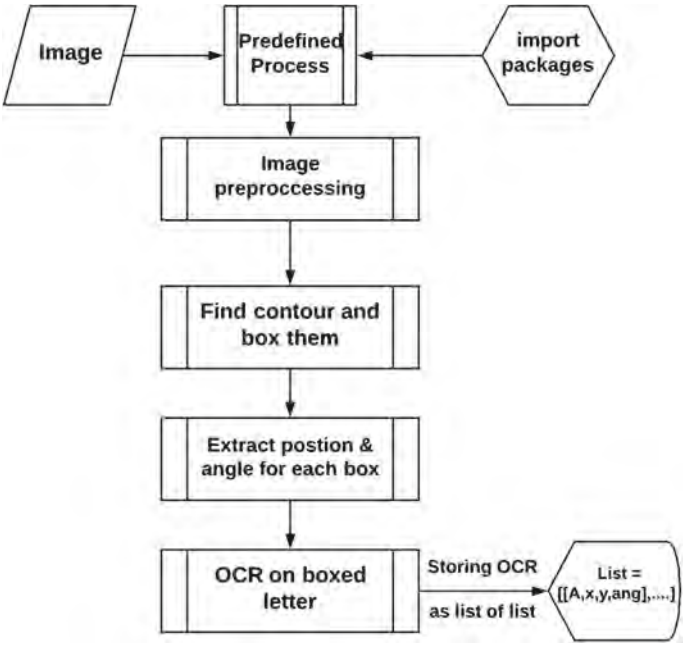


Fig. 3 Workflow of the proposed alphabets OCR

the input image. Then, it converts the image to grayscale and thresholds the image using the OpenCV library. After the recognition of the alphabet through CNN, the image is processed to find the contour and then they are boxed. After that, it extracts the position and the angle for each box. The midpoint for each box is then found, to have the arm fixed at the center of each box. Then in a repeated manner, it will perform the OCR on boxed alphabets to get the output as a list. The list will contain [*alphabets*, x and y coordinates, *rotation*]. The *alphabet* is one of the 26 possible English alphabets, the x and y coordinates represent the position on the workspace and the *rotation.* is the angle at which the alphabet is arranged. There are three main functions of the alphabets image processing, which are described below. They are preprocessing step, contour identification step, and the storage step.

Preprocessing: There is a need to preprocess the input which got as an image of the alphabets available on the workspace. The image is converted into black and white. Then threshold it using binary inversion. Also, there is a need to dilate the alphabets so that the OCR can better detect the desired alphabets.

Contour: After preprocessing our image, each alphabet has to be boded to figure out the rotated angle of each alphabet, locate the x-position and y-position of the alphabets, and perform an OCR to detect the alphabets shown in the box.

Data Storage: Once the OCR completes detecting each alphabet, it will store them as a list where each list contains four data variables: the alphabets detected, x-position, *y*-position and the angle. This array will then be passed to the robotic arm to loop through the alphabets and arrange them accordingly.

3.3 Alphabets Recognition Through CNN

Alphabets recognition was done with the convolutional neural network-based machine learning approach. Figure 4 shows the proposed CNN model used for image recognition where the image size is 48 × 48 and it passes through convolution 2D layer consisting of 48 filters with a kernel size of 3 × 3 and ReLU activation function [18, 19]. Then it passes through a max pooling 2D layer with 32 filters having a kernel size of 2 × 2 which returns the important features present in the image. This will result in the reduction of the image size. Then, it goes through another convolution layer of 24 filters with kernel size 3 × 3 and having a ReLU activation function. Later the image passes through a third convolution layer of 12 filters having a kernel size 3 × 3 and consisting of ReLU activation function. Then, having the classification layer that flattens the matrix and converts the vectors into a fully connected layer. Finally, the softmax function is used to classify the image into alphabets based on the predicted probability distribution value.

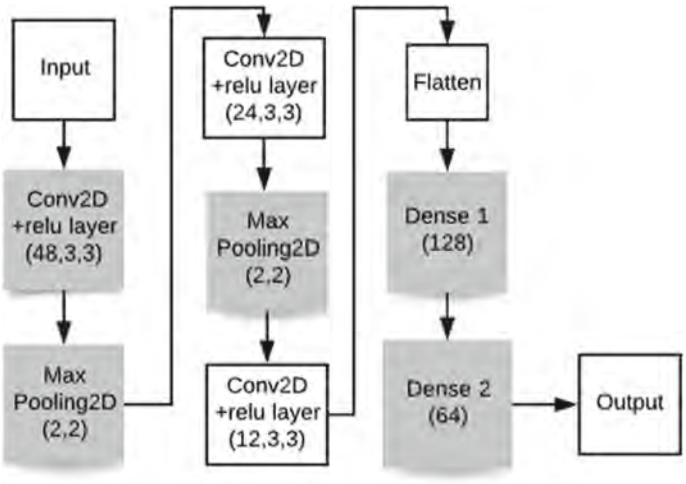


Fig. 4 Proposed CNN architecture

3.4 Hardware Design and Implementation

In the proposed system, different hardware components are integrated with the robotic arm which includes Raspberry Pi, digital camera, and RC servo motors. Figure 5 shows the circuit diagram of the hardware components. The hardware components are Raspberry Pi model B+, SSC-32U USB servo controller, six different types of HS servo motors, Pi camera, and two DC power supplies. The six HS servo motors are connected into six different channels in SSC-32U USB servo controller. Each channel contains three inputs which are pulse width modulation (PWM) pin, VCC pin, and ground pin. These inputs are arranged from top to bottom for each channel. Each servo motor has three outputs which are pulse width modulation (PWM) pin, VCC pin, and ground pin. SSC-32U USB servo controller requires a 12 V power supply connected to the VS2 because the channels that are used are connected to VS2. However, Raspberry Pi3 model B+ requires a 5 V power supply. The positive side of the DC power supply is connected to pin 4 in the Raspberry Pi, and the negative side of the DC power supply is connected to pin 6 in the Raspberry Pi. Raspberry Pi is connected with the SSC-32U USB servo controller through the USB cable. There is a special port in the Raspberry Pi (number 24) which is used to connect the Pi camera. The type of cable that is used to connect Raspberry Pi and Pi camera is CSI.

Robotic Arm: AL5D-PLTW arm is part of Lynxmotion's collection of AL5 robotic arm as shown in Fig. 6. This robotic arm has four degrees of freedom (4-DOF).

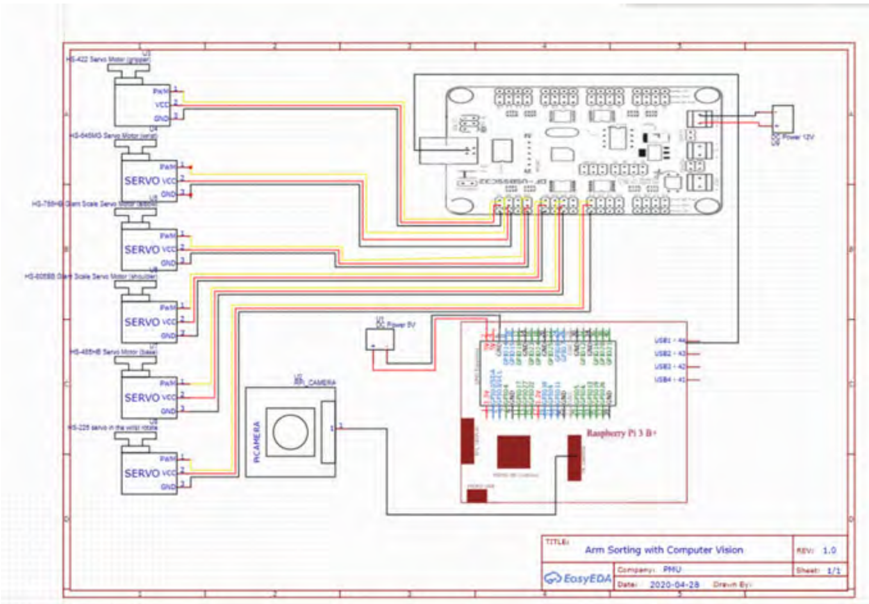


Fig. 5 Circuit diagram of the hardware components

Fig. 6 AL5D robotic arm

The AL5D robotic arm can perform repeatable movements with high accuracy [20]. One important feature is that the arm can move at high speeds with precise positional placements [21]. Dimensions of the AL5D robotic arm: shoulder to elbow: 14.605 cm. Elbow to wrist: 18.7325 cm. Wrist to tip of gripper: 8.5725 cm. Height: 18.415 cm. Height (reaching up): 48.26 cm. Median forward reach: approx. 26.035 cm. Gripper opening: 3.175 cm. Weight: 0.878 kg. Range of motion per axis: 180°.

Servo Controller: SSC-32U is a dedicated servo controller, the core of R/C servo controller is an ATmega328p chip which has a Harvard architecture with an 8-bit RISC processor core as shown in Fig. 7 [22]. The servo controller was not supposed to be programmed but was meant to receive and execute commands sent to it from an external system such as a computer or microcontroller like Raspberry Pi. The R/C servo controller has many features including control up to 32 servo motors, USB, and serial input.

Fig. 7 SSC-32U servo controller

Raspberry Pi Microcontroller: It is a small credit card-sized computer. It is not only limited to perform routine home automation tasks, but can also be used for various other applications such as home entertainment, a video game console, or anything that is programmable [23]. Latest models of Raspberry Pi have in the better processing capabilities, new features, such as a wireless and Bluetooth chip. For our research work, it is found useful based on its relation to some of the system requirements desired, such as moving the robot arm to arrange the alphabets and executing a computer vision libraries which can recognize the alphabets. Therefore, the Raspberry Pi 3 B+ model is decided to use and installed the Raspbian OS to initiate our research work.

Workspace Setup: The workspace is made up of a square wooden board. It consists of two crossed planks, which are supported on top by three wooden planks as shown in Fig. 8. These crossed planks support the camera which is placed in the middle to capture the wooden board (workspace). The size of our wooden board is roughly around $20 \times 20\,\text{in}^2$. The workspace contains the alphabets which are scattered around having random angular rotations. The robotic arm is placed at the other end of the wooden board, so it will only be able to reach the alphabets of the workspace in front of it. The distance between the camera and the wooden board is around 64 in. The alphabets are written on cubic-shaped wooden boxes. Each side of the cubic wooden box is exactly 1 in., so the robotic arm grabber (end effector) will be able to grasp it and place it at the desired location on the workspace. The position information will be given as an input to the robotic arm from the output of the OCR module explained earlier.

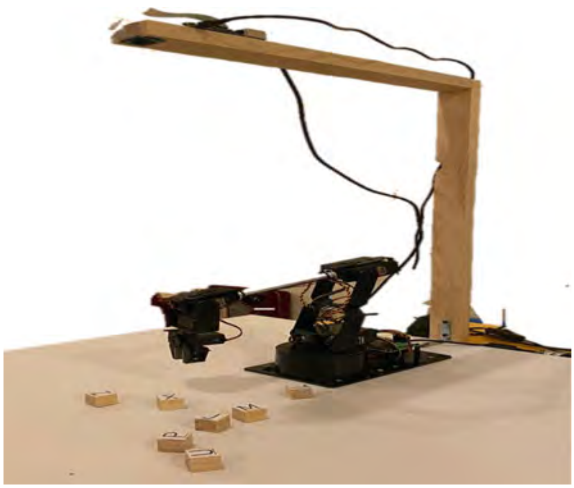


Fig. 8 Workspace setup for the robotic arm system

4 Experimental Results

In this section, the results are now presented from the experiments which were performed on the proposed system. Even though our system has two modules (hardware and software) focused on the results related to the software module, the hardware module was able to work properly based on the control signals provided to the Raspberry Pi controller from the CNN-based image classification module.

Two important accuracy measures which focused on were training accuracy and validation accuracy and also measured training loss and validation loss. These are important measures when evaluating the performance of CNN-based image classifier. The number of epochs was the independent variable which was used in this study. The number of epochs was varied from 1 to 120 and its effect on the four performance metrics is presented in Table 3. As it is expected, the accuracies increase with the number of epochs, whereas the loss functions decrease. Figure 9 shows that the training accuracy steadily increases from 21% to a maximum value of 99%, whereas

Table 3 Experimental results for alphabets recognition using CNN with different epochs

Epochs	Train Acc (%)	Train loss	Val. Acc (%)	Val. loss
1	21.46	2.7689	16.82	2.9277
15	65.83	1.0079	27.19	3.1293
30	77.81	0.6876	31.49	3.3836
60	87.50	0.4657	78.06	0.8680
120	99.06	0.2444	98.79	0.2408

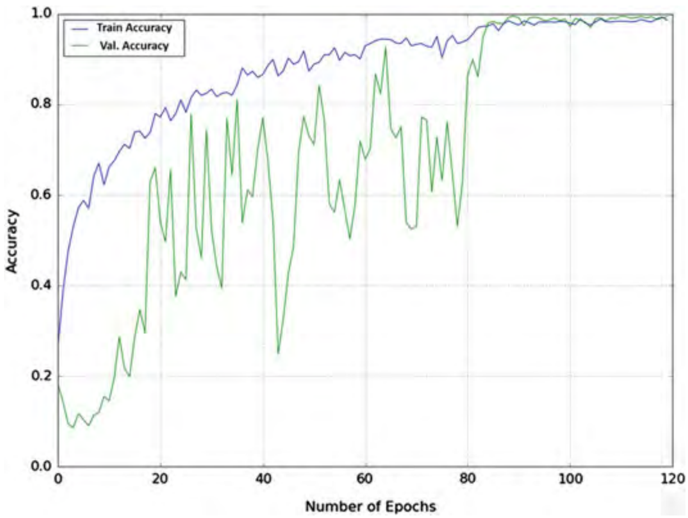


Fig. 9 Accuracy for alphabets recognition using CNN

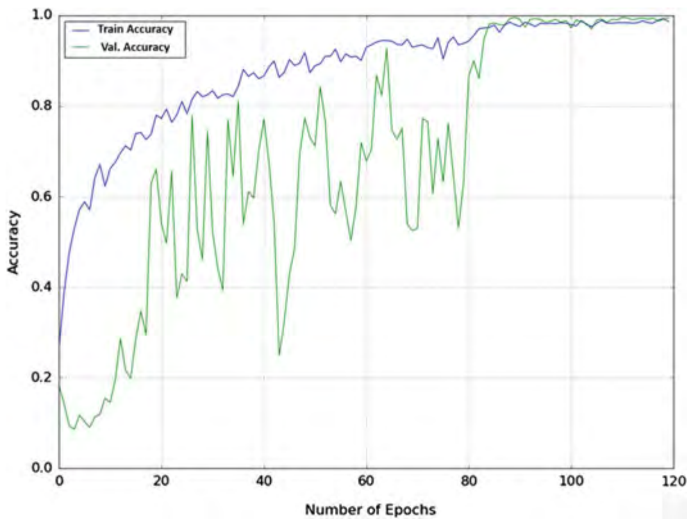


Fig. 10 Loss measurements for alphabets recognition

the validation accuracy initially has a fluctuated behavior and later on stabilizes to a maximum value of about 98.8%. Figure 10 provides a loss measurements for the recognition of the alphabet. As can be observed, both the training loss and the validation loss steadily decrease from a value of 2.77 to 2.93, respectively. At an epoch value of 120, the loss values reach a minimum value of 0.24 for both cases. By training the CNN model with 120 epochs, it found out that our system was able to correctly identify and classify the alphabets, as the accuracy was close to 99%. Further, it is observed that the robotic arm was able to correctly arrange the alphabets in the desired order according to the word which was given as an input to the system. Based on a variety of words, our system was able to demonstrate the achievement of the proposed objective which was English alphabet sorting.

5 Conclusions

In this paper, a deep learning-based CNN model has been proposed, designed, and implemented for recognizing and sorting English alphabets. Hardware and software modules are combined to provide a solution for arranging alphabets based on the desired word given as an input by the user. The hardware part of the system consists of a Raspberry Pi 3 Model B+, SSC-32U servo controller, and RC servo. The software part of the system was programmed in Python and using a convolutional neural network (CNN) machine learning model. The CNN model was trained on a dataset consisting of 3898 images of the 26 English alphabets randomly rotated at certain angles. The experimental results show that our model had a training and validation accuracies of 99.06% and 98.79%, respectively. As a part of future work, a system is planned to implement for Arabic alphabets. The changes that need to be implemented is the recognition of 28 Arabic alphabets and for this, the required layers to be used in the CNN model have to be studied. Another extension would be to speed up the training process by optimizing the various CNN layers and to reduce the identification time through faster processing microcontrollers.

Acknowledgements Our sincere thanks to the College of Computer Engineering and Science at Prince Mohammad Bin Fahd University for providing the resources and facilities for conducting this research.

References

1. Latif G, Alghazo J, Alzubaidi L, Naseer MM, Alghazo Y (2018) Deep convolutional neural network for recognition of unified multi-language handwritten numerals. In: 2018 IEEE 2nd ınternational workshop on Arabic and derived script analysis and recognition (ASAR). IEEE, pp 90–95
2. Al-Hmouz A, Latif G, Alghazo J, Al-Hmouz R (2020) Enhanced numeral recognition for handwritten multi-language numerals using fuzzy set-based decision mechanism. Int J Mach Learn Comput 10(1)
3. Alghazo JM, Latif G, Alzubaidi L, Elhassan A (2019) Multi-language handwritten digits recognition based on novel structural features. J Imaging Sci Technol 63(2):20502–20511
4. Alghazo JM, Latif G, Elhassan A, Alzubaidi L, Al-Hmouz A, Al-Hmouz R (2017) An online numeral recognition system using ımproved structural features–a unified method for handwritten Arabic and Persian numerals. J Telecommun Electron Comput Eng (JTEC) 9(2–10):33–40
5. Latif G, Iskandar DA, Alghazo JM, Mohammad N (2018) Enhanced MR image classification using hybrid statistical and wavelets features. IEEE Access 7:9634–9644
6. Latif G, Mohammad N, AlKhalaf R, AlKhalaf R, Alghazo J, Khan M (2020) An automatic Arabic sign language recognition system based on deep CNN: an assistive system for the deaf and hard of hearing. Int J Comput Dig Syst 9(4):715–724
7. Latif G, Mohammad N, Alghazo J, AlKhalaf R, AlKhalaf R (2019) ArASL: Arabic alphabets sign language dataset. Data Brief 23:103777
8. Alghmgham DA, Latif G, Alghazo J, Alzubaidi L (2019) Autonomous traffic sign (ATSR) detection and recognition using deep CNN. Procedia Comput Sci 163:266–274
9. Khan AH, Al-Asad JF, Latif G (2017) Speckle suppression in medical ultrasound images through Schur decomposition. IET Image Proc 12(3):307–313
10. Latif G, Iskandar DA, Alghazo J, Butt M, Khan AH (2018) Deep CNN based MR image denoising for tumor segmentation using watershed transform. Int J Eng Technol 7(2.3):37–42
11. AlSaid H, AlKhatib L, AlOraidh A, AlHaidar S, Bashar A (2019)Deep learning assisted smart glasses as educational aid for visually challenged students. In: Proceedings of ınternational conference on new trends in computing sciences (ICTCS 2019), Amman, Jordan, pp 1–6
12. AlZamil M, AlBugmi R, AlOtaibi S, AlAnazi G, AlZubaidi L, Bashar A (2020) COMPASS: IPS-based navigation system for visually ımpaired students. In: Proceedings of 9th IEEE ınternational conference on communication systems and network technologies (CSNT 2020), Gwalior, India, pp 161–166
13. Jia Y, Yang G, Saniie J (2017)Real-time color-based sorting robotic arm system. In: 2017 IEEE ınternational conference on electro ınformation technology (EIT), Lincoln, NE, pp 354–358. https://doi.org/10.1109/EIT.2017.8053385

14. AlSalman Z, AlSomali N, AlSayari S, Bashar A (2018) Speech driven robotic arm for sorting objects based on colors and shapes. In: 2018 3rd ınternational conference on ınventive computation technologies (ICICT), Coimbatore, India, pp 6–11. https://doi.org/10.1109/ICICT43934.2018.9034306
15. Szabó R, Lie I (2012)Automated colored object sorting application for robotic arms. In: 2012 10th ınternational symposium on electronics and telecommunications, Timisoara, pp 95–98. https://doi.org/10.1109/ISETC.2012.640811
16. Pereira V, Fernandes VA, Sequeira J (2014)Low cost object sorting robotic arm using Raspberry Pi. In: 2014 IEEE global humanitarian technology conference—South Asia Satellite (GHTC-SAS), Trivandrum, pp 1–6. https://doi.org/10.1109/GHTC-SAS.2014.6967550
17. Sarma P, Chourasia CK, Barman M (2019)Handwritten Assamese character recognition. In: 2019 IEEE 5th ınternational conference for convergence in technology (I2CT), Bombay, India, pp 1–6. https://doi.org/10.1109/I2CT45611.2019.9033603
18. Butt MM, Latif G, Iskandar DA, Alghazo J, Khan AH (2019) Multi-channel convolutions neural network based diabetic retinopathy detection from fundus images. Procedia Comput Sci 163:283–291
19. Shaikh E, Mohiuddin I, Manzoor A, Latif G, Mohammad N (2019) Automated grading for handwritten answer sheets using convolutional neural networks. In: 2019 2nd ınternational conference on new trends in computing sciences (ICTCS). IEEE, pp 1–6
20. LynxmotionLynxmotion SSC-32U USB Servo Controller Board. Retrieved from: https://www.robotshop.com/media/files/pdf2/lynxmotion_ssc-32u_usb_user_guide.pdf32u_usb_user_guide.pdf. Last accessed on 28 Apr 2019
21. Montiel-Vázquez EC, Torres-Rosique AE, Park G, Garcıa-Cavazos I, González-Flores U, y Mejia-Rosete RS, González-Hernández HG MOBMA: gesture driven mobile manipulator
22. Saleh T, Khan MR (2019) Hexapod robot for autonomous machining. In: IOP conference series: materials science and engineering, vol 488, no 1. IOP Publishing, p 012003
23. Agrawal N, Singhal S (2015) Smart drip irrigation system using raspberry pi and arduino. In: International conference on computing, communication and automation. IEEE, pp 928–932

Lung Cancer Detection from LDCT Images Using Deep Convolutional Neural Networks

Shahad Alghamdi, Mariam Alabkari, Fatima Aljishi, Ghazanfar Latif, and Abul Bashar

Abstract Lung cancer is second cancer common to men and women as well as it is one of the world's highest cause of death. Reports in recent years have shown that standard X-rays are not effective in diagnosing lung cancer. It has clinically established that low-dose computed tomography (LDCT)-based diagnosis helps to decreases mortality from lung cancer by 20% relative to normal chest X-rays. Deep learning is considered as one of the most beneficial techniques for lung cancer diagnosis. This technique used in many fields, including healthcare, which helps to facilitate complex tasks, analyze medical images, promote reliable diagnosis, and improve diagnostic accuracy. One of the deep learning algorithms is the convolutional neural network (CNN) and in this paper, different deep CNN based models are proposed for lungs cancer detection. The experiments are performed using dataset acquired from Data Science Bowl 2017 (KDSB17). The dataset consists of 6691 LDCT lung images. For testing the efficiency of the model, the accuracy is reckoned, which represents 91.75%. However, due to the sensitivity of this process, other techniques are also used to assess the model's performance including specificity, sensitivity, recall, precision, and f1-score.

Keywords Deep learning (DL) · Convolutional neural networks (CNNs) · Low-dose computed tomography (LDCT) · Lung image database consortium (LIDC)

1 Introduction

Lung cancer or lung carcinoma, regardless of race, is the principal cause of death from cancer among both men and women. More people die each year from lung cancer than from other popular cancers [1]. It has the lowest several rates among colorectal cancer (65%), prostate cancer (99%), and breast cancer (89%). In 2018, 1.76 million deaths have been estimated [2, 3]. For several decades, lung cancer was

S. Alghamdi · M. Alabkari · F. Aljishi · G. Latif (✉) · A. Bashar
College of Computer Engineering and Sciences, Prince Mohammad Bin Fahd University, Al Khobar, Saudi Arabia
e-mail: glatif@pmu.edu.sa

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_27

the most prevalent and deadliest form of cancer. It is the most common cancer among men and is the third most common cancer among women worldwide as two million new cases have been reported in 2018 [4]. Statistics show that over 50% of patients with lung cancer die within one year of diagnosis. In 2014, lung carcinoma was Saudi men's fourth most frequent cancer and Saudi women's 7th frequent cancer [5].

Medical imaging generates an immense volume of data, and thousands of images are involved in each medical study. Deep learning is used in healthcare to improve diagnosis accuracy, resolution, and promote reliable and fast diagnosis [6–9]. It can extract and process medical images at a speed and scale that exceed human capabilities and analyze more efficiently. Even though lung cancer is the deadliest type of cancer, it is highly curable if diagnosed early. Catching lung cancer before spreading can add years to human life. It can increase the possibility of survival for five years or more by 55% [1], so it is important to invest in a system that assists in early lung cancer detection. England statistics show that 88% of lung cancer patients who diagnosed at the first stage survived for at least one year compared to 19% of those diagnosed at the fourth stage [2]. Even though it is the number one cause of death from cancer, less money is spent on research into lung cancer as compared with other rising cancers.

Deep learning algorithms consists of several layers which extract higher-level characteristics from the raw input. This consists of several layers for extracting characteristics of higher levels from the raw inputs [10, 11]. Each layer in the network transforms its input into a more complex and abstract representation. The first representational layers learn to detect simple feature filters such as edges and corners while the middle representational layers learn more complex feature detection filters such as part of an object. The last layer will learn to recognize the full object. The research puts in a profoundly convolutional neural network model (CNN) to diagnose lung cancer patients into two classes: have cancer and does not have cancer. The network comprises two convolution layers, two max-pooling layers, two drops out layers, two fully connected layers, and a flattened layer.

In recent years, doctors found that normal X-rays are not appropriate for the diagnosis of lung cancer. They found that low-dose computed tomography (LDCT) decreases mortality from lung cancer by 20% relative to normal chest X-rays. comparison to regular chest radiography [2]. LDCT can detect pulmonary cancer early on, which not measurable with a standard X-ray. Thus, a data set consisting of 6691 LDCT images have been used for the experiments in this research [12].

2 Literature Review

Serj, Lavi, Hoff, and Valls gave a new deep convolutional neural network. (dCNN) model for learning high-level image representation to achieve highly accurate results with low variance using low-dose computed tomography (LDCT) images [13]. The researchers' goal was building a binary classification model that learns discriminant compact features at the beginning of the network to detect lung cancer. The researchers proposed a new deep convolutional neural network architecture consists of three convolution layers, two max-pooling layers, a full-connected layer, and a binary soft-max layer. The proposed model commences with several sequential convolution layers to generate high-order convolutional features. The researchers used a data set from the Data Science Bowl 2017 (KDSB17) to confirm the model results. The dataset consists of 63,890 (LDCT) images of cancer patients and 171,345 images of non-cancers. The researchers divided the data set into testing, training, and cross-validate set; 50% for the training and 25% for the validation and the rest for the testing. The researchers used cross-entropy as a loss function to maximize the probability of having cancer by maximizing the multinomial logistic regression objective. The results of the model were impressive. The performance of the model is measured using specificity (0.991), sensitivity (0.87), and F1 score (0.95).

Chon and Balachandar have used the deep convolutional neural network for lung cancer detection of CT scans [14]. They used a data set from Date Science Bowl 2017 along with updated U-Net that trained on data set LUNA16, which represents CT scans with marked nodules. At first, they started to pass the CT scans into the 3D CNNs directly for classification, which gave them a poor result. Then, to insert only the regions of interest into the 3D CNNs, they had to perform further pre-processing. For this to be the case, U-Net that trained on the LUNA16 data set was used (CT scans with labeled nodules) for identification in CT scans of nodule candidates. This process produced many false positive predictions, so they used the CTs scans to specify where the nodules located as determined by the U-Net outputs were fed into 3D convolutional neural networks. That ultimately helped to identify CT scans for lung cancer either as positive or negative.

Several kinds of research have applied lung cancer diagnosis to manipulating images and machine learning approaches. Makaju, Prasad, Singh, and Alsadoon had two models, their best model was not reliable and failed to identify the findings of cancer found in the nodules [15]. Therefore, they also introduced a new method for detecting cancer nodule from the CT scan image using the watershed segmentation to identify along with SVM to classify a nodule as malignant or benign. This model has identified 92% accuracy of cancer, this is better than the previous model which had 86.6% accuracy. Regarding their data set, they used actual CT scans from the Lung Image Database Consortium (LIDC) archive of patients. This collection of pulmonary cancer CT images for computer-aided diagnostic methods for the identification and treatment of lung cancer, which was introduced by the National Cancer Institute. The dataset was composed of 1018 cases which supported 7 research centers and 8 medical imaging firms. Images were in DICOM format with 512 * 512 pixels in size, and as DICOM format was hard to process; those images have been converted to JPE grayscale images the aid of MicroDicom software, which converts the DICOM CT scan images to JPEG format.

Sharma and Jindal obtained their CT images from NIH/NCI Lung Image Database Consortium (LIDC) data set that is provided for research purposes [16]. They obtained an automated CAD program to diagnose lung cancer early they achieved that by examining the CT images in multiple steps, their method started by extracting pulmonary areas from the CT picture using multiple image processing strategy, including bit image slicing, erosion, and Weiner filter. Instead of using the thresholding technique, they used the bit plane slicing technique that was used to convert the CT image into a binary image during the first stage of the extraction process. This strategy is faster and user autonomous compared to the thresholding technique. Afterwards, the lung regions extracted were segmented using region-widening segmentation algorithms. Lastly, the field has been used to classify cancerous regions and to obtain an objective result with a high sensitivity level of 90%, with a fair number of false positives per picture with 0.05 false positives per picture.

Medial image processing gained more importance in last decade due to the availability of high computational power and advancement in the imaging techniques such as medical image enhancement [17], noise removal [18], medical image classification [19], and medical image segmentation [20].

3 Methodology

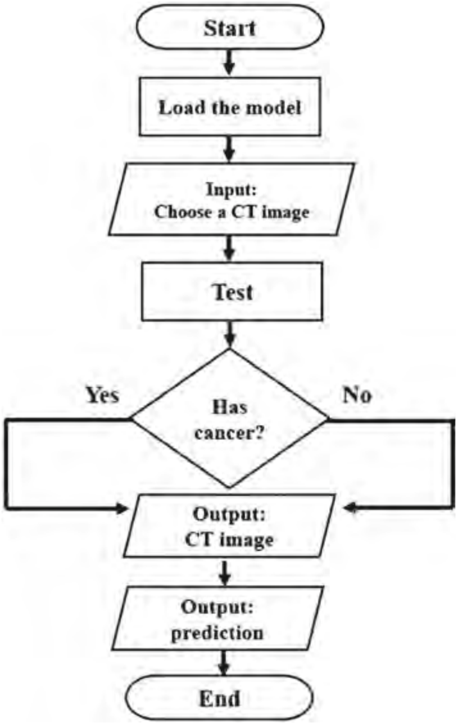
In this section, the models that have been tried will be discussed. The deep convolutional neural networks (dCNN) is generated by an input layer, hidden layers, and an output layer. It composes of two major parts: feature extraction and classification. The first layers learn basic detection filters while the complexity increases in the middle layers. The biggest advantage of dCNN, the developer is not expected to extract features manually from the image. During the training, the network learns to extract features. The classification decision will be guided in the last layers based on the features extracted from the preceding layers. Figure 1 represents the process of lung cancer detection system.

3.1 Classification Through Deep CNN

Medical images are often corrupted by noise, lighting, and affected by artifacts; this could have an impact on model accuracy. So, to limit these phenomena, the data set was divided by 255. 75% of the images were used for the training, and 25% for the testing. To balance the classes in each set, the data has been divided based on the labels, and the random state is set to 42. Before deciding our model, 4 deep convolutional neural network architectures were tried then compared the results and chosen the most accurate, which is the fourth model. Figure 2 shows the workflow of the deep convolutional neural network for lungs cancer detection.

The first model consists of three sequential convolution layers, two max-pooling layers, two fully connected layers, and a flattened layer. Firstly, the network begins with an input layer which is the first convolution layer; the first layer will take the image with an input size of 120×120 pixels. It consists of 50 filters and convolution kernel of 11×11. Secondly, another convolution layer has been added. The second layer consists of 120 filters and convolution kernel of 5×5. Thirdly, a max-pooling

Fig. 1 Process of lung cancer detection system



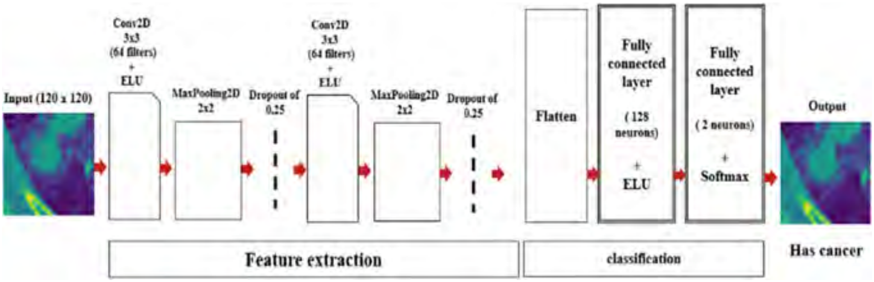


Fig. 2 Workflow of the deep convolutional neural network for lungs cancer detection

layer has been added with a pool size of 2 × 2 and strides of 2. Fourthly, a third convolution layer with 120 filters and convolution kernel of 3 × 3 is added. Fifthly, a max-pooling layer has been added with a pool size of 2 × 2 and strides of 2. RELU is used as an activation function in all the convolution layers. This model did not get the best results, this might be because the flattened layer is added after the first dense layer.

In the second model, the first convolution layer is the input layer. The input layer consists of 32 filters and with the convolution kernel of 2 × 2. After that, the batch normalization technique is used to speed up the training, reduce overfitting, and to put all our data on the same scale. Similar to batch normalization, the drop-out layer will help in reducing overfitting. After that, a convolution layer consists of 64 filters is added with the convolution kernel of 2×2. After that, the batch normalization is used and the drop-out layer again. After the feature extraction and normalization, the output of the final convolution layer will be the input of the flatten layer. The flatten layer will convert the data into a one-dimensional array and pass it to the first fully connected layer. After that, a drop-out layer is added between the dense layers because they have the largest number of parameters and could cause overfitting. Finally, the output layer will pass its output to the last fully connected layer, and the classification decision will be made.

The third model consists of an input layer that takes an image input (120, 120, 1); this convolution layer applies 50 kernels (filters) of size 50 with kernel size (11, 11). Then, the process of batch normalization is used to improve training and prevent overfitting. As with batch normalization, the drop-out layer helps to reduce overfitting. After that, a max-pooling layer is used with pool size 2×2 and strides of 2. A layer of convolution that consists of 80 filters with 3×3 kernel size is applied. Following that, batch normalization is used and the drop-out layer again. Afterwards, max-pooling layer with pool size 2×2 and strides of 2 is added. Finally, the output of the last layer will pass through the flatten layer which will convert the data into a 1-dimensional array and pass it to the first fully connected layer.

Because having one input and one output, the network begins with a sequential convolution layer consisting of 64 filters with a convolution kernel of 3×3. The layer uses the exponential linear unit (ELU) as an activation function. In addition, it takes the LDCT image of an input size of 120×120 pixels. Secondly, a max-pooling layer with a 2×2 pool size has been added. This layer was used to minimize the number of parameters along with minimizing the spatial size of the representation. Therefore, the computational cost and overfitting will be reduced [18]. Similarly, a dropout of 0.25 has been added to prevent the model from overfitting. Fourthly, a convolution layer consists of 64 filters a convolution kernel of 3×3 is added. The fourth layer uses ELU as an activation function. Fifthly, a max-pooling layer with a 2×2 pool size is added. Sixthly, a dropout of 0.25 is added. After that, two flatten layers is added to convert the pooled feature map to a single column to input it to the fully connected layer. Finally, two fully connected layers are added to take the results of the convolution and pooling process and use them to drive a classification decision. As shown in Fig. 1, the first fully connected layer uses ELU as an activation function while the second fully connected layer uses SoftMax. The SoftMax is usually applied in the last layer of the neural networks instead of using RELU, Tanh, or Sigmoid. It is used because it converts the input into values between 0 or 1, so they can be interpreted as probabilities [19]. After constructing the layers, the model is trained for 100 epochs. In addition, 10% of the training set is used for validation. Because Adam optimizer is used, the accuracy increased in each epoch.

3.2 Experimental Data

The data is acquired from the cancer imaging archive (TCIA) as a file of un-labeled scans. The file contains two datasets, 'ct_slices' which had 6691 slices of the original scans to help with the sharpness of the image when resizing it [12]. All the images are sorted in a hierarchical data format (HDF5). The second dataset 'slice_class' had the labels of each slice. The labels used were 0 and 1 to indicate the presence of cancer. The 6691 images consist of 2526 cancerous images, and 4165 non-cancerous images. Sample images of the CT lung cancer are shown in Fig. 3.

The original size of the images was 64 × 64 pixels, yet all the images were resized to 120 × 120 pixels by using resize function from cv2 library. This helped to obtain clear slices and detect cancer cells accurately while using the model. NumPy library which contains reshape function is used to reshape the images from [6691, 64, 64] to the new shape [6691, 120, 120, 1]. Reshaped it from a three-dimensional array to a four-dimensional array to use it with the model. 6691 represents the number of the images, 120 represents the height and width of the images, and 1 represents the number of channels in the images as all the images are in grayscale. Training the model by using grayscale images may increase the model performance because the model will focus on the shape of the images rather than the colors of each image [21]. Also, used to categorical function to convert the labels to a matrix to use it with the model since having two classes 1: cancerous and 0: non-cancerous. The dataset

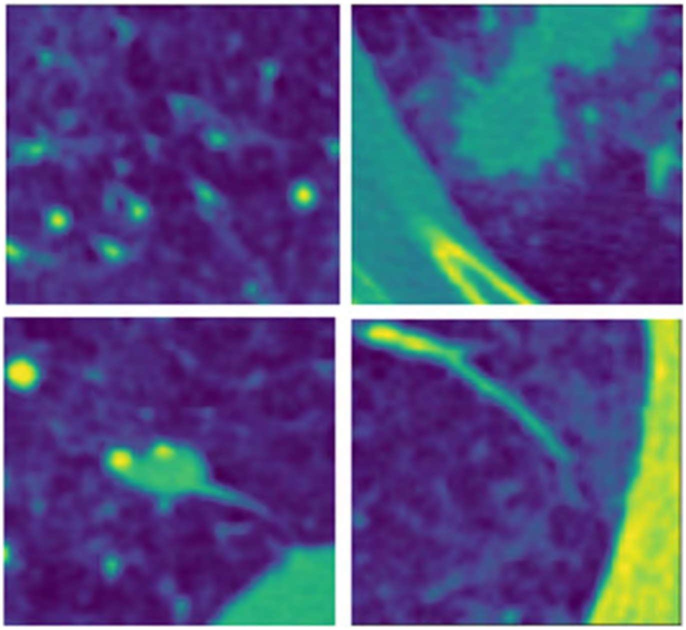


Fig. 3 Sample of low-dose computed tomography (LDCT) lung images

has been divided randomly into three categories: 65% of the images used for the training, 10% used to validate the model during the training process, and 25% used to test the model.

4 Experimental Results

The dataset has been divided randomly into three categories: 65% of the images used for the training, 10% used to validate the model during the training process, and 25% used to test the model. The dataset has been divided based on the labels to guarantee that each category includes images of both classes 0: non-cancer and 1: cancer. To achieve the objective of this research, four different experiments are conducted along with different features for the best results. In the first experiment, the model is trained with 200 epochs, which means 200 training cycles. In addition, the model was complicated, because it contained many layers with a high number of filters and big kernel size. Thus, it was the cause of overfitting. In the second and third experiments, tried to improve the model by decreasing the number of epochs from 200 to 100 cycles. The number of layers used is decreased along with reducing the number of kernels and kernel size. Different parameters are also applied to avoid the overfitting and increase the model performance. This helped to avoid overfitting; however, it affected the accuracy of the model. An accuracy of 91.75% is achieved in the fourth experiment, as long with high model performance.

The accuracy of the four experiments is calculated to evaluate the overall efficacy of the classification process. The accuracy is simply representing the percentage of the predictions that the model has been getting correct. Formally, the accuracy determined by using the formula below for binary classification:

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

Due to the sensitivity of this process, several techniques is used to evaluate our model's efficiency, including specificity, sensitivity, recall, precision, and f1-score. This will help us to calculate the percentage of the right and wrong predictions, which will help to avoid having wrong predictions since the software could have an impact on society as it could be used in medical diagnosis. A wrong diagnosis will affect individuals' health. As shown in the table below, a recall is determined showing the right positive rating average from all the real positive ratings. In addition, the precision is calculated from the cases predicted as positive that represents the rate of correct positive classification. The f1-score helped us to combine the recall and precision values to balance and see the overall results.

In medical image analysis, model performance usually measured using sensitivity and specificity, which is the percentage of true positive and true negative. The geometric mean rate shows a combination of sensitivity and specificity in a single matric.

Table 1 Comparison of the accuracies of different proposed DCNN Models

Epochs	Accuracy (%)
Proposed dCNN model M1	83.02
Proposed dCNN model M2	77.23
Proposed dCNN model M3	71.73
Proposed dCNN model M4	91.75

Different models were implemented to compare their performances and choose the best accuracy among them where the fourth model had the highest accuracy of 91.75% as shown in Table 1. The variance of the accuracy was because of the changes made in each model including, the type of layers added, number of kernels, kernel size, number of epochs (training cycles), and type of activation functions. In addition, more parameters have been added to boost model efficiencies such as kernel initializer, Adam optimizer, Elu activation, and some padding.

A sequential model is created to add multiple layers to the model, where each layer has only one input and one output. The proposed deep convolutional neural network model consists of two convolution layers, two max-pooling layers, two drop-out layers, a flatten layer, and two fully connected layers. Multiple layers are applied and repeated some of the layers multiple times for higher performance, along with extract more features from the images which may help in the classification process.

The two convolution layers applied 64 kernels of the size of 3 × 3. The 3 × 3 filter is the smallest and the most commonly used. In the fourth model, the size of the filter is tried to reduce from 11 × 11 to 3 × 3. This is because a smaller filter size of an odd number is preferred over large filter size of an even number. This will help simplify the image processing for better performance of the convolution layers, along with the number of kernels (filters) used to extract useful features from the images.

Elu activation function is used which more likely to converge cost to zero faster. In addition, Elu tends to produce more accurate results, and it combines the good features of ReLU and Leaky ReLU. Moreover, Elu does not have the main problems of ReLU. The same padding in both convolution layers is applied, which helped to keep the dimensions of the output the same as its input. That means the convolution layers output size is the same as the input. For instance, the input shape of the first convolution layer was (120, 120, 1) and the output shape was also (120, 120, 64). This helped to not reduce the features of the images and lose any important feature, which may help in the classification process.

Two drop-out layers are added after applied the Elu activation function inside the convolution layers; this technique helped us to avoid having overfitting. Each hidden unit (neuron) is set to 0 with a probability of 0.25 at passing 0.25. This means that there is a change of 25% forcing the neuron output to 0. The detailed proposed model's performance analysis is presented in Table 2 which compares different proposed models experimental results based on PrecisionRecall F1-score Specificity.

Adam optimizer is used from Keras library which helped to improve the model and increase the accuracy during the training process. This will guarantee that a

Table 2 In-depth proposed model's performance analysis

#		Precision	Recall	F1-score	Specificity
M1	Non-cancerous	0.87	0.86	0.86	0.79
	Cancerous	0.77	0.79	0.78	0.86
M2	Non-cancerous	0.74	0.98	0.84	0.42
	Cancerous	0.94	0.42	0.58	0.98
M3	Non-cancerous	0.70	0.97	0.81	0.30
	Cancerous	0.86	0.30	0.44	0.97
M4	Non-cancerous	0.92	0.95	0.93	0.86
	Cancerous	0.92	0.86	0.89	0.95

higher accuracy will get at each training cycle. In addition, Keras initializer is used to pass initializers to the layers. In addition, the number of dense is increased, which represents the number of neurons the full connection layer will connect to from 10 to 128 connections; this helped to get a perfect test error.

Lastly, the epoch number was reduced from 200 to 100 to prevent overfitting, because the first model is trained with 200 epochs, and the model started to memorize the images which gave us 1 accuracy during all the training cycles. During this process, the model has trained with only 6691 (LDCT) images, so no need to have a high number of epochs unless the number of images used is increased.

5 Conclusion

Studies have shown that computed tomography (LDCT) based diagnosis helps to decreases mortality from lung cancer by 20% relative to normal chest X-rays. The deep learning technique helped to analyze the (LDCT) medical images and promoted reliable diagnosis and improved the diagnosis accuracy. The paper proposed profoundly convolutional neural network architecture for binary lung cancer classification. Convolutional neural network (CNN) is one of the profound learning algorithms, which used to achieve high classification accuracy for tasks involving medical images. The convolutional neural network (CNN) model consists of two convolution layers, two max-pooling layers, two drop-out layers, a flatten layer, and two fully connected layers. CNN layers were used to extract features from the images, which permit the model to be trained using these features to be able to make the prediction. The models presented, predicted, and classified the cancerous and non-cancerous slices with 91.75% accuracy. However, due to the sensitivity of this process, a lot of methods are used to assess our model's performance, including specificity, sensitivity, recall, precision, and f1-score. These techniques helped to calculate the percentage of true positive, true negative, false positive, and false negative levels, which will help to avoid having wrong predictions since the software could have an impact on society as it could be used in medical diagnosis. This process will help to promote a reliable diagnosis along with improving the diagnosis accuracy.

References

1. Desai A, Gyawali B (2020) Fall in US cancer death rates: time to pop the champagne? EClinicalMedicine 19
2. Rindi G, Klimstra DS, Abedi-Ardekani B, Asa SL, Bosman FT, Brambilla E et al (2018) A common classification framework for neuroendocrine neoplasms: an International Agency for Research on Cancer (IARC) and World Health Organization (WHO) expert consensus proposal. Mod Pathol 31(12):1770–1786
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68(6):394–424
4. Latif G, Butt MM, Khan AH, Butt O, Iskandar DA (2017) Multiclass brain Glioma tumor classification using block-based 3D Wavelet features of MR images. In: 2017 4th ınternational conference on electrical and electronic engineering (ICEEE). IEEE, pp 333–337
5. Jazieh AR, AlGhamdi M, AlGhanem S, AlGarni M, AlKattan K, AlRujaib M et al (2018) Saudi lung cancer prevention and screening guidelines. Annals Thor Med 13(4):198
6. Kim M, Yun J, Cho Y, Shin K, Jang R, Bae HJ, Kim N (2019) Deep learning in medical imaging. Neurospine 16(4):657
7. Butt MM, Latif G, Iskandar DA, Alghazo J, Khan AH (2019) Multi-channel convolutions neural network based diabetic retinopathy detection from fundus images. Procedia Comput Sci 163:283–291
8. Latif G, Iskandar DA, Alghazo J, Butt MM (2020) Brain MR ımage classification for glioma tumor detection using deep convolutional neural network features Current Med Imag
9. Latif G, Iskandar DA, Alghazo J, Butt M, Khan AH (2018) Deep CNN based MR image denoising for tumor segmentation using watershed transform. Int J Eng Technol 7(2.3):37–42
10. Alghmgham DA, Latif G, Alghazo J, Alzubaidi L (2019) Autonomous traffic sign (ATSR) detection and recognition using deep CNN. Procedia Comput Sci 163:266–274
11. Latif G, Alghazo J, Alzubaidi L, Naseer MM, Alghazo Y (2018) Deep convolutional neural network for recognition of unified multi-language handwritten numerals. In: 2018 IEEE 2nd ınternational workshop on Arabic and derived script analysis and recognition (ASAR). IEEE, pp 90–95
12. Jadhav S (2020) Lung cancer detection using classification algorithms (Doctoral dissertation, Dublin, National College of Ireland)
13. Frank M, Drikakis D, Charissis V (2020) Machine-learning methods for computational science and engineering. Computation 8(1):15
14. Chon A, Balachandar N, Lu P (2017) Deep convolutional neural networks for lung cancer detection. Standford University
15. Makaju S, Prasad PWC, Alsadoon A, Singh AK, Elchouemi A (2018) Lung cancer detection using CT scan images. Procedia Comput Sci 125:107–114
16. Sharma D, Jindal G (2011) Identifying lung cancer using image processing techniques. In: International conference on computational techniques and artificial ıntelligence (ICCTAI), vol 17, pp 872–880
17. Khan AH, Al-Asad JF, Latif G (2017) Speckle suppression in medical ultrasound images through Schur decomposition. IET Image Proc 12(3):307–313
18. Al-Asad JF, Khan AH, Latif G, Hajji W (2019) QR based despeckling approach for medical ultrasound images. Current Med Imag 15(7):679–688
19. Latif G, Iskandar DA, Alghazo JM, Mohammad N (2018) Enhanced MR image classification using hybrid statistical and wavelets features. IEEE Access 7:9634–9644

20. Latif G, Iskandar DA, Jaffar A, Butt MM (2017) Multimodal brain tumor segmentation using neighboring image features. J Telecommun Electron Comput Eng (JTEC) 9(2–9):37–42
21. Joulin A, Cissé M, Grangier D, Jégou H (2017) Efficient softmax approximation for gpus. In: International conference on machine learning, pp 1302–1310

Global Integration and Distribution of Data Through Machine Learning for COVID-19

E. Laxmi Lydia, Jose Moses Gummadi, Chinmaya Ranjan Pattanaik, G. Jaya Suma, A. Krishna Mohan, and Ravuri Daniel

Abstract COVID-19 is referred to as a broad disaster struck out in the society as a challenge. The large quantity of statistical information requires machine tools to improve the knowledge and accelerate COVID-19 forecast, analysis, and its corresponding remedial measures. But, to evade global hazards in these applications, open research will be made mandatory. This article uses machine learning model to integrate COVID-19 data and distribute the data globally. Machine Learning (ML) solutions that rely on COVID-19 data use the SIR model and logistic regression model to analyze how the pandemic cycle propagates all around the inhabitants. The figures of the SIR model concentrate on time-driven case scenarios to predict the behavior of infection, whereas the cumulative cases have become more reliant on significant-data plots.

Keywords COVID-19 · Machine Learning (ML) · SIR model · Logistic Regression (LR)

E. Laxmi Lydia (✉)
Department of Computer Science and Engineering, Vignan's Institute of Information Technology (A), Visakhapatnam, Andhra Pradesh, India
e-mail: elaxmi2002@yahoo.com

J. M. Gummadi
Department of CSE, VFSTR (Deemed To Be University), Guntur, India
e-mail: josemoses@gmail.com

C. R. Pattanaik
Department Computer Science and Engineering, Ajay Binay Institute of Technology, Cuttack, Odisha, India
e-mail: chinmaya.pattnaik@rediffmail.com

G. Jaya Suma
JNTUK-UCEV, Vizianagaram, India

A. Krishna Mohan
Department of Computer Science and Engineering, JNTUK, Vizianagaram, Andhra Pradesh, India

R. Daniel
Department of Computer Science and Engineering, Bapatla Engineering College (Autonomous), Bapatla, India
e-mail: danielravuri@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_28

1 Introduction

The first coronavirus case has been originated from the city of Wuhan, China and infected many humans across the globe since December 31, 2019. Organizations from China and World Health Organization (WHO) declared the existence of novel coronavirus and also reported the related diseases, especially SARS-CoV-2 and COVID-19. 213 countries got infected to date by crossing 2.4 million virus cases worldwide [1] with death cases 4, 18,135 till June 10, 2020. COVID-19 was diagnosed and pointed as a great danger to global health. The unusual drastic increase of virus and its roots escalation was underlined unknowingly within the society. Global distribution of experimental COVID-19 data overture assuring tools to flap against the disease. More than 12,400 editorials were printed in the preceding five months [2], and the information regarding COVID-19 thousands of patients were recorded and broadcasted [3]. The maximum investigation of the data was gathered with blood sampled and interpreted. Certainly, the standard mechanism to grab success over this pandemic circumstance experiences two common asserts. The demand for proper input and substantial expertise is necessary. Secondly, it should maintain timely assumptions. These two assets were meet by machine learning pathologies [4, 5] comprising of disease infections [6]. The following are the two statistical approaches using machine learning for the COVID-19 pandemic and also the process to share data and uphold global collaboration.

2 Literature Survey

Machine learning phase medication for COVID-19 is an interpretive framework that usually takes a few assertions into account at a time. For comparative purposes, out of over 1,200 pre-clinical studies documented for symptoms for COVID-19, the majority of the population continues to focus is on a single medicine or a few drugs, hand-selected based on different relations [7]. The machine learning process can be broadened to choose a set of alternative antiviral; several patterns of DNAs and/or protein molecule that can anticipate interrelationships among medicines and the virus, including potential SARS-CoV-2 drug sites, to allow impressive treatment methods for candidates [8, 9]. ML has indeed been employed similarly in certain contagious disorders [10]; for instance, a deep neural network has been specially managed to monitor the operation of the *Escherichia coli* [11] activity of more than 100 million organic compounds. In a certain case, a broad spectrum of vaccination patients could also be tested individually in order, for obvious reasons by conveying the dramatic drop protein S observing a SARS-CoV-2 infection [12], to develop successful immune reactions. Such prospective pathways do not, however, mask the complexities of pharmacological ML-based work. Initially, ML still could not stimulate fundamental biology, and it a remarkable problem even without the prediction of protein [13].

Therefore, there is indeed mandatory time duration in the particular instance of vaccines. Secondly, a massive essential problem is the desire to avoid appropriate laboratory tests: The latest hydroxychloroquine-based therapeutic research [14] has indeed documented functioning with very small samples, which don't use sufficient model or obfuscating the eligibility criteria. With ML algorithms, this risk could significantly increase. Methodologies, along with deep neural networks, are "broad sense classifiers." They could be organized to match each primary goal on a dataset by memorizing medications of all patients. Analytical techniques are sometimes assessed only in a compelling way by evaluating their ability to estimate an independent test set.

3 Methodology

3.1 Alleviate the Workload of Medical Experts Through Machine Learning

Although conventional statistical techniques might yield the very first required responses through a disaster, they also enable significant information systems that are distinctly insufficient during this scenario. Health care institutions easily became overloaded, and the capacity for statistical analyses was restricted mostly beyond the amount of effort expected, related to clinical investigation. Predictive models may reduce the period deemed necessary through analytical research and assist clinicians via artificial intelligence practitioners. Diagnostic imaging analyses, for instance, demonstrate that chest verified tomography (CT) diagnostic tests will be utilized to pinpoint COVID-19 infections [15, 16]. Conversely, this research method involves a skilled and experienced radiologist to evaluate each sample, who otherwise would be operating across the front lines. ML can help solve this task: for the diagnosis of 14 distinct lung infections, subsequently supervised classifiers selected around a set of data of 400,000 chest X-Rays managed to achieve a normalized area under its patient operating characteristic curve (true positive rate vs. false positive) of 94% [17].

Also, observational research focused on several hundred chest cartridge diagnostic tests indicates whether COVID-19 can be diagnosed often with ML software automatically [18]. Consequently, clinical images ML is currently restricted to extremely minor comparisons to predict or diagnose COVID-19; many such investigations, therefore, have little influence over the many confusions (e.g., age, corpulence) here which methodologies can find from images in the chest. One suitable strategy is to pre-train machine learning models from bigger image datasets and develops learning computational attributes that might be used to promote COVID-19 image training. In recent years, this technique has been used consistently in computer vision to produce successful outcomes with very few case studies [19].

3.2 Essential Data Integrity

ML is pertinent to speed up the analytics of challenging and complicated statistical techniques, including large genomic or medical imagery data samples, although standard statistical analyses are assimilated to many clinical and epidemic issues. Cumulative ML is also planned instead of replacing traditional diagnosis, forecast, and treatment methods. Nevertheless, the environmental effect of ML is currently constrained by two major issues. Initially, it is notorious that ML algorithms are not easy to understand. While visualization tools may outline the set of variables that led technology to find a certain estimation.

Systemic perceptions (e.g., measuring tool, personage, etc.) can eventually influence ML. Distinctive analytic activities are therefore made to create normal concern in scientific articles as well as in clinics concerning ML results. Furthermore, the absence of huge corporate repositories for public health, clinic, imagery, and genetics leads each organization to localize their analytical platform within its limited data, which noticeably restricts results interpretation. Although this problem does not relate to ML, sophisticated methodologies should lead to disparate datasets (1) accessing at every new treatment of de-anonymized statistical data and (2) the creation of large samples. The ISARIC project intends to choose a robust and popular clinical database of COVID-19 patients [20]. The International Severe Acute Infection Consortium (ISARIC) data-sharing protocols are signed by several other managements to make ensure data which is extensively and easily shared [21, 22] such that innovative concepts are updated, and this is only accomplished in part, making it very hard to use the frequent data captured throughout the deadly virus.

This will be important to facilitate successful initiatives across cultures and various types of health care facilities [6], and perhaps, even directly relies on the scale, functionality, and interpretability of these databases. The interactive exchange of clinical databases calls for careful management of regulatory problems and privacy. Throughout the event that many government agencies do not ordinarily operate, rapid resolution of such problems can be exceedingly difficult. Yet machine learning cannot maintain its commitment which will counteract the virus once to overcome many challenges.

3.3 SIR Model with Machine Learning Regressions

The main objective was to develop a predictive framework for the interpretation of the vital factor influencing the data transfer of COVID-19. SIR is a current idea that takes into account the population from one of the countries listed:

(a) Susceptible (S). The patient has not developed symptoms, but that can be contaminated by affected individuals
(b) Infected (I) with bacteria. This individual does have the infection.

(c) Recovered (R) or Dead. Anyone of two individual lives seems to be the epidemic: whether another individual persists or the patient has become resistant to this disease or has passed away.

Multiple implementations of this model are available, concerning birth and death (SIRD including demographic).

Due to the advancement in COVID-19 aids to be concerned everywhere to enhance the future predictions, furthermore, it is important to discover that individuals gain immunity (a prolonged probability of losing immunity and of returning COVID-19 up in a particular variability like influenza virus) even though there is no transformation from being regained to any of the two States. The differential equations governing the device are as follows:

$$\mathrm{d}S/\mathrm{d}t = -\frac{\beta \mathrm{SI}}{N}$$

$$\mathrm{d}I/\mathrm{d}t = \beta \mathrm{SIN} - \gamma I$$

$$\mathrm{d}R/\mathrm{d}t = \gamma I$$

Here, β denotes the contagion rate of the pathogen, and γ denotes the recovery rate of the epidemic.

The strategy of the SIR has been applied across several aspects: inside a medium-range exponential function of the differential equations and also in the virtual community (graph) further with dynamics. For consistency purposes, the first alternative approach was preferred and implemented for the differential equation system using a numerical method (Runge–Kutta). Consequently, it is necessary to classify the key parameters to get the progression of the disease and notify the rk4 approach.

3.4 Procedure for Standard Logistic Regression Model

(a) Initially choose appropriate features.
(b) Filter data from a certain period.
(c) Perform log transformations to reported cases and deaths.
(d) Substitute logarithm infinity by 0. With the asymptotic logarithm behavior of log(0), a reverse (exponential) transformation has been achieved.
(e) Split the data for training, validating, and testing.
(f) Prediction results.
(g) Finally, report the outcome within that appropriate sequence and execute an exponential transformation to the reverse log.

4 Results

The rate of deaths in India is predicted by the cumulative number of confirmed COVID-19 contexts. Results were carried out using Kaggle data collection, i.e., population by country 2020.csv. Figure 1 shows the SIR model; Fig. 2 shows logistic regression for Indian infected cases.

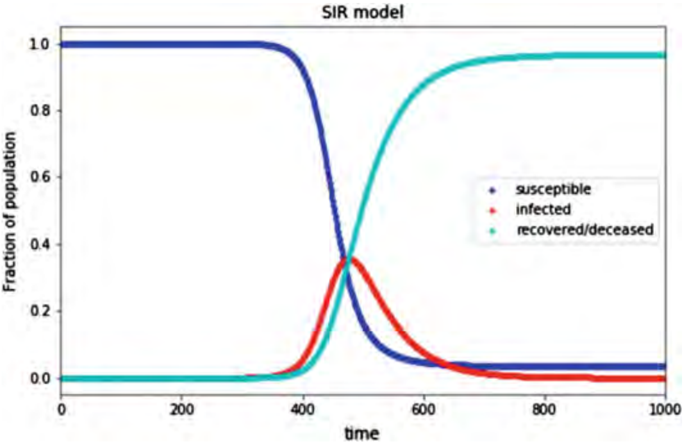


Fig. 1 Plot for SIR model over a fraction of the population

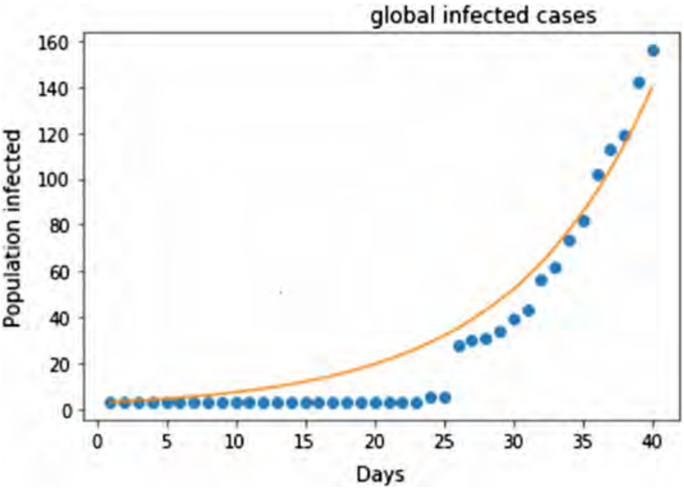


Fig. 2 Logistic regression plot for global infected cases

5 Conclusion

The COVID-19 contagious disease spreading the virus all over but rather it is highly improbable to be the last, all over the world. Worldwide health statements from World Health Organization association and teamwork with several other evidence-based platforms also happen to a single platform to cooperate as a unified system. However, this result's high quality as it entirely depends on machine learning. Its efficiency and consistency are often more focused on developing worldwide partnerships. Transmitting significant information which always accelerates innovation and promoting optimistic interventions. It has been observed that with the spread of the virus, the susceptible percentage of the community will gradually fall to zero, and also from a significant period, where the number of infected cases tends to increase. However, it starts decreasing as people started to recover from the disease.

References

1. Dong E, Du H, Gardner L (2020) Lancet Infect Dis 20:533–534
2. Dimensions COVID-19 publications, data sets, clinical trails. Figshare https://dimensions.figshare.com/articles/Dimensions_COVID-19_publications_datasets_and_clinical_trails/11961063 (2020)
3. Wu Z, McGoogaan JM (2020) JAMA 323:1239–1242
4. Claassen J et al (2019) N Engl J Med 380:2497–2505
5. Sitt JD et al (2014) Brain 137:2258–2270
6. Peiffer-Smadja N et al (2019). Clin Microbiol Ingect. https://doi.org/10.1016/j.cmi.2019.09.009
7. Belhadi D et al (2020) Preprint at https://doi.org/10.1101/2020.03.18.20038190
8. Liu X, Wang X (2020) J Genet Genom 47:119–121
9. Computational predictions of protein structures associated with COVID-19. Deepmind https://deepmind.com/research/open-source/computational-predictions-of-protein-structuresassoicated-with-COVID-19 (2020)
10. Peiffer-Smadja N et al (2020). Clin Microbiol Infect. https://doi.org/10.1016/j.cmi.2020.02.006
11. Stokes JM et al (2020) Cell 180:688–702e13
12. Weiskopf D et al (2020) Preprint at https://doi.org/10.1101/2020.04.11.20062349
13. Senior AW et al (2020) Nature 577:706–710
14. Gautret P et al (2020) Int J Antimicrob Agents. https://doi.org/10.1016/j.ijantimicag.2020.105949
15. Ai T et al (2020) Radiology. https://doi.org/10.1148/radio1.2020200642
16. Chen Z et al (2020) Eur J Radiol 126:108972
17. Pham HH, Le TT, Tran DQ, Ngo DT, Nguyen HQ (2019) Preprint at https://arxiv.org/abs/1911.06475
18. Zheng C et al (2020) Preprint at https://doi.org/10.1101/2020.03.12.20027185
19. Chen T, Kornblith S, Norouzi M, Hinton G (2020) Preprint at https://arxiv.org/abs/2002.05709
20. COVID-19 Clinical Research Coalition Lancet 395:1322–1325 (2020)
21. Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak. Wellcome Trust https://ac.uk/coronavirus-covid-19/open-data (2020)
22. Open-access data and computational resources to address COVID-19. National Institutes of Health https://datascience.nih.gov/covid-19-open-access-resources (2020)

Role of Internet of Things and Machine Learning in Finding the Optimal Path for an Autonomous Mobile Robot

Dadi Ramesh, Mohmmad Sallauddin, Syed Nawaz Pasha, and G. Sunil

Abstract Path planning for the mobile robot is an emerging area in today's world; the development of autonomous vehicles like driverless cars and mobile robots has tremendously enhanced researchers to work more on path planning strategies using emerging technologies, like the Internet of things and machine learning. These technologies will provide optimal solutions than classical problem-solving algorithms. The article deals with path planning for mobile robots in an unknown environment using deep learning and the Internet of things. These technologies are adopted to work in different strategies like environmental prediction, object detection/ obstacle detection, and finding a path. For this, an innovative model is proposed to detect static and dynamic obstacles from the input data and finding a path from one point to another point.

Keywords Internet of things (IoT) · Machine learning (ML) · Deep learning (DL)

1 Introduction

Path planning is a very crucial task for autonomous mobile robots. The robots always try to find a path from source to destination by overcoming the obstacles and finding an optimal route. This mobile robot should concentrate on predicting the environment and constraints to reach the goal. Today, path planning for driverless cars and autonomous robots is a challenging task. Researchers are working on different types of algorithms to find the optimal path. All the path planning algorithms use traditional approaches like a greedy approach and heuristic approaches to some other algorithms. These are working well, but computational time is more and more space to find a path, and there are not using artificial intelligence techniques [1, 2]. In

D. Ramesh (✉)
Center for Artificial Intelligence and Deep Learning, Computer Science and Engineering, S R Engineering College, Warangal, India
e-mail: dadiramesh44@gmail.com

M. Sallauddin · S. N. Pasha · G. Sunil
Computer Science and Engineering, S R Engineering College, Warangal, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_29

this article, the proposed system tends to analyze the role of IoT and deep learning techniques to find the best path.

This paper is organized in the following manner; the related works are explained in Sect. 2. Section 3 discusses the proposed system and role of deep learning and IoT in path planning. Results and analysis are depicted in Sect. 4. Finally, Sect. 5 describes the conclusion of the proposed method.

2 Related Work

The path planning algorithms that will use different technologies to find a path which can classify into three categories, namely the traditional approach, cognitive-based, and artificial intelligence-based methods (Table 1).

2.1 Blend Search

It is an uninformed search where it does not have an idea about adjacent nodes blindly; it searches for goal node. And it does not provide the optimal path. It searches for all possibilities to find the goal node. Such algorithms are breadth-first search [3] and depth-first search [4] which blindly search for goal node without knowing the environment. The computational time is also high, and some no guarantee to find the goal node.

2.2 Grid-Based Method

It follows the concept of state-space search also called graph-based search, representing the environment in an $n * n$ grids and starting cells and goal state in that. It starts searching from starting cell to goal cell. The robot can move cell by cell obstacles are also represented in a cell. It is also an uninformed search, and it searches for the goal in all directions to find the optimal path. This method follows the algorithm which solves the graph and does not try to learn the patterns. And the computational

Table 1 Autonomous mobile robot path planning algorithms

Traditional approach	Cognitive-based methods	AI-based methods
Blend search	Cognitive-based adaptive path planning algorithm (CBAPPA)	D* and A* algorithm
Grid-based method		Ant colony optimization
Divide and conquer method		BEE colony optimization and gray wolf optimization

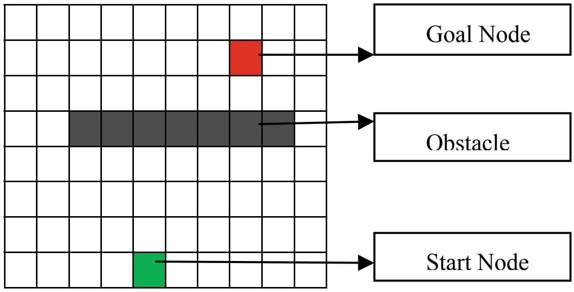


Fig. 1 Environment representation

time is more in finding a path. Figure 1 illustrates the environmental representation of a mobile robot.

2.3 Divide and Conquer Method

The divide and conquer method is a popular method to solve the complex problems. The main aim of the divide and conquer [5] way is dividing the problem into a small number of parts and solving the pieces one by one. The same concept was used in mobile robot path planning. It divides the environment into small parts called the local area. First, it finds the path for in local area; later, it merges all local paths to find the final path.

2.4 Cbappa

Cognitive-based adaptive path planning algorithm [6, 7] is an algorithm that finds a path from source to destination based on cognitive methods like how a human being perceives the target when he/she does not know the environment. It will choose one direction (goal direction) [8] and starts moving to reach the destination by overcoming the obstacles. It does not search for all possible paths to find an optimal path. It also searches more environments to find the path. But it works in both environments like known and unknown environments.

2.5 D and A* Algorithms*

A* [9] and D* [10] algorithms are heuristic search algorithms; A* works for reducing the length of the path by considering all possible paths. And D* practices on reducing the cost of the path from source to destination. A* works on heuristic function $h(x)$. Both are finding optimal paths but searching for more areas to find the final path.

2.6 Ant Colony Optimization

Marco Dorigo proposes this concept. ACO [11] is used to solving computational problems like robot path planning and network path planning. ACO system works on chemical essence, which is released by ants. This phenomenon works how an ant colony finds the best path from their original place to the goal state. The first ant colony starts moving in all directions by releasing pheromones until reaching the goal state. Then, it will decide the optimal path; in the second iteration, ant colony will move in the best path. So ACO's first ants are searching for more areas to find the best path. Almost ants search 60–70% other areas from the actual path to find the optimal path.

2.7 Bee Colony Optimization

The bee colony optimization [12] algorithm proposed by D. Karaboga is based on how the bees find an optimal path from source to destination. The bee colony algorithm works the local search concept first. It searches in the local area later; it increases the local area step by step until reaching the goal state.

2.8 Gray Wolf Optimization

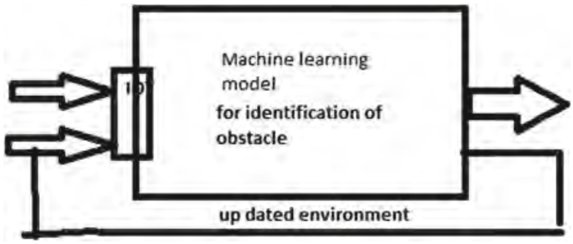
Gray wolf optimization [13] algorithm is based on natural human life how a human being will solve the problems in real time. It works based on a chain pattern. But these two algorithms are finding an optimal path by searching more areas. BCO and GWO algorithms are exploring almost 60% more area than the actual path.

The traditional algorithms can find the path from source to destination, but the complexity is high and not optimal. The AI-based algorithms are finding an optimal way, but the searching area is high. And the CBAPPA algorithm finds the optimal path in some cases. All paths finding algorithms' efficiency are almost less than 50% though they are finding an optimal route. An algorithm's ability is calculated as the area searched for the final path and length of the closing route.

3 Proposed Approach

An IoT and machine learning-based system are proposed for finding a path from source to destination. The selection is based on one path, which is toward the goal point; this path is the primary path. A refined path is identified from a primary path

Fig. 2 System architecture



by overcoming the obstacle and updating the environment. Figure 2 illustrates the autonomous agent architecture with sensors and effectors.

3.1 Obstacle Size Prediction

The input and output modules are controlled by the Arduino microcontroller [14, 15]. It will take the input in a video format and make them into frames. These frames are given to deep neural networks for finding the output path. When the model detects an obstacle, it tries to predict the obstacle size with the help of the prediction algorithm. The algorithm takes the input of three parameters such that the parameter denotes the angle from the local path, the second parameter denotes the distance traveled from a local path, and the third parameter denotes the required angle to reach the local path. In path finding systems, the obstacle size prediction [16] is an important part that will reduce the searching area by predicting obstacle size early. Figure 3 illustrates the methodology of obstacle prediction with the side angle method.

Here, in Fig. 3:
b—is obstacle,
A—is an angle made by a robot corresponding to the primary path.
c—is the distance traveled by a robot from the primary path,
Predicting obstacle size from the image:
$a = \mathrm{BC}$, $b = \mathrm{AC}$, $c = \mathrm{AB}$.
$C = 180 - (\mathrm{A} + \mathrm{B})$

$$b = \frac{asinB}{\sin A}$$

$$\frac{a}{sinA} = \frac{c}{sinC}$$

$$a = \frac{csinA}{sinC}$$

$$\Delta ABC = \frac{1}{2}\mathrm{ac}\sin B$$

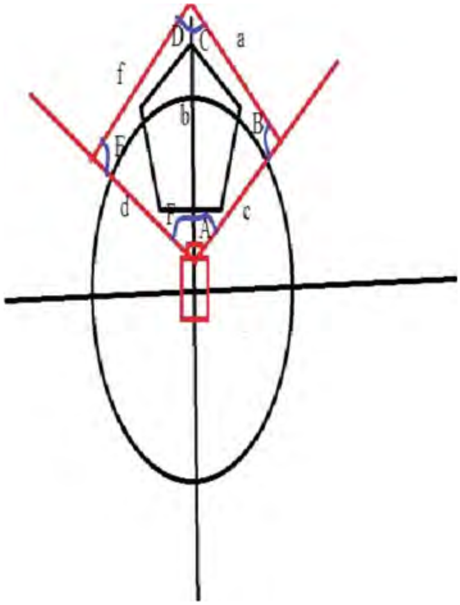


Fig. 3 Obstacle size prediction

From Fig. 3 angles $B \equiv E$.
And $b = \mathrm{DF}, f = \mathrm{DE}, d = \mathrm{EF}$.

$$f = \frac{b \sin F}{\sin E}$$

$$\Delta DEF = \frac{1}{2} \mathrm{bf} \sin D$$

Obstacle size $= \Delta ABC + \Delta DEF$.

3.2 Deep Neural Network Model

The OpenCV and TensorFlow libraries are utilized for implementing. With OpenCV, the video is converted into several frames [17, 18]; after that, each frame is converted to RGB or grayscale image type. Then, it will pass to the input layer.

3.2.1 Training

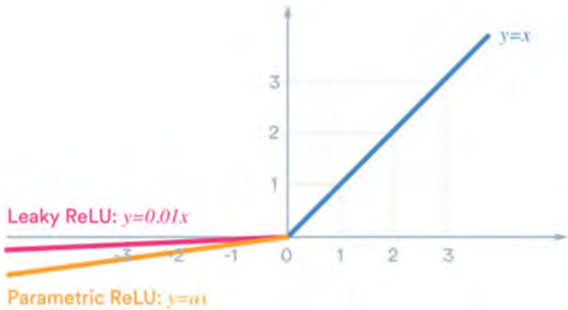
The AlexNet pre-trained model is used to train the images with five convolution layers [19] with max-pooling and zero padding. And the activation function is Relu.

Based on the robot moving, the images will be passed to the model to select the right direction. If the robot moves on a primary path, then the model will find the only obstacle on the primary path until it finds any impediment on the primary path. If it finds any barriers on the primary path, it selects the images from the left and right to move and change the direction.

The trained AlexNet [20] for 12 epochs with dropout layers at each time, the nodes in the convolution layer will be dropped to train the model effectively. The AlexNet model is trained on the dataset, obstacles, free, left direction, and right direction images. Whenever it finds the obstacle first, it calculates the obstacle size and then moves either left or right direction. Whenever it finds the obstacle, it will choose the right or left direction concerning the primary path. And it does not consider the other path.

The test is not executed directly in the robot to find a path from source to destination. Figures 4 and 5 illustrates Relu activation function and CNN architecture corresponding.

Fig. 4 Relu activation function



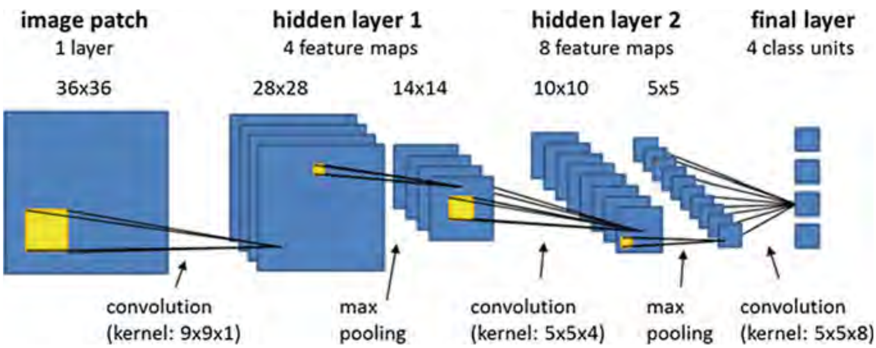


Fig. 5 Sample CNN model

4 Result Analysis

The proposed method is concentrated only on reducing the searching area for finding the final path. The model is trained in a way that whenever the ambiguity occurred, choose to move left or right, the robot will move based on the size of the obstacle. With this, the robot's efficiency will be increased, and the searching area will be reduced by up to 20%.

$$\text{searched area} = \frac{\text{no.of cells searched}}{\text{total no.of cells}} \times 100$$

$$\text{efficiency} = \frac{\text{length of the finalpath}}{\text{number cells searched}} \times 100$$

5 Conclusion

The proposed method presents an approach with Internet of things and a deep learning-based model. It works on unknown environments and finds an optimal path from source to destination, by reducing the searching area. The size of the obstacles is predicted by using trigonometric concepts. When the obstacle size is observed, then the robot will feel like it is moving on known environment, and it is easy to reduce the searching area. It just searches 20% of the new area than the final path.

References

1. Al SNP et al (2019) Variation analysis of artificial intelligence, machine learning and advantages of deep architectures. Int J Adv Sci Technol 28(17):488–489
2. Yue P et al (2019) Experimental research on deep reinforcement learning in autonomous navigation of mobile robot. In: 2019 14th IEEE conference on industrial electronics and applications (ICIEA), IEEE, pp 1612–1616. DOI.org (Crossref). https://doi.org/10.1109/ICIEA.2019.8833968
3. Hansen EA, Zhou R (2007) Anytime heuristic search. J Artif Intell Res (JAIR) 28:267–297
4. Al SM et al (2019) A comprehensive study on traditional Ai and Ann architecture. Int J Adv Sci Technol 28(17):479–487
5. Stentz A (1996) Optimal and efficient path planning for partially-known environments. In: Proceedings IEEE international conference on robotics and automation, pp 3310–3317
6. Ramesh D, Pasha SN, Sallauddin MD (2018) Cognitive based adaptive path planning for mobile robot in dynamic environment. In: Artificial intelligence and cognitive computing
7. Pasha SN, Ramesh D, Roopa G A novel approach to path planning of robots by electing dynamic obstacles
8. Manoharan S, Ponraj N (2019) Precision improvement and delay reduction in surgical tele robotics. J Artif Intell 1(01):28–36

9. Goldberg AV, Harrelson C (2005) Computing the shortest path: A* search meets graphtheory. In: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics, pp 156–165
10. LaValle SM (2006) Planning algorithms. Cambridge University Press, Cambridge, UK; Stentz A (1995) The focussed D* algorithm for real-time replanning. Int J Rob Autom
11. Gambardella LM, Dorigo M (1997) Ant colony system: a cooperative learning approach to the traveling salesman problem. IEEE Trans Evol Comput
12. Teodorović D (2009) Bee Colony optimization (BCO). In: Lim CP, Jain LC, Dehuri S (eds) Innovations in swarm intelligence. Studies in computational intelligence, vol 248. Springer, Berlin, Heidelberg
13. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. J Adv Eng Softw 69:46–61
14. Li Q, Chen L, Li M, Shaw SL, Nüchter A (2013) A sensor-fusion drivable-region and lanedetection system for autonomous vehicle navigation in challenging road scenarios. IEEE Trans Veh Technol 63(2):540–555
15. Smys S, Ranganathan G (2019) Robot assisted sensing, control and manufacture in automobile industry. J ISMAC 1(03):180–187
16. Adapting Best Path for Mobile Robot By Predicting Obstacle Size (2019). Int J Innov Technol Explor Eng 8(9S2):200–202. DOI.org (Crossref). Blue Eyes Intelligence Engineering & Sciences Publication. https://doi.org/10.35940/ijitee.I1039.0789S219.
17. Fujiyoshi H et al (2019) Deep learning-based image recognition for autonomous driving. IATSS Res 43(4):244–252. DOI.org (Crossref). https://doi.org/10.1016/j.iatssr.2019.11.008
18. Harshavardhan A et al (2020) 3D surface measurement through easy-snap phase shift fringe projtion. Springerprofessional.De. https://www.springerprofessional.de/en/3dsurfacemeasurement-through-easy-snap-phase-shift-fringe-proj/15447362. Accessed 26 Mar 2020
19. Levine S, Finn C, Darrell T et al (2016) End-to-end training of deep visuomotor policies. J Mach Learn Res 17(39):1–40
20. Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39(6):1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.

Overview of Blockchain Technology: Applications and Use Cases

J. S. Shyam Mohan, Vedantham Hanumath Sreeman, Vanam Venkata Chakradhar, Harsha Surya Abhishek Kota, Nagendra Panini Challa, M. U. M. Subramanyam, and Surekuchi Satya Swaroop

Abstract Blockchain technology has revolutionized agriculture, finance, education, supply chain, health care, and many sectors. Blockchain has extended its advantages to the non-financial sector too. Researchers in this area are constantly trying to explore many applications of blockchain that can provide fruitful results. This article discusses various use cases and real-time applications of blockchain technology in various sectors and subsequently, explores the state-of-the-art developments in blockchain use cases in various sectors and the scope for further research aspects in the future. It is intended for audience those who are interested to learn use cases and real-time applications of blockchain technology.

Keywords Distributed ledger technology (DLT) · Federated Byzantine Agreement (FBA) · Monetary Authority of Singapore (MAS)

J. S. Shyam Mohan · V. H. Sreeman · V. V. Chakradhar (✉) · H. S. A. Kota ·
M. U. M. Subramanyam · S. S. Swaroop
Department of CSE, SCSVMV, Kanchipuram, India
e-mail: vanamvenkatachakradhar@gmail.com

J. S. Shyam Mohan
e-mail: jsshyammohan@kanchiuniv.ac.in

V. H. Sreeman
e-mail: hanumathsreeman@gmail.com

H. S. A. Kota
e-mail: khsabhishek1335@gmail.com

M. U. M. Subramanyam
e-mail: mani12061999@gmail.com

S. S. Swaroop
e-mail: satyaswaroop058@gmail.com

N. P. Challa
Department of IT, SVECW, Kovvada, India
e-mail: paninichalla123@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_30

1 Introduction

With the tremendous improvements in blockchain technology and with the increasing utilization of bitcoins and continuous improvement in blockchain technology, various academic and financial sectors are continuously exploring the practical applications of blockchains. Today, the majority of the financial, sales, and clinical fields' blockchain has shown its impact. In this paper, the authors focus on the applications and use cases of blockchain technology in various sectors. Transactions in blockchain are handled in the distributed ledgers using tokens. Apart from using tokens, Ethereum and Hyperledger Fabric are also used for making digital transactions [1, 2]. Blockchain has been commercially adopted, influenced the world's currency market, and facilitated the illegal dark web. Blockchain additionally has been a critical factor influencing the expansion of monetarily determined digital assaults, for example, ransomware and denial of service against retailers and other online associations. The usage of blockchain has brought the world's first decentralized cryptocurrency. Mostly, this is because of hype, powered by the rising and dropping estimation of bitcoin. Blockchain applications play a vital role in 5G and 6G technologies. The impacts of distributed technology on business systems give off an impression of being like dis-intermediation and digital intervention impacts in e-commerce (Laudon and Traver 2018). The blockchain applications may have an assortment of consequences for business systems and business relations, remembering impacts for trust and consequences for the system structure.

Research in distinctive business regions and use cases where blockchain technology leads to the transformation of a new era. The implementation of blockchain technology in various fields will be helpful for advancement and a trustworthy environment. Many applications in blockchain are based on distributed ledger technology (DLT). Blockchain can increment money-related productivity by diminishing manual control and tampering of the existing data. In intercompany trades, blockchain will make one type of the record allowing intercompany straightforwardness and settlement at a comparative second. Blockchain is an open record that keeps up records of the impressive number of trades held tight a blockchain mastermind while working in a scattered manner. This blockchain coordinator is a conveyed framework that does not waste time with any central capacity to confirm or settle the trades in the crucial framework and thusly empty go-betweens and bring straightforwardness and improved security. Blockchain technology has been adopted for use in many government sectors. Because of digitalization, character the board has consistently involved worry for all open just as private associations. Blockchain applications and use cases by industry are shown in Table 1.

This paper is organized into the following sections: Sect. 2 describes the blockchain technology and use cases for payments in financial services, and Sect. 3 records the compliance and mortgage. Section 4 represents the blockchain for Global Trade Logistics. Section 5 outlines the blockchain in Healthcare, and Sect. 6 explains the blockchain in energy markets. Section 7 depicts the blockchain in government. Finally, Sect. 8 concludes the research work.

Table 1 Blockchain applications and use cases by industry

Financial services	Insurance	Retail	Supply chain and logistics	Public sector
Trade finance	Claims processing	Supply chain	Supply chain and finance	Asset registration
Cross-currency payments	Risk provenance	Loyalty programs	Maintenance tracking	Citizen identity
Mortgages	Asset usage history	Information sharing	Provenance	Medical records
KYC	Claims file		Supply chain compliance	Medicine supply chain
Cross-border taxes				

2 Blockchain Use Cases—Payments and Securities Trading in Financial Services

Blockchain technology has been applied to banking and money-related administrations in different manners and getting various advantages. Keen agreement administration helps in leading budgetary exchanges without a go-between. It can oversee protections, deeds, settlements, and cases in a robotized way. With the advanced development in the financial area, blockchain innovation models have just been executed in worldwide installments which are profiting banks as far as diminishing expenses and to abbreviate handling times. The economy purely depends on the circulation of the respective country currency, and the overview of financial services that inherited blockchain technology is as follows [3].

2.1 Cross-Border Payments—Stellar Open Network

A decentralized, hybrid blockchain platform that makes it possible to create, send, and trade digital settlements on a single network with Lumens as a native asset is based on Federated Byzantine Agreement (FBA) that takes approximately 2–5 s for each transaction clearance. It has anchors that act as bridges between a given currency and a Stellar network that consists of a distributed exchange, viz., pay in EUR with INR balance, and the network will automatically convert it at the lowest available rate.

Figure 1 shows the Stellar performance that it could be the best solution and might this blockchain technology can replace the interbank transfer method which was vulnerable in some cases, and with the cryptocurrency security, these types of frauds could be nullified. The Stellar network stock fluctuations represent the currency circulations and depend on the particular country currency value at the international

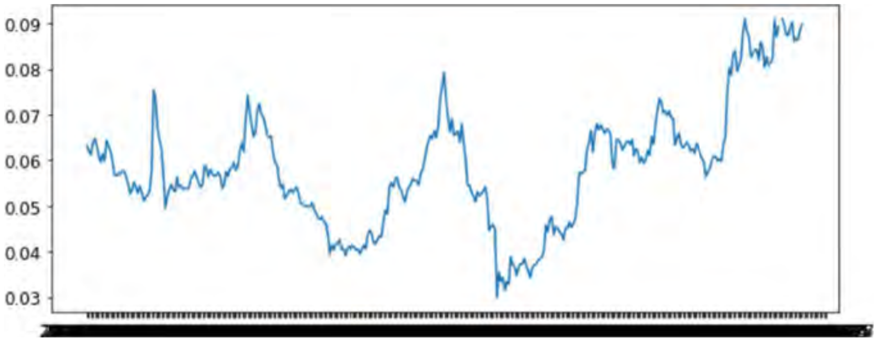


Fig. 1 Graph of Stellar EUR performance

level and the conversions made by the users, and these all factors involved in the performance of the Stellar.

2.2 Permissioned Network for Payments

Maintains Nostro ("Ours") and Vostro ("Your") accounts and transactions. Only member banks are permitted to transact, ensuring privacy and confidentiality of transactions, and eliminate the need for reconciliation and errors that happen, reducing costs and delays in clearing and settlement (like RTGS). The treasury has an instantaneous view of the currency position of their Nostro accounts across the globe, allowing for optimal use of capital. It provides reduced foreign exchange and capital exposure, lower fees, and increased compliance and security [4].

2.2.1 Project Ubin: Central Bank Digital Money Using Distributed Ledger Technology

A synergistic venture with the business to explore the utilization of blockchain and distributed ledger technology (DLT) for clearing and settlement of installments and protections. Project Ubin aims to assess the ramifications of having a tokenized type of the Singapore Dollar (SGD) on a distributed ledger (DL) and its expected advantages to Singapore's monetary biological system. Project Ubin was considered as an open door for Singapore to play the main job in the exploration of national bank money on a DL and Central Bank Digital Currencies (CBDCs). Monetary Authority of Singapore (MAS) is Singapore's central bank. MAS goes about as a settlement specialist, administrator, and manager of an installment, clearing, and settlement frameworks in Singapore that attention on wellbeing and effectiveness. MAS embraces this job as a confided in an outsider (trusted third party) and effectively connects with banks in Singapore, just as with the open and private sector, for example, the Singapore Clearing House Association (SCHA) and the Association of Banks in Singapore (ABS). Since MAS has become a more trusting third party, this introduced an incredible open door for MAS to work together with the banks and survey the worthiest that blockchain could bring to this current relationship [5]. Some of the phases of Project Ubin are Tokenized are SGD, Re-imagining RTGS, Delivery versus Payment (DvP), Cross-border Payment versus Payment (PvP), Enabling broad ecosystem collaboration [6].

2.3 Ripple Protocol and Network—Financial Services

Ripple protocol is used for banks to clear and settle payments in real time through a distributed network without a centralized clearance house. It takes an average of 5 s for confirmation. Gateway nodes convert fiat currencies to XRP (currency in Ripple) (Fig. 2).

Interbank payment depends on the daily fluctuations in the currency of a country in international trades and makes a linked path which confuses the process and the centralized system monitoring the transactions of the investors/organizations. The security of the transactions is not compromised, but the personal information and transaction details of an individual investor are recorded, and there might be a chance of losing investors due to monitoring their records. The traditional method does have some time-taking process because every transaction made by the person or investor

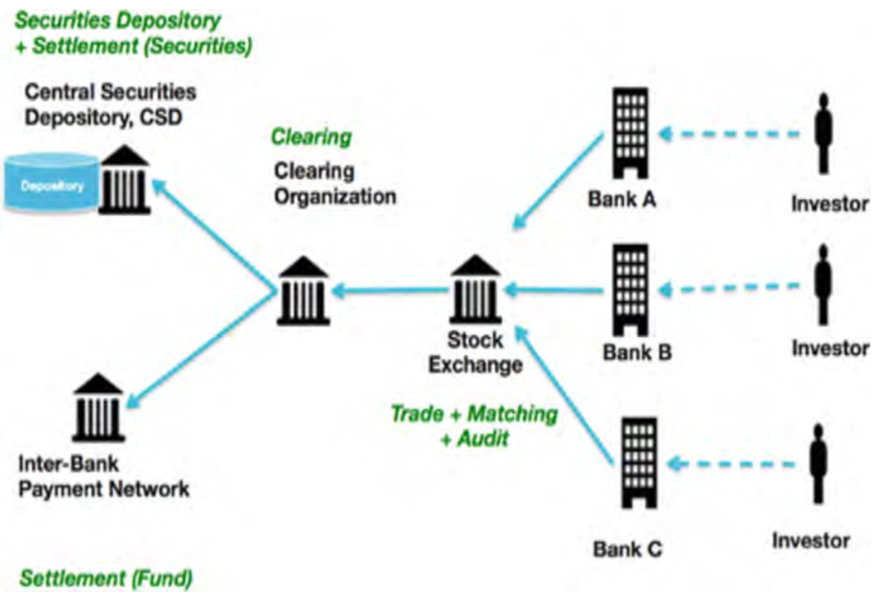


Fig. 2 Securities settlement—today's process [7]

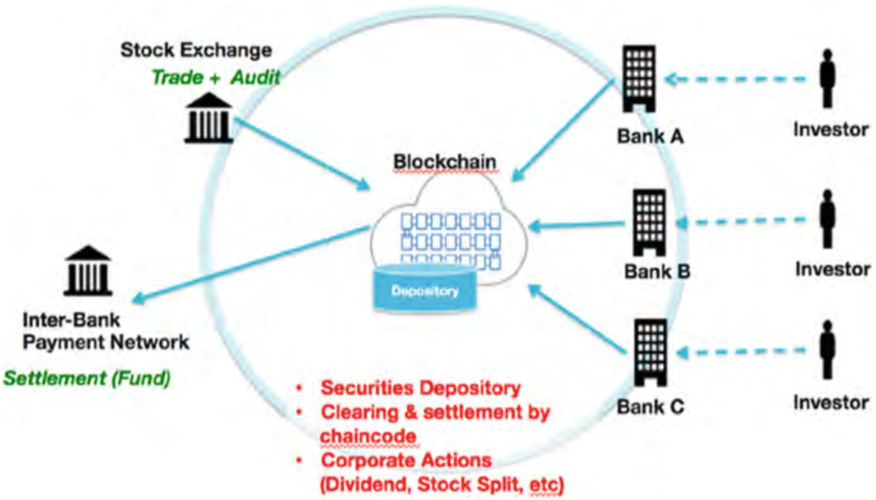


Fig. 3 Securities settlement—blockchain process

is centralized and because many people involved brags the transaction time more [7] (Fig. 3).

The blockchain process provides a more secure way of digital transactions with all the cryptographic records, and every investor's transaction details are secured in between them, and no one is there to monitor the transactions. The inter-border fund transactions would be easier and secure than the complex traditional way, for example, Stellar can precede the transaction conversions according to the currency trades between the persons involved in the transaction. This method also provides a method of unified currency options like bitcoins which then converted into their corresponding country currency. Blockchain is a type of distributed ledger, distributed ledger method, which provides the cryptographic documents/transactions between the two persons or investors or dealers. There is no central administrator in a distributed ledger, mainly focused on digital transactions, records, and security. Distributed ledgers are like databases that are synchronized, duplicated, and shared; there are no others to change and record the data between the shared persons [7].

3 Compliance and Mortgage

A Mortgage is a kind of agreement that tends if the client neglects to reimburse the credit sum with interest taken from the associations, and then, the associations have the right to take the property referenced as an affirmation in the advancement procedure. Blockchain can be utilized to make a computerized ID for every property, in this way making the property trackable on the system. Aside from making the land showcase more fluid, simply from a home loan application point of view, this computerized ID would incorporate a chain of possession and a current market valuation that will permit banks to rapidly check the current proprietorship status or affirm the market cost, conceivably relieving the need of experiencing title deeds and drawing in with assessors.

3.1 Compliance (KYC and AML)

Know Your Customer is a process of verifying the due diligence of a client. Sharing KYCs on a blockchain would provide financial institutions to deliver better compliance outcomes and improves customer experience [8, 9]. KYC compliance is of the following steps they are gathering of information, verification and due diligence, initial risk assessment, and continuous monitoring (Fig. 4).

Blockchain innovation takes into account the making of a distributed ledger that is then mutual with all clients on the system. Ethereum is a blockchain stage that utilizes a smart contract for handling every single exchange. This factor implies that nobody has single power, and consequently, a state of shortcoming, as in the client/server model, and also the information are immutable. This implies blockchain databases have an inbuilt changelessly that makes the information that they contain unquestionably more dependable. Such databases can be utilized to store ID subtleties of people which would be dependable. On the off chance that the money-related administration area, for instance, executes blockchain for KYC confirmation, they will have the option to check clients rapidly and dependably by using an application, and so on. Because of the unwavering quality of blockchain databases, government establishments and organizations could depend on the information totally, something which would evacuate the requirement for any further ID checks [10].

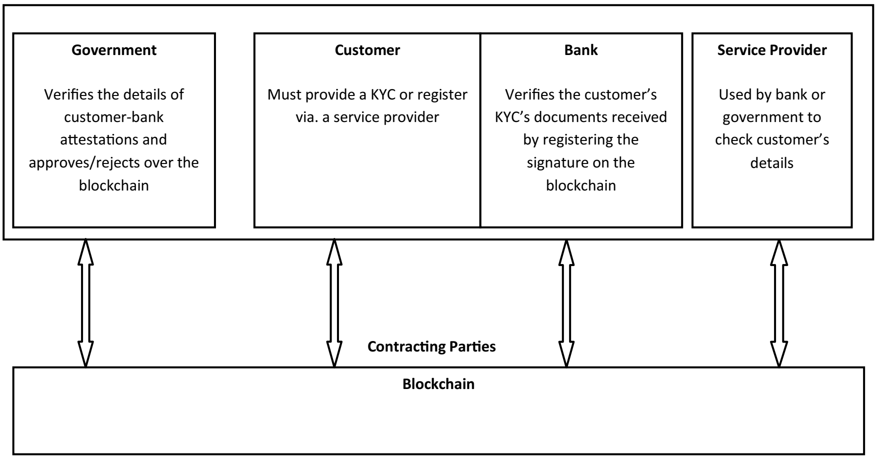


Fig. 4 Compliance (KYC and AML)

3.2 Trade Finance Network

The platform is built on the IBM Blockchain Platform using Hyperledger Fabric that provides easy access to banking services to customers. It takes into consideration adaptability that takes into consideration fast universal extension as a business, administrative, and security opportunities converge [11, 12].

4 Blockchain for Global Trade Logistics

To work internationally, the shipments are required to proceed onward time. Blockchain conveys the continuous following of the considerable number of members in the flexible fasten to see the exactness, adjust verification data, and information. DLT helps in making the best stage for the administration of every single part of delivery coordinations. Since the data is immutable, there is no possibility of data tampering. Data in a blockchain cannot be hacked or forged and is promptly "trusted" and along these lines, acknowledged by anybody with access to your chain [13].

4.1 IBM Blockchain for Trade Logistics

IBM Blockchain for trade logistics provides an effective way of transporting goods from one location to another location [14]. Some of the services provided by IBM Blockchain are container logistics, food supply, procurement, responsible sourcing, counterfeit prevention, and supply chain visibility.

4.2 Tradelens

Tradelens is powered by IBM Cloud and IBM Blockchain. It gives each substance associated with a worldwide exchange with the computerized instruments to share data and team up safely.

5 Blockchain: Healthcare

Distributed ledger technology is used to store the information in an immutable format and updates the data progressively which has been reshaping the human services area entirely. The traditional models in this scene end up being exceptionally wasteful as far as conveying quality human services which are moderate in nature of the people. Blockchain technology-based human services applications are fit to be utilized and change the social insurance establishments over the world Blockchain technology works for the improvement as far as straightforward and proficiency; different gatherings are related to the medicinal services framework, and patients get profited [15].

5.1 GuardTime

Guardtime HSX crosses over any barrier between patients, suppliers, payers, controllers, and pharma via flawlessly shipping information over various human services partners, conveying secure utilization of a solitary, honest form of wellbeing information [16].

5.2 Loyyal

With Loyyal's Blockchain-as-a-Service (BaaS), customers have comprehensive access to undertake grade facilitating administrations, advancement apparatuses, bolster administrations, and ever-developing system of accomplices. Comprehensive month-to-month permit expenses can begin as low as $5 K every month suitable for particular needs. The following are some of the features of loyal, namely unlimited API access to the loyal platform, entry to loyal's network of earning and redemption partners, personalized node dashboard, monthly support services, and unlimited support for severity level 1–2 issues.

6 Blockchain in Energy Markets: Gridchain

PONTON has developed imaginative pilot programming dependent on blockchain innovation that simulates future forms for constant lattice the executives, called Gridchain. The next stage will be that for all intents and purposes, test this procedure in the field "with a choice of market members, i.e., an alliance of the willing." Also, Gridchain is a commitment to the European institutionalization of between process correspondences when planning keen lattices of things to come [17].

6.1 Bloomberg New Energy Finance (BNEF)

Bloomberg New Energy Finance (BNEF) produces investigation on ventures on the move, concentrating on clean vitality, propelled transport, advanced industry, creative materials and, wares. BNEF bits of knowledge help corporate methodology; money and strategy experts separate the truth from the publicity, explore change, and produce openings. The BNEF controls the accessibility, portability, and the terminal by 250 examiners in 18 areas around the world. Figure 5 shows detailed information about the BNEF [18].

The network edge, where the person connects, and the distribution grid balance the supply and demand with no central controller, and Bloomberg New Energy Finance helps to create opportunity very useful while including in the corporate strategy, finance, and policies. Bloomberg is always interested in fields like clean energy, digital industry, advanced transportation, and innovative products. For example, the bulk generation in Fig. 5 had solar panels, a windmill plant registry with fraud-proof, the carbon tax for the transport, and energy sources collected directly from appliances, and all transmission networks resolve all the payments peer to peer.

7 Blockchain in Government

Blockchain [19] helps the government in the following ways such as access and verification of central data, sharing of data, and sharing of data and access control.

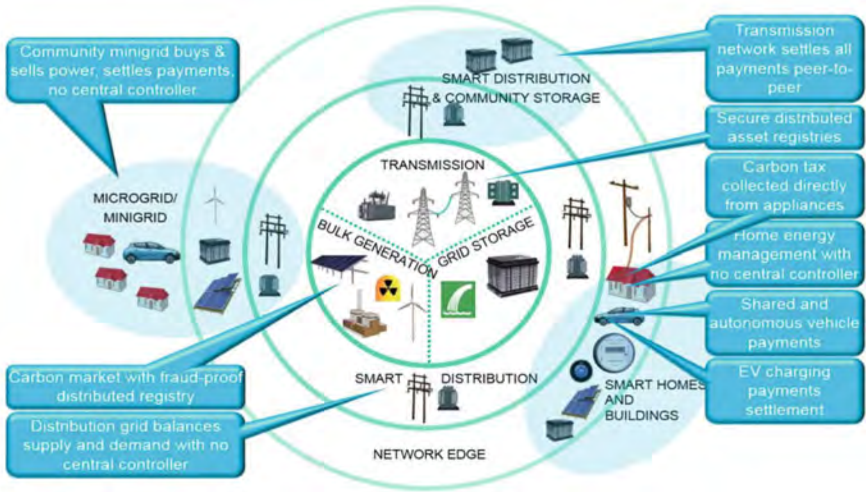


Fig. 5 Bloomberg new energy finance

7.1 *Blockchain Use Cases in Government—Worldwide*

7.1.1 Russia

The first government-level blockchain implementation is completed officially in Russia. The state-run bank Sberbank announced today that it is partnering with Russia's Federal Antimonopoly Service (FAS) to actualize record move and capacity via blockchain.

7.1.2 South Korea

South Korea's leading blockchain project is a network designed to interconnect independent chains and their communities; in essence, ICON is a blockchain of blockchains. Consensus models are continuously developing and one of the most Delegated Proof-of-Contribution (DPoC) which takes delegated proofs of stake above and beyond.

7.1.3 Singapore

The government has initiated a Project Ubin that is built on the distributed ledger for clearing and settlement of installments and protections. DBS Bank has cooperated with the Singapore government for the blockchain trade platform.

7.1.4 India

IndiaChain is the first blockchain initiative proposed by NITI Aayog in India. It is used to maintain India's public records. This has collaborated with UPI for transactions. Andhra Pradesh has entered into a partnership with Chromaway to use blockchain technology to maintain land registry records [20].

7.1.5 USA

Federal agencies in the USA are trying to assess and receive appropriate record advances that utilizes encryption and coding to improve straightforwardness, proficiency, and trust in data sharing like financial management, trademarks, and many more services.

Table 2 Blockchain applications and use case for other sectors

S. No.	Blockchain application	Description
1	Hyperledger Indy	Distributed ledger platform for decentralized identity management
2	Hyperledger Indy—Plenum	Distributed ledger platform that uses Redundant Byzantine Fault Tolerant algorithm for consensus
3	SecureKey	SecureKey is a driving personality and confirmation supplier that rearranges shopper access to online administrations and applications
4	Sovrin	Personally manage individual IDs online
5	IBM Blockchain	Helps to create, operate, and maintain permitted decentralized identity networks built using Hyperledger Indy DLT

7.1.6 Estonia

e-Estonia is like a digital Id card and a decentralized distributed system [21]. Some of the advantages of e-Estonia are file taxes within 5 min, sign a contract electronically, register a business within 30 min, I-voting, and registering as a citizen (Table 2).

8 Conclusion

An overview of blockchain technology applications and use cases in various sectors and the study of Stellar's data shows a better understanding of blockchain technology because some countries like India banned blockchain technology in some fields. Blockchain is a kind of distributed ledger, which gives the cryptographic documents/transactions between the two people or speculators or dealers. Many applications in blockchain depend on distributed ledger technology (DLT). With the serious advancement in the budgetary territory, blockchain development models have quite recently been executed in overall portions which are benefitting banks similar to reducing costs and to abbreviate handling times. However, this article provides insights into modern-day applications of blockchain technology in real time. The interested audience can refer to the references and web links for more details.

References

1. Aste T, Tasca P, Di Matteo T (2017) Blockchain technologies: the foreseeable impact on society and industry. Computer 50(9):18–28
2. Zheng Z, Xie S, Dai H, Chen X, Wang H (2017) An overview of blockchain technology: architecture, consensus, and future trends. In: 2017 IEEE international congress on big data (BigData Congress), p 557564

3. Laroiya C, Saxena D, Komalavalli C (2020) Applications of blockchain technology. In: Handbook of research on blockchain technology, pp 213–243
4. Nakamoto S (2008) Bitcoin: a peer-to-peer electronic cash system. www.Bitcoin.Org. https://bitcoin.org/bitcoin.pdf [Online]. Available
5. Project Ubin—https://www.mas.gov.sg/-/media/MAS/ProjectUbin/Project-Ubin--SGD-on-Distributed-Ledger.pdf
6. Project Ubin—https://www.mas.gov.sg/schemes-and-initiatives/Project-Ubin
7. Blockchain charts—https://www.blockchain.com/en/charts
8. JPX—https://www.jpx.co.jp/english/
9. Blockchain Applications in Supply chain—https://www.ibm.com/blockchain/industries/supply-chain
10. Sinha P, Kaul A (2018) Decentralized KYC system. Int Res J Eng Technol (IRJET)
11. Loyal—https://loyyal.com/
12. Skuchain—https://www.skuchain.com/about/
13. David MD et al (2018) Blockchain for and in logistics: what to adopt and where to start. Multidisciplinary Digital Publishing
14. Kan L, Wei Y, Hafiz Muhammad A, Siyuan W, Linchao G, Kai H (2018) A multiple blockchains architecture on inter-blockchain communication. In: 2018 IEEE international conference on software quality, reliability and security companion (QRS-C), pp 139145
15. Zeadally S, Siddiqui F, Baig Z, Ibrahim A (2019) Smart healthcare: challenges and potential solutions using internet of things (IoT) and big data analytics. PSU A Rev J 1–17
16. Guard time—https://guardtime.com/health
17. Gridchain—https://enerchain.ponton.de/index.php/16-gridchain-blockchain-based-process-integration-for-the-smart-grids-of-the-future
18. Bloomberg report—https://www.bloomberg.com/impact/products/bloombergnef
19. Blockchain for government—https://digital.gov/communities/blockchain/
20. Indiachain—https://www.businessinsider.in/what-is-indiachain-a-blockchain-system-that-could-soon-be-the-heart-of-governance-in-india/articleshow/64676670.cms
21. E-estonia—https://e-estonia.com/

Tuberculosis Detection from CXR: An Approach Using Transfer Learning with Various CNN Architectures

P. Anu Priya and E. R. Vimina

Abstract Tuberculosis (TB), a major public health threat, is preventable and curable if identified at its earlier stage. Advances in artificial intelligence lead convolution neural network (CNN) to focus on TB elimination by using popular diagnostic tools like chest X-rays. However, the limited availability of publicly accessible chest X-ray datasets remains a challenge that can be resolved using the transfer learning technique. In this work, to detect TB, transfer learning is used with different CNN architectures such as VGG-19, RestNet50, DenseNet121, and InceptionV3 on Montgomery, Shenzhen, and combined dataset. For performance evaluation, the area under the ROC curve (AUC) and accuracy (ACC) along with the confusion matrix is considered. The results show that the VGG-19 model achieved the highest AUC score of 0.89, 0.95, and 0.95 for Montgomery, Shenzhen, and combined datasets, respectively.

Keywords Tuberculosis · Deep learning · Convolutional neural network · Artificial intelligence · Transfer learning

1 Introduction

Artificial intelligence (AI) is becoming the future of mankind, especially in the field of medical science. The standardization of diagnostic measures in disease identification to be specific is always a developing and promising ground for technological advancements like AI. But in the world of evidence-based medicine, all the possibilities need to be tested to its core. At the same time, potentials of this technology should be focused on the eradication of major public health threats like tuberculosis.

Tuberculosis (TB) has always been a persistent public health threat caused by Mycobacterium tuberculosis. Being an airborne disease, TB is transmitted from person to person through cough, sneeze, and spits of infected individuals. The

P. Anu Priya (✉) · E. R. Vimina
Amrita School of Arts and Sciences, Amrita Viswa Vidyapeetham, Kochi Campus, Kochi, India
e-mail: priyaanu29@gmail.com

E. R. Vimina
e-mail: vimina.er@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_31

disease is said to be under control in developed countries, but the global condition stays perilous as the developing and underdeveloping countries are still struggling to manage the malady. This statement can be indicated by World Health Organization (WHO) global report statistics which is getting updated every year that says the disease affects nearly 10 million of the world population and kills an estimate of 1.6 million annually [1]. The major contributor for the global burden for TB in India as the country's caseload when added up will make up to at least one-third of the global TB cases. Figure 1 shows the year-wise TB incidents in India from 2010 to 2018. In 2016, 28 lakh Indian citizens became ill with mycobacterium tuberculin and 4.5 lakh death due to the disease occurred [2]. WHO Report from 2019 revealed that about 2.6 million cases occurred in the country, and 0.45 million death occurred due to the disease.

A major cause of high mortality and spread is due to the persistent gap occurring in the detection of this disease. Therefore, it is a known fact that more than one-third of the estimated 10 million cases are not undergone timely diagnosis or reporting. The widely used two diagnosing measures for this disease are Ziehl–Neelsen (ZN) staining test and chest X-ray. Though CXR is having low specificity, it is still a popular measure used for pulmonary TB detection by resource-limited settings of primary-level clinicians and mass screening programs worldwide. Chest X-ray is said to make about 37% over diagnosis when no other diagnostic measures are used along with it to confirm the disease occurrence [3–5]. These can be caused due to factors like unavailability of experts to read the X-ray film, compromised quality of the machine or film used, the gender of the patients, etc. Inter-reader and intra-reader variations in CXR reading are also a notable factor when it calls out for the need of involving artificial intelligence in the disease identification [6, 7]. The current lack

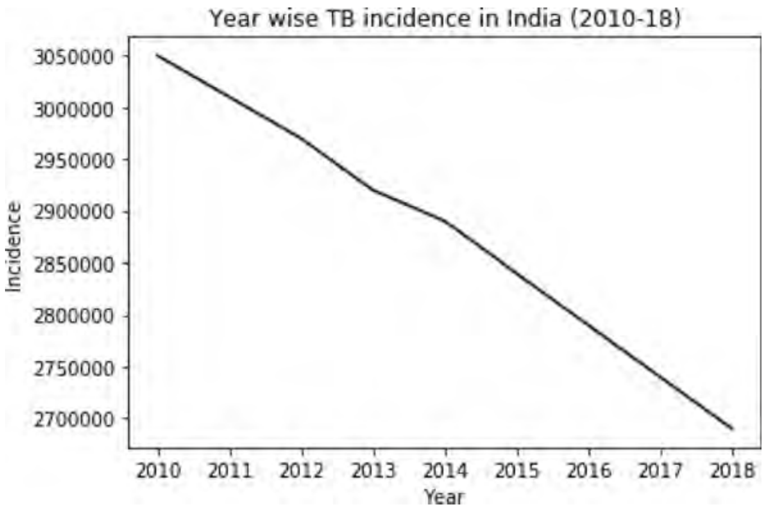


Fig. 1 Year-wise TB incidents in India

of evidence is preventing authority bodies like the World Health Organization from developing recommendations regarding the use of automated reading systems for TB detection [8].

Though deep learning was proving its ground in several studies for the detection of cancer cells and diabetic retinopathy, lack of studies based on TB detection from CXR using deep learning systems made a demand for its involvement in recent years. The researchers were looking forward to the expectation that there would be less inter-reader variability, provision of radiological services in places where experts were unavailable, and reproducibility of valid results. Now, deep neural networks provide new possibilities for TB detection from chest radiographs. Developments in neural networks to classify images, sound, and text with algorithms that use nodes and layers to learn like that of human cognitive functioning have always contributed to this advancement. This is achieved by recurrent training sessions using different datasets [9, 10].

CNN is a technique commonly used in deep learning that arranges nodes in tiles to build a visual reception area. The layers for training extract significant visual definitions of images from data, while sets of metrics to be prepared using fully connected layers classify the characteristics to target groups, e.g., TB or normal [11]. As CNN is programmed to identify and extract the most discriminatory features [12, 13]; based on the target objectives from the data itself, it eliminates the requirement for manual feature inputs that rely on domain-specific knowledge. It is practically difficult to provide enough data to train the deep networks due to the lack of publicly available datasets. This issue can be overcome to a level using transfer learning which uses pre-trained CNN models on large datasets; for example, ImageNet has 1.2 million of 1000 class images. In this technique, pre-trained CNN models use initial weights that are prepared with a large-scale dataset to train CNN models later, even with a small database [14].

The objective of this study is to evaluate the efficiency of advanced pre-trained CNN in TB detection from chest X-ray. To accomplish this, X-ray images from publicly available datasets are processed and used to train and test CNNs. Since TB-related samples are small, the favored technique for training deep CNN is transfer learning. Therefore, the transfer learning approach with four different CNN architectures was used to detect the presence of TB and later to compare the results.

2 Methods

2.1 Dataset

The publicly accessible Montgomery dataset and Shenzhen dataset of NIH Tuberculosis Chest X-ray dataset is used. The National Library of Medicine developed Montgomery dataset by associating with the Health and Human Services Department,

Table 1 Dataset summary

	Montgomery	Shenzhen	Combined
Subjects	138	662	800
Without TB	80	326	406
With TB	58	336	394
Male	63	442	505
Female	74	213	287
Other/unknown	1	7	8

Montgomery County, Maryland, USA, and created Shenzhen dataset in collaboration with the Shenzhen No. 3 Hospital of China [15, 16]. Both sets contain normal and abnormal X-rays, the latter containing tuberculosis manifestations. The dataset of Montgomery and dataset of Shenzhen have 138 and 662 subjects, respectively, with the presence and absence of TB. Table 1 provides information regarding the datasets. For taking advantage of the discrepancies in processing between various sets of data, a combination of both datasets has been created. The model is trained on these datasets to acquire robust functionality.

In CXR images, tuberculosis is seen in the form of infiltrates, cavitations, blunted costophrenic angles, consolidations, pleural effusion, opacities, pneumonia, horizontal fissure displacement, widely spread nodules, and in many other radiological forms [17]. Figure 2 shows CXRs of subjects chosen from the dataset. Figure 2 a exhibits CXRs of subjects uninfected with TB and b, c, d shows CXRs of patients infected with TB with different manifestations.

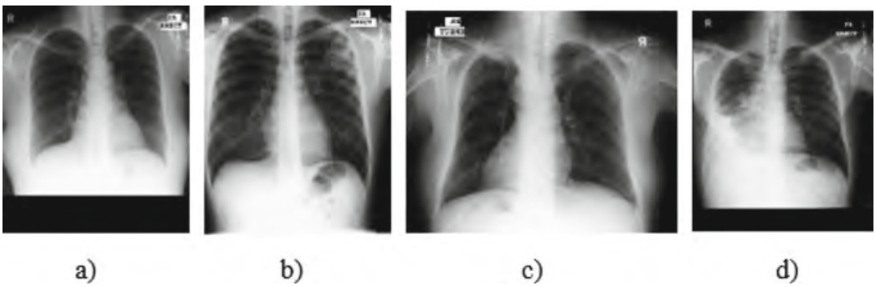


Fig. 2 **a** Normal CXR, **b** cavitary infiltrates, **c** infiltrates, **d** pleural effusion

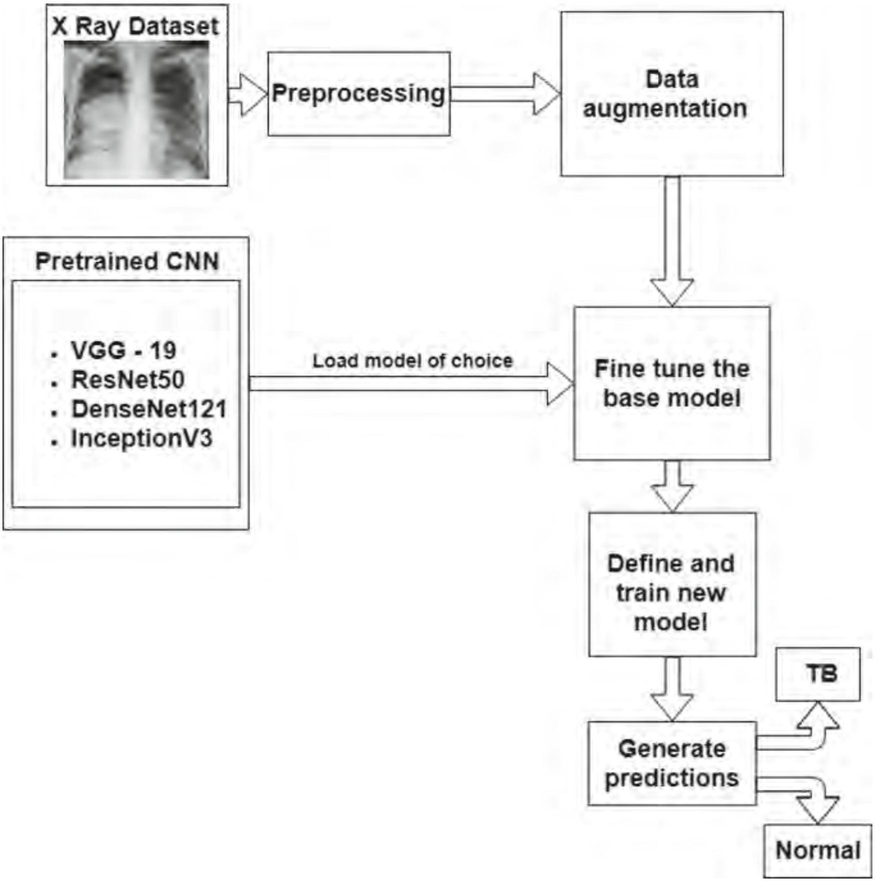


Fig. 3 Conceptual diagram

2.2 *Methodology*

The overall procedure is visualized through a conceptual diagram in Fig. 3. The following steps are included: loading CXR images, preprocessing of CXR, augmentation transformation on training and validation images, transfer learning with VGG-19, DenseNet121, InceptionV3 or ResNet50 architechures, fine-tuning of the base model, defining and training new model, and generate predictions.

2.2.1 Data Preprocessing and Augmentation

All images used in the datasets for this study are frontal thoracic chest X-rays and include regions beyond the lungs that do not apply to TB detection. Preprocessing is performed to diminish the features that are irrelevant to TB detection that distorts the final results. For this purpose, few of the available options were used to initialize data generators for training and validation datasets. In training, data generator transformations are performed on images to produce new images by applying rescaling on images and using the fill mode parameter to fill in new pixels after applying transformations with the nearest surrounding pixel values. All the models pre-trained were very large to carry the dataset of this size, which could easily overfit the model. To prevent this, data augmentation is performed. An iterator is created to train the network model by specifying the target size of the image as 128 * 128, batch size as 64, shuffle to a true value, and class mode set as categorical. No other preprocessing techniques have been carried out.

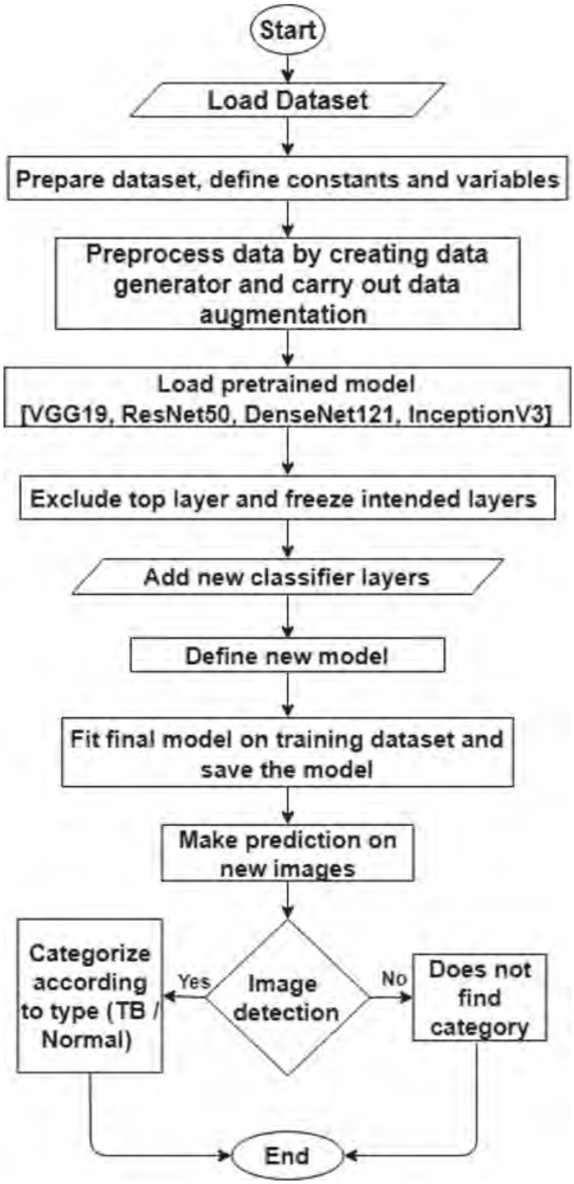
2.2.2 Transfer Learning with CNN

In several computer vision tasks including classification and segmentation, CNN has demonstrated its promising performance. But, since deep learning is a data-driven process, learning the rules involves a great deal of data. If there are not enough training data, it is difficult for deep learning algorithms to optimize the parameters of the prediction models. Working on the related issue of computer vision, the best approach is to involve transfer learning and to use pre-trained ones rather than the training of models from scratch. Many research institutions release models for large challenging datasets, and these pre-trained models can be used for extracting features from the new network. Because of the limited volume of data, this automated TB diagnosis problem embraces transfer learning as the solution and strives to explore, assess, and evaluate the impact of different CNN architectures.

There are two widely used methods to embrace pre-trained CNN's capabilities. The first technique does feature extraction where the extracted features are introduced into a new network to perform classification. This approach is widely used to maintain the valuable feature extractors trained during the initial stage. The second technique adds modifications to the pre-trained model that may include improvements to the architecture and parameter tuning. The basic information derived from the previous task is preserved, and new trainable parameters are introduced into the network. This is the technique opted for in the present study.

Various models like VGG-19 [18], DenseNet121 [19], ResNet50 [20], and InceptionV3 [21], which is already trained on ImageNet dataset, are utilized. After instantiating the model, pre-trained weights were loaded automatically, and images are resized to 128 * 128 with three channels without including top layers. The weights of the initial layers are made frozen to retrain only the higher layers to fine-tune the model. After several experiments, the parameters were established. Choices of parameters are possible and can be explored in the future for improving the results. After flattening the layer, a dense layer is added with arguments of 1024 neurons and the rectified linear unit (ReLU) [22] as an activation function. To avoid overfitting, dropout [23] of probability 0.4 is introduced. One more dense layer is added with 256 neurons with a dropout probability of 0.2. Subsequently, the dense layer is enabled by

Fig. 4 Flowchart



the softmax activation function to classify the image input as normal or infected. The learning rate of the model was set to 0.001. SGD [24] as the optimization method is used to compile CNN, and training was carried out using a categorical cross-entropy loss function of batch size 64. Early stopping was implemented to monitor the test loss at each epoch and to interrupt the training once the model improvement is stopped. The workflow is given, and a detailed process is included in Fig. 4 as a flowchart.

Procedure followed in flowchart:

1. Examine and prepare a dataset
2. Preprocess data by adding data augmentation using ImageDataGenerator class
3. Compose new model by loading the pre-trained model and add new classifier layers on top.
4. Train the model
5. Fit the final model on the training dataset and save it.
6. Generate predictions using the saved model: normal or tuberculosis.

2.2.3 Metrics

Specific metrics were recorded for CNN classification task which are as follows: (a) correctly TB cases recognized (TP, true positives), (b) incorrectly recognized TB cases (FN, false negatives), (c) correctly recognized no TB cases (TN, true negatives), and (d) incorrectly recognized no TB cases (FP, false positives). Based on these metrics, accuracy, sensitivity, and specificity of the model are calculated.

$$\text{Acurracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \tag{1}$$

$$\text{Sensitivity} = \text{TP}/(\text{TP} + \text{FN}) \tag{2}$$

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FP}) \tag{3}$$

3 Results

The primary aim of using the transfer learning technique was to diagnose tuberculosis precisely in CXR images. For this, prepare and train all the models separately. In the Montgomery dataset, VGG-19 was trained, and it achieved ACC of 89% and AUC value was 88%. ResNet50 performs better than DenseNet121 and InceptionV3 in terms of accuracy and AUC. It achieved an AUC of 77% and an accuracy of 80%. As these measures are highly dependent on the number of samples representing each class, their subjective evaluation leads to incorrect conclusions. Because of this reason, the criterion for selecting the best model has to be the combination of sensitivity and specificity. Table 2 shows the achieved results for each CNN in terms of accuracy, sensitivity, and specificity. The best results were attained by the VGG-19 network.

In the Shenzhen dataset also, VGG-19 performed better than all other networks by achieving the best accuracy of 95% and AUC of 95% which is presented in Table

Table 2 Results of CNN with transfer learning-Montgomery dataset

Network used	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
VGG-19	89	98	77	88
ResNet50	80	91	65	78
DenseNet121	80	98	55	77
InceptionV3	70	83	51	67

Table 3 Results of CNN with transfer learning-Shenzhen dataset

Network used	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
VGG-19	95	96	93	95
ResNet50	85	79	91	85
DenseNet121	75	91	60	75
InceptionV3	73	93	43	74

3. The results attained in the Shenzhen dataset outperform the results achieved by the Montgomery dataset, probably due to the limited sample size and class imbalance of the samples. The Montgomery dataset has 60% samples negative with 40% positive, and in Shenzhen dataset, the sample class balance is almost 50% which adds as a favorable aspect.

VGG-19 outperformed in the combined dataset with the best results, followed by ResNet50 shown in Table 4. The accuracy and AUC achieved by VGG-19 are almost similar to Shenzhen dataset but vary in the ResNet50 model.

The confusion matrix of VGG-19 shown in Table 5 depicts true positives, true negatives, false positives, and false negatives of the model which helps to compare the best models further. The best outcomes are those with the lowest FN. A real-life perception of a false negative case will lead to the absurd conclusion that the patient is normal, which gives opportunities for public transmission of bacteria. Table 5 shows the low value of FN, especially for VGG-19, which is a good result. In terms of specificity also, VGG19 outperforms ResNet50 and other models and thus proves to be the most powerful model for the particular classification task in all the three datasets.

To visualize the performance of the model, the receiver operation characteristic (ROC) curve is plotted, as it is the most significant assessment metrics for testing

Table 4 Results of CNN with transfer learning-combined dataset

Network used	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
VGG-19	95	98	91	94
ResNet50	87	95	79	87
DenseNet121	74	97	50	74
InceptionV3	76	80	71	75

Table 5 Confusion matrix of VGG-19 and ResNet50

CNN	Dataset	TP	FP	FN	TN
VGG-19	Montgomery	79	13	1	45
	Shenzhen	316	21	10	315
	Combined	400	34	6	360
ResNet50	Montgomery	73	20	7	38
	Shenzhen	259	30	67	306
	Combined	386	82	20	312

the efficiency of any classification model. The ROC curve plotting true positive rate (sensitivity) against false positive rate ($1 - \text{specificity}$) of the VGG-19 model is presented in Fig. 5 for all the three datasets. It is a proven fact that the higher the AUC, the better the model is to differentiate between patients with TB and normal. Classifiers give curves closer to the top-left corner, indicating that a better accuracy and AUC were achieved.

4 Discussion

The findings of other papers that use the same dataset as ours were compared [25]. Proposed a CNN network to reduce computational as well as memory requirement and achieved 79%, 84.4%, 86.2% in terms of accuracy without pretraining the model [26] obtained 90.3% and 0.96 in accuracy and AUC with transfer learning after training a set of CXRs around 10,848 [27] proposed a model with 19 layers and showed an accuracy of 94.73% with validation accuracy of 82.09% using Adam optimizer. Though in other papers full data is not provided for assessment, the proposed model performed better in terms of accuracy even after training the model on a smaller size database (Table 6).

To achieve more validity to the findings in a clinical-based environment, it is ideal to train the system with more images. This work has faced the limitation of sufficient publicly accessible datasets, and the results obtained could be improved by using larger datasets.

5 Conclusion

This study aims to propose a deep learning approach with the transfer learning technique for classifying and identifying TB from CXR images. To extract features from X-rays, the pre-trained architectures VGG-19, RestNet50, DenseNet121, and InceptionV3 are used and trained on the ImageNet dataset. The approach of deep learning can be considered for bidirectional system validation in a clinical setting

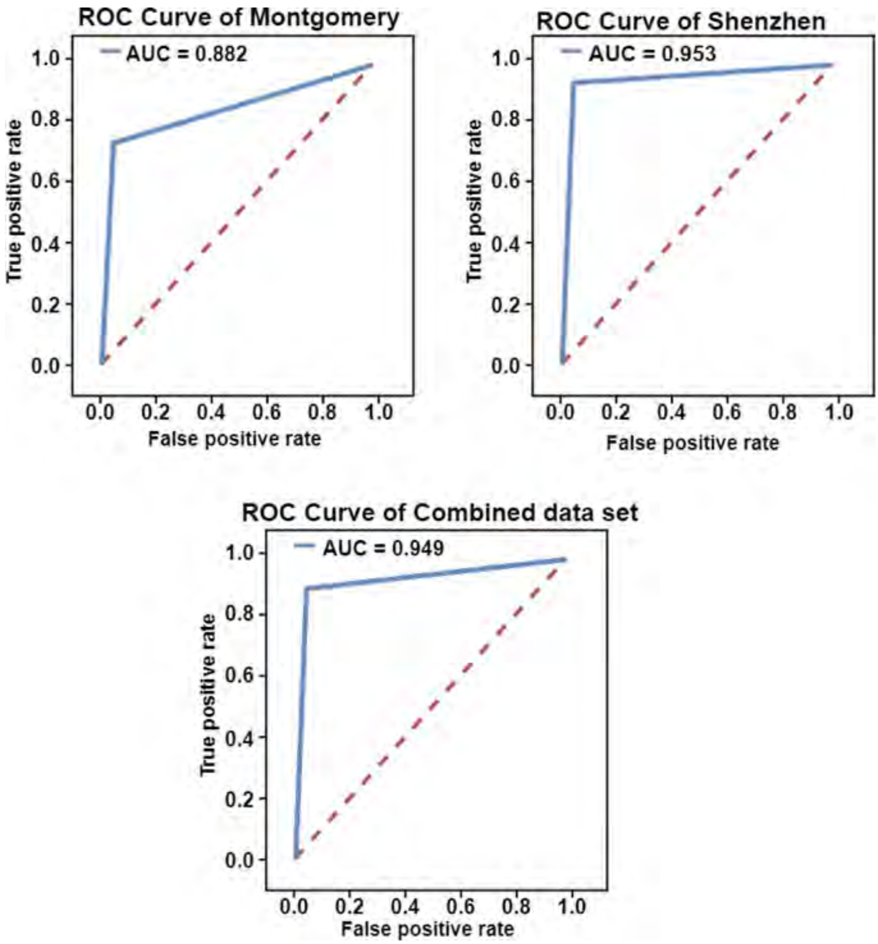


Fig. 5 ROC curves of VGG-19 on three different datasets. The AUC score is **a** 0.88 for dataset Montgomery, **b** 0.95 for dataset Shenzhen, **c** 0.94 for combined dataset

Table 6 Comparative results

Model	AUC (%)	Test accuracy (%)
Pasa [25]	92	86
Hwang [26]	96	90
Rahul Hooda [27]	–	94
Proposed approach	94	95

to detect TB by which both human and system error can be drastically minimized. This could be helpful for early diagnosis of TB, and to achieve this, the system must be trained with larger datasets. Thus it is recommended that high burden developing countries like India use this situation to share their chest X-ray dataset to the public, which may help to facilitate the development of much better tools.

References

1. Tuberculosis (2020) World Health Organization. https://www.who.int/news-room/fact-sheets/detail/tuberculosis
2. Global TB Report 2017. https://www.who.int/tb/features_archive/global_tb_report_2017/en/
3. Global Tuberculosis Report (2018) World Health Organization. https://apps.who.int/iris/bitream/hadle/10665/274453/9789241565646eng.pdf
4. Shah S (2009) Intensified tuberculosis case finding among HIV-infected persons from a voluntary counseling and testing center in Addis Ababa, Ethiopia. J Acquired Immune Def Syndromes 537–545
5. Bakari M (2008) Basis for treatment of tuberculosis among HIV-infected patients in Tanzania: the role of chest X-ray and sputum culture.BMC 8
6. Harries AD, Maher D, Graham S (2004) TB/HIV: a clinical manual. WHO
7. Ndugga K, Klatser PR (2005) Sex-specific performance of routine TB DiagnosticTests. Int J Tuberc Lung Disease 294–300
8. WHO (2016) Chest radiography in tuberculosis detection. WHO
9. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press
10. Krizhevsky A, Hinton GE, Sutskever I (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems
11. Rouse M (2020) Techtarget. https://searchenterpriseai.techtarget.com
12. Bindhu V (2019) Biomedical image analysis using semantic segmentation. J Innov Image Process 91–101
13. Manoharan S (2019) Smart image processing algorithm for text recognition, information extraction and vocalization for the visually challenged. J Innov Image Process 31–38
14. Huang Z, Pan Z (2017) Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data. MDPI 9
15. Jaeger J (2014) Automatic tuberculosis screening using chest radiographs. IEEE Trans Med Imaging 33(2):233–245
16. Jaeger J (2013) Automatic screening for tuberculosis in chest radio graphs: a survey. Quant Imaging Med Surg
17. Weinberger J (2013) Principles of pulmonary medicine. Elsevier Health Sciences
18. Bansal S (2018) Kaggle. https://www.kaggle.com/shivamb/cnn-architectures-vgg-resnet-inception-tl
19. Keras. https://keras.io/api/applications/densenet
20. Dwivedi P (2019) Towards data science. https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33
21. Keras. https://keras.io/api/applications/inceptionv3
22. Brownlee J (2020) A gentle introduction to the rectified linear unit. https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks
23. Brownlee J (2019) A gentle introduction to dropout for regularizing deep neural networks. https://machinelearningmastery.com/dropout-for-regularizing-deep-neural-networks
24. Keras. https://keras.io/api/optimizers/sgd
25. Pasa F, Golkov V, Pfeiffer F (2019) Efficient deep network architectures for fast chest X-Ray tuberculosis screening and visualization. Nature
26. Hwang S, Kim HE, Jeong J (2016) A novel approach for tuberculosis screening based on deep convolutional neural networks. In: SPIE medical imaging
27. Hooda R, Sofat S, Kaur S (2017) Deep-learning: a potential method for tuberculosis detection using chest radiography. IEEE

E-commerce Logistic Route Optimization Deciphered Through Meta-Heuristic Algorithms by Solving TSP

M. Soumya Krishnan and E. R. Vimina

Abstract E-commerce business is now becoming more popular and has become a critical need of the day for every customer. When considering the revenue and expense factors, each e-commerce business spent their major outflow in its logistics activities. Logistics costs are the focused area to be optimized so that the overall business can be dealt with in a controlled and safe manner. To with the logistics part, let us first consider the problem of vehicle routing. The Vehicle Routing Problem (VRP) is an intricate situation where multiple constraints can be taken into consideration. Thus, before dealing with such a complex situation, originally to move on with the study, the simple Traveling Salesman Problem is considered here in the paper. To solve such a hard NP problem and to get an optimal solution, many alternate methods have been tried out by many. In this paper, some of the meta-heuristic algorithms are considered and each of their performance is compared depending upon the length of the route, number of cities, time of execution of each algorithm, and their error rates. These algorithms are implemented and verified using Python 3.8.

Keywords Traveling salesman problem (TSP) · Logistic optimization · Meta-heuristics · Swarm intelligence · E-commerce · Artificial fish swarm algorithm (AFSA)

1 Introduction

Logistics costs are classified into many types. In this study, the outbound logistics cost is taken into consideration by taking into view the transportation activities. The workflow of the study can be depicted as shown in Fig. 1. The flow graph depicts the correct flow of work to follow based on the research point of view. The main concentration of the original research study is to optimize the logistic cost for an

M. Soumya Krishnan (✉) · E. R. Vimina
Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Kochi, India
e-mail: soumyamahesh15@gmail.com

E. R. Vimina
e-mail: vimina.er@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_32

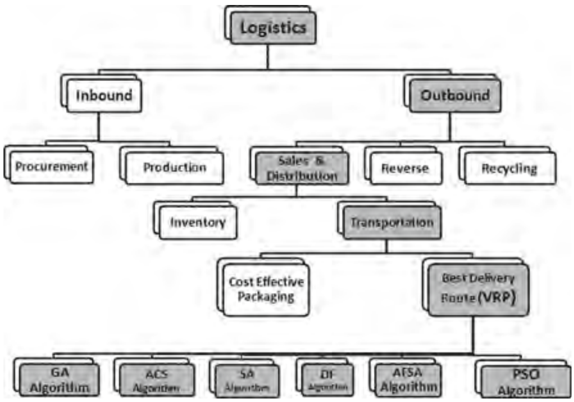


Fig. 1 Workflow of the study

e-commerce company whose major concentration is on outbound logistics. Even though other areas are involved in this, our study mainly focuses on the sales and distribution part. This is again managed by the transportation unit. When transportation is considered as the focus area of study, it was noticed from various literature reviews and other sources of study that VRP plays an imperative role in scheming logistics. Thus, VRP is taken as the main area of study based on which one can try to optimize the cost of logistics.

Some of the common objectives of a VRP can be projected out as follows minimize the total transportation cost based on the overall distance traveled as well as the costs allied with the drivers and the vehicles used, minimize the count of the vehicles required to serve all clients, calculate the travel time and capacity (vehicle load), and restrain penalties for poor quality service.

Also, VRP is a complex system where various variants as well as specializations are to be considered during the study. Some of the commonly considered VRP variants can be portrayed as VRP with pickup and delivery, with LIFO, with time windows, capacitated VRP, with multiple trips as well as open VRP. Along with these, multiple constraints need to be considered in solving each variant. Some of the possible constraints are shown in Fig. 2. Solving VRP by considered these multiple constraints is a complex process. But, one of the proven methodologies is by using various meta-heuristic algorithms. Thus, to start with the study, instead of the complex VRP, simple TSP is taken into consideration, using which, optimization of transportation cost can be done. When there's only one constraint such as a single-vehicle traveling through multiple cities is considered for the study, the VRP gets reduced to the TSP. Based on this concept; the study is performed using various meta-heuristics algorithms to get the most optimized result.

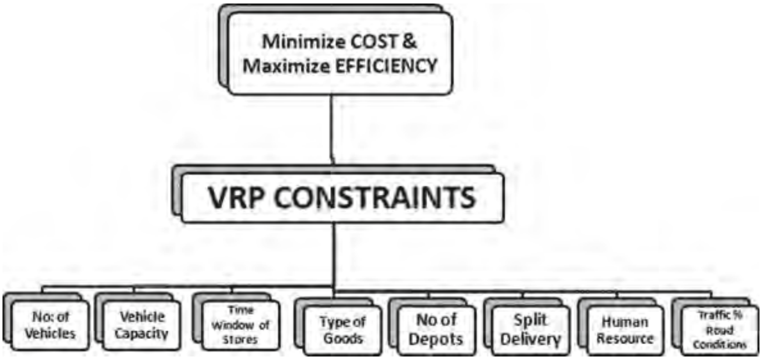


Fig. 2 Factors affecting VRP

1.1 Traveling Salesman Problem

Traveling Salesman Problem (TSP) is intended to explore the best possible roadmap for a traveler or a salesman where he tries to visit a set of cities based on the condition that every city should be touched exactly once, except the city where the travel was initiated and that must be the last city to visit. Such methodology is implemented in areas like logistics and transportation [1]. TSP is always well thought-out as an NP-hard problem in combinatorial optimization. Solving such problems using conservative methods is complex. This process of computations is liable to consume enormous time. Thus, try to compromise for approximated results which comparatively consumes less time, ending up with not an optimal solution but near to it.

Mathematically, TSP is described as follows:

$$\min T = \sum_{\mathrm{i}=1}^{n-1} d(\mathbf{x}_{\mathrm{i}}, \mathbf{x}_{\mathrm{i}+1}) \tag{1}$$

Here, X_i denotes the ith town or city in which i can be assigned values from 1 to n. Also, T is the total of the distance covered in the entire trip represented as $d(X_i, X_{i+1})$, in which X_i and X_{i+1} is the length between the town i and the next immediate $i + 1$th town [2].

The basic steps to solve a TSP problem can be illustrated as follows:

1. Generate a list of city coordinates
2. Initialize a function object using a coordinate list
3. Create a list of distances between pairs of cities
4. Initialize a function object using distance list
5. Define an object for the optimization problem.
6. Select and execute any chosen randomized optimization algorithm
7. Get the output as optimal route and minimal distance covered.

The steps mentioned above can be modified using different meta-heuristic algorithms.

TSP can be solved by various meta-heuristics approaches in a much efficient and faster method than any other proved customary methods by sacrificing aspects like optimality, speed, accuracy as well as precision [3]. Heuristics approaches have classifications like specific heuristics as well as meta-heuristics [1, 4]. Those heuristics which are used to solve specific problems are known as specific heuristics, whereas meta-heuristics are a certain class of algorithms that are used commonly to resolve almost any type of problems related to optimization [5]. One such popular meta-heuristic algorithm is swarm intelligence algorithms (SI), where a huge number of individual agents interact together and are used to portray the collective behavior of the system [6]. Some of the popular and widely used SI algorithms include the Genetic Algorithm, Artificial Bee Colony Algorithm, Ant Colony Optimization Algorithm, Differential Evolution Algorithm, Particle Swarm Optimization Algorithm, Cuckoo Search Algorithm, and so on.

The paper starts with an epigrammatic explanation on six SI-based algorithms as mentioned. Following that, experimentation is tried out to measure the performance of the prescribed algorithms based on the efficiency check and the time of execution. The results are conferred comprehensively along with the statistical analysis in the section followed. Based on the outcome, one best-performing algorithm is selected against the other five algorithms opted. The conclusion segment is projected at the end of this paper.

2 Related Work

In every e-commerce activity, it is found that the organization's main expense is in its logistics. Thus, considering such a factor into consideration, in this paper, the problem is primarily considered as a TSP issue and the study is carried out with minimum constraints (as mentioned in the introduction part). In this paper, the efficiency of various swarm intelligence algorithms is compared on the basis of factors like the length of the trip and the total time taken for the same. Here, the datasets are generated randomly in which details like city coordinates are given in a pre-specified range from 10 to 200 or more. The datasets are retrieved from TSPLIB [7].

TSPLIB is a library of TSP sample instances where a lot of such samples are found in which varying sizes of dataset related to city counts and corresponding locations are found.

Here, similar problems from various sources are of different types. Each dataset in the library is categorized based on the number of cities. The attributes defined in each one are the type of data, dimension, tour section, and their corresponding (x, y) coordinates indicating the city positions. For our study purpose, the dataset opted are having a dimension of 10, 51, 100, and 200 cities.

In this experiment, the performance of optimization techniques selected is assessed on some factors like the shortest length of the route explored and the execution time of the algorithm. The experiment was implemented using the Python 3.8 version on a Sony VAIO Laptop, Core i5 processor, and operating system of version Windows 10. The algorithm was set to execute an average of around 30 iterations. The result was evaluated based on the mean value of the time taken (in a sec) to complete the execution process of each algorithm. Also, the efficiency is measured in terms of the minimal route length which each algorithm gave as the result based on the criteria of each algorithm set while coding using Python 3.8. If it is found that the mean value is less than 1.000E−10, then the result is stated as 0.000E+00. In this experiment, only the fundamental versions of swarm intelligence techniques are well thought-out and no specific modifications are introduced. Algorithm codes are tailored from multiple sources and updated to adhere to our experimental setup. In the forthcoming section, each SI algorithms taken for the study purpose is elaborated.

2.1 Differential Evolution (DE)

DE is a new type of progression algorithm which is alike GA. DE is a type of heuristic algorithm proposed by Kenneth Price and Dainer Storn in 1995 [2]. DE is a tool for optimizing a given population. It is used to obtain optimal solutions from a given set of constraints in various numerical problems [8]. DE is considered similar to the Genetic Algorithm as it has many similar basic operations. The mutation is one such to say. Similarly, crossover and selection are also operations of both. The fundamental difference between these two algorithms is in producing many improved solutions. When DE works on mutation, GA shows its expertise in crossover operation [6, 9]. DE works by introducing a mutant vector. During the mutation operation, it computes the difference in weights between two vectors which are selected randomly and then adding it to a third vector [2]. The procedure of the DE algorithm starts with *Initialization* followed by *Mutation*, then again the process of *Recombination,* and finally deals with *Selection.* The DE pseudocode for TSP [9] is explained using Fig. 3.

Fig. 3 Pseudocode of DE for TSP implementation

1. Generate an initial dataset from TSPLIB
2. Initialize the initial population of individuals
3. Evaluate the objective function value
4. Perform Mutation operation
5. Apply Crossover operation
6. Update the generation's counter by 1 step
7. Verify the stopping criterion until condition is met.

Fig. 4 Pseudocode of GA for TSP implementation

1. Generate a standard dataset from TSPLIB
2. Initialize the genes population;
3. Apply local search and optimize the population
4. Perform population evaluation
5. While (termination criteria not met)
6. Do Selection;
7. Do Crossover;
8. Do Mutation;
9. Perform population optimization
10. Calculate and Update population

2.2 Genetic Algorithm (GA)

GA is a time-honored search-based heuristic optimization algorithm that operates on both the genetic strategy and the natural selection basis. The whole idea of GA was put forward by John Holland in the year 1975. The algorithm tries to mimic biological progression or evolution [6, 10]. GA is a specific type of algorithm that use techniques which are motivated by certain biological operations like inheritance. Mutation is another step in the process along with a selection followed by crossover or recombination. The working of the algorithm begins with finding problem solutions for the population of candidates. These candidate solutions are then evaluated on the basis of their problem-solving capability [11]. It is found that only the fittest survives. This then combines with each other, and finally, various possible solutions of next-generation are created. The algorithm is repeated and finds that this population of patterns "evolves" through operations like reproduction, mutation, and natural selection. The algorithm terminates when maximum numbers of generations are reached, or when it approaches a sensible fitness level for the considered population. The pseudocode of the Genetic Algorithm for TSP [10] is as shown in Fig. 4.

2.3 Particle Swarm Optimization Algorithm (PSO)

PSO again falls under the category of optimization algorithm enthused by swarm intelligence. In 1995, the algorithm was proposed by Kennedy and Eberhart. The algorithm is based on a simple mathematical model that tries to illustrate the communal behavior of birds, fishes, etc. The model relies mostly on the basic principles of self-organization and utilizes a simplified model based on social behavior to solve the optimization problems [11, 12]. PSO algorithm fundamentally works by taking into consideration a population or a group known as "*swarm of candidate solutions*" also identified as *particles*. This particle movement is restricted within the particular space of search. It works based on certain formulae to calculate the velocity of particle notated as (2) and also the position of the particle which is denoted by (3) [13]. Here, the best position of an individual particle is indicated as *pBest* and on the whole, the best position of the particle is recognized as *gBest*. These are then tracked

Fig. 5 Pseudocode of PSO for TSP implementation

1. Generate a standard dataset from TSPLIB
2. Calculate fitness value for each particle
3. Set value based on the Best value *(pBest)*
4. Choose best fitness value of all particles and set as *gBest*
5. For each particle, Calculate Particle Velocity
6. Update Particle position
7. Repeat until termination condition is met
8. Update the current best value
9. Evaluate the best solutions

by analyzing the activities of the particles in the swarm. Based on the improved positions, the movement of the swarm is guided forward [14]. The process continues until a satisfactory solution is discovered. The formula for the PSO algorithm is described as shown:

$$\boldsymbol{v}_{k+1}^{i} = \boldsymbol{v}_{k}^{i} + \boldsymbol{c}_1 \boldsymbol{r}_1 \left(\boldsymbol{p}_{k}^{i} - \boldsymbol{x}_{k}^{i}\right) + \boldsymbol{c}_2 \boldsymbol{r}_2 \left(\boldsymbol{p}_{k}^{q} - \boldsymbol{x}_{k}^{i}\right) \tag{2}$$

Gives the velocity of individual particles whose position is given by

$$\boldsymbol{x}_{k+1}^{i} = \boldsymbol{x}_{k}^{i} + \boldsymbol{v}_{k+1}^{i}. \tag{3}$$

where

$\boldsymbol{x}_{k}^{\prime}$	Particle position
$\boldsymbol{v}_{k}^{i}$	Particle velocity
$\boldsymbol{p}_{k}^{i}$	Best "remembered" individual particle position (***pBest***)
$\boldsymbol{p}_{k}^{q}$	Best "remembered" swarm position (***gBest***)
c_1, c_2	Cognitive and social parameters
r_1, r_2	Random numbers between 0 and 1.

The PSO algorithm for TSP [12] is shown in Fig. 5.

2.4 Simulated Annealing (SA)

SA is found out as another effectual optimization algorithm that focuses on simulating the annealing of metals [15, 16]. Here, a solid metal's temperature is made to rise until it fuses in a heat bath. The metal is then slowly cooled until the particles are rearranged in the solid ground state. Thus, the metal's physical properties change with its internal structure. This will happen if the temperature is kept high enough, and it slowly decreases [17, 16]. The entire procedure is known as the simulation of the process of annealing. The method begins by having the heating cycle simulated with a temperature variable. A high initial value is attained to decrease the value gradually as the algorithm runs. Then, the algorithm is allowed to jump out of any local optimums during its execution. The opportunity to consider worse options will diminish as the temperature drops [17, 18]. Therefore, the algorithm is allowed to

Fig. 6 Pseudocode of SA for TSP implementation

```
Generate a standard dataset from TSPLIB to initialize population
While (termination criteria not met)
Create new solutions
Access new solutions
If new solution is accepted, update storage
Try adjusting temperature
Calculate and update solutions
```

concentrate gradually on a region in the search space from which a near optimum solution is sought. Figure 6 depicts the pseudocode of SA for TSP [18].

2.5 *Ant Colony Optimization Algorithm (ACO)*

ACO is another SI-based algorithm that comes under the meta-heuristic category which was put forward by Marco Dorigo in the year 1992 [10, 19]. ACO works on the principle of indirect communication using pheromones, which is released by ants. Pheromones are the chemical substances that attract additional ants when looking for food. The magnetism of a given direction depends on the magnitude of the pheromones the ant detects. The excretions of pheromone follow some rules. It is not always the same strength that they display. Each quantity of excreted pheromones depends on the nature of the path they are navigating. Evaporation is the main pheromone mannerism, and the entire process is time-dependent. When a track is not used, these pheromones quickly evaporate and they start using some other route [20]. A colony or group of ants moves through diverse states of the problem which are influenced by two decision rules, namely *trails* and *attractiveness*. Also, they follow the other two mechanisms called *trail evaporation* and *daemon actions*. As mentioned, the main aim of the algorithm is to find the shortest path that is optimal based on the actions of ants searching for a path between the colony and the food source point. Thus, every single ant slowly builds a solution to the problem [11, 21]. Figure 7 depicts the pseudocode of ACO for TSP [20].

Fig. 7 Pseudocode of ACO for TSP implementation

```
Generate a standard dataset from TSPLIB;
Initialize the pheromone trails
While (termination criteria not met)
Build Solutions
Apply Local Search;
Update Trails
Evaluate the best solutions
```

2.6 Artificial Fish Swarm Algorithm (AFSA)

Another approach that is widely accepted in swarm intelligence is AFSA, which works based on stochastic search as well as the population under consideration [22, 23]. The school of fishes shows intellectual societal manners like random, searching, swarming, chasing, and leaping type of behaviors. The AFSO was first proposed in 2002. Since the system works on population, it is initialized first into a series of possibly randomly generated solutions and then iteratively searches for the optimum one [22]. The atmosphere in which the artificial fish lives is known as the solution space. Each fish's behavior depends on its current state and environmental situation. The environment is influenced by its own activities as well as the other companions' activities [22]. The three basic behaviors of AF are *prey* followed by the *swarm* and *then follow* [24]. Behaviors of each fish vary with the situation and are described as—*foraging, huddling, random, bulletin board,* and so on [25, 26].

The three basic behaviors and working of Artificial Fish can be illustrated [24] as:

(a) ***Prey_AF***: The fish perceives the concentration of food in water to assess the movement through vision or sensation and then selects the pattern [22, 26]
(b) ***Swarm_AF***: The fish must usually gather in groups during the moving cycle, which is a kind of living habits to guarantee the colony's survival and to escape dangers [26].
(c) ***Follow_AF***: The community partners will track and easily enter the food in the moving phase of the fish swarm when a single fish or multiple fish find food [24, 26].

The working of AFSA can be elaborated by considering the total number of AF as N, and its individual state as X whose range can be defined as $x_1, x_2, \ldots, x_n$, in which x_i gets the value from $1, 2, 3, \ldots, n$. These are the variables considered to be optimized. Next, assume that the best ever moving step of AF as *Step*. Also, consider the perceived_distance of the AF as *Visual*, δ to be the congestion_factor and the distance of the AF denoted as (i, j) as $\mathbf{d_{ij}} = |\mathbf{x_i} - \mathbf{x_j}|$. The concentration of food for the AF is depicted in terms of $Y = f(x)$. Here, the objective function value is taken as Y, and the number of tries is assumed as *Try_Num* [25, 26].

Similarly, the behavior of the fish differs from the situation. They can be described as follows:

2.6.1 Foraging _Behavior

This is a fundamental behavior of each AF, which is mentioned as the activity of moving toward the food. Here, the AF gets attracted to the food in water by perceiving the concentration of the food with the help of its vision [25, 27]. The current state of every AF is measured as x_i. Let x_j be a randomly chosen perception state which is given by:

$$x_j = x_i + \text{Visual} \cdot \text{Rand}() \tag{4}$$

Here, Rand( ) is any random number in the range between 0 and 1.

Also, consider the condition when $Y_i < Y_j$. Then, the step forward will be performed by AF in this direction notated as

$$x_i^{t+1} = x_i^t + \frac{x_j - x_i^t}{\left\| x_j - x_i^t \right\|} \cdot \text{Step} \cdot \text{Rand}(). \tag{5}$$

Else, repeat the try *Try_Num* times by randomly selecting status x_j to ensure whether the forward condition is satisfied. Still, if a forward condition is found dissatisfied, randomly select step by using (6).

$$x_i^{t+1} = x_i^t + \text{Visual} \cdot \text{Rand}(). \tag{6}$$

2.6.2 Huddling_Behavior

In this case, the fish will naturally seek to form a cluster to ensure the safety of groups and to avoid swimming hazards. AFSA says every fish should migrate as far as possible to the center of the neighboring partners and not be overcrowded [27]. For this, consider the current state of AF to be x_i. Next, assume that the search for the number of partners as n_f and x_c as its central location. If it is found that $Y_c/n_f > \delta Y_i$, it can be concluded that the partner at the center gets more food and the situation as not crowed. Thus, the forwarding step made toward the center of the partner is given by:

$$x_i^{t+1} = x_i^t + \frac{x_c - x_i^t}{\left\| x_c - x_i^t \right\|} \cdot \text{Stcp} \cdot \text{Rand}(). \tag{7}$$

If not, try performing the *foraging_behavior* again.

2.6.3 Following_Behavior

In this process, the neighboring partners will try swimming quickly to reach the point of food by following one fish that finds food. Rear-end behavior is just a chase behavior with the highest fitness for AF nearby. This process is also known as advancing toward a near-optimal partner. During this period, assume the current state of AF to be x_i. If x_j is set as the current neighborhood, then the maximum value of the partner to be searched is given as y_j. And if, $y_j/n_f > \delta y_i$, the situation is assumed as the state of x_j has a higher food concentration and is not found crowded. Thus, the forwarding step should be made toward x_j. Hence, it is given in (8) as shown:

Fig. 8 Pseudocode of AFSA for TSP implementation

1. Generate a standard dataset from TSPLIB
2. Calculate fitness value
3. Perform Clustering
4. Do foraging
5. Update the current best value
6. Update the distance among fish swarm
7. Exit after achieving maximum evolution algebra
8. Evaluate the best solutions

$$x_i^{t+1} = x_i^t + \frac{x_j - x_i^t}{\left\| x_j - x_i^t \right\|} \cdot \text{Step} \cdot \text{Rand}(). \tag{8}$$

Otherwise, repeat the steps of foraging behavior.

Some other behaviors exhibited by the AF are illustrated as follows:

2.6.4 Random_Behavior

This is usually considered as a default behavior of foraging. This action is to randomly pick a location, and then switch to the direction chosen.

It is meant for recording the individual state of the optimum AF position and the concentration of food. In this procedure of optimization, the own state as well as the billboard status measured per activity is checked for each individual AF. If the own state is found to be better than the status of a bulletin board, translate the billboard's own status and record the history of its optimal state.

This is used to record the autonomy behavior of every AF. When used for TSP implementation, it can be utilized to carry out cluster as well as to show rear-end behavior. Later, the same is used to evaluate the value of actions by selecting the optimal behavior to detect the value to be executed. Foraging_behavior is again found as its default behavior.

The pseudocode of AFSA for TSP [25] is given in Fig. 8.

These are considered as the basic behavior of a normal Artificial Fish Swarm Algorithm. Improved versions of AFSA [22] are also depicted in many other papers. But for our purpose of study, only local AFSA is used as per the explanation done in this paper.

3 Proposed Work Along with Its Experimental Results

The various meta-heuristic algorithms mentioned and discussed in this paper are implemented with a few alterations in some parameters in order to acclimatize and to decipher the problem of a traveling salesman. The investigational situation is executed using Python 3.8 programs implemented using a Sony VAIO Laptop, Processor is Intel Core i5, and the operating system is Windows 10. The TSP data used here are downloaded from the TSPLIB [7], and four test cases of a different number of cities were considered for the test to measure the performance of each meta-heuristic algorithm used [28]. These algorithms were iterated until they showed a congregated return value. An algorithm is usually said to converge when it shows the same result which can be considered best in its last 20–25 iterations. Here, the algorithms were tested by iterating it until their termination criteria were satisfied.

Figure 9 projects the route length comparison, and Fig. 10 shows the execution time taken by the six meta-heuristic algorithms referred here by considering a sample value of 51 cities from the dataset ***eil51***. It is clear that out of the six algorithms used in a sample dataset of 51 cities, the AFSA algorithm gives the optimal result with the least route length of 570.51. And comparatively the worst performance is shown by the SA algorithm. The result also shows that the AFSA algorithm again beats other algorithms in the execution process by considering 30 iterations with an average value of 2.283.

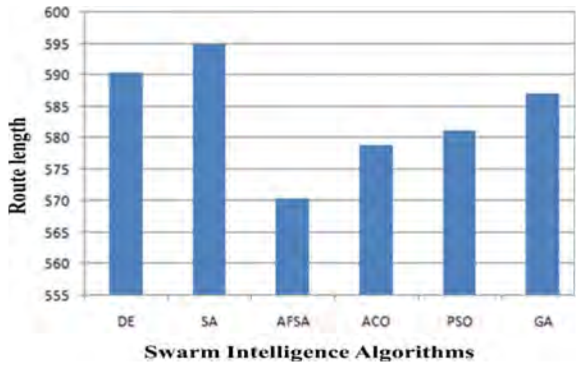


Fig. 9 Comparing route length obtained from different algorithms using eil51 dataset

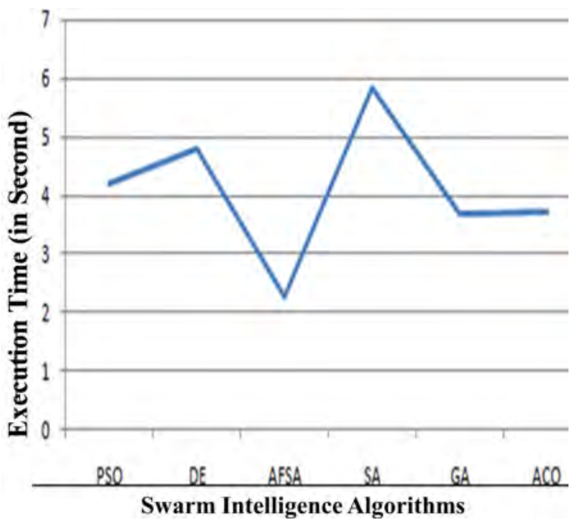


Fig. 10 Comparing the computation time of different algorithms using **eil51** dataset

During implementations, two main factors have to be considered to achieve better results. Locally searching within the problem space known as exploration and searching around the intended solution called extraction. Better results are achieved when one could balance these two factors. In the ACO algorithm, parameters such as α and β help to control the extraction of the algorithm which are empirically chosen as 1 and 2, respectively. In the SA algorithm, a high temperature is considered first by setting the exploration to large and as temperature decreases the extraction gradually becoming more. Therefore, first temperature T is considered to be high and then based on that, a cooling function is designed, which helps to lower the temperature slowly to obtain a high extraction. In the PSO algorithm, exploration and extraction are controlled by changing the weight of matrices accordingly. In the GA algorithm, crossover operation is used to increase exploration, and mutation operation helps for a better solution by implementing a local search on a solution by performing extraction [29, 30]. In the AFSA algorithm, the step size of the artificial fish and the number of the feeding process indicated as try-num helps to control the exploration and extraction. This adjustment is clearly mentioned in AFSA in Sect. 2 of this paper. Finally, with the DE algorithm, to control the exploration and extraction, it requires the six fundamental factors such as step size function indicated as *scaling factor*, *population number, initialization, crossover probability, evaluation function, and termination.*

Table 1 picturizes the performance or efficiency comparison of the different meta-heuristic algorithms mentioned here based on the four datasets selected. They are averaged and set over 30 runs. The corresponding time of execution of each algorithm was also measured based on a particular dataset of 51 cities. The result was evaluated based on the mean value of the time taken (in a sec) to complete the execution process of each algorithm. And the error rate of each was also noted. If it is found that the mean value is less than 1.000E−10, then the result is stated as 0.000E+00. Likewise, the output of each algorithm was taken, and the best case considering the dataset ***eil51*** was considered for finalizing the output. The corresponding results are displayed here within Table 2.

From the output received, it is very evident that an AFSA algorithm shows a better result than others when considering certain factors as mentioned. Thus, the Artificial Fish Swarm Algorithm works better than other algorithms mentioned here in this paper when considering a sample TSP dataset of 51 cities. The termination criteria considered here are the number of iterations which are manually set to 30. Tables 1 and 2 show the results so obtained by the algorithms.

Table 1 Performance comparison of algorithms based on route length

TSP Dataset	No. of Cities	AFSA	GA	ACO	PSO	DE	SA
burma10	10	271.628	300.75	319_95	325.69	342.12	321.28
eil51	50	570.515	537.18	578.8	581.24	590.47	594.85
rd100	100	801.746	847.57	809.51	837.52	874.56	869.63
KroA200	200	3007.48	3218.04	3092.96	3146.36	3345.28	3324.27

Table 2 Performance comparison of various algorithms based on its execution time, route length, and error rate for a sample dataset of 51 cities (eli51)

Meta-heuristic algorithms	Execution time	Route length	Error rate
AFSA	2.283	570.51	4.09E−01
GA	3.673	587. 18	1.78E−15
ACO	3.727	578.8	4.38E−01
PSO	4.214	581.24	8.21E−08
DE	4.805	590.47	2.00E+00
SA	5.821	594.85	1.37E−10

The output results represent the performance of each algorithm based on the selected dataset eli5, which is a dataset of 51 cities taken as an average among other datasets used for the study. The figures show the performance of each algorithm based on its route length and its execution time. Figure 11 projects the efficiency of the ACO algorithm based on its execution time against the number of iterations as well as the route length obtained against the set number of iterations.

Figure 12 emphasizes the efficiency of GA based on the route length and execution time against the number of runs of the algorithm. Figure 13 points out the performance

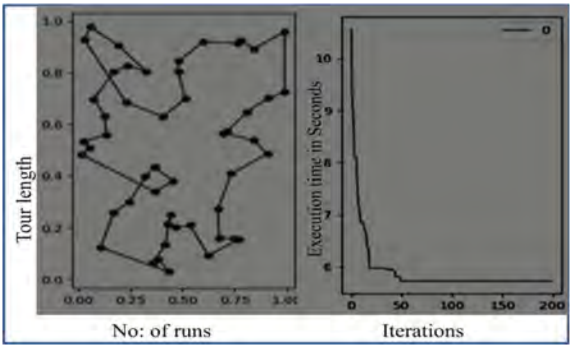


Fig. 11 Performance of **ACO** based on eli51

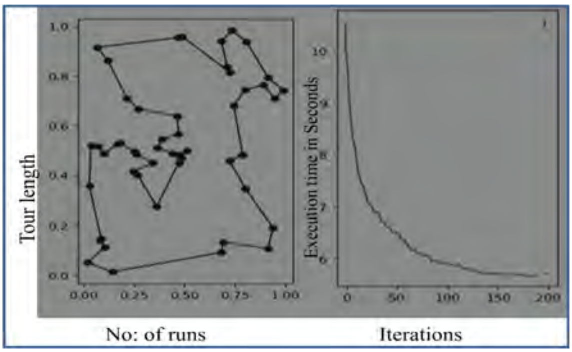


Fig. 12 Performance of **GA** based on eli51

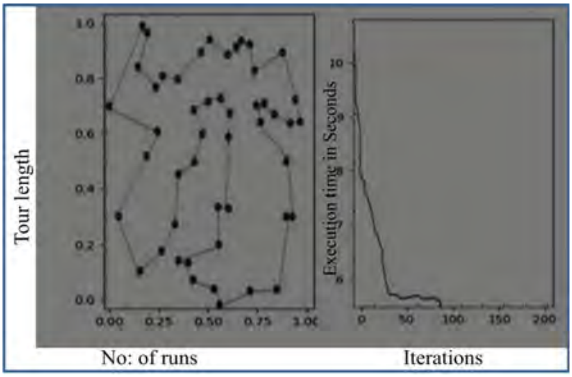


Fig. 13 Performance of **AFSA** based on eli51

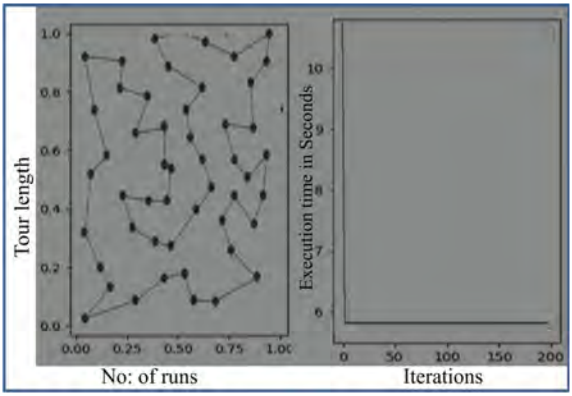


Fig. 14 Performance of **DE** based on eli51

of AFSA based on execution time and route length set against the iterations performed on the algorithm. Figure 14 represents the performance of DE algorithms based on the number of times the algorithm is executed against the time it takes to execute and produce the result along with the optimized route length. Similarly, Fig. 15 depicts the performance of PSO, and Fig. 16 showcases the efficiency of SA based on the number of iteration set to 30 against the route length and the execution time. Based on these findings, it can be concluded from the results that the AFSA algorithm is better than the other five algorithms used herein the analysis, both in terms of the length of the resulting routes as well as in terms of their execution time. AFSA is thus proved to have the advantages of fast execution speed as well as a strong ability to show the global convergence.

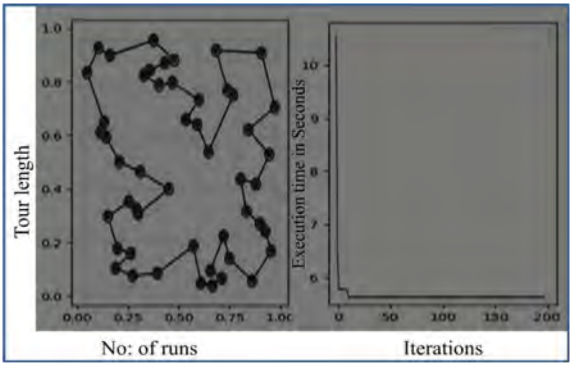


Fig. 15 Performance of **PSO** based on eli51

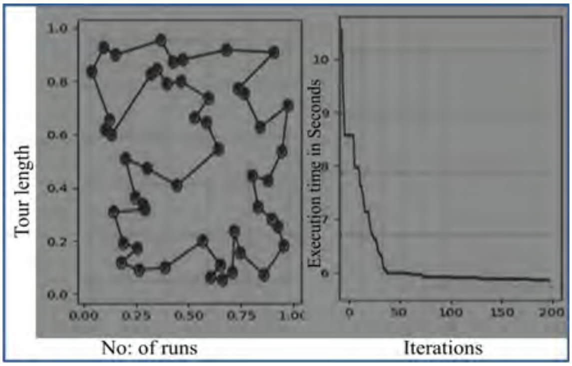


Fig. 16 Performance of **SA** based on eli51

4 Conclusion

The study was initiated to find an optimal solution to minimize the overall logistic cost incurred by an e-commerce firm. Thus, to move on with the work, it was decided to make minor improvements in the Vehicle Routing Problem (VRP) which can create a huge difference in the overall logistic cost. But, VRP problems deal with multiple constraints and factors. Initially, TSP was taken into consideration where it was dealt with only a single constraint "Number of Cities." But, finding an optimal route for a TSP is also critical to saving time and cost. In this work, six meta-heuristic swarm-based algorithms were considered and implemented to discover the finest minimum length of the route for the problem of traveling salesman. For this, four random datasets were selected from the TSPLIB consisting of cities from 10 to 200 in numbers. Each of the algorithms was individually tested to measure its performance based on the optimal route length along with its execution time. The set of algorithms considered for the study are the Genetic Algorithm, Simulated Annealing, Ant Colony Optimization, Artificial Fish Swarm Algorithm, Particle Swarm Optimization, and the Differential Evolution. These were implemented using Python 3.8 to measure the individual performances of these approaches. The results showed the superiority of AFSE with the ability to outperform others by covering the number of cities with minimum route length as well as with its minimum execution time of the algorithm to conclude with the result. The results identified AFSA as the best-performing algorithm to solve the TSP followed by ACO, then GA, PSO, SA, and finally DE. Every algorithm has got its advantages and shows its best performances which are selected based on the condition. Though AFSA demonstrates smarter behavior and produces more efficient performance than other swarm intelligence algorithms, it has got some shortcomings like showing a fall in local optimum points and advance in convergence in many situations while using multiple constraints. The conclusion is that the AFSA algorithm is one of the best swarm intelligence techniques, with key advantages like high convergence speed, versatility, resistance to errors, and high precision. Thus, researchers can use this along with its modified versions for their future research which involves multiple constraints such as in Vehicle Routing Problems in e-commerce to optimize the logistic cost.

References

1. Shima S, Mohammad S, Fardad F (2016) A comparison between swarm intelligence Algorithms for routing problems. Electr Comput Eng Int J (ECIJ) 5(1):17–33
2. Mei M, Xue H, Zhong M, Gu Y (2010) An improved differential evolution algorithm for TSP problem. In: International conference on intelligent computation technology and automation, pp 544–548
3. Vincent K, Matthew N, Spencer S (2014) Heuristics algorithms. (ChE 345 Spring)
4. Zharfi V, Mirzazadeh A (2013) A novel metaheuristic for travelling salesman problem. J Ind Eng 2013, Article ID-347825:5
5. El-ghazali T (2009) Meta-heuristic from design to implementation. Wiley Inc Publications, NY
6. Ab Wahab MN, Nefti-Meziani S, Atyabi A (2015) A comprehensive review of swarm optimization algorithms. PLoS ONE 10(5):e0122827
7. TSPLIB datasets http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95/tsp/
8. Vanita GT (2013) Travelling salesman problem using differential evolutionary algorithm. IOSR J Eng (IOSRJEN) ISSN (e): 2250-3021, ISSN (p): 2278-8719, pp 63–67
9. Xiang W, Guoyi X (2011) Hybrid differential evolution algorithm for traveling salesman problem. Adv Control Eng Inf Sci, Procedia Eng 15:2716–2720
10. Sabry AH, Benhra J, El Hassani H (2015) A performance comparison of GA and ACO applied to TSP. Int J Comput Appl (0975–8887) 117(1):28–35
11. Hosam HAM, Ashraf YAM (2015) Performance comparison of simulated annealing, GA and ACO applied to TSP. Int J Intel Comput Res 6(4):647–654
12. Appiah MY, Xiong Q (2019) Route optimization in logistics distribution based on particle swarm optimization. Int J Comput Appl (0975–8887) 178(30)
13. Yudong Z, Shuihua W, Genlin J (2015) A comprehensive survey on particle swarm optimization algorithm and its applications. Math Prob Eng 2015, Article ID 931256:38
14. Thirachit S (2018) Enhancing particle swarm optimization using opposite gradient search for travelling salesman problem. Int J Comput Commun Eng 7(4):167–177
15. Ai-Hua Z, Li-Peng Z, Bin H, Song D, Yan S, Hongbin Q, Sen P (2019) Traveling-salesman-problem algorithm based on simulated annealing and gene-expression programming. MDPI Inf J

16. Sumathi M, Rahamathunnisa U, Anitha A, Druheen D, Nallakaruppan. MK (2019) Comparison of particle swarm optimization and simulated annealing applied to travelling salesman problem. Int J Innov Technol Explor Eng 8(6):1578–1583 ISSN: 2278-3075
17. Xiutang G, Zehui S (2009) An effective simulated annealing algorithm for solving the traveling salesman problem. J Comput Theoret Nanosci 6:1680–1686
18. Ai-Hua Z, Li-Peng Z, Bin H, Song D, Yan S, Hongbin Q, Sen P (2019) Traveling-salesman-problem algorithm based on simulated annealing and gene-expression programming. Information 10(7)
19. Sapna K, Ibraheem AQA (2015) Ant colony optimization: a tutorial review. In: Conference paper-national conference on advances in power and control, at faculty of engineering and technology, International University, Haryana
20. Ivan B Jr, Zuzana Č (2011) Solving the travelling salesman problem using the ant colony optimization. Manag Inf Syst 6(4):010–014
21. Hui Yu (2014) Optimized ant colony algorithm by local pheromone update. TELKOMNIKA Indo J Electr Eng 12(2):984–990
22. Mehdi N, Ali A, Ghodrat S, Mehdi S, Adel NT (2012) A review of artificial fish swarm optimization methods and applications. Int J Smart Sens Intell Syst 5(1): 107–148
23. Nitesh MS (2020) Solving random travelling salesman problem using firefly algorithm. Int J Innov Technol Explor Eng 9(4):1037–1041. ISSN: 2278-3075
24. Yun C (2010) Artificial fish school algorithm applied in a combinatorial optimization problem. I.J. Intell. Syst. Appl 1:37–43
25. Teng F, Liyi Z, Yang Li, Yulong Y, Fang W (2014) The artificial fish swarm algorithm to solve traveling salesman problem. Adv Intell Syst Comput 679–685
26. Nitesh MS, Sanjay PP (2020) Solving a combinatorial optimization problem using artificial fish swarm algorithm. Int J Eng Trends Technol 68(5):27–32. ISSN: 2231-5381
27. Nurezayana Z, Azlan MZ, Safian S (2015) Overview of Artificial Fish Swarm Algorithm And Its Applications In Industrial Problems. Appl Mech Mater Trans 815:253–257
28. Haider AA, Ibrahim FA (2015) Comparison of algorithms for solving traveling salesman problem. Int J Eng Adv Technol 4(6):76–79. ISSN: 2249–8958
29. Gamal Abd ENS, Abeer MM, El-Sayed MEl-H (2014) A comparative study of meta-heuristic algorithms for solving quadratic assignment problem. Int. J Adv Comput Sci Appl 5:1
30. Elham D, Arash M (2017) Meta-heuristic approaches for solving travelling salesman problem. Int J Adv Res Comput Sci 8(5). ISSN No. 0976-5697

A Comparative Study on the Performance of Deep Learning Algorithms for Detecting the Sentiments Expressed in Modern Slangs

Vivank Sharma, Shobhit Srivastava, B. Valarmathi, and N. Srinivasa Gupta

Abstract Sentiment analysis is a text investigation technique that distinguishes extremity inside the text, regardless of whether an entire document, sentence, etc. Understanding individuals' feelings are fundamental for organizations since customers can communicate their considerations and emotions more transparently than any other time in recent memory. In this paper, the proposed model is the sentimental analysis on Twitter slangs, i.e., tweets that contain words that are not orthodox English words but are derived through the evolution of time. To do so, the proposed model will find the root words of the slangs using a snowball stemmer, vectorizing the root words, and then passing it through a neural network for building the model. Also, the tweets would pass through six levels of pre-processing to extract essential features. The tweets are then classified to be positive, neutral, or negative. Sentiment analysis of slangs used in 1,600,000 tweets is proposed using long short-term memory (LSTM) network, logistic regression (LR), and convolution neural network (CNN) algorithms for classification. Among these algorithms, the LSTM network gives the highest accuracy of 78.99%.

Keywords Snowball stemmer · Classifiers · Sentiment analysis · LSTM · CNN · Logistic regression · Twitter · Word2Vector

---

V. Sharma · S. Srivastava
Department of Information Technology, Vellore Institute of Technology, Vellore, Tamil Nadu, India
e-mail: vivanksharma@ymail.com

S. Srivastava
e-mail: shobhit.sri0108@gmail.com

B. Valarmathi (✉)
Department of Software and Systems Engineering, School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India
e-mail: valargovindan@gmail.com

N. Srinivasa Gupta
Department of Manufacturing, School of Mechanical Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India
e-mail: guptamalai@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_33

1 Introduction

The Web has significantly changed how individuals express their perspectives and assessments. Sentiment analysis refers to the task of recognizing suppositions, positivity decisions, and other data identified with the emotions and frames of mind communicated in ordinary language writings. Some of the emotions are happy, sad, frustrated, angry, and so on.

Slang is the utilization of exceedingly casual words, shortened forms, and articulations that are dismissed when asked to be taken as a significant aspect of the traditional language. The flood of online communications, for example, electronic mail, text messages, and microblogs administration made use of Internet slang practically omnipresent. It has turned out to be critical to gauge feeling extremity of the opinion or sentiment arranged slang present on the Internet that shows up in the investigation or analysis. Opinion analysis attempts to choose the sort of opinion (affirmative or not) in a given content. Emotion or feeling-based classification has a couple of basic characteristics, including different tasks, highlights, strategies, and context. A new strategy is proposed to classify the sentiment expressed in the microblog slangs in this paper. At that point, the proposed model offers a technique to aid in deciding the extremity available in Slangs.

2 Related Works

Almost all existing sentiment analysis algorithms to date are designed in such a way, that it classifies any content into two binary class, i.e., either it is positive or else negative [1–3]. Few recently emerged algorithms were designed in such a way that it extends binary classification to multipoint rating system, i.e., rating inference or multi-category problem [4–6]. Recently few researchers have proposed models for analyzing the reviews of a product or service with respect to all its features [7, 8]. For slangs, a slang dictionary is used to calculate the TF-IDF score to decide the polarity [9]. They made an underlying stride at programmed identification and recognizable proof of slang from normal sentences utilizing deep learning techniques. They showed how phonetic highlights joined with deep learning algorithms offer interpretability. They found that the bidirectional LSTM with feature-based inputs and character-based convolutional embeddings utilizing multilayer perceptron yields the best performance in position recognizable proof, and the model with comparative components aside from with conditional random field has better execution in distinguishing regardless of whether a source sentence contains a slang term [10]. Proposed Twitter information to identify depression [11].

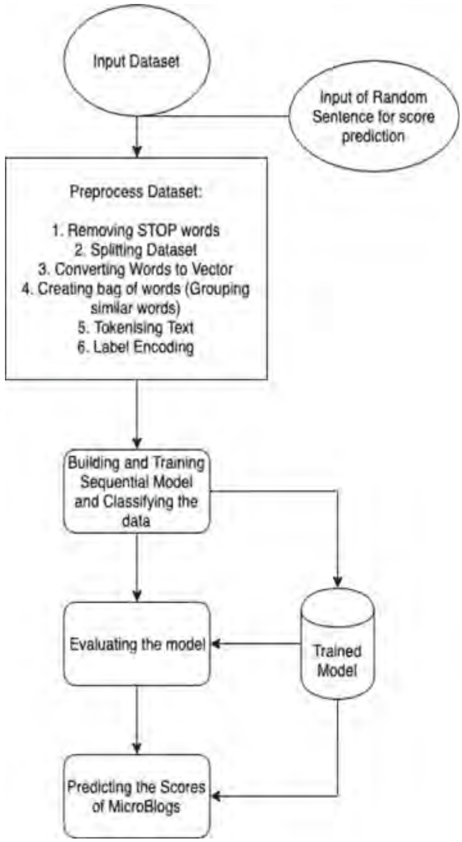
Better data pre-processing methods like changing over emojis to message structure, changing overstretched words to normal form, etc. were utilized to improve the identification accuracy. They extracted the highlights utilizing BOW, TF-IDF with n-grams, and Word2Vec procedures and used these highlights to the methods of classification. Logistic regression classifier (TF-IDF with n-grams) provided a 81% of the most extreme accuracy. Proposed a novel method called Representative Term-Document Matrix (RTDM) [12]. The given text document was transformed into a vector consisting of eight terms like bad, very bad, disgusting, never recommended, good, very good excellent, and recommended. A classification method is described using the Mahalanobis Distance (MD) [13]. The classifier name was Mahalanobis Distance Classifier (MDC). For 25,000 movie reviews, MDC achieved a 70.8% of accuracy. The hybrid classifier (MDC + MLP) performed with 98.8% of accuracy for 25,000 movie reviews. Deep learning technique is utilized to figure out sentiment analysis issues, for example, sentiment polarity [14]. Models utilizing term frequency-inverse document frequency (TF-IDF) and word embedding were applied to various datasets like Sentiment140, IMDB Movie Reviews, Cornell Movie Reviews, etc. CNN, DNN, and RNN were used in thisarticle.

3 Proposed Method

Current algorithms for sentimental analysis work great on formal English literature, but it fails when it comes to slangs. As slangs have no definite list and have no exact meaning, it is all based on contextual and scenario. So, the proposed model, a new algorithm, is developed which not only focuses on keywords but also takes into consideration every single word and tries to overcome slang by making custom datasets and using stemmers to find its original keyword and meaning. Hence, it makes our model different in the form and capability that it can also detect and analyze informal English, which is mostly used in microblogs. For example, for love, one can write many variants which are not in the English dictionary like "Luv," "Lub," etc. and the same one word can be even used to describe sentences also. "Shoulda" is replaced by should have. Similarly "lol" is replaced by "laugh out loud." So, to overcome this, the proposed model is taking more than 1 million Twitter data into account and training our model over these slangs and hence predicting it. Our model is achieving accuracy on the range of 70–80% on different datasets of different microblogs. The proposed model analyzes the microblogs by calculating a score by considering every word and slang and a sentiment score is calculated on a scale of 0-1 to assign sentiment lable. This can aid an organization to better comprehend the social sentiment of their service or their product by analyzing online conversations.

The first step is to preprocess the dataset. The second step is building and training the sequential model. The third step is to evaluate the model. The next step is to predict the score of Microblogs. The proposed model is summarized in the flowchart given in Fig. 1. The sentiment 140 dataset can be found in the link https://gofile.io/d/YS7U1r. It consists of 1,600,000 tweets that were obtained through the Twitter API. 1,280,000 tweets are used for the training and 320,000 tweets are used for testing. In the dataset, the class label "0" represents a negative tweet, whereas "4" represents a positive tweet. This can be used to detect sentiments [15]. The dataset label distribution which is depicted in Fig. 2 displays that both the positive and

Fig. 1 Proposed work flow chart



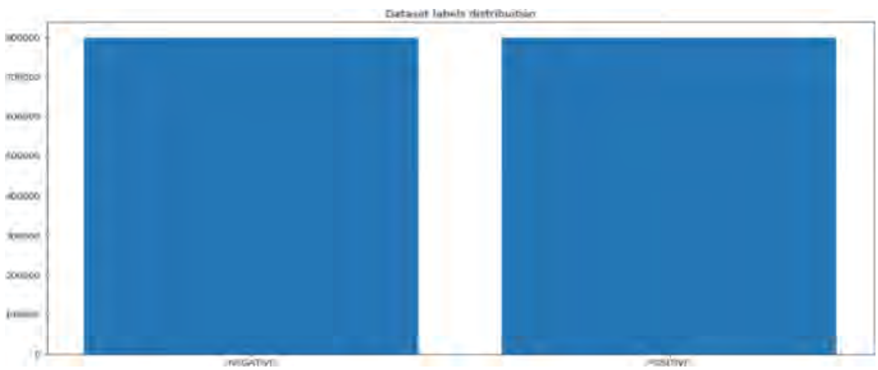


Fig. 2 Dataset distribution

negative labels are almost equal to the preprocessed data by removing the stop words. Content may contain stop words like "the," "is," "are." Stop words can be separated from the material to be prepared. There is no complete rundown of stop words in natural language processing inquire about; anyway, the NLTK module holds a group of stop words. All the stop words present in the dataset are removed. Then, the proposed model has used snowball stemmer for reducing inflected words in the dataset to their word stem, base, or root form. A stemmer is a calculation that works on the guideline of perceiving "stem" words implanted as such. These are useful for lexical purposes, for instance, in online lexicons, for heuristics in a record the board, or anyplace else that semantic apparatuses can help make a request.

Stemmers get the consideration of a center or stem word inside a more drawn out term. For instance, a stemming calculation may take a gander at a name like "planning," and accurately perceive that the root word or stem word is "plan." This can be a useful component of something that parses crude content for investigation, either for a site or some other venture. Snowball stemmer is famous for its compatibility with slangs. The next step is to convert all the text to vectors. These vectors are used by deep learning models over millions of words. The Word2Vec model helps in considering the context for further processing.

Word2Vec is a gathering of models which infers relations between a word and its relevant words. The two significant models inside Word2Vec are skip gram and continuous bag of words (CBOW).

The proposed model takes an inside word and a window of setting (neighbor) words and the model endeavor to foresee setting terms out to some window estimate for each middle name in skip gram model, Along these lines, our model will characterize a likelihood appropriation, for example, the likelihood of a word showing up in the setting given an inside word, and the next step will pick our vector portrayals to expand the possibility.

CBOW is just the inverse of the skip gram model. Try to forecast core-word by adding, vectors of neighboring words.

Skip gram as well as CBOW model output is shown in Fig. 3.

An example of a data pre-processing model is given in Fig. 4.

After this, tokenize the text, and then label encodes it for the further training process. Now, built the sequential model and it consists of four layers. Figure 5 shows the sequential model summary.

In the proposed model, the sigmoid activation function is used for the dense hidden layer. A sigmoid function exists between (0–1) and hence is the best model for predicting probability. The formula for the sigmoidal function is shown in Eq. (1).

```
similarity between 'alice' and 'wonderland' - CBOW :  0.9994316
similarity between 'alice' and 'machines' - CBOW :  0.99209344
similarity between 'alice' and 'wonderland' - Skip Gram :  0.9007044
similarity between 'alice' and 'machines' - Skip Gram :  0.86795944
```

Fig. 3 Example of data processed with skip gram as well as CBOW model

```
w2v_model.most_similar('love')
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:1: DeprecationWarning: Call to dep
recated `most_similar` (Method will be removed in 4.0.0, use self.wv.most_similar() instead).
  """Entry point for launching an IPython kernel.
2019-02-17 08:50:44,280 : INFO : precomputing L2-norms of word weight vectors

[('luv', 0.5840025544166565),
 ('loves', 0.5525496602058411),
 ('loved', 0.5403332710266113),
 ('adore', 0.5374413728713989),
 ('looove', 0.5025960206985474),
 ('amazing', 0.494439959526062),
 ('loooove', 0.47303086519241333),
 ('awesome', 0.46765822172164917),
 ('loveee', 0.456807404756546),
 ('lovee', 0.45561471581459045)]
```

Fig. 4 Data pre-processing grouping similar words

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (None, 300, 300)          87125700
_________________________________________________________________
dropout_1 (Dropout)          (None, 300, 300)          0
_________________________________________________________________
lstm_1 (LSTM)                (None, 100)               160400
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 101
=================================================================
Total params: 87,286,201
Trainable params: 160,501
Non-trainable params: 87,125,700
_________________________________________________________________
```

Fig. 5 Sequential model summary

$$\sigma(x) = 1/(1 + \exp(-x)) \tag{1}$$

Now, finally train the proposed model on the dataset, which consists of 1,600,000 tweets that were obtained through the Twitter API. In the dataset, the class label "0" represents negative tweets, whereas "4" represents positive tweets. This can be used to detect sentiments [15]. So, that it can further be used for predicting scores of the content to be analyzed.

4 Result and Discussion

On evaluating the test dataset, the proposed model gets an accuracy of 78.99% (LSTM Network), which is excellent compared to earlier achieved accuracy of 56.58–76.69% considering slangs.

Dang et al. [14] used sentiment140 dataset. Accuracy of DNN, CNN, and RNN algorithms with TF-IDF was 76.50%, 76.69, and 56.58. Among these algorithms, CNN was performing well. In the proposed model, the LSTM network, logistic regression, and CNN algorithms are used for processing. Among these algorithms, LSTM network was performing well. LSTM network gets an accuracy of 78.99%. Figure 6 shows the training, as well as the validation accuracy graphs.

Figure 7 shows the training and validation loss graph.

The confusion matrix is shown in Fig. 8. Hence from the confusion matrix, true positive, false positive, true negative, and false negative values are shown as true positive is 0.77, false positive is 0.19, true negative is 0.81, and false negative is 0.23.

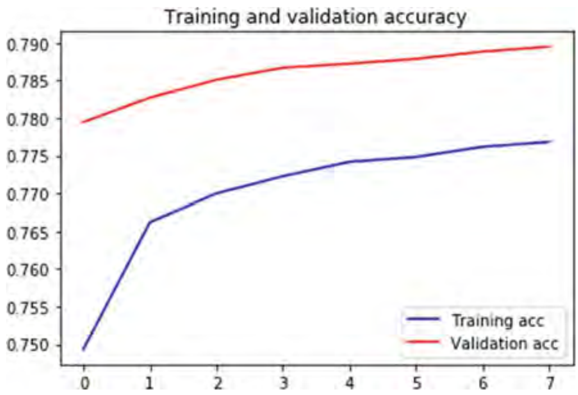


Fig. 6 Training and validation accuracy graph

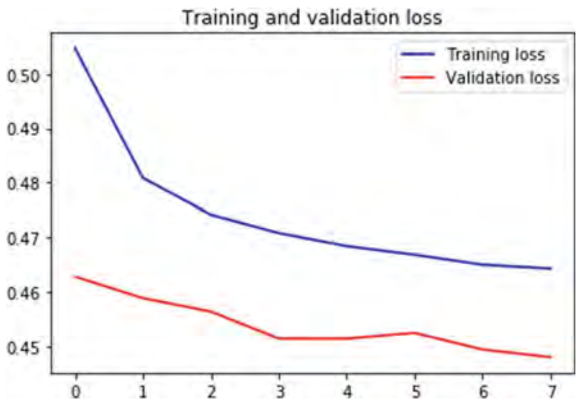


Fig. 7 Graph for training and validation loss

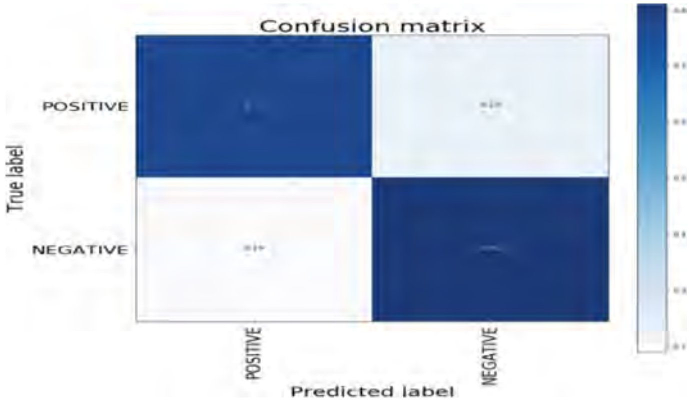


Fig. 8 Confusion matrix

This shows that our model is predicting scores and classifying sentences containing multiple slangs with very fewer errors. The proposed model is performing quite well-considering slangs in the dataset [16].

Now, let us predict some text and see how well out model is doing. Figures 9, 10, 11 and 12 show some examples of the score predicted by our model.

Hence, our model is giving scores from 0 to 1 up to 8 decimal places which can be further extended to rating inference from 0 to 10, where 0 being worst and ten being the best and can be applied on microblogs for analyzing and mining contents. Though our model performing great, there are still some future works that can be done like

```
predict("I love the music")
```

```
{'label': 'POSITIVE',
 'score': 0.963658332824707,
 'elapsed_time': 0.4383208751678467}
```

Fig. 9 Predicted score of “I love music”

```
predict("is upset that he can't update his Facebook by texting it... and might cry as a result
 School today also. Blah!")
```

```
{'label': 'NEGATIVE',
 'score': 0.03207983821630478,
 'elapsed_time': 0.2546839714050293}
```

Fig. 10 Predicted score of “is upset that he can’t update his Facebook by texting it … and might cry as a result School today also. Blah!”

```
predict("I love the music")
```

```
{'label': 'POSITIVE',
 'score': 0.963658332824707,
 'elapsed_time': 0.4383208751678467}
```

Fig. 11 Predicted score of "I love the music"

```
predict("I hate the rain")
```

```
{'label': 'NEGATIVE',
 'score': 0.015783820301294327,
 'elapsed_time': 0.2579636573791504}
```

Fig. 12 Predicted score of "I hate the rain"

Table 1 Comparison of the accuracies of various classifiers of the proposed method

S. No.	Name of the classifier	Accuracy (%)
1	LSTM network	78.99
2	Logistic regression	71
3	CNN	73

extending its from microblog to excellent contents like books, documentary, etc. and further improving it for multiple grams content.

Perform a sentimental analysis of Twitter slangs using LSTM Network, logistic regression, and CNN. Table 1 shows a comparative performance of the different techniques or classifiers used in the proposed method. Figure 13 shows accuracy and comparison of the LSTM network, logistic regression and CNN algorithms. Among these algorithms, the LSTM network gives the highest accuracy of 78.99% and it is shown in Fig. 13.The accuracy comparison of the proposed and existing methods is shown in Fig. 14.

5 Conclusion and Future Scope

The proposed work utilized Twitter data (sentiment140 dataset) to identify sentiment. After pre-processing the data, develop an algorithm that not only focuses on keywords but also takes into consideration every single word and tries to overcome slang by making a custom dataset and using stemmers to find its original keyword and meaning. Next, trained and tested the model by using various algorithms like long short-term memory network, logistic regression, and convolution neural network algorithms. Among these algorithms, the long short-term memory network gives the

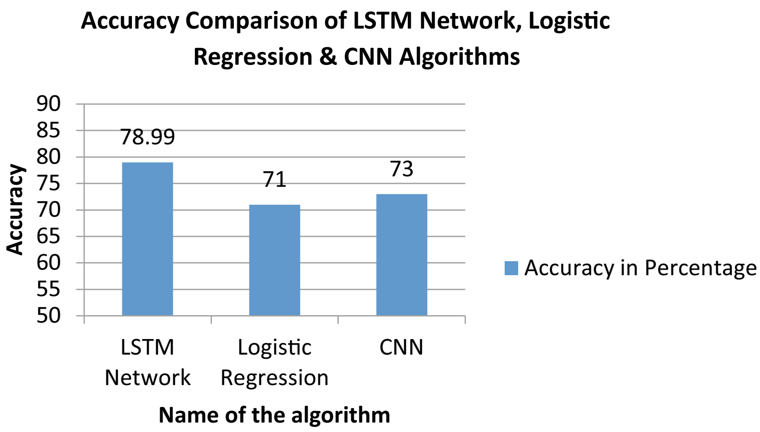


Fig. 13 Accuracy comparison of the LSTM network, logistic regression and CNN algorithms

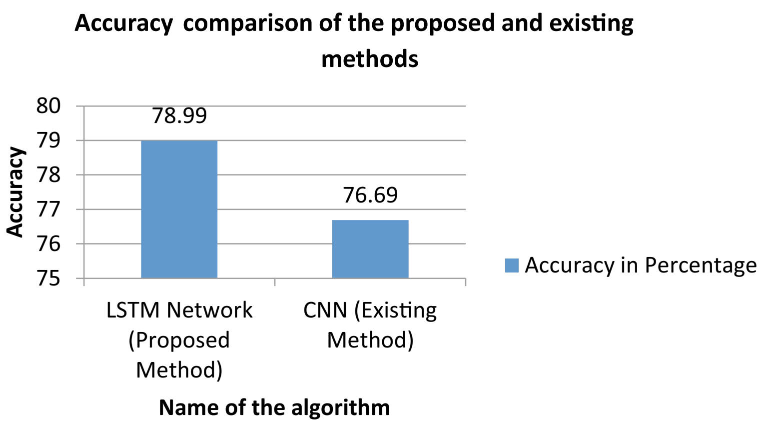


Fig. 14 Accuracy comparison of the proposed and existing methods

highest accuracy of 78.99%. In the future, to increase the accuracy of the proposed model, use bidirectional encoder representations from transformers and embeddings from language models algorithms for classification.

References

1. Dave K, Lawrence S, Pennock DM (2003) Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In: WWW '03: proceedings of the 12th international conference on World Wide Web, pp 519–528

2. Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment classification using machine learning techniques. In: EMNLP '02: Proceedings of the ACL-02 conference on empirical methods in natural language processing. Association for computational linguistics, pp 79–86
3. Turney PD (2002) Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: ACL '02: Proceedings of the 40th annual meeting on association for computational linguistics, pp 417–424
4. Goldberg AB, Zhu X (2006) Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization. In: Text Graphs '06: proceedings of text graphs: the first workshop on graph based methods for natural language processing on the first workshop on graph based methods for natural language processing, association for computational linguistics, pp 45–52
5. Ki Leung CW, Fai Chan SC, Lai Chung F (2006) Integrating collaborative filtering and sentiment analysis: a rating inference approach. In: ECAI 2006 workshop on recommender systems, pp 62–66
6. Pang B, Lee L (2005) Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. In: ACL'05: proceedings of the 43rd annual meeting on association for computational linguistics, pp 115–124
7. Gamon M (2004) Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. In: COLING '04: proceedings of the 20th international conference on computational linguistics, pp 841
8. Hu M, Liu B (2004) Mining and summarizing customer reviews. In: KDD '04: proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp 168–177
9. Manuel K, Varma K, Radha I, Krishna P (2010) Analyzing internet slang for sentiment mining. In: IEEE, 2010 second vaagdevi international conference on information technology for real world problems
10. Sun Z, Xu Y (2019) Slang detection and identification. In: Proceedings of the 23rd conference on computational natural language learning, pp 881–889
11. Mounika M, Srinivasa Gupta N, Valarmathi B (2020) Detection of depression related posts in tweets using classification methods—a comparative analysis. In: Proceeding of the international conference on computer networks, big data and IoT (ICCBI-2019), vol 49, pp 620–630
12. Srinivasa Gupta N, Valarmathi B, Sanju J (2012) Sentiment analysis using representative terms—a grouping approach for binary classification of documents. J Theor Appl Inf Technol 44:161–165
13. Valarmathi B, Srinivasa Gupta N, Palanisamy V (2016) Mahalanobis distance—the ultimate measure for sentiment analysis. Int Arab J Inf Technol 13:252–257
14. NhanCach D, María NM-G, Fernando De la P (2020) Sentiment analysis based on deep learning: a comparative study. Electron Open Access J 9:1–29
15. Go A, Bhayani R, Huang L (2009) Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, vol 1
16. https://towardsdatascience.com/sentiment-analysis-of-transliterated-texts-in-hindi-and-marathi-languages-c29eab19fae2

Exploration of Varied Feature Descriptors for Diabetic Retinopathy Through Image Classification

K. Sreekumar and E. R. Vimina

Abstract Diabetic retinopathy (DR) is a disorder affecting retinal blood circulation; it is precipitated by diabetes mellitus. Since DR can lead to complete blindness, its early detection is of utmost importance. The computational methods for detecting DR include segmentation, feature extraction, and classification. This study uses different feature descriptors—local directional pattern (LDP), local binary pattern (LBP), and histogram-oriented gradients (HOG) for extracting features by processing diabetic retinopathy images. A set of supervised learning classifiers—support vector machine (SVM), random forest (RF), and K-nearest neighbor (KNN)—are used to classify the retina images based on the features extracted. The study indicates that a combination of LBP and HOG coupled with RF is the most accurate form among the various candidate approaches and gives the best accuracy of 87.50%.

Keywords Diabetic retinopathy · Retinal fundus image · Machine learning · Classification · Image processing · Feature extraction

1 Introduction

Diabetes is a chronic disease formed by an abnormal increase in blood sugar in the body. In 2015, as per estimates made by the International Diabetes Federation (IDF), there were 415 million diabetes affected persons in the world by 2040; the number is projected approximately to surge to 642 million that is 10% of all adult humans. Diabetic patients often face severe complications of diabetic retinopathy (DR) [1]. It affects the retina of the patients, and if left untreated, it can lead to blindness. A popular domain where computer vision is applied in object recognition or object classification the principal intent of this domain is extraction of features from the

K. Sreekumar (✉) · E. R. Vimina
Department of Computer Science and IT, Amrita Vishwa Vidyapeetham, Kochi, India
e-mail: sreekumar4@gmail.com

Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Kochi, India

E. R. Vimina
e-mail: vimina.er@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_34

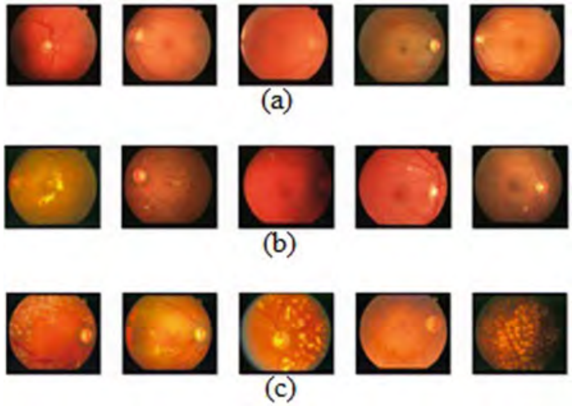


Fig. 1 **a** Normal, **b** non-proliferative diabetic retinopathy, **c** proliferative diabetic retinopathy

images and classification of images into suitably defined classes for this, any one of several possible classifiers or methods of classification could be used [2]. The various classification methods used here are SVM, RF, and KNN. A STARE dataset, containing around 400 images related to eye diseases, is used for the analysis. Around 131 retina images from the above set belonging to three categories, namely proliferative diabetic retinopathy (PDR), non-proliferative diabetic retinopathy (NPDR), and normal are selected interrelated to the diabetic retinopathy detection. Figure 1 shows a sample of retinal fundus images that are taken from the STARE dataset. The sample images on the first row belong to (a) normal category, the second row contains images showing (b) non-proliferative diabetic retinopathy, and the last row contains the images which show (c) proliferative diabetic retinopathy.

2 Related Work

Diabetic retinopathy can be detected by screening retinal fundus images of normal and diabetic patients. The optic disk, blood vessels, and fovea are the standard features of the fundus images. The blot hemorrhages and exudates are the main uncharacteristic features of diabetic retinopathy [3]. The features that mark NPDR are microaneurysms, dot and blot hemorrhages, flame-shaped hemorrhages, and retinal edema. The different features indicating PDR are new vessels at the disk (NVD) and new vessels elsewhere (NVE). The exudate can be detected by using strategies such as thresholding, edge detection, and classification. Backpropagation in a neural network can be used for feature selection to detect hemorrhage areas in the retinal fundus image.

A lot of work has already been done on methods for classification of DR and results have been largely encouraging. Pratt et al. proposed a convolutional neural network (CNN) approach, an area of deep learning for diagnosis of DR by analysis of digital fundus images as well as quantification of its severity [4]. They set up a network based on CNN architecture and incorporating data intensification and this resulted in a sensitivity of 95% from among 5000 images for validation. The accuracy was 75%.

Bhatia et al. presented a paper that seeks to arrive at decisions about the presence or absence of the disease by employing a collection of machine learning and classification algorithms [5]. Chandran et al. proposed an algorithm for spot wise extraction of features from the retinal shallow [6]. A random forest classification procedure was carried out on the features resulting from image spots. The approach displays enhanced specificity in comparison to previously existing methods and is suitable for deployment as a tool that performs screening for diabetic retinopathy.

Bui et al. described a method of automated dissection and used it to identify cotton wool patches on retina images as a marker of diabetic retinopathy [7]. Identification of cotton wool spots could help in the prevention of serious damage which could even lead to a total loss of vision. Their study demonstrates that cotton wool could be segmented with the accuracy was 85.54%; sensitivity, 85.9% and specificity, 84.4%. Carrera et al. proposed a diagnostic technique that uses digital processing performed on retinal images to achieve early detection of diabetic retinopathy. The objective is to achieve an automatic classification in terms of non-proliferative diabetic retinopathy (NPDR), grade found in a given retinal image [8]. This proposal has been tested and achieved 94% predictive capacity and a maximum sensitivity of 95%.

Arora et al. proposed a solution to overcome the problem generated using a machine learning method employing techniques of deep learning along with CNN automatically does pattern identification and achieves classification of imageries of retina among 5 classes [9]. The model's learning rate was 0.001 resulting in 74% validation accuracy. Alzami et al. proposed a method that works with the MESSIDOR dataset and employs a random forest classification and found that fractal dimensions are capable of distinguishing between healthy persons and those affected with diabetes retinopathy [10].

Bindhu proposed a method employing the semantic segmentation model for the biomedical images reducing the diagnostic time [11]. Samuel proposed a smart algorithm used for image processing which will be beneficial for the visually challenged by utilizing the text recognition, extraction of vocalization, and information [12]. Abraham proposed a method of assessment based on medical imaging technology (IoT-based) for the application used in health care [13].

Different types of techniques, architectures, simulations, and frameworks introduced by many researchers have played an important role in detecting lesions in the early stage of DR. Extraction of features of retinal fundus image and application of machine learning techniques to them is one of the methods to detect DR. Another method for classification of retinal fundus images, deep learning can be used in the field of DR. Deep learning is a class of machine learning algorithms that uses multiple layers to gradually excerpt higher-level features from the raw input. The convolutional neural network (CNN) approach, an area of deep learning, is an advanced method used for image classification to detect DR.

Thus, it is found that the various feature descriptors play a major role in extracting the required features and help the classifiers to perform the classification on retinal fundus images. In this study, diverse feature extraction methods like LBP, LDP, and HOG along with analysis of the various classification methods like SVM, KNN, and random forest are done on a set of retina images and the net accuracy for each combination is observed to select the best one.

3 Proposed Work

A comparative analysis is performed for classifying a set of diabetes retinopathy images by using different feature descriptors like LBP, LDP, and HOG along with machine learning algorithms. The classifiers like SVM, KNN and random forest will quantify the net accuracy of the experiment. A combination of LDP + HOG and LBP + HOG is also tested with the above mentioned classifiers to find the best accuracy rate.

The retina images are taken from a STARE dataset, which contains around 400 images and in that 131 images are related to DR. After the image acquisition, the same are preprocessed by applying green channel extraction to avoid noisy data. Various feature descriptors are applied to perform the feature extraction followed by classification done by the classification algorithms to obtain the various category labels, and finally, the net accuracy is retrieved. Figure 2 depicts the flow chart for the proposed work.

In this experiment, retina images which belong to three different classes namely proliferative diabetic retinopathy (PDR), non-proliferative iabetic retinopathy (NPDR), and normal are used. The feature descriptors LBP, LDP, and HOG are used for feature extraction along with machine learning classification algorithms SVM, RF, and KNN to retrieve the accuracy rate. 70% of the total 131 dataset images are trained in which the preprocessing and the feature extraction are executed. The extracted features are stored in a feature database. The remaining 30% of the dataset images are tested and classified into three categories.

3.1 Preprocessing

The preprocessing is done on the retina images collected from the STARE dataset. The retinal fundus images are RGB color images, which consist of three channels such as red, green, and blue. The green channel is alone extracted from this concatenated RGB, since the green channel will have less noise compared to the other two channels.

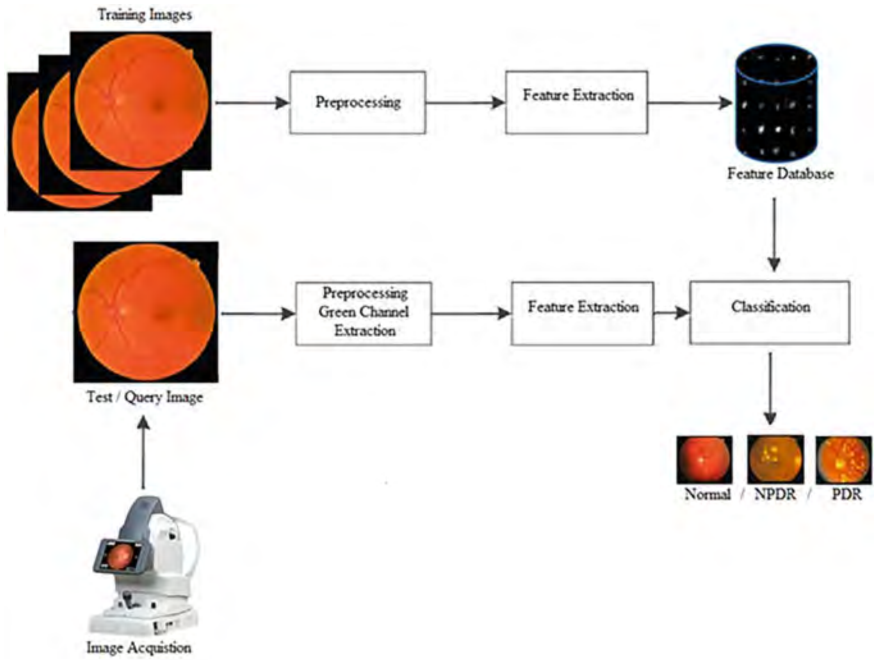


Fig. 2 Flow chart

3.2 Feature Extraction

Feature extraction comprises revealing and segregation of features of a given image such as shape, shading, baseline—and these features are combined to create a vector. The extraction of features is most crucial, as the specific features made available for selection determine the effectiveness of classification.

3.3 Local Binary Pattern (LBP)

In computer vision, the LBP, a visual descriptor, is used for classification. It marks image pixels by relating a threshold to the neighborhood of the pixel and extracts a specific binary number as result. This texture operator has gained acceptance in diverse applications, primarily due to its power of discrimination and essential simplicity [14].

Local binary pattern (LBP) is a non-parametric descriptor; it aims to summarize with efficiency, local structures in images. LBP works among adjacent pixels by fixing a center threshold. Checking if the neighboring value of the pixel exceeds or equals that of the pixel selected as the center which indicates it as 1, else 0. The decimal equivalent of the 8-digit binary number obtained is set to the central-pixel threshold. By applying a threshold to every pixel, obtain an LBP code corresponding to each pixel value.

3.4 Local Directional Pattern (LDP)

The various challenges of the LBP such as noise and illumination change can be overcome easily by using the feature extraction method local directional patterns (LDP). It is an image feature that calculates the address values corresponding to the strings in various directions and then applies these values to work out the encoding of image textures [15, 16]. It is a code (8 bit) assigned to each pixel constituting the picture and the same is worked out in multiple methods.

Kirsch masks m_i—a typical edge detector, processes the mask values, with i ranging among values 0–7, founded on their position concerning 8 possible orientations, given a central-pixel within the picture. Figure 3 shows those masks. Along with some directions, the intervention of an edge or corner is indicative of greater reaction values is to understand the k's almost all significant routes for the above descriptor. The b_i is assigned to the highest retort shown by the guiding bit, k. All the other bits are set to 0. Finally, it jumps to Eq. (1). m_k indicates the kth ranked among relevant directional responses. The m_i is the eight directional edge responses and b_i is the bit response.

$$\mathrm{LDP}_K = \sum_{i=0}^{7} b_i(m_i - m_k) \times 2^i \tag{1}$$

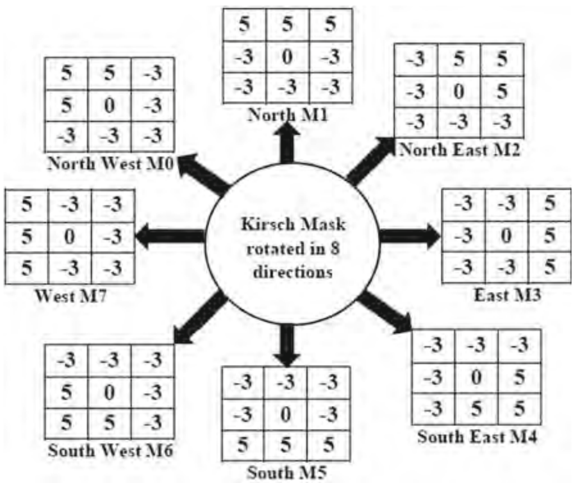


Fig. 3 Kirsch masks

$$b_i(a) = \begin{cases} 1 & a \geq 0 \\ 0 & a < 0 \end{cases}$$

3.5 Histogram-Oriented Gradients (HOG)

Image appearance and shape can be denoted with HOG, which is a simple and efficient technique for extracting features. The shape and appearance of an image are denoted with HOG; the image is divided into tiny cells and edge directions calculated. To enhance precision, normalize the histograms. The HOG descriptor is centered on the design or form of the objects and the decision is made whether the pixel is part of an edge or not. The edge directions given by HOG are found in confined segments. It splits the full image into constituent gradients and parts; then directions corresponding to every region are calculated. The HOG yields a unique histogram for each of these areas and these histograms are formed out of pixel value gradients and orientations [17].

Each tiny patch is chosen from the images and then the gradients are worked out. The pixel values are needed for the patch. To estimate the x-axis alteration (G_x), subtract the pixel value to the left from the rightward pixel value. Next, subtract the beneath pixel value from the overhead pixel value of the selected pixel to find the y-direction gradient (G_y). The same cycle is repeated for every pixel in the image. Now, display the magnitude as well as the direction of every value of a pixel with the help of the x and y gradients computed in the prior stage. The magnitude, M—overall gradient vector for a point (x, y) is obtained from its Euclidean vector norm in Eq. (2).

$$M = \sqrt{\left(G_x^2 + G_y^2\right)} \tag{2}$$

The gradient vectors, direction, or angle (α) for a point (x, y) can be computed using Eq. (3).

$$\alpha = \tan^{-1}\left(G_y / G_x\right) \tag{3}$$

The histograms yielded by the HOG function descriptor for the full image are not established instead it subdivide the picture into 8 × 8 cells and for every cell determine the histogram of each directed gradient. Thus, for smaller patches, the histogram (or features) represents the full image. This value here certainly gets altered to 16 × 16 or 32 × 32 from 8 × 8 and the maximum benefit occurs in the path of the bin with respect to the pixels' location.

Algorithm:
Step 1: Image Acquisition (Input Image: Retinal Color Fundus Image)
Step 2: Green Channel Extraction
Step 3: Feature Extraction
  Method 1: LDP
  Method 2: HoG
  Method 3: LBP
  Method 4: LDP+HoG
  Method 5: LBP+HoG
Step 4: Classification
  Method 1: SVM
  Method 2: KNN
  Method 3: Random Forest
Step 5: Calculate Net Accuracy

Fig. 4 Algorithm

3.6 Classification Methods

The classifier algorithms like SVM, RF, and KNN are used for classifying various retina images. Support vector machine (SVM) was introduced by Vapniket et al. [5, 18]. Training lengthy decision trees with several train patches are the main feature of the random forest algorithm. [5, 6]. KNN classifier works according to the distance/similarity function that classifies based on similarity measure [19].

Figure 4 explains the algorithm for the proposed work. It begins with image acquisition, followed by green channel extraction and then the feature extraction using LDP, HOG, LBP, LDP + HOG, and LBP + HOG, and finally the classification using the machine learning algorithms SVM, KNN, and random forest.

4 Result Analysis

4.1 Data Set and Implementation

A STARE dataset is used for this analysis. STARE stands for structured analysis of the retina; the University of California initiated this project. From around 400 images, 131 images belong to three different classes, namely PDR, NPDR, and Normal. Table 1, shows the total number of images in each category. The analysis is done on every

Table 1 Category of images

Category	No. of images
Normal	40
Proliferative diabetic retinopathy (PDR)	22
Non-proliferative diabetic retinopathy (NPDR)	69

image and they are grouped among the above categories based on the specifications. The computational algorithms for the classification of retina images using machine learning techniques have been formulated and implemented using MATLAB 2018a. A GUI interface has been created for selecting the corresponding feature descriptor and the classifier combination to find the net accuracy rate.

A classifier evaluation is done based on the true positive (TN), true negative (TN), false positive (FP), and false negative (FN), which are the values obtained based on a two-class domain. The TP is the positive tuples and the TN is the negative tuples correctly labeled by the classifier. The FP is the negative tuples incorrectly labeled as positive and FN is the positive tuples incorrectly labeled as negative by the classifier. The accuracy measure is computed using Eq. (4) as shown below.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4}$$

The entire dataset is divided into a training set (70%) and a testing set (30%) and the net accuracy is computed. In our experiment, the net accuracy of different feature descriptors like LBP, LDP, and HOG are coupled along with a set of classifiers—SVM, RF, and KNN are compared. Our study clearly shows that the combination of LBP and HOG coupled with RF gives the maximum accuracy (87.50%) among the various approaches. The observations shown in Table 2 describe experimental outcomes for the comparison of the net accuracy percentage for various classification methods and feature extraction combinations used. It is found that the combination of LBP + HOG along with the random forest algorithm is giving the highest accuracy of 87.50%.

The SVM classifier is combined with the different feature descriptors LDP, HOG, LBP, and the net accuracy of the respective combination is shown in the above table. The SVM is also used to calculate the accuracy rate by combining LBP + HOG and LDP + HOG and the application result is also shown in the above table. Similarly, the KNN classifier is combined with LDP, HOG, and LBP to get the respective accuracy rate as shown. The KNN is also applied with the combination of LBP + HOG and LDP + HOG to get the accuracy rate as shown in the above table. Finally, the random forest classifier is used along with LDP, HOG, and LBP to get the results. It is also used with the combination of LBP + HOG and LDP + HOG and the respective results are recorded in the above table.

Table 2 Comparison of net accuracy rate

	SVM (%)	KNN (%)	RF (%)
LDP	50.00	47.50	47.50
HOG	52.50	50.00	55.00
LBP	52.50	57.50	82.50
LDP + HOG	50.00	47.50	52.50
LBP + HOG	52.50	45.00	87.50

4.2 Discussions

The main advantages of SVM are that it works well with a clear margin of separation, memory efficient, and is very effective in high dimensional spaces. It is also suited for extreme case binary classification. The disadvantages of SVM are a higher training time needed for large datasets, unsatisfactory performance with noisy data, and that it will not directly provide probability estimates. An advantage of the KNN algorithm is the ease of understanding and implementation. It tags the new entry data based on the history available. It quickly responds to the changes in the input during real-time use and it automatically adjusts to multi-class without extra effort. The disadvantage of the KNN algorithm is that despite its simplicity, its efficiency and speed degrade for large data sets. KNN is very sensitive to outliers since it chooses the neighbors using the distance criteria. Random forest algorithms are a group of decision trees for prediction; they take input from all the trees and gives the prediction. This reduces the overall variance and error and thus improves accuracy. The other advantages of the random forest algorithm include the capability to handle huge amounts of data, to handle missing data, and robustness to outliers it is very useful for feature extraction. It has the following disadvantages: the predictions of the trees need to be uncorrelated and the features should have some predictive power else they give some error. The complexity and longer training period are some other minus points of the random forest algorithm.

Figure 5 depicts the net accuracy rate of classification approaches for a combination of feature descriptors. The chart clearly indicates the net accuracy rate for LDP, HOG, LBP, LDP + HOG, and LBP + HOG combined with the three classification algorithms SVM, KNN, and random forest. The combination of LBP + HOG along with the random forest algorithm gives the highest accuracy of 87.50%.

70% of the dataset has been used for the training of the model and the remaining 30% of the dataset, for testing. The implementation of the above comparative study is done using MATLAB 2018a. A graphic user interface has been created using

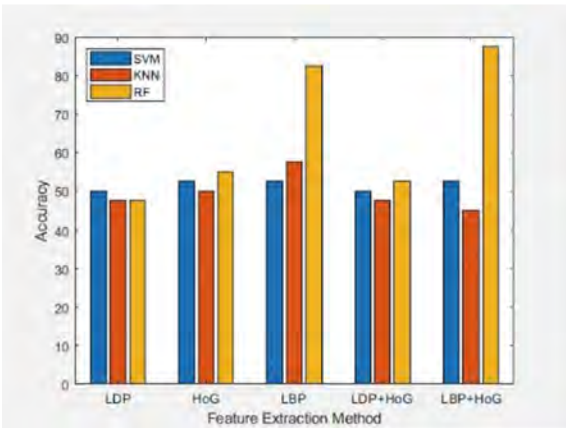


Fig. 5 Net accuracy rate of classification approaches for combination of feature descriptors

MATLAB to enable the selection of various simulation parameters. It is possible to select the classifier along with the feature extractor combination with the help of controls placed in the graphic user interface. The interface enables the display of the accuracy of the selected combination by a button click and there is also an option to print the net accuracy rate graph.

5 Conclusion

This article aimed to formulate an efficient feature extraction method for retina images by coupling different feature extraction methods with various machine learning techniques. The process was tested using a structured analysis of the retina dataset which contains 131 images, further grouped into 3 classes. The experiment results show that the combination of local binary pattern and histogram-oriented gradients feature extraction methods coupled with the random forest algorithm gives the best accuracy (87.50%), which compares favorably with the other combinations tested. It is suggested that this combination could be used in the analysis of retina images in particular to achieve accurate and timely classification that enables ophthalmologists to prioritize cases and identify emergencies that need immediate intervention.

References

1. Marina T et al (2002)Perceptions of diabetic retinopathy and screening procedures among diabetic people. Diab Med 19(10):810–813
2. Carmen V et al (2016)Automated detection of diabetic retinopathy in retinal images. Indian J Ophthalmol 64(1):26
3. Dilip Singh S, Nair S, Khobragade P (2017)Diabetic retinal fundus images: preprocessing and feature extraction for early detection of diabetic retinopathy. Biomed Pharmacol J 10(2):615–626
4. Harry P et al (2016)Convolutional neural networks for diabetic retinopathy. Procedia Comput Sci 90:200–205
5. Karan B, Arora S, Tomar R (2016)Diagnosis of diabetic retinopathy using machine learning classification algorithm. In: 2016 2nd international conference on next generation computing technologies (NGCT). IEEE
6. Anaswara C, Nisha KK, Vineetha S (2016)Computer aided approach for proliferative diabetic retinopathy detection in color retinal images. In: 2016 international conference on next generation intelligent systems (ICNGIS). IEEE
7. Toan B, Maneerat N, Watchareeruetai U (2017)Detection of cotton wool for diabetic retinopathy analysis using neural network. In: 2017 IEEE 10th international workshop on computational intelligence and applications (IWCIA). IEEE
8. Carrera EV, González A, Carrera R (2017)Automated detection of diabetic retinopathy using SVM. In: 2017 IEEE XXIV international conference on electronics, electrical engineering and computing (INTERCON). IEEE

9. Mamta A, Pandey M (2019)Deep neural network for diabetic retinopathy detection. In: 2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon). IEEE
10. Farrikh A, Megantara RA, Fanani AZ (2019)Diabetic retinopathy grade classification based on fractal analysis and random forest. In: 2019 international seminar on application for technology of information and communication (iSemantic). IEEE
11. Bindhu V (2019) Biomedical image analysis using semantic segmentation. J Innov Image Process (JIIP) 1(02):91–101
12. Manoharan S (2019) A smart image processing algorithm for text recognition information extraction and vocalization for the visually challenged. J Innov Image Process (JIIP) 1(01):31–38
13. Chandy A (2019) A review on iot based medical imaging technology for healthcare applications. J Innov Image Process (JIIP) 1(01):51–60
14. Prithaj B et al (2018)Local neighborhood intensity pattern–a new texture feature descriptor for image retrieval. Exp Syst Appl 113:100–115
15. Zhou J, Tianwei Xu, Gan J (2013) Feature extraction based on local directional pattern with svm decision-level fusion for facial expression recognition. Int J Bio-Sci Bio-Technol 5(2):101–110
16. Taskeed J, Kabir MH, Chae O (2010)Local directional pattern (LDP)–a robust image descriptor for object recognition. In: 2010 7th IEEE international conference on advanced video and signal based surveillance. IEEE
17. Navneet D, Triggs B (2005)Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol 1. IEEE
18. Vladimir V (1998)The support vector method of function estimation. In: Nonlinear modeling. Springer, Boston, MA, pp 55–85
19. Wu X, Kumar V (eds) (2009) The top ten algorithms in data mining. CRC Press

News Topic Classification Using Machine Learning Techniques

Pramod Sunagar, Anita Kanavalli, Sushmitha S. Nayak, Shriya Raj Mahan, Saurabh Prasad, and Shiv Prasad

Abstract News topic classification is a method of classifying news articles available in text data into some predefined classes or labels. This is one of the applications of text classification. Text classification can be applied in the fields of spam filtering, language recognition, segmenting customer feedbacks, segregating technical documents, etc. This paper discusses news topic classification on AG's News Topic Classification Dataset using machine learning algorithms such as linear support vector machine, multinomial Naive Bayesian classifier, K-Nearest Neighbor, Rocchio, bagging, and boosting. This paper discusses three steps for classification, namely pre-processing of text, then applying feature extraction techniques, and finally implementing machine learning algorithms. These algorithms are compared using evaluation metrics like Accuracy, Recall, Precision, and F1 Score.

Keywords Text Classification · Natural language processing (NLPs) · Term frequency–inverse document Frequency (TF-IDF) · Support vector machine (SVM) · K-nearest neighbours (KNN) · Naïve Bayes · Rocchio

---

P. Sunagar (✉) · A. Kanavalli · S. S. Nayak · S. R. Mahan · S. Prasad · S. Prasad
Department of Computer Science & Engineering, Ramaiah Institute of Technology, Bangalore 560054, India
e-mail: pramods@msrit.edu

Visvesvaraya Technological University, Belagavi, Karnataka, India

A. Kanavalli
e-mail: anithak@msrit.edu

S. S. Nayak
e-mail: nayaksushmitha90@gmail.com

S. R. Mahan
e-mail: shriyarajmahan@gmail.com

S. Prasad
e-mail: saurabhprasad12@gmail.com

S. Prasad
e-mail: shivpsmy227721@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_35

1 Introduction

Text classification or text labeling is the method of categorizing text into prearranged groups [1]. Due to the recent breakthroughs in Natural Language Processing (NLP) and text mining, the application of text classification has reached many real-time applications. The machine learning algorithms can be implemented to automatically evaluate text and then allocate a set of pre-defined tags or categories based on its content. Due to the digital era and social media influence, the unstructured text is in large volumes. It is difficult to analyze such a large amount of text data if not organized. Unstructured textual data is nothing but the text which is available in its natural free form. Machines try to analyze this textual data by implementing pre-processing techniques and extract the information by implementing specific algorithms. Using NLPs, it has proven to be fast, cost-effective, and scalable. Text classifications are widely used in real-time applications like spam filtering, sentiment analysis, automated responses to customers, fraud detection, etc. [2, 3]. Today, the amount of social media data is increasing day by day as more and more people have started using the social sites and data is in great abundance, so this data is of no use until and unless it is made in some useful way of being classified under some categories which later can be studied on, this is the classifications that the text classifier provides us with. In this work, the AG's News Topic Classification Dataset was considered which had four classes, namely Business, Sci/Tech, Sports, and World. Text classification works like this, the training set $N = (N1, \ldots, Nn)$ consists of news articles that are already labeled with a class Business, Sci/Tech, Sports, and World. The task is then to establish a classification model which will assign the exact class to a new news article.

2 Literature Survey

Text classification or text categorization is the process of categorizing text into organized categories. By using NLPs techniques, text classifiers can automatically analyze text and then assign a set of previously defined tags or classes. During recent years, studies are ongoing on the classification of unstructured texts. Papers were published regarding text classification to classify papers using text mining which classified text using keywords, expectation maximization, and hierarchical shrinkage [4]. There were also studies to classify text using bootstrapping methods with keywords. This naive Bayesian classifier works by first selecting a class of document, and then document words are to be generated based on class-specific traits. After this, make use of the Expectation–Maximization (EM) algorithm for the entire dataset to give weighted class labels for all sets of document. This model can improve the model iteratively by first determining the missing values and then using the found values to improve the model [5]. Also papers on some text classification techniques, using WordNet Hypernyms. In this classification technique, different binary classification problems of changing difficulty are to be defined, and discrimination rules are produced for each problem so defined using ripper system and hypernym density representation. Rules can also be given without any use of WordNet by using the knowledge of recently used words. The last problem faced after doing the above mentioned steps is whether the improvements made by the process above can generalize some machine learning algorithms or not. The changes defined above in the process will grant some semantic character. More revealing ambiguity for words can provide us with more accurate features for hyponym [6]. Text categorization is the process of categorizing or classifying text in different groups based on keywords or phrases and then giving the categories that are already previously defined. There are many applications of this such as content classification, handling loads of unstructured data from social media, the spam filtering in emails, etc. Here main keywords are used from the documents to classify and assign them some category of documents. Such keywords provide almost every information of the document's content. Such keywords can be called a summary of its content [7].

Later, another method for text classification was proposed which was using text segmentation which was based on a feature selection method that gives its utmost emphasis on the document structure. Then, features were also extracted from the topic of the paper or document rather than just the content of the document. Some authors proposed to extract features as short as possible such as from titles, summary, abstract, etc. Various methods of obtaining new features were then introduced [8]. The next method studied is for making an efficient text classifier was PDF text classification using information retrieval. In a PDF, the content of it is usually unstructured or semi-structured and often mixed which makes it challenging to classify using NLP algorithms. Information Extraction (IE) systems can help us in getting the useful information that are needed to classify but most of these systems are not built to work with the PDFs which is a very important source for information retrieval. In a systematic review development, information extraction is a time-consuming process and can also be prone to so many errors. The main vision behind this is to categorize the documents using these information extraction systems. Normal machine learning algorithms occupy us with less accurate results than multi-pass NLP algorithms. These algorithms can also improve the performance of information extraction systems [9]. The next method so used in the classification is using support vector machines. Support vector machines have great importance in real-world problem solving and shown us some significant results. Next is the study of a new method of classification with a support vector machine which is an active learning process. The SVM understands the instances that are going to be used in the next stage and can request for it significantly. For this method, as it is described as active learning, it makes use of a training set that is already categorized to train the model that is being created. For this process, version space can be used. The center of the mass of a version space can be approximately found by using Bayes point machines. It is always better to minimize the labeled instances because labeling them is pretty much time taking process and can also be a bit costly. Make use of active learning to determine the predefined class labels rather than choosing random labels from a training set. Now text classification plays a very important role as in the modern world where the amount of unstructured data is huge and there is a need for a system so it can categorize this data and have easy access to it [10].

3 Implementation Details

3.1 Pre-processing

(a) **Tokenization**: It is a mechanism where a text or string is split into a set of tokens [11, 12]. Large text documents can be split into sentences, sentences that can be split into words, etc.
(b) **Stop Word Removal**: Once the text is tokenized, the next step is to remove the stop words which help in differentiating words with specific meaning from the ones which add a specific structure to the sentence.
(c) **Stemming**: This technique helps in reducing a word to a stem or base form [13, 14]. Snowball stemmer is a popular stemmer for the English Language.
(d) **TF-IDF**: Term frequency–Inverse Document Frequency suggests how important or relevant a word is for a document in the group of documents. TF-IDF for a word from a document is calculated by multiplying term frequency and inverse document frequency metrics [15]. The term frequency metric will give you the count of instances of a word in the document. Inverse document frequency specifies how important a word is to a document if it is appearing only for one particular document or how common the word is to all the documents if it appears in all the documents.

This section briefly explains the design of the proposed method. State diagrams indeed consist of a finite number of states and sometimes show a reasonable abstraction. All states eventually lead to a final state where the class label of a text document being given in input is predicted and helps to categorize the text document to which label it should belong to. The training data is first evaluated and represented using feature vectors for further help in the classification process and eventually create a model that can certainly predict the classes for text documents. As in Fig. 1, the process of classification begins with data gathering or acquisition and feeding it to the text classification algorithm which then trains itself according to the data being fed to it. Based on the training data being given to the algorithm, it attains prediction accuracy and thus classifies a text under a certain category. More training data are being fed for better accuracy.

The new classification task is carried out by preprocessing the dataset first. For the process of processing the raw text, the snowball stemmer (a string-processing language) is used. In the pre-processing stage, the unwanted text is cleaned by removing exclamations, punctuations, semi-colons, etc. Then, the next step is to start filtering the data by removal of such data which does not add much useful information. The parameter and the feature extraction are based on the algorithms. In this

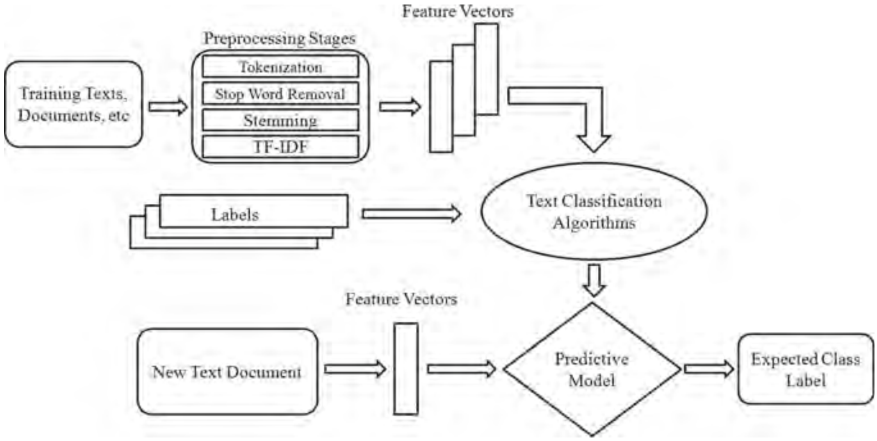


Fig. 1 Proposed architecture

work, a total of six machine learning algorithms are implemented and a comparative analysis is done on them using different evaluation metrics. The linear support vector machine (SVM) algorithm had shown the highest accuracy among the other algorithms.

The K-Nearest Neighbor (KNN) [16] algorithm classifies new data points depending on the similarity measures. The third algorithm which was implemented for the classification is the Rocchio algorithm [17]. This was developed using the vector space model and relevance feedback method. Then, bagging [18] and boosting [19, 20] algorithms were implemented for the classification. In bagging, the estimation uncertainty was reduced by producing additional data from the dataset to learn using variations of repetitions to generate multi-sets of the initial data. Boosting is an iterative procedure that fine-tunes the weight of an observation based on the final classification. Then, multinomial Naive-Bayes algorithm [21] was implemented which is based on the principle that every pair of features being classified is independent of each other. Finally, the linear SVM was implemented and found out that it is more accurate and has got better results of classification.

The support vector machine is a dominant and supervised learning algorithm commonly used for classification purposes. The SVM divides the classes by creating different hyperplanes. Then, the SVM [22] classifies the new samples using the hyperplanes. It was also considered how frequently a data was occurring or a word was occurring in the given text data. Only the important data was considered for the classification process. The important data is nothing but the data that came after the filtering stage and after the removal of useless and noisy data. Thus, a well-transformed data gave a good classification with the help of SVMs.

3.2 Dataset Used

The dataset used in this work was AG's News Topic Classification Dataset. This dataset consisted of four classes, namely Business, Sci/Tech, Sports, and World. There were 30,000 training samples and 1900 testing samples per class. The dataset consists of 120,000 training samples and 7600 testing samples. The files train.csv and test.csv contain all the training samples as comma-separated values. There are three columns in them, corresponding to class index (1–4), title, and description.

4 Evaluation Metrics

Initially, the pre-processing of the text was carried out, then feature extraction techniques were applied, and then the six machine learning algorithms were implemented on the test and training datasets. Now to figure out how effective these algorithms were while classifying, there was a need to assess them using a few evaluation metrics. For this purpose, the different metrics were used such as Accuracy, Recall, Precision, and F1 Score to exhibit the effectiveness of the algorithms.

(a) **Accuracy**: It is one of the simplest metrics. It is calculated as the number of appropriately predicted classes divided by the total number of predictions.

$$\text{Accuracy} = \frac{\text{NumberofCorrectPrediction}}{\text{TotalNumberofPredictionsMade}} \tag{1}$$

(b) **Recall**: It is another important metric and is calculated as the number of True Positives divided by the sum of True Positive and False Negative.

$$\text{Recall} = \frac{\text{TruePositive}}{(\text{TruePositives} + \text{FalseNegatives})} \tag{2}$$

(c) **Precision**: When the dataset has an imbalanced class distribution, then precision is the preferred metric compared to accuracy. It is calculated as the number of True Positives divided by the sum of True Positive and False Positive.

$$\text{Precision} = \frac{\text{TruePositives}}{(\text{TruePositives} + \text{FalsePositives})} \tag{3}$$

(d) **F1 Score**: This metric combines the Recall and Precision into one metric. It is the harmonic mean of recall and precision. It is defined as follows:

$$\text{F1Score} = 2 * \frac{1}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} \tag{4}$$

5 Results

5.1 Dataset Used

The performance of the machine learning algorithms is evaluated using four different evaluation metrics. The performance of all the algorithms is compared to the Accuracy, Recall, Precision, and F1 Score. The same is shown in Fig. 2 and in Table 1.

It was found out that the linear support vector machine performed better than all the other algorithms considered in this work. The support vector machine had consistently demonstrated 90% and above while predicting the correct labels. Multinomial Naïve Bayes comes next in the order when it comes to the prediction of correct labels. Boosting and Rocchio algorithms have demonstrated poor performance in predicting the true labels.

The training time and prediction time of all the algorithms are presented in Table 2. Boosting displayed the worst performance in terms of training time with 71.573469 s. While K-Nearest Neighbor algorithm is classified in very less time, i.e., 1.195145 s to train. Bagging displayed the worst performance in terms of prediction time with 157.886281 s. While the Naïve Bayes algorithm took very less time, i.e., 0.075812 s to predict.

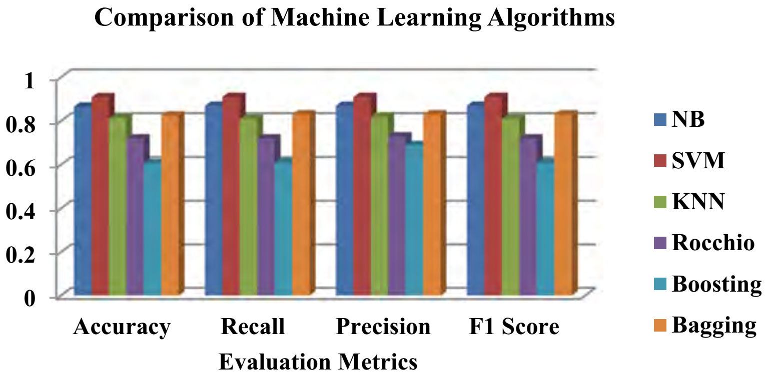


Fig. 2 Comparison of machine learning algorithms on various evaluation metrics

Table 1 Comparison of machine learning algorithms on different evaluation metrics

	NB	SVM	KNN	Rocchio	Boosting	Bagging
Accuracy	0.8662	**0.9085**	0.8124	0.7199	0.6075	0.8257
Recall	0.87	**0.91**	0.81	0.72	0.61	0.83
Precision	0.87	**0.91**	0.82	0.73	0.69	0.83
F1 Score	0.87	**0.91**	0.81	0.72	0.61	0.83

Table 2 Training time of different machine learning algorithms

	NB	SVM	KNN	Rocchio	Boosting	Bagging
Training time	1.482	10.974	1.195	1.226	71.573	1.520
Prediction time	0.075	0.586	16.167	0.0861	0.099	157.886

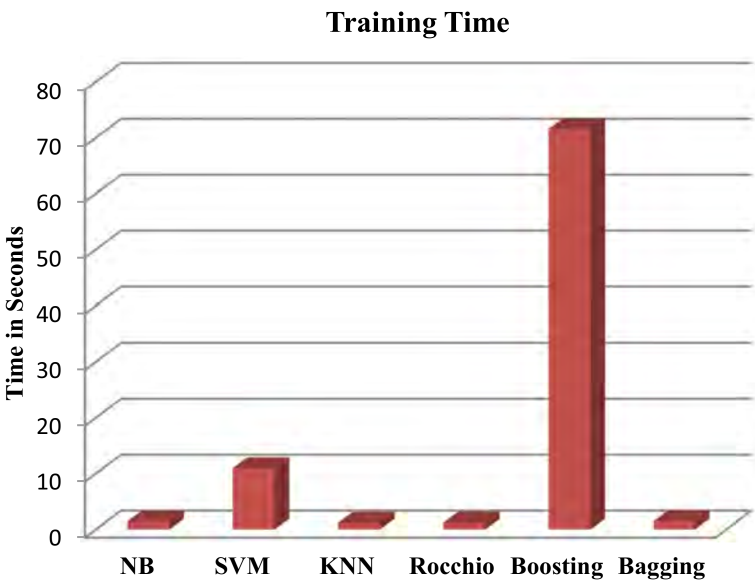


Fig. 3 Training time of different machine learning algorithms

In this work, the bar graphs are used for demonstrating the training time for all the algorithms considered. All the algorithms have demonstrated the short training time except for the boosting algorithm. The boosting algorithm is an iterative algorithm that keeps building strong learners. Due to this iterative nature of the boosting algorithm, it takes more time to train the model. The same has been shown in Fig. 3. The choice of algorithm for any real-time application depends on how fast the algorithm will predict the output. The more the prediction time, the less is the application of the algorithm in real-time applications. The prediction time for all the six machine learning algorithms is as shown in Fig. 4. The Naive Bayes algorithm took 0.075 s time to predict and bagging took 157.886 s to predict.

5.2 Confusion Matrix

The confusion matrixes are one more way to test the correctness of the model for any classification work. The matrix will give you the percentage of the correctly classified

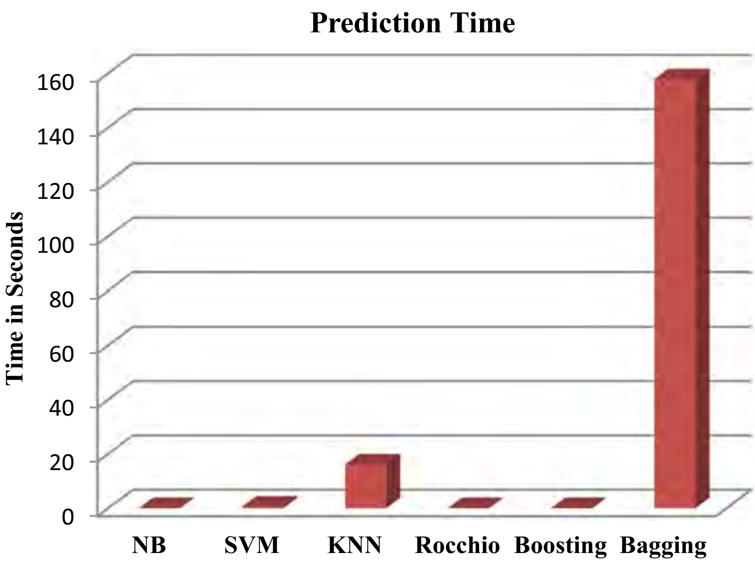
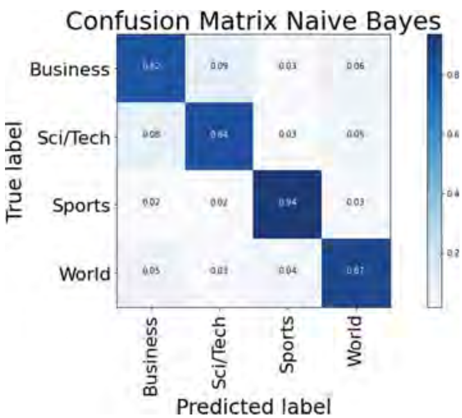


Fig. 4 Prediction time of different machine learning algorithms

data point. In the testing dataset, there are a total of 1900 rows per class. The Naive Bayes algorithm correctly predicts 1558 out of 1900 news topics with business labels. Similarly, the algorithm correctly predicts 1596, 1786, and 1653 news topics with Sci/Tech, Sports, and World labels, respectively. The same is shown as the confusion matrix in Fig. 5.

The linear support vector machine algorithm correctly predicts 1577 out of 1900 news topics with business labels. Similarly, the algorithm correctly predicts 1615, 1786, and 1672 news topics with Sci/Tech, Sports, and World labels, respectively. The same is shown in Fig. 6.

Fig. 5 Confusion matrix for multinomial Naïve Bayes



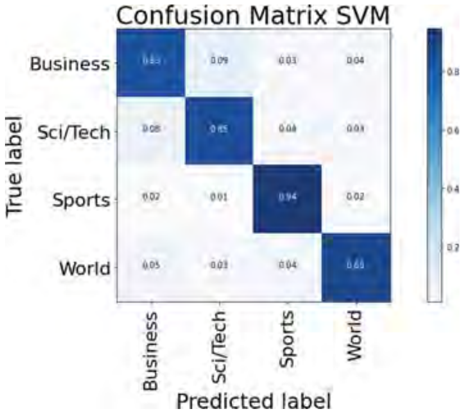


Fig. 6 Confusion matrix for linear support vector machine

The K-Nearest Neighbor algorithm correctly predicts 1539 out of 1900 news topics with business labels. Similarly, the algorithm correctly predicts 1501, 1653, and 1482 news topics with Sci/Tech, Sports, and World labels, respectively. The same is shown as the confusion matrix in Fig. 7. The Rocchio algorithm correctly predicts 1197 out of 1900 news topics with business labels. Compared to other algorithms, this algorithm has shown less accuracy due to the inconsistencies in the laws of classification between the various processes. Similarly, the algorithm correctly predicts 1387, 1596, and 1311 news topics with Sci/Tech, Sports, and World labels, respectively, and the same is shown in Fig. 8.

The Bagging algorithm correctly predicts 1501 out of 1900 news topics with business labels. The algorithm correctly predicts 1520, 1691, and 1539 news topics with Sci/Tech, Sports, and World labels, respectively, and the same is shown as the confusion matrix in Fig. 9. The boosting algorithm correctly predicts 893 out of 1900 news topics with business labels. The algorithm correctly predicts 969, 1672, and 1102 news topics with Sci/Tech, Sports, and World labels, respectively. This

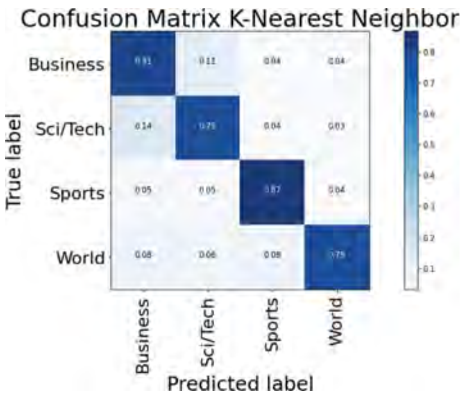


Fig. 7 Confusion matrix for K-Nearest Neighbor

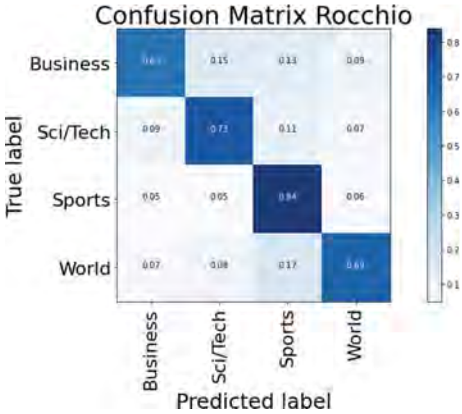


Fig. 8 Confusion matrix for Rocchio

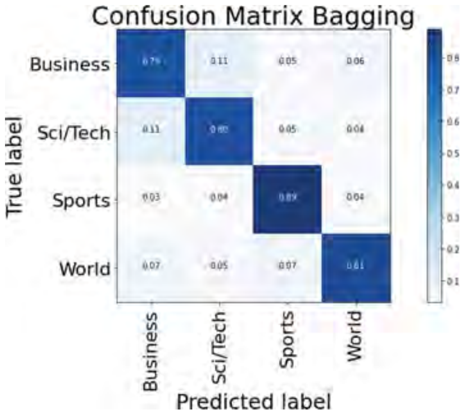


Fig. 9 Confusion matrix for bagging

algorithm displays the least accuracy compared to other algorithms, and the same is shown in Fig. 10.

5.3 Receiver Operating Characteristic (ROC)

A ROC curve or receiver operating characteristic curve is a graphical chart that indicates a binary classifier system's diagnostic capabilities as the threshold for discrimination is varied. Typically, the ROC curves are often used in binary classification to evaluate a classifier's performance. To extend the ROC curve and ROC region to multi-label classification, the output must be binarized. The ROC curve is formed by the plotting at different threshold settings of the True Positive rate (TPR) against the False Positive rate (FPR). There are four labels in the dataset, and the algorithms like Naive Bayes, SVM, KNN, bagging and boosting are implemented to generate ROC

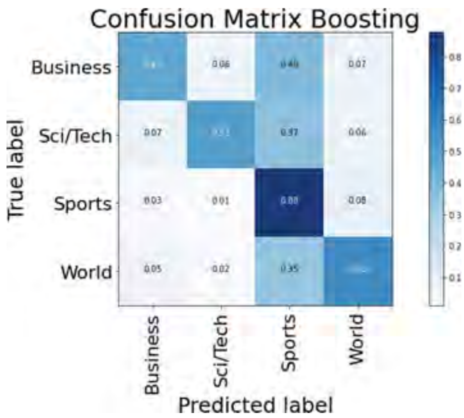


Fig. 10 Confusion matrix for boosting

for Business labels. To generate ROC for Rocchio, the algorithm has to compute a score. But by using the nearest centroid classifier, the algorithm displays the limitation to compute the score. Hence, the ROC curve for the Rocchio is not shown in Fig. 11.

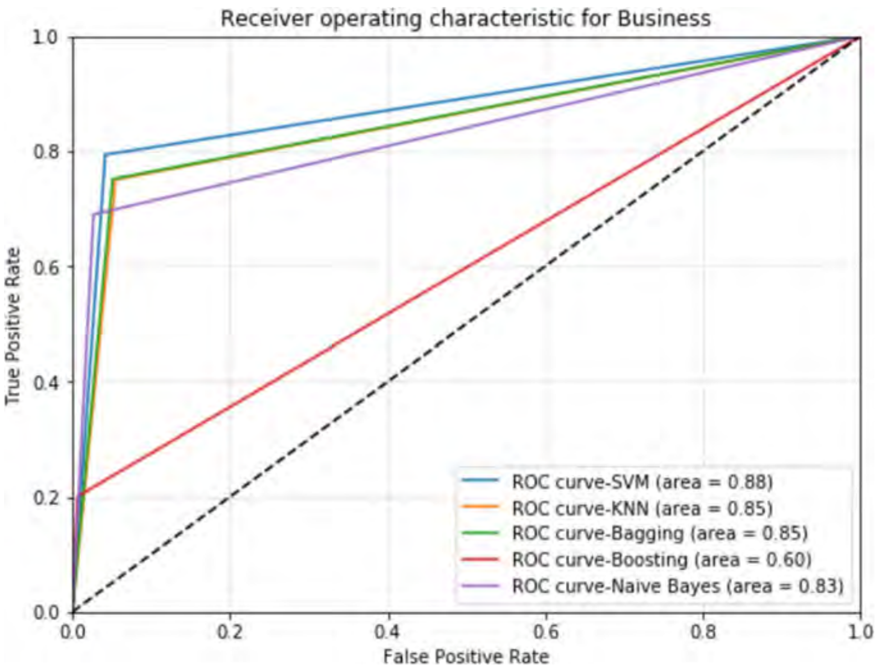


Fig. 11 ROC curve for machine learning algorithms for business label

6 Conclusion

The classification of News Topic was implemented using AG's News Topic dataset under predefined class labels. An approach was proposed to classify news texts comprising three steps. Text preprocessing was carried out first with a Snowball stemmer. The second step involved feature extraction based on the count vectorizer and TF-IDF vectorizer. Classification based on SVM, MNB, Rocchio, K-Nearest Neighbor, boosting, and bagging algorithms was the final step. For comparative analysis of these machine learning algorithms, different evaluation metrics like Accuracy, Precision, Recall, and F1 Score were used. The F1 Score was used to obtain a single value between precision and recall. As per the results, the SVM showed the highest accuracy of 91% among the algorithms considered, also requiring low training time and average prediction time. Multinomial Naïve Bayes showed the second-best accuracy among the algorithms, also requiring average training time and low prediction time. If the time requirements are the main focus, then the KNN would be efficient considering that the algorithm will deal with datasets of higher size. Linear SVM works poorly when the dataset is quite large and has more noise. The Naive Bayes classifier presumes highly about the structure of the data distribution. KNN is limited by data storage barriers for exploring nearest neighbors for large search problems. Boosting and bagging approaches have weaknesses such as the numerical complexity and lack of interpretability, which indicate that these models do not decide the function value. The Rocchio algorithm exhibits poor classification precision, and the linear combination is too uncomplicated for classification.

Acknowledgements This work was supported by Ramaiah Institute of Technology, Bangalore-560054, and Visvesvaraya Technological University, Jnana Sangama, Belagavi -590018.

References

1. Kowsari K, JafariMeimandi K, Heidarysafa M, Mendu S, Barnes L, Brown D (2019) Text classification algorithms: a survey. Information 10(4):150
2. Aggarwal CC, Zhai C (2012) A survey of text classification algorithms. In: Mining text data. Springer, Berlin/Heidelberg, Germany, pp 163–222
3. Aggarwal CC, Zhai CX (2012) Mining text data. Springer, Berlin/Heidelberg, Germany
4. Sulova S, Todoranova L, Penchev B, Nacheva R (2017) Using text mining to classify research papers. Int Multidisc Sci GeoConf Surv Geol Min Ecol Manag SGEM 17(21):647–654
5. McCallum A, Nigam K (1999) Text classification by bootstrapping with keywords, EM and shrinkage. In: Unsupervised learning in natural language processing
6. Scott S, Matwin S (1998) Text classification using WordNet hypernyms. In: Usage of WordNet in natural language processing systems
7. Menaka (2014) Text classification using keyword extraction technique
8. Nguyen TH, Shirai K (2013) Text classification of technical papers based on text segmentation. In: International conference on application of natural language to information systems. Springer, Berlin, Heidelberg, pp 278–284

9. Bui DDA, Del Fiol G, Jonnalagadda S (2016) PDF text classification to leverage information extraction from publication reports. J Biomed Inform 61:141–148
10. Tong S, Koller D (2001) Support vector machine active learning with applications to text classification. J Mach Learn Res 2:45–66
11. Verma T, Renu R, Gaur D (2014) Tokenization and filtering process in RapidMiner. Int J Appl Inf Syst 7(2):16–18
12. Aggarwal CC (2018) Machine learning for text. Springer International Publishing, Cham
13. Spirovski K, Stevanoska E, Kulakov A, Popeska Z, Velinov G (2018) Comparison of different model's performances in task of document classification. In: Proceedings of the 8th international conference on web intelligence, mining and semantics, pp 1–12
14. Singh J, Gupta V (2016) Text stemming: approaches, applications, and challenges. ACM Comput Surv (CSUR) 49(3):1–46
15. Jones KS (1972) A statistical interpretation of term specificity and its application in retrieval. J Doc
16. Jiang S, Pang G, Wu M, Kuang L (2012) An improved K-nearest-neighbor algorithm for text categorization. Expert Syst Appl 39(1):1503–1509
17. Korde V, Mahender CN (2012) Text classification and classifiers: a survey. Int J Artif Intell Appl 3(2):85
18. Breiman L (1996) Bagging predictors. Mach Learn 24(2):123–140
19. Freund Y (1992) An improved boosting algorithm and its implications on learning complexity. In: Proceedings of the fifth annual workshop on computational learning theory, pp 391–398
20. Bloehdorn S, Hotho A (2004) Boosting for text classification with semantic features. In: International workshop on knowledge discovery on the web. Springer, Berlin, Heidelberg, pp 149–166
21. Tong S, Koller D (2001) Support vector machine active learning with applications to text classification. J Mach Learn Res 2:45–66
22. Kim SB, Han KS, Rim HC, Myaeng SH (2006) Some effective techniques for Naive Bayes text classification. IEEE Trans Knowl Data Eng 18(11):1457–1466

Blockchain for Management of Natural Resources Using Energy Trading as a Platform

S. Thejaswini and K. R. Ranjitha

Abstract Nowadays, the need and necessity of renewable energy are one of the most important techniques which play a major role in our day-to-day life. Renewable energy is one of the best natural resources. In the earth, energy is available in the form of light, breeze, rain, surfs, and thermal. Resources are naturally renewable and it can be renovated. The proposed work mainly concentrates on blockchain technology in rainwater management because the limitation of water is one of the major reasons which is mainly caused by the climatic changes in the weather, floods, and amplified pollution. Blockchain helps to make a perfect decision by the assortment of data of water quantity and water quality. It empowers clean water projects to tap investors around the world with unparalleled breadth and ease. By giving a strong fireproof cabinet, safeguards transparency, and distributed ledger to document relations among the festivities, this branch of knowledge might deeply indicate a change in the way of rainwater resources which can be achieved and traded easily and one example is the use of the smart water grid (SWG) which is a two-way mechanism. One of the best is network transmission between the sensors and the mechanism that is constant and distant used for observing the water supply mechanism. The other one is the smart grid that can display many alternative parameters such as coercion, an improvement in the quality, rate of flow, a temperature which is more efficient toward this technology. By making use of this technology, a model based on an energy trading system helps the growth of industries from small-scale industries to large-scale industries that provides outstanding practice for energy efficiency measures is proposed. It mainly helps in monitoring the truth value of water. By the use of smart grid and sensors, the data can be fetched within a fraction of seconds. Finally, blockchain collects the exact record of water transaction and this makes as an account which helps in the management of water resource.

S. Thejaswini · K. R. Ranjitha (✉)
Department of CSE, Siddaganga Institute of Technology, Tumkur, Karnataka, India
e-mail: ranjithakr.1si18scs08@sit.ac.in

S. Thejaswini
e-mail: thejaswinis@sit.ac.in

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_36

Keywords Blockchain · Smart grid (SG) · Smart water grid (SWG) · Rain water · Machine learning (ML) · Intrusion detection system (IDS) · Deep learning (DL)

1 Introduction

India is the land of agriculture. Without water, there is no farming. Similarly, rainwater is a very important renewable energy asset because rainwater is always evergreen for the survival of plants and animals. It conveys new water to the earth's crust. If the scarcity of rainfall is less, then gradually the problem arises. Natural energy resources are solar power, wind power, small hydropower, biomass, etc., whereas renewable resources are natural resources which will regenerate naturally after consumption. But the time duration is high to refill or reload the resources. Hydropower generation is one of the examples of renewable energy which is able to generate few megawatt of power by converting the water potential into electrical energy with the help of turbines. Natural sources of water are categorized as rainwater, underground water, and surface water. Rainwater is another source which improves the water cycle above the surface and below the surface as groundwaters. Still large amount of water is present in a hidden state in different forms as glaciers and ice caps. This is how natural resources play an important role in the world. Rainwater is a mixture of electrolyte, abundant ions, sodium, and various minerals. Renewable energy resources are one of the main techniques which helps to convert from one form of energy to another form of the energy that is in the form of water cycle, water fades through solar energy and advances potential energy when the water hastens. This cycle of evaporation provides a mechanism for the conversion of solar energy into electrical energy. This rainwater can be stored as energy in many other forms. For example, in South Africa, the volume of wate**r** in storage is smaller in excellence due to the rise in inhabitants of a particular place and the area of the land where the water collects when it rains is surrounded by hills, truly generated by the modernization of urban, cutting of the plants and trees, damages of the rivers, the devastation of swamps, economic activity concerned with the processing of raw materials, obtaining coal or other minerals, from a mine, area of cultivation, usage of the work and unexpected poison of water pollution. This rain supports the blockchain technology in a greater way by providing a secure, transparent, and distributed ledger to record transactions between parties; blockchain-based technology [1] can primarily exchange the path of water utilities which can be succeeded and marketed easily. This smart grid technology helps in rainwater management in a greater way than by providing an electric grid which includes operations like smart meters, smart appliances, etc. It has a strategy called a smart grid policy which is systematized in Europe as Smart Grid European Technology Platform [1, 2]. Finally, this shows that rainwater is one of the renewable resources because rain is usually in a liquefied form which is molded as a result of condensation in the distinctive water vapors.

2 Renewable and Non-renewable Resources

Renewable energy sources are abundantly present in the nature such as solar, wind, hydro, thermal, and biomass energy and these energies are never depleting. The non-renewable resources such as coal, fossil fuel, natural gas, and petroleum do not successfully satisfy the needs of the people. But it is very efficient to use renewable energy resources than non-renewable. Nowadays, human beings are wasting water every day [2]. After all, they do not know the true value of water. And the rainfall has become reduced due to fewer forests, fewer trees, and deforestation which deliveries in the rainfall reduction. Water recycling is one of the best processes. This routine is almost done by the process of the flock and turns toward to form rain. The abrupt or immediate cause of a rainfall scarcity may be due to one or more issues including a non-appearance of available moisture in the atmosphere; large-scale collapsing which defeats convective activity; and the non-appearance or non-arrival of rainwater systems. Cutting down of trees will cut a significant amount of vapor from the atmosphere. This reduces the amount of rainfall received in the atmosphere.

3 Blockchain Technology in Rainwater Management

It is very hard to imagine our lives without water. Fine, 4 million people started their daily life's terrifying in Cape Town, South Africa, last year. Bureaucrats projected the town that there are only six periods of the time between the same dates in successive calendar months of investments is present. Through the shared process of doing something and some to a remarkable resultant rain in an area, they closely declined successively move away from water. The deadline for South Africa's water situation never ends. Weather conversion and deprived water system endures towards pay to water dearth, high-level pressures at activities, not to waste and quarter scale of the amount. In the next few decades, some of the major cities like Tokyo, Beijing, London, Mexico, Barcelona, and Istanbul are going to face day zero which represents zero level of water in the cities. Report states that in 2018 an average of 2.1 billion people is facing difficulty to access water for their regular needs.

The total volume of water is 1.386 billion km^3 loads of water on earth, in which 97% is of saline water and almost 2.5% remaining water is absorbed below the surface of the ground. The situation never goes roughly 0.4% to divide among one billion having 9 zeros human beings. It also concentrates on our whole commerce and cultivation sectors which provide a way for provisions toward adequate water to function. Weather reports change an industrialized organization at the last evolving world wannabe to devour like to walk or progress slowly into water scarceness. Expressing motion toward pillories is even worse, hominids are whispery celebrated their helplessness to share resources equally. Conflicts have been war nowadays for water during the legend. This sequence of actions continues until now. Blockchain technology can provide two ways of solutions which are as follows, water transparency and cooperation for smart water management.

3.1 Water Transparency

Blockchain is always safe, transparent [3], and scattered for concerning the record of events for communications across a formally constituted social gathering**.** If public blockchain is used for water and its excellence facts, then the evidence cannot be secreted or altered by the crooked behavior of politics, firms, or strong folks. The degree of excellence and the accent can be run through a conclusion in the period of growing water scarcity. Visualize ménages, industry, consumers, water supervisors, and policymakers use this evidence to conserve or use water. Activities of water supervisor are disposed to sleaze and devolved benefits, by giving the impression of being something perfect movement toward smart water system. Hong Kong has stated research toward water systems through WATERIG to identify rainwater through various points. These central parts of the wheel (hubs) are into contact with the water handling systems and they mainly focus on bids characteristics of vertical farming. Since a set of things is reorganized, humanities can come or bring to a resolution in which everything uses the blockchain to bulky account for their own colorless, transparent, odorless centers. WATERIG is chiefly contained with its colorless, odorless, and tasteless supply. This provides a ruinous study of Flint. About $28 million was given as a support by Flint's governor in the year 2016 to identify alternate to the natural calamity. Authorizing societies and slight firms to revenue toward rainwater encourages the market and networking system to a chance for employment. Rainwater under normal conditions gets wasted, but when societies provide openings to program their rainwater using the blockchain, this motivates the people toward recycling. The World Economic Forum recounted that rainwater might be cast-off to casual and influence factories or to create fertilizer. The municipal is contented and density is condensed on the system without special interventions. UN world record in 2017 reports that 80% of rainwater is wasted and does not treated for future use. As water becomes more insufficient for demand with weather change and growth of the world populace, societies will be further fortified to take benefit of their water gathering hubs.

3.2 Cooperation for Smart Water Management

Blockchain collaborating with the Internet of things (IoT) helps water systems cleverer, harmless, and more effectual. IoT centers for the operations of the water organization system, objects would look identical changed. The urban area water dispersal grid indicates the consequence of an imagined event of automotive and industrial radars will collect the records on the degree of the accent of water. Through online sensors communicate and investigate for seepages, channel torrents, and the state of making or being made impure by pollution or poisoning. But they swiftly send an alarm to indicate water executives. Based on the alerts, executives alter the pressure of water to avoid huge damage. In this view, Internet of things (IoT) provides better coverage for the betterment of every part of cities in a water interconnected system. The same knowledge is dropped in the backgrounds. Smart sensors interconnect one another to establish uncommon freshwater taken from the ground or surface. These observations help to see the water feasting visibly, choosing to select your performance and rescue currency. This alarms the mobile instantaneously for leaks even if the person is not at home and stops the water supply. In Internet of things (IoT) systems, all particular groups move from a lone view of safety intelligence. Entire systems conclusions are made here, producing it defenseless to equitation and maneuver. Just visualize, nowadays the water mechanization is erroneous in hands which gives rise to a threat to the national refuge. Blockchain accepts the undesirable failure in an Internet of things (IoT) scheme, by authorizing and securing the network. The combination of Internet of things and blockchain will be a fantastic job in water management.

4 Smart Grid in Rainwater Management

The smart grid is a shared term functional to an assembly of technologies, counting smart meters and meter reading apparatus, wide-area checking systems, lively line evaluation, electromagnetic autograph extent and investigation, time-of-use and physical stretch estimating tools, innovative shifts and cables, radio transportations technology, and numeral caring relays. A smartgrid [2] includes records. Each record has information that helps the clients to manage conveyance and detect incongruities and fiddling. Figure 1 represents the applications of the smart grid in various energy resources. The smart grid impression also embodies a trend in technology and the commerce prototypical for a convenience's connection with clients and extra sponsors. In the ancient model, a utility merely fashioned energy, transporting it over broadcast and spreading linkage to customers. In the new model, the helpfulness and its patrons become allies in dealing with the supply/demand connection. The smart grid provides the machinery to enable this connection. Using advanced metering infrastructure (AMI), smart meters have the power which regularly interconnects with the utility, electric efficacy can set charges which track costs that diverge by the time of day, swelling the price of power used during high demand periods, and decreasing it during truncated claim periods. Clients can retort to that material by altering their power weights, consequently, tumbling demand during peak usage periods, or buying utilizations that can retort to such information. For water utilities, there are foremost equivalents with power creation, transfer and buyer amenity, as well as significant variances. While there are major variances between water and electrical service, water conveniences involve many of the identical works and

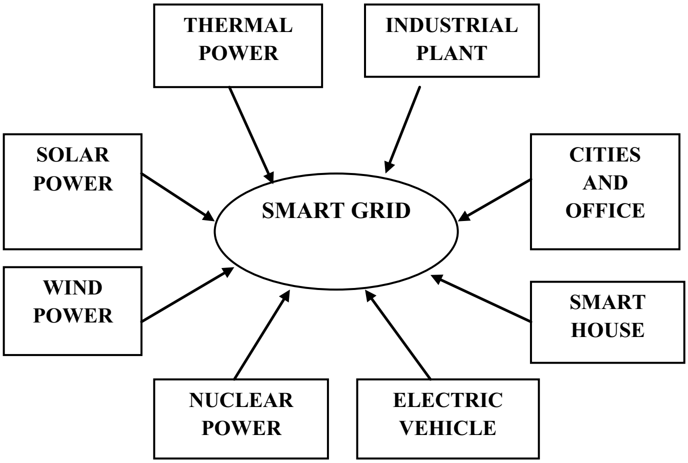


Fig. 1 Smart grid applications

topics that drive electric abilities, counting the necessity to grasp costs down while preserving system steadiness, powerful wealth necessities, the necessity for maintenance, conservation worries, and enrolment matters. Besides, water and wastewater benefits apply the foremost parts of their profitable resources on power. Mainly used for seriatim pump engines, which adds the measurement of responsive power. Vitality charges for water and wastewater can be tierce of a metropolis's energy bill. Portable water and drain water in the USA employ about US$4 billion a year on imparting the fluid energy, distribution, fetch, and sparkling water. Portable water and drain water conveniences justification for a predictable 75 billion KW/h of overall US power demand and tons are expected to grow by 20% over the next 15 years due to amplified populaces and more harsh guidelines. While water efficacies have more elasticity than electrical utilities in merchandise storage, many are theme to eventual anxieties that necessitate propelling when the cost of power is great. So, wastewater services stand to be jammed by the dawn of smart grid machinery.

4.1 User Feasting Outline Demonstrating a Leak

Water conveniences can use smart metering to display and device alternative day scattering vetoes. Water conveniences can launch periodic or regular period of usage or highest freight assessing, as well as tiered estimating for water resources. Such assessment is already being recognized in California and at other services around the world, expressly in thirstier zones. Smart metering offers the material to help clients retort to such worth gestures previous to getting the bill. In belief, a water value might send load resistor signs to irrigation managers and extra strategies, just as plug-in utilities would guide regulator signals to air conditioners and other tackle to shut off or dodge rotary on through certain periods. By using claim running techniques, a water efficacy could diminish the sum of water it needs to put into burden system storing (elevated tanks); this is tentatively equivalent to rechargeable efficacies dropping the sum of rotating fall-back that they obligate to retain operationally.

4.2 Need for Smart Grid in Water Utilities

Nowadays, a variety of operations and energy came into existence. So in rainwater management smart grid can be referred to as an apple because the one-half side is used to improve water infrastructure and the other half side is used for energy efficiency. Water conveniences will be a theme to more cultured assessing and claim side organization by their power providers. Smart grid retailers will adapt rechargeable smart grid machinery, particularly advanced metering infrastructure (AMI) and meter records organization systems, and arcade them to the water effectiveness arcade along with novel commercial models. Regulators will hearten water conveniences to assume power utility and commercial models. Their outcome will bear water convenience construction, dispersal, and client facility.

Smart grids are used to distribute the energy generated from renewable energy sources by connecting distributed energy sources. So this smart grid can be effectively used in renewable resources. One of the main reasons to include a smart system in the renewable resources is to gather information and communication technology (ICT) into the water network structure wherein which it increases the efficiencies of all elements in the water network. Smart grids support to manage the energy resources and reduces the power demand when any one of the energy sources gets dips by balancing the excess from other sources. Due to this energy management process, a smooth and efficient energy utilization could be achieved in an easy manner.

5 Major Problems of Water

In India, there are many water problems. Some of them are water conflict and water monopolies.

5.1 Water Conflict

Water conflict is a term relating to between republics, states, or assemblies over admittance to water incomes. A malicious process of water misrule devises fired radical disaster and controlled to humanity extremely feels that something cannot rely upon the central government's capability to resistor their water supply. This is an example of a clash between Karnataka and Tamil Nadu states in India for water conflict. Progressively thirstier region of the earth having the specified climatic conditions over the previous decades it has been igniting or cause to ignite again in the countries. Powerful combats generated gigantic objection as persons struggled for the scarcity of water which retains their folks active. The counterpole of the biosphere is the transformation of water by the Bolivian Government in 2000. This worsened into the "Water War of Cochabamba." After a broad conflict, the metropolis's water was re-nationalized and established the creation of lawful backing. However, on account of current water lack, some areas of the Bolivian countryside lost 90% of woodland during the drought in 25 years. For many societies and relatives, this is a matter of life and death. So the humble request is to save water and save lives which is most important for the biosphere.

5.2 Water Monopolies

Water is a natural resource that is controlled almost globally. Even a monopoly cannot charge any price it wishes. Once the patent expires, similar products will enter the market and the price of the product will fall. Monopolies are mostly measured to have numerous drawbacks (advanced value, fewer motivations to be effective).

By making use of rainwater as a renewable source of energy, the rainwater helps the energy storage device in a very efficient way. This rainwater comes in contact with the smart grid which includes records and bids to supply and investigate massive sizes of circulation and feeding information, offer it provides helpfulness work and clients to benefit them succeed transport and custom, and detect modifications and shifting. These routines a blockchain to provide transparency [3] and it will provide cooperations for the generation of smart meters.

5.3 Prediction

Machine learning models are widely used for prediction operations as a better prediction model provides better decisions based on analysis. In blockchain technology, machine learning models are used to predict the price of bitcoins. Valenkar et al. reported a prediction model using Bayesian regression analysis and random forest method to predict the bitcoin price through various parameters like size of the block, number of transactions, total number of bitcoins, and its trade volume. Normalization of trained dataset is achieved through box-cox, log and z-score. Recently, various price prediction models are evolved for predicting the values of Ripple, Dash, Ethereum, Litecoin, and Bitcoin cryptocurrencies.

6 Proposed Model

The energy trading system is proposed by using rainwater which is depicted in Fig. 2. Rainfall is one of the best factors of the water cycle and it is liable for leaving most of the renewed water on the earth. The situation offers opposite environments for several sorts of ecologies, as well as water for hydroelectric power plants and harvest irrigation. So this proposed model chooses the best renewable resources. When the air temperature falls to its dew point, water mist shrinks to form clouds. The air chills and the water mist shrinks, creating rain droplets. So this rainwater can be stored as energy. The concept of security of management of natural resources is proposed in which it includes air, water, soil, minerals, plants and animals are just used as a natural resource wherein which the humans use to produce the energy and make the things what the people use. The natural resources can be protected by cutting down on what you throw away, conserve water, plant a tree, choose sustainable, volunteer for cleans ups in your community and many more help to protect the natural resources. In this proposed model, a blockchain and smart grid for energy transactions using cryptocurrency is represented which is as in Fig. 2.

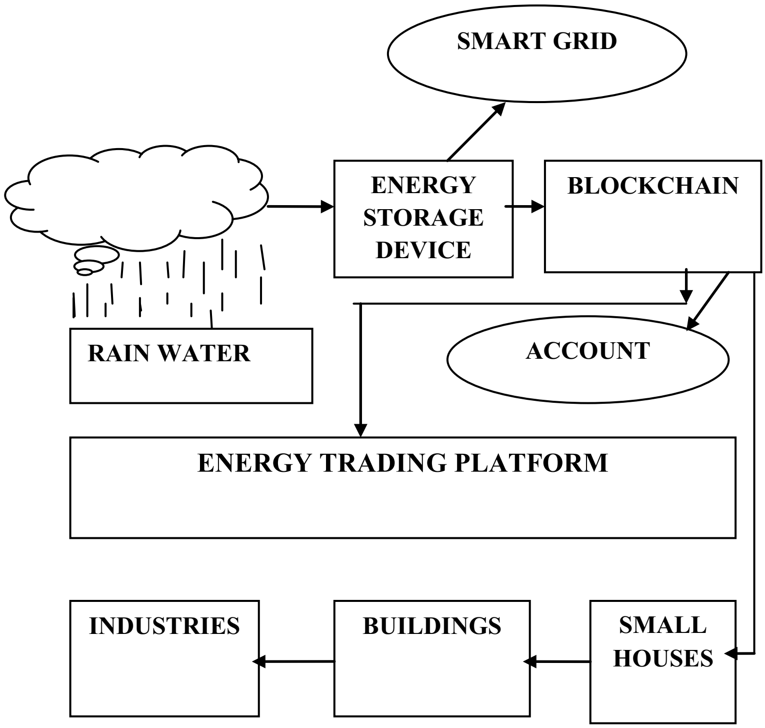


Fig. 2 Energy trading system using rainwater

It is more reliable for energy transactions. It uses deep learning (DL) and blockchain as a major resource [4, 5]. Byzantine fault tolerance is the feature of a distributed network to reach consensus even when some of the nodes in the network fail to respond or respond with incorrect information. In blockchain, byzantine fault tolerance can continue even if some of the nodes fail or act maliciously. Byzantine fault tolerance algorithm [6] is mainly used toward the yield of high structure speed. Blocks are created by hash functions and a dumpy signature. The proposed model consists of five ways infrastructures: buyer, seller, blockchain, intrusion detection system (IDS) [7], and machine learning [1].

The buyers trade the vitality with the seller. The buyer displays sufficient cryptocurrency which is required to purchase the sufficient asset. The buyer can neither be a person nor a building, so that the buyer could be situated in a dedicated network connecting devices in the home such as display, load control devices and finally smart applications or it can be a collection of smaller local area networks. Seller always gives rise to energy from renewable resources which include tidal energy, wind energy, biomass, etc. The seller can be a fellow citizen. Blockchain in rainwater management has more advantages for water scarcity. In this proposed system, the blockchain clearly collects the data of water quantity, and blocks are generated through its short signatures and hash functions. Based on recurrent neural network, this intrusion detection system works and identifies the fraud transactions and attacks. Further quality and efficiency can extend to make an account. This collection of data can be used to evaluate data analysis [8] which provides a platform for energy trading platforms. Energy storage device uses a smart grid because grid management is always cost saving which as a significance toward water management. Small houses, buildings, and industries are the major examples because it collects the data in an efficient way, and this helps the blockchain to list the records in a very efficient way. This unit is a distributed digital ledger [9], by which it is surrounded by all vigor relations in the smart grid system. Seller produces a vigor from the renewable energy resources which is used for itself and in the second step seller left the energy resources. Finally, the seller distributes per element fees which is a measure of power liveliness on the blockchain network. In peer-to-peer (P2P) [10], the energy trading system, the desirable energy buyer resolve aspect on the top of the printed worth. Frequently all the energy buyer drafts his justification which takes sufficient cryptocurrency stability for energy trading. Formerly, if the vigor claim of energy buyers and per-unit price coordinated their condition of energy at that instant, it pushes an acquisition appeal to the energy seller continuing in time the blockchain network. Once the successful authentication is performed, it provides option to buy energy resource through cryptocurrency. In this peer-to-peer (P2P) [10], energy trading system obtains the seller and buyer details in order to avoid hurting transactions. A spiteful block or lump can impress the peer-to-peer (P2P) system and contest for energy trading.

An energy trading platform has the ability to transform the energy from one form to another form. The proposed approach requests the buyer to consider the price per unit of energy and checks buyer account has sufficient credits for trade. Based on the demand of user and amount, a purchase request is forwarded to the service provider through blockchain network. Upon validation of buyer, seller provides resource to the user and accepts cryptocurrency. This P2P trade requires data analysis in order to identify the frequent buyer and seller, so that malicious transactions could be identified in the network as it will produce huge impact to the P2P system and energy trade. Blockchain is a decentralized security system that prone to attacks [11]. The important factor to be considered in blockchain is its consensus protocol, so that the blocks are controlled by few farms. Public blockchain has these features, whereas private blockchains are not affected by the attacks. Since each node is identified in this network through suitable consensus protocol, blockchain provides better control and efficiency in energy management. Also, real-time update and security features are provided through distributed ledger.

Intrusion detection system (IDS) [7] can figure out the untrustworthy connections and connection occurrences on the energy trading platform. The intrusion detection system mainly helps to alert the teams when leaks have defected. The intrusion detection system can also be set to detect water leaks, fire, an open window, or any other anomaly which may put property or personnel at risk. In intrusion detection network, intrusion detection systems can be observed in this proposed system because they easily identify the anomalies with the aim of catching the hackers before they do real damage to a network.

Machine learning (ML) [1] can be implemented to blockchain technology to make the petitions smarter. By making use of machine learning, distributed ledger can be maintained with high security. This machine learning supports to obtain suitable paths for data sharing. Various sources of data are collected and processed in the smart applications and blockchain will be a part of this smart application. Machine learning can be used to analyze the collected information and predicts the values and the collected information are stored in blockchain network. This avoids the issues such as redundant data, missing values, noise, and errors. Most of the data-related issues could be eliminated through blockchain. Integrating machine learning with blockchain classifies the essential data instead of analyzing the entire information in the dataset. Based on this, various approaches and applications could be developed in future for fraud detection.

7 Conclusion

Advancement in blockchain technology made a tremendous path in the technical environment. In this proposed work, detailed information on blockchain technology in the smart grid by using renewable resources that are rainwater. The architecture mainly represents energy trading system for data analysis using deep learning and blockchain technology. The energy trading system plays a major role which helps in the storage of energy using renewable resources. In this proposed work, rainwater is the best natural resource, this energy is stored in an energy storage device, and this energy storage device comes in contact with the smart grid because of smart meters and sensors which provide a dynamic way of monitoring the system. A distributed ledger is one of the main roots of the smart applications such as smart health care and smart grid. Making use of this smart grid (SG) and blockchain can evaluate the data analysis using an energy trading platform. This major activity helps the industries, buildings, and small houses through account so that it is greatly benefited in an easy manner. In future, this account can be replaced by the digital applications and since the paperwork concentrates mainly on the proposed model further this can be extended to implement it efficiently.

References

1. Tanwari S, Bhatiai Q, Patel P, Kumari A, Singh PK, Hong W-C (2016) Machine learning adoption in blockchain-based smart applications: the challenges, and a way forward
2. Bracciali A, Chatzigiannakis I, Vitaletti A (2019) Citizens vote to act: smart contracts for the management of water resources in smart cities
3. Abe R, Watanabe H, Ohashi S (2018) Storage protocol for securing blockchain transparency
4. Sun L, Chen X, Yang Z, Ke D, Meng, Qiu J, Cao Y (2019)Energy storage strategy in a non-agent energy trading platform: energy bank system
5. Wu H-T, Lu C-Y (2019) A deep learning application system based on blockchain technology for clicks-and-mortar businesses
6. Matsumoto Y, Kobayashi H (2010) A speculative byzantine algorithm for P2P system
7. Borkar A, Donode A,Kumari A (2017) A survey on intrusion detection system (IDS) and internal intrusion detection and protection system (IIDPS)
8. Nugent M, Lennon RG (2019) Blockchain for decentralized data analysis
9. Zhang K, Jacobsen H-A (2018) Towards depenadable, scalable and pervasive distributed ledgers with blockchains
10. Dorriab A, Hillab A, Kanherea S, Jurdakb R, Luoc F, Yang Z, Donga (2019) Peer-to-peer energy trade: a distributed private energy trading platform
11. Sivaganesan D (2019) Blockchain enabled internet of things. J Inf Technol Dig World

Machine Learning Model for Anomaly Detection in Big Data for Health Care Applications

M. G. Sharavana Kumar and V. R. Sarma Dhulipala

Abstract Recently, enormous amounts of data are increased by the essentials of data security and investigation for big data. Anomaly detection system monitors the data to analyze and detect any intrusion/misbehave in the network or devices. With the traditional detection techniques, it is highly complex to perform analysis and detection process. Techniques for big data are widely incorporated for anomaly detection system in achieving efficient data analysis. Hence, it is attempted to introduce fuzzy logic-based anomaly detection. The efficiency of the result has been obtained with greater accuracy with high performance with the added advantage of reduced training time.

Keywords Big data · Anomaly detection system · Machine learning · Fuzzy logic

1 Introduction

In the last decade, internet applications and communication have been tremendously developed in the medical field of information and communication technology. An application that falls in that category could generate huge generation, various difficult multifaceted structured data usually known as big data [1]. Since this era is full of automatic data collection, systematic measurements could not even know about the relevant information [2]. An example of this kind of application is E-Commerce; every transaction in E-Commerce includes buying, selling, and investing [3]. Hence, it generates health related data with complex structure to process and store which leads to the situation that traditional data storage techniques become fatal. This leads to handle the data and analytics it enables big data into the platform [4]. Nowadays, wearable sensors and systems along with people use the web with an exponential generation of the massive size of data, size of the data might be measured in exabyte

M. G. Sharavana Kumar · V. R. Sarma Dhulipala (✉)
University College of Engineering BIT—Campus, Anna University, Tiruchirappalli, India
e-mail: dvrsarma@aubit.edu.in

M. G. Sharavana Kumar
e-mail: mgsharavanakumar02@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_37

(EB) and petabytes (PB). It is expected that in 2025, there might be some yottabyte [5]. Roger Magoulas a researcher who coined the term big data for the first time to describe the propensity.

The expanded amount of health data provides a general overview of large databases through big data systems and computational processes [6]. Large data are semi-structured and unstructured data relative to standard data sets which tend to be more real-time analyzed [7]. Big data allows to define emerging principles and permits the fully grasp which is the secret ideals in greater detail, although it also raises new obstacles. Unanticipated massive cloud and internet computing (IoT) growth advanced the growth of data collected from medical devices [8]. Cloud infrastructure is now well established to provide for the collection of data and business connectivity for massive computer holdings [9]. In IoT, wearable medical sensors require the data to be obtained and data sent to the cloud to be analyzed further. These data usually surpass existing processing power and contributes to issues with the collection or reservation for the unique hardware and software infrastructures of large heterogeneous datasets.

Big data is difficult to store due to the rapidness and problematic to achieve and examine dynamically using the traditional storage techniques both software and hardware [10]. Big data has its characteristics of high volume and velocity, data variety, which needs emerging techniques to handle it. Anomaly detection system (ADS) monitoring is both hardware and software monitoring that analyze and detect data by any form of attack in system or network. Existing old-style anomaly finding techniques is complex and not up to the level of efficient, when it happens to be big data [11]. Time constraint is the major thing to deal with, since existing system is prone to harm due to time delay, there is a need to introduce the efficient techniques to store and analyze the data in ADS. Thus, the anomaly detection calculation time and alerting period can be in actual/on time.

ADS has three methods for detecting attacks namely, signature, identification of anomalies, and combination detection. Signature-based uncovering is one of the effective methods to detect in such a way that preloaded in the ADS database, thus it will be accurate enough to identify the attempt of an anomaly in a known type of attack [12]. The disadvantage of this detection mechanism only pre-defined attacks can be detected it cannot detect the new type of attacks, since its signature not in the anomaly database to increase the efficiency of the detection the database needs to be updated frequently/periodically [13]. Even the updated database effectiveness is not up to the level, since it has a high false-positive rate.

The hybrid-detection mechanism is the combination of two or more detection mechanism of anomaly detection enables the hybrid-detection technique. To overcome the pitfalls in the mono detection mechanism, it is preferred to use a hybrid-detection mechanism [14]. To handle the big data along with machine learning is one of the effective methods ADS is for anomaly detection to reduce the false-positive rates and to increase the effectiveness of the detection accuracy.

2 Methods and Background Studies

Machine learning (ML) is an emerging field in the research area that primarily focused on the theory and performance of the algorithms [15]. Those algorithms are highly interdisciplinary filed such as artificial intelligence, optimization, data, knowledge, and commerce. Almost any technological area, which has already had a significant influence on study and development, has already covered the way the data are treated in a broad variety of applications, Machine learning, and artificial intelligence. This was used for different issues, including engines. Machine learning is generally divided into three categories: supervised learning, uncontrolled learning, and enhanced learning as shown in Table 1.

The huge amount of data in real-time has hindered network analysts' success in scaling up the plentiful data rate. It is necessary to generate actual and near-in realistic, and not from the periodic log data available on the network protection review [16]. The current anomaly detection technologies, however, are unable to store, identify, and report on the real-time or near real-time data [17]. Additionally, more and more different attack styles are discovered every day in networks, but the existing detection systems have struggled to identify these assaults due to the immense number, size, range for study.

In the era of big data, the network security issues have been many and still not covered, particularly in the areas of modeling, data analysis, and anomaly detection [18]. This includes the fields in which the data process, banking, education, and transport are still not addressed. However, conventional surveillance systems that exist do not manage large data and are thus unable to constantly track the network architecture and recognize irregularities and vulnerabilities are necessary for rapid analysis of the data generated to recognize possible risks to the network [19]. Thus, a fusion of deep learning algorithms combined with large data technology is necessary to overcome the holes in the bottlenecks of current techniques and to effectively

Table 1 Methods and characteristic of big data

Types	Methods	Characteristic norm	Name of the algorithm	Types of data	Approach
Supervised learning	Classification & Regression	Computational classifiers Statistical classifier	SVM Naïve Bays	Labeled data	Maps the labeled inputs to the known outputs
Unsupervised learning	Cluster formation & prediction	Parametric classifier	K-means neural networks	Unlabeled data	Understands patterns & discovers the output
Reinforcement learning	Decision making	Modeling & non-modeling	Adaptive learning based	No Predefined data	Trail & error method

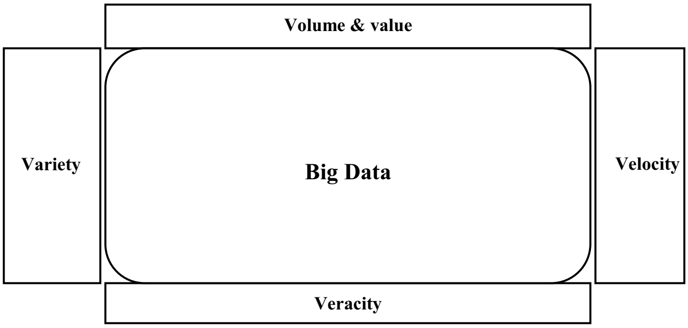


Fig. 1 Key issues in big data

process large real-time data to identify anomalies. Machine learning knowledge helps to detect using algorithms by analyzing the data collected [20]. With the help of back propagation algorithms to train the RNN, the work projected in such a way to improve the system reliability by using the SVM algorithm, thus the robust prediction can be ensured for different applications [21]. Besides, various big data technologies can contribute to the real-time and almost real-time processing and streaming of enormous amounts of network data. There are many issues faced by machine learning techniques such as large-scale data, various types of data, high-speed streaming data (including multimedia data), uncertain and incomplete data, and low-value density data as shown in Fig. 1.

The various terms referred to in Fig. 1 are namely, volume—large scale, variety—heterogeneous, non-linear, velocity—real-time and high-speed stream, veracity—uncertain, value—low value, and diverse.

3 Challenges of Anomaly Detection System (ADS)

There are challenges for ADS devices that are deployed on a large network is in the ADS components communication in and around of the sub-networks, in some case finished firewalls and entries also. Different amounts of the system and different network devices their data formats which are usually in unique and different protocols with different data formats. Another major challenge for ADS in a huge network is effective traffic monitoring. Network anomaly detection system (ADS) uses many different components that are speckled through a system, if those components are not placed in a specified way, many attacks may intrude in network anomaly detection system (NDS) as shown in Fig. 2.

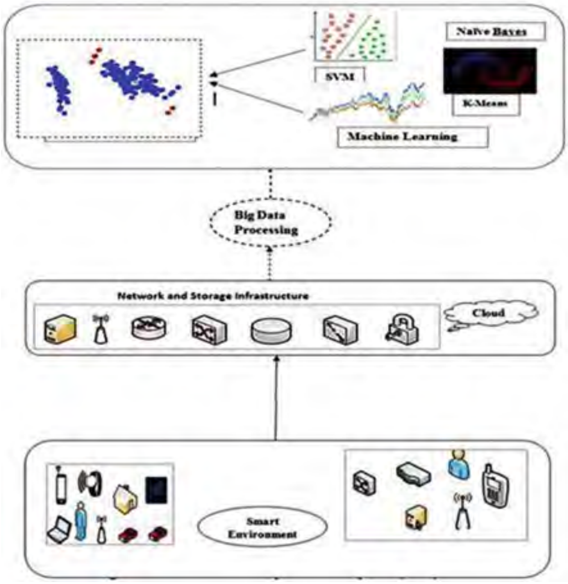


Fig. 2 Flow chart of anomaly detection system (ADS)

3.1 Fuzzy-Based Detection Factors

The concentrated logic detection block makes choices concerning healthcare and the presence of anomaly and therefore proposes a suitable fuzzy level implementation. It is carried out based on human logic, which imitates the same strategy of reasoning. The intensity of the system and progression of the system are based not on personal health, but also on the complete detection of an anomaly. In the fuzzy systematic evaluation logic, there are overlapping fluctuating changes in the interval which mimic the way people perceive modification along with reasoning. Fuzzy logic describes the way of making use of fuzzy if/then or or/and rule of imprecise dependencies and commands [22].

The linguistic variable, membership function, and the speech are the concepts that play a vital role in the application of the systems. A fuzzy set is a set of real numbers with part of the set. The layers of the group are calculated by a value of membership function, absolute non-members in 0 functions in the system, and half of the community by a value of 0 to 1. The MF is a curve that sets the form of benefit in a matrix for each point inside the input space. The member's role is for each input to find the standard range, with MF specifying a set.

Consider the scheme under which a vector in a discourse of anomaly and is an actual number in which denotes the fuzzy collection described. The member function associated with these conditions is a function that maps into [0, 1] and provides the function rank of anomaly detection. The MF used in the fuzzy logic system is triangular, provided that it is within [0, 1]. It then defines the function of the deltoid anomaly function in which the parameters are on the input the fuzzy logic works to achieve the crisp output α.

A multiple input and one output structure are built for the study of health performance. The model is designed with several entrances and every if-then rule is defined by fuzzy input sets and output set of parameters.
The linguistic variables are defined by the value of the output. Accordance with the data of the fuzzy system, the de-fuzzing framework is necessary, to get the crisp out of the method, since the output of each assumed fuzzy group is independently measured by a characteristic value of classification.

3.2 Dataset

KDD99 dataset is widely used for anomaly detection system performance evaluation. The dataset consists of nearly 5 million records including normal and abnormal events that are already collected to differentiate the anomalies synthetic anomalies has been injected and tested. İt is taken 5000 samples of data to check and validate among the huge number of records.

3.3 Model Classification SVM

One of the supervised learning methods in support vector machine (SVM), it examines data for organization and reversion. Usually, SVM categorizes data into divergent lessons. In the binary case, this method does classification in such a way that given data into a couple of classes with the help of a linear hyperplane. Consider a scenario if suppose a vector x exists, scalar q and r as shown in Eqs 1 and 2.

$$x^{\mathrm{T}} x + q \geq 1 \tag{1}$$

$$x^{\mathrm{T}} x + r \leq 1 \tag{2}$$

where $x-$ the weight vector is q and $r-$ is the bias value

SVM facilitates to reduce the error rate in classification, maximizing the margin. It achieves better execution when maximizing the margin involving the vectors of the two classes, which could result in the classifier in the rate of maximum margin. Equation (3) is used to find the best possible separating hyper lance

$$w = x, y = b, w = x, x = a, c = b, a = q, b = r \tag{3}$$

$$a = q, b = r, \tag{4}$$

$$\min \frac{1}{2} ||x||^2 \tag{5}$$

Subject to

$$b_i(xa_i + c) \geq 1; \quad \forall(a_i, b_i) \in D \tag{6}$$

In order to reduce the possessions of outliers and errors of misclassification, soft margin has been used in SVM. It enables the show in Eq (7) for the non-negative slack variable which is used to understand the trade-off between the margin and misclassification error, which usually end-user entity defined.

$$\min \frac{1}{2} ||x||^2 + C \sum_{i=1}^{O} \mathcal{E}_i \tag{7}$$

Subject to

$$y_i(wx_i + b) \geq 1 - \mathcal{E}_i \quad \mathcal{E}_i \geq 0, \quad 1 = 1 \ldots N \tag{8}$$

where $\mathcal{E}_i$ signify slack variable, $C-$ indicate a consequence parameter usually reins the trade-off between the cost of classification margin and errors of misclassification.

3.4 Genetic Operators

Efficient optimization takes place when genetic operators are taken into account, which includes basic genetic operators. In order to optimize the parameters, an efficient strategy needs to be incorporated. The genetic algorithm has the adaptability to fine-tune in the dynamic fashion of altering that operator with respect to the evolutionary situation in the population, in order to retain the variety of the populace and to connect early and enhance the exploration.

3.5 Selection Operator

Sort comparative technique has been used to verify the value that is in the fitness of every individual. The fitness value is proportional when each probability of selection and the corresponding value of fitness is proportional. If the group size is n, if the adapt degrees of the individual is f_i, probability pi which can be chosen with the help of the following formula.

$$Pi = \frac{f_i}{\sum_{i=1}^{N} f_i} \tag{9}$$

Cross operator intended of a grouping of linear vectors combined to numerical the value of the character. In case two individuals $T_b T_c$ cross, then the result of spring by an entity in the following formula.

$$T_b^{'} = \lambda T_b + (1 - \lambda) T_c \tag{10}$$

$$T_C^{'} = \lambda T_c + (1 - \lambda) T_b \tag{11}$$

where λ is a random number between (0, 1)

4 Role of Mutation

Using the probability Pm mutation operator randomly selects a value then it adds the value of entering point, it applies to all offspring chromosomes. Whenever GA into the technique, there are two considerations to take care of, primarily to prevent the premature phenomenon of population diversity in the initial mutation operation that can be large. Secondary consideration, a secondary consideration is the near-optimal solution neighborhood, how to make reduced the variations operator. To ensure the optimal solution, random search capability and acceleration of convergence are to be ensured. Hence, the adaptive mutation probability in the following formula solve the above-said problems.

$$Pm = \frac{\exp(-1.5Xt/2)}{\text{pop\_ size}X\sqrt{L}} \tag{12}$$

The ADS neural network anomaly identification parameters are the number of neurons in cached layers, learning duration, epochal amount, and the momentum of learning as shown in Fig. 3. Depending on the issue, the number of secret layers may be increased. An ADS with three secret layers is therefore sufficient to map every continuous feature to satisfy the necessary complexity by inserting a certain number of neurons. This is why the amount of secret layers used in our experiments is 3. Training pace and momentum are two main metrics for the effective training of the ADS network.

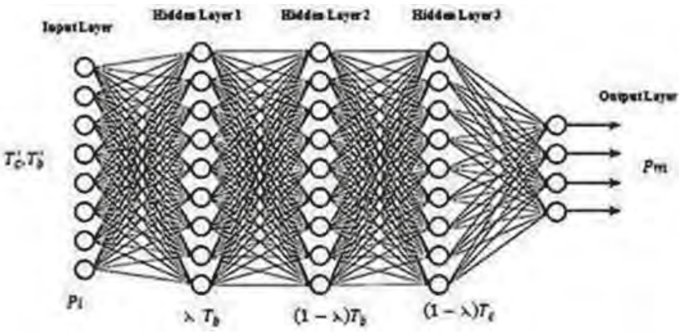


Fig. 3 ADS neural network anomaly and parameter

5 Result and Discussion

Changing parameters for neural networks in hidden layers is the distribution of neurons: [12,… 35] The first step is to be considered as a step with phase 3 parameter for the pace of learning: [0.03, 0.06, 0.08, 0.1, 0.5], for the number of epochs: Only one parameter change principle with fixation of all other parameters is created to search neural structure parameters.

Big data is used for the neural network detection of anomalies. Results are achieved by changing network parameters similar to the results of ADS neural networks with help of MATLAB software environment. The parameters used are the number of neurons, rate of education, number of epochs, and momentum in hidden layers. Big data's initial parameters are identical to ADS' initial neural network (the secret layer of neurons is 15; hidden layers 3 remain, and so on). Some of the improvements are still based on the same idea. The epoch variations [60… 300] and the amount of momentum [1.0… 0.5].

The network parameter modifications were collected for comparison of data-driven mapping neural networks (SVM, EB, and ADS). The problem is the global optimization function in which each goal is described as a shift in the given neural network parameter, the optimum forehead is present in Table 2. ForSVM, the distribution coefficient is just one target in building a neural network structural search space. All the scales (Accuracy, Performance, and Efficiency) have been measured as shown in Fig. 4. Instead of the neural network ideal parameters, the more complicated approach called the neural network architectures are considered based on the investigation.

The ADS provides great outcomes for all output indicators for both chosen pieces. Namely, for the chosen best parameters of EB modeling, the total precision scales (Accuracy, Performance, and Efficiency) by 98.82% as shown in Fig. 5. The ADS produces marginally better results (92–99%) compared to the neural SVM network (84–96%) with test information.

In comparison with the section (92–99.32%), neural network types (SVM and EB) give better rating results for the steam drums section (88–99%) as shown in

Table 2 Comparison of accuracy, performance, and efficiency

S.NO	SVM			EB			**ADS**		
	Accuracy	Performance	Efficiency	Accuracy	Performance	Efficiency	**Accuracy**	**Performance**	**Efficiency**
1	87.25	90.12	82.36	91.23	92.36	94.52	95.62	96.12	98.25
2	88.62	91.23	84.62	93.25	93.75	95.12	96.12	97.62	97.36
3	89.32	94.62	88.62	93.75	91.62	94.62	97.65	98.36	98.61
4	91.62	96.52	92.65	94.62	94.65	96.72	98.25	98.98	99.65
5	94.62	95.62	94.67	95.62	96.75	97.12	99.42	99.12	99.87

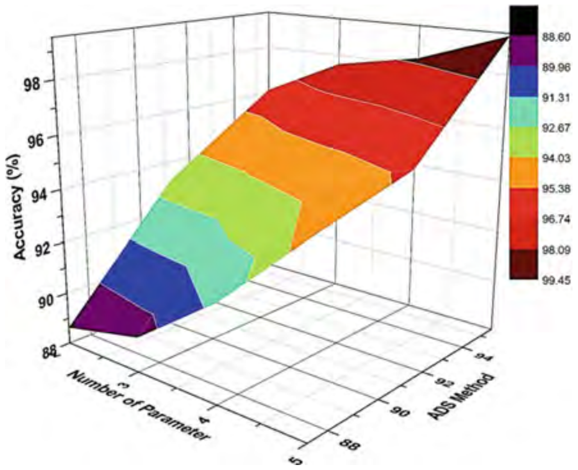


Fig. 4 Comparison of accuracy in ADS method

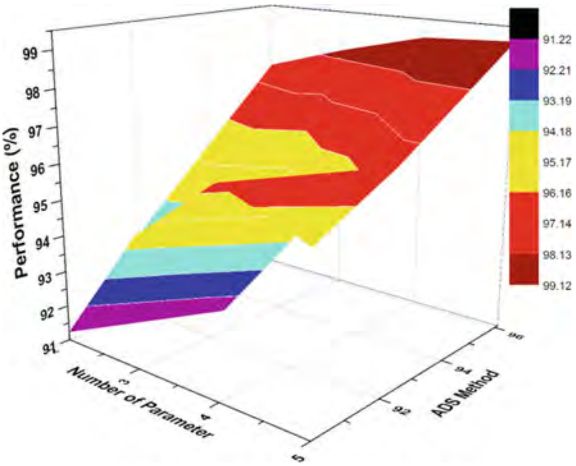


Fig. 5 Comparison of performance in ADS method

Fig. 6. For the steam drum portion, the accuracy and performance values are higher than the segment in specific parameters. It is anticipated mostly as there is a greater gap between data that is anomalous and data that is usual for steam drums. That is predicted.

The identification of abnormalities is a significant issue, which has been studied in numerous fields of study and implementation. Thanks to continual use, manufacturing systems become affected and this should be identified as early as possible to prevent losses. This is also important to provide a good device anomaly detector due to how they perform analysis, and the detection process method is one of the most complicated fluid structures that will still work appropriately at a minimal expense. Many fatalities have been attributed to heart attacks for the last ten years. Many scientists have developed many techniques of data-mining to diagnose these cardiac

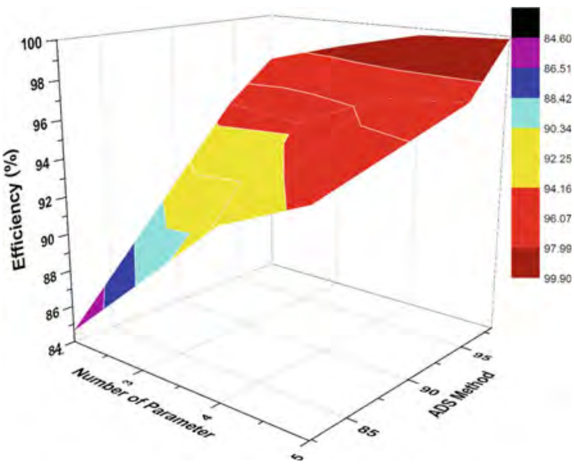


Fig. 6 Comparison of efficiency in ADS method

disorders for medical practitioners. When coping with cardiovascular disorders, the most popular methods for machine learning are ADS. Within this article, the procedure is used to treat cardiac disease, which incorporates k-means clustering and ADS. The findings of the trial indicate better specificity in the treatment of heart failure. Our process is 99.82 percent accurate; it is much stronger than many older methods.

6 Conclusion

With the conventional detection procedures, it is exceptionally complex to execute the investigation and detection process. Procedures for big data are broadly consolidated for a neural network framework in the discovery of anomalies. Machine learning (ML) is a growing field in the research of regions that fundamentally centered around the theory and performance of the algorithms. KDD99 dataset is generally utilized for the anomaly detection system execution assessment. One of the supervised learning methods in support vector machine (SVM), inspects information for identification and inversion. This work analyzes different models to investigate anomaly detection choices in selected parts of health care. Experimental results demonstrate that early anomaly detection in neural networks is highly efficient, with 99.52 percent overall precision (Accuracy, Performance, and Efficiency) modeling parameters.

References

1. Habeeb RA, Nasaruddin F, Gani A, Hashem IA, Ahmed E, Imran M (2019) Real-time big data processing for anomaly detection: a survey. Int J Inf Manage 1(45):289–307 (Apr)

2. Abdelghafar S, Darwish A, Hassanien AE (2020) Intelligent health monitoring systems for space missions based on data mining techniques. In: Machine learning and data mining in aerospace technology 2020. Springer, Cham, pp 65–78
3. Kumar MS, Dhulipala VS, Baskar S (2020 Jun) Fuzzy unordered rule induction algorithm based classification for reliable communication using wearable computing devices in healthcare. J Ambient Intell Humaniz Comput 24:1–2 (Jun)
4. Yuan G, Zhang C, Hu S, Guo J, Wang X (2019) Big data based bridge anomaly detection and situational awareness. In: 2019 Chinese automation congress (CAC) 2019 Nov 22. IEEE, pp 3864–3868
5. Garg S, Kaur K, Kumar N, Rodrigues JJ (2019 Jan 16) Hybrid deep-learning-based anomaly detection scheme for suspicious flow detection in SDN: a social multimedia perspective. IEEE Trans Multimedia 21(3):566–78 (Jan 16)
6. García NM (2019) Multi-agent system for anomaly detection in Industry 4.0 using Machine Learning techniques. ADCAIJ: Adv Distrib Comput Artif Intell J 8(4):33–40
7. Haripriya AP, Kulothungan K (2019 Dec 1) Secure-MQTT: an efficient fuzzy logic-based approach to detect DoS attack in MQTT protocol for internet of things. EURASIP J Wirel Commun Netw 2019(1):90 (Dec 1)
8. Abdelaziz A, Salama AS, Riad AM, Mahmoud AN (2019) A machine learning model for predicting of chronic kidney disease based internet of things and cloud computing in smart cities. In: Security in smart cities: models, applications, and challenges 2019. Springer, Cham, pp 93–114
9. Parvin P, Chessa S, Kaptein M, Paternò F (2019 Jan 1) Personalized real-time anomaly detection and health feedback for older adults. J Ambient Intell Smart Environ 11(5):453–69 (Jan 1)
10. de Sousa LD, Giommi L, Tisbeni SR, Viola F, Martelli B, Bonacorsi D (2019) Big data analysis for predictive maintenance at the INFN-CNAF data center using machine learning approaches. In: Conf of Open Innovations Association (FRUCT), Helsinki 2019, pp 448–451
11. Ahmed H, Younis EM, Hendawi A, Ali AA (2019) Heart disease identification from patients' social posts, machine learning solution on Spark. Future Gener Comput Syst (Oct 5)
12. Ramasamy B, Hameed AZ (2019 Jul 1) Classification of healthcare data using hybridised fuzzy and convolutional neural network. Healthc Technol Lett 6(3):59–63 (Jul 1)
13. Goyal KK, Paray AH (2019) A survey of different approaches of machine learning in healthcare management system. Int J Adv Netw Appl 11(03):4270–6
14. Larriva-Novo X, Vega-Barbas M, Villagrá VA, Rivera D, Álvarez-Campana M, Berrocal J (2020 Jan) Efficient distributed preprocessing model for machine learning-based anomaly detection over large-scale cybersecurity datasets. Appl Sci 10(10):3430 (Jan)
15. Yousefi S, Derakhshan F, Karimipour H (2020) Applications of big data analytics and machine learning in the internet of things. In: Handbook of big data privacy 2020. Springer, Cham, pp 77–108
16. Abd Ghani MK, Mohammed MA, Arunkumar N, Mostafa SA, Ibrahim DA, Abdullah MK, Jaber MM, Abdulhay E, Ramirez-Gonzalez G, Burhanuddin MA (2020) Decision-level fusion scheme for nasopharyngeal carcinoma identification using machine learning techniques. Neural Comput Appl 32(3):625–38 (Feb 1)
17. Razzak MI, Imran M, Xu G (2019 Mar) Big data analytics for preventive medicine. Neural Comput Appl 16:1–35 (Mar)
18. Selvaraj A, Patan R, Gandomi AH, Deverajan GG, Pushparaj M (2019 Nov) Optimal virtual machine selection for anomaly detection using a swarm intelligence approach. Appl Soft Comput 1(84):105686 (Nov)
19. Qi B, Zhang P, Rong Z, Li C (2020 Oct) Differentiated warning rule of power transformer health status based on big data mining. Int J Electr Power Energy Syst 1(121):106150 (Oct)
20. Bayrak EA, Kirci P (2019) Intelligent big data analytics in health. In: Early detection of neurological disorders using machine learning systems 2019. IGI Global, pp 252–291

21. Raj JS, Ananthi JV (2019) Recurrent neural networks and nonlinear prediction in support vector machines. J Soft Comput Paradigm (JSCP) 1(01):33–40
22. Kumar MS, SarmaDhulipala VR (2020) Fuzzy allocation model for health care data management on IoT assisted wearable sensor platform. Measurement 108249

Pilot Decontamination Algorithm with Iterative Weighted Graph Coloring Scheme for Multi-cell MIMO System in 5G Applications

V. Baranidharan, S. Karthikeyan, R. Hariharan, T. Mugunthan, and S. Vhivek

Abstract A multi-cell massive multi-input multi-output (MIMO) technology is the most important and promising technology in 5G wireless systems. The base stations in this 5G system are always provided with an abundant number of antennas. These base station antennas are simultaneously served more data to the set of users. Due to this more number of high-speed users, there will be high data traffic, the antennas suffer different types of pilot contamination in the various adjacent cells. A modified EIWG-WGC-based pilot contamination scheme is proposed to mitigate and avoid the unnecessary pilot contamination from the near adjacent cells. In the first phase, in order to find the potential PC, relationships among the users will be calculated by the construction of a modified edge-weighted graph coloring algorithm (EWIG). This graph coloring helps to find whether the two users are connected in the different cells by the weighted edge. This also indicates potential PC available and its strength for reuse of the same pilot. After the construction EWIG, the WGC-based pilot contamination method is widely used to segregate the pilot by assigning colors and finding the vertex of each user. This segregation helps to mitigate the PC by assigning weights based on different pilots. This proposed scheme is simulated by using MATLAB software. The simulated results show that this proposed scheme (EWIG-based WGC-PC) outperforms the existing system.

V. Baranidharan · S. Karthikeyan · R. Hariharan (✉) · T. Mugunthan · S. Vhivek
Department of Electronics and Communication Engineering, Bannari Amman Institute of Technology, Sathy, India
e-mail: hariharanr.ec16@bitsathy.ac.in

V. Baranidharan
e-mail: baranidhar@hotmail.com

S. Karthikeyan
e-mail: karthikeyan@bitsathy.ac.in

T. Mugunthan
e-mail: mugundhan.ec16@bitsathy.ac.in

S. Vhivek
e-mail: vhivek.ec16@bitsathy.ac.in

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_38

Keywords Edge-weighted graph coloring algorithm (EWIG) · Multi-input multi-output (MIMO) · Weighted graph coloring-based pilot decontamination (WGC-PD)

1 Introduction

In the 5G wireless communication system, Multi-input multi-output (MIMO) systems are widely analyzed to meet the rapid and exponential increase in a large number of users and their high data traffic. In this 5G mobile wireless communication, the base stations (BS) are provided with an abundant number of transmitting and receiving antenna for supporting the more number of the set of users and support large data traffic [1]. If the number of antennas used is always fixed in the 5G massive MIMO wireless systems in which it gives BS as the total number of the antenna will goes reduce the inter-cell interference between the set of users and uncorrelated noise among the user's channels [2].

By the exponential increase in the number of antennas, it will lead to a massive MIMO system. This massive MIMO system is giving more benefits in the context of received signal detection scheme from the different antennas and estimation of the desired signal [3]. In this context, the contamination of pilots will be caused for the effective reuse of various pilots in the adjacent cells because of the limitation of less number of pilot's contaminations in a system. This pilot resource-effective utilization does not diminish the size of the BS antennas and values. So, the pilot contamination is to be recognized properly and it is being bottlenecked of multi-cell massive MIMO system [4].

There are numerous effects that have been taken to resolve these PC problems in these massive MIMO systems [5, 6]. These are different new methods that are evolved to solve PC issues. They are, time sifting pilot scheme: This scheme is more suitable for asynchronous transmission among the adjacent cells. Smart pilot assigning scheme: In this method, the pilots are assigned sequentially for each and every cell. PC precoding scheme uses multi-cell joined signal processing methods in a greater way. AOA method will assign the pilots based on a geographically separated by non-overlapping AOA values.

In all these existing systems, the severity value PC may carry for all the users and all this solution is also to try to mitigate and avoid the pilots for all the users, while the other will enjoy to negligible PC. In this proposed work, the modified EWIG–WGC-based reweighted pilot contamination scheme is designed and proposed to avoid and mitigate the contaminations of the pilots for this various multi-cell massive MIMO system is formulated the optimization problem in order to maximize and convergences of the uplink achievable rate. In the second phase, the modified EWIG–WGC construction is formed from the classical pilot assigning and contamination algorithms of Multi Cell massive MIMO systems. This algorithm greedily comprises and assigns a different pilot connect to multi-users for large iteratively re-weight pilots over the users.

2 Literature Survey

Many researchers have taken many efforts to resolve challenging pilot contamination (PC) problems. Some of the recent related works are explained in this section.

Wei et al. [7] have suggested the PARAFAC-based channel estimation method in order to overcome the pilot contamination over the uplink process for the massive multiple input multiple output (MIMO) systems for the exploitation of the low-rank property in a physical environment. In this scheme, the proposed parallel factor called PARAFAC, which derives massive MIMO systems parameters of the fading co-efficients, the direction of arrivals and delay characteristics that are caused by scattering in the environment. After the estimation of PARAFAC values, they formulate the estimation of optimization problems occurring in the three variables and get an optimal solution through the ALS algorithm. The major limitation in this proposed scheme is achieving high estimation accuracy by the AS is not possible.

Zhao et al. [8] have proposed a novel strategy for assigning the optimized pilot to minimize pilot contamination effect in the TDD mechanism-based massive multi-input multi-output-based wireless systems. The scheme consists of Chu-based sequences with a perfect matched auto-correlate property that is used for effective design and assigns the optimal value of the pilot sequences for gathering the channel state information (CSI) at the transmitter (CSIT). They presented analytical expressions with the NMSE value-based channel estimation scheme. From that, they calculate a strategy to reduce the contamination of pilot signals. This strategy of this pilot assignment scheme is widely used for improving the effective performance measures of this wireless system, by which it can attain optimal solutions. Thus, the improvised system can be used to enable reducing pilot contamination. The main disadvantage is the uplink achievable data rate versus SNR performance of TDD-based massive MIMO needs to be improved.

Muamer Hawej, et al., proposed this scheme in order to reduce the contamination of the pilots in TDD multi-cells massive MIMO system. In this scheme, this pilot contamination effect occurs only during the non-orthogonal TDD-based multiplexing over the pilot contamination over the users. This scheme comprises two estimation methods that are given on the low-rank matrix value-based approximation techniques [9]. The channel estimation is constituted of the new metrics called nuclear norm vector based on the optimization problem to improve the accuracy by mitigating the pilot contamination. The estimation method IWNN is used to increase the performance of the nuclear norm estimation value. This method needs to improve the present problems in pilot contamination with a small number iteration cost. Jiaming Li, et al., introduced a novel estimation technique as a hybrid pilot (TM pilot, which is time division-based multiplying of the pilot and TS pilot, which is time superimposed pilot) by the massive MIMO systems-based uplink channel will divide the closed form of its optimization and approximation techniques for improving the uplink achievable rate [10]. This hybrid pilot can be used with the help of both TM pilot contamination and TS pilot contamination. With the hybrid pilot, they can provide a superior range of performance to do work with TS pilot or TM pilot. The power and time ratio between pilot and data needs to be optimized.

Hayder Al-Salihi et al. have proposed a new pilot design and assigning scheme to minimize the NMSE estimators of massive MIMO system base station (BS). In this scheme, at first, the originally transmitted information signal is decomposed into convex problems into different distributed optimization problems for each and every individual BS. After this optimization, the successive approximation technique is converted into a linear matrix-based inequality form based on the optimization algorithm [11–13]. The individual BS antennas can be widely used to optimizing their own pilot data with the help of other signals from all the BS. The major disadvantage is that the proposed algorithm computational complexity is comparatively high.

3 Proposed Pilot Decontamination Algorithm

Let the multi-cell massive MIMO system model is considered with consisting of the L values of the number of hexagonal cells chosen, and every hexagonal cell widely consists of a base station with a different value of M number of antenna and K single transmitting antenna users [14]. From the j_{th} cell used in the adjacent kth user to the BS of the ith cell the channel vector $h_{(j,k),i} \in C^{M \times 1}$ will be given as,

$$h_{(j,k),i} = g_{(j,k),i}\sqrt{\beta_{(j,k),i}} \tag{1}$$

From the above expression, the large- and small-scale fading vector is represented by $\beta_{(j,k),i,}$ and $g_{(j,k),i}$, respectively. This small-scale values of the fading vector with some distribution CN(0,I_M). At the time of the coherence period, $h_{(j,k),i}$ is the channel vector will be unaltered when it will make use of mostly used block-fading channel model. Consider that the total available $S(S \geq K)$ pilots $\varphi_i \in C^{T \times 1}$ $(1 \leq i \leq S)$ are orthogonal to each other when the length T can be used in one cell, $\phi = [\varphi_1, \varphi_2, \ldots, \varphi_S]^T \in C^{S \times T}$, $\phi\phi^H = I_{S,}$ due to limited pilot resource, the similar pilot group ϕ is reutilized [15].

By Rayleigh fading channels, the pilot reutilizes technique that can be occurring within the one cell that can be exploiting correlatively. The pilot assignment methods are used to allocate pilot $p_{(j,k)}$ to the users at the different timeslots as (j, k) non-specifically, $P(j, k)$ and that ensure that the pilot's contaminations have not to be reutilized within one of its adjacent cell $p_{(j,\,k)} \neq p_{(j,\,k')}$, The user (j, k) has the uplink SINR value for the station antennas. So, it is represented by (SINR) will be based on the matched-filter receiver is assuming to the base station. This can be solved pilot assignments as,

$$\mathrm{SINR}^{UL}_{(J,K)} = \frac{h^{H^4}_{(j,k)j}}{\sum_{(j\circ,k\circ)\in I_{(j,k)},\, j\circ \neq j'} h^{H^4}_{(j,k)j} + \sigma^2_{(j,k)}/\rho^2} \tag{2}$$

$$\approx \frac{\beta_{(j,k),j}^2}{\sum_{(j_\circ,k_\circ)\in I_{(j,k)},j_\circ\neq j'}\beta_{(j_\circ,k_\circ),j}^2}\mathrm{M} \rightarrow \infty \tag{3}$$

The same pilot with a set of users has the user (j, k) and potential of uncorrelated interference is represented by $I_{(j,\,k)}$ and $\sigma_{(j,k)}^2$, respectively. The additive white Gaussian noise is always be the given essentially decreased by the rapid increasing the number of base station antennas M, $\sum_{(j',k')\in I(j,k),j_\circ\neq j'}\beta_{(j',k'),j}^2$ indicates PC that is occurring by the pilot reused ρ indicates the power for the transmission. The calculated average uplink achievable rate of the user (j, k) is,

$$C_{(j,k)}^{UL} = (1-\mu_s)E\left\{\log_2\left(1+\mathrm{SINR}_{(J,K)}^{UL}\right)\right\} \tag{4}$$

Uplink pilot transmission can because of the loss of spectral co-efficiency which can be solved by μ_s. That is the ratio of channel coherence time $\mu_\mathrm{s} = {}^{\tau}\!/_{l}$ and pilot length T. From the above equations, it is said that the small-scaling d and thermal noise are equal to fading effects. This can be calculated as M; it leads to infinity by pilot contamination.

4 Proposed WGC-Based Pilot Contamination Scheme

4.1 Problem Formulation for Optimization

The pilot assignments based for N slots for S pilot and jth cell is $A_s^k = S!/((S-K)!)$ for the K users. When only one cell is considered, there is no change in the A_s^k kinds of pilot assignments. A specific structure with L number of the hexagonal cells, the whole number of importantly various numbers of the pilot assignment is high as $(A_s^k)^{L-1}$. The allocation of the pilot $\varphi_{p(j,k)}$ to user randomly (j, k) by classical pilot assignment. In L hexagonal, the maximization of KL users can be achieved. This can be formulated as the accompanying improvement issue P_1:

$$\max_{p(j,k)}\left\{\sum_{(j,k)}\log_2\left(1+\frac{h_{(j,k)j}^{H^4}}{\sum_{(j_\circ,k_\circ)\in I_{(j,k)},j_\circ\neq j'}h_{(j,k)j}^{H^4}+\sigma_{(j,k)}^2/\rho^2}\right)\right\} \tag{5}$$

where $\{P(j,k)\}$ indicate the types of various pilot assignment and (j, k) indicate all K and L number of the users. Since the base station has not given accurate channel information. This optimization problem P_1 is impossible to solve. A huge amount of fading coefficient $\beta_{(j,k)}$ can represent the limit of total uplink throughput. The optimization problem P_1 can be approached by P_2,

$$\max_{p(j,k)}\left\{\sum_{(j,k)}\log_2\left(1+\frac{\beta^2_{(j,k),j}}{\sum_{(j_\circ,k_\circ)\in I_{(j,k)},j_\circ\neq j'}\beta^2_{(j_\circ,k_\circ),j}}\right)\right\} \tag{6}$$

The optimization problem P_2 has been solved by the complete search in $(A^k_s)^{L-1}$ kinds of the pilot assignment. The classical multi-cell massive MIMO system with various numbers of the L value is 7 and both S and K value is 8 the search difficulty is $(A^k_s)^{L-1} = (8!)^6 \approx 4.3 \times 10^{27}$.

4.2 Modified EWIG Construction

Two users in the various cells can be considered with a similar pilot, (j, k) and the inverse pilots for $P_{(j,\,k)} = P_{(j',\,k')}$, Now, modify the uplink rate of users (j, k) generated as,

$$C^{UL}_{(j,k)} \propto \log_2\left(1+\frac{\beta^2_{(j,k),j}}{\beta^2_{(j\prime,k\prime),j\prime}+\varepsilon_{j,k,j\prime,k\prime}}\right) \tag{7}$$

where $\varepsilon_{j,k,j\prime,k\prime} = \sum_{(j_\circ,k_\circ)\in I_{(j,k)},j_\circ\neq j\prime}\beta^2_{(j\prime,k\prime),j}$ denotes the pilot contamination. When the same pilot with different users can cause pilot contamination. $C^{UL}_{(j,k)}$ could be comparably represented by using pilots of the assigned users. To verify the original measurement of pilot contamination with the various $\varepsilon_{j,k,j\prime,k\prime}$ of user (j, k) is very complicated.

The pilot contamination ratio is always between the user used in the vector $(j\prime, k\prime)$ and users (j, k) with a similar pilot are nearly the same. The ratio of pilot contamination between the users is $\beta^2_{(j\prime,k\prime),j}/\beta^2_{(j,k),j}$ and $\beta^2_{(j,k),j\prime}/\beta^2_{(j\prime,k\prime),j\prime}$. The potential impact of the contamination of pilots is measured always who lies in between a large number of users in various cells, (j, k) and (j', k'),$j \neq j'$. For this operation, define a metrics is known as pilot contamination. This can be represented as,

$$\zeta_{(j,k),(j\prime,k\prime)} = \beta^2_{(j\prime,k\prime),j}/\beta^2_{(j,k),j} + \beta^2_{(j,k),j\prime}/\beta^2_{(j\prime,k\prime),j\prime} \tag{8}$$

where $\zeta_{(j,k),(j\prime,k\prime)}$ represents the variation of the proportion between impedance channel quality and powerful channel quality, and larger ζ represents that increasingly serious pilot contamination will be established between ore number of the user (j, k) and user over the multi-cell user (j', k') when the similar pilot is always an issue. Unmistakably pilot contamination metric as equal property $\zeta_{(j,k),(j\prime,k\prime)} = \zeta_{(j,k),(j\prime,k\prime)}$, $\forall j \neq j'$. Thus, modified EWIG has widely used be interspersed as a directionless weighted graph $G = (V, E)$, where $V = \{(j, k)\colon \leq j \leq L,\ 1 \leq k \leq K\}$, and $E = \{\zeta_{(j,k),(j\prime,k\prime)}\colon j \neq j'\}$, thus the V represent the vertices in all users and E represent the edges in all users have potential pilot contamination. When the edges with removable weight lies between two users in various cells are negligible for resemblance.

4.3 Modified EWIG and WGC-Based Pilot Contamination Scheme

There are many users in various cells that are placed in the EWIG. This type of assigning a similar pilot is assigned it is clear that it has the large pilot contamination $\zeta_{(j,k),(j\prime,k\prime)}$. This represents to introduce more number of pilot contaminations. When the pilot contamination $\zeta_{(j,k),(j\prime,k\prime)}$ is sufficiently low, then the performance loss is negligible for the different users in various cells has repeated similar pilot.

To investigate the important pilot assignments with significantly decreased pilot contamination by manually allocating a various number of the pilots to the attached users with a bigger iterative re-weight mechanism. For this purpose, it can use EWIG construction. EWIG is an essential tool for this operation. When the pilot resource is usually limited only K $\ll$ KL represents the more number of the N number of the orthogonal pilots of the vector placed are applicable in classical multi-cell massive MIMO systems. Under the N number of the restriction of minimal pilot resources, the WGC-PD method is proposed to achieve a change between pilot overhead and the reduction of pilot contamination and substantially reduces pilot contamination.

On the base of the typical DSATUR method which will always sort the value of the vertices range in more descending order based on its degrees and colors them as sequential based as possible with reused colors. The EWIG users are linked with the proposed WGC-PD method manually allocated various pilots with a broad weight. Unlike the DSATUR method which ensures that unused vertices are allocated with similar color, there are two users in the separate adjacent cells with a minimum weight may be allocated to the similar contaminated number of the pilots to the more and new restriction of finite pilot resources. This proposed weighted graph coloring-based pilot decontamination (WGC-PD) method will be always considered for the different variants based on the standard DSATUR method edge-weighted graph over the constrained strategies. The proposed WGC-PD involves three important main steps namely, initialization, user selection, and pilot assignment.

4.3.1 Initialization

User $(j_{1,}\ k_1)$, and user $(j_{2,}\ k_2)$ are two users. Such two users are selected in various cells of the EWIG. The two users are assigned their pilots φ_1 and φ_2, respectively; they are added to assigned pilots and prepare for the initialization. After that, the rest of the users are picked and sequentially allocated to all the users.

4.3.2 User Selection

An important parameter $\delta_{(j,k)}$ for select users in order of specification is introduced, which is defined as users in various cells within Ω and the weighted aggregate of the edges interface user (j, k). The user (j_0, k_0) with the largest effective pilot contamination intensity from allocation set Ω will determine whose pilot assignments must be treated favorably.

4.3.3 Pilot Assignment

After the selection of the user (j_0, k_0), the proposed reweighted-based WGC-PD method focused to pick the pilot which causes the lower efficient pilot contamination to their users from the present pilot resources. The alternative pilot set A is initially developed, which consists of the free pilots in the j_0th cell to make sure that none pilot is repeated within a similar cell $p_{(j,k)}$.

By assuming that user (j_0, k_0) is assigned to pilot *S* and also defined η_s to describe the effective pilot contamination potential between the adjacent cell users with various assigned pilot *S* in the range and user (j_0, k_0) . The pilot has the lowest effective pilot contamination potential η_s is pick to be allocated to the user (j_0, k_0) and user (j_0, k_0) will be always added into the allocated within the number of the set Ω. The process will be done in a linear manner with all users being allocated to their respective pilots.

5 Simulated Results and Discussion

The chapter gives a detailed description of the performance of the proposed modified weighted graph coloring-based pilot decontamination algorithm is investigated by using Monte Carlo simulations. In this massive MIMO system, the typical hexagonal type of cellular structures with *L* number of the cells is considered, where each cell has *K* number of single-antenna users and base station (BS) with *M* number of receiving and transmitting antennas. For the spectral efficiency, loss value with *S* is always equal to *K* is set as $\mu_0 = 0.05$. So the corresponding μ_s value is calculated as $\mu_s = \left(\frac{S}{K}\right)\mu_0$. The parameters for the simulation are tabulated in Table 1. The co-efficient $\beta_{<j,k>,i}$ modeled as,

$$\beta_{<j,k>,i} = \frac{z_{<j,k>,i}}{\left[\frac{r_{<j,k>,i}}{R}\right]^{\infty}} \tag{9}$$

where $Z_{\langle j,k\rangle,i}$ gives the shadow fading models and it gives a log normal probability distribution is always equal to Gaussian distributed with mean value as zero and the σ_{shadow} standard deviation $r_{\langle j,k\rangle,i}$ is measured as the distance vector matrix between *K*th number of the user of *j*th number of the cell and βS values is the *i*th cell and cell radius is given as *R*.

Table 1 Simulation parameters

Parameters	Values
Total number of adjacent cells considered (L)	3,7
Total number of base station considered (K)	8–256
Total number of end users in the each hexagonal cells (K)	4,6
Total number of orthogonal contamination pilots (s)	$K \leq S \leq KL$
Radius of communication range of each cells (R)	500 m
Power of transmission (P)	[5,30] dB
Minimum spectral efficiency loss considered	$\mu_0 = 0.05$
Minimum path loss exponent considered	3
Log normal shadow model value	8 dB

5.1 CDF Comparison

The uplink achievable rate cumulative distribution function (CDF) curve is calculated simulated by using the MATLAB with the following parameters. The system parameters considered for the simulations are tabulated below,

The proposed WGC-PD method is simulated and comparing the proposed systems with all the other existing systems. Initially, the pilot assignment scheme is assigning the various pilots to the users randomly. In this classical pilot assignment strategy, the pilots are not having any co-operations among the cells. Here, it is considered the two cases, the BS is provided with more number of transmission and receiving antennas. These two cases are $M = 32$ and $M = 256$. By considering the optimum solution of the exhaustive searches, this simulated result shows that the proposed scheme is ensured the outperforms that all the other existing classical pilot assignment strategies (Figs. 1 and 2).

The data statistics of the curves are tabulated in Table 2 and Table 3. These values form the output result that this proposed method outperforms the other existing system.

5.2 SNR Comparison

The difference between the average powers of the transmitted signal to the average power of the noise in the communication system is called SNR. The figure shows that the average uplink achievable rate per user against the power required for the transmission or the power received at the base station (Figs. 3 and 4).

The data analytics of these curves are tabulated in Table 4. These values show that the proposed work will outperform the existing pilot assignment schemes is given in Table 5.

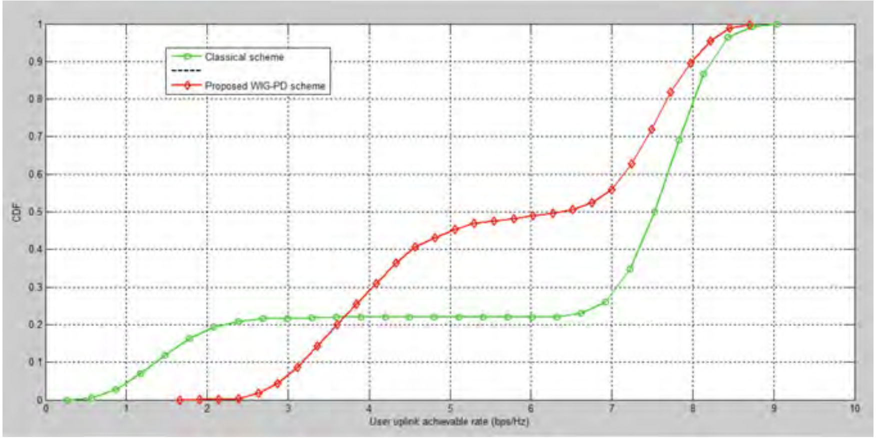


Fig. 1 CDF of the user's uplink achievable rate ($M = 32$)

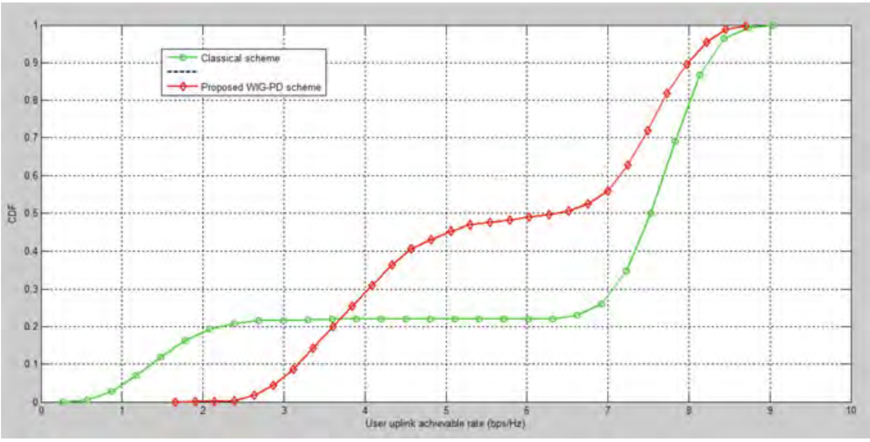


Fig. 2 CDF of the user's uplink achievable rate ($M = 256$)

In this proposed pilot scheme S and K values are the same, this will utilize the pilots effectively so this proposed scheme significantly outperforms than all other existing classical schemes. In this classical pilot assignment strategy, the pilot resources are not utilized effectively rather than assigned randomly. But in this proposed scheme, the pilots are assigned by using graph theory and weight based.

Table 2 Comparison of statistical data of CDF ($M = 32$)

Parameters	$M = 32$			
	Classical scheme		Proposed method	
	User uplink achievable rate (bps)	CDF	User uplink achievable rate (bps)	CDF
Minimum	0.2406	0	0.3365	0
Maximum	3.743	0.9975	3.479	0.9975
Mean	1.992	0.616	1.908	0.6035
Median	1.992	0.8016	1.908	0.7719
Mode	0.2406	0	0.3365	0
STD	1.063	0.3988	0.954	0.3996
Range	3.502	0.9975	3.143	0.9975

Table 3 Comparison of statistical data of CDF ($M = 256$)

Parameters	$M = 256$			
	Classical scheme		Proposed method	
	User uplink achievable rate (bps)	CDF	User uplink achievable rate (bps)	CDF
Minimum	0.2686	0	1.659	0
Maximum	9.042	0.9987	8.7	0.9975
Mean	4.655	0.3161	5.18	0.4239
Median	4.655	0.2194	5.18	0.4609
Mode	0.2686	0.2194	1.659	0
STD	2.663	0.287	2.137	0.311
Range	8.774	0.9987	7.041	0.9975

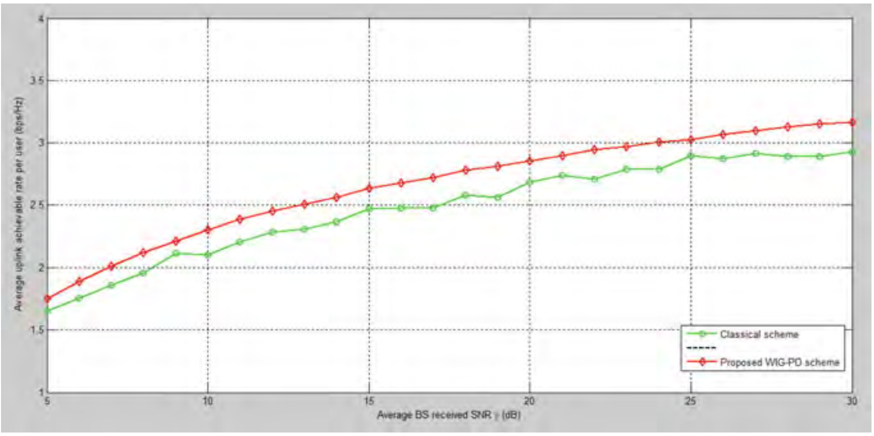


Fig. 3 Average uplink achievable rate for $M = 32$

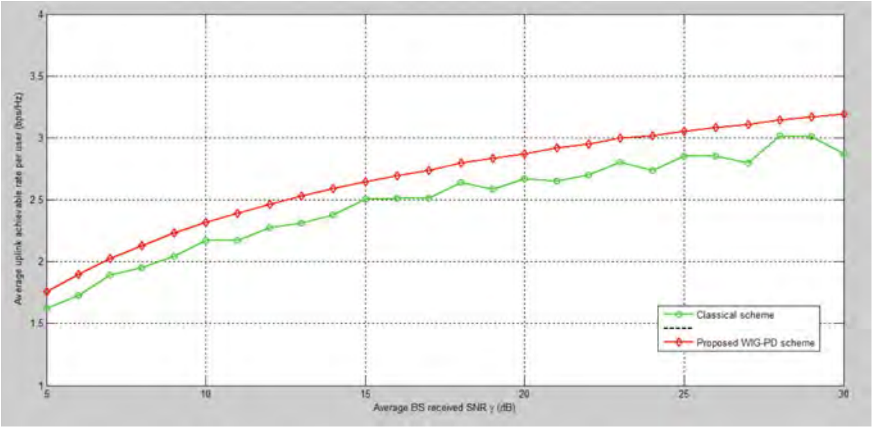


Fig. 4 Average uplink achievable rate for $M = 256$

Table 4 Comparison of statistical data of SNR ($M = 32$)

Parameters	Average uplink achievable rate ($M = 32$)		
	Average BS received SNR (dB)	Classical scheme	Proposed scheme
Minimum	5	1.652	1.75
Maximum	30	2.925	3.166
Mean	17.5	2.472	2.558
Median	17.5	2.52	2.749
Mode	5	1.652	1.75
STD	7.649	0.3845	0.4116
Range	25	1.273	1.416

Table 5 Comparison of statistical data of SNR ($M = 256$)

Parameters	Average uplink achievable rate ($M = 256$)		
	Average BS received SNR (dB)	Classical scheme	Proposed scheme
Minimum	5	1.621	1.755
Maximum	30	3.016	3.192
Mean	17.5	2.472	2.675
Median	17.5	2.548	2.766
Mode	5	1.621	1.755
STD	7.649	0.3903	0.4156
Range	25	1.394	1.437

6 Conclusion

The proposed work investigates the existing pilot assignment methods. The weighted graph coloring-based pilot decontamination assignment scheme is proposed for massive MIMO systems with 5G wireless systems to overcome and mitigate the contamination of the pilots in the assignment related issues. In this scheme, EWIG metrics are first constructed based on the available potential contamination of the PC relationships among the adjacent cell users. After the construction of EWIG, PC assigned different pilot contamination to the users in which it having the large weight in iteratively reweighted EWIG. This will reduce the unwanted pilot contains and ensures the efficient utilization of the available resources. The simulation result shows that the proposed WGC-PD scheme outperforms all the other existing systems.

References

1. Larsson EG, Edfors O, Tufvesson F, Marzetta TL (2014) Massive MIMO for next generation wireless systems. IEEE Commun Mag 52(2)
2. Rusek F et al (2013) Scaling up MIMO: opportunities and challenges with very large arrays. IEEE Sig Process Mag 30(1):40–60
3. Dai W, Milen kovic O (2009) Subspace pursuit for compressive sensing signal reconstruction. IEEE Trans Inf Theory 55:2230–2249
4. Ciuonzo D, Salvo Rossi P, Dey S (2015) Massive MIMO channel-aware decision fusion. IEEE Trans Sig Process 63(3):604–619
5. Zhu X, Wang Z, Dai L, Qian C (2015) Smart pilot assignment for massive MIMO. IEEE Commun Lett 19(9):1644–1647
6. You L et al (2015) Pilot reuse for massive MIMO transmission over spatially correlated Rayleigh fading channels. IEEE Transm Wirel Commun 14(6):3352–3366
7. Wei X, Peng W, Chen D, Ng DWK, Jiang T (2019) Joint channel parameter estimation in multi-cell massive MIMO system. IEEE Trans Commun 67(5)
8. Zhao J, Ni S, Gong Y, Zhang Q (2019) Pilot contamination reduction in TDD-based massive MIMO systems. IET Commun
9. Hawej M, Shayan YR (2019) Pilot decontamination in massive multiuser MIMO systems based on low-rank matrix approximation. IET Commun
10. Li J, Yuen C, Li D, Wu X, Zhang H (2019) On hybrid pilot for channel estimation in massive MIMO uplink. IEEE Trans Veh Technol 68(7)
11. Al-Salihi H, Van Chien T, Le TA, Nakhai MR (2018) A successive optimization approach to pilot design for multi-cell massive MIMO systems. IEEE Commun Lett 22(5)
12. Hassan N, Fernando X (2017) Massive MIMO wireless networks: an overview. MDPI J Electron Spe Issue Smart Antennas MIMO Commun 6:63
13. Al-Juboori S, Fernando X (2018) Multiantenna spectrum sensing over correlated Nakagami-m channels with MRC and EGC diversity receptions. IEEE Trans Veh Tech 67(3):2155–2164
14. Biswas A, Gupta VR (2020) Design aspects of 5G: frequency allocation, services and MIMO antennas. Eng Appl Sci Res 47(1):103–110
15. Albreem MA, Juntti M, Shahabuddin S (2019) Massive MIMO detection techniques: a survey. IEEE Commun Surv Tutorials 21(4):3109–3132

Synthesis Approach for Emotion Recognition from Cepstral and Pitch Coefficients Using Machine Learning

S. Vaijayanthi and J. Arunnehru

Abstract Emotion Recognition is a significant research domain for speech emotion identification, which includes various emotions like happy, calm, sad, angry, anxious, depressed, fearful, etc. Hence, speech emotion analysis is gaining more admiration due to the upcoming challenges in Computer Vision. Speech commands are remaining as the most prominent way for expressing ourselves as humans. There are many new methods to analyze vocal emotion synthesis using machine learning methods. This paper proposes a synthesis approach to combine the Mel Frequency Cepstral Coefficients (MFCC) with the vibration rate(PITCH) in order to characterize the emotion according to its respective vocal speech signals. The RAVDESS dataset is utilized here and the extracted features are modelled using the K-Nearest Negibhour and Decision Tree classifier for recognizing the eight emotions. The experimental results, show the efficacy of the proposed method with an overall mean accuracy rate of 87.12% for K-NN and 77.39% for Decision Tree, which outperforms the state-of-the-art results.

Keywords Speech emotion · Mel frequency cepstral coefficients · Pitch · Feature extraction · Machine learning

1 Introduction

Speech Emotion Recognition (SER) is a knowledge that extracts sensitive features from speech signals by machine. The variations and interpretations of the specific parameters and the emotional change also acquired. Almost many proposed systems combine two processing steps. The initial step is to separate the input wavelets and

S. Vaijayanthi (✉) · J. Arunnehru
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus, Chennai, Tamilnadu, India
e-mail: vaijayanthisekar@gmail.com

J. Arunnehru
e-mail: arunnehru.aucse@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_39

extract specific features (parameters) from it. The feature extraction [1] usually means it includes a significant knowledge compression. The next step performs a grouping of similar audio files based on the extracted features. Feature extraction is a necessary part of speech emotion recognition, and it mainly associates in SER problems; hence this research introduces a novel approach for feature extraction, using MFCC to extract emotional features in dialogue signal automatically. SER is an emerging research topic in artificial intelligence, artificial therapy and pattern recognition. The broad application of this analysis is in security fields, auto supervision in healthcare areas, interactive teaching, human-computer interaction, entertainment, and so on. Speech emotion processing and recognition systems comprise of three broad sections; they are speech signal retrieval, feature extraction, and emotion recognition. A significant difficulty in this area is the automated classification of audio files. The MFCC + PITCH technique helps in extracting features from raw audio files and generates a solid description of the content. The objective is to discover the most suitable sequence of emotions, relevant to the average and vigorous-intensity of the audio file. The features are extracted based on eight different emotions.

1.1 Outline of the Work

This research paper deals with speech-based emotion recognition, which aims to identify emotions from the speech signal. The evaluation of proposed work is carried out by using RAVADESS dataset with the statement 1 is “Kids are talking by the door”, and statement 2 is “Dogs are sitting by the door”. The feature extraction techniques involve MFCC + PITCH to obtain the audio features from human speech. K-Nearest Neighbor and Decision trees classifiers are helpful in feature extraction and undergo training and testing with a 5-fold cross-validation approach.

The rest of the paper is structured as follows. Section 2 summarizes the related work. Section 3 gives an overview of the proposed technique of feature extraction. Section 4 illustrates the K-Nearest Neighbor and Decision tree classifier. Section 5 describes the datasets used, and the experimental setup Sects. 6 and 7 subsequently reports the evaluation outcomes and comparative study with various state-of-the-art models. At the end, Sect. 8 concludes the paper.

2 Prior Research

The speech data consists of emotion recognition [2], which is essential to extract the features precisely and to signify the emotional phase of different speech waves. The prime task here is to mine proficient features for the most excellent categorization of Emotions. The prior methods include the category of synthesis and analysis of speech emotion [3].

Mel Frequency Cepstral Coefficients (MFCCs) have undergone many strategies in the area of speaker identification [4] and emotional speech recognition. Comparison of Existing studies proved that MFCCs has a better way to analyze emotions with other traditional speech features. (e.g., linear predictive coding, loudness, PLP, etc.) [5]. Hansen and Bou-Ghazale [6] verified that the features built on cepstral based analysis, overtake the Linear Predictive coding (LPC) in refining the stressed speech emotion in SUSAS database. Liu [7] revealed that there is an average rise in the accuracy of 3.6% in the Gammatone Frequency Cepstral Coefficients (GFCCs) featureset aganist MFCCs for identifying emotion. Besides, different vocal prosodic and spectral features such as shimmer and glottal parameter, etc. Are also interconnected with speech emotion [8]. Liu et al. [9] interconnected voice quality parameters such as shimmer and jitter with MFCCs features in the SUSAS database to identify emotions. He proposed, a feature selection algorithm, created by using the Fisher Correlation Coefficient and correlation analysis. He classifies the Chinese speech database by using the Extreme Learning Machine (ELM) decision trees from the Institute of Automation of the Chinese Academy of Sciences (CASIA). The Fisher Criterion helps in removing redundancy features of similar audio sources extracted from the speech emotion recognition.

Recent times, a variety of different types of features has majorly used for the identification of voice-based emotion recognition. Pan et al. [10] unveiled the mixture of Pitch, Mel-energy spectrum dynamic coefficients (MEDCs), LPCC with MFCCs in SVM classifier with the help of Berlin EmoDB and Chinese emotional database (SJTU). The main difference between MFCCs and MEDC is that MFCCs requires log after the filter bank, while MEDC needs a filter bank with a logarithmic average of energies. Chen et al. [11] pulled out the energy, Pitch, Zero crossing rate (ZCR), Correlation density, Spectrum centroid and cut off frequency with five level Melfrequency energy bands.He used fractal dimension from the three-level speech recognition model which in turn helps in resolving the difficulty of speaker-independent speech emotion recognition model. The three-level categories include, pairwise six vocal emotions, representing the input as Fisher rate at each level and providing a better Classification than the previous one.

Schuller et al. [12] combined linguistic and acoustic information of emotional features with multiple-stage classifier in SVM over seven different emotional classes. The Emotional key-phrases are spotted by Belief Network using phrase-spotting. The performance comparison with a variety of classifiers include Nearest Neighbors algorithm, Support Vector Machines (SVM), Linear Classfiers and Gaussian Mixture Model (GMM) is stated with FERMUS III emotional corpus. Rao et al. [13] deliberated his work in recognizing emotions by fusing the acoustic features and facial expressions from video-based speech signal in the real-life emotional database and gives the improved performance in the Hidden Markov Model (HMM), SVM, and GMM.

Fahad et al. [14] enhanced DNN-HMM adaptive model for identifying four emotions with epochs features based on the strength of excitation (SoE), change of phase, and instantaneous pitch is combined with MFCCs using IITKGP-SEHSC and IEMOCAP databases. The primary purpose is to extract and utilize the speech features and to avoid rapid change in the vibrations of the vocal cords of the speaker. In general prosodic features, these changes are not captured, and hence the speech signal assumed to be static. The significant challenges in the prior work are limited to 4–5 emotions since many databases do not prefer a different emotions. The pre-processing steps that effectively refine data and improve the accuracy of the classifier are missing in the context of previous SER techniques. This paper proposes an efficient SER technique to process the emotions with eight different intensity labels in the vocal speech tract, to recognize emotions in real-time.

3 Feature Extraction

The extraction of individual person emotion is the most prominent part of human speech emotion recognition [15] which represents the most significant facts that are essential for future study, the following sections ensembles the representation of the feature extraction methodology with eight different input emotional states, such as happy, sad, angry, calm, disgust, neutral and surprised, are used for two dissimilar statements in this paper. Figure 1 provides an overview of the proposed work.

3.1 Mel-Frequency Cepstral Coefficients (MFCC)

The most traditional and robust audio feature extraction methodologies are the Mel-frequency cepstral coefficients (MFCC), which have 13 features. In that, eight dif-

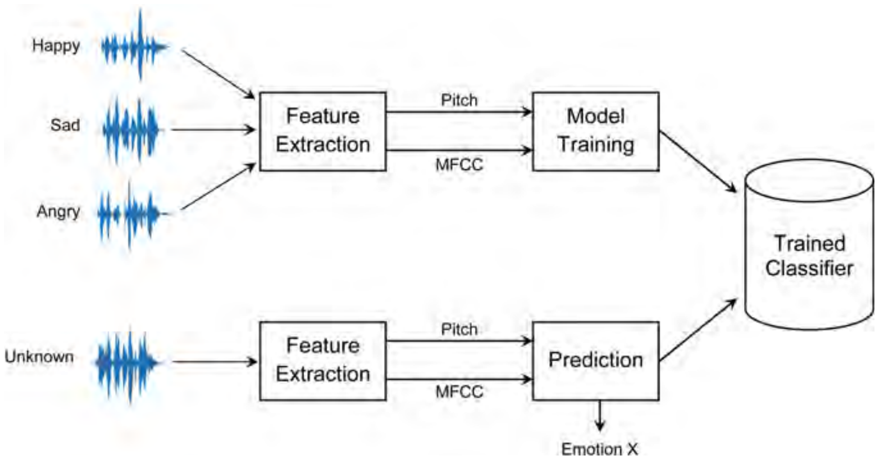


Fig. 1 Overview of the proposed work

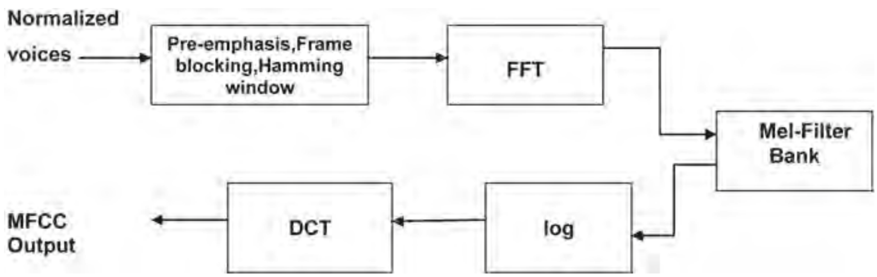


Fig. 2 MFCC feature extraction process

ferent emotional speeches are considered for classification. The feature sum is inadequate to force us to train the data of the audio. The Eight emotions were sad, happy, calm, neutral, angry, disgust, fearful and surprised etc. with two different statements. MFCC focus on interpretations of human sound perceptions, where it cannot recognize frequencies above 1 Khz [16]. It mainly involves the process of windowing the speech signal by doing triangular overlapping of the windows to take Fourier transform and map the spectrum powers to the mel scale. The third step is to find the log of the magnitude, and apply Discrete Cosine Transform to the list of mel log power, resulting is MFCC amplitudes. The complete outline of the MFCC features and the extraction steps is shown in Fig. 2.

3.1.1 Pre-emphasis, Frame Blocking, and Windowing

Pre-emphasis involves boosting the energy level and filtering to higher frequencies. The idea behind this is to poise the spectrum of vocal sounds like vowels that have a sudden roll-off in the high-frequency Zone compared with lower frequency energy levels. Hence it is called a spectral tilt, which relates to the glottal source. The glottal source in turn has −12 dB/octave drop for vocal speech signals [17]. The radiation of the sound (acoustic) energy level from the lips creates an approximately +6 dB/octave boost to the log spectrum. This results in the improvement of phone detection accuracy. On the other hand, when a vocal speech signals recorded with the help of the microphone from an average position have a downward 6 dB/octave curve corresponding to the actual spectrum of the vocal region. As a result, pre-emphasis eradicates some of the vocal tracts parameters in the glottal effects. The succeeding transfer function specifies the most frequently used pre-emphasis filter by using Eq. 1.

$$H(Z) = 1 - bz^{-1} \tag{1}$$

Here the rate of b has control over the slope of the filter between 0.4 and 1.0 [17].

The most common approaches in speech signal processing mainly relate to short-time analysis [18]. The vocal signal is a moderate quasi-stationary or time-varying signal. Regular examination of speech signals, based on steady acoustic characteristics, are examined sufficiently for a short period. The speech analysis is performed in short segments across the assumption of the speech signal is expected to be stationary. The pre-emphasis signal develops the frames of N samples. The high-frequency temporal characteristics formants process and it relates to the amplitude compared to a low frequency to attain a similar amplitude for all the formants. In all the frame, a window can minimize the speech signal with respect to frame borders. Typically, Hamming or Hanning windows are most probably used [17] to improve edge smoothing and it also decreases the edge impact of Discrete Fourier Transform (DFT) signal.

3.1.2 Fast Fourier Transform

In each frame, the Fast Fourier Transform (FFT) is imposed after windowing to discover the power spectrum of the entire frame structure. Here filter bank processing is passed out on the power spectrum, using mel-scale. Finally the Discrete Cosine Transform helpful in the conversion of windowed frames into the magnitude spectrum by using Eq. 2.

$$x(k) = \sum_{n=0}^{N-1} x(n) \; e^{-\frac{j2\pi nk}{N}} \tag{2}$$

Here $x(k)$ is the frequency domain samples, $x(n)$ is the time domain samples, N exhibits the FFT size, and k ranges from $(k = 0, 1, \ldots, N-1)$.

3.1.3 Mel Scale Filter Bank

The Fourier transform moves through a series of band-pass filters to compute the Mel spectrum range known as Mel–Filter bank. Mel is a measurement of human ears received frequency in units. The Mel scale is one of the major types of linear frequency ranging below 1 khz and has a log value above 1 khz [4]. Mel estimation of physical frequency is as follows

$$f_{\mathrm{Mel}} = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \tag{3}$$

In Eq. 3 (f) represents the physical frequency in Hertz, and (f_{Mel}) indicates the perceived frequency [19]. In MFCC, Filter bank's implementation carried out using frequency domain. In the frequency domain axis, the centre bank filters spaced uniformly, to identify the human ear's perception.

3.1.4 Discrete Cosine Transform

Discrete Cosine Transform is performed in MFCC to generate a pair of cepstral coefficients because the verbal stretch is smooth, and the energy level tends to be correlated. The time domain signal of the Mel log Spectrum is identified and hence the result of high frequency and low-frequency pitch signal is easily distinguished. The MFCC separates the signal coefficient information at the beginning itself, so it is easy to extract and truncate the higher-order DCT derivatives. MFCC measurement is computed in Eq. 4 as

$$C\,(n) = \sum_{m=0}^{M-1} \log_{10}\,(s\,(m))\cos\left(\frac{\pi n\,(m - 0.5)}{M}\right) \tag{4}$$

where n ranges from $n = 0, 1, 2, \ldots, C - 1$, C is the number of MFCCs, $C(n)$ represents the cepstral coefficients.

3.2 *Pitch*

Pitch features comply with the frequency information of the signals from our ear's response [20]. In vocal speech, the passage of air flow from the lungs is normally regulated by vocal cords, resulting in quasi-periodic excitment. A low-frequency oscillation majorly rules the subsequent sound at the end referred to as a pitch. In the non-vocal speech, the air in the lungs shrinks in the vocal tract and turns into a turbulent, noisey arousal. Depending upon the frequency of the sound wave, it is predictable as male or female. It is an essential feature of emotion recognition. Usually, the female pitch have a high pitched wave, then the males. The gender of the speaker is analyzable easily from the signal pitch by using Eq. 5. Pitch uses the auto-correlation method [21].

$$S(n) \simeq \sum_{t=x_1}^{x_2-n} y(l).y(l+n) \tag{5}$$

$S(n)$ implies the correlation of the signal n, l is the index of the signal, and y represents the signal source. Here x_1 and x_2 refer to the frame boundaries.

4 Machine Learning Algorithms

Machine learning is a subset of artificial intelligence which enables a system to attain the knowledge from data rather than through explicit programming. However, machine learning is a complex process for large dataset. As there are many strong algorithms for training data in machine learning, it is then possible to produce more precise models based on that data. This work has used K-NN and decision tree algorithms for training and testing the cepstral and pitch features.

4.1 K-Nearest Neighbors Algorithm

K-Nearest Neighbors Algorithm is a unique form of algorithm that works on the principle of supervised learning. The KNN has a set of related samples falling in a similar class of high probability. In contrast, the general idea behind the KNN algorithm is to choose K nearest neighbors for every test sample case, to predict the sample by using learnt K nearest neighbors. KNN is the non-parametric learning algorithm were no explicit training data are required. This Lazy Learner Algorithm performs well on both classification and regression related issues. During the training phase, it stores the data set, and when new data arrives, it classifies data based on a category that is very similar to new data. Here Eq. 6 uses Euclidean distance (ED) to compute the distance between two different points.

$$\mathrm{ED}(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{6}$$

The range of K value in the K-NN algorithm is still adamant and complicated [22], For example, Lall and Sharma [23] have identified the value of k is $k \simeq \sqrt{n}$ for the training datasets of training samples greater than 100 [24].

4.2 Decision Tree

A decision tree [25] is a form of supervised learning approach,which splits the voiced data spontaneously according to the specific parameter with the help of decision, problem and with the outcomes of each decision.Ambiguity is reduced in decision making, here representation of the tree is entities, called nodes and leaves. Decision tree classifier helps in classifying multi-class classification in a dataset. It undergoes the process of attribute selection with information gain and Gini index. It practices all possible results of a decision tree and traces every route to find a conclusion.

5 Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)

The Ryerson audio-visual database of emotional speech [26] is a speech emotion dataset in the English language. This dataset contains both emotional vocal speech and song for recognition. The dataset consists of emotions of both male and female genders like happy, sad, angry, neutral, disgust, calm, surprised and fearful etc. The propsoed work has used emotional vocal speech signals alone for the proposed feature extraction approach. The sample speech signals of the eight emotions is shown in Fig. 3.

6 Experimental Results

The experiments are done using MATLAB 2019b in Windows 10 Operating System with Intel Core i7 Processor 3.40 GHz with 16 GB build-in RAM. The eight emotions are used for speech emotion recognition system viz Happy, sad, fearful, disgust, neutral, calm, surprise and angry from RAVDESS dataset [26]. The performance of the proposed method on KNN and Decision Tree classifier assessed by using a 5-fold cross-validation approach.

The satistical metrics like Accuracy (A), Precision (P), Recall (R) and F_1-measure has opted for evaluation of performance. Where tp (positive prediction) is true positive, tn (Negative prediction) is true negative, fp (mispredicted as positive) is the false positive, fn (mispredicted as negative) is the false negative. Accuracy (A)

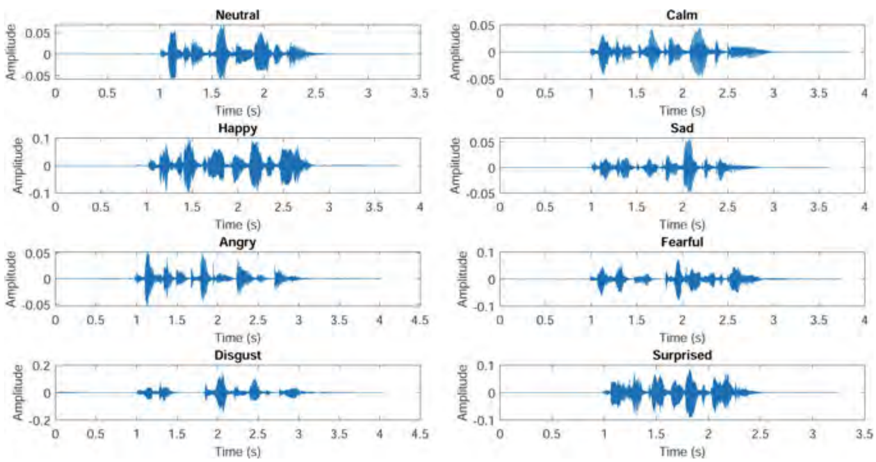


Fig. 3 Sample speech signals of the eight emotions

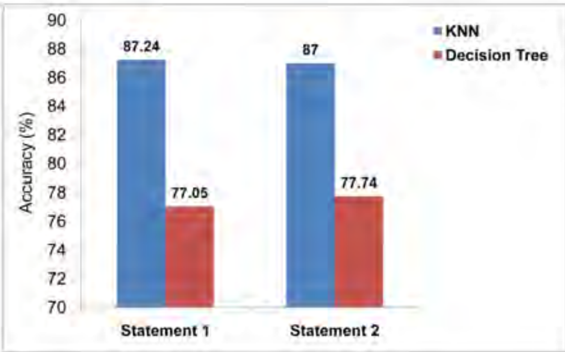


Fig. 4 The performance of K-NN and decision tree classifiers for the two statements "Kids are talking by the door" and "Dogs are sitting by the door" speech

$= \left[\frac{\text{tp+tn}}{\text{tp+fp+tn+fn}}\right]$ ensembles the overall correctness of the emotion recognition system. Precision $(P) = \left[\frac{tp}{tp+fp}\right]$ is the measure of perfection. Identification of emotions in the correct way defined by Recall $(R) = \left[\frac{\text{tp}}{\text{tp+fn}}\right]$ F_1-measure $= 2\frac{P \times R}{P+R}$ gives the mean of both Precision and Recall.

The proposed synthesis method obtained an overall accuracy rate for statements "Kids are talking by the door" and "Dogs are sitting by the door" are 87.24 and 87% for K-NN classifier and the accuracy results for Decision tree are 77.74% and 77.05% respectively. From the experimental results K-NN classifier performs well on the two speech statements when compared to decision tree. The performance graph of the KNN and decision tree for the two statements is shown in Fig. 4.

6.1 *Results on K-NN Classifier*

This section discusses the K-NN [27] and decision tree classifier performance in terms of Accuracy (A), Precision (P), Recall (R), and F_1-measure for the two vocal speech statements. The K-NN and Decision tree performance measures are shown in Tables 1 and 2. From the results, K-NN performs well when compared to decision tree and the corresponding confusion matrix for the eight different emotions using the two statements "Kids are talking by the door" and "Dogs are sitting by the door" are shown in Figs. 5 and 6 respectively.

The main diagonal of the confusion matrix represents the number of instance/ samples that was classified precisely/correctly. Rows represent the emotional class instance, and column represents the speech emotion class predicted by the KNN classifier. The emotions like Happy, Sad, Fear, Disgust, Neutral, Calm, surprise and Angry are classified with greater accuracy. On average, the emotion recognition rate of K-NN classifier in RAVDESS dataset is 87.12%. Here, some of the emotions like neutral and calm are misclassified as sad, since it is tough to distinguish the emotions, and it needs further attention.

Table 1 Performance measures obtained for the KNN classifier

Emotion	"Kids are talking by the door"				"Dogs are sitting by the door"			
	Precision	Recall	Specificity	F-score	Precision	Recall	Specificity	F-core
Angry	0.9227	0.8847	0.9873	0.9033	0.9123	0.8883	0.9847	0.9002
Calm	0.8449	0.8645	0.9826	0.8546	0.8419	0.8723	0.9794	0.8568
Disgust	0.8793	0.8766	0.9812	0.8779	0.8756	0.8731	0.9823	0.8743
Fearful	0.8819	0.8708	0.9824	0.8763	0.8697	0.8738	0.9807	0.8717
Happy	0.8790	0.8715	0.9804	0.8753	0.8767	0.8697	0.9820	0.8732
Neutral	0.8545	0.8695	0.9810	0.8619	0.8451	0.8503	0.9784	0.8477
Sad	0.8209	0.8633	0.9769	0.8416	0.8369	0.8534	0.9802	0.8451
Surprised	0.8780	0.8735	0.9827	0.8757	0.8877	0.8729	0.9840	0.8802
Mean (μ)	0.8702	0.8718	0.9818	0.8708	0.8682	0.8692	0.9815	0.8686

Table 2 Performance measures obtained for the Decision tree classifier

Emotion	"Kids are talking by the door"				"Dogs are sitting by the door"			
	Precision	Recall	Specificity	F-score	Precision	Recall	Specificity	F-Score
Angry	0.8368	0.7757	0.9740	0.8051	0.8311	0.7896	0.9721	0.8098
Calm	0.7502	0.7837	0.9714	0.7666	0.7570	0.7883	0.9669	0.7723
Disgust	0.7615	0.7615	0.9628	0.7615	0.7794	0.7747	0.9670	0.7770
Fearful	0.7735	0.7852	0.9652	0.7793	0.7788	0.8041	0.9672	0.7912
Happy	0.7885	0.7816	0.9657	0.7850	0.7822	0.7719	0.9693	0.7770
Neutral	0.7462	0.7618	0.9668	0.7539	0.7303	0.7545	0.9620	0.7422
Sad	0.7264	0.7676	0.9645	0.7464	0.7532	0.7739	0.9687	0.7634
Surprised	0.7658	0.7470	0.9674	0.7563	0.7983	0.7599	0.9724	0.7786
Mean (μ)	0.7686	0.7705	0.9672	0.7693	0.7763	0.7771	0.9682	0.7765

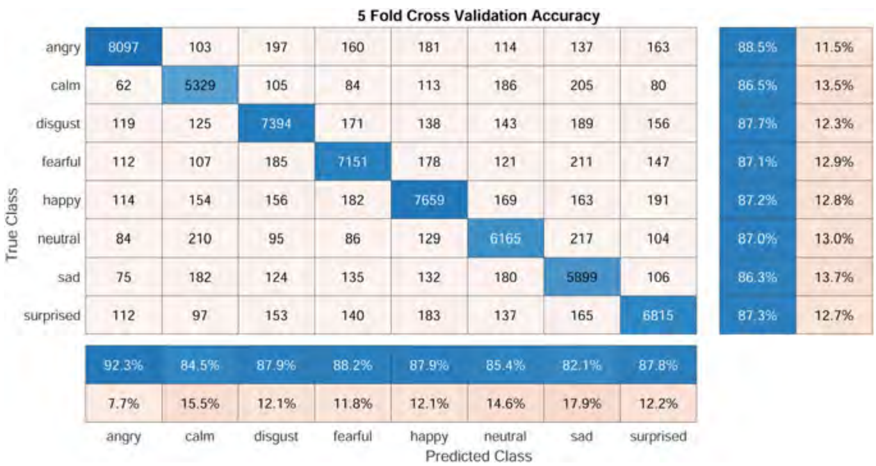


Fig. 5 Confusion matrix obtained for the KNN classifier for Statement 1 "Kids are talking by the door"

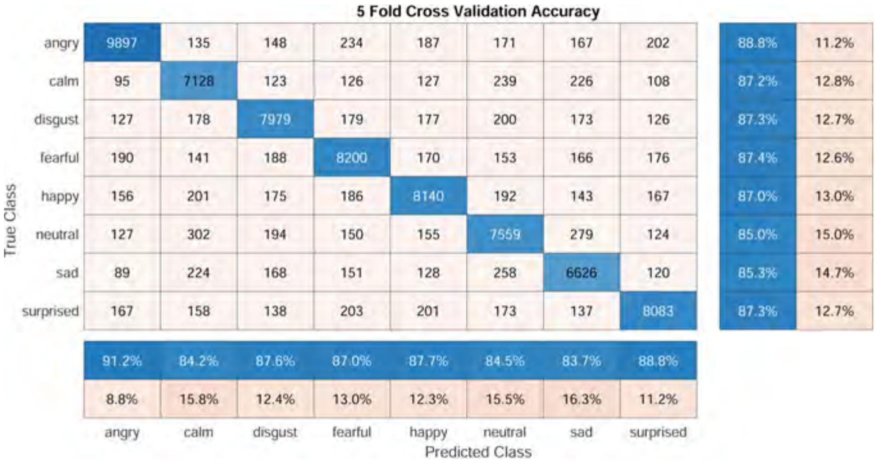


Fig. 6 Confusion matrix obtained for the KNN classifier for Statement 2 "Dogs are sitting by the door"

7 Comparative Study

The results obtained from the proposed work have compared with the state-of-the-art results to show the effectiveness of the proposed technique and comparison is presented in Table 3. Based on the study, in RAVDESS database, Zeng et al. [28] has proposed a multi-task model using GresNets Spectrogram generations, with an accuracy of 64.48%. Bhavan et al. [2] evaluated the recognition of emotion considering Bagged SVM using MFCC derivatives and Spectral centroids and resulted in the best accuracy of 75.69%. Kwon et al. [29] proposes Deep Stride CNN model using Raw and Clean Spectrograms and provides a better accuracy rate of 79.50%. It is observed that our proposed approach tends to increase the recognition accuracy of 87.12% and gives excellent results in the RAVDESS dataset.

Table 3 State-of-the-art results on RAVDESS dataset

Method	Classifier	Feature	Accuracy (%)
Zeng et al. [28]	DNNs	Spectrograms	64.48
Bhavan et al. [2]	Bagged SVMs	MFCC	75.69
Kwon et al. [29]	DSCNN	Spectrograms	79.50
Ours	KNN	Cepstral + Pitch	87.12

8 Conclusion

This research paper proposes an efficient framework for speech emotion recognition using MFCC + Pitch features with K-NN and decision algorithm. Experimental analysis has conducted on vocal speech statements “Kids are talking by the door” and “Dogs are sitting by the door” from RAVDESS emotion dataset with eight different emotions. The proposed approach based on speech emotion on different statements is extracted by using MFCC + Pitch from the vocal speech signals. The extracted features are trained and modelled using K-Nearest Neighbour and Decision tree classifiers for identifying emotions. The experimental results shows that RAVADESS dataset revealed the feasibility of the proposed method with an overall accuracy of 87.12% for K-NN and 77.39% for Decision tree. Further, it has concluded that the experiments based on K-NN performs better than decision tree. The accuracy of various Quantitative evaluations is computed with metrics like precision, recall, accuracy and F-measure. The findings concluded that the system could not be able to distinguish neutral and calm with high precision. Our further research extends to recognize the identification of emotional patterns by manipulating gender in the perceived speech signal.

References

1. Koduru A, Valiveti HB, Budati AK (2020) Feature extraction algorithms to improve the speech emotion recognition rate. Int J Speech Technol 23(1):45–55
2. Bhavan A, Chauhan P, Shah RR et al (2019) Bagged support vector machines for emotion recognition from speech. Knowl-Based Syst184:104886
3. Kim EH, Hyun KH, Kim SH, Kwak YK (2009) Improved emotion recognition with a novel speaker-independent feature. IEEE/ASME Trans Mechatron 14(3):317–325
4. Hasan Md R, Jamil M, Rahman MGRMS et al (2004) Speaker identification using mel frequency cepstral coefficients. Variations 1(4) (2004)
5. Dave N (2013) Feature extraction methods IPC, PLP and MFCC in speech recognition. Int J Adv Res Eng Technol 1(6):1–4
6. Bou-Ghazale SE, Hansen JHL (2000) A comparative study of traditional and newly proposed features for recognition of speech under stress. IEEE Trans speech Audio Process 8(4):429–442 (2000)
7. Liu GK (2018) Evaluating gammatone frequency cepstral coefficients with neural networks for emotion recognition from speech. arXiv:1806.09010
8. Shashidhar G, Koolagudi K, Sreenivasa R (2012) Emotion recognition from speech: a review. Int J Speech Technol 15(2):99–117
9. Liu Z-T, Min W, Cao W-H, Mao J-W, Jian-Ping X, Tan G-Z (2018) Speech emotion recognition based on feature selection and extreme learning machine decision tree. Neurocomputing 273:271–280
10. Pan Y, Shen P, Shen L (2012) Speech emotion recognition using support vector machine. Int J Smart Home 6(2):101–108
11. Chen L, Mao X, Xue Y, Cheng LL (2012) Speech emotion recognition: features and classification models. Digital Signal Process 22(6):1154–1160

12. Schuller B, Rigoll G, Lang M (2004) Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. In: 2004 IEEE international conference on acoustics, speech, and signal processing, vol 1. IEEE, pp I–577
13. Sreenivasa Rao K, Koolagudi SG (2015) Recognition of emotions from video using acoustic and facial features. Signal Image Video Process 9(5):1029–1045
14. Fahad Md, Yadav J, Pradhan G, Deepak A et al (2018) DNN-HMM based speaker adaptive emotion recognition using proposed epoch and MFCC features. arXiv:1806.00984
15. Arunnehru J, Kalaiselvi Geetha M (2017) Automatic human emotion recognition in surveillance video. In: Intelligent techniques in signal processing for multimedia security. Springer, pp 321–342
16. Muda L, Begam M, Elamvazuthi I (2010) Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. arXiv:1003.4083
17. Picone JW (1993) Signal modeling techniques in speech recognition. Proc IEEE 81(9):1215–1247
18. Benesty J, Mohan Sondhi M, Huang Y (2007) Springer handbook of speech processing. Springer
19. Deller JR, Proakis JG, Hansen JHL (2000) Discrete-time processing of speech signals. Institute of Electrical and Electronics Engineers
20. Kurpukdee N, Kasuriya S, Chunwijitra V, Wutiwiwatchai C, Lamsrichan P (2017) A study of support vector machines for emotional speech recognition. In: 2017 8th International conference of information and communication technology for embedded systems (IC-ICTES). IEEE, pp 1–6
21. Selvaraj M, Bhuvana R, Padmaja S (2016) Human speech emotion recognition. Int J Eng Technol 8:311–323
22. Kang P, Cho S (2008) Locally linear reconstruction for instance-based learning. Pattern Recogn 41(11):3507–3518
23. Meesad P, Hengpraprohm K (2008) Combination of kNN-based feature selection and kNN based missing-value imputation of microarray data. In: 2008 3rd International conference on innovative computing information and control. IEEE, pp 341–341
24. Lall U, Sharma A (1996) A nearest neighbor bootstrap for resampling hydrologic time series. Water Resources Res 32(3):679–693
25. Badshah AM, Ahmad J, Lee MY, Baik SW (2016) Divide-and-conquer based ensemble to spot emotions in speech using MFCC and random forest. arXiv:1610.01382
26. Livingstone SR, Russo FA (2018) The Ryerson audio-visual database of emotional speech and song (RAVDESS): a dynamic, multimodal set of facial and vocal expressions in North American English. PloS One 13(5):e0196391
27. Arunnehru J, Yashwanth A Shammer S (2017) Canonical correlation-based feature fusion approach for scene classification. In: International conference on intelligent systems design and applications. Springer, pp 134–143
28. Zeng Y, Mao H, Peng D, Yi Z (2019) Spectrogram based multi-task audio classification. Multimedia Tools Applications 78(3):3705–3722
29. Kwon S et al (2020) A CNN-assisted enhanced audio signal processing for speech emotion recognition. Sensors 20(1):183

Performance Analysis of Deep CNN Assisted Optimized HIV-I Protease Cleavage Site Prediction with Hybridized Technique

Navneet Kaur and Wiqas Ghai

Abstract In recent years, human immunodeficiency virus infection and acquired immune deficiency syndrome (HIV/AIDS) has emerged as a global health issue. The disease is caused by a virus that affects the CD4 cell in the human body that lowers the immune system in the human body. HIV-protease is the agent that replicates itself and affects the CD4 T cells in the human blood. To overcome the problem of replication, inhibitors can be analyzed and designed that can bind the active sites in the proteases. To design efficient protease inhibitors, the knowledge about the specificity of cleavage sites is essential. Several encoding techniques and classifiers have been proposed to study and analyze the active cleavage sites in proteases. This paper proposes a new model and comparatively analyses the performance of Hybridized SVM_Genetic modeling with Deep CNN assisted optimized prediction of Cleavage sites. For optimal tuning of activation functions, two metaheuristic algorithms such as moth search and dragonfly are proposed in this work. The performance of both the methodologies is compared based on different parameters such as accuracy, specificity, F1 score, sensitivity, and NPV. To authenticate the performance of the proposed model, standard data from machine learning algorithms called UCI repository is processed for experimentation. The performance measured is compared with existing available techniques for predicting cleavages.

Keywords Moth search · Dragonfly · Deep convolution neural network · Support vector machine · Acquired immunodeficiency syndrome

1 Introduction

Acquired immunodeficiency syndrome (AIDS) is caused due to the presence of human immunodeficiency virus present in the body. Human immunodeficiency virus

N. Kaur (✉) · W. Ghai
RIMT University, Mandi Gobindgarh, Punjab, India
e-mail: Bawa.navneet@gmail.com

W. Ghai
e-mail: ghaialpha@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_40

(HIV) shows symptoms of the lower body's immune system and leading to the death. The first case of HIV-1 was reported in 1981 in the center for disease controlling the USA. It is been 39 years and still the human immunodeficiency virus (HIV-I) is a global health issue. According to the World Health Organization, 38 million people living in the world are infected with infectious disease by the end of 2019 and have claimed 33 million lives so far. However, with concerted efforts of various researchers, a new technique antiretroviral therapy is devised. This kind of treatment of HIV-I smacks and suppresses the action of HIV-I proteases [1, 2]. The first reported antagonist of HIV-1 proteases was discovered in 1987. When HIV-AIDS enters the human body and targets CD4 T cell present in the blood, and replicates itself producing its copies in the blood, which lowers the body's immune system which is shown in Fig. 1(L-R) the long viral proteins are being cut by proteases, to prevent the cleavage protease inhibitors can be developed, but the problem lies in predicting the cleavage site in various viral proteins. The various antagonist and inhibitors were proposed like Saquinavir and Nelfinavir have been proposed by FDA. The basic mechanism lies in the fact is to replace the peptide linkage consisting of –NH–CO with hydroxyl ethylene group (–CH_2–CH (OH)–) which proteases is difficult to cleave [3].

To develop a standard benchmark inhibitor for HIV-proteases, so the study of HIV-1 proteases cleavage specificity is a major concern to researchers. Proteases get attached viral protein and replicate itself to produce a large number of viral proteins.

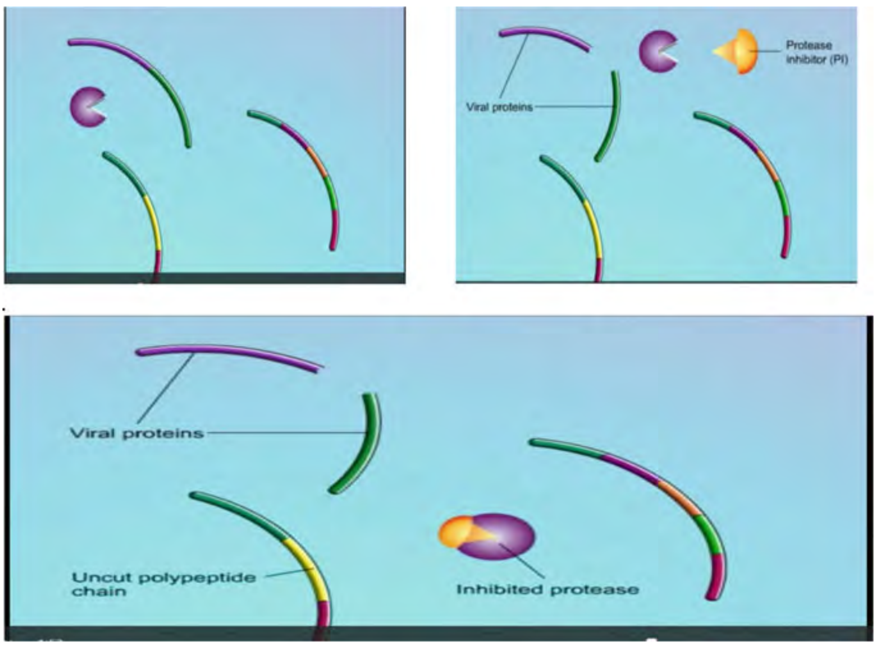


Fig. 1 (L-R) various stages of protease effecting CD4 T cells and viral protein replicating itself

2 Literature Survey

Extensive work has been done by the researchers and organizations to explore the potential of data provided in the repository for analysis, and resulted in substantial classification accuracy. Singh et al. in 2018 proposed an evolutionary-based ensemble framework, genetic algorithm to attain the optimally coupled feature and classifier sequence. The simulation showed significant enhancement of the proposed model over the conventional models because of average accuracy, sensitivity, and precision. The average accuracy achieved without cross-validation is 87.14% [4]. Fatehi and sadegi in 2018 introduced a novel technique in spatial and structural features are taken into account, SVM and genetic programming have been used for modeling, and the prediction accuracy achieved is 91.1% [5]. In 2018, Singh et al. developed an optimal formation of encoding-classifier pair selection by an evolutionary algorithm. Natural selection is based on several learners and optimal data-learner mapping. The accuracy achieved in this case was quite high [6]. In 2009, Ogul constructed a model entirely based on a generalization of variable order Markov chains (VOMC) for protein sequence; the model developed predicted cleavability by some proteases. The novel method, called variable context Markov chains (VCMC), tries to analyze the context equivalence based on the similarities of specific amino acids. The result shows that it outperforms the performance of existing models in terms of prediction accuracy on a benchmark dataset [7]. In 2016, Singh et al. proposed a prediction method in which certain properties like sequential, structural, and chemical features are added in certain machine learning methods. The main features are extracted by various encoding schemes and as input to decision trees, regression and ANN, a three-way approach was applied for prediction which achieved an accuracy of 80.0–97.4% [8]. In 2005, Lummini and Nanni constituted an idea from various machine learning algorithms to develop a knowledge base [9]. Song et al., in (2019) proposed a web server for the prediction of cleavage sites by many different proteases, using SVR with a combination of different features. Bi-profile Bayesian was used for feature extraction, and the Gini score was used for calculation. The data was used from a larger dataset from Schilling and other sources published data for cleaving whole proteins [10].

3 Material and Methods

Deep convolution neural network (DCNN) is employed for HIV-1 protease cleavage specificity prediction. It is used for classification and two metaheuristic approaches moth search and dragonfly is used for the optimization of the results that predicts the specificity, accuracy, sensitivity of the input data. The proposed model consists of following different phases such as *s*electing the dataset from the repository, data simplification, and preprocessing, feature selection model based on discrete wavelet transform, training and classification of the dataset using deep convolution neural network, and finally, optimal tuning of activation function using hybridized metaheuristic algorithms (moth search and dragonfly).

A. **Selecting Dataset from Repository**

Collecting and selection of data from the domain that is authentic, informative, and useful are very challenging, especially when you are dealing with biological data. In this study of the cleavage site, the dataset is taken from the UCI machine learning repository is a collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms. The archive was created as an FTP archive in 1987 by David Aha and fellow graduate students at UC Irvine. In the proposed work, 4 datasets are used Data-746, Data-1625, data-Schilling, Data-Impens.

Each input query consists of two parts, first eight-letter alphabetic strings that represent 8 different amino acids, −1 and 1 denote the cleave and non-cleave sites in octamer. The character string that are allowed for encoding consists of {A,B,N,D,C,Q,E,H,I,L,K,M,F,P,S,T,W,Y,V} each denoting different amino acids (Table 1).

The encoding of an octamer sequence requires much attention and is necessary for interpretation by different machine learning approaches. Different encoding techniques are designed by researchers in the past like orthonormal encoding (OE), consisting of a 20-bit vector and but the main drawback is information loss. Another encoding technique includes combining BLOSUM50 and BLOSUM62 matrices with orthonormal encoding [9], Taylor Venn diagram encoding is also experimented by Zvelebil et al. The proposed encoding model in this research work consists of orthonormal encoding that combines the structural and chemical features of every amino acid. The features polarity, acidity, hydropathy index, aliphatic, aromatic, proline, etc., are taken and their values are normalized using the following formula

$$nv_i = \left(\left(\frac{r_i - \min(r_i)}{\max(r_i) - \min(r_i)} \right) \times 2 \right) \tag{1}$$

r_i and nv_i denote the original and new value in the range $1 < nv_i < 152$ (Fig. 2).

B. **Feature Selection model based on wavelet transform**

Different methodologies are applied by various researchers for feature extraction like SVM, genetic programming, ANN, KNN, etc., but the technique applied in the

Table 1 Dataset with their cleavage specificity

Sr.No	Dataset	No. of cleavages	No. of non-cleavages
1	Data_746	401	345
2	Data_1625	374	1251
3	Data_schlling	434	2838
4	Data_impens	149	798

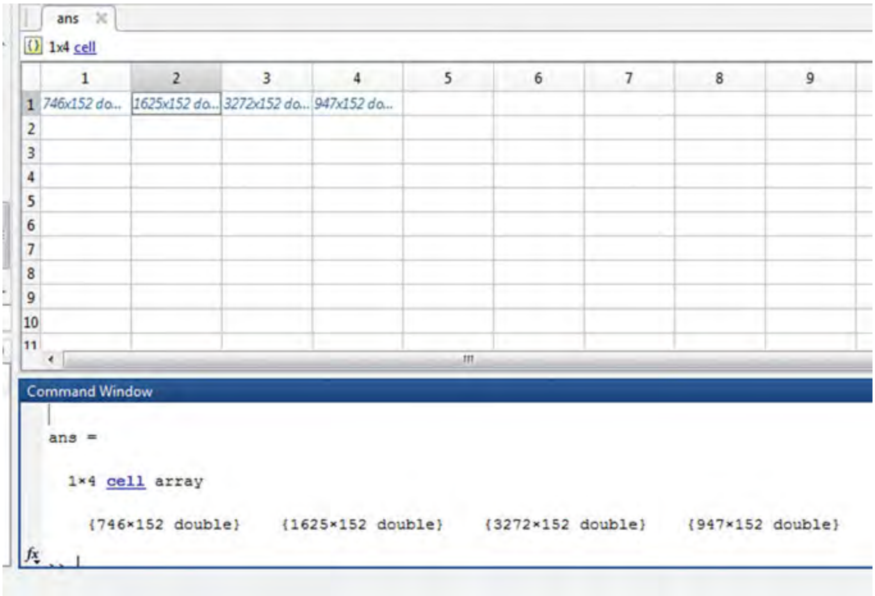
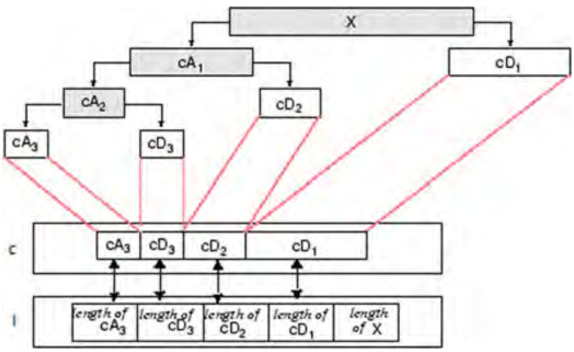


Fig. 2 Snapshot of MATLAB showing feature preprocessing

proposed model for extracting features is wavelet-based. There are few reasons, this technique is applied including precise details of the dataset or signal that can be extracted easily, it provides a good approximation of the result and can recognize breakpoints, trends, and discontinues. Wavelet decomposition is a wavelet transform or decomposition of the signal or data into 1 dimension or 2 dimension. Suppose a signal x is divided into n levels using different wavelets. The output decomposition structure contains the wavelet decomposition vector c and the bookkeeping vector l, which maintains the coefficients by levels. The structure is shown in the decomposition diagram (Fig. 3).

Fig. 3 Representation of 1D wavelet transform for extracting features



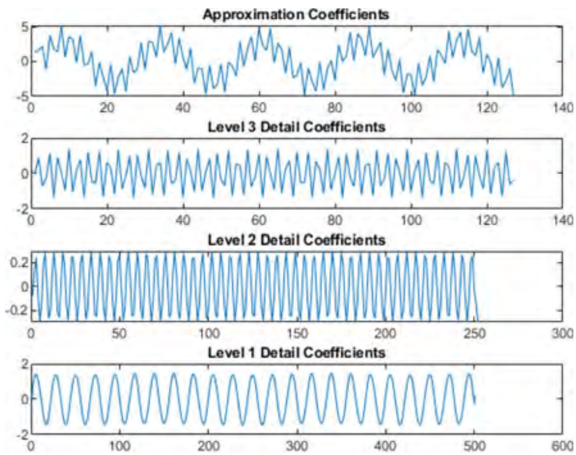


Fig. 4 Wavelet transformation after preprocessing of the data

$$[\mathrm{c}, \mathrm{l}] = \mathrm{wavedec}(x, n, \mathrm{wname}) \tag{2}$$

The wavelet decomposition equation is given by

$$[\mathrm{c}, \mathrm{l}] = \mathrm{wavedec}(x, n, \mathrm{LoD}, \mathrm{HiD}) \tag{3}$$

From Eqs. 1–3, the wavelet transformation of dataset is shown in Fig. 4.

C. **Training and Classification of the dataset using deep convolution neural network**

After extracting features, using wavelet transform the output extracted for the previous phase is used to train the deep convolution neural network. Classification involves a supervised learning technique, classifier learns through various attributes present in the input data, the classification is done in two steps first is training the data and then predicting the samples. The classifier used in research work is a deep convolution neural network. Classification is done by training the known sample set (UCI repository), the performance of the model is measured in terms of accuracy, which is achieved while predicting the unknown sample. CNN is divided into two parts: feature detection layers and classification layers. The pseudocode for deep CNN training is shown in Fig. 5.

D. **Optimization of activation function using moth search and dragonfly algorithm**

After training and optimization, the next step involves optimization of the output after the classification phase. Two metaheuristic algorithms are used to optimize the performance of deep convolution neural network. Moth search algorithm works on the phenomenon of phototaxis and Levy walk. Moths tend to fly near to the source of light, this behavior is known as phototaxis, Levy walk are one of the most important style of moths to fly in natural environments. Many animals and species,

```
Matlab code for Training dataset using DCNN


function[acc,net] = Deep_CNN(train_data,train_lab,act_fn,sol)

if act_fn == 1
   a = reluLayer;
elseif act_fn == 2
   a = leakyReluLayer;
elseif act_fn == 3
   a = clippedReluLayer(10);
elseif act_fn == 4
   a = eluLayer;
elseif act_fn == 5
   a = tanhLayer;
else
   a = preluLayer;
end
layers = [
   imageInputLayer([size(train_data,1) size(train_data,2) 1])
   convolution2dLayer(sol(1),sol(2),'Padding','same')
   convolution2dLayer(sol(1),sol(2),'Padding','same')
   convolution2dLayer(sol(1),sol(2),'Padding','same')
   convolution2dLayer(sol(1),sol(2),'Padding','same')
   a
   fullyConnectedLayer(384) % 384 refers to number of neurons in next FC
hidden layer
   fullyConnectedLayer(384)
   fullyConnectedLayer(length(unique(train_lab)))
   softmaxLayer
   classificationLayer
   ];
options = trainingOptions('sgdm',...
   'MaxEpochs',100, ...
   'Verbose',true);
labels = categorical(train_lab);
net = trainNetwork(train_data,labels',layers,options);
predictedLabels = classify(net,train_data)';
```

Fig. 5 Pseudocode for deep CNN

like Drosophila, fly in the form of Levy flight that can be rounded to a power law. It is basically distributed over varied scale with the feature of exponents close to 3/2 [11, 12]. It is one of the commonly used metaheuristic optimization algorithm based on swarm intelligence. The main approach used in this is based on static and dynamic behaviors of swarming in dragonflies in the environment. The two most important phases of optimization are exploration and exploitation which are configured by modeling the behavior of dragonflies, i.e., navigating, hunt for food, protecting themselves from enemies [13, 14] (Fig. 6).

```
Matlab code of moth search & dragonfly

disp('moth search')
        [bestfit,fitness,bestsol,time] = MS(initsol,fname,xmin,xmax,itermax);
        Mso{i}.bf = bestfit; Mso{i}.fit = fitness; Mso{i}.bs = bestsol; Mso{i}.ct = time;
save Mso Mso

        disp('Dragon fly')
        [bestfit,fitness,bestsol,time] = DA(initsol,fname,xmin,xmax,itermax);
        Da{i}.bf = bestfit; Da{i}.fit = fitness; Da{i}.bs = bestsol; Da{i}.ct = time; save
Da Da

        disp('moth search_dragonfly')
        [bestfit,fitness,bestsol,time] = MS_DA(initsol,fname,xmin,xmax,itermax);
        Prop{i}.bf = bestfit; Prop{i}.fit = fitness; Prop{i}.bs = bestsol; Prop{i}.ct =
time; save Prop Prop
```

Fig. 6 MATLAB code of moth search and dragonfly

4 Results and Discussion

The experimental setup deployed for the intel® core i5 CPU@ 1.8 GHz with 5 GB RAM, MATLAB 2018(9.4.0) is used that combines a desktop environment tuned for iterative analysis and design processes the expression matrix and array mathematics directly. Dataset is used for training and testing; i.e., it is taken from the UCI repository. The performance of the proposed CNN model is compared with the existing state-of-the-art models, work done by researchers is evaluated based on accuracy, specificity, precision, FPR, F1 score, NPV, MCC, FDR, etc.

1. Accuracy:T_p denotes correctly predicted cleaved and T_n denotes correctly predicted non-cleaved sites and F_p and F_n number of incorrectly predicted cleaved and incorrectly predicted non-cleaved sites [13, 15].

$$\text{Accuracy} = \frac{T_\mathrm{p} + T_\mathrm{n}}{T_\mathrm{p} + T_\mathrm{n} + F_\mathrm{p} + F_\mathrm{n}}$$

The accuracy achieved by various datasets is shown in Table 2.

Average accuracy achieved using a benchmark dataset in the proposed methodology, i.e., deep CNN assisted HIV-1 protease cleavage site prediction is 93%, which is quite high as compared to other states-of-the-art techniques (Fig. 7 and Tables 3, 4, 5).

2. Sensitivity: It specifies the likelihood of false positives. It is defined as the True positive rate given by the following formula. In the proposed algorithm, the average sensitivity achieved is 95.53% which is comparatively better as compared to other processes. The sensitivity is given by the formula

Table 2 Performance classifier accuracy achieved with various techniques (existing and proposed)

Technique dataset	DCNN	SVM_GA	MSO	DA	DA_MSO (proposed method)
Data746	84.375	81.696	89.286	90.625	92.411
Data1625	93.634	81.696	92.197	93.84	94.661
DataShilling	92.974	81.696	93.177	93.279	93.89
DataImpens	88.732	81.696	90.493	88.028	91.197

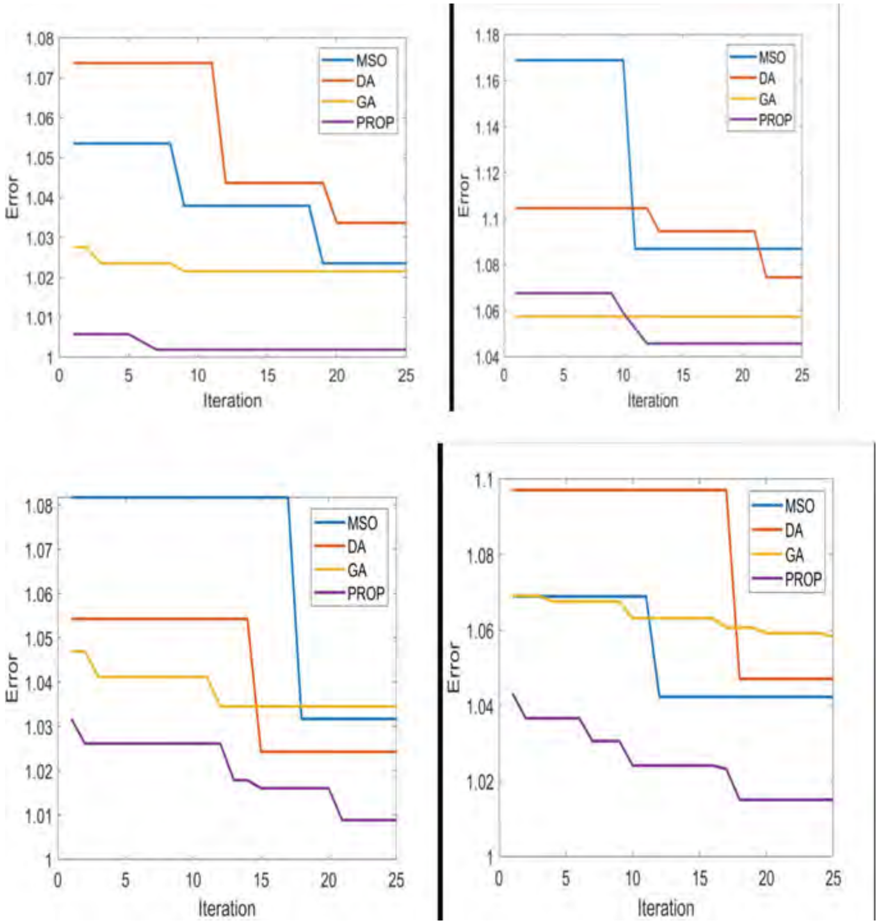


Fig. 7 (L-R) performance parameters (convergence data) showing—accuracy, precision, sensitivity, specificity

Table 3 Performance classifier sensitivity achieved with various techniques (existing and proposed)

Technique dataset	DCNN	SVM_GA	MSO	DA	DA_MSO(proposed method)
Data746	78.571	76.531	80.612	89.796	92.857
Data1625	95.137	76.531	93.658	95.137	95.983
DataShilling	95.982	76.531	94.834	95.637	96.556
DataImpens	93.878	76.531	94.286	91.837	96.735

Table 4 Performance classifier precision achieved with various techniques (existing and proposed)

Technique dataset	DCNN	SVM_GA	MSO	DA	DA_MSO (proposed method)
Data746	88.889	85.714	96.032	91.27	92.063
Data1625	42.857	85.714	42.857	50	50
DataShilling	69.369	85.714	80.18	74.775	72.973
DataImpens	56.41	85.714	66.667	64.103	56.41

Table 5 Performance classifier specificity achieved with various techniques (existing and proposed)

Technique dataset	DCNN	SVM_GA	MSO	DA	DA_MSO (proposed method)
Data746	84.615	80.645	94.048	88.889	90.099
Data1625	98.253	80.645	98.226	98.468	98.482
DataShilling	96.092	80.645	97.406	96.748	96.556
DataImpens	93.117	80.645	94.672	94.142	96.556

$$\text{Sensitivity} = {}^{T_\text{p}}\!/_{T_\text{p}} + T_\text{n}$$

3. Precision: Precision is defined as the fraction of relative instances among the retrieved instances. It is defined as the number of true positives divided by the number of true positive plus the number of false positives. The average precision achieved in the proposed methodology is 67.8%. The formula for sensitivity is given below

$$\text{Precision} = {}^{T_\text{p}}\!/_{T_\text{p}} + F_\text{n}$$

4. Specificity: It specifies the likelihood of false negatives. It is also defined as a True Negative rate. The average specificity achieved in the proposed methodology is 95.42%. Sensitivity is given by the formula

$$\text{Specificity} = {}^{T_\text{n}}\!/_{T_\text{n}} + F_\text{p}$$

5 Conclusion

Deep convolution neural network hybridized with moth search and dragonfly is one of the efficient tools in machine learning and optimization. The datasets are classified and trained to achieve higher accuracy with optimization techniques. HIV-protease cleavage site prediction model proposed here will assist the researcher in predicting cleavage specificity with greater accuracy as compared to other methods and automated decision support systems are one such method in medicine. The proposed will give a boost to the researcher with user-friendly, fast, and robust methods for cleavage specificity.

6 Future Scope

In this paper, data considered is replicated at some places and is very limited, replicated data should be eliminated before processing and also fast and efficient encoding technique can be devised in the future that can achieve higher accuracy in the prediction of cleavages in proteomics.

References

1. Brik A, Wong C-H (2003) HIV-I protease: mechanism and drug discovery Org Biomol Chem 1(1):5–14
2. World Health Organization. http://www.who.int/gho/hiv/en/
3. De Clercq E (2009) The history of anti retro viral: key discoveries over the past 25 years. Med Virol 19(5):287–299
4. Singh et al (2019) Clean: evolutionary based ensemble framework for realizing transfer learning in HIV-1 Protease cleavage sites prediction. Appl Intell 49:1260–1282
5. Fathi et al (2018) A genetic programming method for feature mapping to improve prediction of HIV-1 protease cleavage site. Appl soft Comput 72:56–64
6. Singh et al (2018) Evolutionary based optimal ensemble classifiers for HIV-1 protease cleavage sites prediction. Expert Syst Appl 109:86–99. https://doi.org/10.1016/j.eswa.2018.05.003
7. Ogul et al (2009) Variable context Markov chains for HIV protease cleavage site prediction. Bio Syst 96(3):246–250. https://doi.org/10.1016/j.biosystems.2009.03.001
8. Singh O, Su ECY (2016) Prediction of HIV-1 protease cleavage site using a combination of sequence, structural, and physicochemical features. BMC Bioinform 17(Suppl 17):478. https://doi.org/10.1186/s12859-016-1337-6

9. Nanni L, Lumini A (2008) Using ensemble of classifiers for predicting HIV protease cleavage sites in proteins. Amino Acids 36(3):409–416. https://doi.org/10.1007/s00726-008-0076-z(2008)
10. Jiangning Song Hao Tan,Andrew J. Perry,Tatsuya Akutsu,Geoffrey I. Webb,James C. hisstock, Robert N. Pike.: PROSPER: An Integrated Feature-Based Tool for Predicting Protease Substrate Cleavage Sites. Briefings Bioinform 20:638–658
11. Gai-Ge Wang (2016) Solar Moth search algorithm: a bio-inspired metaheuristic algorithm for global optimization problems. Memetic Comput 10:151–164
12. Li Z, Zhou Y, Zhang S, Song J (2016) Lévy-flight moth-flame algorithm for function optimization and engineering design problems. Math Probl Eng https://doi.org/10.1155/2016/1423930
13. Mirjalili S (2016) Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. Neural Comput Appl 27:1053–1073. https://doi.org/10.1007/s00521-015-1920-1
14. Mafarja MM, Eleyan D, Jaber J, Hammouri A, Mirjalili S (2017) Binary dragonfly algorithm for feature selection. In: 2017 international conference on new trends in computing sciences (ICTCS). https://doi.org/10.1109/ictcs42043
15. Rahamn S (2019) Dragonfly algorithm and its applications. Appl Sci Surv

Artificial Neural Network for Identification and Classification of Natural Body Marks

Dayanand G. Savakar, Danesh Telsang, and Anil Kannur

Abstract Natural and artificial body marks like mole and tattoos are used to identify the victims, such as suspected, and unidentified bodies like in mass death in a plane crash and the tsunami it is a very complex situation to identify the body; in recent years, classification and identification have taken a lot of attention. This paper presents the classification and identification of natural and artificial body marks like mole and tattoo. Active contour segmentation is used to segment the image. There are 28 features extracted from each mole and tattoo image, where 18(color features), 4(texture features), 6(shape features). The artificial neural network is used to classify natural and artificial body marks, and classification accuracies obtained 88.7%. The designed algorithm works based on the features that are being extracted. Several different forms of the process exist to notify the different forms of the naturally identified body marks. The designed and proposed algorithm within this paper incorporates such kind of techniques to identify the natural and artificial body marks.

Keywords Artificial neural network · GLCM—gray-level co-occurrence matrix · BPNN—backpropagation neural network

1 Introduction

Moles are little sores in the skin. They are normally tanned and found on any portion of the body that is chest, face, hands, leg, and so on. A few moles are a lot darker, and some others are skin shaded. They might be harsh, level, raised, round, and

D. G. Savakar · D. Telsang (✉)
Department of Computer Science, Rani Channamma University, Belagavi, Karnataka, India
e-mail: dtelsang@gmail.com

D. G. Savakar
e-mail: dgsavakar@gmail.com

A. Kannur
Department of Computer Science & Engineering, Rajarambapu Institute of Technology, Islampur, Maharashtra, India
e-mail: anilkannur1978@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_41

oval shape. Moles are normally found in individuals who are more presented to the sun instead of less uncovered [1]. Sun consumes are not moles [2]. Identification of the several different marks over the body is an application that acts as a major area mainly in the case of the medical field. In case of mass death due to floods, plane crash, tsunami, and earth quick, it is very tough to identify the person, so the victim identification bases upon their body mark like mole (natural body mark) and tattoo (artificial body mark) are useful to identify the victim [3, 4]. Every person has its identification mark like birthmark (mole), as well as if he has an artificial body mark like a tattoo [5]. Artificial body marks resemble tattoo is a type of body modification, made by embeddings permanent ink into the dermis layer of the skin to change the shade tattoo. A tattoo may be found in any part of the body that is leg, shoulder, face, hand. Promising ongoing work has established the achievability of figuring out mole and tattoo. For about 500 years people have been using steps to identify and represent themselves using tattoo [6]. At first, the tattoo was limited to certain groups, such as motorbike riders, sailors, and criminal groups, but in the modern world, it is natural to have tattoos, as about 36% of people in the world have at least 1 tattoo. For a person, an identification tattoo is a useful tool in forensic applications.

The layout of the paper is organized as follows, Sect. 2 describes the survey report of various methods. Section 3 defines the problem statement. Section 4 explains the methodology adopted for the proposed method. Section 5 depicts the results and discussions. Finally, Sect. 6 concludes the paper with future scope of the research.

2 Literature Review

The backpropagation neural network (BPNN) is used for the ANN to deploy the process of recognition and classification on almost the same images of food grain like cumin seeds, fennel seeds, mung beans, black gram, finger millet, mustard, soybeans, and black-eyed beans. The color and texture features are selected for classification, considering color, and texture 18 and 27 features are extracted, respectively, and combined color and texture 4 and 5 features are extracted [7]. Automatic detection of melanoma skin cancer using texture analysis, gray-level co-occurrence matrix (GLCM) used for feature extraction, and used multilayer perceptron network. It is a feed-forward network and author set two types of classifier, namely automatic MLP and will be traditional MLP [8].

The different neural network models are used, namely Elman's network, cascade network, and feed-forward network, watershed segmentation is used to identify the area and equivdiameter features of diverse seed varieties like lentils, wheat, redgram, groundnut, rice, bengal gram, jowar, and metagi, here 11(includes 9 area, color, and equivdiameter) features, 18 only color and 20 features are extracted, in which 18 are colors and 2 are boundaries. Elman's network remains time-consuming in the training process, and also the cascade-forward network consumes increased time in training; but it greatly reduces the memory size and feed-forward network takes less time to train accompanied with reduced memory size when compared with Elman's and cascade-forward networks [9].

To identify melanoma skin cancer, firstly edge detection technique is used to has been segmented melanoma area, then features are extracted like asymmetry, border, color, diameter (ABCD). ANN is used to classify melanoma, and finally, the backpropagation algorithm is used to identify melanoma skin cancer area [10]. Active contour and CBIR method were implemented for segmenting tattoo; for texture analysis, they used haar wavelet decomposition, and for color representation, hue-saturation-value histograms are used at the end and they got result using their new approach glocal(glocal-local) image feature methods to test the dataset [11].

The author discussed the automatic skin cancer detection system, and they used the thresholding method for segmentation and 2D wavelet transform have used for the feature extraction, and in this proposed system, they used ANN classifiers with feed-forward multilayer network and backpropagation algorithm was used for training [12].

The significant impact caused by foreign bodies in the identification and classification of the largest food grain images like rice, jowar, wheat, groundnut, and green gram in the variety of grains, plant leaves, stones, pieces of stems, weed, soil lumps, and other types of whole grains for the sake of sorting here the author used the color and texture features which are selected for classification, considering color and texture 18 and 27 feature are extracted, respectively, and combined color and texture 4 and 5 features are extracted in this work feed-forward ANN model and backpropagation algorithm are used in the training process [2, 13].

3 Problem Statement

The identification of natural and artificial body marks remains one among the challenging tasks present in image processing. Identification of criminals is a challenging job; in recent years, one of the approaches to identify the criminals based on the natural and artificial body mark like mole and tattoo. In day to day, life tattoo is attracted to everyone, some of them have a tattoo on hand, leg, neck, etc., and everyone has at least one mole in their body. It is very difficult to identify the body where mass death occurred in the case of tsunami and plane crash in that time body marks like mole and tattoo will help to identify the person. To overcome this problem, proposed a method to identify the criminals and mass death persons, the artificial neural network has been used. The proposed method steps are shown in Fig. 1.

4 Methodology

This section presents a detailed description of image acquisition, image preprocessing, segmentation, and feature extraction; finally, an artificial neural network is

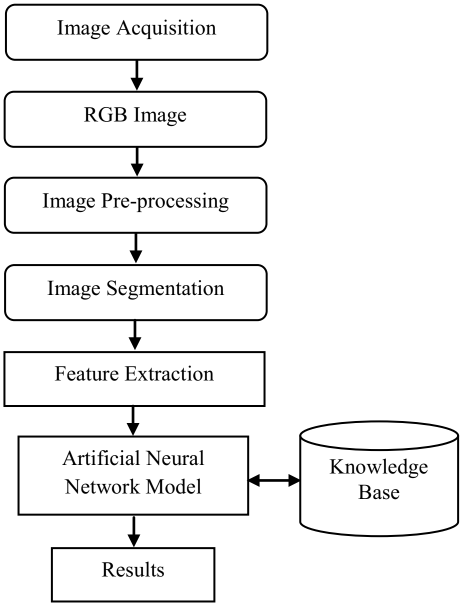


Fig. 1 Proposed system architecture

used to classify the natural and artificial body marks as explained in the subsequent section.

4.1 *Image Acquisition and Preprocessing*

It is a process of capturing an image from a camera, and it is a first step in the workflow sequence. The images are captured by canon 1300D camera with 1.5X to 10X lens zooming capacity and keeping object and the camera in a fixed position with maintaining clock word distance between them and some of the images were collected from an online source [14]. In the preprocessing stage, unnecessary noise is removed from the image and resized too [300 400]. The DULL RAZOR Software [15] is used to alteration, and it removes unnecessary hairs from images to improve the quality of image; it is a medical imaging software and it uses an algorithm the same as that of average filtering. The filtering technique also plays a major role in image processing to upgrade the images, sharpening, edge detection, and noise reduction. For convolution operation, spatial filtering is used as shown in Eq (1).

$$S(x, y) = \sum_{m=-M/2}^{M/2} \sum_{n=-N/2}^{N/2} h(m, n) f(s - m, y - n),$$

where $h(m, n)$ is the Gaussian filtering mask of size $M \times N$.

4.2 Segmentation (Active Contour)

Active contour is a refinement of an object boundary. It will get a curve formed by connecting their edged points, so there is no clear cut boundary for that need to connect those edge points to formed by a curved, then set of connected points, which moves to minimize a specified energy function, there is also another name for this one called as a snakes method, the contour over here is something which can flow along with the different perspectives on the image itself and then it became down to convergence, and from this particular attribute of the contour itself, it gets its name called as an active contour. The model is elected as an active model for the segmentation process. The first step, it should have an image which is extracted from its background, then, will take it as energy function like gray-level value, gradient and the initial step is segment a boundary by a general segmentation technique object with boundary has been found. The next step refines that boundary with the technique used to wiggle the snake in this it compares the pixel on each point on the boundary with energy calculated for the point in its neighborhood, it moves the boundary to neighborhood point that has the lowest energy and operates once on all points on the boundary after that repeat the iteration until it goes no for the movement the process will get stopped. The advantages of the active contour are computational efficiency and relative simplicity, and disadvantage is decision criteria worship complexity. Active contour is used in various field of segmentation based on the application. The mole and tattoo images are segmented with several segmentation techniques in that active contour has given better results as compared to the other methods [16]. Hence, in this work active contour is used for segmenting images as shown in Fig. 2.

Algorithm 1: Active contour (Region-based)-based image segmentation

Input:RGB image
Output: Color Segmented image
Start:
Step1: Read RGB image
Step2: Convert RGB to gray
Step3: Create a mask
Step4: Initialize number of iteration
Step6: Cover the object of an image using a region using a mask value based on the number of iteration.

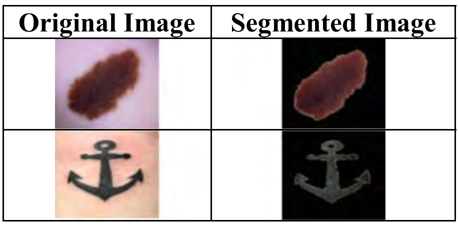


Fig. 2 Samples of the original image and segmented image using active contour segmentation technique [2]

Step5: Set, background = 0 and foreground = 1, (covered area of the an image is 1)
Step7: Segment the covered area of an image
Step8: Reshape to the original image.
Stop

4.3 Feature Extraction

The image which is segmented by active contour is introduced to perform feature extraction (refer Table 1) in the process of identifying different body marks, which include natural mark like mole and artificial mark like a tattoo. The color features are selected because the mole and tattoo are representing different colors. The texture feature also extracted from the images based on the surface area of the image. Finally, shape features are extracted to identify the mole and tattoo images based on shape area and shape solidity, etc. From each image 28 features are extracted and stored as a feature vector. Table 1 represents the total number of features like color features (18), texture features (04), shape features (06).

The detailed description of color, texture, and shape futures are discussed in the subsequent section.

4.3.1 Color Feature Extraction

Color image consists of several components like hue, saturation, value, red, green, and blue. Three features are selected from each component like mean, variance, and range. Based on the color of the mole and tattoo image, eighteen color features are extracted from each image.

Algorithm 2: Color Features Extraction

Input: color image
Output: Eighteen Color features
/Start
Step 1: color image separate from the RGB components.
Step 2: Established HIS components from RGB Components using Eqs. 2–7
Step 3: From each RGB and HSI component extracted range, variance, and mean.
Stop.

$$H = \cos^{-1}\left\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{1/2}}\right\} \tag{2}$$

Table 1 List of color, texture, shape, features

Sl.No	Color features
01	Red mean
02	Red variance
03	Red range
04	Green mean
05	Green variance
06	Green range
07	Blue mean
08	Blue variance
09	Blue range
10	Hue mean
11	Hue variance
12	Hue range
13	Saturation mean
14	Saturation variance
15	Saturation range
16	Value mean
17	Value variance
18	Value range
Texture feature	
19	Texture contrast
20	Texture correlation
21	Texture energy
22	Texture homogeneity
Shape feature	
23	Shape area
24	Shape convex area
25	Shape eccentricity
26	Shape solidity
27	Shape filled area
28	Shape equivdiameter

$$S = 1 - \frac{3}{(R + G + B)}[\mathrm{Min}(R, G, B)] \tag{3}$$

$$I = \frac{1}{3}(R + G + B) \tag{4}$$

$$\text{Mean} \quad \mu = \sum_{x,y} x p(x, y) \tag{5}$$

$$\text{Variance} = \sum_{x,y} (x - \mu)^2 p(x, y) \tag{6}$$

$$\text{Range} = \text{Max}(p(x, y)) - \min(p(x, y)) \tag{7}$$

4.3.2 Texture Feature Extraction

It refers to characteristics appearance of an object given by the density, shape, size, arrangement; the proportion of its elementary parts of natural and artificial body marks in a preliminary stage. The analysis of the mole and tattoo image thought the texture analysis, then extracted four features from each mole and tattoo image like contrast, correlation, energy, homogeneity.

Algorithm 3: Extraction of textural feature

Input: Original Image RGB
Output: 4 Texture Features
Start
Step1: 24-bit input color image separate from RGB components, obtain the Gray-level co-occurrence matrices (GLCM).
Step 2: Compute the Co-occurrence Matrix
Step 3: GLCM features (Texture features) Energy, Contrast, Homogeneity, correlation
Stop.

$$\text{Contrast} = \sum |x - y|^2 p(x, y) \tag{8}$$

Contrast refers to the degree of intensity of the pixels and its neighbor, which is calculated by using the difference in the color & brightness of the object and objects present within the similar field view is given by Eq. 8.

$$\text{Correlation} = \sum_{x,y}^{\sigma_x \sigma_y} [(xy)P(x, y)] - \mu_x \mu_y \tag{9}$$

μ_x and μ_y: mean; and σ_x and σ_y: standard deviation.

Correlation is referred as the degree of similarity between the data that relate the processing of spatial domain with the frequency-domain as given in Eq. 9.

$$\text{Energy} = \sum_{x,y} p^2(x, y) \tag{10}$$

Here, energy is referred as the degree of pixel pair repetitions at the extent, where it also leverages the image uniformity as represented by Eq. 10.

$$\text{Homogeneity} = \sum_i \sum_j \frac{P(i, j)}{1 + |i - j|} \tag{11}$$

{P (d, θ) (i, j)} (i, j)	represents the probability in occurrence of gray-level pair separated by a specified distance *d* at angle θ as given by Eq. 11.

4.3.3 Extraction of Shape Feature

The mole and tattoo features are extracted by different measures like shape area, convex area, eccentricity, solidity, filled area, equivdiameter. The area of the object present in the image is measured by shape area and measured area of the particular object as consider as a feature, stored in the feature vector. The other measures also extract the feature of the mole and tattoo images based on their properties. There are six features are extracted from each image.

Algorithm 4: Extraction of shape feature

Input: RGB Image
Output: 6 Shape Features
Start
Step1: RGB to Gray conversation
Step 2: Detection of an object area from an Image
Step 3: Shape Features (Shape Area, Shape Convex Area, Shape Eccentricity, Shape Solidity, Shape Filled Area, Shape Equivdiameter)
Stop.

$$\text{Shape Area } \mathrm{A}(R) = |R| = \mathrm{N} \tag{12}$$

It is a scalar value used to find out the number of pixels in the region as given by Eq. 12.

$$\text{Shape Convex Area } C = \left\{ \sum_{i=1}^{S} \alpha_i \chi_i \middle| (\forall i : \alpha_i \geq 0) \wedge \sum_{i=1}^{S} \alpha_i = 1 \right\} \tag{13}$$

It is a scalar value that specifies the number of pixels in the convex image as given by Eq. 13.

$$\text{Shape Eccentricity } E = \sqrt{1} - {b^2}/{a^2} \tag{14}$$

Here, eccentricity is obtained by using the length associated with the semimajor axis a and semiminor axis b of the object present in the image and is represented by using Eq. 14.

$$\text{Shape Solidity} = \frac{\text{Area}}{\text{Convex area}} \tag{15}$$

Solidity measure is obtained as the ratio of the object area to the area associated with the convex hull of the object Eq. 15.

$$\text{EquivDiameter} = \sqrt{\frac{4 \times \text{Area}}{\pi}} \tag{16}$$

It is a scalar quantity which denotes the circle diameter with the identical area as a region and can be computed by Eq. 16.

4.4 Artificial Neural Network

Artificial neural network [ANN] is the most powerful network architecture in today's word. In this work, ANN used to classify the natural and artificial body marks such as mole and tattoo based on color, texture, and shape features. There are three different layers with feed-forward architecture that are used to the formation of the neural network. The numbers of input features are considered equally with the available input layers. Outcome with number of body marks remains equal to the output layers shown in Fig. 3. From each image 28 features are extracted in that color (18), shape (6), texture (4), based on that mole and tattoo samples are recognized. The output layers have 2 nodes in all cases. The output is represented as a pattern vector of 2 bits Q(q1,q2) is set to 1 and remaining bits to 0 s, the image sample belongs to the ith type of body images. The vector Q1 (1 0), Q2(0 1) represents a mole and tattoo, respectively. The number of hidden layers is calculated using the formula as shown in Eq. (17).

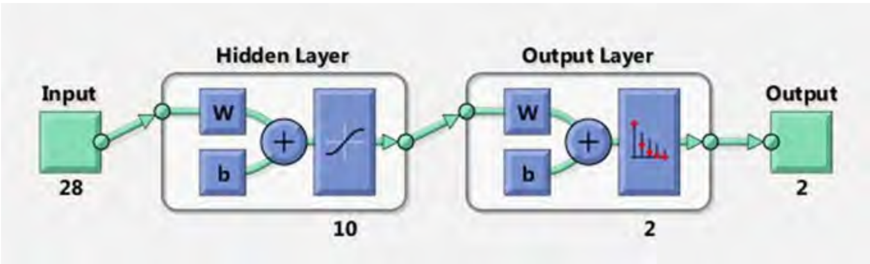


Fig. 3 Structure of ANN

$$n = \frac{I + O}{2} + y^{0.5} \tag{17}$$

where

- n number of nodes in the hidden layer
- I number of input features
- O number of outputs
- y number of input patterns in the training set.

4.4.1 Training and Testing

The identification and classification of natural and artificial body marks on different datasets like mole and tattoo are summarized as follows: Training and testing of the neural network are accomplished using body mark datasets; these datasets are divided into two parts: first one is training and testing is the second one. Totally, 266 sample images were taken for an experiment; in that, 135 images are mole and 131 images are a tattoo; 50% dataset are taken for training and 25% are taken for both validation and testing purposes. The ANN models performance for training and testing of the selected dataset shown in Fig. 4, the blue color indicates the training, red color indicates the testing, and green color indicates the validation of the experiment. The graph shows the performance of the ANN model achieving the target with many iterations in a particular time period.

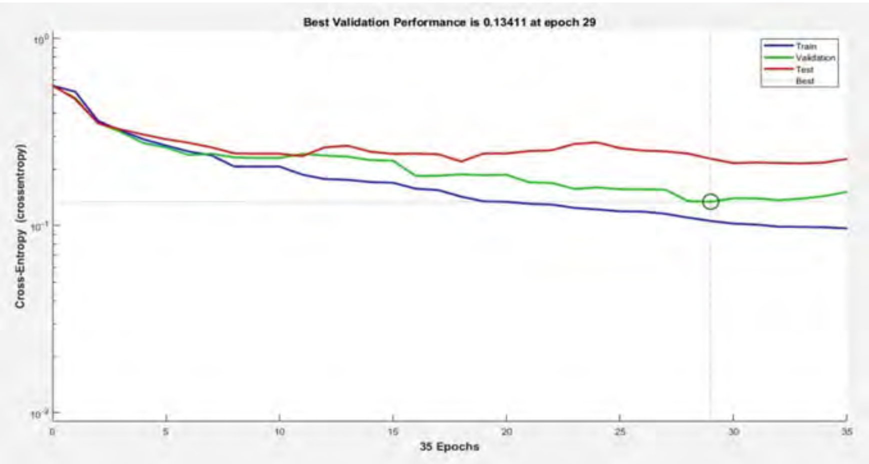


Fig. 4 Best validation performance of ANN

5 Results and Discussions

The experiment is evaluated on 266 images in which 135 contain mole images and 131 tattoo images. The total number of features are extracted from each image which is 28. The numbers of input layers are 28, the number of hidden layers is 10, and 2 output layers are set to experiment. Then 50% of the dataset is used for training and the remaining 25% for validation and 25% for testing. The proposed experimental results are shown in the confusion matrix, the prediction value of mole 86.7% true and failure case 13.3% out of 135 samples, whereas in tattoo 90.8% true and the failure case 9.2% out of 131 samples and the overall accuracy of the proposed method is 88.7% true and 11.3% failure out of 266 samples present in Fig. 5. The detailed description is given in the subsequent section.

This is a commonly used graph that summarizes the performance of a classifier's overall possible thresholds. It is generated by plotting the true positive rate (y-axis) against the false positive rate (x-axis) as you vary the threshold for assigning observations to a given class. The blue color represents class one (mole) and tan color represents the class two (Tattoo). The curves are presented in all ROC the performance of the proposed method shown in Fig. 6.

Fig. 5 Target class confusion matrix between class1 mole and class2 tattoo

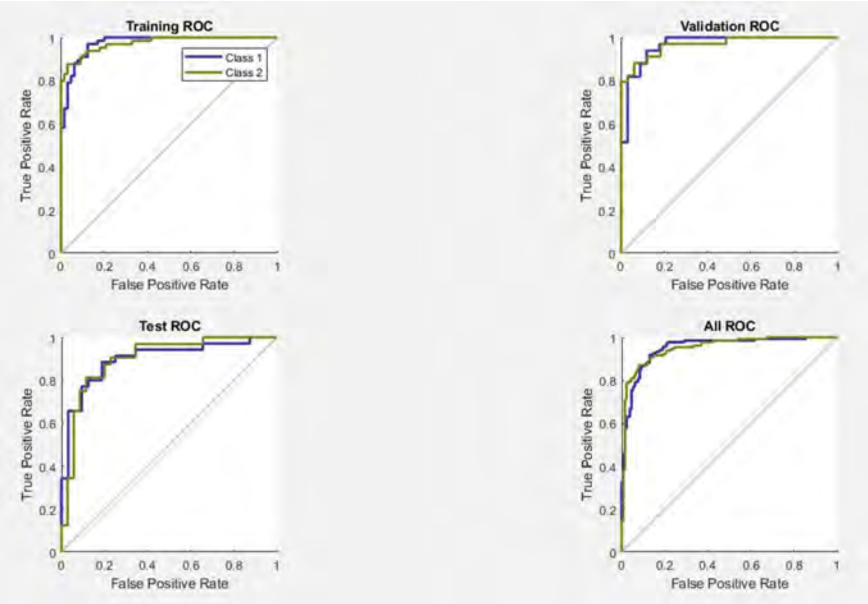


Fig. 6 Classifier accuracies for the selection of different mole and tattoo classes

5.1 Experimental Results of the Work

Mole and tattoo images are collected from various hospitals and websites; the total number of images that are considered for the experiment is 266. These images were preprocessed with the following techniques like hair removal using DULL RAZOR software and spatial filtering is used to reduce edge detection, noise reduction, and segmentation processes and finally get segmented images from the preprocessing technique. ANN toolbox is used from MATLAB platform, in that random for data division and scaled conjugate gradient for training and cross-entropy for performance, and the numbers of epochs are set to 35 iterations has been set. For the identification of mole is set to network 1 and for the tattoo is network 0. The result of mole identification is about 86.7% and the failure case is about 13.3% of 135 samples. In tattoo, the result 90.8% and the failure case 9.2% of 131 images, and the combined images overall accuracy result are 88.7%, and the failure case 11.3% of total 266 samples.

5.2 Comparative Analysis

The experiment is evaluated on artificial and natural body marks; it contains 131 and 135 images. The proposed method results are compared with other methods, whereas

Table 2 Comparison of the proposed method results with other methods

Sl.No	Methods	Results (%)
1	Neural network with backpropagation	75
2	Neural network with backpropagation	75.6
3	Auto-associative neural network	80.8
4	Auto-associative neural network	82.6
5	**Proposed method**	**88.7**

only compared with methods not with an object because the proposed method is novel. Till now no one has come across this work. The proposed method compared with the neural network with backpropagation [14, 17], an auto-associative neural network [18, 19]. The results of the proposed method are compared based on the method accuracy shown in Table 2.

The proposed method has given better results as compared to the other methods presents in the literature review.

6 Conclusion and Future Scope

The proposed method has shown that the classification and identification of body marks using ANN. The proposed method obtained 86.7% accuracy for mole and 90.8% for tattoo and overall accuracy is about 88.7% when both images were combined. Using color, texture, and shape features set for mole and tattoo, respectively. The proposed method is very useful to classification and identification between natural and artificial body marks like mole and tattoo. The proposed research work is more relevant to the body marks type recognition and classification, and it will integrate both pattern recognition and image processing technologies. The future research work will be progressed to classify and identify the artificial body marks and accidental body marks, which is useful in several areas like identifying the criminal and identify the bodies where mass death occurred.

References

1. Savakar DG, Telsang D (2016) A survey on identification and analysis of body marks. Int J Innov Res Sci Eng Technol 5(5) (May)
2. A Brief History of Tattoos. http://www.designboom.com/history/tattoo_history.html
3. Jain AK, Lee JE, Jin R (2007) Tattoo-ID: automatic tattoo image retrieval for suspect and victim identification. PCM, pp 256–265
4. Lee J-E, Jain A, Jin R (2008) Scars, Marks, and Tattoos (SMT): soft biometric for suspect and victim identification. In: Biometrics symposium, Sept 2008
5. Ngan M, Grother P (2015) Tattoo recognition technology—challenge (Tatt-C): an open tattoo database for developing tattoo recognition research. National Institute of Standards and Technology (NIST)

6. Lee T, Ng V, Gallagher R, Coldman A, McLean D (1997) Dullrazor A software approach to hair removal from images. Comput Biol Med 27(6):533–543
7. Savakar DG (2012) Recognition and classification of similar looking food grain images using artificial neural networks. J Appl Comput Sci Math 13(6) (Suceava)
8. Sheha MA, Mabrouk MS, Sharawy A (2012) Automatic detection of melanoma skin cancer using texture analysis. Int J Comput Appl 42(20) (March)
9. Kannur A, Kannur A, Rajpurohit VS (2011) Classification and grading of bulk seeds using artificial neural network. Int J Mach Intell (IJMI) 3(2):62–73
10. Kanimozhi T, Murthi A (2016) Computer-aided melanoma skin cancer detection using artificial neural network classifier. Singaporean J Sci Res (SJSR) J Sel Areas Microelectron (JSAM) 8(2):35–42
11. Acton ST, Rossi A (2008) Matching and retrieval of tattoo images: active contour cbir and glocal image features. In: IEEE SSIAI, Mar 2008
12. Abdul Jaleel J, Salim S, Aswin RB (2012) Artificial neural network-based detection of skin cancer. Int J Adv Res Electric Electron Instrum Eng 1(3) (Sept)
13. Anami BS, Savakar DG (2009) Effect of foreign bodies on recognition and classification of bulk food grains image samples. J Appl Comput Sci 6(3) (Suceava)
14. Mhaske HR, Phalke DA (2013) Melanoma skin cancer detection and classification based on supervised and unsupervised learning. Circuits, Controls and Communications (CCUBE). In: 2013 Int. Conf. IEEE Dec 2013, pp 1–5
15. Tschandl P et al (2018) The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data 5:180161. https://doi.org/10.1038/sdata.2018.161
16. Savakar DG, Telsang D, Kannur A (2020) Comparative analysis of segmentation technique for different body marks. Int J Adv Sci Technol (IJAST) 29(4):8605–8614
17. Mahmoud MKA, Al-Jumaily A, Takruri M (2011) The automatic identification of melanoma by wavelet and curvelet analysis: study based on neural network classification. In: Hybrid intelligent systems (HIS), 2011 11th Int. Conf. IEEE, Dec 2011, pp 680–685
18. Lau HT, Al-Jumaily A (2009) Automatically early detection of skin cancer: study based on neural network classification. In: International conference of soft computing and pattern recognition, IEEE, pp 375–380
19. Srivastava S, Sharma D (2016) Automatically detection of skin cancer by classification of neural network. Int J Eng Tech Res 4(1):15–18

Improvised Distributed Data Streaming Scheduler in Storm

J. Geetha, D. S. Jayalakshmi, Riya R. Ganiga, Shaguftha Zuveria Kottur, and Tallapalli Surabhi

Abstract Apache Storm is one of the most widely used platforms for processing of data streams due to its properties of being distributed, highly scalable, and fault-tolerant. It provides real-time processing, is fast and stateless, and uses master–slave architecture with ZooKeeper. In the Hadoop ecosystem, Apache Storm is the one that fills the present real-time functionality and provides strong coupling with many tools and technologies. In Storm framework, the Storm default scheduler is commonly used to schedule the task or data to be processed, whose basis for scheduling the task is the time quanta or time slots, which leads to increase in context switches and longer response time. In Storm default scheduler, the workload of a topology equally distributed among worker processes or Java virtual machine (JVM) all over the cluster using a simple round-robin algorithm without considering any priority or criteria. The proposed algorithm addresses the above-specified issues. An improvised the custom Storm scheduler was developed where the scheduling is based on the workload, which is calculated based on the total memory utilized per task and the total processing unit utilized per task, thereby resulting in lesser context switches and faster response time of distributed streaming applications.

Keywords ZooKeeper · Java virtual machine (JVM) · Self-timed periodic (STP) model

1 Introduction

Apache Storm is a distributed stream processing computation framework. It utilizes spouts and bolts to specify information sources and manipulations to allow distributed batch processing of streaming data. The topology consists of spouts and bolts which

J. Geetha (✉) · D. S. Jayalakshmi · R. R. Ganiga · S. Z. Kottur · T. Surabhi
Department of Computer Science and Engineering, Ramaiah Institute of Technology, Bangalore 560054, India
e-mail: geetha@msrit.edu

D. S. Jayalakshmi
e-mail: jayalakshmids@msrit.edu

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_42

are interconnected. It is a directed graph consisting of vertices which are computational nodes and edges which are data streams. A simple topology begins with spouts, emitting the data to one or more bolts. In the topology, bolt replicates a node having the small processing logic unit and the output of a bolt can be emitted as an input into another bolt. Storm topology stops only after it is killed. The data that is generated continuously by different sources is called Streaming data. Such data has to be processed by an incremental procedure using Stream processing techniques without allowing access to all of the data. In addition, data drift must be considered, which means that the properties of the stream may change over time. The schedulers existing in the Storm are: Isolation scheduler, Multitenant Scheduler, Default Scheduler (also known as the Even Scheduler), and Pluggable Scheduler. The isolation scheduler provides an easy and safe mechanism for sharing Storm cluster resources among many topologies. Within the Storm cluster, the isolation scheduler helps to allocate or reserve the dedicated sets of nodes for topologies. A Storm default scheduler assigns component executors as equally as possible between all the workers. In the pluggable scheduler, one can replace the default scheduler by implementing their scheduler which assigns executors to workers.

2 Background

Apache Storm [1] is a system used for real time, distributed big data processing [2, 3]. It is being used at for analysis of social media at Twitter [4] and Klout, for weather data [5], in telecommunication industries [6], etc. A Storm cluster consists of master–slave architecture where the Nimbus daemon is run on the master node and Supervisor daemon is run on the worker nodes. Coordination between the Supervisors and Nimbus daemons is managed by the ZooKeeper cluster as shown in Fig. 1. The responsibility of the Nimbus daemon includes distribution of code around the cluster, task assignment to machines, and failure monitoring. The supervisor daemon listens for work assigned to its worker machine and starts and stops worker processes as necessary based on what Nimbus has assigned to it. To perform real-time computations, topologies are created which is a graph of computation. Each worker process executes a part of the topology. The node of a topology possesses processing lodges and the edges indicate the direction of data transfer between nodes [7].

A stream is an unbounded sequence of tuples, which is the data that the topology works upon. Storm provides the primitives for the transformation of streams in a reliable and distributed way. The basic primitives provided by Storm are “spouts” and “bolts.” A spout is a source of streams and a bolt consumes input streams, performs some processing, and possibly emits new streams. An executor is a thread created by a worker process.

For a specific spout or bolt, the executor can run one or more tasks. Networks of spouts and bolts form a topology. Each node is a spout or bolt. Edges indicate the subscription of streams by the nodes which are either spouts or bolts. When a tuple

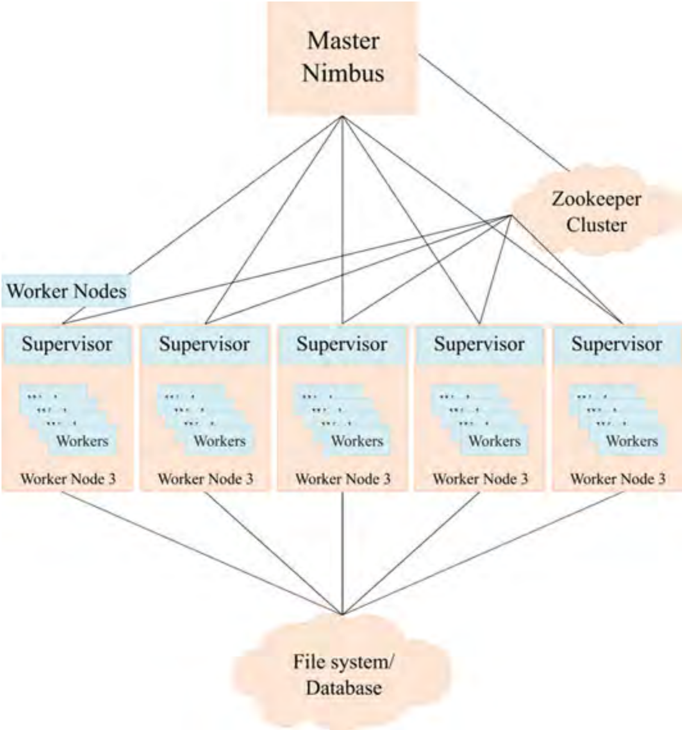


Fig. 1 Apache Storm Cluster Architecture depicting master node with Nimbus daemon, worker nodes with supervisor daemon and various worker slots; coordination is managed by ZooKeeper cluster

is generated, it is forwarded to all the nodes which have subscribed to the particular stream. A topology runs forever until it is killed.

The schedulers existing in the Storm are: Isolation scheduler, Multitenant Scheduler, Default Scheduler (also known as the Even Scheduler), and Pluggable Scheduler. The Storm default scheduler aims to produce an even allocation and hence utilizes the round-robin strategy. It iterates through the executors in the topology and uses round-robin to assign them to the configured number of workers in the topology. Then, according to slot availability of the nodes, these workers are then assigned to worker nodes. Storm provides the capability to implement custom schedulers which need to implement the "prepare" and "schedule" methods of the IScheduler interface. The custom scheduler takes as input the topology structure as a weighted graph and a set of additional user-defined parameters. Then a deployment plan or schedule is computed which outlines the mapping of executors to workers and the assignment of workers to slots, as shown in Fig. 2. The Storm scheduler can either be made to execute periodically or only when the topology is submitted [8].

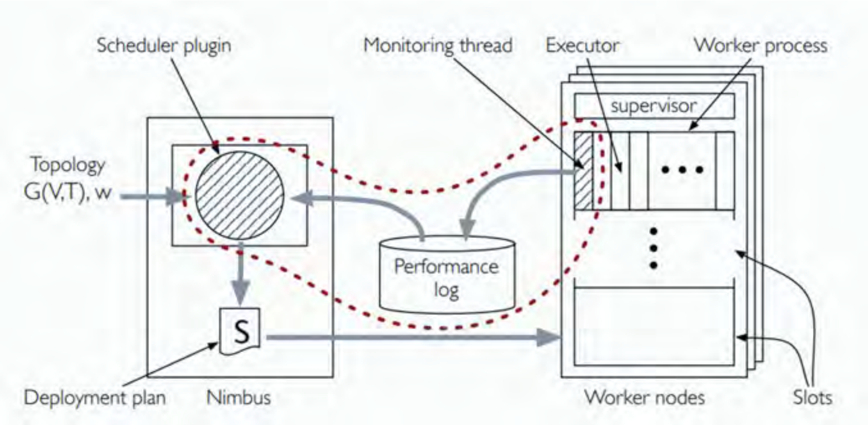


Fig. 2 Apache Storm Architecture with a depiction of scheduling methodology—the scheduler plugin holds the scheduling logic, based on the topology a deployment plan (schedule) is created which is then executed by the communication between Nimbus and Supervisors [8]

3 Related Work

The problems with the current scheduling practices in Apache Storm and associated traffic-aware online scheduling challenges [9] are discussed in this paper. The conclusions inferred from the experiments include the significant impact of inter-node/interprocess traffic on performance and the counter-effect of overloading a worker node. Along with the default scheduler, two more schedulers are used: an offline scheduler and an online scheduler. The design and implementation of T-Storm for enabling traffic-aware online scheduling were done by revision of certain Storm components and the addition of a schedule generator, a custom scheduler and load monitors. Scheduling in T-Storm operates as follows: (1) Periodic collection of runtime workload and traffic load information is done by the load monitor and stored in the database. (2) The schedule generator periodically reads load information from the database and computes a schedule using a traffic-aware online scheduling algorithm. (3) The current scheduler is regularly fetched by the custom scheduler and executed by assigning executors. The desirable properties of T-Storm are: Consolidation for Cost Reduction, Traffic-aware Online Scheduling, Worker Node, Optimization for Re-assignment Overhead, Optimization for the number of slots, Hot-Swapping of Scheduling Algorithms, and 12 Storm User Transparency. The experiments are conducted using Word Count, Throughput Test, and Log Stream Processing.

For dynamic scheduling required in streaming applications, stable scheduling algorithms are preferred over efficient scheduling algorithms. In [10], a Stable Online scheduling strategy with Makespan Guarantee (SOMG) is discussed which includes mathematical dependencies between system stability, resource utilization, and response time; specification of acceptable response time objectives and high system stability requirements to be met; structural optimization of a data stream graph by quantification and adjustment of vertices and heuristic critical path scheduling of a data stream graph. The system architecture consists of space for graph, Storm, hardware, and user and the topology includes subsystems of Nimbus, ZooKeeper, and Supervisor. The performance evaluation is done using a simulation environment consisting of the master node, ZooKeeper node and 16 worker nodes, and parameter settings based on system throughput, response time and vertices live migration ratio.

For prediction of a given scheduling solution, the accurate prediction of the average tuple processing time of an application can be done using a topology-aware [11] method. For scheduling, assigning threads to machines using an effective algorithm under the guidance of prediction results is used. To evaluate and validate, Storm was tested with three representative applications. The proposed framework consists of modules such as performance predictor, data collector, data store, schedule generator, time synchronizer, data pre-processor, and custom scheduler. Performance prediction, schedule generation, and the custom scheduling constitute for the important modules of the framework proposed. Assigning threads to machines is served by a prediction which is a ground rule for a scheduling algorithm and the objective of the scheduling problem over an application graph G is to minimize the average tuple processing time. From the experimental results, the following observations are made: The accuracy prediction for the Word Count Topology (Stream Version) is 83.3%; the accuracy of prediction for the Log 13 Stream Processing Topology of 84% is slightly lower than the accuracy of individual predictions; And for the Continuous Query Topology, the prediction accuracy is 85.2%.

Self-Timed Periodic [12] is an execution model, which is a combination of self-timed scheduling and periodic scheduling. In self-timed scheduling, actors are fired soon after the data dependency is observed. This schedule, since a long time, is noticed as the most appropriate policy for streaming applications. To resolve the problem with respect to the Periodic Schedule model of static nature regarding latency increase in unbalanced graphs can be resolved by using the self-timed periodic schedule. Self-timed periodic schedule is a hybrid execution process model based on the integration of self-timed schedule and periodic schedule while considering varying IPC times. To demonstrate the performance of self-timed periodic schedule model, two classes of self-timed periodic schedules based on two different granularities are used; the first schedule is denoted as STPIqi, and the other schedule is denoted as STPI ri. Under periodic scheduling, the effect of self-timed periodic schedule can be modified by putting back with actor period in every stage with its worst-case execution time. In worst case, execution time is the summation of all actor's computational and communication time. The performance of newly proposed scheduling policy is evaluated based on the experiment conducted on a set of 10 real-life streaming applications: DCT, FFT, Beamformer, Filterbank, MP3, Sample-rate, H.263 Encoder, H.263 Decoder, Bipartite, and Satellite.

Distributed Stream Processing Systems [13] are gathering interest as they provide the capability to scale invisibly by using the distributed resources of a cloud environment. One of the major challenges faced by DSPS is the development of effective scheduling mechanisms which can manage resources and their allocation to the concurrently running data analytics tasks. The proposed model aims to develop a scheduling mechanism that allocates resources to applications according to their priority by attaching it to the vertices of the processing flow graphs which represent the processing components and mapping operator priority to tuple priority. To achieve this, a meta-scheduler is used whose task is to intercept tuples, assign them priority according to their destinations, reorder them based on priority, and then forward them to their destination. The approach is implemented on Quasit, an open-source DSPS, by developing two new QoS specifications, corresponding to priority schemas: absolute or proportional priority and priority specifications. The experimental evaluation based on meta-scheduler overhead shows that the performance of the meta-scheduler depends mainly upon the number of priority classes. Evaluation of vehicle traffic proved that using the meta-scheduler allowed the application to handle the high input rate of the data streams.

Heuristic scheduling [14] algorithm operates by using graph partitioning algorithms and mathematical optimization software package to obtain reliable and efficient tasks involving more communication. Though optimal solution for efficient scheduling exists, for large size problems characterized by huge search space and high complexity of computation makes it infeasible to use optimization scheduling technique to obtain such an optimized solution. I-Scheduler is an iterative heuristic-based graph partitioning algorithm focusing on cutting down the total number of tasks for the specified problem so that the optimization software can operate in the scheduled period. Here, our performance metric is the average throughput (average of the total number of instances) accomplished in every bolt's task per 10 s period. The two real-world topologies used are—Smart Homes application based on Load Prediction model technique and NYC Taxi Data operated by Top Frequent Routes method. Although for the top frequent Routing topology there is no significant performance upgrade by the usage of I-Scheduler on top of R-Storm, it(R-Storm)needs a substantial amount of tuning. On comparison, it has been observed that similar performance results were achieved using I-Scheduler without requiring such extensive tuning.

Suitability of scheduling algorithms depends on the parameter of importance [15–17]. Real-time applications can generally tolerate failed messages with a higher margin compared to other applications. With the requirement of higher quality in streaming applications, this tolerance reduces and hence scheduling should also take into account this development.

4 Proposed Algorithm

The objective of this project is to develop an improvised scheduler in the Storm for distributed data streaming application based on the workload, thereby resulting in faster response time of distributed streaming applications. The following are some of the objectives of the project:

1. To optimize the existing Storm scheduler to get faster response time and less waiting time.
2. To reduce the number of context switches, by including the workload calculation criteria for the assignment of executors to worker slots.
3. To prioritize the tasks, such that a task with the maximum workload is executed first.

The algorithm is a modification to the existing Even Scheduler. Here the task having maximum workload is taken and scheduled. A task is defined as either the execution of a spout or a bolt. There exists workload calculation before choosing which task has the highest priority. The score is calculated as the maximum among Total memory consumed per task and total processing unit consumed per task. Figure 3 depicts the algorithm (Fig. 4).

The steps of the entire algorithm are explained below:

1. Score to Executor Mapping—From a cluster, topologies are selected which need scheduling, then for each such topology, the TotalMemoryPerTask and Total-CpuPerTask required are calculated, which are available in the topology specifications. Find the maximum among them; this is known as score calculation. Then the score is mapped to an executor of the topology arbitrarily. Hence, the first step of Score to executor mapping.
2. Scheduler Assignment—From the above mapping, the task with the maximum workload is chosen, and check for free WorkerSlot, which has already been

```
---------------------------------------
Algorithm 1 Proposed Algorithm
---------------------------------------
Procedure BASIC_SCHEDULER <parameters: Topologies Clus-
ter>
Loop1 - Topology wise:
  Choose the Topology which needs Scheduling from the
  cluster
  Loop2 - Executer wise:
    Calculate the score <- max (TotalMemReqTask,
    TotalCpuReqTask)
  Loop2 end
  Loop3 - WorkerSlot wise
    Map the executors to workerSlots only if:
      Executer is free && WorkerSlot is free
    End if
  Loop3 end
  Loop4: Reverse Mapping and Cluster Assignment
    Map nodePort and Executor
    Assign topology to cluster
  Loop4 ends
Loop1 ends
Scheduled Topologies are returned
```

Fig. 3 Proposed algorithm of custom scheduler

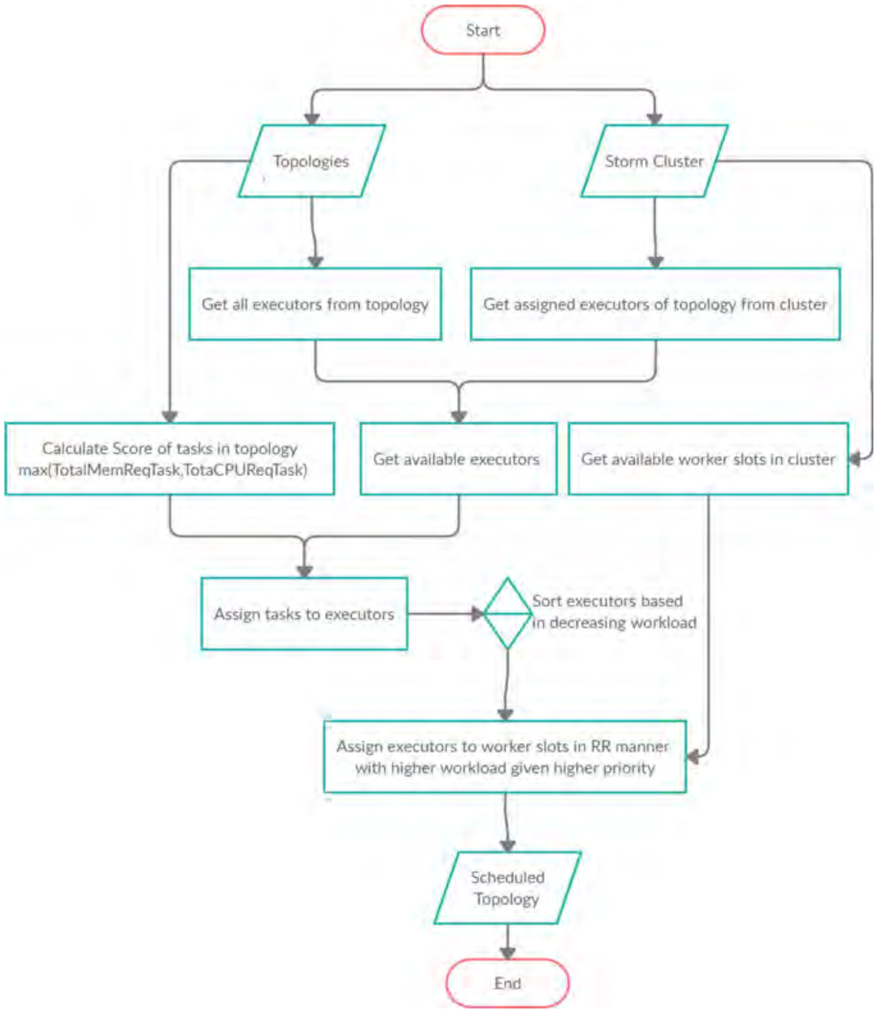


Fig. 4 Flow diagram for the proposed scheduling algorithm

mapped to an Executor. This is known as Scheduler Assignment. This is iterated till either the slots are not occupied or the tasks are finished. The task with a higher workload has higher priority.

3. NodePort to Executor Mapping—Here nodePorts (of the worker nodes) are mapped to the WorkerSlots which in turn are mapped to the Executors. This is also known as Reverse Mapping of WorkerSlots and executors.
4. Cluster Assignment of Topologies—At the end, the scheduled topologies are assigned to the cluster. This topology assignment is carried out in the Cluster.java class. There are many ways of assigning the topology back to the cluster. After this assignment, the topology is said to be scheduled.

The proposed custom Storm Scheduler involves a calculation of the workload based on memory and CPU utilized per task of a topology. This reduces the context switch present in round-robin, which is based on time quanta. This leads to a reduction in response time, increase in the number of acknowledgement of the stream and a decrease in the average latency.

5 Results

The *FastWordCountTopology* was executed with both the Default Scheduler and the Custom Scheduler which takes in a string input and calculates the *WordCount* in a non-traditional way. The metrics are compared across the uptime. The Metrics include Acknowledged, Average Latency, Failed, and Acknowledged per second which are obtained from the *FastWordCountTopology.printMetrics* method. Uptime represents the time duration for which the topology has been running and Acked denotes the number of messages/stream packets acknowledged in a given window. Similarly, Failed denotes the number of messages failed in a given window. The topology keeps track of all statistics related to it for a given time window and has an attribute called *completeLatency* which denotes the total latency for processing the message in a given window. The average latency is calculated by taking the average of the sum of *completeLatency* times the Acked messages in a window, for all windows for the corresponding uptime. The below snapshots are a comparison of the default scheduler and the custom scheduler based on different metrics have been chosen.

From Fig. 5, it can be inferred that the Custom Scheduler can acknowledge more and more stream packets as the uptime increases as compared to that of Default Scheduler. From Fig. 6, it is inferred that the Custom Scheduler and Default Scheduler have approximately the same Average Latency. Though it is observed that initially,

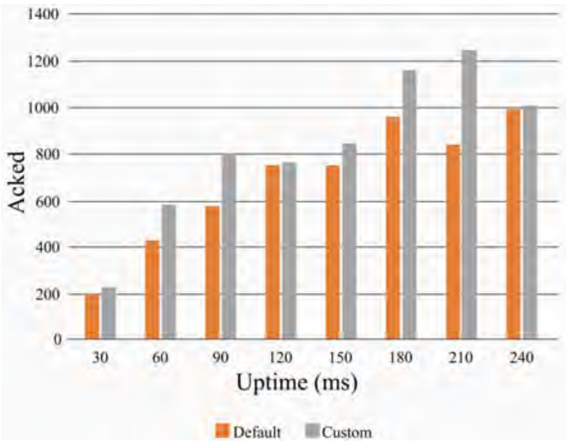


Fig. 5 Asked-uptime for default versus custom scheduler

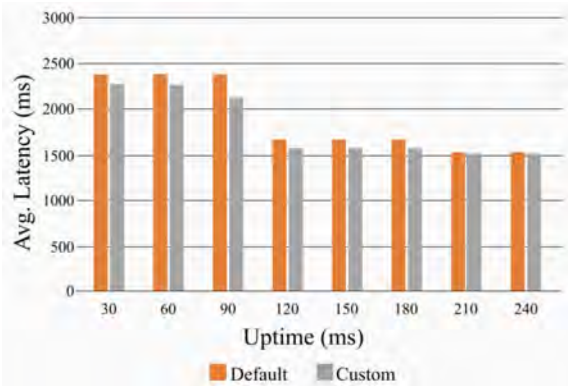


Fig. 6 Average latency of default versus custom scheduler

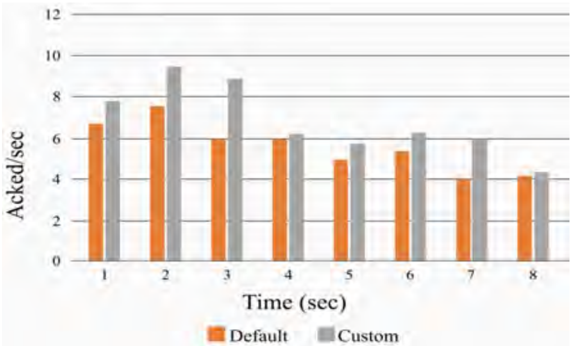


Fig. 7 Asked per second for default versus custom scheduler

the latency of the Custom Scheduler was slightly lower than the Default Scheduler; later on, it is seen to become almost the same.

In Fig. 7, the comparison of the Acknowledged per second across uptime in Storm default scheduler and Storm custom scheduler can be seen. This shows that the rate at which the custom scheduler acknowledges the stream packets increases faster than the default scheduler as uptime increases.

Acknowledgements in Custom Scheduler are decently higher that Default Scheduler, similarly average latency is slightly less in the custom scheduler. So when Ack/sec is considered, it is observed that in some cases Custom Scheduler outperforms the Default Scheduler.

6 Conclusions and Future Work

The project presents a new Custom Storm Scheduler, which is a modification of the Default Storm Scheduler. The proposed Storm scheduler works on the concept of finishing off the larger tasks of a topology first. The size of the tasks is calculated based on the workload of memory and CPU per task of a topology. This reduces the context switch present in round-robin, which is based on time quanta. The Storm custom scheduler designed as compared to that of the Storm default scheduler, it has been noted that; Storm custom scheduler made visibly outperforms the Storm default scheduler in the number of acknowledgements made per second. Storm custom scheduler still has approximately the same average latency when analyzed in comparison with the Storm default scheduler. This leads to a future scope of reducing it henceforth. The Storm custom scheduler was successful in acknowledging more streams than the Storm default scheduler, hence leading to quick data processing in steps after scheduling.

The proposed Storm custom scheduler does not differ much in average latency in comparison with the Storm default scheduler. Thus, in future, a more advanced methodology can be introduced to overcome this issue. In future, the Storm scheduler can be further modified to make it more fine-grained by introducing a new grouping methodology or any other useful criteria based on topology or source generating data just like workload calculation in the current project.

References

1. Iqbal MH, Soomro TR (2015) Big data analysis: Apache storm perspective. Int J Comput Trends Technol 19(1):9–14
2. Van Der Veen JS, van der Waaij B, Lazovik E, Wijbrandi W, Meijer RJ (2015) Dynamically scaling apache storm for the analysis of streaming data. In: 2015 IEEE first international conference on big data computing service and applications. IEEE, pp 154–161 (March)
3. Yang W, Liu X, Zhang L, Yang LT (2013) Big data real-time processing based on storm. In: 2013 12th IEEE international conference on trust, security and privacy in computing and communications. IEEE, pp 1784–1787 (July)
4. Toshniwal A, Taneja S, Shukla A, Ramasamy K, Patel JM, Kulkarni S, Jackson J, Gade K, Fu M, Donham J, Bhagat N (2014) Storm@ twitter. In: Proceedings of the 2014 ACM SIGMOD international conference on Management of data, pp 147–156 (June)
5. Gu S, Yao L, Tunc C, Akoglu A, Hariri S, Ritchie E (2016) An autonomic workflow performance manager for weather research and forecast workflows. In: 2016 international conference on cloud and autonomic computing (ICCAC). IEEE, pp 111–114 (September)
6. Schaefer C, Manoj PM (2015) Enabling privacy mechanisms in apache storm. In: 2015 IEEE international congress on big data. IEEE, pp 102–109 (June)
7. Apache Software Foundation (2019) Tutorial Apache Storm Version 2.2.0, Apache Software Foundation 2019, viewed on Sept 2020. https://storm.apache.org/releases/current/Tutorial.html
8. Aniello L, Baldoni R, Querzoni L (2013) Adaptive online scheduling in storm. In: Proceedings of the 7th ACM international conference on Distributed event-based systems, pp 207–218 (June)
9. Xu J, Chen Z, Tang J, Su S (2014) T-storm: traffic-aware online scheduling in storm. In: 2014 IEEE 34th international conference on distributed computing systems. IEEE, pp 535–544 (June)
10. Sun D, Huang R (2016) A stable online scheduling strategy for real-time stream computing over fluctuating big data streams. IEEE Access 4:8593–8607
11. Li T, Tang J, Xu J (2016) Performance modeling and predictive scheduling for distributed stream data processing. IEEE Trans Big Data 2(4):353–364

12. Dkhil A, Do XK, Louise S, Rochange C (2015) A hybrid scheduling algorithm based on self-timed and periodic scheduling for embedded streaming applications. In: 2015 23rd Euromicro international conference on parallel, distributed, and network-based processing. IEEE, pp 711–715 (March)
13. Bellavista P, Corradi A, Reale A, Ticca N (2014) Priority-based resource scheduling in distributed stream processing systems for big data applications. In: 2014 IEEE/ACM 7th international conference on utility and cloud computing. IEEE, pp 363–370 (December)
14. Eskandari L, Mair J, Huang Z, Eyers D (2018) Iterative scheduling for distributed stream processing systems. In: Proceedings of the 12th ACM international conference on distributed and event-based systems, pp 234–237 (June)
15. Eskandari L, Huang Z, Eyers D (2016) P-Scheduler: adaptive hierarchical scheduling in apache storm. In: Proceedings of the Australasian computer science week multi conference, pp 1–10 (February)
16. Liu Xm Buyya R (2017) D-storm: dynamic resource-efficient scheduling of stream processing applications. In: 2017 IEEE 23rd international conference on parallel and distributed systems (ICPADS). IEEE, pp 485–492 (December)
17. Peng B, Hosseini M, Hong Z, Farivar R, Campbell R (2015) R-storm: resource-aware scheduling in storm. In: Proceedings of the 16th annual middleware conference, pp 149–161 (November)

Efficient Mining of Rare Itemsets

Shwetha Rai, Geetha M., and Preetham Kumar

Abstract Itemset mining discovers interesting patterns in the dataset. The itemset may be frequent or it can be rare based on its occurrence in the database. It has been observed that most of the algorithms are designed for mining frequent itemsets. However, discovery of rare itemsets is equally important since they play a major role in making decisions in some situations. The efficiency of the algorithms depend on the way in which the data structures are designed to store and retrieve the data. Hyperlinked Rare Pattern Mining algorithm discovers all rare itemsets and is suitable for sparse dataset. In this algorithm item_id and its support count are stored in Support and Header tables. This redundancy is removed in the proposed algorithm to improve the time efficiency. An experimental analysis is conducted to discover the rare itemsets. It is observed that while there is an improvement in time efficiency, there is a tradeoff for space efficiency.

Keywords Association rule mining · Data engineering · Data structure · Hyper-link · Rare itemsets · Redundant data · Sparse data

1 Introduction

Data mining, an important field in computer science, is the process of discovering patterns, that are both interesting and useful, and relationships among the patterns in huge amount of data [5]. Pattern mining can be used to extract patterns that are

S. Rai (✉) · G. M.
Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India
e-mail: shwetha.rai@manipal.edu

G. M.
e-mail: geetha.maiya@manipal.edu

P. Kumar
Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education,Manipal, Karnataka 576104, India
e-mail: preetham.kumar@manipal.edu

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_43

either frequent or rare from the bulky dataset. Rare Pattern mining is a technique used to extract rare patterns, the patterns of which the support count is between two user-defined thresholds, freqSup and rareSup, from the dataset.

Most of the research works on pattern mining are carried out only on the extraction of frequent patterns for a considerable period disregarding the extraction and analysis of rare patterns. In recent years, pattern mining techniques showed a key role in solving numerous tasks in Data mining. Substantial extent of research has been accomplished for the discovery of rare itemsets.

Data structures in the implementation of pattern mining algorithms play a key role in discovering the itemsets of interest. A tree data structure is more suitable when the dataset is dense but the same data structure may not be efficient to store itemsets from a sparse dataset. Hyper-Linked RPM algorithm is an attempt to store sparse dataset in the main memory.

2 Literature Survey

Since the inception of the research work on rare pattern mining, a significant extent of research has been accomplished for the discovery of rare patterns. Different authors have proposed different techniques to extract patterns from the dataset. There are different techniques for discovery of rare patterns such as level-wise approach or candidate generation approach like Apriori [2] or a non-candidate generation method, like FP-Growth [7] which is efficient than candidate generation algorithm. The rare itemsets are also discovered based on other constraints such as multiple support counts [3, 8–12] and based on high utility, a weight given to each item in the database [6, 15].

An Apriori approach, a candidate generation method, for the discovery of rare itemsets was first proposed in the paper [13]. Algorithms such as ARIMA [16] and AfRIM [1] used a single support value for identifying and extracting the rare itemsets. It is known from the existing literature for rare itemsets mining that there are many drawbacks of in this approach and hence to overcome those drawbacks other techniques were developed.

Rare Pattern Tree or RP-Tree, implemented in [17] is a tree based algorithm to store rare itemsets. This algorithm supports two threshold values and the transactions that has at least one rare itemset were considered for further processing. RP-Tree algorithm showed a better performance when multiple support thresholds were used for the itemsets [3]. Since dense datasets contain many frequent itemsets, tree-based approaches are considered to be more suitable to discover frequent itemsets. The tree-based approaches that gave a good result for the data having long patterns, failed for data having short patterns.

In the paper [4], a queue-based approach to mine infrequent patterns, "Hyper-Linked Rare Pattern Mining (HLRPM)", was proposed. The authors claimed that it generates better results when the dataset provided is sparse and has long patterns. It is space efficient when compared to other tree based rare pattern mining algorithms. The algorithm uses hyper-link data structure [14] to store the rare transactions. Since it employs memory-based data structure, it outperforms the level-wise and pattern growth approaches in various cases.

Each of the techniques discussed in the previous section has its own advantages and drawbacks. In the paper [4], as claimed by the author, the algorithm is faster compared to other approaches but it takes two database scans and one additional scan through the support count table which affects the time consumption.

3 Research Methodology

3.1 HLRPM Algorithm

Algorithm 1 represents the step wise method mining rare itemsets using hyperlink structure.

Algorithm 1 HLRPM()

Input: Complete original transaction_database(DB), rareSup, freqSup
Output: All rare_itemsets in the DB

∀ items in the database, generate the number of times it appears in the DB
for each item, $IT \in DB$ **do**
  if the support count (IT) > rareSup and support count (IT) < freqSup **then**
    RareItem ← IT
  end if
end for each
for each transaction $TR \in DB$ **do**
  if ∃ r | r is Rare ∧ r ∈ TR **then**
    RareItemTransaction ←TR
  end if
end for each
Create header table (HT) containing the following fields: item_id, sup_count and hyperlink to link the transaction items of TR.
Construct different queues, Q_j that has following fields: item_id and hyper-link that stores the items of RareItemTransaction. Use a hyperlink to link all transactions with same first item.
for each item β in HT, create rare itemset projections from β-projected database.

The steps in discovering rare itemsets are as follows:

1. The transaction DB is scanned once to find the total number of occurrences of each item in the DB and it is entered in the support table.
2. Initialise two support threshold values, freqSup and rareSup. The itemsets I, having support count between these thresholds are considered to be rare.
3. If the support count of itemsets are lesser than freqSup but greater than rareSup then such items are considered as rare items and it will be included in RareItem table.
4. The DB is scanned again to identify the transactions that contains at least one item that is rare. Only these transactions are included in subsequent steps. Transactions that is a collection of frequent itemsets are excluded from future processing.
5. A header table is created with 3 fields to store item_ID, sup_Count and hyper_Link to link all transactions beginning with the same first item.

6. For each item β in the header table, rare itemset projections will be generated in the new β-projected database.

3.2 Proposed Algorithm

In the proposed algorithm, the header table in [4] is modified such that it consists of following fields: item_id, support_count, Type and hyperlink. Type is an additional field to indicate whether the item is noise(zero), rare(one) or frequent(two). During the initial scan of the database each item is read and its support count is incremented if the item is present in the table. Otherwise, a new entry is created for item in header table if it is not present. The support count is initialised to 1 and Type is set as 0. During each scan of the database, if freqSup > SupCount(I) > rareSup, then its type will be 1 which is referred as Rare item. If SupCount(I) > freqSup then Type is marked as 2 indicating the item is frequent. The additional scan of the database is reduced using this technique. Then rest of the algorithm follows same approach as mentioned in the HLRPM.

Algorithm 2 Efficient Hyper-Linked_Rare_Pattern_Structure (EHLRPS)

Input: Complete original transaction_database, rareSup, freqSup **Output**: Complete set of rare_itemsets

```
Create header table (HT) containing the following fields: item_id, sup_count, Type to indicate
rare/ frequent and hyperlink
for each item IT ∈ DB do
  if item IT is in the HT then
    support count(IT)← support count(IT)+1
    if freqSup >support count (IT) > rareSup then
      mark Type (IT) as Rare
    else
      mark Type (IT) as Frequent
    end if
  else if item IT not in the HT then
    create entry for item IT in HT
    support count(IT)← 1
    Type (IT) ← 0
    hyperlink ← NULL
  end if
end for each
for each transaction TR ∈ DB do
  if ∃ r | r is Rare ∧ r ∈ TR then
    TR ← RareItemTransaction
  end if
end for each
Create different queues, Q_j with following fields: item_id and hyper-link that stores the items
of j^th RareItemTransaction. Use a hyperlink to link all transactions with same first item.
for each item x in HT, create rare item projections from x-projected database.
```

A schematic representation of EHLRPS algorithm is shown in Fig. 1.

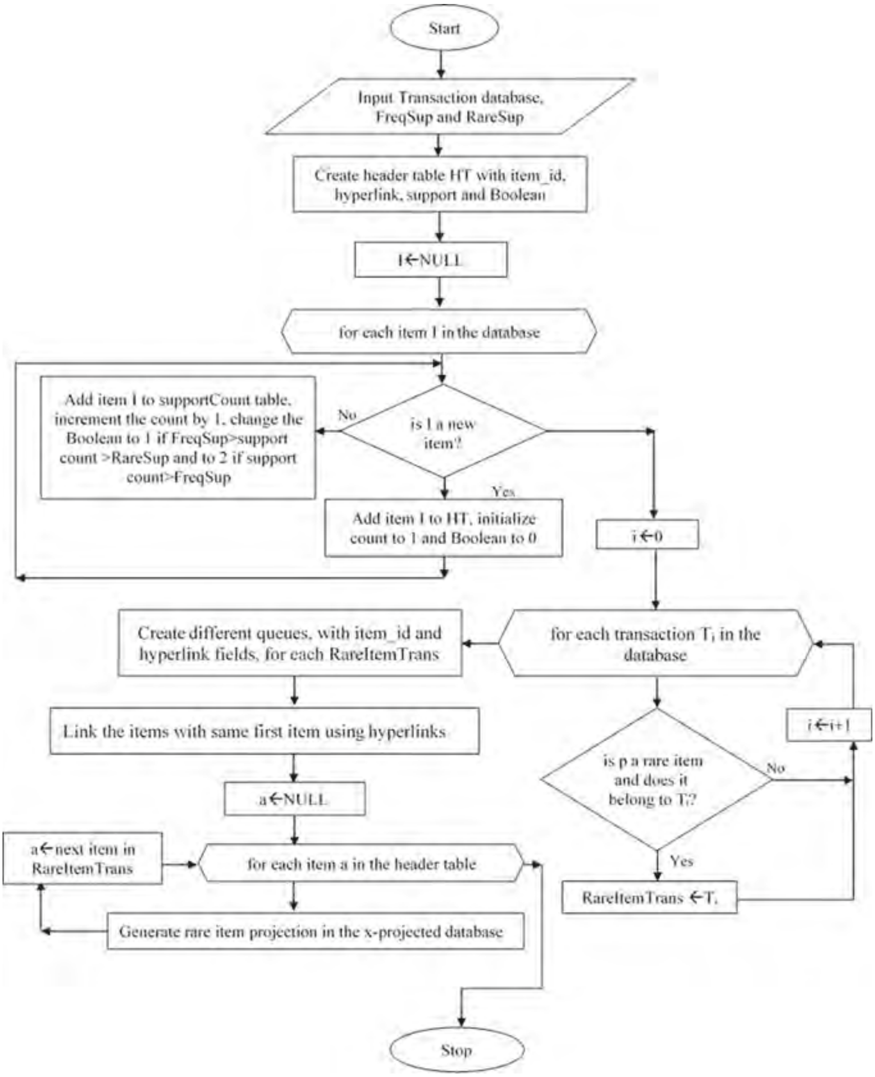


Fig. 1 Schematic representation of efficient hyper-link RPM algorithm

4 Illustration of the Proposed Algorithm

The implemented method is illustrated by considering a sample database with eight transactions as shown in Table 1.

Table 2 shows the header table HT constructed with four fields, Item_id, Sup_count, Type and Hyperlink, for the transaction database in Table 1.

Table 1 Sample transaction database

T_ID	Itemsets					
Tr1	d	f	c	a		
Tr2	a	d	g			
Tr3	f	e	b	d	a	
Tr4	d	a	h	g		
Tr5	g	a	c	f	e	d
Tr6	d	a				
Tr7	d	a	f			
Tr8	a	e				

Table 2 Header table created based on Table 1

Item_id	Sup_count	Type	Hyperlink
h	1	0	Ø
b	1	0	Ø
c	2	1	Ø
e	3	1	Ø
g	3	1	Ø
f	4	2	Ø
d	7	2	Ø
a	8	2	Ø

Table 3 Header table with rare transactions

Item_id	Sup_Count	Type	Hyperlink						
h	1	0	→	h	g	d	a		
b	1	0	→	b	e	f	d	a	
c	2	1	→	c c	f e	d g	a f	 d	 a
e	3	1	→	e	a				
g	3	1	→	g	d	a			
f	4	2	Ø						
d	7	2	Ø						
a	8	2	Ø						

Withe respect to the user defined rareSup set to 20% and freqSup set to 40% the rare items identified are **g, e** and **c**. Table 3 shows the output of second database scan that discovers rare transactions.

Table 4 Execution time of hyper-linked RPM and modified algorithm

No. of transactions	Execution time (s)	
	Hyper-linked RPM	Modified algorithm
100	0.545	0.427
200	1.04	0.83
300	1.913	1.339
400	2.159	1.6
500	2.8	2.214
600	3.363	2.446
700	3.683	2.811
800	4.439	3.348
900	4.889	3.811
1000	5.492	4.093

5 Results and Analysis

A comparison between "HyperLinked RPM" algorithm and proposed algorithm is made with data set of 1000 transactions starting with 100 transactions. The experiment was conducted using Visual Studio 2012 on Intel core i5 at 2.30 GHz with 4 GB RAM.

To check the output of sample input transaction database, the algorithm was run with minimum support threshold i.e., rareSup = 20% and maximum support threshold i.e., freqSup = 40%. A total of 10 outputs were analyzed with increment of 100 transactions each time and results were analyzed based on execution time of both the algorithms. Table 4 shows the observed output of both the algorithms. The difference in the time taken for the execution of the algorithms can be noted and using the same values, a line graph representation of the execution time is shown in Fig. 2.

From Fig. 2 it is found that as the number of transactions increases, the modified algorithm is faster and performs 24% better than the "Hyper-Linked Rare Pattern Mining" algorithm.

6 Conclusion and Future Enhancement

The Hyper-Linked Data Structure approach (HLDS) used in discovery of rare itemsets used two tables: Support count table and Header table, which leads to redundancy. This was solved by removing the redundant Support Count table and retaining Header table with an additional field, to indicate the type of the itemset i.e., noise, rare or frequent. The overhead of copying the itemset and its support count to the Header table is also removed which resulted in improvement of the running time.

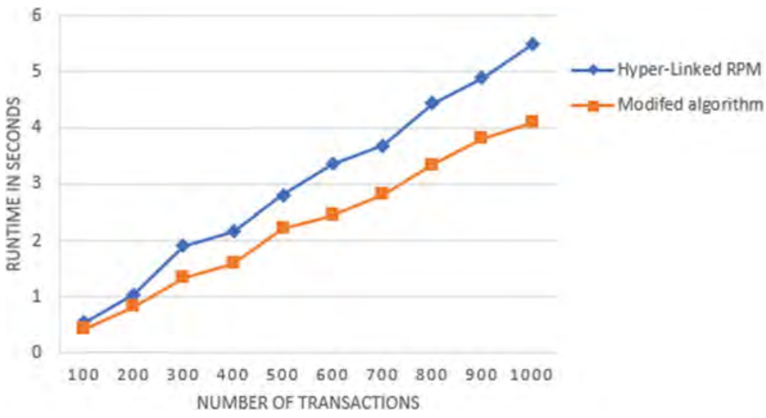


Fig. 2 Execution time of hyper-linked RPM and modified algorithm

The EHLRPS algorithm is a modification of the algorithm given in the paper [4] for mining rare patterns using a queue based HLDS. Performance evaluation, as shown in Fig. 2, explains that the modified method is 24% faster than the algorithm given in the paper [4].

The proposed algorithm works better only when the database is sparse but it is not a good option when the database is dense. It scans the database twice to discover the rare items and then to store the rare transactions in the main memory, a method can be developed to do the same in a single scan as a future enhancement. Also, since the algorithm uses a sequential approach to discover rare itemsets, it is not suitable for big data. Hence a parallel approach to discover the interesting patterns may be implemented.

References

1. Adda M, Wu L, Feng Y (2007) Rare itemset mining. In: Sixth international conference on machine learning and applications (ICMLA 2007). IEEE, pp 73–80
2. Agarwal R, Srikant R, et al (1994) Fast algorithms for mining association rules. In: Proceedings of the 20th VLDB conference, pp. 487–499
3. Bhatt U, Patel P (2015) A novel approach for finding rare items based on multiple minimum support framework. Procedia Comput Sci 57:1088–1095
4. Borah A, Nath B (2017) Mining rare patterns using hyper-linked data structure. In: International conference on pattern recognition and machine intelligence. Springer, pp 467–472
5. Clifton C (2020) Data mining. https://www.britannica.com/technology/data-mining. Last accessed 9 Aug 2020
6. Goyal V, Dawar S, Sureka A (2015) High utility rare itemset mining over transaction databases. In: International workshop on databases in networked information systems. Springer, pp 27–40
7. Han J, Pei J, Yin Y, Mao R (2004) Mining frequent patterns without candidate generation: a frequent-pattern tree approach. Data Mining Knowl Discov 8(1):53–87

8. Kiran RU, Re PK (2009) An improved multiple minimum support based approach to mine rare association rules. In: 2009 IEEE symposium on computational intelligence and data mining. IEEE, pp 340–347
9. Kiran RU, Reddy PK (2009) An improved frequent pattern-growth approach to discover rare association rules. In: KDIR, pp 43–52
10. Kiran RU, Reddy PK (2010) Mining rare association rules in the datasets with widely varying items' frequencies. In: International conference on database Systems for advanced applications. Springer, pp 49–62
11. Kiran RU, Reddy PK (2010) An efficient approach to mine rare association rules using maximum items' support constraints. In: British national conference on databases. Springer, pp 84–95
12. Koh YS, Ravana SD (2016) Unsupervised rare pattern mining: a survey. ACM Trans Knowl Discov Data (TKDD) 10(4):1–29
13. Liu B, Hsu W, Ma Y (1999) Mining association rules with multiple minimum supports. In: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pp 337–341
14. Pei J, Han J, Lu H, Nishio S, Tang S, Yang D (2001) H-mine: hyper-structure mining of frequent patterns in large databases. In: Proceedings 2001 IEEE international conference on data mining. IEEE, pp 441–448
15. Pillai J, Vyas O, Muyeba M (2013) Huri–A novel algorithm for mining high utility rare itemsets. In: Advances in computing and information technology. Springer, pp 531–540
16. Szathmary L, Napoli A, Valtchev P (2007) Towards rare itemset mining. In: 19th IEEE international conference on tools with artificial intelligence (ICTAI 2007), vol 1. IEEE, pp 305–312
17. Tsang S, Koh YS, Dobbie G (2011) RP-tree: rare pattern tree mining. In: International conference on data warehousing and knowledge discovery. Springer, pp. 277–288

EDoS-BARRICADE: A Cloud-Centric Approach to Detect, Segregate and Mitigate EDoS Attacks

S. B. Ribin Jones and N. Kumar

Abstract Cloud computing through realizing the height of virtualization offers service models that can meet dynamic demands through performing auto-scaling of resources [1]. This helps the cloud service providers to broaden the grasp across sectors and the computing service market. Though it follows stretchable and elastic service models, it implements a rigid pay-per-use utility pricing model [Ribin Jones and Kumar in J Adv Res Dyn Control Syst 11(9):541–553, 2019 2]. The idea of dynamically scaling across platform makes it more vulnerable to security threats and makes room for easy exploits [Ribin Jones and Kumar in J Adv Res Dyn Control Syst 11(9):541–553, 2019 2]. Among various security threats, the economic denial-of-service (EDoS) attack presents a serious threat, since it exploits auto-scaling feature to impact the utility pricing model [Ribin Jones and Kumar in IEEE Xplore third international conference on trends in electronics and informatics, pp 1003–1008, 2019 3]. In this paper, a real-time cost incurring EDoS attack is performed against a cloud data center hosted Web page with simple Structured Query Language (SQL) manipulation method for experimental research. The experimental observations are applied to define an effective EDoS-BARRICADE that performs detection, segregation and mitigation specific to EDoS attack. The detection algorithm considers metrics that are associated with the auto-scaling feature to detect a suspicious increase in VM activities. The segregation algorithm implements linear SVM to isolate attack VMs optimally and rapidly. The results show that the developed EDoS-BARRICADE algorithms perform detection and segregation with 100% accuracy.

Keywords Cloud computing · Economic denial of service (EDoS) · DDoS · IDPS · SLA

S. B. Ribin Jones (✉) · N. Kumar
Department of Computer Science and Engineering, VISTAS, Chennai, India
e-mail: ribinjones@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_44

1 Introduction

Cloud computing through achieving multi-layer virtualization provides resources for provisioning of on-demand access through Internet access [2]. The resources involve, but not limited to programming platforms, software services and hardware capabilities such as storage, processing and networks [3]. The cloud has the inbuilt brokering capabilities, through which the service users negotiate various characteristics termed as QoS parameters to reach a service-level agreement (SLA) with cloud service providers (CSPs) [3]. Once the agreement is reached, the service offered has advanced and unique features such as on-demand access, pay-per-use, dynamic resource scaling and other characteristics that ensure cloud elasticity [3]. This helps to meet the end-users' varying service demands, with minimal resource wastage. In short, the utility-based pricing model and dynamic resource assignment model of cloud ensures that the user is only paying for the resources used [3]. The auto-scaling feature of cloud helps to realize the on-use resource scaling requirements through auto-allocation or de-allocation of resources; it can scale up or down based on the usage [4, 5]. The auto-scaling at VM level applies to the number of processing units, memory, networking components, etc [6]. Most presumably used auto-scaling metrics from the performance standpoint are threshold and duration [7]. These two metrics are considered necessary for triggering auto-scaling based on the need.

The value-added features of the cloud computing that distinguish it from other distributed and parallel computing model open an unfathomable number of possibilities for exploits and security breaches [8]. Among those, the improvised DDoS attack in the form of EDoS can exploit the cloud pay-per-use model to incur a huge financial loss to the customer or service provider [9]. The evolution of EDoS attack and the limitations in detecting and preventing such attacks has been explained in our previous works [2, 3]. The most threatening of EDoS attack involves exploiting the auto-scaling feature to cause scaling up of VMs in such a way to stealthily evade the security systems and the casual observations. The scaling EDoS attack if persists for a period of time runs VMs with junk data that adds to the offered service and impact heavily on service bill. This can damage the credibility of the service provider and can impact the business heavily and yet goes unnoticed. EDoS attack can adapt botnet model or applies any of the traditional DoS and DDoS attack model on paid services and can impact as EDoS attack.

1.1 Methodology

Cloud computing nowadays is used extensively on Web servers to enhance the request processing potential. This work therefore intents to perform an EDoS attack through iterating SQL-targeted DoS attack. This component does not compromise computers or host any malicious components in server; this helps to avoid security risks and to make sure that the components function without any security hiccups after the completion of the experiment. Moreover, this helps to understand how a simple attack module can constitute a harmful EDoS attack. The obtained results are then used to develop the detection and segregation algorithms that work hand-in-hand to monitor, detect and isolate the compromised VMs from genuine VMs. The Snort rules are then configured to filter and block the isolated VMs. This work improvises the segregation accuracy by incorporating simple and quick linear SVM model. The combined effort is then fitted into a framework to offer a wholesome EDoS handling service.

1.2 Organization

The full article is arranged in the following manner. Section 1 describes the introduction of cloud computing and its security implications from EDoS point of view. Section 2 presents a literature survey of EDoS detection, segregation and mitigation-related works. In Sect. 3, an experimental setting to perform a cloud Web server-targeted EDoS attack is narrated. In Sect. 4, the results of the performed experiment have been analyzed. Section 5 presents the proposed algorithms to detect and segregate attacking VMs or instances using linear SVM. Section 6 discusses the obtained results. Section 7 concludes the article by pointing out the future direction for research.

2 Literature Survey

[1] EDoS-ADS uses the threshold and the duration as auto-scaling parameters, an average CPU utilization threshold with two time triggers scaling-up and down are applied to differentiate VM scaling due to attack from normal. The mode switches to the suspicion mode from normal mode when the cloud CPU utilization exceeds the scaling-up utilization thresholds; then if the attack is found, it switches to attack mode if not it switches to flash crowd mode if the increase continues. In NAT run network, it can block the entire range of attacking nodes from its framework. In [10] EDoS-Shield, the main components of the architecture are virtual firewalls (VF) and verifier cloud nodes (V-nodes). The virtual firewalls are VMs with filtering and routing capabilities that work as filter mechanisms based on the white list to allow authenticated sources and blacklists to block unduly sources. The verifier cloud nodes (V-nodes) are a pool of VM that can verify its legitimacy through Turing tests, and the corresponding white or blacklists are updated based on the results of the verification process. In [4] enhanced EDoS-Shield, as an improvement to EDoS-Shield, it presents two algorithms which enhances both from the perspective of TTL field. Algorithm 1 describes the actions to be taken at VF node upon receiving a packet. A packet can be forwarded to the destination if only its source IP address and TTL value match the white list. Algorithm 2 describes the actions to be taken by a V-Node for not forwarded packets. In [11] EDoS ARMOUR, it is a Multilayered defense Architecture. At the first level, it applies the port hiding mechanism to deter

attackers from knowing the port to perform attacks. At the second level, a learning algorithm is used to monitor user behavior. If inappropriate behavior is spotted, then the corresponding user will get a slower service response. This helps to mitigate application-level DDoS attacks. It also entails a challenge mechanism, admission control, congestion control, user classification and client classification mechanism. Cloud eDDoS mitigation scrubber service in [12] is modeled as an on-demand EDoS-specific security service. Its core functional component implements crypto puzzle to be auto-generated and verified by the users or clients. The users are expected to solve the generated crypto puzzle through brute force to avail the requested service. This helps the CSP to follow the legitimacy of the users. The APART in [8] is a pattern recognition-based EDoS attack mitigation model. The number of packets sent by a user falls in the frequency of 400–800 per second; the APART model gets active and applies pattern recognition to detect EDoS attack. As a deterrence mechanism, it includes time-based and key-sharing post-setup authentication scheme to prevent the replication or replay attacks. It also provides a pre-shared security mechanism to ensure the access of legitimate users on the cloud services. In [9], an enhanced EDoS mitigation mechanism EDOS-EMM has been presented. The design involves three sequential modules, namely (1) data preparation, (2) detection and (3) mitigation modules. Module 1 involves flow monitoring, data collection and processing the flow to segregate it by protocol type and implements the sFlow agent algorithm. Module 2 detects the attack and extracts packet fields such as source, destination IP, port number and no. of packets per second. In this module, Hellinger distance and entropy methods are implemented for anomaly detection to improve accuracy. Module 3 generates alert, initiates rule update process and blocks the attacking IP. In [7] an entropy-based architecture is proposed for the detection of EDoS attacks. The proposed multilayered architecture involves monitoring and aggregation of metrics that affect the cost model, the novelty detection procedures to detect EDoS attack, and the decision-making and action response procedures. It is a predictive mode; it applies entropy variations related with considered metrics such as per-client CPU time when deciding auto-scaling actuations ahead of time for forecasting the EDoS attack possibility.

3 Experimental EDoS Attack

Performing EDoS attack involves components that can corrupt the servers, propagate across the Internet or flood the network or server with traffic with or without the knowledge of the attacker. Hence, performing EDoS attack is not safe in any platform. Therefore, even performing a simple DoS attack over the Internet is prohibited by law. However, without real-time experimentation of EDoS attack, the research cannot be effective. Consequently, an attack which targets the cloud Web server and performs VM scaling, but does not involve compromise, propagation and aftermath exploitation possibility, becomes ideal for our research. However, nowadays, Web servers run in cloud platform to meet any number of request; most of them use pay-per-hit model [13]. Therefore, if the Web servers are forced to process more requests, then it starts more VMs and ends up as a complicated EDoS attack. This requires writing a Web page and hosting it in a cloud Web server, and to use a minimal number of client computers to start consecutive DoS attacks together, they cause an EDoS attack. This way, single client computer can start any number of VMs in the server that stays and performs EDoS attack on server without compromising the facility.

The objective of this section is to perform a real-time EDoS attack on a Web hosting data center. The experimental setup involves a Windows Server 2016 and 25 machines connected through a LAN. The server runs in a cloud platform. The experimental scenario involves setting up a client Web page and creating db tables in server for Web access. Initially, the database for Web application is loaded in the server with various user names. This helps to ensure that none of the server which runs Web pages is targeted by the attack. The client sends the request; the server uses VMs to create a separate instance to process every request. Usually, once the request is complete processing, the VM is released. However, to perform database-targeted EDoS attack, it involves three steps. Step 1: The blah application is used to insert a user name and password from the client browser. Blah is a database manipulation command which helps to perform SQL injections and DoS attacks [14]. Step 2: To start a DoS attack, a 'blah' shell command is delivered to start a separate VM for every request as well as to performs a half-open TCP SYN-based DoS attack. The command **blah'; exec master.xp_cmdshell 'ping** target Web page **-165000 –t'**; – create a VM in server-side and that VM acts on behalf to reply to the client. It then performs a half-open TCP ping-based DoS attack. Step 3: The attack performs VM scaling by starting five VMs every 5-min interval from five computers through repeating step 2 for 5 times. This avails five VMs every interval. The cost of availing every VM for one minute is set to 0.0011 [8]. The hypervisor access is prohibited to avoid single-point failure because if it fails, the entire cloud facility becomes inaccessible. The VM or workloads of the experiment is monitored by task manager and Wireshark applications.

4 Experimental Results of EDoS Attack and Observation

The Windows Server on which the EDoS attack is performed is hosting the cloud-based file sharing application termed as 'ownCloud.' ownCloud becomes inaccessible due to the attack from the local client. Since every VM responds individually on behalf of the server, it becomes difficult for the firewalls to detect. The experiment has been conducted in a confined environment only after making sure that the Windows Defender Firewall and Web Shield are turned ON. However, they failed to detect the attack due to two reasons; one is the projected traffic which happens with in the server, that is, VM or frontal instances and Web server runs within the server due to the cloud incorporation. The firewall that looks for anomaly from the boundary of the server, therefore, fails to detect. Another reason is that the VMs send traffic in an individualized manner, so collectively, it will not shoot up at firewall observation point. Therefore, the firewalls fail to register the flood.

The results are observed from the perspective of traffic, processing and memory. According to Figs. 1 and 2, it is evident that the temporary memory usage is increasing gradually, i.e., from 1.9 (48%) to 2.1 (53%). Initiation of application involves starting and allotting space for VMs; therefore, more memory usage is evident in Fig. 1. However, in Fig. 2, due to the similarity of data processed and traffic generated, it did not reflect on memory or processing even after starting additional 20 frontal instances. However, for genuine application, each VM may process unique tasks, so it consumes more processing capability and reflects upon processing and memory.

The Web server due to the cloud implementation allows any number of frontal instances or VMs to be deployed by a single click from the client computer. It also allows every VM to act as the server to the client. At the same time, it manages to receive tremendous TCP hits that are generated from the VMs. In spite of all this, the server functions smoothly except for the subordinate application such as ownCloud.

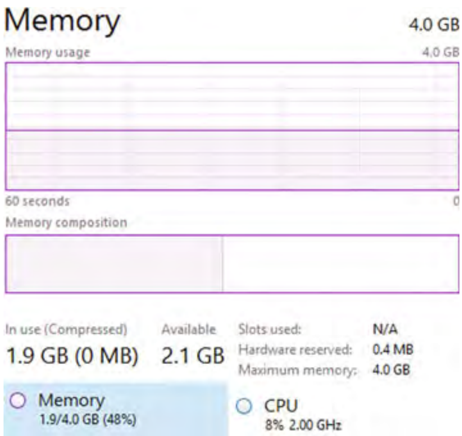


Fig. 1 Memory observation after starting of five VMs

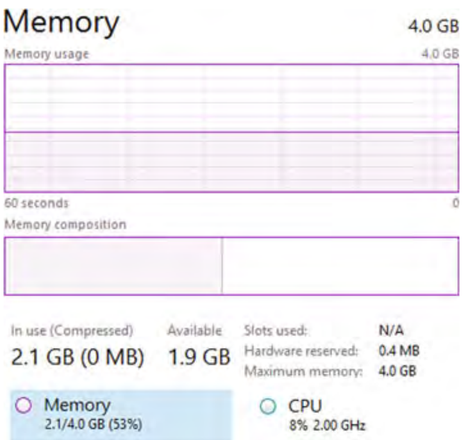


Fig. 2 Memory observation after starting of 25 VMs

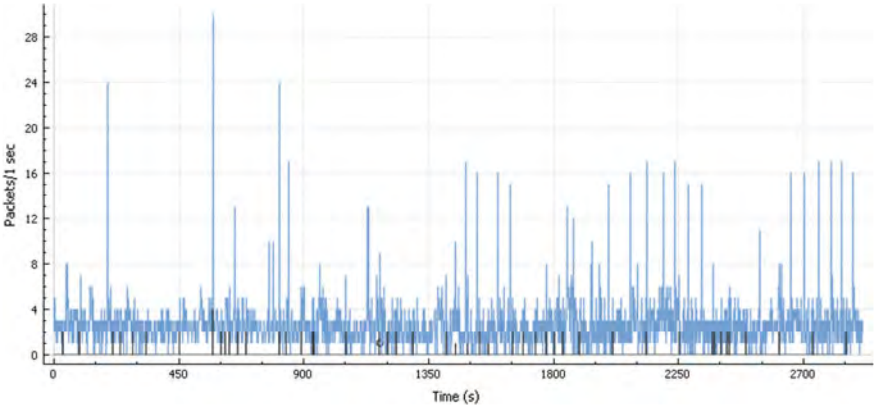


Fig. 3 Horizontal increase of incoming TCP packets from 25 VMs toward the server

This is because the Web service is prioritized over the ownCloud sharing application. The traffic anomaly is however observed with Wireshark, and the observation is presented as the following graph.

According to Fig. 3, the TCP traffic increases in horizontal manners in specific to the VMs, which is uncustomary for a flood attack. Moreover, occasional hits from the database (DB) or Web server were also visible from the graph, denoted as black bars. Consequently, the results show the requirement for individualized monitoring of VMs both from the perspective of the following auto-scaling metrics and behavior of VMs. Based on the experimental results, the following detection and segregation algorithms were designed.

5 Proposed EDoS-Barricade

The complications of cloud security arise from its unique features. The proposed EDoS-Barricade is therefore designed to look into the cloud features closely to eliminate the most threatening form of cloud attack, i.e., EDoS. Evidently after close examination of the experimental results, for initial EDoS detection, following the number of VMs with respect to a predetermined interval becomes necessary to deter the auto-scaling-based attack from exploiting the cloud pay-per-use model. The detection algorithm, therefore, requires to follow the VM to server communication and capturing to-and-fro packets. However, to achieve that, the Snort firewall has been chosen. The Snort is a widely used firewall for Linux, and it suits both research and commercial purposes and helps to build OS kernel-level solution for security implementation. However, Snort is tedious to work with, after careful tuning of Snort SO rules, pre-processor rules, etc.; it still lacks the proper command to detect VM to server traffic. After careful inspection, the following command is tailored to detect VM to server communication; so far, there is no reference to do that .

Table 1 Snort VM to server detection procedure

Innovative Snort command to detect in-server VM to server hits
alert tcp $HOME_NET any any -> $HOME_NET any flow:to_server; Seq: N; ack: A;
win: W;length: l; Options:O; classtype:attempted-user;

The above command helps to detect VMs with its IDs, as well as to follow all sorts of in-house traffic between VM and server communication; it can also follow server to VM communications.

5.1 Auto-Scaling-Based EDoS Detection

The 15 number of VMs is identified as the reasonable number for presetting detection metric from our experimentation standpoint. The chosen interval is 5 min. This number can be deduced by looking into cloud history. For instance, the average number of VMs running with respect to the time interval for a period of history, such as for a month and the year, can be computed to set this parameter. If the number of VMs increases for three consecutive intervals, then this algorithm raises the detection alert as well as switches mode to suspicious and calls the EDoS-BARRICADE Segregation () algorithm.

5.2 Support Vector Machine (SVM)

The EDoS segregation requires classifying data into two binary sets {EDoS, normal}. However, to achieve that, the SVM classifier is an ideal choice. SVM offers both linear and nonlinear mapping techniques to project and classify the dataset or input into 2-D or 3-D plane. This projection helps to deduce a hyperplane that can segregate data into its corresponding classes. The linear SVM can offer quick and accurate classification compared to other classifiers. Moreover, SVM can stretch to big data processing since it can handle outlier and overfitting effectively. The dataset during the preprocessing is usually depicted as follows;

$$(x_1,\ y_1),\ (x_2,\ y_2), \ldots (x_i,\ y_i), \ldots (x_n,\ y_n) \tag{1}$$

where x_i denotes the metrics that can help to classify the given data; from our observation, they are incoming and outgoing packets in specific to VMs, n denotes the maximum number of values, and i denotes the iteration of values from 1 to n.

$$\text{Class}x_i \in \{\text{Incoming\_pkts},\ \text{Outgoing\_pkts}\} \tag{2}$$

The class y_i serves as a label and helps to separate the VM data into its corresponding class.

$$\text{class}\, y_i \in \{\text{EDoS,normal}\} \tag{3}$$

Given the datasets x_i and $y_{i,}$ the SVM classification engine naturally creates a hyperplane as follows

$$w^{\mathrm{T}} * x + b = 0 \tag{4}$$

The hyperplane is a separator that determines the maximum margin between two classes, where w is a weight vector and b is a bias. The SVM automatically creates weight vector and bias to group the data using a hyperplane. The grouping is done as follows.

$$\text{if} \quad w^{\mathrm{T}} x + b \geq +1 \text{ then normal} \tag{5}$$

$$\text{if} \quad w^{\mathrm{T}} x + b \leq -1 \text{ then attack} \tag{6}$$

However, the closest point called a support vector from the respective classes can only establish the optimally separating hyperplane. Support vector from the given dataset determines the hyperplane by using:

$$\min(\tau) = \left\{ {}^{1}\!/\!{}_{2} ||w||^{2} \right\} \tag{7}$$

It has been shown that the optimal separating hyperplane can be found by minimizing (7), with respective to

$$y_i (w,\ x_i + b) \quad \text{for all} \quad i = 1, \ldots, n \tag{8}$$

This is a convex optimization problem; the SVM classifier solves it with Lagrange multipliers

Algorithm:1 EDoS-BARRICADE Detection ()
Step 1: Monitor increase or decrease in workloads or VM scaling Step 2: If no. of VM increases above the default limit then initiate EDoS detection Step 3: If it continues to increase for consecutive intervals, raise detection alert and spontaneously call the segregation procedure

(continued)

(continued)

Algorithm:1 EDoS-BARRICADE Detection ()
Input: {List of VMs with unique identifier, Interval} Output: {List of VMs, No. of, Interval, Total No. of incoming & Outgoing pkts} Define VM_Number; \\ Number of workloads Define interval = 5 min; \\ Time Interval Define i = 0; \\Increment Variable For each application; if VM_number > 15 then Monitor; For each interval; if current_VM_No > previous_VM_No; { i +1 if i > 3 then raise detection alert Switch Mode to Suspicious Call EDoS-BARRICADE Segregation () }

5.2.1 Linear SVM

The dataset contains the non-separable case of data; however, by introducing a cost for violating constraint (8) such as a positive slack variable ξ, a hyperplane can be realized for optimal separation.

$$y_i.(x_i.w + b) \geq 1 - \xi_i. \quad \text{For all } i \tag{9}$$

It can be furthered for erroneous cases with cost parameter C, and such a scenario is not required for this research.

5.2.2 Nonlinear SVM

If the data is not mapable using linear decision function, the data will be mapped into higher dimensional feature space through a nonlinear transformation which incorporates kernel functions. Vastly used kernel functions for binary classification problems are as follows;

$K(x_i,x_j) = x_i x_j$: *linear SVM*
$K(x_i,x_j) = (x_i x_j + 1)^p$: *polynomial of degree p*
$K(x_i,x_j) = \tanh(a.x_i x_j + b)$ *and* $a > b$: *multi-layer perceptron* (*MLP*) *classifier*
$K(x_i,x_j) = \exp\{||x_i - x_j|^2/^2\}$: *radial basis function* (*RBF*) *classifier.*

5.3 Behavior Heuristics-Based EDoS Segregation

The training dataset has been generated during the monitoring with respect to VMs. The data is provided to an SVM-based EDoS-BARRICADE segregation algorithm to differentiate malicious VMs from normal VMs.

Algorithm:2 EDoS-BARRICADE Segregation ()

Step 1: Upon invocation from the detection algorithm; Gather the data.
Step 3: Perform anomaly-based segregation by applying relevant SVM.

Input : {List of VMs, No. of, Interval, Total No. of incoming & Outgoing pkts}
Output: {Scatter plot, Attacking VM list for snort to filter and block }
Call SVM Classifier ()
For each VMs;
{ if $wT\ x + b \geq +1$
then project as normal
if $wT\ x + b \leq -1$
then project as EDoS }
For all VMs;
{if linearly separable
Deduce hyperplane $yi.(xi.w + b) \geq 1 - \xi i$
else apply kernel}
project result in scatter plot
Generate Attack VMs list
Call Snort()
Perform Filtering and Blocking

The above algorithm presents the result for input dataset instantaneously; it can be plotted into two classes separated with a hyperplane. Moreover, the output presents a list of attacking VMs which can be blocked and filtered out. The following sections examine the result of EDoS-BARRICADE through varying the input.

6 EDoS-Barricade Result Observation and Verification

The experimental data along with normal VM dataset obtained from GitHub has been fed into RapidMiner. The linear SVM has been simple enough to classify the normal flow from the attack flow as follows.

According to Fig. 4, attacking VMs distinguishes itself with the following heuristics, i.e., for attacks, 'Accumulated incoming packet > Accumulated normal Incoming packets' and 'Accumulated Outgoing packet < Accumulated normal Incoming packets.' Moreover, to intensify the testing, data migration dataset obtained from GitHub has been incorporated into the dataset, and then, testing is performed. For that, the SVM is required to transform data into a higher dimensional plane using a kernel as shown in Fig. 5.

According to Fig. 5, the attack instances become as blue dots becomes classifiable using SVM kernel when the data migration dataset is involved. However, this is an

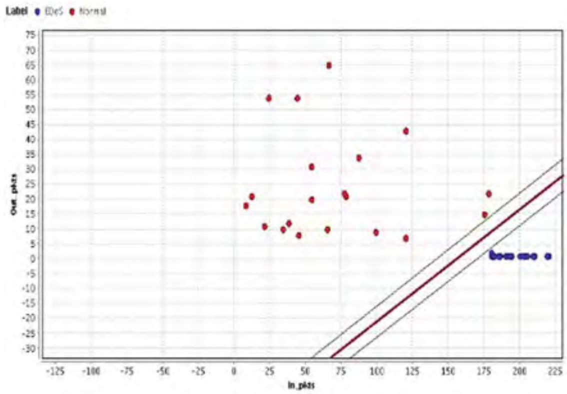


Fig. 4 Linear SVM segregation result after the third interval

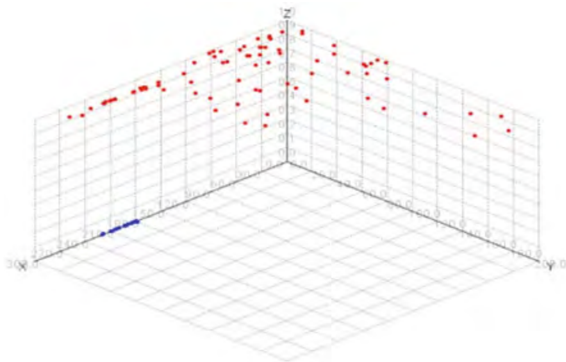


Fig. 5 Nonlinear SVM segregation result for a dataset with data migration instances

unusual and rare happening; for other cases, simply linear SVM proves sufficient to detect EDoS attack. The accuracy achieved for EDoS detection after the third interval is 100%. The obtained SVM-based EDoS-BARRICADE result is then interpreted as the list of attacking VMs and forwarded to the Snort firewall for filtering and blocking.

7 Conclusion and Future Scope

The EDoS attack is performed on any cloud Platform through exploiting the auto-scaling feature to increase the number of VMs, which in reality processes junk data and floods the server with junk requests. This phenomenon manages to exploit the cloud usage-based pricing model and impact heavily on customer bills. However, to have a real-time perspective, a Web server-based EDoS attack is performed, and the impact of the attack is observed. The result shows that the EDoS attack is detectable by following the auto-scaling and segregate-able by following incoming and outgoing packets. In accordance with the results and through making necessary improvement to the Snort, a quick auto-scale-based EDoS-BARRICADE detection algorithm has been proposed. The tedious part involves isolating attack VMs from the normal ones without false detections. An EDoS-BARRICADE segregation algorithm which involves linear SVM as its base classifier is proposed and tested with datasets involving normal and attack data. The linear SVM performs classification with 100% accuracy since the normal data has more number of incoming packets and less number of outgoing packets than the attack data. However, more rigorous testing has been performed by including data migration samples to the dataset. In such cases, the algorithm switches to nonlinear kernel-based SVM classifier to isolate the attack class in higher dimensional plane. In the future, the presented work will be extended into a framework with more added characteristics.

References

1. Shawahna A, Abu-Amara M, Mahmoud ASH, Osais Y (2018) EDoS-ADS: an enhanced mitigation technique against economic denial of sustainability EDoS attacks. IEEE Trans Cloud Comput (Feb)
2. Ribin Jones SB, Kumar N (2019) Unraveling the security pitfalls that stem from core cloud benefits through analyzing various DoS attacks, detection and prevention. J Adv Res Dyn Control Syst 11(09):541–553 (Aug)
3. Ribin Jones SB, Kumar N (2019) Precursory study on varieties of DDoS attacks and its implications in Cloud Systems. In: IEEE Xplore third international conference on trends in electronics and informatics, pp 1003–1008, Apr 2019
4. Sqalli MH, Al-Haidari F, Salah K (2012) Enhanced EDOS shield for mitigating EDoS attacks originating from spoofed IP addresses. In: IEEE eleventh international conference on trust, security and privacy in computing and communications
5. Kumar D (2019) Review on task scheduling in ubiquitous clouds. J ISMAC 1(01):72–80
6. Bulla S, Basaveswara Rao B, Gangadhara Rao K, Chandan K (2018) An experimental evaluation of the impact of the EDoS attacks against cloud computing services using AWS. Int J Eng Technol 7(1.5):202–208
7. Monge MAS, Vidal JM, Villalba LJG (2017) Entropy-based economic denial of sustainability detection. MDPI Entropy 19(12) (Nov)
8. Thaper R, Verma A (2015) Adaptive pattern attack recognition technique (APART) against EDoS attacks in cloud computing. In: IEEE third international conference on image information processing
9. Singh P, Rehman SU, Manickam S (2017) Enhanced mechanism to detect and mitigate economic denial of sustainability (EDoS) attack in cloud computing environments. Int J Adv Comput Sci Appl 8(9)
10. Sqalli MH, Al-Haidari F, Salah K (2011) EDOS shield—a two-step mitigation technique against EDoS attacks in cloud computing. In: IEEE fourth international conference on utility and cloud computing
11. Masood M, Anwar Z, Raza SA, Hur MA (2013) EDoS Armor: a cost effective economic denial of sustainability attack mitigation framework for e-commerce applications in cloud environments. In: IEEE sixteenth international multitopic conference
12. Kumar MN, Sujatha P, Kalva V, Nagori R, Katukojwala AK, Kumar M (2012) Mitigating economic denial of sustainability (EDoS) in cloud computing using in-cloud scrubber service. In: Fourth international conference on computational intelligence and communications networks

13. Abusitta A, Bellaiche M, Dagenais M (2018) An SVM-based framework for detecting DoS attacks in virtualized clouds under changing environment. Springer J Cloud Comput: Adv Syst Appl 7(9):1–18 (April)
14. Mazrekaj A, Shabani I, Sejdiu B (2016) Pricing schemes in cloud computing: an overview. Int J Adv Comput Sci Appl 7(2):80–86
15. Duraipandian M, Vinothkanna R (2019) Cloud based internet of things for smart connected objects. J ISMAC 1(02):111–119
16. Udhayan J, Hamsapriya T, Vasanthi NA (2012) DDoS attack detection through flow analysis and traffic modeling. In: SPIT 2011, LNICST 62, pp 89–94
17. Baig ZA, Sait SM, Binbeshr F (2016) Controlled access to cloud resources for mitigating economic denial of sustainability (EDoS) attacks. Elsevier Comput Netw 97:31–47 (Mar)

User Query-Based Automatic Text Summarization of Web Documents Using Ontology

K. Selvakumar and L. Sairamesh

Abstract Web document summarization is a very important function for knowledge management when documents are very huge and dynamic. This helps the document readers to easily read and understand. In general, summarization is done based on the extraction of sentences from the retrieved documents and assembled with the needed proposition. Text content plays a major role in the Web, and extraction of text content is a challenging task. Recently, one of the problems that arise with the rapid growth of the Web and general information availability (sometimes referred to as an information overloading) is the increased need for effective and powerful text summarization. In this paper, a system is built for automatic text summarization from the source documents which is retrieved from the Web. The proposed method considers the ontology approach to extract the text summary according to query terms, which are existing in the ontology graph based on their depth. Also, attributes are used to improve the semantic representation of a sentence's information from contents.

Keywords Information retrieval (IR) · Text summarization · Ontology · Automatic text summarization (ATS)

1 Introduction

Automatic text summarization (ATS) is a precise version of a text provided by a computer program [1]. Normally, summarization is the one that summarizes the content of the document without changing the actual meaning of the document. In this task, summarizing the content with the same synonyms is a challenging task. To solve this challenge, text summarization tries to summarize the text content which

K. Selvakumar (✉)
Department of Computer Applications, National Institute of Technology, Trichy, India
e-mail: kselvakumar@nitt.edu

L. Sairamesh
Department of IST, CEG Campus, Anna University, Chennai, India
e-mail: sairamesh.ist@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_45

has to provide brief information on the content. Some of the tools are available for summarization which searches the headings and subtopics relevant to the key points given in the document and gives the content [2, 3].

Mostly, text summarization can be classified into two different ways. One is the extractive method which consists of important sentences from the given paragraph. It extracts the important sentence based on the frequency of words and linguistic features of the word. The next is abstractive summarization, which is mainly based on language concepts or linguistic features of the context. It uses the natural language processing approach to examine the concepts and express them in the shortest context than the available version. In this paper, the context is summarized based on extractive summarization.

In the operation of extractive summarization, the key segments are formulated based on a statistical analysis of the text. The features of sentences are extracted based on word frequency and location of words. The frequent words are taken for the context by treated as the most important word in the context. This type of extraction makes the user to easily understand the concept without having deep knowledge of the topic. The process of extractive summarization is classified as sentence boundary identification, stop word elimination, stemming, tagging, frequent word identification, and scoring of sentence [4, 5].

In the processing, features of the sentences are computed and selected based on the influence of the words in the sentences, and weights are assigned based on the context features. The ranking is done by using the feature—weighting equation and sentences chosen for the final summary.

Problems with the existing techniques for generating extractive summary are:

(a) The sentence which is ranked high may also contain some unwanted information.
(b) Relevant information is available across the sentences, and hence, the extractive summaries will not be able to capture this sometimes due to the length of the sentence.
(c) Sometimes, conflicting information may not be presented precisely.

Interpretation of anaphors among similar sentences may lead to include some unwanted information in summary. Similar challenges are present with time-based expressions. These problems most commonly arise in multi-document summarization, since information is collected from different sources. In such a scenario, post-processing can be used to extract anaphors by replacing the pronouns with their antecedents by replacing the relative temporal expressions represented using dates.

2 Current Practice and Research

Many works are proposed by most of the linguistic and information retrieval researchers for multi-document summarization. The work proposed in [4] used A* search technique with discriminative training approach for summarizing the multiple documents. In this approach, A* is used to search and provide the relevant words for the given keywords and make the summary relevant to the query. In [6, 5], multi-document summarization is done by using the ontology which easily extracts the meaning of the keyword and finds the appropriate words for the keyword and summarizes the content. Some other works in [7–9] explain various summarizations for the single and multi-document. In [10], the author describes how to evolve the user groups which easily evaluate a summary for the group of people. In [11], summarizations are carried out based on the relationship between the contexts using natural language processing.

3 Ontology Approach in Text Summarization

The usage of ontology is increased due to the increase in word usage in different ways. Gruber's define "ontology as an explicit specification of the topics." In general, ontology is a formal representative of the vocabulary based on the logical definition of the words and statements where it is used. It provides a formal way to easily get the relevant information for the given keywords. Assume that large articles contain many subtopics within the main topic. So, the keywords available in the subtopic have to be identified, and it makes the summarization easier and meaningful.

Figure 1 shows the architecture of the text summarization system proposed in this paper. In this proposed method, user query will be preprocessed (such as stop word removal and stemming), and key terms are extracted. Each term is compared with the existing ontology graph, and weight will be assigned according to the depth where they exist in the ontology graph [12]. Here, the ontology graph can be constructed using the protégé tool. The key terms which are present in the ontology graph will be estimated and ranked; otherwise, the value of those terms will be assigned as −1. Based on the rank, the key terms are chosen, and their corresponding statements will be extracted from various multiple Web documents. Before this step, the Web documents are preprocessed. After extracting the sentences from the documents, their similarity can be measured by using similarity measures (such as cosine similarity). The above process can be applied to specific sports domain areas.

Let A be the selected features. For each candidate document x_i, its dynamic features used for scoring (Eq. 1) include the similarity measure between the sentence and user query.

$$\mathrm{Score}(x_i) = \frac{(\Psi \mathrm{Sim}_1(x_i, Q) + (1 - \Psi)\mathrm{Sim}_2(x_i, A))n(x_i)}{N} \tag{1}$$

where

N Total number of documents.

$n(x_i)$ Number of sentences having similarity measure above the predefined threshold level.

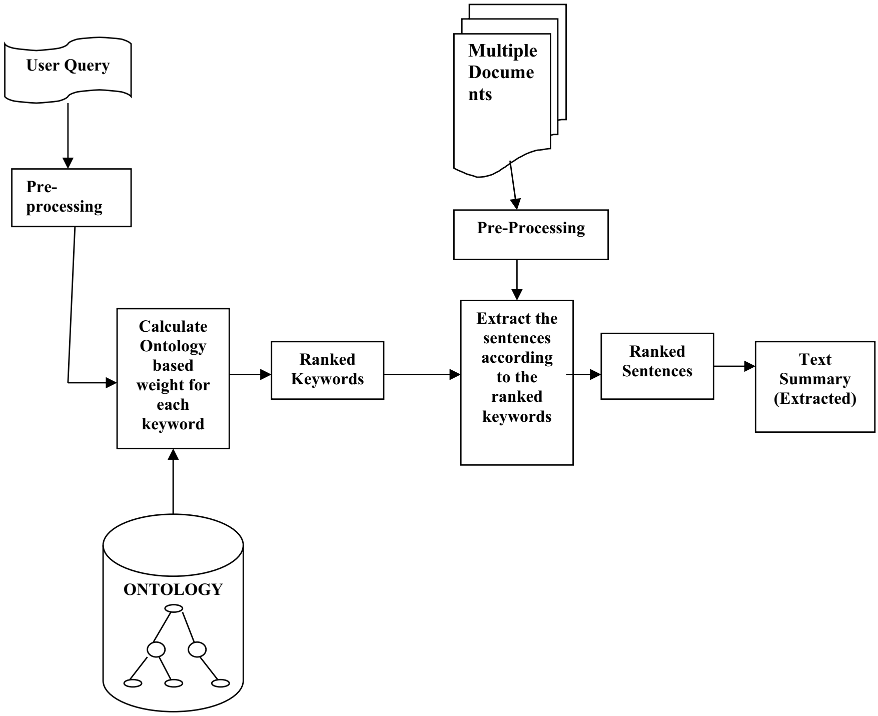


Fig. 1 System architecture for text summarization

Ψ Parameter with values [0, 1] that control relative importance given to relevance versus redundancy.

Sim_1 and Sim_2 are similarity measures (Eq. 2) with documents (x, y) selected feature (research area) and user query.

$$\mathrm{Sim}_1(x, y) = \mathrm{Sim}_2(x, y) = \frac{\langle x, y\rangle}{\|x\| \cdot \|y\|} \tag{2}$$

To represent the structural information of the words and their taxonomy, it propagates feature constraints from the leaf nodes to the parent nodes by recursively aggregating the children feature weights and assigning them to the parent:

$$\omega'_c(t) = \omega_c(t) + \sum_{i \in \mathrm{children}(c)} \omega_i(t) \tag{3}$$

where

$\omega_i\ (t)$ is the TFIDF weight of term t in child category i.

4 Results and Discussions

The results for text summarization are carried out with the different set of documents which are varied in size (MB). Table 1 shows the summary of the data set which is collected from newspaper sites under different topics such as sociology, sports, education, entertainment, and medicine.

The content of the document is taken as the average sentence, and experiments are executed, and the results are compared with the existing systems. Table 2 shows the summary accuracy of the proposed and existing systems. The proposed system shows better accuracy than the available system for summarization. Moreover, the result shows that the proposed system performs better both semantically and also in content summarization than [2] which follows latent semantic analysis and [13] uses hierarchical summarization for multiple documents. Table 3 shows the comparison of the summarized content of the existing and proposed systems. The proposed system summarizes the content of different documents as a single document and provides a semantically acceptable summarization.

Both the accuracy and summarization are better proposed than the existing system. And also, the time taken for summarization is calculated for all experimental systems as shown in Table 4, where most of the researchers not concentrated on the timing constraints in summarization. In this manner, our proposed system is performing better than the existing approaches.

Table 1 Data set for summarization

Data set (DS)	Number of documents	Average number of sentences per document
DS1	35	1230
DS2	58	1080
DS3	75	950
DS4	63	750
DS5	70	930

Table 2 Summary accuracy

Data set	Existing system [2] (%)	Existing system [13] (%)	Proposed system (%)
DS1	84	92.4	96.5
DS2	90.8	93.5	97.4
DS3	92.3	94.7	96
DS4	87.5	91	97.8
DS5	91.5	95	98.4

Table 3 Comparison of summarized content

Data set	Avg. no. of sentences per document	No. of sentences in summarized content		
		[2]	[13]	Proposed system
DS1	1230	1750	1680	1500
DS2	1080	980	1100	930
DS3	950	2360	2180	1800
DS4	750	1950	1800	1730
DS5	930	1680	1550	1420

Table 4 Time is taken for summarization

Data set	Existing system (ms) [2]	Existing system (ms) [13]	Proposed system (ms)
DS1	21.4	18	9.5
DS2	15	16.5	8.4
DS3	57.5	56.3	14.8
DS4	54	48	13
DS5	46	35	12.6

5 Conclusion and Future Work

This user query-based automatic text summarization is an old challenge, but the current research direction leans toward emerging trends in biomedicine, product review, education domains, emails, and blogs. This is because there is information overload in these areas, especially on the World Wide Web. This paper examines recent advances and challenges of automatic text summarization in general and explores some emerging trends on automatic text summarization using the ontology approach. Pertinent issues persist in automatic text summarization, especially that of achieving summarizations that are close to those produced by human linguists. However, even expert summarizations have slight differences. Future work includes extractive summarization using a fuzzy neural network approach to enrich its efficiency.

References

1. Mitkov R (2005) The Oxford handbook of computational linguistics. Oxford University Press
2. Ozsoy MG, Cicekli I, Alpaslan FN (2010) Text summarization of Turkish texts using latent semantic analysis. In: Proceedings of the 23rd international conference on computational linguistics, 2010. Association for Computational Linguistics, pp 869–876
3. Ramesh LS, Ganapathy S, Bhuvaneshwari R, Kulothungan K, Pandiyaraju V, Kannan A (2015) Prediction of user interests for providing relevant information using relevance feedback and re-ranking. Int J Intell Inf Technol (IJIIT) 11(4):55–71

4. Aker A, Cohn T, Gaizauskas R (2010) Multi-document summarization using A* search and discriminative training. In: Proceedings of the 2010 conference on empirical methods in natural language processing. Association for Computational Linguistics, pp 482–491
5. Li L, Wang D, Shen C, Li T (2010) Ontology-enriched multi-document summarization in disaster management. In: Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. ACM, pp 819–820
6. Guran A, Bekar E, Akyokus S (2010) A comparison of feature and semantic-based summarization algorithms for turkish international symposium on innovations in intelligent systems and applications. Kayseri Cappadocia, Turkey, pp 21–24
7. Litvak M, Last M, Kisilevich S, Keim D, Lipman H, Gur AB (2010) Towards multi-lingual summarization: a comparative analysis of sentence extraction methods on English and Hebrew corpora. In: Proceedings of the 4th workshop on cross-lingual information access, pp 61–69
8. Naderi N, Witte R (2010) Ontology-based extraction and summarization of protein mutation impact information. In: Proceedings of the 2010 workshop on biomedical natural language processing. Association for Computational Linguistics, pp 128–129
9. Foong OM, Oxley A, Sulaiman S (2010) Challenges, and trends of automatic text summarization. Int J Inf Telecommun Technol 1(1)
10. Selvakumar K, Ramesh LS, Kannan A (2015) Enhanced K-means clustering algorithm for evolving user groups. Indian J Sci Technol 8(24):1
11. Chakrabarti P, Basu JK (2010) Text summarization, and discovery of frames and relationship from natural language text-A R&D methodology. Int J Comput Sci Eng 2(3):487–492
12. Wan X, Yang J (2008) Multi-document summarization using cluster-based link analysis. In: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. ACM, pp 299–306
13. Christensen J, Soderland S, Bansal G (2014) Hierarchical summarization: scaling up multi-document summarization. In: Proceedings of the 52nd annual meeting of the association for computational linguistics, pp 902–912

Quantum Chaos-Based Encryption Technique for Transmission of Medical Images

R. Anitha and B. Vijayalakshmi

Abstract With the fast advancement of networked devices and growth in usage of digital data transmission, hawk-eye attention is needed to provide the secured transmission for various vital data such as medical images and digital signatures. Even though many encryption algorithms were proposed, achieving high-security transmission remains in the darker side of the research because of its low sensitivity and less complexity. Recently, usage of quantum chaos has gained more insight in the encryption world due to its complex nature. A new encryption scheme quantum chaos-based network-centric encryption for data transmission (Q-CNEST) is proposed to integrate with the network characteristics as its initial conditions for generation of high complex encrypted data. As a first step, cipher keys are generated by the quantum chaotic logistic maps using the random network parameters. Following that, diffusion and permutation of each pixel with the high randomness cipher keys have been performed. Finally, sensitivity analysis is done. Simulation has been implemented in the Qiskit packages, and its results have proven that the proposed algorithm has good randomness which will be more suitable for the prevention of conventional attacks over the medical information. Moreover, the proposed encryption algorithms can provide more limelight for the secure transmission of medical image transmissions.

Keywords Cipher keys · Qiskit packages · Digital imaging and communication in medicine (DICOM) · 2D spatiotemporal chaotic-based encryption · Internet of things (IoT) · Long-term evaluation (LTE) · Received signal strength (RSSI)

1 Introduction

In the current scenario, most of the industrial applications are highly associated with the revolution of technology. The advancements in technology have revolutionized medical field and telemedicine. The security of patient health record is imperative

R. Anitha (✉) · B. Vijayalakshmi
B.S.Abdur Rahman Crescent Institute of Science & Technology, Chennai, India
e-mail: anitharajesh29@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_46

from security threats. Digital imaging and communication in medicine (DICOM) is a standard that focuses on security issues of telemedicine. It insists on security services like privacy, reliability and authenticity. The secrecy of transmitted images is achieved by symmetric encryption algorithms [1], whereas reliability and authenticity are achieved by hashing and digital signatures. But all these classical encryption techniques encounter a complex computation. Hence, the development of an efficient algorithm has turned out to be an interesting research field. Chaos theory has many attractive features, like sensitivity to initial values, pseudorandomness and ergodicity and low cost in the computer operating system. Hence, a chaotic system suite best for encryption process. In recent years, chaos-based image encryption algorithms have become an interesting research domain.

However, the current chaotic system still needs improvisation in terms of high randomness and more complex cipher keys [2] because they may not resist too many conventional attacks such as brute force attacks. Hence, improving the randomness and keyspace requires brighter light of research. Also, a discrete chaotic system is often adopted as a catalyst for encryption schemes which lags in the complex behaviors, in which the generated sequences can be predicted easier.

Generally, the chaotic systems are classified as discrete and continuous chaotic systems. For all digital applications, continuous chaotic systems have to be converted into discrete data, and then, it should be digitized. There are many discretization methods like Euler method, Runge–Kutta method [3], etc. But, discrete chaotic systems are more attractive toward digital application. The behaviors of these systems slowly degenerate leading to the limited accuracy. In contrast, the continuous system has a complex structure, and this complex variable chaotic system can be applied in an encryption algorithm to improvise the security performance. Also, it can expand the variable space of the system that improves the dynamical characteristics. The complex chaotic system creates a challenging domain for malicious encoding. Hence, this system can be considered to enhance security performance against different malicious attacks [4, 5].

The classical chaos systems which hang about universally differ from the quantum chaos systems that rely on quantum mechanics theory. Quantum logistic system is built from the classical logistic system. Also, this logistic map uses the repetition procedure resulting in a chaotic behavior depending on its initial parameter. This system results in high dimension and complex dynamic behavior. Hence, the quantum logistic system is suitable for an efficient encryption algorithm. Enormous research papers have been proposed based on the quantum logistic system [6].

Considering all the above analysis, the novel encryption scheme Q-CNEST has been proposed which is based on a quantum chaotic system whose initial conditions are designed based on the dynamic varying network characteristics such as signal strength and power consumption of the system followed by the double tier permutation [7]. The following steps were adopted for the formation of novel encryption schemes.

1. Dynamic quantum chaotic key formation using network characteristics such as signal strength and power consumption modes.

2. Formation of new encrypted data with permutation between the cipher keys and information. (Tier-I mechanism)
3. Formation of new encrypted data by the process of diffusion (Tier-II mechanism).

The organization of the paper is as follows.

Section 2 deals with related works. Section 3 deals with the background work of the logistic maps. Section 4 describes the proposed architecture. Experimentation setup of the proposed system is explained in the Sect. 5, and the results with comparative analysis are discussed in Sect. 6. Conclusion along with future scope has been discussed in Sect. 7.

2 Related Work

Yi He et al. proposed a 2D spatiotemporal chaotic-based encryption algorithm. Both linear and nonlinear chaotic map lattices are considered. 2D-coupled map lattices and permutation process are incorporated to enhance security [8].

Moatsum et al. proposed a hybrid chaotic system using the perturbation process. In this system, confusion and diffusion processes are implemented involving pixel shuffling and substitution. Security and performance analysis are observed using various analyses [9].

Qiang Lai et al. proposed a 4D chaotic system. An infinite number of chaotic attractors is produced. The initial conditions for a chaotic system are obtained using a sine function. The attractors depend on the equilibria to determine the attraction location. They have suggested that increasing the equilibrium points may generate multiple attractors [10].

Ranjeet Kumar Singh et al. presented a novel method to utilize a security system for the transmission of computerized substance over the open system. The proposed strategy utilizes a staggered picture encryption/unscrambling calculation dependent on quantum confusion map and meager inspecting. In the underlying stage, a unique picture is partitioned into squares of equivalent estimate, and each square is additionally isolated into sub-square utilizing DWT method. The pixels of the nearby sub-squares are traded arbitrarily by an irregular network. Every recurrence band of the squares is encoded by modulus capacity and consolidated each to get the new square. Next, pixels of the neighbor squares are traded haphazardly by arbitrary network, and along these lines, each square is scrambled utilizing modulus work. The outcome given in graphical and unthinkable structure shows the properness of the calculation [11].

Guodong Ye et al. proposed a picture encryption calculation dependent on a confused guide and data entropy. In contrast to Fridrich's structure, the proposed technique contains the change, regulation and dissemination tasks. This strategy stays away from the weakness in conventional plans of carefully rearranging the pixel positions before dispersion encryption. Data entropy is utilized to impact the age of the keystream. The underlying keys utilized in the change and dispersion stages communicate with one another. Thus, the calculation goes about as a resolute substance to upgrade security. Test results and security investigations exhibit the great execution of the proposed calculation as a safe and viable specialized strategy for pictures [12].

3 Background Work

3.1 Logistic Maps

Chaotic systems are characterized by initial sensitivity, randomness and high unpredictability. The mathematical expression for the logistic chaotic maps is given by

$$Xn + 1 = \mu Xn(1 - Xn) \tag{1}$$

where $0 \leq \mu \leq 4$ represents a bifurcation parameter. Generally, bifurcation occurs when a small change leads to an unpredictable system. A variation in a parameter leads to a change in differential system. An equilibrium may be unstable and results in either periodic solution or new stable equilibrium that makes the previous equilibrium unstable. The parameter which creates this change of state is called a bifurcation parameter. The initial value $x_0 \in (0,1)$ iterates the sequence $x1$, $x2$, …, xn, when $3.5699456 < \mu \leq 4$ leads the logistic system *i* sin a chaotic state. Moreover, quantum logistic chaos which was proposed by Goggin et al. 1990 has the following mathematical expressions

$$\left\{ \begin{array}{c} a_{n+1} = C * (a_n - |a_n|) - C * f_n \\ b_{n+1} = -f_n * h^{-2\alpha} + h^{-\alpha} C[(2 - 2a_n) f_n - 2a_n d_n] \\ e_{n+1} = -d_n * h^{-2\alpha} + h^{-2\alpha} C[2(1 - a_n) d_n - 2a_n f_n - a_n] \end{array} \right\} \tag{2}$$

where C is the control parameter, α is the dissipation constant, a_n and d_n are considered to be the complex conjugate part of their real counterparts. By varying the initial conditions, the characteristics of the above-mentioned logistics chaotic maps also vary. When compared with the chaotic maps, quantum maps add the disturbance at the end of the process. Because of the sensitivity of the initial conditions, a very small change in information may lead to the production of different sequences. Due to the variation in each iteration, the above chaotic logistic maps will produce more nonlinear characteristics which suit the best for applications related to image encryption [13, 14].

4 Proposed Encryption Process

The whole quantum encryption process for the proposed scheme is as follows,

(a) Generation of pseudo-random generators using network-centric initial conditions in quantum logistic chaotic maps.
(b) Permutation process between the quantum chaotic maps and image information.
(c) Chaotic diffusion process.

4.1 Network-Centric Logistic Chaotic Maps

Since the paper proposes the encryption process for the next-generation networks such as the Internet of things (IoT) and long-term evaluation (LTE), the pseudo-random keys are generated by using the network-centric MAC parameters such as received signal strength (RSSI) and ID of the channel. The mathematical expressions for calculating the RSSI are given as follows

$$\text{Received Signal Strength Indicator (RSSI)} = -(10.\eta.\text{log(d)} + \text{A}) \tag{3}$$

where

$$d = 10\left[\frac{(P_o - F_m - P_r - 10n\log(f) + 30n - 32.44)}{10n}\right] \tag{4}$$

P_o Power of the signal (dBm) in the zero distance,
P_r Signal power (dBm) in the distance d,
F Signal frequency in MHz,
F_m Fade margin,
N Path-loss exponent.

These are the initial cmaps given in Eq. (2). Figure 1 represents different values of RSSI which are used as the initial conditions for chaotic behavior. Figure 1 illustrates the chaotic behavior of the proposed systems with different network parameters. From Fig. 1, it is clear that the system has unrelated chaotic circular behaviors with the different parameters which find its suitability in the encryption process.

4.2 Quantum Permutations

To implement strong protection, quantum-based XOR operation has been done. It has been implemented with the help of Hadamard basis and CNOT gate in QISKIT tool. The sequences which are obtained from the above process are sorted in ascending order without changing the nature of the bits [15, 16]. Let S = {S1, S2, S3, … Sn}

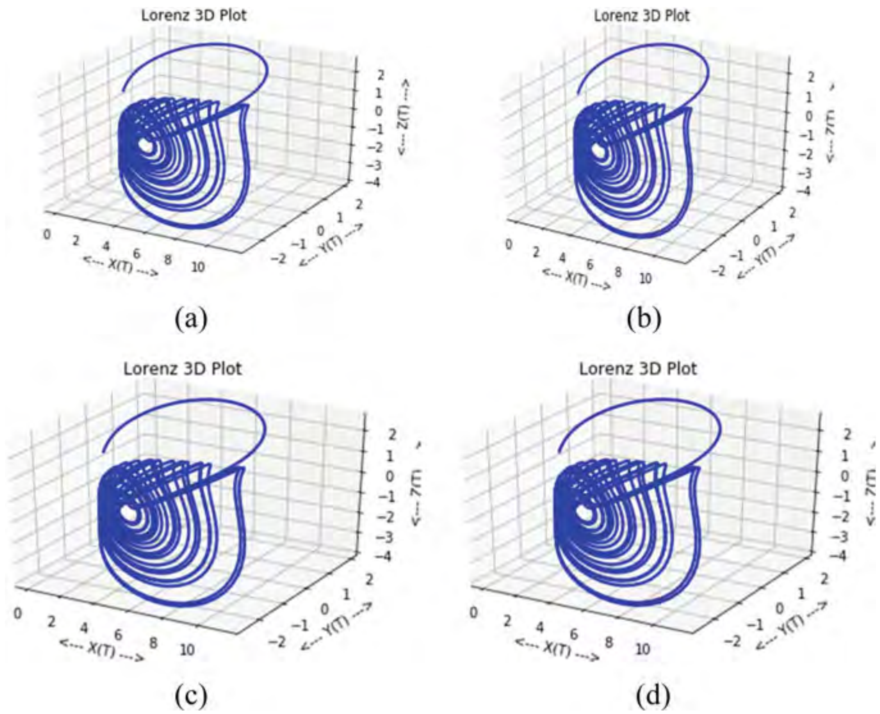


Fig. 1 Chaotic behaviors **a** 48 dbm, **b** 38 dbm, **c** 24 dbm, **d** 19 dbm

be the pseudo-random sequences and sorted sequences are given by S″ = {S1, S2, S3, S4, … Sn}. These sequences are then XORed with bit image information whose quantum implemented circuit is given in Fig. 2.

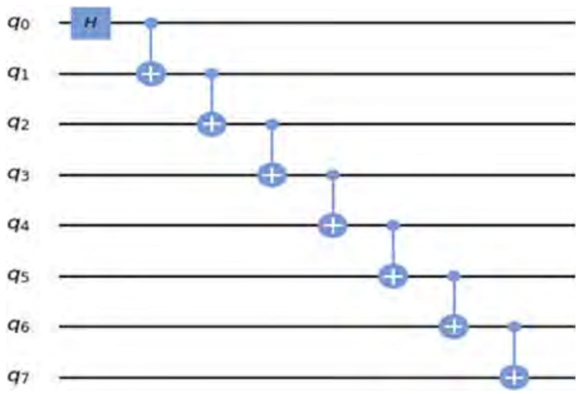


Fig. 2 Quantum circuit implementation of the permutated XOR bits for the secured transmission

4.3 Chaotic Diffusion

To ensure the security process of the data, the diffusion process is employed between the new encrypted key from the above process and the input data streams [17, 18]. Before the diffusion process, the scaling process has to be done to satisfy the input data transmission time, such that all the output matrix elements must be scaled to 0–255 using Eq. (5).

The new dynamic constant β and diffusion operatorμ have been introduced in the diffusion process and input date streams, respectively, Eqs. (5)and (6).

$$\beta = \text{mod } 256 \left\{ \sum D(i) \right\} \quad \text{where} \quad i = 0, 1, 2, 3, \ldots 25 \tag{5}$$

$$\mu = Fi + \beta + \text{mod } 256\{\, d(i)\} \quad \text{where} \quad 0, 1, 2, 3, \ldots M \tag{6}$$

Hence, the new cipher data has been obtained through the diffusion process between the dynamic keys and the data streams.

4.4 Complete Encryption Process

Step 1: Medical images as input data streams
Step 2: Divide the input data streams into the length eight bytes which are scaled in accordance with the application (Our Case is 256).
Step 3: Formation of quantum logistics maps
Step 4: Measurement of MAC parameters such as RSSI, distance and channel ID.
Step 5: Formation of 'S' vector which has been formulated by quantum chaotic maps which depends on network-centric initial conditions.
Step 6: Rearranging the S-vectors in ascending orders.
Step 7: Formation of a new complex key by XORing the obtained the 'S' vector along with the image datasets.
Step 8: Rearranging the data as data matrix $d(i)$ and G as F matrix which are scaled to 256.
Step 9: Diffusion process is performed.
Step 10: Encryption process is completed.

5 Experimental Setup

For evaluating the proposed encryption scheme, the MIAS mammogram datasets have been used as the input images which consist of 322 mammogram images which are stored as PNG format. All the proposed algorithms are simulated in QISKIT

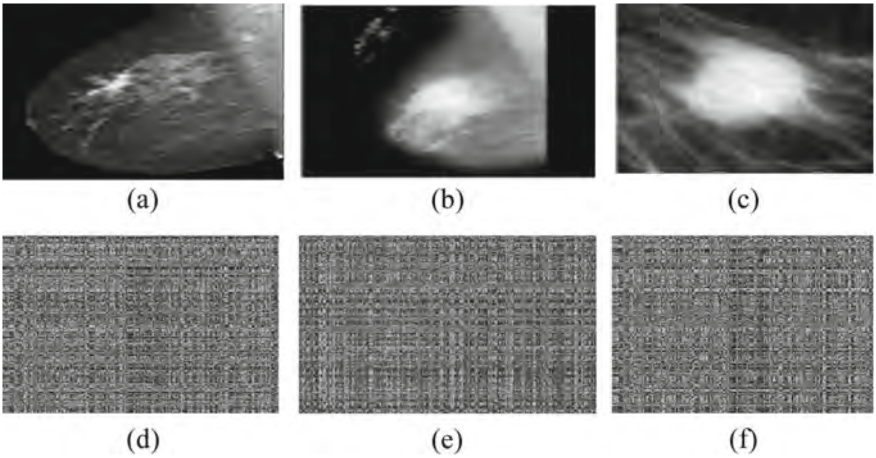


Fig. 3 **a** Normal Mammogram medical image, **b** Benign image, **c** Malignant image, **d** Encrypted normal image, **e** Encrypted benign image, **f** Encrypted malignant image

packages which run on Python 3.6 environment. The whole experimentation is implemented in computer which runs I7 CPU, 16 GB RAM, 2 TB HDD and 2 GB AMD Radeon GPU. Figure 3 shows different images used as the input for the proposed encryption scheme.

6 Various Analysis

6.1 Key Sensitivity Analysis

In this methodology, the secret key should be very sensitive even for minute changes resisting brute-force attack. To test the key sensitivity, the change in the number of bits has been observed with the change in the initial value of a chaotic system. The sensitivity of the key is calculated by the measurement of NPCR and UACI which measures the performance of this algorithm against the differential attack [19–21]. For a plain image, the mean value of NCPR and UACI can be calculated through an iterative procedure for nearly 20 times in a random manner. The mathematical expression for calculating the NPCR and UACI is given by the following equation. The analysis results are tabulated and discussed in Tables 1, 2 and 3.

$$\mathrm{NPCR} = \frac{\sum_{i,j} E(i, j)}{L} * 100 \tag{7}$$

Table 1 Illustration of NPCR and UACI for MAIS normal image datasets

S. No	No of bits change (%)	NPCR (%)	UACI (%)
1	10	99.4	89.90
2	20	99.3	85.57
3	30	99.2	86.90
4	40	99.1	85.56
5	50	99.0	85.89
6	60	99.1	86.90
7	70	99.2	89.90
8	80	99.1	89.56
9	90	99.2	88.90
10	100	99.1	88.34

Table 2 Illustration of NPCR and UACI for MAIS benign image datasets

S. No	No of bits change (%)	NPCR (%)	UACI (%)
1	10	99.34	88.80
2	20	99.25	85.47
3	30	99.2	86.00
4	40	99.10	85.06
5	50	99.10	84.79
6	60	99.23	83.00
7	70	99.28	87.00
8	80	99.19	88.00
9	90	99.10	84.90
10	100	99.45	83.34

Table 3 Illustration of NPCR and UACI for MAIS malignant image datasets

S. No	No of bits change (%)	NPCR (%)	UACI (%)
1	10	99.30	87.90
2	20	99.56	89.57
3	30	99.40	88.90
4	40	99.34	81.56
5	50	99.25	84.89
6	60	99.10	85.90
7	70	99.12	86.90
8	80	99.5	89.56
9	90	99.2	88.90
10	100	99.0	88.34

$$\text{UACI} = \frac{1}{L}\sum_{i,j}\frac{|f(i,j) \neq f(i,j)|}{256} * 100 \tag{8}$$

where

$$E(i,j) = \begin{bmatrix} 1, & f(i,j) \neq f(i,j) \\ 0, & f(i,j) = f(i,j) \end{bmatrix} \tag{9}$$

From the above tables, it is clear that the NPCR and UACI image datasets can resist different resistant attacks effectively. From the above tables, it is clear that the proposed chaotic encryption has good NPCR ranges from 99.56 to 99.0% even though the bit values are changed at different proportions. Entropy remains to be as low even for the different bit changes.

6.2 Statistical Performance Analysis

In this section, the statistical performance of the encryption algorithm has been observed to measure the degree of confidentiality of the image datasets. The histogram chart reflects the distribution of pixel values in the image. The pixels in closer result in better encryption. It illustrates the statistical figures in the image shown in Fig. 4.

Figure 4a represents the image before encryption which shows the uniformity of image used is very less in almost all parts of areas. Figure 4b represents the image after encryption which shows the uniform distribution of the images. The chi-square test is a method to measure the uniformity. The cipher images' distribution is calculated with critical thresholds with 25,15,10, 5 and 1% probability of changes whose values are given as 235.378, 244.29 and 249.44. Table 4 shows the chi-square value of different plaintext images and cipher images. From Table 4, it is clear that the

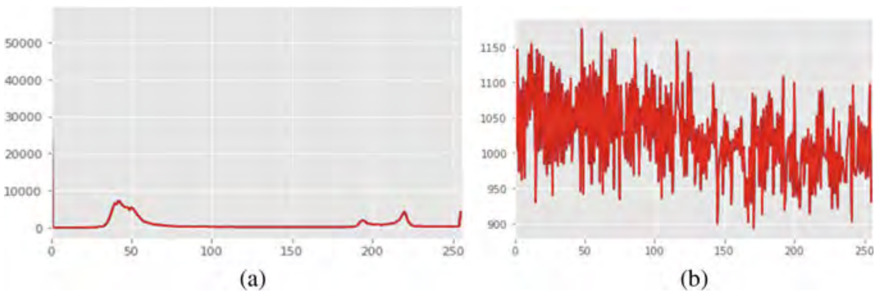


Fig. 4 Histogram analysis results **a** image before encryption, **b** image after encryption

Table 4 Chi-square (α) distribution for the different image datasets

Image_details	Plain image	Cipher-images	Critical values				
			$\alpha_{(0.1)}$	$\alpha_{(0.5)}$	$\alpha_{(0.01)}$	$\alpha_{(0.02)}$	$\alpha_{(0.05)}$
Normal medical images	1.3456×10^5	235.378	Pass	Pass	Pass	Pass	Pass
Benign images	3.456×10^5	233.89	Pass	Pass	Pass	Pass	Pass
Malignant images	3.567×10^5	244.89	Pass	Pass	Pass	Pass	Pass

encryption has passed the above thresholds which consequently results in resistance of the statistical brutal force attack.

6.3 Adjacent Pixel Point Correlation Analysis

The adjacent pixel point correlation of the image can be analyzed by the following mathematical expressions.

$$R_{xy} = \frac{\text{cov}(a, b)}{\sqrt{E(x)E(y)}} \tag{10}$$

$$\text{cov}(a, b) = D\{[a - D(a)][b - D(b)]\} \tag{11}$$

$$e(a) = \frac{1}{n}\sum_{i=1}^{n} a_i \tag{12}$$

$$L(x) = \frac{1}{n}\sum_{i=1}^{n} [a_i - a(x)]^2 \tag{13}$$

where $e(a)$ and $L(x)$ represent the expectations and variance of the plain text images and cipher image datasets. The correlation between the plain images and cipher images are given in Table 5.

Table 5 Representation of the correlation coefficient analysis between the plain image and cipher images

Image_details	Plain images			Cipher images		
	Horizontal	Vertical	Diagonal	Horizontal	Vertical	Diagonal
Normal image	99.567	99.456	99.900	0.2333	1.6779	2.9000
Benign images	90.788	90.444	90.23	0.03445	0.56778	0.7889
Malignant images	89.890	89.00	85.56	0.00900	1.456	3.9090

From the above Table 5, it is clear that correlations between different views of the plain images remain to be same in the most viewing angles. But to contrast, encrypted images do not correlate with the other parts of the images. The difference between the correlation coefficients of the different views of normal images is found to be 0.001, but in case of encrypted images, it is found to be 1.5. Also, the difference of coefficient of malignant and benign images is found to be 1.4,10, whereas coefficient of encrypted images is greater than 1.4. From the table, it clear that correlation is high as 1.5 which increases the complexity in the location of the encrypted images to change.

6.4 Information Entropy Analysis

Data entropy is the measure of uncertainty which reflects the highest degree of uncertainty of image information. The higher values of entropy prove higher randomness of cipher images. The mathematical expression for the entropy calculation is given by

$$g(m) = \sum_{l}^{l-1} q(m) \log_2 \frac{1}{q(m_i)} \tag{14}$$

where l represents the gray level. $q(m) \rightarrow$ the probability of gray value that appears in the image matrix. As per theoretical concepts, for a 8-bit gray image, it is good to have an entropy value greater than or equivalent to 9 that shows the rate of unpredictability. Table 6 illustrates the value of information entropy for a different image.

The test results of entropy are listed in the above Table 6. For efficient data encryption, the value of entropy should be close to 10, and from the above table, the value of entropy is found to be 9.90 for normal images, 9.789 for benign and 9.78 56 malignant images. Based on the results obtained, the encrypted images cannot be decoded by an attacker.

Table 6 Illustration of different information entropies for image sets used

Image_datasets	Entropy	Local entropy (block of images)
Normal image sets	9.900	9.678
Benign	9.789	9.402
Malignant images	9.7856	9.211

Table 7 Computational complexity analysis

Image datasets	Computational time complexity
Normal image datasets	0.06 s
Benign image datasets	
Malignant image sets	

6.5 Computational Complexity Analysis

The time consumption for the proposed algorithm depends on two processes, namely permutation and diffusion process. The total time consumption for different image sets is depicted in Table 7.

The time of computation is calculated by the Python timers used for experimentation, and it is found to be 0.06 s which is considered to be faster operation and proves to be utilized in medical image transmission.

7 Conclusion and Future Scope

A novel network parameter-based quantum encryption process QCNET has been suggested for secured medical image transmission. The proposed encryption algorithm utilizes the permutation process to shuffle the image pixels. Furthermore, logistic mapped-chaotic diffusion is also accomplished to protect the images. The complex control factors of the logistic map help to develop a strong keyspace to resist against brute-force attack. The quantum circuit is designed, and numerical simulation results show that the proposed scheme is used to secure the information and resist various attacks. Moreover, the computational difficulty is lesser than other traditional encryption schemes.

The proposed algorithm has been tested only for one attack in an IoT environment. Still, the algorithm needs intelligence for different categories of attacks. To increase the encryption efficiency for various attacks, hybrid chaotic encryptions for multiple images can be researched further.

References

1. Liu X, Xiao D, Xiang Y (2019) Quantum image encryption using intra and inter bit permutation based on logistic map. IEEE Access 7:6937–6946. https://doi.org/10.1109/ACCESS.2018.2889896
2. Xu J, Li P, Yang F, Yan H (2019) High intensity image encryption scheme based on quantum logistic chaotic map and complex hyperchaotic system. IEEE Access 7:167904–167918
3. Zhang J, Huo D (2019) Image encryption algorithm based on quantum chaotic map and DNA coding. Multimedia Tools Appl 78:15605–15621

4. Hua Z, Zhou B, Zhou Y (2018) Sine-transform-based chaotic system with FPGA implementation. IEEE Trans Ind Electron 65(3):2557–2566 (March)
5. Axenides M, Floratos E, Nicolis S (2018) The quantum cat map on the modular discretization of extremal black hole horizons. Eur Phys J C 78(5):412–427
6. Zhang X, Seo S-H, Wang C (2018) A lightweight encryption method for privacy protection in surveillance videos. IEEE Access 6:18074–18087
7. Lia C, Zhanga Y, Xie EY (2019) When an attacker meets a cipher-image in 2018: a year in review. arXiv:1903.11764v2 [cs.CR] 5 June 2019
8. He Y, Zhang Y-Q, Wang X-Y (2020) A new image encryption algorithm based on two-dimensional spatiotemporal chaotic system. Neural Comput Appl 32:247–260
9. Alawidaa M, Samsudina A, Teha JS, Alkhwaldehb RS (2019) A new hybrid digital chaotic system with applications in image encryption. Sig Process 160:45–58 (July)
10. Lai Q, Chen C, Zhao X-W, Kengne J, Volos C (2019) Constructing chaotic system with multiple coexisting attractors. IEEE Access 7:24051–24056 (March)
11. Singh RK, Kumar B, Shaw DK, Khan DA (2018) Level by level image compression encryption algorithm based on Quantum chaos map. J King Saud Univ Comput Inf Sci
12. Ye G (2018) A chaotic image encryption algorithm based on information entropy. Int J Bifurcat Chaos 28(1):1850010 (11 pages)
13. Wang X, Pham V-T, Jafari S, Volos C, Munoz-Pacheco JM, Tlelo-Cuautle E (2017) A new chaotic system with stable equilibrium: from theoretical model to circuit implementation. IEEE Access 5:8851–8858 (June)
14. Aqeelurrehman X, Kulsoom LA, Ullah S (2016) A modified (Dual) fusion technique for image encryption using SHA-256 hash and multiple chaotic maps. Multimedia Tools Appl 75(18):11241–11266
15. Bakhshandeh A, Eslami Z (2013) An authenticated image encryption scheme based on chaotic maps and memory cellular automata. Optics Lasers Eng 51(6):665–673
16. Akhshani A, Akhavan A, Lim S, Hassan Z (2012) An image encryption scheme based on quantum logistic map. Commun Nonlinear Sci Numer Simul 17(12):4653–4661 (December)
17. Belazi A, El-Latif AAA, Belghith S (2016) A novel image encryption scheme based on substitution-permutation network and chaos. Sig Process 128:155–170
18. Kulsoom A, Xiao D, Aqeelurrehman, Abbas SA (2016) An efficient and noise resistive selective image encryption scheme for gray images based on chaotic maps and DNA complementary rules. Multimedia Tools Appl 75(1):1–23
19. Ji X, Bai S, Guo Y, Guo H (2015) A new security solution to JPEG using hyper-chaotic system and modified zigzag scan coding. Commun Nonlinear Sci Numer Simul 22(1):321–333
20. Wang XY, Zhang Y-Q, Zhao Y-Y (2015) A novel image encryption scheme based on 2-D logistic map and DNA sequence operations. Nonlinear Dyn 82(3):1269–1280
21. Ahmed F, Anees A, Abbas V, Siyal M (2014) A noisy channel tolerant image encryption scheme. Wirel Pers Commun 77:2771–2791

Secure Voting for Democratic Elections: A Blockchain-Based Approach

Hardik Ruparel, Shraddha Hosatti, Mahesh Shirole, and Sunil Bhirud

Abstract Blockchain technology, due to its lucrative features like immutability, transparency, and security, is definitely at the forefront of an impending digital revolution. Numerous organizations have already started integrating blockchain with their applications to enhance security. Currently, blockchain is being used in a wide spectrum of industries ranging from the energy domain to the automobile industry to the financial sector. Blockchain technology can be used in any domain that requires transparency, flexibility, and immutability as a fundamental requirement. One such domain targeted in this paper is the voting domain. Building a secure electronic voting system that offers transparency, immutability and security is the challenge that has been faced for a long time. This paper proposes a novel Blockchain-as-a-Voting-Service (BaaVS) solution that solves all the issues that are currently existing in the electronic-based voting system by delivering a secured and transparent voting approach.

Keywords Blockchain · E-voting system · Democratic elections · Secure immutable transparent vote · Flexible vote recounting · Ethereum

H. Ruparel (✉) · S. Hosatti · M. Shirole · S. Bhirud
Computer Science Engineering and Information Technology Department, Veermata Jijabai Technological Institute, Mumbai 400019, India
e-mail: hardikruparel14@gmail.com

S. Hosatti
e-mail: shraddha.hosatti@gmail.com

M. Shirole
e-mail: mrshirole@it.vjti.ac.in

S. Bhirud
e-mail: sgbhirud@ce.vjti.ac.in

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_47

1 Introduction

Electronic votin g systems have been the subject of active research for decades, intending to minimize the cost of running an election, while ensuring the election integrity by fulfilling the security, privacy, and compliance requirements [6]. Although the evolution from pen paper-based voting to electronic voting systems has decreased the cost and enhanced the security manifold, there still exist some disadvantages like dependency on a centralized system, lack of transparency and no feature to ensure immutability, to name a few. Blockchain, with its inherent nature to provide transparency and immutability, perfectly fits as a solution to this problem. Blockchain technology has become the trending topic in the software world since the inception of Bitcoin [8] in 2008. A blockchain can be defined as a digital ledger of transactions that are duplicated and shared across the entire network of computer nodes. Conceptually speaking, it is a data structure that is built by a chain of blocks where each block contains a set of transactions. Along with the transaction data, cryptographic data like the hash of the previous block, timestamp of the transaction, Merkle root and nonce value is also stored in the blocks.

The process of adding a block to a blockchain is called *mining*. The mining of the blocks is typically a computationally expensive task that requires miner nodes to solve a cryptographic puzzle. The node that solves this puzzle first will propose the block to all the nodes in the network for verification. If the majority of the nodes verifies it positively and reach on a consensus, then the block will be added to the blockchain. It is the responsibility of the consensus algorithm to bring the blockchain network on to a decision.

Voting information needs to be unaltered and verifiable. Information stored in the blockchain is immutable and verifiable, hence blockchain is a good option for the voting system. Once a block is added to a blockchain, it cannot be amended retrospectively without the revision of all the subsequent blocks. This requires a consensus with a majority of the nodes in the blockchain network which is practically impossible [3], thereby making the blockchain immutable. Apart from being immutable, blockchain also provides the feature of easy verifiability. Since the blockchain ledger is distributed and duplicated across the nodes in the network, the nodes can easily verify the new block against their version of the ledger. These features are in part achieved through advanced cryptography, providing a security level greater than any previously known record-keeping system. Blockchain technology is therefore considered by many [1], including us, to have substantial potential as a tool for implementing a new modern voting process.

Although electronic elections are preferred for fast and secure operations, it is under threat from malicious actors that can infiltrate voting machines, alter voter registration databases, and more. A novel Blockchain-as-a-Voting-Service (BaaVS) architecture is proposed to leverage the benefits of the blockchain system in the voting system, thereby fabricating a new end to end secured voting system for democratic elections. The proposed system is fast, tamper-free, transparent and auditable. It is developed with the following objectives:

1. To store all stakeholders information in transparent and secure way
2. To enable secure anonymous voting with no double voting problem
3. To provide distributed, transparent, and re-countable voting Blockchain-as-a-Voting-Service (BaaVS) architecture.

The paper is organized as follows: Sect. 2 describes the related work in this domain. In Sect. 3 provides details about the proposed system. Section 4 presents the implementation details of the proposed system. In Sect. 5, the results of the implementation of the proposed system is presented. Finally, Sect. 6 concludes the paper.

2 Background and Motivation

Migrating the voting process online has been a challenging task for a long time. Many systems and architectures have been proposed to contest the voting process online. Each of these systems provides a different degree of privacy protection, transparency and risk mitigation. Various cryptographic techniques have been integrated to further fortify security. However, all the solution involves the presence of either a centralized system or a dependency of a third-party entity which makes the entire process vulnerable. Blockchain technology can be integrated into multiple areas such as cryptocurrencies, financial industry, games, anti-counterfeiting, supply chain management, etc. In blockchain-based secure voting system, a voter can utilize vote-token to cast his/her vote. There are myriad tokens [11], which are either fungible or non-fungible tokens are used to represent different types of assets in different industries. Some of the blockchain-based solutions that provide decentralized voting are as follows.

Open Vote Network [10] is a decentralized self-tallying voting protocol that is designed for a small-scale boardroom meeting. Unlike other voting protocols, Open Vote Network protocol does not rely on a third party for tallying. Each voter is in control of the privacy of their own vote. This voting protocol is a two-round protocol in which first the voters register their intent to vote in the elections post which they cast their vote in the second round. In [7], the open vote network protocol is implemented by developing smart contracts and deploying them on the Ethereum blockchain.

BitCongress [4] is another voting ecosystem that is created in conjugation with Bitcoin, Counterparty and Smart Contract blockchains using a distributed model to verify elections, votes and voters on separated blockchain networks. The smart contract blockchain is used to create elections as smart contracts that contain election protocols like election time, candidates, legislation and custom election rules. The counterparty blockchain is used to create addresses for the election smart contracts along with the addresses of all the actors of the system. It also keeps a tally of every vote cast in the election. Finally, there is the Bitcoin blockchain which is used for mining and storing the hash of the votes cast in the election. Tallying of the votes is done using the Borda count [4] mechanism.

In [3], a new blockchain-based voting system is proposed which provides security, privacy control measures and the ability to audit the votes. This system consists of three main modules viz. Identity Management, Cryptographic Privacy and Aggregation and Auditing. The identity management module is responsible for providing access control by requiring the voters to register first. The voter's registration details are stored on to a private blockchain which is then broadcasted to the election authorities. The cryptographic privacy module helps maintain the privacy of the voters by using the blind signature mechanism. The aggregation and auditing module comes into the picture after the election is over. All the valid votes are tallied, and the results are displayed. The voters will then be provided with the inverse function of the signature to verify and audit the votes.

A blockchain-based voting system has been proposed in [5] which allows the voter to change the vote during the election-time window even after the vote has been cast. The solution consists of a centralized system containing the list of all the qualified people who can vote. This system then distributes a security token to each of the qualified people which is then used as a digital signature. A vote that is cast is first encrypted using this digital signature and then is stored on to the blockchain.

Privacy Preserving Voting Protocol on Blockchain [13] consists of two main components viz. clients and smart contracts. The client component consists of the voting operations that will be performed by the voter, While the smart contract contains the voting logic and protocols. Each vote is first encrypted using hashing algorithms post which the validation process starts. In case of the vote has been proved invalid, the devoting steps are triggered. All the valid votes are then tallied and aggregated in the end.

After reviewing the related works in this domain, a new online voting system is proposed using blockchain to target the shortcomings of the current e-voting systems and provides benefits like increased security, transparency and ease of verifying the results.

3 Proposed Methodology

Our proposed system aims at bridging the shortcomings of the current online voting systems by providing a new Blockchain-as-a-Voting-Service (BaaVS) solution. Blockchain will ensure that once a transaction has been made, which is a vote, in this case, cannot be tampered with. Proposed Blockchain-as-a-Voting-Service (BaaVS) System's block diagram is shown in Fig. 1. The proposed voting system consists of six actors: voters, candidates, political party and Electoral Registration Officer (ERO), Polling Officer (PO) and Election Commission (EC).

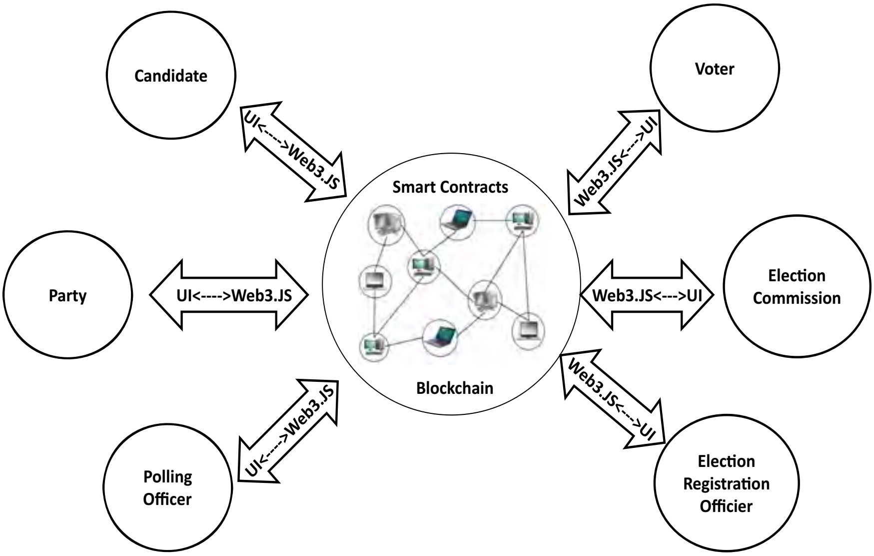


Fig. 1 Proposed blockchain-as-a-voting-service (BaaVS) system—Block diagram

3.1 Roles and Responsibilities

Voters, candidates and political parties have to register themselves by filling up the required information which is then validated by the contract deployed for registration. The ERO will then verify the details of the voters, the candidates and the parties. After a successful verification, the details entered by the actors are made permanent on blockchain for further usage. A smart contract is deployed on blockchain to ensure that a voter can vote only once thereby eradicating the double counting problem. The smart contract will also verify that the candidate has met all the conditions necessary to contest in the elections. The PO's responsibility is to tally the votes using proper cryptographic mechanisms.

Voter—A person is classified to be a voter only if he/she is the citizen of the nation and is above the voting age. The role of the voter is to register him/herself by entering the necessary details on the blockchain. After the successful verification, a voter will receive his/her voting credentials. The voter will get a vote-token for the respective upcoming election through which the voter will cast his/her vote that will then be stored in the blockchain forever. The voter can cast his/her vote through his/her through registered mobile election wallet application or he/she can approach the polling booths nearby during elections.

Party—A party is an organization that announces its candidates before every election process. The party must first register itself on the blockchain. Party must also announce their representative candidate before the election process starts. Party can also dismiss any of its candidates if the candidate is not verified.

Electoral Registration Officer—The responsibility of an ERO is to verify the details of the voters, candidates and also the parties. After a successful verification, the ERO will provide these actors with their cryptographic signatures which will be used by the system for signing the votes cast by the voters.

Polling Officer The PO is responsible to verify voters on voting booths in case the voter prefers to vote at polling booths. The PO plays a crucial role in the election process during voting and at the time of vote counting. During the voting phase, the system signs the votes cast by the voter by the public key. During the counting phase, he unlocks the transparent secure vote for counting by providing his private key.

Candidate—A candidate contests the election. A candidate is eligible for standing in the election only if he/she is a verified voter. Candidate first needs to register him/herself on the blockchain, after that he/she needs to submit a small security deposit for contesting the election. Candidate must belong to a valid political party. Based on circumstances he/she may withdraw candidacy at any point in time before the election as per the schedule by the Election Commission (EC).

Election Commission— Election Commission (EC) is responsible for declaring the elections, formulating the rules and schedule of the registration, voting and vote counting phases of the election.

Once the users register themselves onto the blockchain, they will be verified and then the voter can be allowed to vote in the election process. As the entire voting ecosystem is hosted on blockchain, the overall election process can be viewed by all the actors of the system. To implement the proposed voting ecosystem, smart contracts are developed using Solidity.

3.2 Overall Working of the Proposed Voting System

The proposed Blockchain-as-a-Voting-Service (BaaVS) ecosystem consists of three main phases namely the Pre-election phase, the In-election phase, and the Post-election phase. Fig. 2 displays the overall working of the proposed blockchain solution.

Pre-election Phase The Pre-election phase further consists of four sub-phases namely Voter Registration, Validation by ERO, Candidate Nomination and Candidate Withdrawal. Each of these sub-phases is discussed in detail below:

- Voter registration is the first sub-phase of the pre-election phase. All newly qualified voters will first have to register themselves on to the blockchain ledger. For registering, a voter must enter its Voter ID along with details like name, address, phone number, date of birth, etc. To enhance the security of the system, the voters are required to register their biometrics.
- Verification of voter by ERO, the second sub-phase, consists of the verification of the voter's details and biometrics by the ERO. Only after a successful verification, the ERO will access the secured Government servers to provide the voter with its

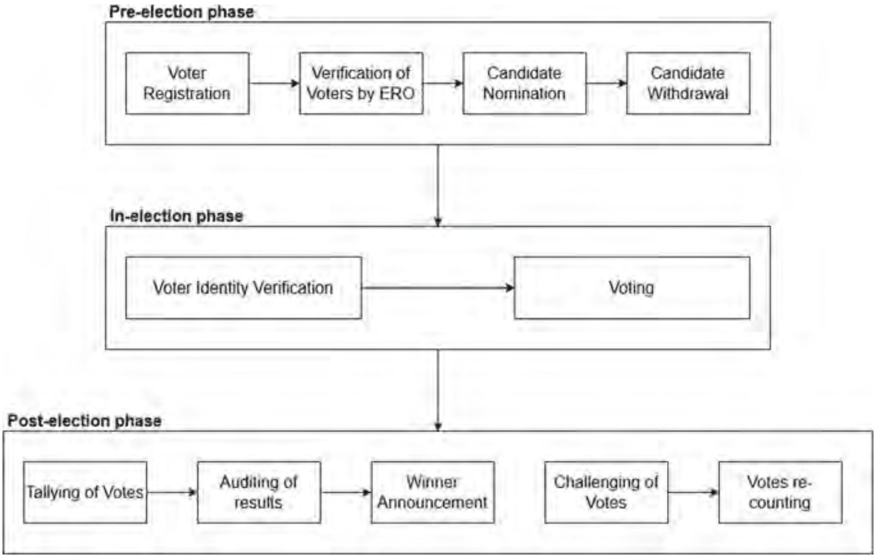


Fig. 2 Proposed system—Working

Private Key and its corresponding Public Key. Only the ERO will have access to the government's secured servers. The proposed system uses the private key of the voter to automatically sign the vote cast by the voter.
- Candidate nomination is the third sub-phase in the pre-election period during which all the political parties release a list of the candidates contesting the elections. Before nominating the candidates, the party must first register itself onto the blockchain. The candidates are also required to register themselves by entering the same details as that of the voter. Along with the above information, the candidate must also include the party he/she belongs to and a security deposit that must be submitted before contesting for any election. The candidate's deposit amount depends on the type of the elections he/she is contesting in.
- Candidate withdrawal is the final but an optional step in the pre-election phase. The party, at any point in time, can withdraw the candidacy of any candidate. This will lead to a refund of the security deposit collected during the candidate's nomination sub-phase.

In-election Phase After the Pre-election phase, all the voters, candidates and parties are registered and verified by the ERO. The in-election phase consists of Voter Identity Verification and Voting as two main sub-phases. The detailed functioning of each sub-phase is mentioned below:

- In the voter identity verification sub-phase, the ERO re-verifies the details entered by the voter during the Pre-election phase. After a successful verification, the ERO will verify if the voter belongs to the constituency he/she is intending to vote. For that, the ERO will verify the address mentioned on the voter's Aadhar Card. If the permanent address and the current address are same and the address belongs to the constituency the voter wants to cast vote in, then the ERO will show a green flag, thereby allowing the voter to vote. If the permanent address and the current address are not the same, but any one of these addresses belongs to the constituency, then too the ERO will allow the voter to proceed to the next sub-phase i.e. the voting sub-phase.

– Voting sub-phase lies at the core of the proposed Blockchain-as-a-Voting-Service (BaaVS). In this phase, the voter can cast his/her vote to any of its preferred candidates. Instead of capturing the votes using the traditional pen-paper approach or the electronic voting machine approach, the proposed system leverages blockchain to capture the votes. Before adding the vote in blockchain, the smart contract will first validate if this private key is used in any previous votes. If the system finds any other vote cast by the same private key, it will reject the vote. This validation allows the system to record only one vote from a particular voter, thereby solving the famous double-spending problem in the blockchain. To further fortify the security, the proposed system also signs the vote using the public key of the PO and at least three candidates contesting the elections as shown in Fig. 3. This Multi-signature authentication measure provides additional security and requires all the keys to tally the votes. Since the proposed voting system uses blockchain to store the votes, it allows voters to cast their vote online using any secured web-based cryptographic wallet. This is hugely beneficial for voters who cannot go to a voting booth to cast their votes. The proposed system uses voter's Aadhar number as a salt while calculating the cryptographic hash. This further strengthens the security by preventing the frequency analysis of the cryptographic hash.

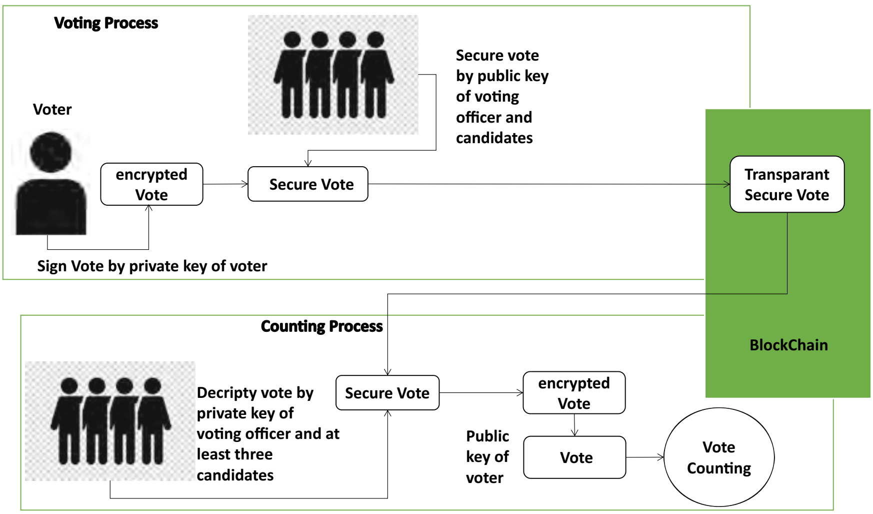


Fig. 3 A systematic view of secure voting and vote counting process

Post-election Phase This is the third and final phase of the proposed voting system. In a normal centralised voting ecosystem, the votes that are cast are tallied by a trusted third-party entity. This centralised dependency has a huge risk of the results getting manipulated since the votes cast by the voters are untraceable as the votes cast by the voters cannot be validated by the voters. To remove the third-party dependency for the vote tallying and to make the votes traceable and re-countable as many times as required, the benefits of blockchain are leveraged to solve these problems. A blockchain-based vote tallying mitigates the risk associated with the final vote count tampering. This phase consists of four sub-phases which are discussed in details below:

- Tally of Votes is the first sub-phase that the system performs after the election period. For the tallying process, the public key of the voter and the private keys of the PO and that of the same candidates are used to unlock the vote and tally it. This will allow removing the central dependency on just the PO, thereby providing a decentralized approach in tallying the votes.
- Auditing of the results provides the voters with an ability to themselves verify the correctness of their votes. Since, blockchain inherently provides features like immutability and transparency, the auditing of the votes can be easily be done by the voters on our proposed system.
- Winner announcement is the final sub-phase of the proposed system, in which the authorities announce the results.
- Challenging of votes is an optional sub-phase in the post-election period. This phase is only activated if anyone wishes to challenge the result and demands the recounting of the votes.
- Vote re-counting is an additional task performed by the BaaVS system to re-verify and validate the final results. Since all the votes are stored on the blockchain, the tallying of the votes can be done easily, cheaply and as many times as required which is a huge advantage over the current voting systems.

The proposed system, thus, solves the prevailing issues like security, tampering of the votes and transparency by using blockchain, which inherently provides all these benefits, to store the votes. The proposed system also provides additional security features like preventing double voting, multi-signature authentication and preventing the frequency analysis of the hashes which are extremely important to maintain the integrity of the elections.

4 Implementation Details

For implementing our proposed Blockchain-as-a-Voting-Service (BaaVS), smart contracts are developed. Smart contracts are developed using Solidity, a statically typed object-oriented programming language for developing smart contracts. These smart contracts are deployed on the ethereum blockchain [12] which are then executed by the miner nodes of the Ethereum network. During the in-election phase when the voter is casting its vote, along with the *candidateId* to which the voter wants to cast its vote to, the system ensures security by automatically signing the vote by the voter's private key, PO's public key and the public key of any two of the candidates. To avoid the frequency analyses of the hashes, the system also adds salt to the vote that needs to be hashed. In our case, the Aadhar Number of the voter are used as a salt. After digitally signing the vote and calculating the cryptographic hash, the hash is now ready to be mined on to the blockchain. The algorithm for the above mentioned steps of the voting sub-phase is given in Algorithm 1.

Since the system uses private/public keys of various actors to store the vote on the blockchain, during the vote tallying phase, the system will automatically use the corresponding public/private keys of the same actors to determine the *candidateId* to which the vote belongs to. Since cryptographic hashing is a one-way function, there is no way to find the *candidateId* to which the vote was cast just by using the cryptographic keys. The systems will loop through the entire list of *candidateId*

Algorithm 1 Casting Votes

1: **function** VOTE(candidateId, voterPrivateKey, POPublicKey, candidate1PublicKey, candidate2PublicKey, salt)
 Input: *candidateId* - ID of the candidate the voter has chosen to vote
 voterPrivateKey- Voter's Private Key
 POPublicKey - Public key of the PO
 candidate1PublicKey- Public Key of any Candidate
 candidate2PublicKey - Public Key of any other Candidate
 salt- Salt for the Hash function (Aadhar number of the voter in our case)
 Output: Cryptographic hash of the vote
2: $hashOfVote \leftarrow$ Hash(*candidateId, voterPrivateKey, POPublicKey, candidate1PublicKey, candidate2PublicKey, salt*)
3: Add *hashOfVote* to the blockchain
4: **end function**

Algorithm 2 Algorithm for Tallying Votes

1: **function** TALLYING_VOTES(voterPublicKey, POPrivateKey, candidate1PrivateKey, candidate2PrivateKey, salt, hashOfVote)
 Input: *voterPublicKey* - Voter's Public Key
 POPrivateKey - Private Key of the PO
 candidate1PrivateKey - Private key of any candidate
 candidate2PrivateKey - Private Key of any other Candidate
 salt - Salt for the Hash function (Aadhar number of the voter in our case)
 hashOfVote - Cryptographic hash of the vote retrieved from blockchain
 Output: Candidate ID for which the vote was cast
2: **for** $candidateId$ in $candidateList$ **do**
3: hash $\leftarrow$ Hash(*candidateId, voterPublicKey, POPrivateKey, candidate1PrivateKey, candidate2PrivateKey, salt*)
4: **if** $(hash = hashOfVote)$ **then**
5: **return** $candidateId$
6: **end if**
7: **end for**
8: **end function**

and perform cryptographic hashing functions on it. The count of the *candidateId* is increased by one if the hash matches the hash that is present on the blockchain. Algorithm 2 describes the procedure for tallying of votes that returns the *candidateId* for which the voter has cast the vote.

5 Results

For implementing the proposed system, five smart contracts- *Registration*, *Voting*, *Counting*, *Election_Configuration* and *Candidate_Nomination* are deployed on the Ethereum blockchain. As a result of the implementation, the performance of the system is measured by using a set of eight parameters which are described in Table 1.

5.1 Comparison with Existing Solution

The performance of the proposed system is compared with the performance of the current non-blockchain based solutions like Electronic Voting Machine (EVM), Electronic Polling Agent [3] and the proposed voting system in [2] against the parameters mentioned in Table 1. Due to blockchain's inherent nature of being immutable and transparent, the proposed system also inherits these features. Since the system is immutable, any vote that was cast cannot be changed at a later point in time, thereby maintaining the integrity of the elections. Hence, the proposed system, Electronic Polling Agent [3] as well as the proposed system in [2] provides integrity and transparency owing to the use of blockchain. Non-blockchain solutions like EVM cannot assure integrity and transparency features which are crucial for the election process.

The proposed system allows the user to verify whether his/her vote is being tallied correctly or not. The Electronic Polling [3] system also allows the voters to verify their votes by using the inverse function of the signatures. The blockchain-based

Table 1 Parameters for determining the performance of the system

Parameters	Definition
Integrity	Votes should not be modified after being cast by the voter
Transparency	Voters should have a general view of the entire process
Verifiability	Voters must be able to verify that the votes are correctly tallied
Cost-effectiveness	System should be affordable and efficient
Votes recounting	System's ability to recount the votes as many times as required
Scalability	System must be able to handle large number of transaction
Enhanced security	System's ability to provide advanced security features
Secrecy	No one should be able to determine how any individual has voted

voting system in [2] provides the voters with an ability to verify his/her votes by using cryptographic mechanisms. However, in non-blockchain solutions, there is no way for the voters to verify that their votes are correctly tallied or not. Leveraging blockchain has not only led to a significant decrease in the cost but also improved the transparency in the voting and the tallying phase. Hence all the blockchain-based solutions are cost-effective while the non-blockchain solutions are not. Votes recounting is a fundamental feature included in the proposed system which leverages the smart contract to allow efficient recounting of the votes as many times as necessary. The smart contract loops through the entire blockchain for getting the cryptographic hashes of the votes and uses Algorithm 2 to determine the *candidateId* for which the vote belongs to. Moreover, in the Electronic Polling Agent [3] and the blockchain-based system in [2], one cryptographic hash of the vote is stored in one block. This affects the scalability of the system as the miner nodes will only be able to process and mine one vote at a time. However, the proposed system allows several hundred cryptographic hashes of the vote to be processed concurrently by the miner nodes, thereby increasing the scalability of the system.

Additionally, the proposed solution provides advanced security features like preventing frequency analysis of the hash and multi-signature authentication. In the absence of salt in the hash function, a malicious user can look at the cryptographic hash of the votes and can identify the trend of the ongoing elections. The proposed system mitigates this risk by mandating the use of the salt while creating the cryptographic hash of the vote. Using the salt also helps maintain the true secrecy of the election as no one will be able to determine how the voter has voted. Thus, the proposed voting system provides advanced security features along with maintaining true secrecy by using additional cryptographic measures which are not present in Electronic Polling Agent [3] and the voting system in [2]. Non-blockchain solutions provide secrecy owing to the absence of transparency in it. Table 2 provides the comparative analysis and the benefits of the proposed system over some of the existing blockchain and non-blockchain based solutions.

However, due to blockchain's limited scalability, additional scalability measures need to be taken to allow the processing of millions of votes. For this, it is required to

Table 2 Comparative analysis with related systems [2, 3]

Parameters	Non-blockchain solutions	[2]	[3]	Proposed system
Integrity	×	✓	✓	✓
Transparency	×	✓	✓	✓
Verifiability	×	✓	✓	✓
Cost-effectiveness	×	✓	✓	✓
Votes recounting	✓	×	×	✓
Scalability	×	×	×	✓
Enhanced security	×	×	×	✓
Secrecy	✓	×	×	✓

implement advanced scalability measures like GeoSharding along with our proposed system. GeoSharding [9] is a sharding protocol that divides the blockchain network in shards and then elects a leader in each shard for parallel processing of the transaction which will help increase the throughput of the blockchain system thereby increasing the capacity of our system to enable millions of voters to vote concurrently.

6 Conclusion

This paper introduced a blockchain-based, secured electronic voting system that utilizes smart contracts to enable secure and cost-efficient election while guaranteeing voter's privacy. Due to blockchain's inherent feature of immutability, the data related to each election conducted using the proposed online voting system is safe, immutable, irreversible, and stored permanently in the blockchain ledger. The proposed solution provides additional advantages like multi-signature authentication, prevention of hash frequency analysis and double voting that is beneficial to all the stakeholders of the system. Leveraging blockchain's security features has offered a variety of new possibilities to overcome the limitations and adoption barriers of electronic voting systems, thereby ensuring election security and integrity and laying the ground for transparency. The proposed system allows the voters to validate their respective votes and candidates to view the tally process themselves. Moreover, using an Ethereum private blockchain has made it possible to send hundreds of transactions per second onto the blockchain, utilizing every aspect of the smart contract to ease the load on the blockchain. For countries of greater size like India, the election process spans out for multiple days for different geographical regions. One day, one election for huge democracies can be possible only if some advanced scalability measures are implemented along with our proposed system.

References

1. Weaver N Secure the vote today. https://www.lawfareblog.com/secure-vote-today
2. Ayed AB (2017) A conceptual secure blockchain-based electronic voting system. Int J Netw Secur Appl 9(3):01–09
3. Babu A, Dhore VD (2020) Electronic polling agent using blockchain: a new approach. In: IC-BCT 2019. Springer, pp 69–77
4. BitCongress: Bitcongress—Blockchain based voting system. http://cryptochainuni.com/wp-content/uploads/BitCongress-Whitepaper.pdf
5. Hardwick FS, Gioulis A, Akram RN, Markantonakis K (2018) E-voting with blockchain: an e-voting protocol with decentralisation and voter privacy. In: 2018 IEEE International conference on Internet of Things (IoT) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE smart data (SmartData). IEEE, pp 1561–1567
6. Hjálmarsson F, Hreiarsson GK, Hamdaqa M, Hjálmtýsson G (2018) Blockchain-based e-voting system. In: 2018 IEEE 11th international conference on cloud computing (CLOUD). IEEE, pp 983–986

7. McCorry P, Shahandashti SF, Hao F (2017) A smart contract for boardroom voting with maximum voter privacy. In: International conference on financial cryptography and data security, Springer, pp 357–375
8. Nakamoto S (2008) Bitcoin: a peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf
9. Ruparel H, Chiplunkar S, Shah S, Goradia M, Shirole M (2020) Geosharding—A machine learning-based sharding protocol. In: IC-BCT 2019. Springer, pp 105–118
10. Seifelnasr M, Galal HS, Youssef AM (2020) Scalable open-vote network on ethereum. IACR Cryptol 2020:33
11. Shirole M, Darisi M, Bhirud S (2020) Cryptocurrency token: an overview. In: IC-BCT 2019. Springer, Singapore, Singapore, pp 133–140
12. Wood G et al (2014) Ethereum: A secure decentralised generalised transaction ledger. In: Ethereum project yellow paper, vol 151, issue 2014, pp 1–32
13. Zhang W, Yuan Y, Hu Y, Huang S, Cao S, Chopra A, Huang S (2018) A privacy-preserving voting protocol on blockchain. In: 2018 IEEE11th international conference on cloud computing (CLOUD). IEEE, pp 401–408

Convergence Analysis of Self-Adaptive Equalizers Using Evolutionary Programming (EP) and Least Mean Square (LMS)

N. Shwetha and Manoj Priyatham

Abstract Digital communication has become an important part of our lives, and technology has been undergoing advancements. With the arrival of the age of digitalization and digital signal, communication has got implemented in a vibrant range of applications but still, they are strongly affected by two basic problems, namely *Noise and Inter-Symbol Interference* (*ISI*). This is caused by the error-creating phenomena which are characteristics between the transmitter and receiver which include the scattering of the transmitted signal. The noise produced in the communication channel is caused by channel characteristics and can be reduced with proper channel selection. The SNR can be improved by improving the transmitter signal strength even in spite of noisy signal at the receiver. By using the adaptive equalization in channels will reduce this effect drastically and can be implemented by using various adaptive algorithms. Hence, an adaptive channel equalizer is used to inverse the effect channel had on the signal to get back the initial information. There are many adaptive algorithms to update the coefficients of equalizers; evolutionary algorithms are used in this paper to do so. The two algorithms used before are Artificial Bee Colony algorithm (ABC) and Ant Colony Optimization (ACO). The latest algorithm is the combination of Evolutionary Programming and LMS algorithm (EPLMS); this gives better solution faster. A comparative study between the algorithms is done in this paper.

Keywords Intersymbol interference (ISI) · Least mean square (LMS) · Artificial bee colony algorithm (ABC) · Ant colony optimization (ACO) · Evolutionary programming (EP) · Maximum likelihood sequence estimation (MLSE)

N. Shwetha (✉)
Department of ECE, Dr. Ambedkar Institute of Technology, Bangalore, Karnataka 560056, India
e-mail: shwethaec48@gmail.com

M. Priyatham
Department of ECE, APS College of Engineering, Bangalore, Karnataka 560082, India
e-mail: manojopriyatham2k4@yahoo.co.in

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_48

1 Introduction

Modern digital communication systems enforce the usage of channel equalization with a high tracking rate and short preparation time. Such constraints focus our interest on adaptive algorithms that are used to unite quickly. The most fundamental advantages of the digital system for video, information, and voice associations are their superior consistency in noise environment conversely with that of their analogy components; unfortunately, the most frequently digital communication of data is complemented with a trend commonly referred to as Inter-Symbol Interference (ISI) [1–3]. Momentarily, this implies that the transferred pulses are covered out so that the pulses that parallel to various signs are not distinguishable. Dependent on the broadcasting media, the most important reasons for ISI are wired communication; in reality, they are band restricted and has multichannel dissemination.

It is important to diminish the impacts of ISI for a dependable digital broadcasting system wherein the adaptive equalization comes into picture. Two of the more intensively developing fields of digital communication, specifically cellular communications and digital subscriber lines, are heavily reliant on the implementation of reliable channel equalizers. The LMS algorithm as seen is one of the extremely prevalent algorithms in adaptive signal handling. It was the emphasis of a lot of research and its realization in several applications because of its robustness and simplicity. One of the potential solutions is the enactment of the equalizer by a filter with finite impulse response (FIR) utilizing the perfectly established LMS algorithm for adapting its coefficients. The recognition stems from its comparatively low computer-based intricacy, good mathematical strength, simple configuration, and simplicity of implementation in accordance with the conditions of the devices. The principle of the LMS algorithm is to modernize the coefficients of adaptive filter recursively together with the adverse descent of the assessment of error surface. The traditional algorithm utilizes a static step size to accomplish the repetition and to find a deal among the divergence of small stable-state MSE and quick convergence. A little step size could make sure little MSE with a sluggish convergence, while a huge step size will offer improved tracking abilities and a faster convergence at the expense of greater stable-state MSE. Consequently, in flexible step-size LMS algorithm, it is impossible to resolve this ambiguity. Therefore, several variable step-size algorithms have been recommended to resolve the trouble [4–7]. Although such algorithms could speed up the convergence and determine stable-state MSE to a certain extent, they have been unsuccessful to examine the optimization of variable step-size LMS additionally. The following equations have been utilized to describe the LMS algorithm.

$$e(k) = d(k) - WT(k) * X(k) \tag{1}$$

$$X(k+1) = X(k) + W(k)\mu e(k) \tag{2}$$

where the dimension of $X(k)$ is the extension of the AF, $e(k)$ is deviation error, $d(k)$ is the anticipated output value, $X(k)$ is coefficient vector of the adaptive filter, $W(k)$ is the *i*/*p* vector at random sampling time n, and μ is learning step.

2 Concept of Inter-Symbol Interference

In the digital communication system, if everything is right at the receiver side then there will be no interaction among successive symbols. Here each of the signals which are arrived is decoded self-reliantly among others. But when it comes to symbol interaction, one of the waveforms will corrupt the values of the next nearby symbols. Due to this, the received signal will be distorted. Because of this, it is difficult to differentiate messages from such a received signal. The shortage is identified as the ISI. The purpose of an equalizer is to reduce the ISI so that a reconstructed signal having from the transmitter side. Due to this, it also reduces the bit rate of the transmitted signal. As assumption made in all pass AWGN is impractical, the lack of frequency spectrum the signal is filtered to minimize the bandwidth so that frequency structured division can be obtained [8, 9]. There are many bandpass channels available in practical but the response varies to the different frequency components. To avoid this, the simplest AWGN model is needed to have for representing the practical channels very accurately. Such commonly available retirement is a dispersive channel model shown in Fig. 1.

$$y(t) = x(t) * h_c(t) + n(t) \tag{3}$$

From the equation, $u(t)$ is the *Tx* signal, $h_c(t)$ is the impulse response of the channel & $n(t)$ is AWGN power spectral density. The dispersive representative of the channel is prototyped by using the linear filter $h_c(t)$. This dispersive channel model is a low-pass filter. By using this low-pass filter as can line the transmitted signal to time causing the effect of symbol difficult to adjust symbols in a practical case while transmitting the signals from the transmitter. Due to this, the ISI will deteriorate the error caused by the transmitted signal to error performance in the communication

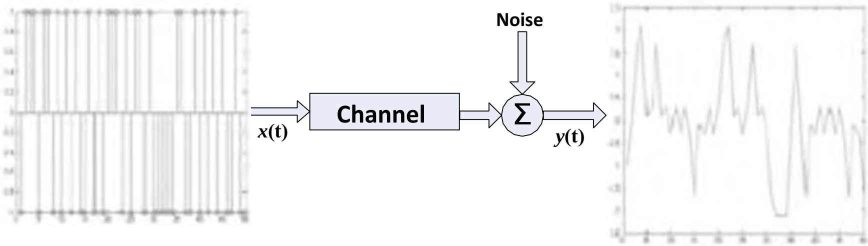


Fig. 1 Inter-Symbol Interference

system. Two main methods are mainly concentrated on which eradicates the ISI deterioration effect. In the first method, the band restricted transmission pulses are used to diminish the ISI. The pulses obtained by the ISI are called free pulsed which are known by its name Nyquist pulses [10–13]. As seen in the second method, the received signal is needed to screen to stop the ISI which was presented by the channel impulse response. This is known as equalization.

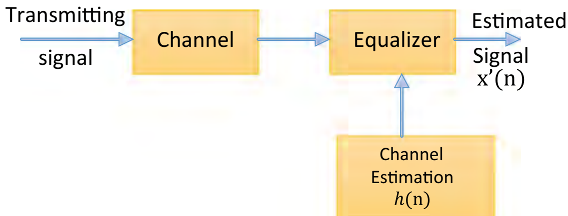
3 Process of Equalization

The equalization process is something like adjusting the balance between frequency components. The receiver signal matching frequency at the receiver is adjusted to decrease the noise and any interference generated during transmission. Hence, post-compensation of ISI is done at the receiver side by equalization. An equalizer is a device which does equalization. There are many applications of equalizers in electronics and communication. Here equalizing the equalizer to the channel will be the main concentration.

3.1 Channel Equalization

Channel equalization is one method of adjusting the equalizer coefficients to channel coefficients to reduce ISI. If a channel is considered as a filter, then equalizer is an inverse filter. An equalizer is not only going to compensate the effect of the channel but also going to compensate all the unnecessary effects transmitted signal went through, i.e., due to pulse shaping, transmitter filter, and receiver filter to get back the initial signal. The block diagram of channel equalization is shown in Fig. 2. When the channel is known, then by sending a known signal through the channel error signal can be calculated [14, 15]. Always the received signals differentiated to desire signal to receiver signal to identify the errors in the signal. The error signal is the driving force for the equalizer. Equalizer will aim to minimize the error signal. Hence, optimization techniques/algorithms are used to achieve this. There are many algorithms used for equalization. The most effective algorithm used before the adaptive algorithm is Maximum Likelihood Sequence Estimation (MLSE), where depending on

Fig. 2 Channel equalization



the MLSE algorithm the channel response for impulse is measured. The equalizer coefficients are adjusted or equalized to nullify the effect by channel. Adjustment of the coefficients is done to reduce the ISI and noise at the output. Hence, from the distorted version of the transmitted signal, the original version can be reconstructed by equalizer [16, 17]. Once the equalizer weights are set, then it won't change and then required information can be sent through the channel.

3.2 Adaptive Channel Equalizer

If the channel is time-variant, then the weights/coefficients used in the equalizer should also be updated as per the varying nature of the channel. Adaptive equalizer works in two modes: exercise mode (Training Mode) and choice directed mode (Decision Mode).

(a) Training mode: In this training, the signal is used, i.e., it is known to both transmitter and receiver. When equalizer gives an output, this is compared with the training signal and difference is the error signal. This error signal is used to update the equalizer weights/coefficients. When the error signal is zero, i.e., the output from the equalizer is equal to the training signal. Hence, the coefficients are saturated. Then, the equalizer changes its mode.

(b) Decision directed mode: Now the actual data signal can be transmitted from the transmitter. Again the equalizer changes its mode to training mode.

Again the weights will be updated and then in decision-directed mode, the next set of actual data will be sent. Therefore, the equalizer will be switching between these modes. The weights are varying with each new training mode. This forms a kind of feedback loop. This makes the communication system reliable. There are many adaptive algorithms which tell how to update the equalizer coefficients.

Figure 3 demonstrates the standard template for a direct channel with an adaptive equalizer. From the figure, it is observed that inverse filtering or channel equalizer comprises assessing a transmission functionality to counteract for the linear deformation triggered by the channel. From a different perspective, the goal is to compel a specified dynamic performance for the cascading of the adaptive filter and the channel

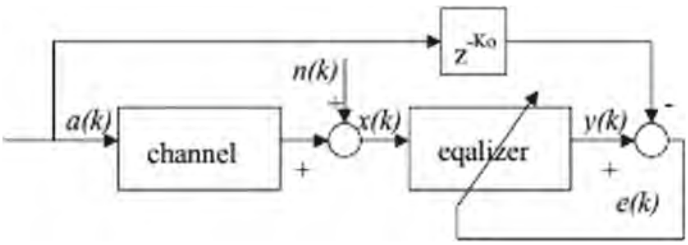


Fig. 3 Adaptive channel equalizer

(an unidentified system), decided by the incoming signal. The initial translation and interpreting are additionally relevant in communications, in which the data is sent via dispersion channels [18–20]. The next is suitable for management applications, in which the inverted filtering program creates control signals that will be utilized in the undetermined system.

Let us describe the vector of weighting coefficients $X(n) = [X0(n)\ X1(n)\ X2(n)\ldots XM - 1(n)]T$ and the signal vector at equalizer input $W(n) = [w(n)\ w(n - 1)\ w(n - 2)\ldots w(n - M + 1)]T$ of the adaptive filter at a moment n, [1–8]. Furthermore, at the equalizer the input signal samples are in the type of:

$$W(n) = W\, j\, h(j)a(n - j) + k(n) \tag{4}$$

where $h(j)$ signifies the response of channel impulse, $k(n)$ signifies the additive noise with the variance $\sigma 2k$, and $a(n)$ signifies the nth data sample.

The data specimens hold on values of ± 1 ($a(n) = \pm 1$) only, and the noise is supposed to be autonomous. At the nth iteration instant, and the equalizer output is given as:

$$y(n) = XT(n)W(n) \tag{5}$$

The outcome $y(n)$ is being utilized in assessing the data transferred symbol $a(n - no)$, with N_0 signifying the interruption. The error sampling of kth output is:

$$e(n) = y(n) - a(n - N_0). \tag{6}$$

The coefficients of weighting in the LMS algorithm are derived from the subsequent statement:

$$X(n + 1) = X(n) + \mu e(n)W(n), \tag{7}$$

where μ is the algorithm step size. The output mean square error (MSE) is:

$$e(n) = E(e2(n)) = XT(n)RX(n) + E(a2(n)) \tag{8}$$

$$2XT(n)E\big(W^{-}(n)a(n - no)\big),$$

with $R = E(W^{-}(k)X^{-}T(n))$.

The regular output MSE after nth iteration may be stated as:

$$\varepsilon\mathrm{avr}(n) = \varepsilon\mathrm{MIN} + E\big(VT(n)RV^{-}(n)\big), \tag{9}$$

where $V(n) = X(n) - X{*}(n)$ is the error vector of the weighting coefficient and εMIN is the minimal MSE (5), for optimum vector for weighting coefficients is $X{*}(n)$, i.e., Wiener vector. The MSE beyond εMIN is referred to as the surplus MSE in the stable state. As demonstrated in [1, 2], for the LMS algorithm the surplus MSE is provided by:

$$\varepsilon\mathrm{e} \;= 12\mu\sigma 2\ \mathrm{ntr(R)}. \tag{10}$$

It can be noticed from (7) that the excess MSE is proportionate to the step size because of the gradient noise. The step size should be chosen to strike an equilibrium among the contradictory objectives and targets of the small stable-state error and fast convergence (large step size), i.e., little surplus MSE (small step size).

4 Problem Formulation

Realizing the period size is enormous, merging rate of the LMS algorithm will be depraved, yet the consistent state MSE, i.e., the mean square error resolves itself. Then, again will be more, if and only if the step size is small, the consistent state MSE will be small, yet the merging rate will be moderate. In this way, the step size gives a trade-off among the merging rate and the consistent state MSE of the LMS algorithm. The other way to increase the efficiency of the LMS algorithm is to make the step size adjustable as opposed to fixed which leads to VSSLMS algorithms. By using this methodology, both the fast merging rate and a little consistent state MSE can be achieved. The step size should satisfy the condition:

0 < step-size < 1/(max Eigenvalue of the input auto-correction matrix).

For fast convergence, step size is set close to its maximum allowed value.

5 Formulation of LMS Algorithm

The LMS, i.e., least mean squares procedure, is one of the utmost famous procedures in adaptive handling of the signal. Because of its robustness and minimalism was the focal point of many examinations, prompting its execution in numerous applications. LMS algorithm is a linear adaptive filtering algorithm that fundamentally comprises of two filtering procedure, which includes calculation of a transverse filter delivered by a lot of tap inputs and creating an estimation error by contrasting this output with an ideal reaction. The subsequent advance is an adaptive procedure, which includes the programmed modifications of the tap loads of the channel as per the estimation error. The LMS algorithm is additionally utilized for refreshing channel coefficients. The benefits of the LMS algorithm are minimal calculations on the sophisticated nature, wonderful statistical reliability, straightforward structure, and simplicity of usage regarding equipment. LMS algorithm is experiencing problems regarding step size to defeat that EP, i.e., evolutionary programming is utilized (Fig. 4).

Basic steps in the LMS algorithm:

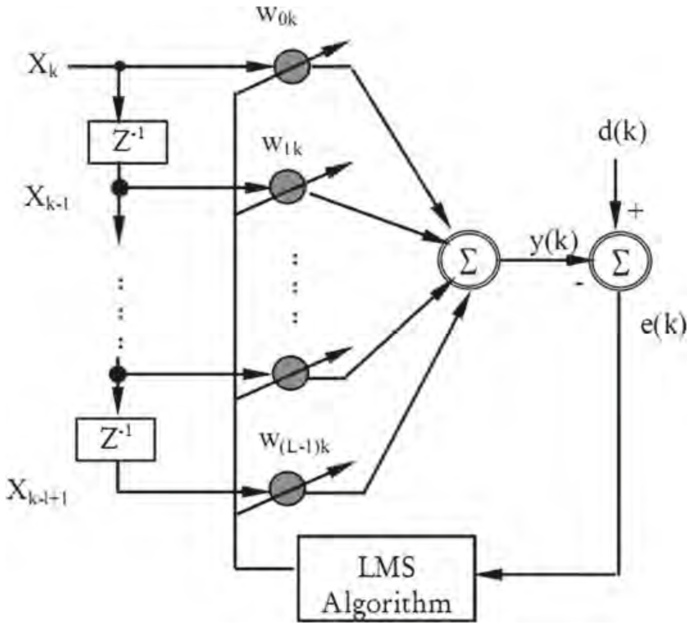


Fig. 4 Adaptive filter using LMS algorithm

1. Fundamentally at the beginning, random coefficients are taken.
2. Error is defined to the present sample using the equation

$$e(n) = d(n) - x^{\mathrm{T}}(n) * W(n)$$

where

$e(n)$ deviation error
$d(n)$ expected output value
$x(n)$ *i/p* vector at sampling time $nW(n)$ coefficient vector

3. Adjustment has defined using the LMS.
4. To get the new coefficient value.

$$W(n+1) = W(n) + \mu e(n)x(n)$$

5. New coefficients will replace the previous coefficients and the process will continue from step 2.

6 Evolutionary Programming

Evolutionary algorithms are stochastic search methods and not the deterministic ones. In 1960, Lawrence J. Fogel utilized the evolutionary programming in the USA to utilize modeled evolution as an educational procedure which is seeking to create AI. The previously existing methods like linear programming, calculus-based methods, for example, nutenian method are having difficulties in delivering the global solution. They are tending to stuck in the local solution. To overcome this problem, nature inspiration computation can be applied. In this approach, some characteristics that are available in nature are taken as a reference to develop the mathematical model. This mathematical model will utilize to find out the solution to the problem. In this paper, the characteristics of nature are taken as evolution [21, 22]. This is one of the most successful characteristics available in nature where the things evolved (the things changed) with the time to adapt the environment in result betterment in fitness value hence, the chances of survival are high, for example, the transformation from a monkey to a human. The mathematical model based on evolution is referred to as evolutionary computation (Fig. 5).

6.1 Basic Steps in Evolutionary Programming

1. Based upon natural evolution, a mathematical model called evolutionary computation has created.
2. In nature, the things change from one time to others to increase its fitness so that chances of survival could be better, for example, human evolution.
3. At the beginning, a random population is defined as a set of solutions.
4. Depends upon the objective a mathematical function called—objective function is defined.
5. With the certain mathematical operator called—mutation offspring's are generated.
6. A mathematical operator called selection operator is applied to define the next-generation members.

The simple and efficient way to achieve the better next generation is

a. Combine parent and offspring population.
b. Get their fitness.
c. Pick up the half members having higher fitness(tournament selection process is used).

7. The next generation will replace the previous generation to produce offspring.
8. This process will keep continuing until some terminating criteria do not satisfy.
9. After termination from the last generation, the member having the highest fitness will consider as a final solution.

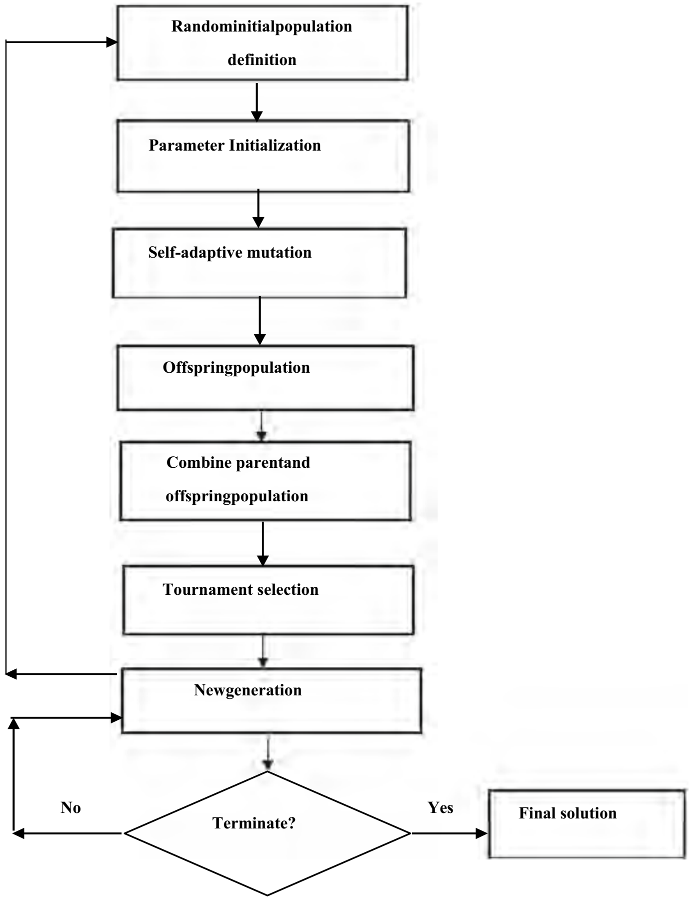


Fig. 5 Flowchart for evolutionary programming

6.2 Tournament Selection

6.2.1 Procedural Steps

1. Take the combined population of parents and offspring.
2. Pick up one member randomly and select ‗n‘no of opponents randomly.
3. Compare fitness.
4. Calculate the score.
5. Sort the score in the ascending order.
6. Select the right half.
7. Then pick up the members corresponding to that score.
8. These members will form the next generation.

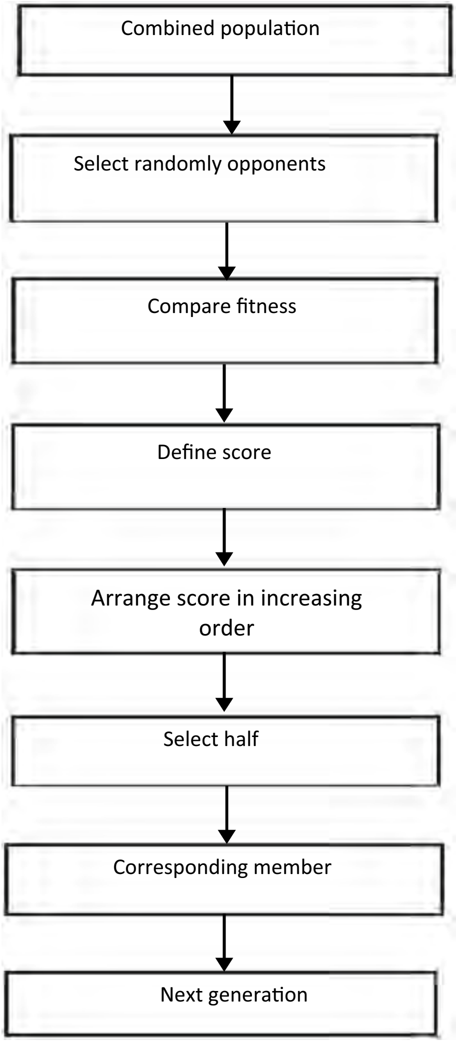


Fig. 6 Flowchart for tournament selection

9. The next generation should replace the previous generation to produce offspring (Fig. 6).

7 Working of EP-Based LMS Algorithms

1. At the beginning of the random step, size is defined as the population.
2. With respect to each step, size applies the LMS and get its corresponding error value (fitness).

3. A step size having the minimum error select it w.r.t current sample point.
4. With the selected step size LMS applied to get the coefficient value.
5. As the new input sample appears, from the previous generation a new population of step size is created in EP and procedure repeated.

8 Simulation Results

MATLAB 2014b was utilized to implement the modelling and subsequent results have been shown in this section. During modeling, a variety of internal factors of the ACO and ABC algorithm was selected by contemplating the articles [23]. For the channel, the number of plugs chosen for the equalizer is 11 taken over to find out the efficiency of EVSSLMS. The incoming signal includes the 500 examples engendered at random through regular dissemination as demonstrated in Fig. 8. Gaussian noise contains 0.01 standard deviation and zero mean which is combined with the incoming signal as demonstrated in Fig. 2, channel features are provided by the vector:

$$[0.06 - 0.0740.099 - 0.237 - 0.360.11580.360\, 0.2370.0490.099]$$

The ability of the recommended structures was determined in accordance with the conditions of its convergence nature as described in Fig. 9.

This is the randomly generated input signal consists of 500 samples. This signal transfer in a bipolar form (+1, −1). To make the system more complex, random information generated between +1 and −1. This makes the information unpredictable at the receiver side.

Figure 7 illustrates the input signal, and Fig. 8 illustrates engendered incoming signal and signal with noise from the channel.

In Fig. 9, comparative convergence faunas of optimized equalizers EPLMS along with equalizers prompted by ABC and ACO algorithms have been implemented for the various input signal. From the figure, it is clear that the proposed EPLMS performs more iterations with reduced time. Additionally, it was observed that among optimized AEs, EPLMS equalizer is the fastest one.

Figure 10 shows that the EPLMS AEs disclose well BER acts in comparison to the AEs with ABC and ACO specifically the greater SNR area. The comprehensive outcomes have been outlined in Table 1.

From Fig. 11 and Table 2, it has been noticed that the EPLMS equalizer effectively retains its superiority over ACO and ABC triggered AEs in improved communication systems. For illustration, EPLMS equalizers express the BER values of 43.125e−02 for a signal-to-noise ratio value of 8 dB, whereas ACO and ABC generated AEs to offer BER values of 5.058 e−02 and 5.442 e−03 for similar SNR value. More obviously, it was distinguished that the order of improvement in BER behavior of the EPLMS-based EP improved communication system is viewed to be 101 over ABC and ACO based improved communication systems; however, the

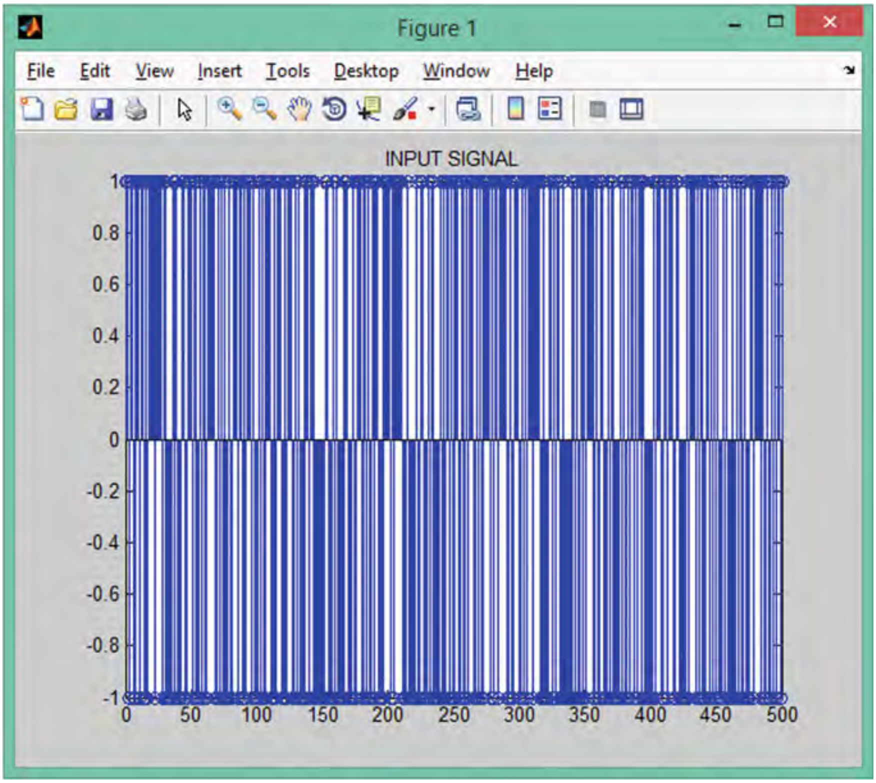


Fig. 7 Input signal for 500 samples

proportion increases in BER value utilizing EPLMS-based improved communication system is experimental to be 23.15%. In connection with this 20 dB SNR level was deemed. Analogous improvements have also been experimental for lower SNR segments. Henceforth, it can remain agreed that EPLMS optimized AEs surpass the AEs synchronized by ACO and ABC methods. However, EPLMS-optimized AE provides the optimal result among all the combinations.

9 Conclusion

Bandwidth-effective data transfer through radio and telephone channels has been made possible through the usage of adaptive equalization to counteract for the dispersal of time launched by the channel. Stimulated by useful applications, a constant research attempt over the past two decades has been producing a wealthy body of fiction in adaptive equalization and the associated more common disciplines of the function of system identification, adaptive filtering, and digital signals. This

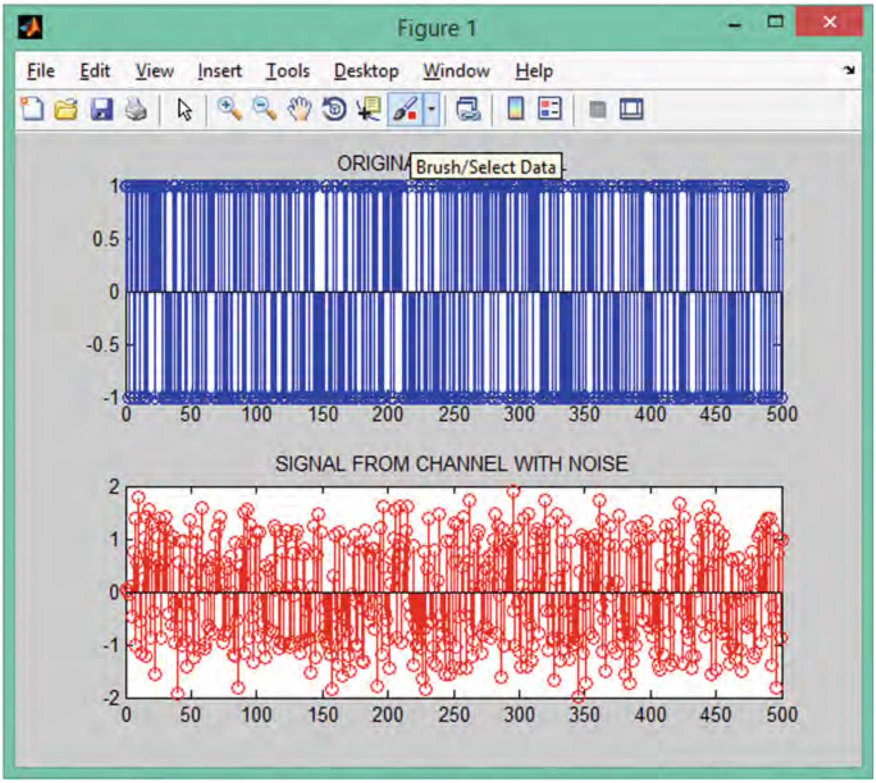


Fig. 8 Engendered incoming signal and signal along with the noise from the channel

article provides a summary of the adaptive equalization. In our design, since evolutionary programming is being used, it will decide what would be the value of step size for a particular application so that mean square error is minimized and convergence is optimal. And also, faster convergence is obtained. Consequently, the effectiveness of an interaction system can be enhanced. A comprehensive review offers that the EPLMS triggered adaptive equalizers to propose quicker merging act in comparison with ABC and ACO structures. Additionally, optimized equalizers are propelled by EPLMS techniques also demonstrate their pre-eminence in accordance with the conditions of BER performance over the ABC and ACO algorithms. Moreover, it has also been observed that EPLMS is better for ABC and ACO optimized AEs. With the layout of the obtaining filters, the impact of Inter-Symbol Interference can be reduced.

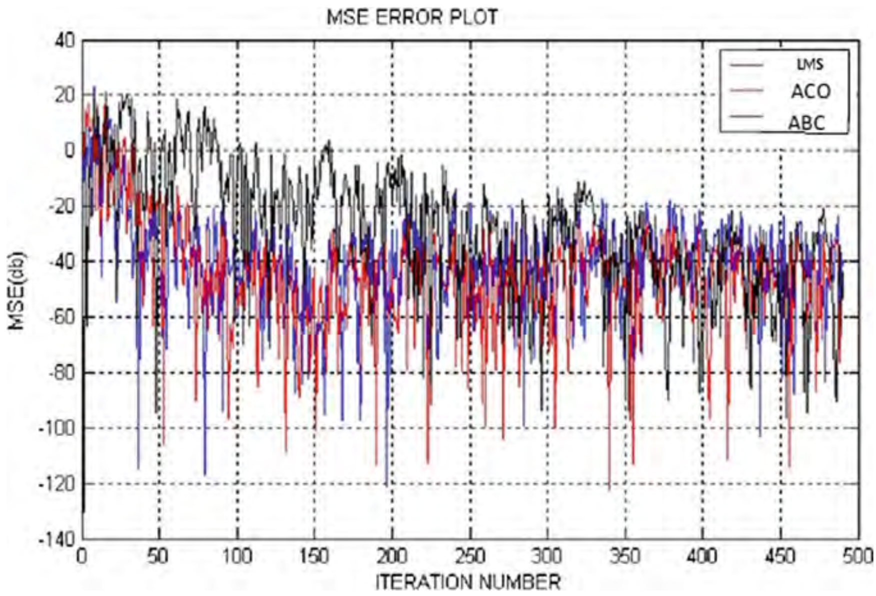


Fig. 9 Comparison of convergence act of LMS with ABC & ACO equalizers

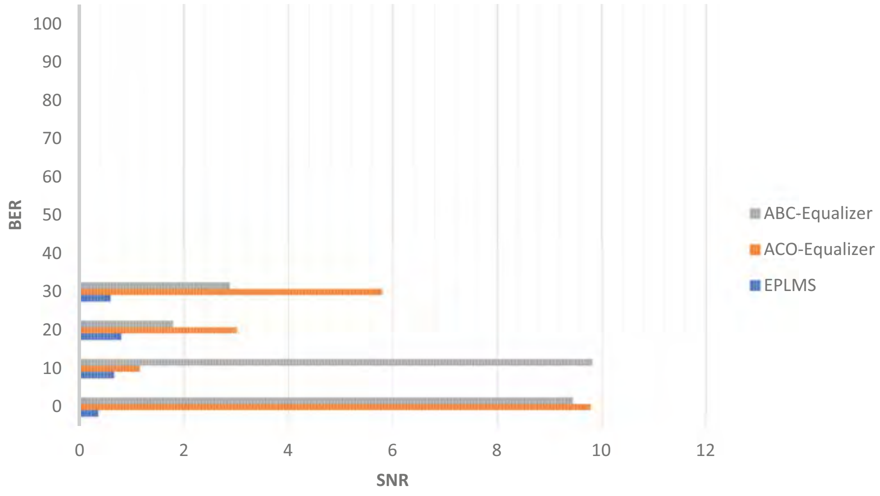


Fig. 10 BER versus SNR analysis of EPLMS optimized equalizer with ABC and ACO equalizers for various input signals

Table 1 BER values input signal

SNR → BER↓	4	8	16	20
EPLMS Equalizer	0.360e−01	0.658e−02	0.799e−04	0.589e−05
ACO-Equalizer	9.784e−02	1.154e−02	3.000e−05	5.789e−06
ABC−Equalizer	9.439e−02	9.820e−03	1.790e−05	2.873e−056

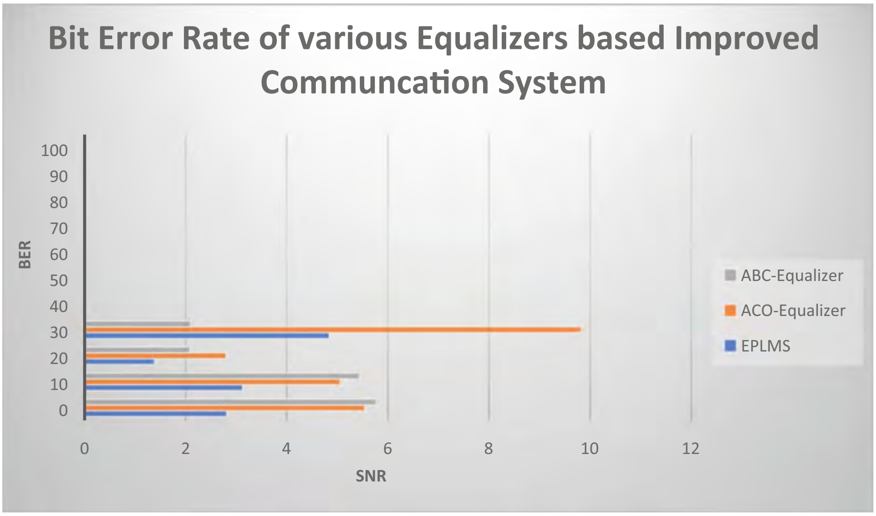


Fig. 11 BER versus SNR analysis acts of the different equalizer-based improved communication system

Table 2 BER values for various optimization algorithms

SNR → BER ↓	4	8	16	20
EPLMS Equalizer	2.811e−03	3.125e−02	1.380e−04	4.837e−08
ACO-Equalizer	5.541e−02	5.058e−02	2.797e−05	9.820e−06
ABC-Equalizer	5.768e−02	5.442e−03	2.085e-05	2.095e−056

References

1. Widrow B, Glover JR, Mc JM. Kaunitz CJ, Williams CS, Hean RH, Zeidler JR, Dong E, Goodlin RC (1975) Adaptive noise cancelling: principles and applications. Proc IEEE 63(12):1692–1716 (December)
2. Harris RW, Chabries DM, Bishop FA (1986) A variable step (VS) adaptive filter algorithm. IEEE Trans Acoust Speech Sign Process ASSP-34(2):309–316 (April)
3. Sethares WA, Mareels IMY, Anderson BDO, Johnson R, Bitmead RR (1988) Excitation conditions for signed regressor least mean squares adaptation. IEEE Trans Circuits Syst

35(6):613–624 (June)
4. Li X, Fan Y, Peng K (2009) A variable step-size LMS adaptive filtering algorithm. In: Proceedings of 5th international conference on wireless communications, networking and mobile computing, IEEE, Sept 2009
5. Wang Y, Bao M (2010) A variable step-size LMS algorithm of harmonic current detection based on fuzzy inference. In: Proceedings of 2nd international conference on computer and automation engineering (ICCAE), IEEE, vol 2, pp 665–668, Apr 2010
6. Xiao Y, Huang B, Wei H (2013) Adaptive Fourier analysis using a variable step-size LMS algorithm. In: Proceedings of 9th international conference on information, communications & signal processing, IEEE, pp 1–5, Dec 2013
7. Schniter P, Johnson CR (1999) Dithered signed-error CMA: robust, computationally efficient blind adaptive equalization. IEEE Trans Sign Process 47(6):1592–1603 (June)
8. Sun L, Bi G, Zhang L (2005) Blind adaptive multiuser detection based on linearly constrained DSE-CMA. IEE Proc Commun 152(5):737–742 (Oct)
9. Al-Awami T, Saif W, Zerguine A, Zidouri A, Cheded L (2007) An adaptive equalizer based on particle swarm optimization techniques. In: Proceedings of 9th international symposium on signal processing and its applications, IEEE, pp 1–4, Feb 2007
10. Boughelala A, Luan X, Leghmizi S (2011) An adaptive channel equalizer using bacterial foraging oriented by particle swarm optimization strategy. In: Proceedings of 3rd international conference on computer research and development, IEEE, vol 2, pp 24–29, Mar 2011
11. Mohammed JR (2012) A study on the suitability of genetic algorithm for adaptive channel equalization. Int J Electr Comput Eng (IJECE) 2(3):285–292 (June)
12. Dey A, Banerjee S, Chattopadhyay S (2016) Design of improved adaptive equalizers using intelligent computational techniques: extension to WiMAX system. In: Proceedings of IEEE Uttar Pradesh section international conference on electrical, computer and electronics engineering (UPCON), pp 305–310, Dec 2016
13. Karaboga D, Basturk B (2008) On the performance of artificial bee colony (ABC) algorithm. Appl Soft Comput 8:687–697
14. Dorigo M, Blum C (2005) Ant colony optimization theory: a survey. Theor Comput Sci 344:243–278
15. Eesa S, Brifcani AMA, Orman Z (2014) A new tool for global optimization problems-cuttlefish algorithm. Int J Math Comput Phys Electr Comput Eng 8(9)
16. Ketonen J, Juntti M, Cavallaro JR (2010) Performance complexity comparison of receivers for a LTE MIMO-OFDM system. IEEE Trans Sign Process 58(6):3360–3372 (June)
17. Paulo SRD (2008) Adaptive filtering: algorithms and practical implementations. Kluwer Academic Publisher, Springer Science and Business Media, LLC. 91
18. Proakis JG (2001) Digital communications. Mc Graw Hill, New York
19. Haykins S (2001) Communication systems. Wiley India Edition
20. Lathi BP (1995) Modern digital and analog communications system
21. Suneel Varma D, Aditya R, Subhashini KR (2013) Synthesis of adaptive antenna with circular geometry employing harmony search and differential evolution techniques. In: 2013 International conference on communication and signal processing, Melmaruvathur, 2013, pp 933– 937
22. Kennedy J (2006) Swarm intelligence. In: Handbook of nature-inspired and innovative computing. Springer, New York, pp 187–219
23. Rao RV, Savani VJ, Vakharia DP (2011) Teaching-learning-based optimization: a novel method for constrained mechanical design optimization problems. Comput Aided Des 43(3):303–315

Modified E-Shaped Resonator-Based Microstrip Dual-Mode Bandpass Filter

Shobha I. Hugar, Vaishali Mungurwadi, and J. S. Baligar

Abstract An modified E-shaped dual-mode resonator is proposed to design two types of dual-mode bandpass filters such as **Filter I** and **Filter II**. The proposed resonator comprises of a half-wavelength hairpin resonator and a meander ring on its symmetrical plane. As the structure is symmetrical, it supports odd and even mode theory. In **Filter I**, based on coupling gaps between feedlines and middle arm of E-shaped resonator, the location of upper stopband transmission zero is controlled. Using two tuning stubs incorporated with inner meander ring, passband center frequency and bandwidth are controlled in **Filter II**. Additional transmission zero is procured in upper stopband using source load coupling scheme to improve the selectivity of the filter. Both filters are designed for passband center frequency 2.5 GHz and fractional bandwidth 0.4.

Keywords Dual-mode · Fractional bandwidth · Transmission zeros · Quality factor · Radio frequency (RF)

1 Introduction

Dual-mode microstrip bandpass filters have been used in front end wireless communication system due to their significant features such as small size, good passband performance, high-quality factor, and low loss. The major limitations of single-mode filters such as narrow bandwidth and low-performance characteristics have made the researchers develop dual-mode filters for radio frequency (RF) and microwave applications.

S. I. Hugar (✉) · J. S. Baligar
Dr. Ambedkar Institute of Technology, Bangalore, Karnataka, India
e-mail: Shobha_hugar@yahoo.co.in

J. S. Baligar
e-mail: jbaligar@gmail.com

V. Mungurwadi
Visvesvaraya Technological University, Bangalore, India
e-mail: Vaishalibm18@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_49

E-shaped resonator was initially proposed in [1], where it was designed by two quarter-wavelength resonators with a quarter-wavelength open stub at center acting as K inverter to control coupling strength and to create transmission zeros at the desired frequency. Doublet and extended doublet coupling schemes were proposed in [2] to design two-pole and three poles bandpass filters, respectively, using E-shaped resonator. But in these filters, the selectivity was improved at the cost of larger filters size. To realize additional transmission zero in upper stopband [3], a capacitor C introduced with source load coupling and location of transmission zero was controlled by adjusting the value of C.

Stepped impedance E-shaped hairpin resonator loaded with T-shaped open stub was proposed to design dual-mode dual-band BPF [4] with wide stopband. Dual-mode ring [5] embedded with quarter-wavelength split open-end resonator was proposed to form extended doublet coupling which has improved the selectivity of the filter by creating one pair of transmission zeros near passband. Center frequency and bandwidth tuning were proposed in [6] by changing the finger width of inter-digital loading element. It is observed that filters reported in [3, 5] have a narrow upper rejection band.

An modified E-shaped dual-mode resonator is presented to design two types of dual-mode bandpass filters **Filter I** and **Filter II**. The proposed resonator comprises of a half-wavelength hairpin resonator and a meander ring on its symmetrical plane. As the structure is symmetrical, it supports odd and even mode theory. The gap between feedlines and mid-arm of E-shaped resonator controls location of transmission zero in upper stopband in **Filter I**. Using two tuning stubs, the passband frequency and bandwidth are controlled in **Filter II**.

2 Modified E-Shaped Dual-Mode Resonator

Figure 1a shows a layout of proposed modified E-shaped dual-mode resonator BPF with source load coupling scheme on a substrate with relative permittivity 10.2 and thickness 0.635 mm. The resonator comprises $\boldsymbol{\lambda_g/2}$ a hairpin resonator with 50 Ω impedance and a meander ring on its symmetrical plane. $\boldsymbol{\lambda_g}$ is the guided wavelength at center frequency 2.5 GHz. The dimensions of the proposed resonator are L_1 = 7.18 mm, L_2 = 12.77 mm, L_3 = 1 mm, L_4 = 1.8 mm, W = 0.59 mm, W_1 = 0.4 mm, W_2 = 0.2 mm. Figure 1b represents a source load coupling scheme where the resonator 1 represents odd mode frequency and resonator 2 represent even mode frequency. The input source is coupled to both modes represented by dark lines and so is the output. The dashed line represents the weak coupling between input and output. The coupling matrix for a given source load coupling scheme can be written as [3] (Fig. 2).

The proposed modified E-shaped dual-mode resonator has the following interesting properties.

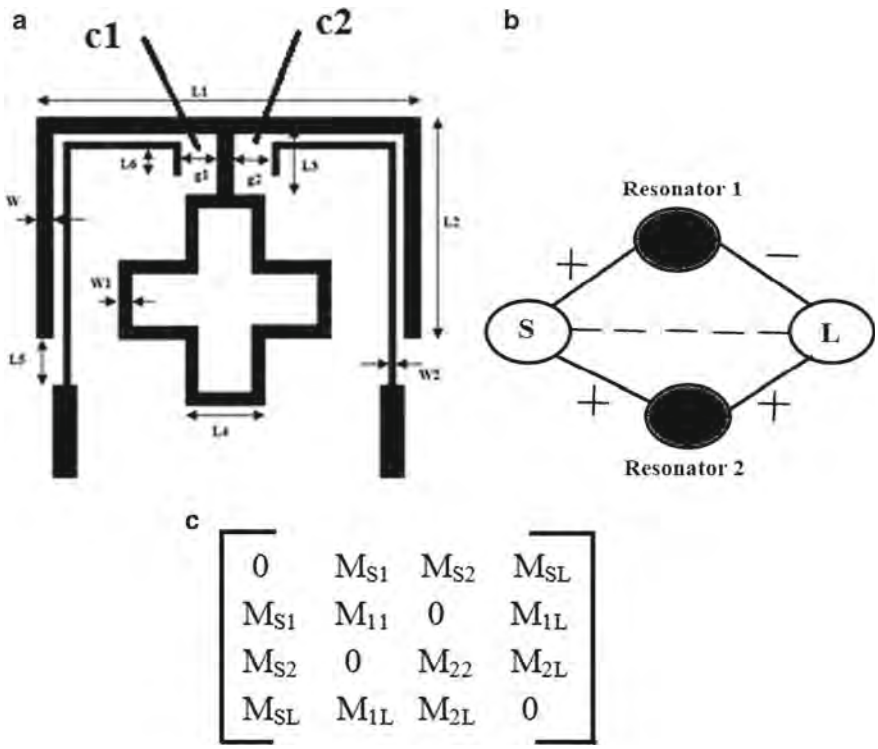


Fig. 1 **a** The layout of the proposed filter **(Filter I)**. **b** Source load coupling scheme. **c** Coupling matrix [3]

(a) The coupling strength between external feeding circuit and odd mode is greater than even mode and hence $\mathbf{M_{s1} > M_{s2}}$
(b) The structure has finite inherent transmission zeros since $\mathbf{M_{s1}}$ and $\mathbf{M_{s2}}$ are not equal.
(c) As $\mathbf{M_{11}} > 0$ and $\mathbf{M_{22}} < 0$, the transmission zeros are greater than zero and appear on the real axis.
(d) Since $f_o > f_{\mathbf{odd}}$ and $f_o < f_{\mathbf{even}}$, the transmission zero appear in the upper stopband.
(e) Two degenerative modes will not couple.

2.1 Filter I

Filter I designed using modified E-shaped dual-mode resonator shown in Fig. 1a exhibits genetic transmission zero in upper rejection band. The proposed filter is designed for passband center frequency $f_o = 2.5$ GHz and fractional bandwidth (FBW) of 0.4. Here, two capacitors C1 and C2 are introduced with source load coupling to control the location of additional transmission zero TZ_2 in the upper

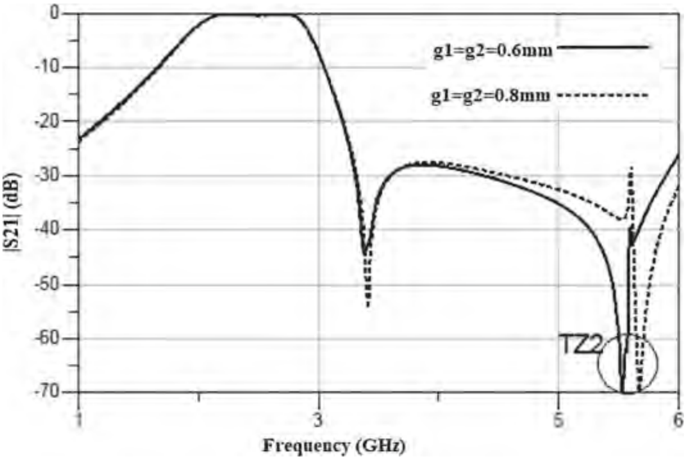


Fig. 2 Change in location of transmission zero TZ_2 with increasing coupling gaps g_1 and g_2

stopband. It is noted that with an equal increase in coupling gaps g_1 and g_2, the equivalent capacitance is reduced which shifts transmission zero TZ_2 to higher frequency location resulting into wider stopband. Figure 2 shows a change in the location of transmission zero TZ_2 with increasing coupling gaps g_1 and g_2. From EM simulation results, it is noted that the designed filter has S21 < 1 dB and S11 > 10 dB in the passband and wide rejection band up to 5.9 GHz with 30 dB attenuation level.

2.2 Filter II

The layout of **Filter II** designed using modified E-shaped dual-mode resonator is shown in Fig. 3. The dimensions of the filter are the same as in Fig. 1a. As the structure is symmetrical, it supports odd and even mode theory. Here, additional two tuning stubs of length L are placed on the inner meander ring to tune even mode resonance frequency. Increase in the length L of two tuning stubs increases the susceptance [7] which shifts even mode resonant frequency and upper passband edge to lower frequencies. From simulation results, it is also noted that the shift of upper passband edge to lower frequency results in narrow passband which increases selectivity and Q of the filter. Figure 4 depicts the shift in even mode resonance and upper passband edge to a lower frequency for various lengths L of tuning stubs. It is observed that these shifts in even mode resonance frequency and upper passband edge results in the tuning of both passband center frequency f_o and bandwidth.

The proposed filter has $S_{21} < 1$ dB and $S_{11} > 10$ dB in the passband for all values of L. From the simulation results, it is noted that with $L = 0$ (without stubs), 1 and 1.5 (in mm) the even mode frequency is at 2.75 GHz, 2.64 GHz, 2.53 GHz, and upper

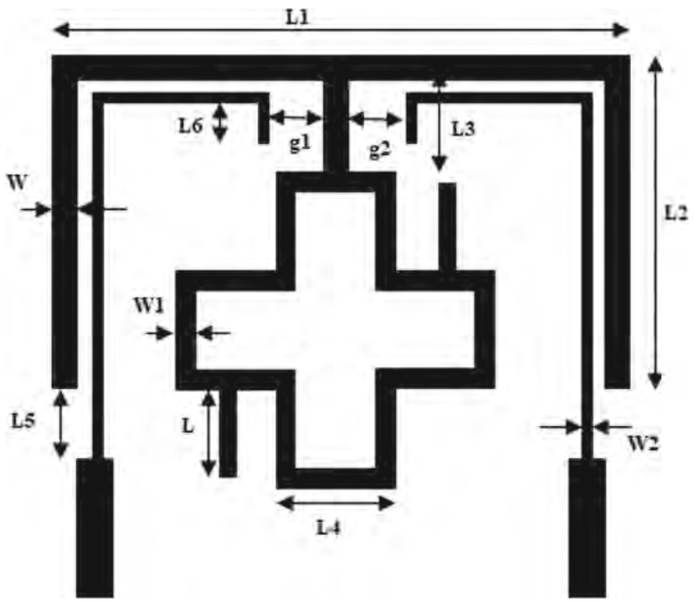


Fig. 3 Proposed dual-mode BPF with two tuning stubs (Filter II)

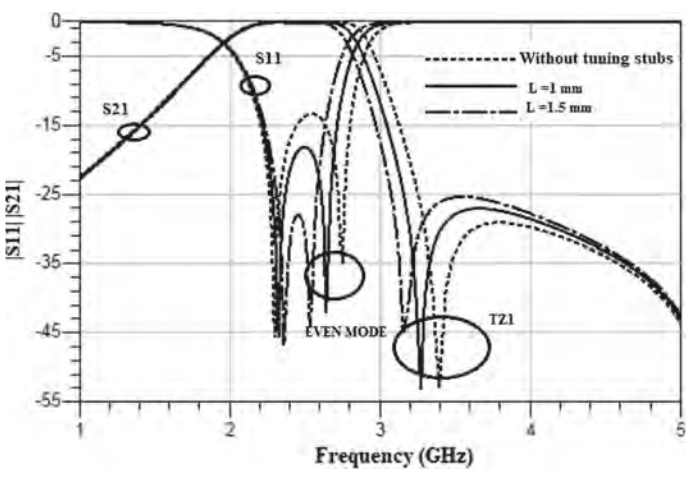


Fig. 4 The shift in even mode resonance and upper passband edge to a lower frequency for various lengths L of tuning stubs

passband edge is at 3.39 GHz, 3.27 GHz, 3.15 GHz, respectively. With $L = 1.5$ mm, the center frequency is tuned from 2.5 to 2.45 GHz and FBW from 0.4 to 0.33.

3 Conclusion

Two types of dual-mode bandpass filters **Filter I** and **Filter II** are designed using modified dual-mode E-shaped resonator at a center frequency 2.5 GHz and fractional bandwidth 0.4. To achieve desired response, first characteristics of the proposed resonator are discussed. In **Filter I**, the location of additional transmission zero TZ_2 is controlled by varying the gaps between feedlines and mid-arm of E shaped resonator. It is noted that wider gaps shifts TZ_2 to higher frequency resulting in a wider rejection band. Both passband center frequency and bandwidth are controlled in **Filter II** using two tuning stubs. The special advantage of **Filter II** is that with an increase in the length of tuning stubs resulted into narrow passband which is important for achieving high selectivity and good Quality factor(Q).

References

1. Lee J-R, Cho J-H, Yun S-W (2000) New compact bandpass filter using microstrip/4 resonators with open stub inverter. IEEE Microwave Guided Wave Lett 10(12) (Dec)
2. Liao C-K, Chi P-L, Chang C-Y (2007) Microstrip realization of generalized Chebyshev filters with box-like coupling schemes. IEEE Trans Microwave Theory Tech 55(1) (Jan)
3. Zhang X-C, Yu Z-Y, Xu J (2008) Design of microstrip dual-mode filters based on source-load coupling. IEEE Microwave Wirel Compon Lett 18(10) (Oct)
4. Wang J, Ge L, Wang K (2011) Compact microstrip dual-mode dual-band bandpass filter with wide stopband. IEEE Electron Lett 47(4)
5. Zhou M, Tang X, Xiao F (2010) Miniature microstrip bandpass filter using resonator-embedded dual-mode resonator based on source-load coupling. IEEE Microwave Wirel Compon Lett 20(3) (Mar)
6. Karpuz C, Özdemir PO, Fırat GB (2016) Design of fourth order dual-mode microstrip filter by using interdigital capacitive loading element with high selectivity. Proc Eur Microwave Conf
7. Kundu AC, Awai I (2001) Control of attenuation pole frequency of a dual-mode microstrip ring resonator bandpass filter. IEEE Trans Microw Theory Tech 49(6) (June)

An Analysis of Air Quality by Vehicular Exhausts

V. Kanpurrani and A. L. Vallikannu

Abstract The level of emission of air pollutant contents is to be monitored to improve the life of the vehicles and improve the quality of the air in the environment. The automatic intelligence system is a dire need to monitor the emission of air pollutants of vehicles, particularly in the smart cities. The proposed system is to monitor the air pollutants in the air and generates the alert for the remedial action when the air pollution exceeds the normal level. The proposed idea has experimented in the simulated smart city environment. The smart city environment is simulated with wireless communications with smart sensors for measuring the air quality, the cloud storage is used for data storage and processing. The experimental results have been recorded and analyzed. The accuracy has been computed and discussed. The proposed system has yielded as high as 11.67% more accuracy than the existing system.

Keywords Radio-frequency identification (RFID) · Internet of things (IoT) · Liquid crystal display (LCD)

1 Introduction

Air pollution control tools are referring to regulate and eliminate the emission of potentially hazardous substances of particulate of gases and matter. The pollution control system is widely understood as air-pollution control system. There are a variety of pollution control systems that focus on different sectors like the air pollution control system and water pollution control system. The pollution control system helps reducing or to prevent polluting particles of hazardous particles to get directly into the environment system.

Nowadays, air pollution control system is a very important concern for industries as many industries releasing toxic waste in the air for that reason every industry

V. Kanpurrani (✉) · A. L. Vallikannu
School of Electrical Sciences, Hindustan Institute of Technology and Science, Chennai, India
e-mail: kanpurrani@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_50

needed to release any kind of air after purification which helps in reducing pollution in the air.

1.1 Major Types of Air Pollution Control

(a) *Particulate Control*: Special machinery is used to filter out particulate pollution from gases. It is a kind of physical matter separation from air.
(b) *Ways of particulate control*: Electrostatic precipitators, cyclone separators, fabric filters
(c) *Gas Control*: In normal chemical methods are used for the separation of pollutant components from gas. It is always useful to remove hazardous gases from waste or else it may lead to acid rain.
(d) *Ways of particulate control*: Scrubbers method, incineration method, carbon capture.

The modeling of air pollution is dependent on the transport and turbulent mixing of emissions. One of the major contributors to pollution is vehicle emissions. Emissions that are released directly from the cars and trucks are the primary source of vehicular pollution. Motor vehicles also pollute the air during the processes of manufacturing, refueling, and from the emissions associated with oil refining [1]. BS are emission standards by the Government of India to regulate the output of air pollutants from combustion engine equipment, motor vehicle. To control the pollution exhausted by vehicles, the amount of air pollution is needed to be calculated, and vehicles causing pollution must be identified. Internet of things may become helpful in cities for monitoring air pollution from vehicles, and the amount of pollution can be gathered and analyzed. This system is specially designed to operate the system using a sensor network and gather information about pollutant levels discharged by the vehicles. To control the emission from the vehicle, the Government has implemented a PUC certification. PUC certification is compulsory for all vehicles on Indian roads. It is to ensure that the vehicle is under the norms of pollution. The validity of the certificate is 6 months. After that, a new certificate must be taken. To overcome this problem, our proposed system has an inbuilt system inside the vehicle that checks the continuous emission values.

2 Literature Survey

The major source of air pollution in the twenty-first century is the road vehicles. The transportation of vehicles exhausted 50% of the carbon monoxide in the air. The increase in population and the busy world, the graph of usage of the personal vehicles in the cities increases in an exponential manner over a decade.

The air pollutant particles include carbon monoxide, nitrogen oxides (NOx), and hydrocarbon (HC) are rapidly increasing in levels in the air. The increase in air pollutants decreases air quality [2]. It also affects the health of the public. The lowest air quality makes people hard to breathe and yields many diseases include cancers. The increase in urbanization and industrialization produces the bad quality of the air. It is very difficult to monitor the vehicles to control air pollution by vehicles. Most of the vehicle users are ignoring to maintain the quality of the vehicles and hence the air quality gets affected.

Though the government has formed many standards and mechanisms to monitor and control air pollution through transportations, controlling air pollution is still not achieved. The ignorance of the vehicle user and the size of populations are the greatest challenges in monitoring vehicle emissions [3]. The dire need of today's air pollution controlling in transport vehicles is an intelligence system that can monitor the air emission particle of the vehicles and analyze the measured data and inform the concerned authorities to take the necessary steps to protect the movement of the vehicles. There are different approaches developed for monitoring vehicular pollution. For instance, in Rushikesh et al. [4] the concept of talking about is to control the pollution level with the help of MQ sensors, RFID, and Arduino. If the pollution level is above the threshold, the authorities will be informed about it. The microcontroller then reads the level and it will be sent to the vehicle owner and the server for future analysis. In Manna et al. [5], the high traffic area is selected to monitor the pollution level with the help of RFID placed at a short fixed distance along with the sensor nodes at the roadside. The tag number of the vehicle is identified by an radio-frequency identification (RFID) reader and is transmitted to the server. When the sensor level is high than the threshold, a message is sent to the owner. According to Priyanka et al. [6], the sensors are placed at the outlets of the vehicles, and the output of it is given to the input of Arduino, and if beyond the threshold value, the information is sent through GSM to IBM Watson internet of things (IoT) platform [7]. The data collected from each of the vehicles are viewed by logging in using user credentials on the dashboard. Considering Pavani et al. [8], the gas sensor used for checking the pollution level is the TGS4161 electrolyte carbon dioxide sensor which needs to be calibrated by exposing them to several concentrations of gas and each sensor output is then plotted accordingly. The goal of Rahman et al. [9] is to interface the output from the sensors used with the online development environment Thing Speak. The signal inputs include temperature variations, sound, potentiometer readings, and so on. It represents the output data as a graphical format as well as in an excel sheet.

3 Proposed System

The proposed work has been designed by sensor module, communication module, cloud storage, learning algorithms, cloud database, and learning algorithms as shown in Fig. 1. The proposed system has experimented in the simulated defined environ-

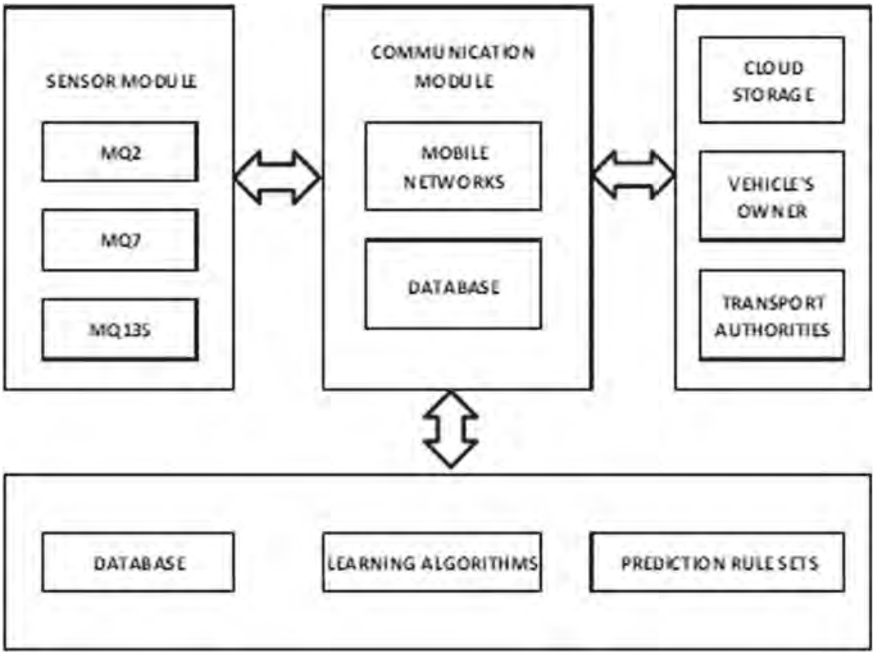


Fig. 1 The block diagram of the air pollution monitoring system

ment. The wireless module was used for communication as in smart city environment [10, 11].

3.1 Vehicle

The vehicle consists of sensors, a controller, a buzzer, and an liquid crystal display (LCD). The sensor has collected the data of pollution, and the level of pollutants has been forwarded to the controller [12]. If the level has exceeded the threshold value, the alert has generated and displayed in the LCD. The controller contains the wireless connection used to send the information to the owner of the vehicle to make him aware of the condition of the vehicle. By collecting the information about the conditions of the vehicular, data has been analyzed from the user-level application. Internet SMS is sent to the vehicle owner by analyzing the data from the application for the use [13].

3.2 Sensors

It measures the emitted values of the exhaust gases and the level of respective gases, like hydrocarbon (mq2), carbon monoxide(mq7), and nitrogen oxide (mq135). The level of gas values has been processed to the algorithm for storage and communication to the cloud storage [14].

3.3 Communication Module with Cloud

It consists of a database and wireless communication, the processing of communication, the sensed data has stored in the database. The stored data has a reference number to be processed with the threshold values of the spatial locations. The reference data has been used for information retrieval between the cloud and the database. The data consist of the level of pollutants, temperature, and id of the owner's vehicle [15]. These databases have been stored in the cloud storage. The cloud is acting as a centralized database for storage and processing. The server will get the information from the vehicle when it attains the threshold value of the level of pollution.

3.4 Processing Module

The processed module consists of a database, learning algorithms, and prediction rule sets. The prediction rule sets are used to predict the near future forecasting to make an advance alert to the users. The sensed data value of the pollutants is compared with threshold value as defined. If the sensed data value exceeds the threshold value, the alert is generated. The generated alert is communicated to the users and other competent authorities for remedial action. The learning algorithm is illustrated as follows.

Pseudocode: Vehicle Pollution Monitoring System – Generating Alert
Input Environment = Transport Vehicle in Smart City
Sensors Used = mq2, mq7, mq135
Time Interval to Measure = 30 min
Communication = Wireless /Mobile Network
Storage space = Data Server, Cloud Setup

For each time interval 't'

1. *The mq2, mq7, mq135 sensors measures the level of gas carbon monoxide (CO), nitrogen oxides (NOx), hydrocarbon (HC).*
2. *The measured data is recorded and compared with the threshold values.*

3. *If the trend of measured data is likely to fall into the threshold value of the geographical locations, the alert is generated and sends to the owner of the vehicles.*
4. *If the measured data is exceeded the threshold value, the alert is generated and send to the owner of the vehicles and the concern authorities.*
5. *Else if Step 2 to Step 4 is repeated in a looping fashion.*

The proposed system has experimented in the simulated smart city. The smart city is simulated with wireless communications and a common cloud for data storage. The government authorities include transport, traffic control, and mobile communications are also simulated in the defined smart city.

4 Result and Analysis

The accuracy of alert is considered as performance parameters, it is defined as

$$\text{Accuracy of Alert } = \frac{\text{No.of\_ Actual\_ Alerts\_ of\_ 100\_ trails}}{\text{No.of\_ Expected\_ Alerts\_ of\_ 100\_ trails}}$$

The experiments are carried out independently for 100 trails and the data is recorded. The average is taken after ignoring the outline's values. The analysis was carried out under four cases as True-True, True-False, False-True, False-False cases as defined in Table 1.

The presented work analyzed the pollution control data in four scenarios, the alert conditions were tabulated in Table 1. The interactive results of predicted accuracy have yielded better than the actual values of various scenarios conditions of the variables.

The accuracy of alert conditional values has tabulated in Table 2. The proposed work yielded a better performance as 98.1% for False–False case as shown in Table 2. However, the proposed work could not get better performance for True–False case as shown in Fig. 2. The accuracy of actual values for the four categories as True–True, True–False, False–True, False–False as shown in Fig. 2.

The results are recorded for prediction analysis. The accuracy on prediction is plotted as shown in Fig. 3. The proposed work was yielded the better performance as 97.98% for False–False case.

Table 1 Recorded scenarios

S.NO	Condition	Alert
1	T_c and T_c	The threshold value has exceeded
2	T_c and F_c	Not generated and the actual data exceed threshold values
3	F_c and T_c	Generated and the actual data exceed threshold values
4	F_c and F_c	Not Generated and the actual data is not exceeded threshold values

Table 2 Accuracy of alert generation on exceed the threshold values

S.NO	Condition	Accuracy (%)
1	T_c and T_c	61.76
2	T_c and F_c	43.21
3	F_c and T_c	71.78
4	F_c and F_c	98.10

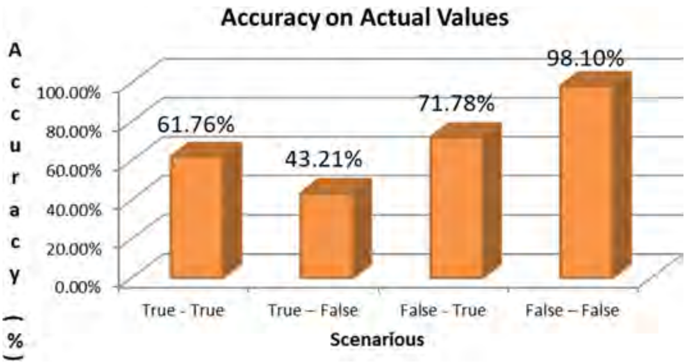


Fig. 2 Accuracy on actual values for alert generation

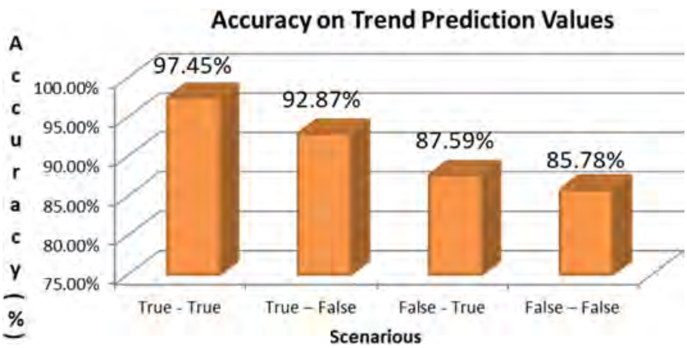


Fig. 3 Accuracy of alert generation on trend prediction

The proposed system is performing well as 97.45% in the True-True scenario. The performance was 92.87%, 87.59% and 85.78% for True-False, False-True, and False–False, respectively. It proved that the proposed work is suitable for the accuracy on prediction while monitoring the pollutants in the air quality.

5 Conclusion

It is a dire need to monitor air pollution in the city to maintain the quality of air. The proposed system has been designed to monitor the air quality in the city. The sensor module has measured the air quality level. The measured values are compared with the threshold values and generate the alert as needed. The learning algorithm plays a vital role in predicting the pollution level shortly. The proposed system has yielded a better performance as 99.61%, 98.10% for prediction, real-time, respectively. The future direction of the work shall focus on the real-time measurement of pollutants in the air.

References

1. Chiti F, Fantaccia R, Gu Y, Han Z (2017) Content sharing in internet of vehicles: two matching-based user-association approaches. Veh Commun 8:35–44 (Apr)
2. Contreras-Castillo J, Zeadally S, Guerrero-Ibáñez JA (2018) Internet of vehicles: architecture, protocols, and security. IEEE Internet Things J 5(5):3701–3709 (Oct)
3. Duangsuwan S, Takarn A, Nujankaew R, Jamjareegulgarn P (2018) A study of air pollution smart sensors LPWAN via NB-IoT for Thailand Smart Cities 4.0. In: 2018 10th international conference on knowledge and smart technology (KST), Chiang Mai, 2018, pp 206–209. https://doi.org/10.1109/kst.2018.8426195
4. Korunoski M, Stojkoska BR, Trivodaliev K (2019) Internet of things solution for intelligent air pollution prediction and visualization. In: IEEE EUROCON 2019—18th international conference on smart technologies, Novi Sad, Serbia, 2019, pp 1–6. https://doi.org/10.1109/eurocon.2019.8861609
5. Kularatna N, Sudantha B (2008) An environmental air pollution monitoring system based on the IEEE 1451 standard for low cost requirements. Sens J IEEE 8:415–422. https://doi.org/10.1109/JSEN.2008.917477
6. Divya A, Kiruthika R, Gayathri D (2019) Detecting and analysing the quality of air using low cost sensors to reduce air pollution in urban areas. In: 2019 IEEE international conference on system, computation, automation and networking (ICSCAN), Pondicherry, India, 2019, pp 1–5. https://doi.org/10.1109/icscan.2019.8878780
7. Kavitha BC, Vallikannu R (2019) IoT based intelligent industry monitoring system. In: 2019 6th international conference on signal processing and integrated networks, SPIN 2019, 8711597, pp 63–65
8. Malky S, Kostanic I, Altheiab K, Alharbai W (2019) Evaluation of precalibrated electrochemical gas sensors for air quality monitoring systems. In: 2019 international conference on internet of things (iThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE smart data (SmartData), Atlanta, GA, USA, 2019, pp 967–973
9. Postolache OA (2009) Smart sensors network for air quality monitoring applications. IEEE Trans Instrum Meas 58(9) (September)
10. Xu Y, Chen X, Liu A, Hu C (2017) A latency and coverage optimized data collection scheme for smart cities based on vehicular ad-hoc networks. Sensors 17(4):888
11. Hemanth J, Fernando X, Lafata P, Baig Z (2018) International conference on intelligent data communication technologies and internet of things
12. He Y, Sun D, Zhao M, Cheng S (2018) Cooperative driving and lane changing modeling for connected vehicles in the vicinity of traffic signals: a cyber-physical perspective. IEEE Access 6:13891–13897

13. Silva CM, Silva FA, Sarubbi JFM, Oliveira TR, Meira W Jr, Nogueira JMS (2017) ‘Designing mobile content delivery networks for the Internet of vehicles’. Veh Commun 8:45–55 (Apr)
14. Meneguette RI, Boukerche A (2017) SERVitEs: an efficient search and allocation resource protocol based on V2V communication for vehicular cloud. Comput Netw 123:104–118 (Aug)
15. Rani VK, Sudalai MT (2015) An innovative design of intelligent cradle for infants. Int J Appl Eng Res 10(8):20867–20873

Secured Data Encryption and Decryption for Transmission in Device Independent Systems

B. N. Lakshmi Narayan, N. H. Prasad, and A. Mariyan Richard

Abstract In general, a company will have its branches at different locations and every branch will have different hardware and software configurations. Different configurations have different issues in transferring the data from various branches to corporate office and vice versa. To overcome the challenge, this application remains as an alternative solution. Data transferred through the network remains insecure due to its plain text format. Regarding the data restoration, it becomes incompatible for software as well as hardware. For security concern, the security process handling in data encryption and data decryption using cryptic service provider for RSA encryption and decryption using public and private key. The keys were stored in server as a XML format. This encryption is an asymmetric encryption. This data also accessed through device independent system like mobile, PDA, etc. The quality of file to trace in progress prominence of the file which users works on and after getting the files, the file will be sent to quality analysis person and it will be obtained back with the status before restoring a database. The quality of file will be checked and sent back to the administrator. The unsatisfactory files are then sent back to client for rectification.

Keywords Data Encryption · Triple Data Encryption · RSA · Asymmetric Encryption

1 Introduction

The paper entitled in secured data transmission in device independent system which deal regarding the issue of data security. The conventional method of data is plain text format. This method will raise the security issue of data when transfer from

B. N. Lakshmi Narayan (✉) · N. H. Prasad · A. Mariyan Richard
Department of MCA.NMIT, Bengaluru, India
e-mail: narayan614@gmail.com

N. H. Prasad
e-mail: naikphd@gmail.com

A. Mariyan Richard
e-mail: mariyanrich01@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_51

source to destination. Any social networking site user, example a Facebook user can login to his account by giving his user id and password. If the user is authorized, only then he would be allowed to get his Facebook home page.

The administrator to get files from the client and register the file information, depending on the information, the encrypted file was decrypted and stored into server database. The restored data was fetched from database and processed by administrator. If the file to be unsatisfactory, then the file is send back to the client for correction. The security process was handling data encryption and data decryption using cryptic service provider.

This paper consist of following module.

- Authentication Module
- Database Selection
- Key Generation Module
- Encryption
- Decryption

Generally, the authentication module deals with the process of authentication. A Facebook user will get his Facebook homepage only after he has cleared the process of authentication [1]. Unauthorized persons were automatically redirected into authentication area [2]. This module provides to create new user registration, change their password, and recover password when they are forget it. The server deals with the admin side [3]. The administrator gets files from the clients and registered to it. The files are in encrypted XML format [4]. The files are decrypted and data will be restored into server database [5]. The administrator fetches data from database and converted into machine independent language then uploaded into server. The client module deals with the client side [6]. The client fetches data from database and provides security to that data and converted into device independent language. This data was uploading to the server for administrator verification [7]. The security deals with handling data encryption and data decryption using cryptic service provider. The public and private keys will be created and stored in server as a XML format [8]. The public key and private key was created and stored into Internet information server [9]. The keys used when security was applied. The conversion module deals with the data convert XML format into plain text format vice versa [10–13]. This conversion was taken place in server and client side. The quality analysis module deals with the data verification of each file before restores [14]. The files are sending to quality analysis persons for data verification. The analyzer was analyzed to it and sends back the file with status.

This module helps us to track the data through various devices such as mobile phone, PDA, etc.

(A) *Demerits Of The Existing System*

- The data transfer is a plain text format.
- Due to plain text, the data will be insecure
- Bulk data will not transfer from source to destination

- The database structure will be copied and restored
- The system occupy huge memory space
- Possible of missing data
- The version of software is incompatible

2 Proposed System

To provide security when the data was transfer in an insecure channel and convert machine or device independent language for accessing the external devices or systems.

The proposed system is a computer-based system which overcomes the existing system. This proposed system provided to fetch all the data from database and converted into encrypted XML format. This XML format do not dependent any operating system, mail server, browser, server, and any other hardware. So, this data was software and device independent to the systems.

(A) *Benefits of Proposed System*

- While fetching data, the database structure will not copied.
- The Rivest–Shamir–Adleman (RSA) Encryption / Decryption standard was used for security.
- The data will be converted into XML format.
- The XML file will be checked for quality analysis person before restore database.
- Movement of data will be traced by both server side and client side persons.
- The data will be accessed through device independent system.

3 System Design

This paper has the process of encryption and decryption methods to send the data from one place to another place. To proceed this process, we have to decide which data has to be sent. After the selection of database, we have to choose the algorithm to generate keys to encrypt and decrypt. The following sections define the algorithms which are used in this paper.

(A) *RSA*

RSA is a one of the first public-key cryptosystems algorithm is used to generate keys for below signature.

- Public Key: (n,e)
- Private Key: (n,d)

(B) *RIJNDAEL/AES*

Rijndael is a symmetric algorithm. The rijndael is also called advanced encryption standards (AES). This has the key size of 128,192, or 256 bits (32 bytes). The keys here are.

- Password,
- Secret Key.

(C) *Ron Rivest Cipher2 (RC2)*

Ron Rivest who designed RC2 replaced DES with his designed cipher. The designed cipher recorded the speed which was 3 times more compared to DES. The block size of input and output are 64 bits each. The key size of the block differ from one byte and it may reach up to 128 bytes. The proposed system and its implementation are used up to eight bytes. The proposed algorithm is designed to make a firmware for 16-bit microprocessors. The keys are.

- Password,
- IV (Initialization Vector)

(D) *Data Encryption Standards (DES)*

The encryption and decryption of electronic data goes to as the input of 64 bit blocks into DES with 64 bits as cipher text. The key length of DES is 56 bit. But, it can accept the input key of 64-bit long. *DES works on bits (0′s and 1′s), where group of four bits forms hexadecimal number.*

Example: (i) 1111 = F.

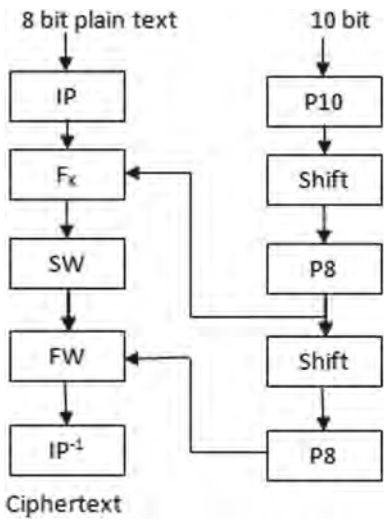
(ii) 0011 = 3.

DES works on bits or binary numbers by encrypting to digital computers. Hexadecimal format is used to encrypt the 64 message bits. DES takes 64 bits of data into each set by converting it into two blocks of 32 bits each.

(a) L = Left half block
(b) R = Right half block.

DES has majorly features into five functions.

(a) Initial Permutations (IP)
(b) Complex functions (FK), which involves both real and imaginary arguments and it produces same structure of complex types.
(c) A combination of one of its permutation function which switches (SW) the controls from one block to other block (i.e., L and R).
(d) Switch Function (SF).
(e) Inverse Initial Permutation (IP-1).



where

Ciphertext = IP-1(Fk2(SW(Fk1(IP(Plaintext))))).

And Key K1 = P8 (Shift(P10(key))).

K2 = P8(Shift(Shift(P10(key)))).
Example: Let P be the plain text message of variable M = 10,111,010, K represents as a key with value of K = 1,001,101,011.
DES is itself is fairly divides the key into 10-bits between the sender and receiver. After receiving the main key, it generates two other sub keys of size 8 bits each.
First, let the 10-bit key is initialized as follows.
K10 (k1, k2, k3, k4, k5, k6, k7, k8, k9, k10).
And permutation P10 is initialized as below:
P10 (k1, k2, k3, k4, k5, k6, k7, k8, k9, k10) = (k3, k5, k2, k7, k4, k10, k1, k9, k8, k6).
So, we consider the initial permutation combination using the IP function as below:
IP (1 2 3 4 5 6 7 8) = (2 6 3 1 4 8 5 7).
So, according to IP, the plaintext gives the output as below.
IP: 11,010,101.
Inverse the initial permutation, as result.
P-1 = 10,101,011.

K = 000,101,111.

CS1 = 00,001 11,111.

CS2 = 10,000 11,111.

As a result of the permutation, we define the mapping variable F, where it takes the 4-bit numbers (n1,n2,n3,n4) and the permutation function is obtained from its first excitation.

Permutation function							
4	1	2	3	2	3	4	1

As explained above, the comparison of the 4-bit values with permutation values, the following deliverables is given below

$$\begin{array}{c|c} n_4 & n_1 \\ n_2 & n_3 \end{array} \qquad \begin{array}{c|c} n_2 & n_3 \\ n_4 & n_1 \end{array}$$

$K_{18=}$ 1000 1111.

K_{28} = 1000 1111.

IP = 1101 0101

$$\begin{array}{c|c} 0+1 & 1+0 \\ 0+1 & 1+1 \end{array} \qquad \begin{array}{c|c} 0+0 & 1+0 \\ 0+1 & 1+1 \end{array}$$

The XOR operations results in: 01,010,101.

Let us rename these 8 bits:

$$\begin{array}{c|cc|c} p_{0,0} & p_{0,1} & p_{0,2} & p_{0,3} \\ p_{1,0} & p_{1,1} & p_{1,2} & p_{1,3} \end{array}$$

10000 1111

(E) *DES Modes of Operations*

The encryption is DES algorithm converts the 64-bit message block M into 64-bit cipher block. If the encryption is processed individually for all the blocks, then this encryption mode is termed as electronic code book mode. Other than ECB mode, there are two other modes used in DES encryption such as cipher feedback (CFB) and chain block coding (CBC). In this modes, each cipher block is dependent on on all the previous messages blocks through an initial XOR operation.

(F) *Triple Data Encryption Standard (TDES)*

TDES is a symmetric key block cipher which is 3 times greater when multiplied with the DES cipher algorithm. The data encryption is performed with keys in a systematic order and TDES considers all the three 64-bit keys for an overall key length of 192 bits.

4 Input Design

The overall system design can be done with the help of input design, which requires a deliberate attention. Input data collection is an essential part of the system which depends on the equipment and number of people.

Making the data entry easy and free from logical error is the main objective of input data.

The objectives of the input design phase are to obtain a cost effective method to achieve the high accuracy level which makes the input as acceptable and understandable.

In this system, the table authentication used to check the user authorization. Then, we can choose tables to encrypt and decrypt.

5 Output Design

The changes made to input records, the normal procedure is to design the outputs in details first and then to work back to the inputs. The input records have to be validated, edited, organized, and accepted by the system before being processed to process to produce the outputs.

The result generated from the system is considered as output, and it is used to evaluate the utilization and effectiveness of the application. To obtain efficiency reports, software systems can be utilized. In this paper can layout the file register and file restore as a reports.

6 Implementation

System implementation phase is concerned with translation of the design specifications into the source code and internal documentation. This helps the testing, debugging, and modifications much simple, and the objective of the research can be obtained by projecting the source code in a clear manner. Generally, elegance and simplicity, clarity are the essential features of a good programs.

(A) *Maintenance*

The primary goal, of various phases of software engineering process discussed so far, is to improve the maintainability of software. That is, to improve the ease with which changes can be accommodated and to reduce the amount of time expanded on maintenance.

During the use of a large program, errors will occur and be reported to the developer. The error reported has to be tracked back to the source code (diagnosis) and corrected. The ease with which an error can be tracked back and corrected depends on how well the program is designed.

It is a software maintenance activity necessitated by change in the operating environment of the software. The environmental changes include, hardware changes, operating system changes. To accommodate these changes, the software has to be modified. Unlike corrective maintenance, this maintenance activity may require a complete design change in the software.

7 Conclusion

In this research work, the proposed algorithm was implemented by front end applications Asp.net, C#, and SQL server 2005 as a database. The data was fetched from database and stored in data table. This data was available in plain text format. This data was converted into bytes then converted into cyber bytes. These cyber bytes were converted into cyber text. These data were put into container and send to server through insecure channel. In other side, the data was received and decrypted, analyzed, and stored into database as single XML tag. The security was applied using encryption standards [Asymmetric and Symmetric Encryption]. In future, all symmetric and asymmetric encryption can be applied to the data as per user choice.

References

1. Nikita Gorasia RR, Srikanth ND, Rupareliya J (2016) Improving security in multi authority attribute based encryption with fast decryption. Procedia Comput Sci 79:632–639
2. Qian Q, Yu Z, Zhang R, Hung CL (2018) A multi-layer information dispersal based encryption algorithm and its application for access control. Sustain Comput Inf Syst 20:76–87
3. Bouhous A, Kemih K (2018) Novel encryption method based on optical time-delay chaotic system and a wavelet for data transmission. Opt Laser Technol 108:162–169
4. Sartakhti JS, Jalili S (2019) On the computational power of the light: a plan for breaking data encryption standard. Theoret Comput Sci 773:71–78
5. Masoumi M (2019) A highly efficient and secure hardware implementation of the advanced encryption standard. J Inf Secur Appl 48:1–12
6. Artiles JAP, Chaves DPB, Pimentel C (2019) Image encryption using block cipher and chaotic sequences. Signal Process Image Commun 79:24–31
7. Zhang K, Chen J, Lee HT, Wang H (2019) Efficient public key encryption with equality test in the standard model. Theoret Comput Sci 755:65–80
8. Crocett L, Baldanzi I, Bertolucci M, Fanucci L (2019) A simulated approach to evaluate side-channel attack countermeasures for the Advanced Encryption Standard. Integration 68:80–86
9. Hiscock T, Savry O, Goubin L (2019) Lightweight instruction-level encryption for embedded processors using stream ciphers. Microprocess Microsyst 64:43–52

10. Nardo LG, Nepomuceno EG, Arias-Garcia J, Butusov DN (2019) Image encryption using finite-precision error Chaos. Solitons & Fractals 123:69–78
11. Sussman D, Ullman C (2006) Beginning ASP.Net with C#, 2006 Edition
12. Nagel C, Skinner M (2004) Bill Evjen "Professional C#" Wiley publishing 2004, Edition
13. Mollin RA (2003) RSA and public-key cryptography, CRC Press 2003 Edition
14. Lewko A, Waters B (2010) Decentralizing attribute-based encryption, Cryptology ePrint Archive: Rep. 2010/351

Electrical Performance of Single Gate and Double Gate MOSFET to Optimize the Gate Length

James Lalthlamuana, Niladri Pratap Maity, L. L. K. Singh, and Reshmi Maity

Abstract This work investigated the electrical performances of a single gate (SG) and double gate (DG) metal oxide semiconductor field-effect transistors (MOSFETs) by varying the gate length. The electrical characteristics are analyzed and parameters like ON-state current, OFF-state current, and transconductance are considered. A similar study was conducted by changing the oxide material to high-k material (HfO_2). It is perceived that for gate length below 25 nm, overall channel potential starts decreasing due to increased source-drain resistance and velocity saturation owing to short channel effects. The drain current decreases as the gate length are decreased for both the SG and DG MOSFET. The transconductance also decreases inversely with the gate length for SG with the graded channel, SG with graded channel using a high-k dielectric, DG with the graded channel, and DG with graded channel using high-k dielectric.

Keyword Single gate (SG) and double gate (DG) metal oxide semiconductor field-effect transistors (MOSFETs)

1 Introduction

MOSFET has to turn into one of the most significant devices in the manufacturing industry because of its successful incorporation into integrated circuits (ICs). Even though the physical structure of MOSFET has not changed in full, its size has been repeatedly reduced by double every 2–3 years in accord with Moore's Law [1]. To look for probable substitutes for bulk MOSFETs beyond the 45 nm technology node, many innovative multi-gate MOSFETs have been suggested. They are gate-all-around (GAA) MOSFET [2], TFET [3], FinFET [4], Tri-Gate [5] and DG MOSFETs [6]. DG MOSFET is a freshly developing device that can more scale down complementary MOSFET (CMOS) technology in the sub-50 nm regime owing to its outstanding control of short channel effects.

J. Lalthlamuana · N. P. Maity (✉) · L. L. K. Singh · R. Maity
Department of Electronics and Communication Engineering, Mizoram University (A Central University), Aizawl 796 004, India
e-mail: maity_niladri@rediffmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_52

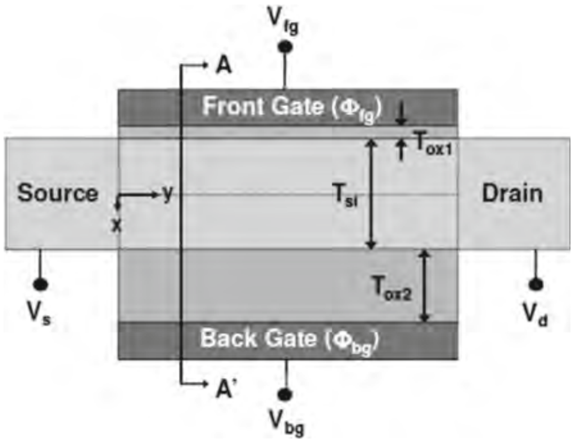


Fig. 1 Configuration of DG MOSFET

2 Configuration of DG MOSFET

Enhanced scalability must be talented by adding another gate terminal on the backside of the body terminal of the standard MOSFET ensuing in the double-gated configuration. The configuration of a double-gated FET is depicted in the inset to Fig. 1.

The downscaling of CMOS technology take on SiO_2 as insulator material has got hold of its maximum; there is an idea in diminishing the electrical oxide thickness more by using high-k materials [7–9]. Among the high-k applicants, HfO_2 is the greatest encouraging material [10–14].

This work focuses on the optimization of the electrical parameters in both SG and DG MOSFETs by the implementation of the graded channel and source/drain engineering and further, improving the performance of the device with the introduction of the high-k material. ON-state current (I_{ON}), OFF-state current (I_{OFF}), and transconductance (g_m) are considered for the analysis. Channel engineering is also employed to achieve the best performance with a graded channel which controls several short channel effects (SCEs) [15–18].

3 Results and Discussion

The electrical performances of SG and DG MOSFETs are analyzed and optimized by varying the gate length. The electrical characteristics are analyzed, and parameters like I_{ON}, I_{OFF}, and transconductance are considered. Polycrystalline silicon (poly-Si) is again used as gate material with work function set as 4.17 eV. Room temperature (300 K) is considered with ohmic contacts for source, drain, and substrate. Further analysis is done using the optimized device parameters. The proposed work has been implemented on different device structures using different dielectric materials and classified into: SG with a graded channel, SG with graded channel using a high-k dielectric, DG with a graded channel, and DG with a graded channel, and high-k dielectric. Figures 2 and 3 represent $I_d - V_g$ characteristic curves of SG MOSFET with the graded channel and SG MOSFET with graded channel using high-k dielectric, respectively. Figures 4 and 5 represent $I_d - V_g$ curves of the DG MOSFET with the graded channel and DG MOSFET with graded channel using high-k dielectric, respectively.

Figure 6 and 7 represent $I_d - V_d$ characteristic curves of SG MOSFET with graded channel and SG MOSFET with graded channel using high-k dielectric material, respectively. Figure 8 and 9 represent $I_d - V_d$ characteristic curves of DG MOSFET with graded channel and DG MOSFET with graded channel using high-k material, respectively.

Figures 10 and 11 represent $g_m - V_g$ characteristic curves of SG MOSFET with graded channel and SG MOSFET with graded channel using high-k dielectric material, respectively. Figures 12 and 13 represent $g_m - V_g$ characteristic curves of DG MOSFET with graded channel and DG MOSFET with graded channel using high-k material, respectively.

From the above graphs and Table 1, it is seen that I_{ON} is maximum at the gate length of 25 nm. For gate length below 25 nm, the overall channel potential starts decreasing due to increased source-drain resistance and velocity saturation owing to short channel effects. Equally seen, the drain current decreases as the gate lengths are decreased for both SG and DG MOSFETs. The transconductance also decreases inversely with the gate length for SG with graded channel, SG with graded channel using a high-k dielectric, DG with graded channel, and DG with graded channel using high-k dielectric. The transconductance has a peak value at the gate length of 25 nm.

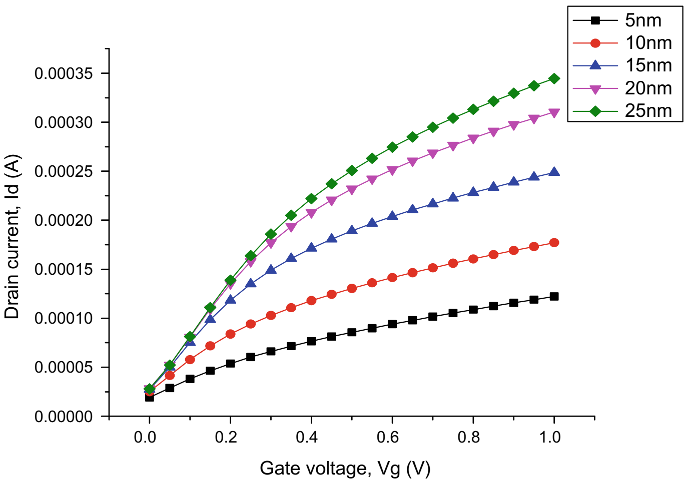


Fig. 2 $I_d - V_g$ plot of SG MOSFET

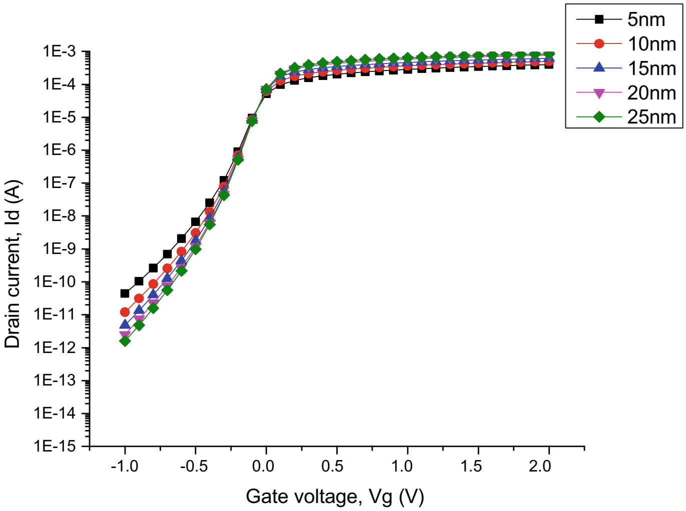


Fig. 3 $I_d - V_g$ plot of SG MOSFET with graded channel using HfO_2

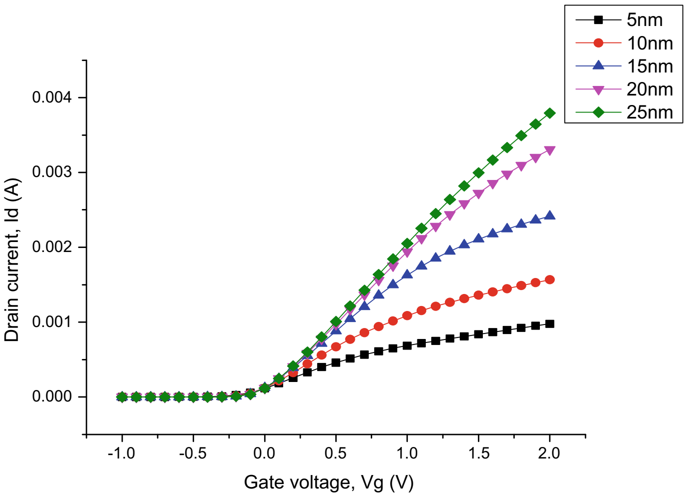


Fig. 4 $I_d - V_g$ plot of DG MOSFET

I_{OFF} should be ideally zero to have good subthreshold characteristics. The I_{OFF} is minimum when the gate length is 25 nm as the gate leakage current through the very thin dielectric material is minimum at this length. Hence, it can be said that the device performance is optimum for gate length, $L_G = 25$ nm for all the four cases.

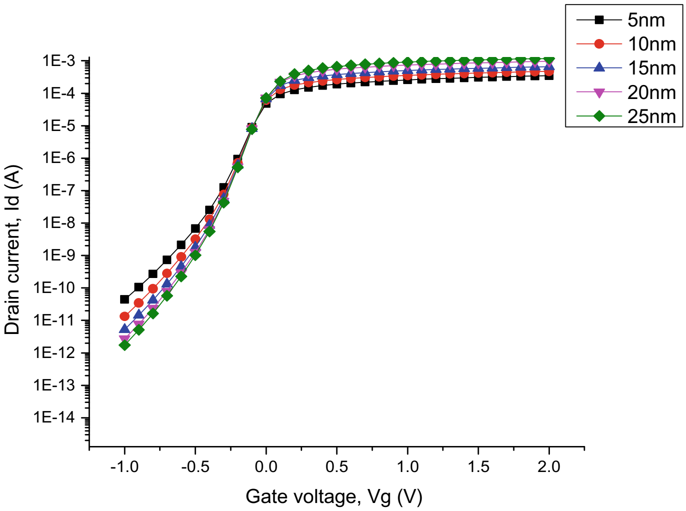


Fig. 5 $I_d - V_g$ plot of DG MOSFET with graded channel using HfO_2

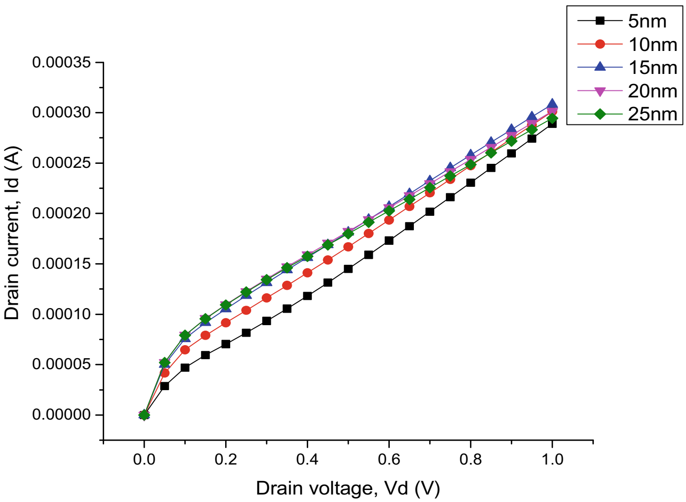


Fig. 6 $I_d - V_d$ plot of SG MOSFET

4 Conclusion

Single gate metal oxide semiconductor field-effect transistors have been optimized for better performance by using the graded channel and source/drain engineering with lightly doped drain structure to maintain satisfactory threshold voltage level, channel mobility, and punch through a mechanism. Channel engineering has been utilized

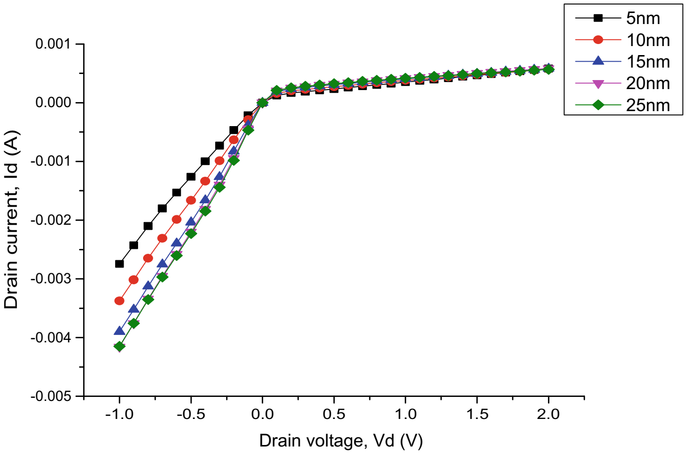


Fig. 7 $I_d - V_d$ plot of SG MOSFET with graded channel using HfO_2

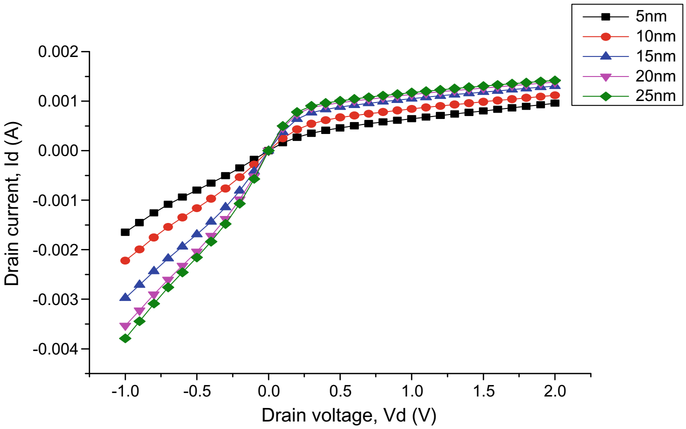


Fig. 8 $I_d - V_d$ plot of DG MOSFET

to obtain the best performance with the graded channel which controls the short channel effects. Then, the device was simulated for obtaining the best performance parameters by varying the gate length. A similar study was conducted by changing the device structure and material with changing the oxide material to HfO_2 (high-k material), the addition of second gate (DG MOSFET) and changing oxide material to HfO_2 in DG MOSFET. For all four cases, the gate length was different from 5 to 25 nm. The high current drive; i.e., peak drain current and peak transconductance values were observed for $L_G = 25$ nm. It was observed that DG has better performance than SG MOSFET in terms of current drive and transconductance. Also, the use of

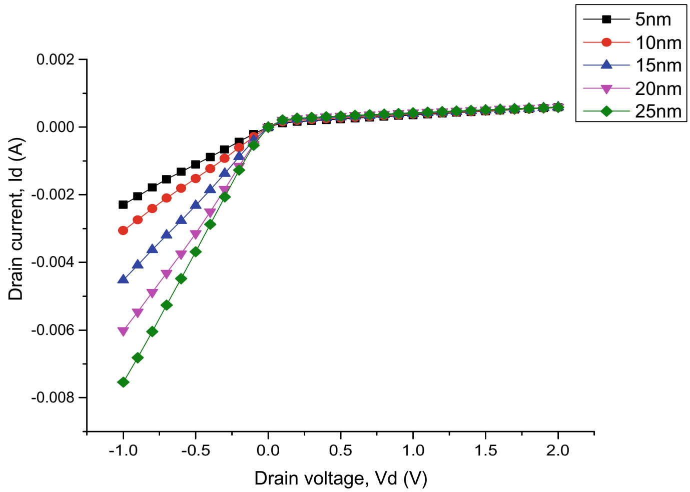


Fig. 9 $I_d - V_d$ plot of DG MOSFET with graded channel using HfO_2

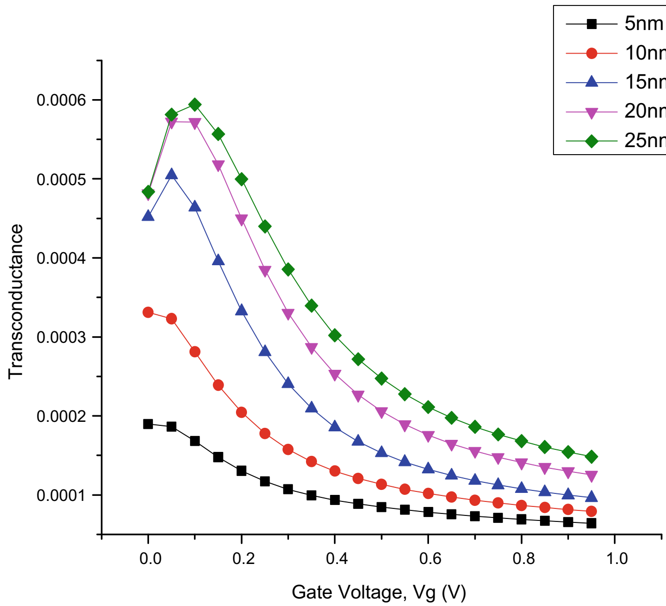


Fig. 10 $g_m - V_g$ plot of SG MOSFET

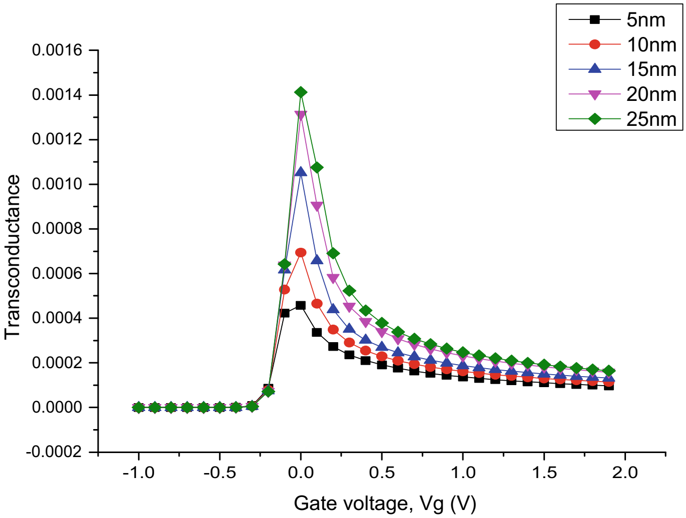


Fig. 11 $g_m - V_g$ plot of DG MOSFET with graded channel using HfO_2

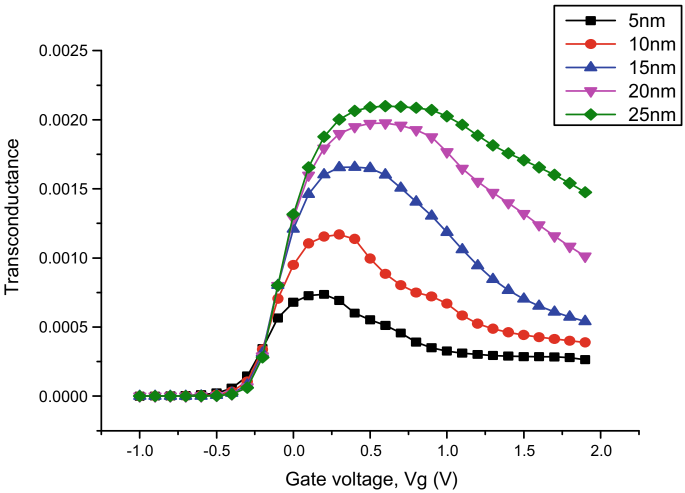


Fig. 12 $g_m - V_g$ plot of DG MOSFET

high-k material, i.e., HfO_2 in both single gate and double gate improves the electrical performance of the MOSFETs.

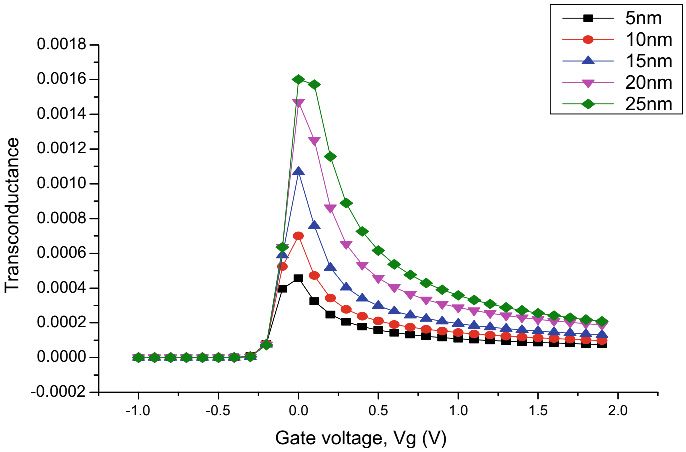


Fig. 13 $g_m - V_g$ plot of DG MOSFET with graded channel using HfO2

Table 1 Optimization of gate length

Description	Para-Meter	$L_G = 5$ nm	$L_G = 10$ nm	$L_G = 15$ nm	$L_G = 20$ nm	$L_G = 25$ nm
SG with graded channel	I_{ON}	0.00012	0.00017	0.00024	0.00031	0.00034
	I_{OFF}	1.919×10^{-5}	2.506×10^{-5}	2.748×10^{-5}	2.812×10^{-5}	2.8×10^{-5}
	g_{max}	6.408×10^{-5}	7.924×10^{-5}	9.642×10^{-5}	1.253×10^{-4}	1.4838×10^{-4}
SG with graded channel using HfO_2	I_{ON}	0.00040	0.00049	0.00062	0.00076	0.00084
	I_{OFF}	4.43×10^{-11}	1.21×10^{-11}	4.84×10^{-12}	2.5×10^{-12}	1.578×10^{-12}
	g_{max}	9.7×10^{-5}	0.0001141	0.0001314	0.0001581	0.00016442
DG with graded channel	I_{ON}	0.00097	0.00156	0.00241	0.00330	0.00379
	I_{OFF}	4.601×10^{-8}	9.332×10^{-9}	2.558×10^{-9}	1.107×10^{-9}	6.569×10^{-10}
	g_{max}	0.0002635	0.0003892	0.0005412	0.0010118	0.0014753
DG with graded channel using HfO_2	I_{ON}	0.00034	0.00047	0.00065	0.00096	0.00118
	I_{OFF}	4.44×10^{-11}	1.34×10^{-11}	5.292e-12	2.73×10^{-12}	1.73×10^{-12}
	g_{max}	7.58×10^{-5}	9.85×10^{-5}	0.0001317	0.000188	0.0002088

References

1. Moore G (1965) Cramming more components onto integrated circuits. Electron 38:114–117
2. Rana V, Ahmed G, Ramesh A, Das S, Singh P (2020) Diameter depended piezoresistive sensing performance of junctionless Gate-All-Around Nanowire FET. IEEE Trans Electron Dev 67(7):2884–2889
3. Li W, Woo J (2020) Vertical P-TFET with a P-Type SiGe Pocket. IEEE Trans Electron Devices 67(4):1480–1484

4. Maity NP, Maity R, Baishya S (2019) An analytical model for the surface potential and threshold voltage of a double-gate heterojunction tunnel FinFET. J Comput Electron 18:65–75
5. Ma J, Matioli E (2017) Slanted tri-gates for high-voltage GaN power devices. IEEE Electron Device Lett 38(9):1305–1308
6. Chakraborty H, Maity R, Maity NP (2019) Analysis of surface potential for dual-material-double-gate MOSFET based on modeling and simulation. Microsyst Technol 25:4675–4684
7. Maity NP, Maity R, Baishya S (2019) A tunneling current model with practical barrier for ultra thin high-k dielectric ZrO_2 material based MOS devices. Silicon 10:1645–1652
8. Maity NP, Maity R, Baishya S (2017) Influence of image force effect on tunneling current density for high-k material ZrO_2 ultra thin films based MOS devices. J Nanoelectron Optoelectron 12:67–71
9. Maity NP, Maity R, Thapa RK, Baishya S (2014) Study of interface charge Ddensities for ZrO_2 and HfO_2 based Metal-Oxide Semiconductor devices. Adv Mat Sci Eng Article ID 497274, 1–6
10. Chaudhry A, Kumar MJ (2004) Controlling short-channel effects in deep submicron SOI MOSFETs for improved reliability: A review. IEEE Trans Dev Mater Rel 4:99–109
11. Maity NP, Maity R, Baishya S (2017) Voltage and Oxide Thickness Dependent Tunneling Current Density and Tunnel Resistivity Model: Application to High-k Material HfO_2 Based MOS Devices. Superlattices Microstruct 111:628–641
12. Jelodar MS, Ilatikhameneh H, Kim S, Ng K, Klimeck G (2016) Optimum high-k oxide for the best performance of ultrascaled double-gate MOSFETs. IEEE Trans Nanotechnol 15:904–910
13. Choi J, Mao Y, Chang J (2011) Development of hafnium based high-k materials-a review. Mater Sci Eng R72:97–136
14. Chakraborty H, Maity R, Baishya S, Maity NP (2020) An accurate model for threshold voltage analysis of dual material double gate metal oxide semiconductor field effect transistor. Silicon, Online published in 9 July 2020https://doi.org/10.1007/s12633-020-00553-8
15. Basak R, Maiti B, Mallik A (2015) Analytical model of gate leakage current through bilayer oxide stack in advanced MOSFET. Supperlattices and Microstructures 80:20–31
16. Jelodar M, Ilatikhameneh H, Kim S, Ng K, Klimeck G (2016) Optimum high-k oxide for the best performance of ultrascaled double-gate MOSFETs. IEEE Trans Nanotechnol 15:904–910
17. Maity N, Maity R, Maity S, Baishya S (2019) Comparative analysis of the quantum FinFET and trigate FinFET based on modeling and simulation. J Comput Electron 18:492–499
18. Narang R, Saxena M, Gupta RS, Gupta M (2013) Impact of temperature variations on the device and circuit performance of tunnel FET: a simulation study. IEEE Trans. Nanotechnol. 12:951–957

Prediction of Diabetes Disease Using Machine Learning Model

Amandeep Sharma, Kalpna Guleria, and Nitin Goyal

Abstract As per the statistics mentioned by the world health organization, four hundred twenty-two million people in the world are suffering from diabetes which has raised the death toll to 1.6 million per year. This unprecedented growth in the number of cases and the number of casualties has led to an alarming situation because the data statistics represent a significant increase in diabetic cases among the young population, 18 years of age. Diabetes leads to various health hazards such as dysfunction of the kidney, cardiovascular problems, lower limb dismembering, and retinopathy. This article builds up a model for the prediction of diabetes using machine learning. The supervised machine learning algorithms used for prediction model such as decision tree, Naïve Bayes, artificial neural network, and logistic regression. Further, the comparison of these methods has been done based on various performance parameters such as accuracy, recall, precision, and F-score.

Keywords Machine Learning(ML) · Artificial neural network (ANN) · Logistic regression · Decision tree · Artificial intelligence (AI) · Naive bayes

1 Introduction

Diabetes is one of the common diseases nowadays. Earlier, the type 2 diabetes cases were reported primarily among the middle and old population. But in recent years, young children have also been reported as diabetic. The pancreas helps in the development of insulin in the human body. If the human body doesn't use the developed insulin efficiently and the pancreas doesn't develop essential quantities of insulin, then the human body gets diabetes [1]. Consequently, diabetes is treated as a major reason in global interest because of several health risks which may cause hyperglycemia [2]. High blood sugar is one of the major causes of diabetes, eye disease, heart attack, neuropathic ulcers [3, 4]. In [5], the authors have investigated worldwide diabetes prevalence and have found it more prominent in urban areas (10.8%)

A. Sharma · K. Guleria (✉) · N. Goyal
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
e-mail: kalpna@chitkara.edu.in

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_53

rather than rural (7.2%) areas. It has also been found that diabetes impacts more in high-income countries (10.4%) rather than countries which have low income (4.0%). Further, the authors have made a future forecast for diabetes worldwide and revealed that there would be 25% increased diabetic cases in 2030 and 51% diabetic cases in 2045. In the medical domain, it is an utmost important task to diagnose a disease at a primary stage because it will, in turn, lead to better treatment and slower progression of the disease or may prevent it further. Nowadays, machine learning plays an important role in the healthcare sector for the prediction of various diseases. So, it is extremely important to apply machine learning models for the initial detection of diabetes to improve the life of the human being. Modern technological improvement in the area of engineering and science relates to different artificial intelligence (AI) applications such as voice recognition, handwriting recognition, pattern matching, recommendation system, self-driving cars, stock market and real estate prediction, and medical healthcare. There are a large number of applications of artificial intelligence in the biocomputing field containing prediction of cardiac stroke, cancer classification, diabetic kidney disease assessment, and analytics [6–9]. In a machine learning process, the system learns from the previous experience and improves its performance [10–12]. It is a subset of AI which creates analytical models to build much statistical analytics.

2 Supervised Learning Algorithms

This section discusses various supervised learning classification algorithms which have been used for the construction of diabetic detection model. The dataset chosen for this work is Pima Indian Diabetic dataset, and it has been downloaded from UCI machine repository [13]. Below-mentioned supervised learning classification algorithms have been applied in this article:

(a) Decision Tree Algorithm
(b) Artificial Neural Network Algorithm
(c) Naive Bayes Algorithm
(d) Logistic Regression Algorithm

2.1 Decision Tree

Decision tree is an algorithm which comes under supervised learning techniques. To deal with continuous and categorical information, this method suits best [14, 15]. Decision trees do population splitting and splitting depends upon the splitter. The splitter can divide two or more than two same types of subsets. It forms a tree after completion of a task and its accuracy is majorly dependent on decisions. Classifying and applying regression on a tree follows dissimilar measures. One of the limitations of using DT is overfitting, which could be prevented by describing rules or restrictions on the building and pruning of a tree. Decision trees have numerous advantages in comparison to other supervised machine learning classification algorithms such as maximum likelihood classification. In addition, they tackle nonlinear relationships between features and classes, allow for disappeared values, and are capable of handling both categorical and numerical data inputs.

2.2 Artificial Neural Network (ANN)

The basic methodology of ANN is to function similarly to the human brain. ANN emulates the operation of the human brain. ANN architecture consists of various nodes known as artificial neurons which function similar to neurons in the human brain [16, 17]. It processes signals and then passes it further to the neurons connected to it. There exist various variants of ANN algorithms, however, the most popular ANN algorithm used is multilayer feed-forward ANN. It has input, output, and hidden layers. ANN also has limitations in finding the minima of global and the slow convergence degree. These layers are connected through many hidden nodes, and each node has a different weight. The root mean square error minimization between expected and predicted results is obtained by applying backpropagation using gradient descent optimization.

2.3 Naive Bayes

It is an approach, which uses the concept of Bayes theorem, by considering non-dependent interactive features in the designated dataset [18, 19]. The features of one class are always non-dependent on the features of another class. In the Naïve Bayes approach, all the assumptions are condition independent. This kind of algorithm could be utilized in the field of building a model where the dataset is having a huge amount of occurrences. Bayes theorem defined as below:

$$P(d/y) = (P(y/d)*P(d))/P(y)$$

.

where

P(d) signifies a class which is former for the P(x/c).

P(y/d) is used to denote likelihood probability.

P(d/y) identifies posterior probability.

P(y) denotes the probability of prior predictor.

It is also called an instant learner approach, which achieves the prediction conclusions very instantly for a class. It exhibits the best results for multiclass classification problems. In comparison with logistic regression, it outperforms as it needs less data for training. It has many applications, such as in-text identification, filtration of spam, recommender model, and sentimental analysis.

2.4 Logistic Regression (LR)

It is an approach which could be used for classification in machine learning model [20, 21]. There are many applications of LR including social studies and health sciences. It could also be used in the field of engineering, to identify and predict the probability of failure. For example, a water tank has two possible states, it could be safety or a failure, if there is an excess water supply, the binary LR will be used as the procedure to identify the relationship between safety and the predictor variable.

3 Predictive Model Construction for Diabetic Detection

The system model has been built with the help of labeled training dataset. The predictive model is required to understand the relationship between the input, output, and system variables [22]. In this paper, different supervised learning classifiers have been applied to construct a model for diabetic detection. The dataset chosen for this work is Pima Indian Diabetic dataset which has been collected from the UCI machine learning repository. This dataset contains data about females patients about 21 years of age and above. This labeled dataset has 768 instances and 9 attributes which include plasma glucose, diastolic blood pressure, triceps skinfold thickness, serum insulin, body mass index, diabetes pedigree, age, diabetes class variable. WEKA 3.8.4 [22, 23] simulator has been used to perform simulations.

Figure 1. shows different steps to create a predictive model for diabetic detection. The first step to developing any predictive machine learning model is a collection of raw data in a structured or unstructured form. After data collection, it is utmost important to perform the data cleaning to remove missing values and outliers. The next step is feature selection which selects various features from the given data set which are required and the most dominant one. Next, step is to train the model using machine learning model using a supervised learning classification algorithm. The percentage split up for training and testing used in this case is 70% and 30%, respectively. Thereafter, testing of the trained model is done using test data. The supervised learning model predicts results from its knowledge base of past experiences [24]. The performance of various classifiers is evaluated based on various performance parameters. The best classifier for this model is obtained based on accuracy value for the respective classifier, however, F-score is also one of the significant measures in this model since diabetic dataset represents an imbalanced class.

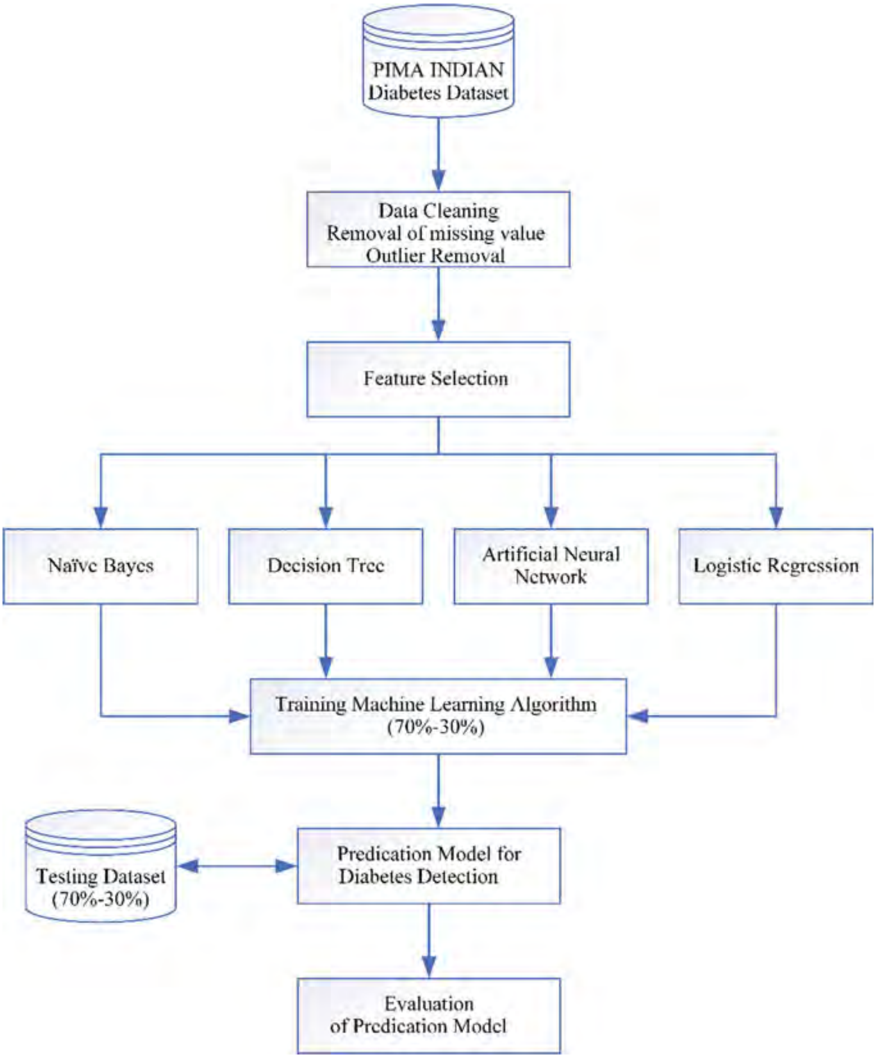


Fig. 1 Process of predictive model construction for diabetic detection

4 Results and Discussion

This section elaborates the results of prediction which has been obtained for various machine learning classifiers on this diabetic dataset.

Table 1. exhibits the simulation results of various classifiers on a diabetic dataset. Accuracy represents the number of right evaluation performed by the model. It identifies if the victim is diabetic or not. Recall (sensitivity) shows the segment of genuine diabetic patients [(tp/(tp + fn)]. Precision displays the percentage of positive diabetic

Table 1 Simulation results of various classifiers on a diabetic dataset

Classification	Naïve Bayes (%)	Decision tree (%)	ANN (%)	Logistic regression (%)
Precision	82.61	87.31	83.01	83.03
Recall	84.18	77.21	80.38	89.86
Accuracy	76.96	73.82	75.21	80.42
F-Score	83.39	81.87	81.66	86.31

cases, those are classified as diabetic or rightly identified [(tp/(tp + fp)]. It portrays the exact results or quality of the predictive model. Though, sensitivity exhibits quality or comprehensiveness of obtained result values. Large precision value shows that an algorithm gives more accurate results and higher sensitivity shows that most of the results are appropriate.

Table 2. shows the results of the Naïve Bayes algorithm where the total number of instances are 230, correctly classified instances are 177, and incorrectly classified instances are 53. The precision of this model is 82.61%, whereas recall for this model is 84.18%. F-measure for the given model is 83.39%. The overall accuracy of this model is 76.96%. The class-wise accuracy statistics about the true positive rate and false positive rate, along with F-score and ROC area has also been represented.

Table 3. represents the result of the decision tree algorithm. There are 176 instances correctly classified, and 54 instances are incorrectly classified. The precision of the given model is 87.13%, and recall for the algorithm is 77.21%. Accuracy of this algorithm is 73.82%. F-measure for the decision tree algorithm is 81.87%. Decision tree algorithms are used to predict the value of the target variable by understanding simple decision rules concluded from training data. Table 3 also presents class-wise accuracy statistics about the true positive rate and false positive rate, F-score, and ROC area.

Table 4 shows the result of ANN. In this classification algorithm, the correctly classified instances are 173 and incorrectly classified instances are 57. Precision for

Table 2 Simulation results of Naïve Bayes

Summary		
Correctly classified instances	177	76.9565%
Incorrectly classified instances	53	23.0435%
Mean absolute error	0.2677	–
Root mean squared error	0.3863	–
Total number of instances	230	–

Detailed Accuracy By Class						
TP Rate	FP Rate	Precision	Recall	F-Score	ROC Area	Class
0.842	0.389	0.826	0.842	0.834	0.845	Tested_negative
0.611	0.158	0.638	0.611	0.624	0.845	Tested_positive
0.770	0.317	0.767	0.770	0.768	0.845	Weighted average

Table 3 Simulation results of decision tree

Summary		
Correctly classified instances	176	76.5217%
Incorrectly classified instances	54	23.4783%
Mean absolute error	0.3206	–
Root mean squared error	0.4239	–
Total number of instances	230	–

Detailed Accuracy By Class						
TP Rate	FP Rate	Precision	Recall	F-Score	ROC Area	Class
0.772	0.250	0.871	0.772	0.819	0.743	tested_negative
0.750	0.228	0.600	0.750	0.667	0.743	tested_positive
0.765	0.243	0.786	0.765	0.771	0.743	Weighted average

Table 4 Simulation results of ANN

Summary		
Correctly classified instances	173	75.2174%
Incorrectly classified instances	57	24.7826%
Mean absolute error	0.3046	–
Root mean squared error	0.445	–
Total number of instances	230	–

Detailed Accuracy By Class						
TP Rate	FP Rate	Precision	Recall	F-measure	ROC Area	Class
0.804	0.361	0.830	0.804	0.817	0.772	tested_negative
0.639	0.196	0.597	0.639	0.617	0.772	tested_positive
0.752	0.309	0.757	0.752	0.754	0.772	Weighted average

this model is 83.01%, and recall (sensitivity) for this model is 80.38%. Accuracy for the ANN algorithm is 75.21%. F-score of the model is 81.66%. The main feature of ANN algorithms is a non-parametric model; other than statistical models are parametric models and require a higher background of statistics. The class-wise accuracy statistics about the true positive rate and false positive rate, along with F-score and ROC area has also been represented.

Table 5. shows the results of logistic regression algorithm applied to Pima Indian diabetic dataset. Correctly classified instances are 185, and incorrectly classified instances are 45. In this classifier, the precision is 83.03% and recall for the applied dataset is 89.86%. The overall accuracy of the classifier is 80.42%. The F-measure score is 86.31% for the logistic regression algorithm. Logistic regression is better

Table 5 Simulation results of logistic regression

Summary		
Correctly classified instances	185	80.4348%
Incorrectly classified instances	45	19.5652%
Mean absolute error	0.2987	–
Root mean squared error	0.3748	–
Total number of instances	230	–

Detailed Accuracy By Class

TP Rate	FP Rate	Precision	Recall	F-measure	ROC Area	Class
0.899	0.403	0.830	0.899	0.863	0.848	tested_negative
0.597	0.101	0.729	0.597	0.656	0.848	tested_positive
0.804	0.308	0.799	0.804	0.799	0.848	Weighted average

than a decision tree because a single line is fit to divide space into two when the data is scattered in such a way so that it can be linearly classified. Table 3 also presents class-wise accuracy statistics about the true positive rate and false positive rate, F-score, and ROC area.

5 Conclusion

This research paper presents the prediction of a diabetic using machine learning models. supervised learning algorithms such as logistic regression, Naïve Bayes, artificial neural network, decision tree has been used to create the analytics models for finding whether the patient is diabetic or not. Accuracy represents the perfection of an algorithm. The prediction model exhibits that the logistic regression displays 80.43% accuracy which is highest among all. Naïve Bayes algorithm and decision tree display very competitive results. The accuracy of the Naïve Bayes algorithm is 76.95%, and decision tree algorithm has an accuracy of 76.52%; so, the final results of both classifiers are very close to each other. Artificial neural network classifier has 75.21% accuracy, which is the lowest among others. Further, rather than accuracy, F-score is also another effective measure to evaluate the prediction model. F-measure value can be represented between 0 to 1. If the F-measure value of any classifier is close to 1 means that the classifier model represents better performance. Logistic regression classifier represents 0.863 F-measure, which is highest among other classifiers and F-measure for the decision tree classifier is 0.817 lowest among other models. F- Measure for Naïve Bayes and ANN classifier is 0.834 and 0.819, respectively. Therefore, it is concluded that for this diabetic dataset, logistic regression represents the highest accuracy and F-score to create an analytical model for diabetes detection among other machine learning algorithms.

References

1. Kawada T (2020) Total dietary antioxidant capacity and risk of type 2 diabetes. Eur J Epidemiol 1–2. https://doi.org/10.1007/s10654-020-00608-5
2. Mamykina L, Heitkemper EM, Smaldone AM, Kukafka R, Cole-Lewis HJ, Davidson PG, Hripcsak G (2017) Personal discovery in diabetes self-management: discovering cause and effect using self-monitoring data. J Biomed Inform 76:1–8
3. Papatheodorou K, Banach M, Edmonds M, Papanas N, Papazoglou D (2015) Complications of diabetes.
4. Soumya D, Srilatha B (2011) Late stage complications of diabetes and insulin resistance. J Diabetes Metab 2(9):1000167
5. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, Shaw JE (2019) Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas. Diabetes Res Clin Pract 157:107843
6. Sapra L, Sandhu JK, Goyal N (2020) Intelligent method for detection of coronary artery disease with ensemble approach. Adv Commun Comput Technol (pp 1033–1042). Springer, Singapore
7. Guleria K, Sharma A, Lilhore UK, Prasad D (2020) Breast Cancer prediction and classification using supervised learning techniques. J Comput Theor Nanosci 17(6):2519–2522
8. Lilhore UK, Simaiya S, Prasad D, Guleria K (2020) A Hybrid Tumour detection and classification based on machine learning. J Comput Theor Nanosci 17(6):2539–2544
9. Babič F, Majnarić L, Lukáčová A, Paralič J, Holzinger A (2014) On patient's characteristics extraction for metabolic syndrome diagnosis: predictive modelling based on machine learning. Int Conf Inf Technol Bio Med Inf (pp 118–132) Springer, Cham
10. Chlingaryan A, Sukkarieh S, Whelan B (2018) Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. Comput Electron Agric 151:61–69
11. Goyal N, Sandhu JK, Verma L (2019) Machine learning based data agglomeration in underwater wireless sensor networks. Int J Manage Technol Eng 9(6):240–245
12. Lilhore UK, Simaiya S, Guleria K, Prasad D (2020) An efficient load balancing method by using Machine Learning-based VM distribution and dynamic resource mapping. J Comput Theor Nanosci 17(6):2545–2551
13. Bay SD, Kibler D, Pazzani MJ, Smyth P (2000) The UCI KDD archive of large data sets for data mining research and experimentation. ACM SIGKDD Explorations Newsl 2(2):81–85
14. Ming H, Wenying N, Xu L (2009) An improved decision tree classification algorithm based on ID3 and the application in score analysis. In 2009 Chinese control and decision conference (pp 1876–1879) IEEE
15. Dong X, Qian M, Jiang R (2020) Packet classification based on the decision tree with information entropy. the Journal of Supercomputing 76(6):4117–4131
16. Momeni E, Nazir R, Armaghani DJ, Maizir H (2014) Prediction of pile bearing capacity using a hybrid genetic algorithm-based ANN. Measurement 57:122–131
17. Koopialipoor M, Fahimifar A, Ghaleini EN, Momenzadeh M, Armaghani DJ (2020) Development of a new hybrid ANN for solving a geotechnical problem related to tunnel boring machine performance. Eng Comput 36(1):345–357
18. Loor M, De Tré G (2020) Contextualizing Naive Bayes predictions. In International conference on information processing and management of uncertainty in knowledge-based systems (pp 814–827). Springer, Cham
19. Zhang H, Jiang L, Yu L (2020) Class-specific attribute value weighting for naive bayes. Inf Sci 508:260–274
20. Yang Y, Chen G, Reniers G (2020) Vulnerability assessment of atmospheric storage tanks to floods based on logistic regression. Reliab Eng Syst Saf 196:106721
21. Shah K, Patel H, Sanghvi D, Shah M (2020) A comparative analysis of logistic regression, random Forest and KNN models for the text classification. Augmented Human Research 5(1):1–16

22. Tiwari S, Kumar S, Guleria K (2020) Outbreak trends of Coronavirus Disease–2019 in India: a prediction. Disaster medicine and public health preparedness, pp 1–6
23. Frank E, Hall M, Holmes G, Kirkby R, Pfahringer B, Witten IH, Trigg L (2009) Weka-a machine learning workbench for data mining. In Data mining and knowledge discovery handbook (pp 1269–1277). Springer, Boston, MA
24. Singh A, Thakur N, Sharma A (2016) A review of supervised machine learning algorithms. In 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom) (pp 1310–1315) IEEE

Infrared Small Target Detection Based on Phase Fourier Spectrum Transform

Sur Singh Rawat, Sashi Kant Verma, and Yatindra Kumar

Abstract Recently, target detection task in infrared imaging is becoming very difficult when the high detection rate as well as low error rate is concerned. As the image background is noisy and cluttered, present methods face issues with detection performance and low error rate. To improve the detection ability of the existing methods, human visual saliency based method is presented in this letter. Firstly, it detects the salient region in the infrared image, which may contain the targets with the help of phase spectrum Fourier transform (PET). Then, saliency map is obtained and at the end simple threshold operation is performed to get the candidate target. Lot of experiments were conducted and the results show that the method presented here can lower the background noise effectively and also predicts the object efficiently with reduced error rate.

Keywords Infrared small target detection · Saliency · Phase spectrum · Fourier transform · Human visual system

1 Introduction

Infrared technology finds a large number of applications in defense and military system. Target detection is found of the well-known task in many areas like space system, warning systems as well as in missile system. The target which is small is normally get remained in a noisy background along with less signal to noise ratio and there is always a long imaging distance and atmospheric transmission.

S. S. Rawat (✉)
Department of Computer Science and Engineering, JSS Academy of Technical Education, Noida, U.P., India
e-mail: sur.rawat@jssaten.ac.in

S. K. Verma
Department of Computer Science and Engineering, GBPIET, Pauri garhwal, Uttrakhand, India
e-mail: skverma.gbpec@rediffmail.com

Y. Kumar
Department of Electrical Engineering, GBPIET, Pauri garhwal, Uttrakhand, India
e-mail: kyatindra@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_54

Although many research work is done in this direction in the past decades but the above said circumstances make the target detection task still a hard problem [1–3]. To tackle with the above said issues, many methods have been addressed in the past. These methods are basically broken into two folds: Single frame methods and multiple frames sequentially methods. The second methods make full use of spatial and temporal features acquired by imaging sensor, but, as the detected target or the sensor system is of high speed in many applications, so this produces some kind of effects due to the effective difference in the image with the background and the target. As a consequence of this, the sequential method does not work well. As the detection of target required detection performance to be high with reduced response time, then a single frame method is a better choice. Gao et al.tried a method to detection of target issue and come-up with a method, patch Image-model (IPI) [1].This method has utilized, non- local self-correlation features of background. Also, expect the background patches are normally belongs to the combination or the single low-rank subspace clusters. The IPI model has the l_1-norm sparsity problem as well as the constant weighting parameter issue which estimates the background inaccurately and this has produced a miss classified target object. Dai et al. [4] Gave a method which was on structural properties of the images in background. This approach although has shown better result when compared with the other approaches, but here in this method, an extra load was created in computation of weights of the columns. Dai et al. [5] presented a non-negative IPI method based on partial sum of minimization of singular values (PSMSV) (NIPPS) so as to approximate the background and retained the singular values. NIPPS has a problem of choosing a proper energy constraint ratio as well as the rank of the matrix. Gao et al. [6], Reweighted IPI (ReWIPI) model was based on work discussed in [7] to adjust the patch image of background and also retain the background edge features. Problem with this approach was the improper tuning of weighting parameter which could affect the calculation of singular value decomposition (SVD) of matrix. To, further improve the IPI based approaches, Non-convex rank approximation minimization (NRAM) [8] was presented. Also, to bring the inner smoothness to the background, a term for regularization called total variation was applied to the method namely: regularization of TV and induction of principal component pursuit (TV-PCP) [9]

Human visual system (HVS) [10] was presented recently to target detection system in infrared imaging, which think that the target is one of the salient information. If we consider methods based on saliency, Laplacian of Gaussian (LOG) [11, 12],Difference of Gaussian (DOG) [13],second-order directional derivative (SODD) [14],contrast measure in local area (LCM) [15],patch-based method on contrast measure and multi-patch image(MPCM) [16], method on difference of Gabor's an improved version (IDOGb) [17],difference measure in local area(LDM) [18],weighted local difference measure(WLDM) [19], spectral residual [20], phase spectrum [21] are the most popular one.

The remaining paper is presented as follows: Section 2 here includes the description the related work, Sect. 3, we describe the process target detection for small objects. In Sect. 4, include the result of method presented and its comparison with the other methods. Section 5 presents the concluding of this paper.

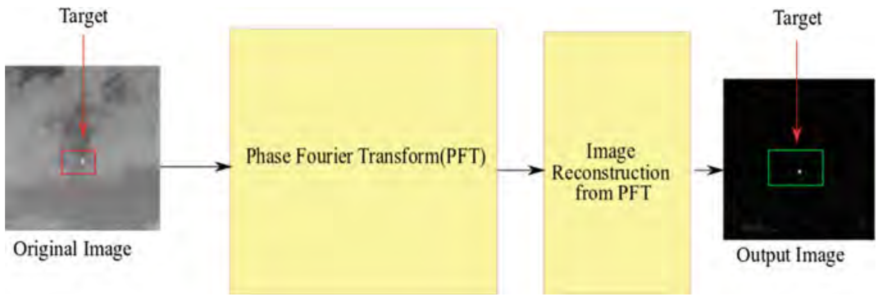


Fig. 1 Proposed method for small target detection

2 Methodology: Method Based on Human Visual System

2.1 *Brief Description of a Human Visual System*

(HVS) basically break the input image into the different small patches and select the salient features through HSV system to do the easy understanding and analyze. We can observe in the HVS system that, important features in the image can be represented not by the amplitude of the visual signal, but it is actually the contrast information with the amplitude at a given point and at its vicinity location. From this, it is concluded that contrast is the valuable feature which is presented in streams of our visual system. Based on the above discussion we proposed a framework for target detection system as shown in (Fig. 1). HVS process will help to separate the target from background noise. Since the background noise are lke infrared target shape and size, so it is very difficult to distinguish them using the traditional methods.

2.2 *Theory of Phase Fourier Transform*

The phase Fourier transform (PFT) [20]of an image signal gives the phase information in an image and tells the location of an object present in an image. Phase spectrum information in an infrared image will detect the target as it the most salient information.

Following steps are needed to calculate the phase Fourier transform of an image. Let $s(x,\mathrm{y})$ is an image

$$f(x, y) = F(s(x, y)) \tag{1}$$

$$p(x, y) = P(f(x, y)) \tag{2}$$

$$SM(x, y) = g(x, y) * \|F^{-1}\big[e^{i.p(x,y)}\big]\|^2 \tag{3}$$

Where F and F^{-1} are the Fourier and its inverse Fourier transform, $P(f(.))$ is the phase spectrum of an image and $g(x, y)$ is the 2-D Gaussian filter at ($\sigma = 8$). $SM(x, y)$ is the saliency map of an image.

2.3 Steps of the Detection Method for Small Object

PFT is used to locate the salient object in an image so in our problem small object is the salient object that we need to detect. Following are the steps to perform in the proposed method.

1. PFT of the input infrared image is computed.
2. Saliency map is obtained after inverse PFT.

Finally the simple threshold operation is adopted to get the object more accurately.

3 Experimental Result and Analysis

We have prepared a dataset of more than 450 images under infrared imaging system of a varying environment such as, sea, sky, cloud and ground. Description of these images sequences are given in the Table 1. All experiments were implemented on MATLAB 2015 software on PC with 4-GB RAM and 2.20 GHz Intel Core 2 Duo processor.

3.1 Suppression Result of Image Sequences for Background

In this sub-section we have presented the background suppression result of the method presented of six different image sequences under the complex background. Figure 2a represents the original infrared images, Fig. 2b shows the background suppression result of our method. The location of target in an image can be seen by utilizing 3D gray map and finally, to show the background suppression result of our proposed method as shown in the Fig. 2c. The proposed approach is also compared to the baseline method shows the high background suppression ability and the high signal to clutter ratio gain. Figure 3 shows the effectiveness of our method with the other baseline method. As shown in the Fig. 3b, c the Max-mean and the max-median methods can predict the object well but in the strong clutter background these methods fail to detect the target effectively. Similarly Top- hat method as shown in Fig. 3d need to adjust the size of the filter to detect the target. Again in the strong clutter background it sometimes fail to detect the target. The

Table 1 Detail description of representative images

Real infrared image sequences	Target type	No of frames	Image size	Features of background	Features of target
Sequence no. 1	A small ship	30	256 × 200	Sea-sky with blurred	Changing target size • Target in small size • Imaging distance is long
Sequence no.2	An airplane	250	256 × 200	Heavy cloudy background containing low local contrast	Changing target size • Target in small size • Imaging distance is long
Sequence no. 3	An airplane	250	256 × 200	With changing background	Changing target size Target in small size Imaging distance is long Low SCR
Sequence no. 4	A Helicopter	100	241 × 200	Changing background	Imaging distance is long • Low SCR
Sequence no. 5	A ship	100	320 × 240	Changing background	• A one to two target • Small in size
Sequence no. 6	A person, moving back and forth in forest	250	280 × 228	Background containing heavy clouds	Changing target size, • Imaging distance is long

method presented compute the phase Fourier transform of the image and detect the location of a salient information. As infrared small target is the salient object in the image so its location can be detected easily by the method presented. Figure 3e shows the detection result.

3.2 *Evaluation Indicators*

In order to judge our method, the two important classical evaluation metrics namely: signal to clutter ratio gain (SCRG) and background suppression factor (BSF) are

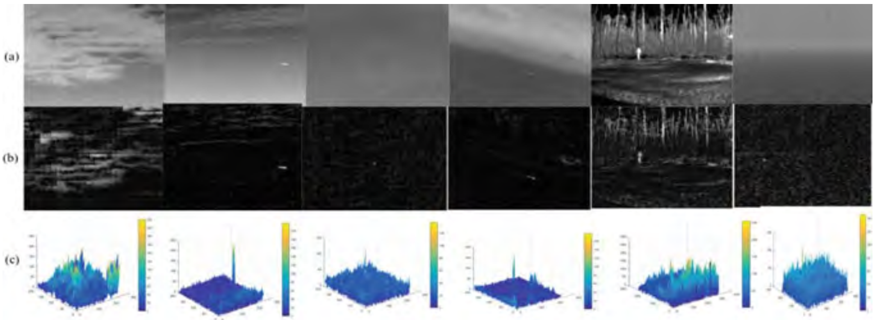


Fig. 2 3-D map of original infrared images and the result of the presented method. **a** Original representative images **b** Detection result of **a**, **c** 3-D map of proposed method respectably

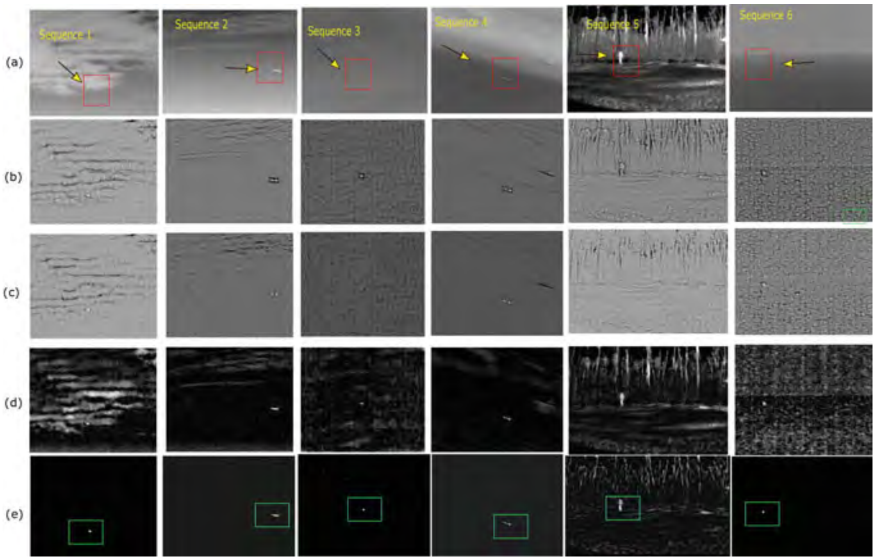


Fig. 3 Result of various methods **a** Input image **b** Max-mean **c** Max-median **d** Top-hat and **e** result of proposed method

utilized to validate the performance of the method. The detail of these metrics is given in [22] and can be written as follows:

$$\mathrm{SCRG} = \frac{\left({}^{S}\!/\!{}_{C}\right)_{\mathrm{out}}}{\left({}^{S}\!/\!{}_{C}\right)_{\mathrm{in}}}, \quad \mathrm{BSF} = \frac{C_{\mathrm{in}}}{C_{\mathrm{out}}} \tag{4}$$

Here, signal of the amplitude and the clutter standard deviation is presented by S and C, *in and out* in the expression are input real image and the output image with target. The signal to clutter ratio gain (SCRG) tells the amplification result of the signal before and after the image is processed. Background suppression factor (BSF) given the level of suppression when no information of target is available. Hence it is expected that, for better efficiency both the indicators should have a large value.

3.3 Experimental Result Comparison with the Other Methods

The method presented here in this paper was compared with the **three (03)** other base methods to evaluate its robustness. Figure 3a depicts the original image sequences. The background suppression results of all the base methods like top-hat, max-mean and max-median along with the proposed method are presented in the Fig. 3b–e . We can observe that due to fractional directional derivatives and the phase Fourier transform the presented method lower down the background and improves the target better than the other base method where a simple filtering approach is performed. We can observe from Tables 2 and 3 which represents the signal to clutter ratio gain (SCRG) and the background suppression factor (BSF) respectably, that the method presented, has got better SCRG and BSF result in comparison with the other base methods. Similarly, as the time complexity is also an important parameter in an algorithm. The proposed method has computational efficiency of **0.051 s** which is better than the baseline methods as can be observe from Table 4.

Table 2 SCRG values of the presented method

Methods	Image sequences					
	A	B	C	D	E	F
Max-mean	1.16	5.73	7.38	12.74	1.32	12.13
Max-median	1.06	1.71	5.41	1.59	0.26	6.75
Top-hat	0.99	1.58	6.46	13.06	0.09	6.92
Proposed method	2.50	10.30	15.85	35.26	4.16	25.50

Table 3 Values of BSF of the presented method

Methods	Image sequences					
	A	B	C	D	E	F
Max-mean	0.49	2.34	0.51	3.93	1.34	0.92
Max-median	1.29	3.90	0.75	7.25	2.12	1.20
Top-hat	1.38	3.39	0.86	10.11	2.43	1.26
Proposed method	2.88	3.96	0.89	14.04	3.82	2.12

Table 4 Computation time of the method presented and base method

Methods	Times(s)
M ax-mean	0.841
Max- median	0.914
Top -hat	0.715
Proposed method	0.0583

4 Conclusion

An approach for target using the infrared images, which is motivated by the human visual system is presented in this paper. The phase Fourier transform is first used to obtain a saliency map of the candidate target and then a simple threshold approach is applied to get the resultant target. The experimental result shows that the method will minimize the noise from the background properly and also detect the candidate object effectively with the lesss error rate.

Acknowledgements The authors would like to thank all the valuable reviewers

References

1. Gao C, Meng D, Yang Y, Wang Y, Zhou X, Hauptmann AG (2013) Infrared patch-image model for small target detection in a single image. IEEE Trans Image Process 22(12):4996–5009
2. Zhao J, Tang Z, Yang J, Liu E (2011) Infrared small target detection using sparse representation. Journal of Systems Engineering and Electronics 22(6):897–904
3. Rawat SS, Verma SK, Kumar Y (2020) Review on recent development in infrared small target detection algorithms. Gurgaon, India
4. Dai Y, Wu Y, Song Y (2016) Infrared small target and background separation via column-wise weighted robust principal component analysis. Infrared Phys Technol 77:421–430
5. Dai Y, Wu Y, Song Y, Gao J (2017) Non-negative infrared patch-image model: Robust target-background separation via partial sum minimization of singular values. {Infrared Physics & Technology, vol 81, pp 182--194
6. Guo J, Wu Y, Dai Y (2017) Small target detection based on reweighted infrared patch-image model. IET Image Proc 12(1):70–79
7. Gu S, Xie Q, Meng D, Zuo W, Feng X, Zhang L (2017) Weighted nuclear norm minimization and its applications to low level vision. Int J Comput Vision 121(2):183–208
8. Zhang L, Peng L, Zhang T, Cao S, Peng Z (2018) Infrared small target detection via non-convex rank approximation minimization joint l2, 1 norm. Remote Sensing 10(11):1–34
9. Wang X, Zhenming P, Dehui K, Zhang P, He Y (2017) Infrared dim target detection based on total variation regularization and principal component pursuit. Image Vision Comput 63: 1--9
10. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20(11):1254–1259
11. Kim S (2011) Min-local-LoG filter for detecting small targets in cluttered background. Electron Lett 47(2):105–106
12. Kim S, Yang Y, Lee J, Park Y (2009) Small target detection utilizing robust methods of the human visual system for IRST 30(9):994--1011

13. Wang X, Lv G, Xu L (2012) Infrared dim target detection based on visual attention. Infrared Physics & Technology}, vol. 55, no. 6, pp. 513--521, 2012.
14. Qi S, Ma J, Tao C, Yang C, Tian J (2013) A robust directional saliency-based method for infrared small-target detection under various complex backgrounds. IEEE Geosci Remote Sens Lett 10(3):495–499
15. Chen CP, Li H, Wei Y, Xia T, Tang YY (2014) A local contrast method for small infrared target detection. IEEE Trans Geosci Remote Sens 52(1):574–581
16. Wei Y, You X, Li H (2016) Multiscale patch-based contrast measure for small infrared target detection. Pattern Recogn 58:216–226
17. Han J, Ma Y, Huang J, Mei X, Ma J (2016) An infrared small target detecting algorithm based on human visual system. IEEE Geosci Remote Sens Lett 13(3):452–456
18. Deng H, Sun X, Liu M, Ye C, Zhou X (2017) Entropy-based window selection for detecting dim and small infrared targets. Pattern Recognition 61:66--77
19. Deng H, Sun X, Liu M, Ye C, Zhou X (2016) Small infrared target detection based on weighted local difference measure. IEEE Trans Geosci Remote Sens 54(7):4204–4214
20. Hou X, Zhang L (2007) Saliency detection: a spectral residual approach. In: IEEE conference on computer vision and pattern recognition IEEE, pp 1--8
21. Guo C, Ma Q, Zhang L (2008) Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In: 2008 IEEE conference on computer vision and pattern recognition, IEEE, pp 1--8.
22. Hilliard CI (2000) Selection of a clutter rejection algorithm for real-time target detection from an airborne platform. In: Proceedings SPIE, vol 4048, pp 74--84

Knowledge Based Analytical Tool for Identifying Children with Dyscalculia

A. Devi and G. Kavya

Abstract Specific Learning Disability is a learning difficulty that has an impact on reading (Dyslexia), mathematical calculations (Dyscalculia), and drawing skill (Dysgraphia), among children. Out of these learning disabilities, the mathematical disability is a major problem for children to do minor calculations like counting numbers, remembering phone numbers and problem in understanding speed, time, distance etc., which also affects their regular academic activities. Dyscalculia is a condition that lacks numerical skill needed for an individual that arises due to heredity or sometimes the brain development itself. Mostly children's who are affected with dyscalculia seem to be good in other learning areas. Due to the lack of identification and assessment, these children severely affected when they are entering into higher grades. This may weaken their mental health condition and make them feel left out and low in confidence. Hence a knowledge-based tool is proposed to identify mathematical difficulties using machine learning decision tree algorithm. The entire proposed model is divided into two sections (i) Knowledge-based tool creation for dataset collection (ii) Decision tree classification. The main aim of this research work is to diagnose and detect the dyscalculia problem in the children at the earlier stage itself.

Keywords Decision tree (DT) · Machine learning (ML) · Random forest (RF) · Content management system (CMS)

1 Introduction

Learning Disability (LD) is a neurodevelopmental disability that affects brain functioning capacity to process the listening information, observing information, etc. [1].

A. Devi (✉)
Department of ECE, IFET College of Engineering, Villupuram, Tamilnadu, India
e-mail: deviarumugam02@gmail.com

G. Kavya
Department of ECE, SA Engineering College, Chennai, Tamilnadu, India
e-mail: kavyavimal@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_55

The learning disability maybe appears in a reading, writing and mathematical skills, hence it is called Specific Learning Disability (SLD). Around worldwide, 10 percent of the students are affected by SLD. Statistics in the west and India show that about 13 to 14% of the children in a regular classroom have SLD. Specifically, around 6 percent of school-going children have math deficits. Such mathematical deficits were common and required equal attention and concern [2]. Effect of math failure with mathematical analphabetism seriously impedes both everyday life and studies. Many factors cause dyscalculia problem in the children. And also several assessment tests are available to screen the children with dyscalculia. Each assessment test focus on the different skills of the children. The relevant assessments can be assessed to [3]

(a) Computation skills—E.g. Woodcock-Johnson IV (WJIV),
(b) Math fluency—E.g. WJ IV Math Fluency, Wechsler Individual Achievement Test (WIAT-III) Math Fluency subtest, Mathematical Fluency and Calculations Tests (MFaCTs)
(c) Mental computation—E.g. Wechsler Intelligence Scale for Children (WISC-V), Paced Auditory Serial Addition Test (PASAT), and Test of Mental Computation
(d) Quantitative reasoning—WIAT-III Math Problem Solving subtest, WJ IV Applied Problems, CMAT Problem Solving.

These assessment tests can help to determine the dyscalculia issue of the children manually. At present Machine Learning is [4] widely used in many fields to predict future results. One of the most important applications of machine learning is to predict the learning disorder problem in children, identify the exact disability and recognise the disability problem at the earlier stage. The fundamental concept behind machine learning is to develop an algorithm that can accept input data and use statistical models to predict an exact output. Classification in machine learning consists of two steps such as learning and the prediction. In the learning step [5], the model is built based on provided training data. The model is used in the prediction process to predict the response to the data provided. For understanding and analysing the datasets, the commonly used classification algorithm is the decision tree algorithm, where the decision tree belongs to supervised learning algorithms. The main objective of the proposed tool is to accurately identify the children with dyscalculia problem at the earlier stage using machine learning Decision Tree algorithm.

2 Related Work

Let us discuss some of the studies carried out to identify the children with dyscalculia problem. And also screening tools developed for diagnosing and identifying the dyscalculia problem is discussed.

3 Proposed Knowledge-Based Tool

E-platform for dataset generation is created for identifying children with dyscalculia, where it consists of many questions which will be plugged in word press Content Management System (CMS) software platform by using the plugin method. The word press is an open-source platform [11], so that website can be customized, updated and managed by the user. Steps for the plugin is given below.

Step 1: Download Word press and log in to word press by using username and password which given during installation.

Step 2: Add details into general settings such as site title, website, email address, time format, date format etc.

Step 3: Install and configure a suitable plugin. In the proposed model, the Quiz Maker plugin and the Audio plugin is configured.

Once the plugin is installed, the questionnaire is set to meet the WJ IV test. The tests are conducted in quiz type (Fig. 1).

The questionnaire is framed by considering the age, computation time, math calculation, math fluency, applied problems, quantitative concepts and number series as parameters according to WJ IV. Some of the skills set questions used for identification of dyscalculia children are.

(a) Math Calculation—Basic arithmetic calculation such as addition, subtraction, multiplication and division.
(b) Math Fluency—Basic arithmetic calculation (Double and triple digits) within the stipulated timing.
(c) Applied problem—Analysis of questions to perform a suitable arithmetic operation.
(d) Quantitate concepts and Number series—Measures the knowledge about math symbols, series of numbers, identification of missing numbers.

Some of the sample math calculation, number comparison and place value questions for the screening of mathematical disability is shown in the Fig. 2a–c (Fig. 3).

4 Decision Tree Classification Methodology

Nearly 50 children's who are in the age of 8 and 9 years participated in the quiz. All the children were allowed to take the quiz in allotted timing. By using the proposed analytical tool, 50 datasets were generated. The datasets were formed by considering the parameters such as average time taken for each question and the aggregated score of each section, where the different sections are math calculation, math fluency, applied problem, Quantitate concepts and Number series. The block diagram of the proposed methodology is shown in Fig. 4.

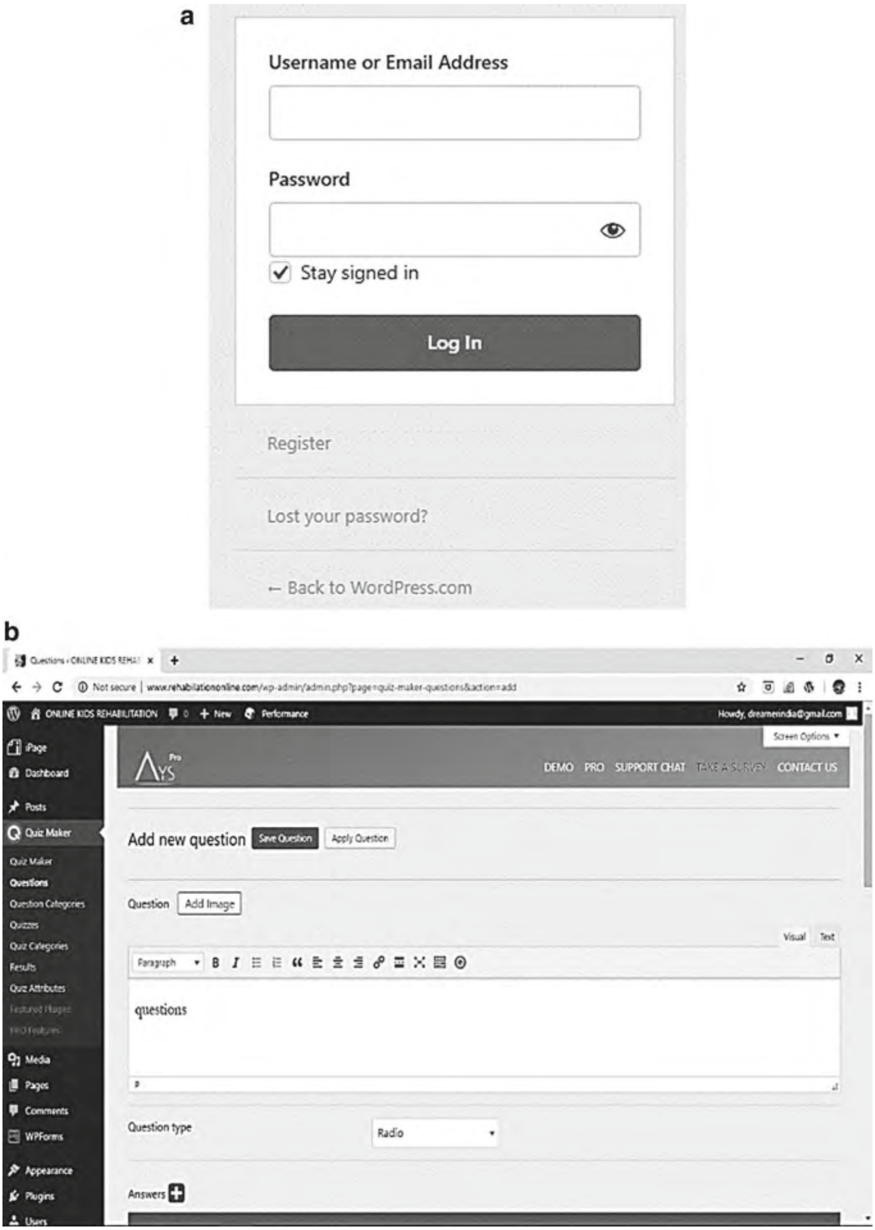


Fig. 1 **a** Login Page **b** Page for adding a questionnaire

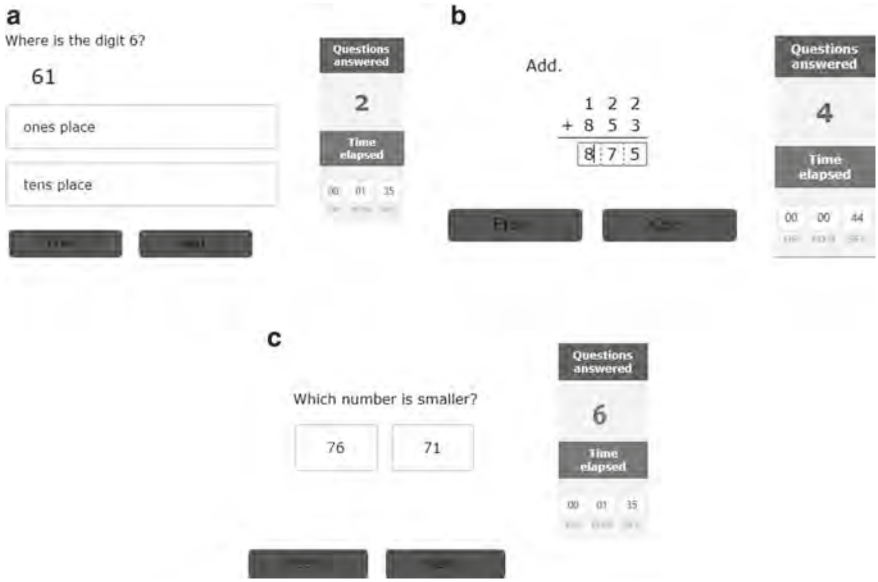


Fig. 2 **a** Place value questions **b** Three-digit addition **c** Comparison question

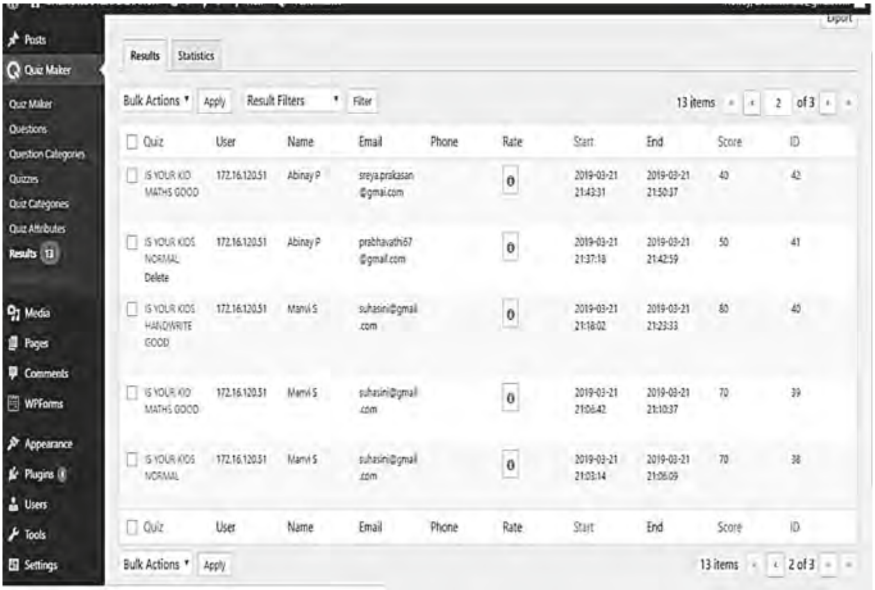


Fig. 3 Test results in word press

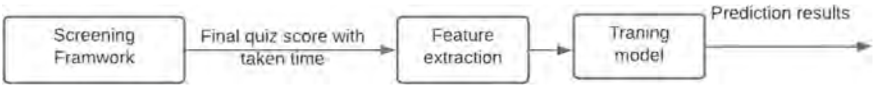


Fig. 4 Block diagram of proposed methodology

In the proposed model, the identification of children with dyscalculia is performed by decision tree classification. The decision is one of the simple and most familiar machine learning algorithms. The decision tree works accurately when the dataset is lesser in size. To test only important parameters (parameters which makes the most difference to classification) first, the decision tree is adopted. The data is filtered and sorted as the datasets are processed, and then datasets are used to identify the children with dyscalculia. The first step is to split the dataset into training and testing, the next step is to perform feature scaling. Then the model is fit in the decision tree. To predict a class label for an object, the decision tree algorithm processes datasets from the root of the tree. The root attribute values are compared with record attributes. Based on the similarity, the branch corresponds to the value and moved to the next node. The random forest [12] is introduced to eliminate the issue of overfitting problem in the decision tree. This method of combining the output of multiple individual models is called Ensemble Learning, where the random forest is an example of ensemble learning. The random forest allows combining several machine learning algorithms to provide better predictive results.

5 Results

The training dataset contains the data set on which the model can be trained, while the testing dataset involves data on which the testing model is applied to assess its accuracy. In the proposed framework, the data to be used as input is the questions set based on the Woodcock Johnson IV test. After training the model, the dyscalculia prediction accuracy is 95.3%. The accuracy plays a significant role as the result produced depends completely on this method. This makes the proposed tool to be reliable and accurately generates results without any manual work (Fig. 5).

6 Conclusion and Future Enhancement

The proposed knowledge-based tool efficiently and accurately generates the dataset to analyse and predict the children with dyscalculia problem. The proposed method is focussed to analyse the generated datasets effectively using decision tree machine learning algorithm. The proposed analytical tool is very interactive and identifies the children with dyscalculia in an efficient manner. The proposed tool will be useful for school teachers to avoid complex manual work carried out, and to identify the

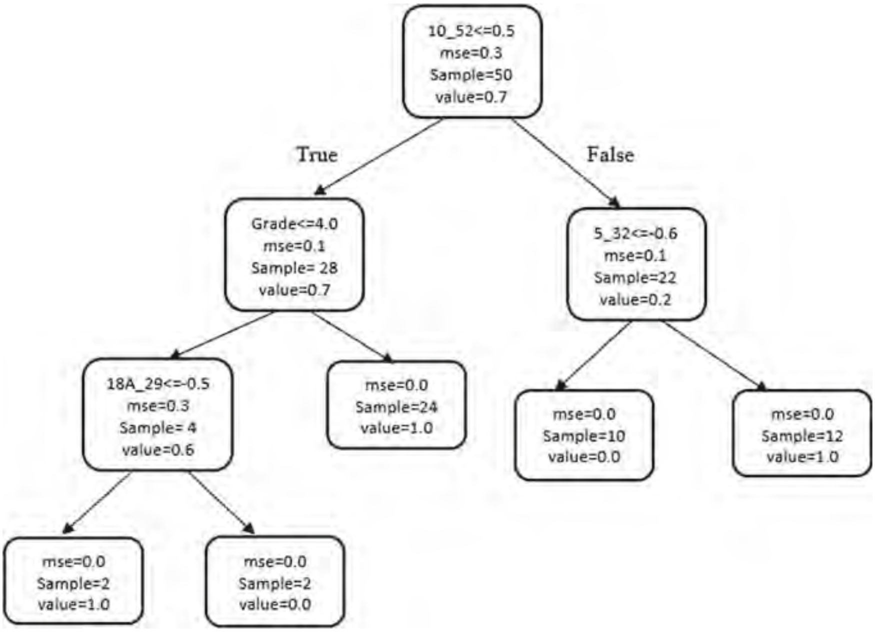


Fig. 5 Structure of Decision Tree based on the prediction result

children with dyscalculia at an early stage. In future, many parameters also can be included to identify the children with dyscalculia problem. The proposed tool can also be used to identify dyslexia and dysgraphia problems by considering the corresponding evaluation parameters (Table 1).

Table 1 Comparative Study based on different Tools

Name of the Author	Study conducted/ Tool	Results	Drawback
Mohd Syah [6]	Computerized play tool for dyscalculia intervention	Total post-test scores found that 57.9 percent of children in the intervention group did slightly higher than the control group during the five-day intervention period	Play model does not support local languages
Rani [7]	Web-based tool	Identify the level of learning capability of mathematics and allow the children to use appropriate the application	Does not suits for all mobile applications
Annemie [8]	Investigated some of the outcomes of studies for predicting the dyscalculia problem at the young children based on the behavioral signs	Dyscalculia has a strong heritability and that the likelihood of dyscalculia is enhanced by prematurity and very low birth weight. Additionally, in kindergarten, dyscalculia is frequently but not always accompanied by language deficits and impaired (procedural and conceptual) knowledge of counting, seriation, grouping, and quantity estimation	Questionnaire-based study
Dazhi Cheng at el [9]	İnvestigated whether deficits in cognitive visual perception were common to both dyslexia and dyscalculia	The findings revealed that both of dyslexia, dyscalculia and comorbid dyslexia with dyscalculia is marked by deficiencies in the understanding of numerosity and visual vision	Difficult to predict

(continued)

Table 1 (continued)

Name of the Author	Study conducted/ Tool	Results	Drawback
Rikard Ostergen [10]	Conducted three studies for examining the connection between cognitive abilities and arithmetic	The results suggested that the number of sensory deficits with memory functions together constitute risks for children with a mathematical learning disability	Suggesting some other basic development model based on von Asters and Shalev's model

References

1. David JM, Balakrishnan K (2009) Prediction of frequent signs of learning disabilities in school age Children using association rules, Proc Int Conf Adv Comput ICAC 2009, MacMillion Publishers Ltd., New York City, pp 202–207
2. https://www.epi.org/publication/bp278/
3. Handbook of Psychological Assessment [4 ed.] 0128022035, 9780128022030.
4. Gupta S, Kaur M, Lakra S, Dixit Y (2020) A comparative theoretical and empirical analysis of machine learning algorithms. Webology 17(1):377–397
5. Julie M, Balakrishnan K (2010) Significance of Classification techniques in prediction of learning disabilities. Int J Artif Intell Appl 1(4):111–120
6. Mohd Syah NE, Hamzaid NA, Murphy BP, Lim E (2016) Development of computer play pedagogy intervention for children with low conceptual understanding in basic mathematics operation using the dyscalculia feature approach. Interactive Learning Environ 24(7):1477–1496
7. Rani MFCA, Rohizan R, Rahman NAA (2014) Web-based learning tool for primary school student with dyscalculia. In: Proceedings of the 6th international conference on information technology and multimedia (pp 157–162) IEEE
8. Desoete A, Baten E (2017) Indicators for a specific learning disorder in mathematics or dyscalculia in toddlers and in kindergarten children. Belgian J Paedicatrics 19(2):122–124
9. Cheng D, Xiao Q, Chen Q, Cui J, Zhou X (2018) Dyslexia and dyscalculia are characterized by common visual perception deficits. Dev Neuropsychol. 43(6):497–507. https://doi.org/10.1080/87565641.2018.1481068 PMID: 29975105
10. Ostergen R (2013) Mathematics learning difficulties, cognitive conditions, development and predictions. Department of behavioral sciences and learning, ISSN 0282-9800
11. Devi A, Kavya G (2019) Intelligent system for identifying Dyscalculia based on Raspberry pi. Int Conf Commun Electron Syst (ICCES)
12. Random Forest Analysis in ML and when to use it - NewGenApps, NewGenApps (2020)

RSA Algorithm Using Performance Analysis of Steganography Techniques in Network Security

B. Barani Sundaram, N. Kannaiya Raja, Nune Sreenivas, Manish Kumar Mishra, Balachandra Pattanaik, and P. Karthika

Abstract The stable network has now become a prerequisite for every relationship. The security hazards are expanding and rendering wired/remote organizations and Internet providers fast, shaky and problematic. Today, safety efforts are all the more important in meeting the cutting edge demands of current emerging businesses. The need is also discussed in areas such as guards, where the key points of concern are safe and checked admission of properties identified with data security. In this paper, the author described the important measures and limits for the establishment of a secure organization with regard to enormous industry / authoritative needs. In providing remote organization access to different assets and interfacing various gadgets remotely, Wi-Fi networks are extremely basic. To deal with Wi-Fi hazards and organizational hacking attempts, different prerequisites are required. This article

B. B. Sundaram (✉)
ICT-CE, Computer Science Department, College of Informatics, Bule Hora University, Bule Hora, Ethiopia
e-mail: bsundar2@gmail.com

N. Kannaiya Raja
College of Informatics, Bule Hora University, Bule hora, Ethiopia
e-mail: kannaiyaraju123@gmail.com

N. Sreenivas
School of Electrical and Computer Engineering, Addis Ababa Institute of Technology, Addis Ababa University, Addis Ababa, Ethiopia
e-mail: ns_maruthi@yahoo.com

M. K. Mishra
Department of Computer Science, University of Gondar, Gondar, Ethiopia
e-mail: mishrasoft1@gmail.com

B. Pattanaik
Department of Electrical and Computer Engineering, Blue Hora University, Blue Hora, Ethiopia
e-mail: balapk1971@gmail.com

P. Karthika
Department of Computer Applications, Kalasalingam Academy of Research and Education, Krishnankoil, Viruthunagar, India
e-mail: karthikasivamr@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_56

discusses important protection efforts identified with different organizational conditions, so that an association may be set up with a fully assured organizational environment. Besides, a contextual investigation was examined to delineate the insignificant arrangement of measures necessary in any association to construct network security.

Keywords Steganography techniques · WAN · Advanced encryption standard (AES) · Open System Interface (OSI)

1 Introduction

Security of organizations can be defined as insurance of organizations and their administrations from unwanted change, annihilation, or disclosure, and assurance arrangements that the organization acts in basic circumstances and has no adverse effects on either customer or representative [1]. Besides, it involves arrangements made in a simple PC network foundation, techniques adopted by the director of the company to protect open assets for the organization and the organization from unauthorized clients. Every organization has data security plans these days, steganography assumes a significant part in ensuring the security of data of innovation applications [2]. Data security is a significant issue, for certain applications. The main concern, for example, web-based business banking, email, clinical data sets, thus some more, every one of them requires the trading of private data. For instance, let us consider an individual named Alice a sender who needs to send an information message which has a length of characters to a beneficiary called Bob. Alice utilizes an unstable correspondence channel. This could be a phone line, PC organization, or some other channel [3]. On the off chance that the message contains mystery information, they could be blocked and perused by programmers.

2 Literature Review

They characterized the centre functional systems' administration parts of security including PC interruption identification, traffic examination, and organization observing parts of organization security. A new methodology has been introduced for the controlled, shared use of dispersed security arrangements, called the security network, in which the gadget network ensures that a gadget is reliable and that interchanges between gadgets can be performed levelled out of the framework strategies [1]. Characterized data security is divided into three parts: information security, the security of the network framework and business security of the organization, and the business security model of the organization. In addition, a hypothetical rationale for safety protection for the programmed development system for large businesses has been created. A public key infrastructure (PKI)-based remote organization protection framework has been characterized [14]. Various instruments and treatments identified with cryptography and organizational security have been characterized in this [4–6]. Besides, the most recent problems found with network security innovation and its relevant implementations such as Advanced Encryption Standard (AES), CMAC validation mode and CCM mode for checked encryption guidelines are addressed in an extremely elaborate way [7]. These days, transferring data around an enterprise more safely and securely has become an important test for the company. The attacks and the organizational security activities characterize how a superior, sound and stable organization can be designed and maintained for an association/industry using organizational security devices [8]. This exploration focuses on the problems through which network security in an association can be supervised and maintained more proficiently. In addition to the security strategy and contextual investigation, the better management of the security control agency in an association would help a great deal to understand.

3 Performance Analysis of Steganography Using RSA Algorithm

The main advancement for a wide range of uses is platform and network technology. In current organizations, it is a fundamental prerequisite, and there is a noteworthy absence of security strategies that can be easily implemented. A "correspondence hole" exists between security innovation engineers and organizational designers. Organization configuration is a loop generated based on the model of the Open System Interface (OSI). When planning network defence, the OSI model has few priorities. It provides seclusion, ease of use, adaptability and convention normalization [9, 10]. To build stacks that allow unique turns of events, the conventions of different layers can be effectively joined. Instead of ensuring, the organizational configuration is not calculated (Fig. 1).

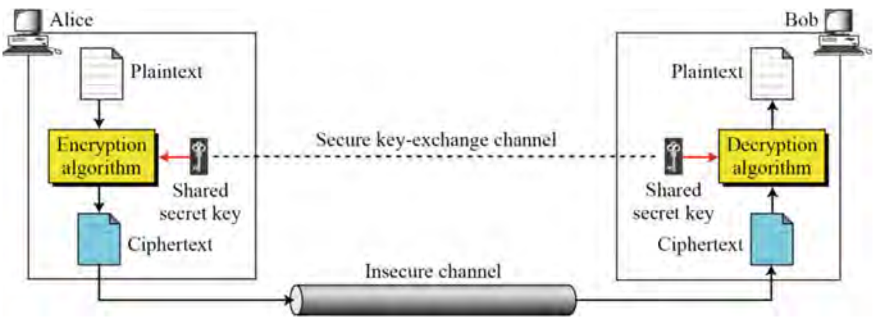


Fig. 1 Steganography technique model

There is no framework to fix the unpredictability of security specifications. When considering organizational protection, it should be stressed that the overall organization is secure. It is not just about the security of the PCs at each end of the correspondence chain. When travelling from one hub to the next hub, the correspondence channel should not be defenceless against attack. A programmer will focus on the channel of correspondence, obtain and decode the content, and re-integrate a print message, although making sure about the organization is similarly as significant as making sure about the PCs and scrambling the message.

On the off chance that P is the plaintext, C is the ciphertext, and K is the key,

$$\text{Encryption} : \mathrm{C} = \mathrm{E_k(P)} \quad \text{Decryption} : \mathrm{P} = \mathrm{D_k(C)} \tag{1}$$

$$\text{In Which,}\ \mathrm{D_k(E_k(x))} = \mathrm{E\ _k(D\ k\ (x))} = \mathrm{x} \tag{2}$$

3.1 RSA Algorithm

RSA represents Rivest Shamir and Adleman name of three innovators. RSA is one of the primary functional public key cryptosystems and is generally utilized for secure information transmission. In such a cryptosystem, the encryption key is public and varies from the decoding key which is left well enough alone. In RSA, this deviation depends on the functional trouble of calculating the result of two huge prime numbers, the considering issue. RSA represents Ron Rivest, Adi Shamir and Leonard Adleman, who first openly depicted the calculation in 1977. RSA includes a public key and a private key. The public key can be known by everybody and is utilized for scrambling messages. Messages scrambled with the public key must be unscrambled in a sensible measure of time utilizing the private key. The keys for the RSA calculation are produced the accompanying way.

Algorithm 1

RSA_Key_Generation

{

Select two large primes p and q such that $p \neq q$.

$n \leftarrow p \times q$

$\phi(n) \leftarrow (p-1) \times (q-1)$

Select e such that $1 < e < \phi(n)$ and e is coprime to $\phi(n)$

$d \leftarrow e^{-1} \bmod \phi(n)$ // d is inverse of e modulo $\phi(n)$

Public_key $\leftarrow$ (e, n) // To be announced publicly

Private_key $\leftarrow d$ // To be kept secret

return Public_key and Private_key

}

Algorithm 2

```
RSA_Encryption (P, e, n)                       // P is the plaintext in Z_n and P < n
{
   C ← Fast_Exponentiation (P, e, n)           // Calculation of (P^e mod n)
   return C
}
```

Algorithm 3

```
RSA_Decryption (C, d, n)                       //C is the ciphertext in Z_n
{
   P ← Fast_Exponentiation (C, d, n)           // Calculation of (C^d mod n)
   return P
}
```

The world is getting more interconnected of the Internet and new system administration innovation. There is enormous measure of individual, military, business and government data on system administration foundations overall accessible. Organization security is happening to extraordinary significance on account of licenced innovation that can be handily procured through the Web.

4 Result Analysis of Network Security

Organization security begins with approval, normally with a username and a secret key. Organization security comprises the arrangements and approaches received by an organization manager to forestall and screen unapproved access, change in framework, abuse or disavowal of a PC organization and organization available assets (Fig. 2).

Fundamentally, network security includes the approval of admittance to information in an organization, which is constrained by the organization administrator. It has gotten more critical to PC clients and associations. If this is handled, then the firewall security arrangements will come into effect, for example, what administrations are permitted to be gotten to for network clients (Fig. 3).

So that to forestall unapproved admittance to a framework, this segment may neglect to check possibly hurtful substance; for example, PC worms or Trojans being communicated over the organization against infection programming or interruption identification frameworks (IDS) help recognize the malware. Today, irregularity may likewise screen the organization like Wireshark traffic and might be logged for review purposes and for later on elevated-level examination in the framework. Correspondence between two hosts utilizing an organization might be utilized encryption to keep up security strategy.

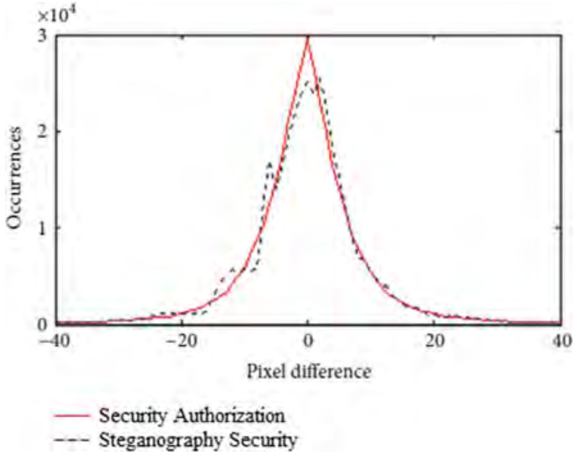


Fig. 2 Network security includes the approval of access

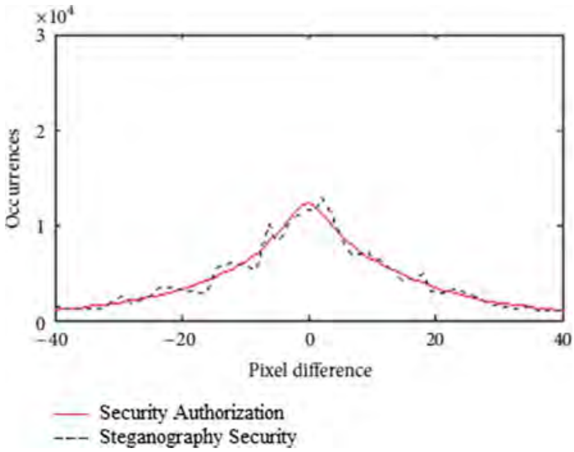


Fig. 3 To check possibly forestall approved admittance to a framework

5 Conclusion

Security is a significant issue besides huge associations. There are different concepts and proposals for protection and risk controls. The safety efforts ought to be planned and given; initially, an organization should know its need for security on various degrees of the association, and later, it ought to be actualized for various levels. Security arrangements ought to be planned first before its execution in such a manner, with the goal that adjustment of selection can worthy with effectively sensible. The security framework should be secure, and also, the end-user should be adaptable to satisfy everyone. The user must believe that the security framework should be appropriate.

References

1. Karthika P, Vidhya Saraswathi P (2017) Content based video copy detection using frame based fusion technique. J Adv Res Dyn Control Syst 9:885–894
2. Karthika P, Vidhya Saraswathi P (2020) Raspberry Pi—A tool for strategic machine learning security allocation in IoT, Apple Academic Press/CRC Press (A Taylor & Francis Group). Proposal has been accepted (provisionally) for the book entitled "Making Machine Intelligent by Artificial Learning", to be published by CRC Press
3. Depren O, Topallar M, Anarim E, Ciliz MK (2019) Anomaly and Misuse detection in machine learning schemes for sensor networks. Expert Syst Appl 29(4):713–722
4. Karthika P, Vidhya Saraswathi P (2020) IoT using machine learning security enhancement in video steganography allocation for Raspberry Pi. J Ambient Intell Human Comput https://doi.org/10.1007/s12652-020-02126-4 Impact Factor 1.91
5. Anuar NB, Shamshirband S, Rohani VA, Kiah MLM, Petkovic D, Misra S (2019) Artificial immune system for detecting intrusion in stego-crptonetworks. J Comput Netw Appl 42(4):102–117
6. Maleh Y, Ezzati A, Qasmaoui Y, Mbida M (2020) Intrusion detection system fora global geometric image using machine learning system. In:Fifth International Symposium on Frontiers in Ambient and Mobile Systems (FAMS 2015) in association with Elsevier-Procedia Computer Science, vol 52, pp 1047–1052
7. Karthika P, Vidhya Saraswathi P (2017) A survey of Content based Video Copy detection using Big Data. Int J Sci Res Sci Technol (IJSRST). 3(5): 114–118. Online ISSN : 2395–602X, Print ISSN : 2395–6011, May-June 2017. https://ijsrst.com/ICASCT2519
8. Ganesh Babu R, Elangovan K, Maurya S, Karthika P (2020) Multimedia security and privacy on real time behavioral monitoring in Machine Learning IoT application using Big Data Analytics. In: Raghvendra K, Sharma R, Pattnaik PK (eds) Multimedia technologies in the Internet of Things Environment (pp 155–177) Springer
9. Shen Y, Liu S, Zhang Z (2020) Detection of hello flood attack caused by malicious cluster heads on efficient machine learning technique. Int J Adv Comput Technol 7(2):40–47
10. Karthika P, Vidhya Saraswathi P (2019) Digital Video Copy detection using steganography frame based fusion techniques. In: International Conference on ISMAC in Computational Vision and bio-engineering. pp 61–68. https://doi.org/10.1007/978-3-030-00665-5_7

Performance of Two-Link Robotic Manipulator Estimated Through the Implementation of Self-Tuned Fuzzy PID Controller

Aditi Saxena, Rishabh Chaturvedi, and Jitendra Kumar

Abstract Robot manipulators become intensified major segments in the production sectors that are utilized in numerous applications such as welding, granulating, mechanical dealing, and assembling because of velocity, accuracy, and repeatability. These applications necessitate a legitimate path plan, appropriate generation of direction, and above all a control design. In a manipulator system, the links are expanded based on efficiency and versatility is upgraded alongside the complexity. This controller tuned to produce the least weighted sum of integral of absolute error and component of positive change in controller output. The controller has the advantage of examining and controlling a highly non-linear and rigid manipulator system. To optimize the system, a genetic algorithm is being used to provide an actual and desired result by incorporating a combination of a genetic algorithm with self-tuning fuzzy. PID is composing the controller much more superior and reliable.

Keywords Self-tuned fuzzy PID · Manipulator · Genetic algorithm

1 Introduction

The term robot is basically a Czech word that means slaved laborers, and it came into existence in 1920. Robots acts like a machine and is dominated by a controller in a robot. Further, the structure of the robot is replicated as similar to human body parts like a head, two arms, and two legs, waist, and so on. Industry-specific robots play out a few tasks, for example, picking and fixing objects, and movement adjusted from seeing how comparable manual tasks are taken care of by a completely working human arm. Such automated arms are otherwise called robotic manipulators. The proposed work aims to have innovation is a multifaceted concept that includes modifications to products, alterations to organizational aspects dependent on innovations, strategies, rebuilding of organizational culture, and revisions to production processes. Accordingly, such optimized controlled robotic manipulator is

A. Saxena (✉) · R. Chaturvedi · J. Kumar
IET Department of Electronics and Communication Engineering, GLA University, Mathura 281406, India
e-mail: aditeesaxenaa29@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_57

appropriate for process innovation since they authorize organizations to concentrate on unique processes for the production of products and services. The second factor is the organizations today are confronted with expanding work costs and a deficiency of laborers, and are thus spending in robotics. Robots nevermore demand promotions and can work nonstop. The third factor is the pressure to increase the production rates to compete the market, and the fourth factor is the increased productivity. The fifth factor explains about the repeatability where the robot's drive product quality or consistency and reduces waste. The last factor represents the speed in which the robotics assist in the increased production and hence reduction in the waiting time.

This controller can easily accomplish all these factors, where the controller is used for underwater controlling, etc. The proposed model can move and grip devices with guidance which can be operated through various programmed motions as it is both programmable and multifunctional devices [1]. Robotics is a combination of two branches [2, 3] that are electronics through the field of control, manufacturing, designing, and kinematics which help in positioning and orientation of the manipulator devices having multi-degree liberty can be positioned with the help of these manipulators. High and improved control strategies are required to achieve accuracy in the trajectory tracking of the manipulator, dynamic model of a robotic manipulator also play a very important role; it is responsible for the performance of a robotic manipulator. To achieve the desire performance and response of the system, a new design is needed for an accurate mathematical model, and after this, the various control strategies are applied to get precise trajectory tracking [4]. For industry purpose, the manipulators that are being used are highly non-linear and coupled; a second-order non-linear differential equation is being used to design a dynamic model for two links rigid robotic manipulator [5]. If the number of links increased simultaneously complexity also increased in terms of position and velocity of the manipulator. To overcome this challenging task, a control system is employed to monitor the manipulator [6]. Especially after designing a dynamic model, it is difficult to attach a precise and powerful controller to the plant that could be able to handle the non-linearities and many other unknown parameters. The controller is managed by adopting a tuning algorithm to have flexibility over the automation, where the entire world is searching for automatic processing in every aspect of life. Many different types of the algorithm are available in current trends such as particle swarm optimization (PSO), optimization method of ant colony [7], cuckoo search algorithm (CSA) [8, 9], genetic algorithm (GA), and so on. In the proposed system, genetic algorithm is used to produce high-quality solutions for optimization and search issues by depending on the process of natural selection such as crossover, mutation, crossover, and selection. The main aim is to reduce the amount of error. The controller will operate with the self-tuned fuzzy logic controller to have the least value of error and can also supervise the non-linearities [10–12]. The limitations of the proposed work are that this controller could be able to control a two-link manipulator having a payload of 0.5 kg [13] (Fig. 1 and Table 1).

2 Dynamic Model for Manipulator

In this study, the configurations of the robotic manipulator is explained and dynamic model was designed by studying the equations of motions for manipulator arm which was given by Lin [14],

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}\begin{bmatrix} \theta_{11} \\ \theta_{22} \end{bmatrix} + \begin{bmatrix} P_{11} \\ P_{21} \end{bmatrix} + \begin{bmatrix} f_{r1} \\ f_{r2} \end{bmatrix} + \begin{bmatrix} fn_{1p} \\ fn_{2p} \end{bmatrix} = \begin{bmatrix} \tau f_{1p} \\ \tau f_{2p} \end{bmatrix}$$

where

$$\begin{aligned} S_{11} = {} & I_{1p} + I_{1p} + m_{11} l_{c1}^2 + m_{22} l_{c2}^2 + 2 m_{22} l_{11} l_{c2} \cos\theta_{22} \\ & + m_{vp} l_{11}^2 + m_{vp} l_{22}^2 \\ & + 2 m_{vp} l_{11} l_{22} \cos\theta_{22} \end{aligned} \tag{1}$$

$$S_{12} = I_{2P} + m_{22} l_{c2}^2 + m_{22} l_{11} l_{c2} \cos\theta_{22} + m_{vp} l_{22}^2 + m_{vp} l_{11} l_{22} \cos\theta_{22} \tag{2}$$

$$S_{21} = S_{12} \tag{3}$$

$$S_{22} = I_{2P} + m_{22} l_{c2}^2 + m_{vp} l_{22}^2 \tag{4}$$

$$P_{11} = -m_{22} l_{11} l_{c2} (2\theta_{11} + \theta_{22}) \theta_{22} \mathrm{Sin}\theta_{22} - m_{vp} l_{11} l_{22} (2\theta_{11} + \theta_{22}) \theta_{22} \sin\theta_{22} \tag{5}$$

$$P_{21} = m_{22} l_{11}\, \theta_{11}^2 l_{c2} \sin\theta_{22} + m_{vp} l_{11} \underset{11}{\theta^2}\, l_2 \sin\theta_{22} \tag{6}$$

$$f_{r1=} b_{1vp}\, \theta_{11} \tag{7}$$

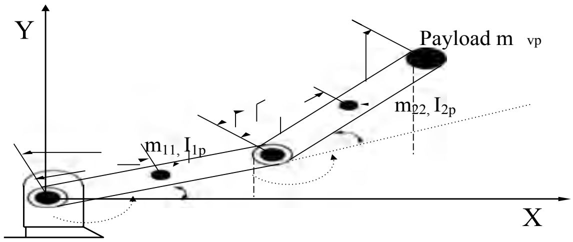
$$f_{r1=} b_{1vp}\, \theta_{22} \tag{8}$$

$$\begin{aligned} fn_{1p} = {} & m_{11} l_{c1} g \cos\theta_{11} + m_{22} g (l_{c2} \cos(\theta_{11} + \theta_{22}) \\ & + l_{11} \cos\theta_{11} + m_{vp} g (l_{22} \cos(\theta_{11} + \theta_{22}) \\ & + l_{11} \cos\theta_{11}\ (9) \end{aligned} \tag{9}$$

$$fn_{2p} = m_{22} l_{c2} g \cos(\theta_{11} + \theta_{22}) + m_{vp} l_2 g \cos\ (\theta_{11} + \theta_{22}) \tag{10}$$

$$\theta_{11} = \frac{\tau f_{1p} - f_{n1p} - f_{r1} - P_{11} - S_{12} * \theta_{22}}{S_{11}} \tag{11}$$

Fig. 1 Diagram of two link robotic manipulator



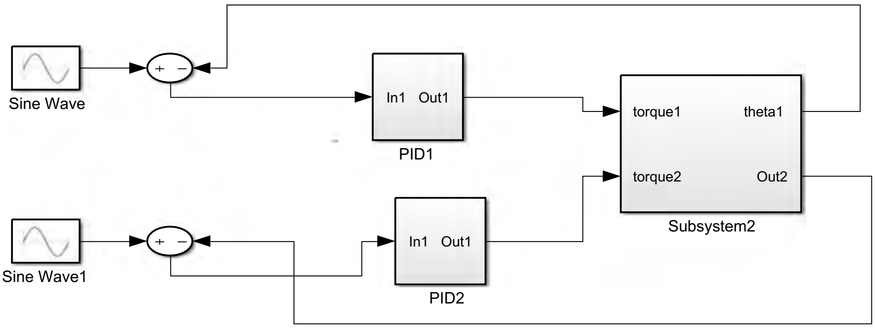


Fig. 2 PID structure

$$\theta_{22} = \frac{\left(\tau f_{2p} - f_{n2p} - f_{r2} - P_{21} - S_{12} * \theta_{11}\right)}{S_{22}} \tag{12}$$

where θ_{11} and θ_{22} shows the links position; τf_{1p} and τf_{2p} are the generated torque; f_{r1} and f_{r2} are the dynamic friction coefficients; m_{11} and m_{22} shows the masses. l_{11} and l_{22} shows the length, I_{1p}, I_{2P} shows inactivity.

Initially, the mathematical model for each equation is constructed as S_{11},S_{12},S_{21},S_{22},P_{11},P_{21},f_{r1},f_{r2},fn_{1p},fn_{2p}. Further, the mathematical model is again converted into a dynamic model for further processing. On the MATLAB SIMULINK, a subsystem for these equations is created and appended with the equations shown in no 11 and 12. This whole mathematical model accommodates various controllers named subsystem 2 that are given in Fig. 2 (Figs. 3, 4, 5 and 6).

2.1 *Various Controller Structure*

2.1.1 PID Controller Structure

This controller structure is designed using the MATLAB SIMULINK. The design of the simple PID controller is constructed by connecting link 1 and link 2. These

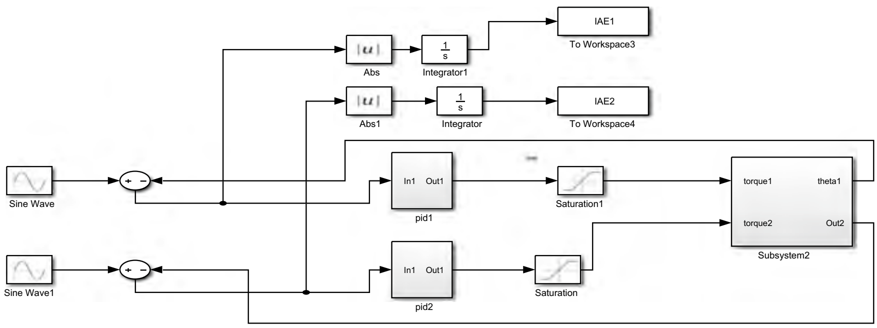


Fig. 3 Optimized PID structure

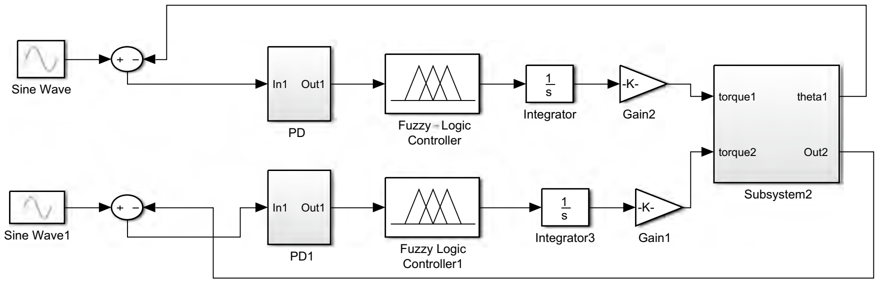


Fig. 4 Fuzzy PD structure

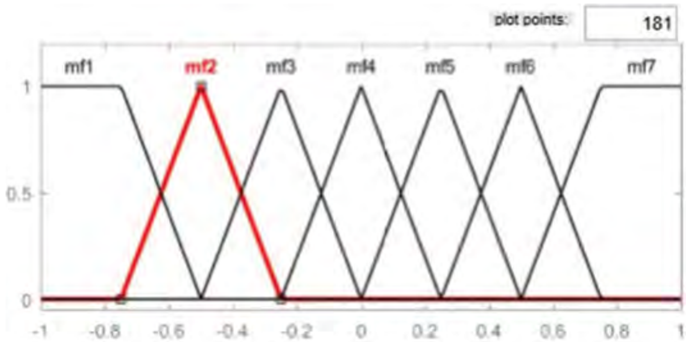


Fig. 5 Membership function plot

links reflect the operations of the equations defined in the mathematical model. Subsequently, a subsystem is being created of the model. The control boundaries of the PID structure are namely, proportional term, integral term, and derivative term. The proportional term refers to a general control activity proportional to the error signal through the steady gain factor. The integral term aims to decrease consistent

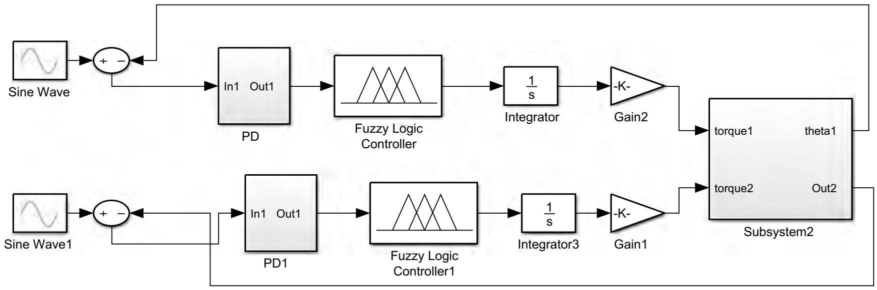


Fig. 6 Fuzzy PI structure

state errors through low-frequency remuneration by an integrator. The derivative term aid to improve transient reaction through high-frequency remuneration by a differentiator. Each link has its own PID controller connected for generating the optimal output. The inputs supplied for each link are in terms of the sine wave, and the controller supports in all sorts of processing, and the final output is matched with the mathematical model equations. If any mismatch occurs between the corresponding desired value and the actual value, then the controller is again fine-tuned towards the gain factor. The genetic algorithm is employed to enhance time and achieve higher precision accuracy.

3 Genetic Algorithm for Optimization

Nature has consistently been an incredible wellspring of motivation for all humankind. The genetic algorithm is associated with search calculations based on the ideas of selection and genetics. In GAs, a pool or a populace of potential clarifications are given for the specified issue. These clarifications at that point go through recombination and transformation (like in genetics), delivering new solutions, and the cycle is rehashed over different circumstances. Every value is specified as a fitness value, and the fitter is given a higher opportunity to yield more fitter. An optimistic approach such as integral absolute error (IAE) is determined to eradicate the number of errors. To operate with the IAE functionality, a small piece of code is written in the command window. The next process is to assign a sim function inside the MATLAB, which is being inbuilt under the optimization tool. By executing the command as optimtool in the command window a pop-up window gets activated, and inside the textbox type the term solver. Finally, run the GA commands with the number of variables associated with the gains. The sequence of steps involved in the genetic algorithm is (1) Population, (2) Fitness scaling, (3) Selection, (4) Reproduction, (5) Mutation, (6) Crossover, and (7) Migration. The population comprises of all the data. The scaling function converts raw fitness scored returned by the fitness function in a given range. The selection function determines parents for next-generation based on their scaled values. The reproduction essentially demonstrates the functionality of creating children for each new generation then progressed by the mutation function, which makes a small random change in the individual which provides genetic diversity. The crossover consolidates two individuals or parents to form a new individual for the next-generation. The values for each operation are assigned using the two commands such as assignin and sim assignin which are defined inside the workspace. The sim function returns a single simulink (simulation output object) that contains all the simulation output. The obtain results aids to plot the output signal value against the time.

3.1 PID Optimized Controller Structure

The genetic algorithm technique that is discussed in the previous section is being used in this model for optimization. This is the reason for the optimized PID controller design; IAE is being attached to the controller input. The rest of the whole structure is the same as the PID controller structure, where it contains only two blocks, and the initial block consists of absolute, and the other is integrator over time. The IAE integrates absolute error value and provides the least value of error.

3.2 Fuzzy PD Controller Optimized Structure

PID controller is alone insufficient to handle all the non-linearities, so there exists a need for adding the fuzzy logic functionalities. The fuzzy controller or logic consists of membership functions which are described in Fig. 7. The membership functions are being shown each has their respective ranges as fuzzy only accept 0 or 1; so accordingly, the range is being decided with the help of a rule base a fuzzy logic. The design is created by typing fuzzy in the command window, a fuzzy logic appears and it is possible to call this logic to the workspace where our structure has been

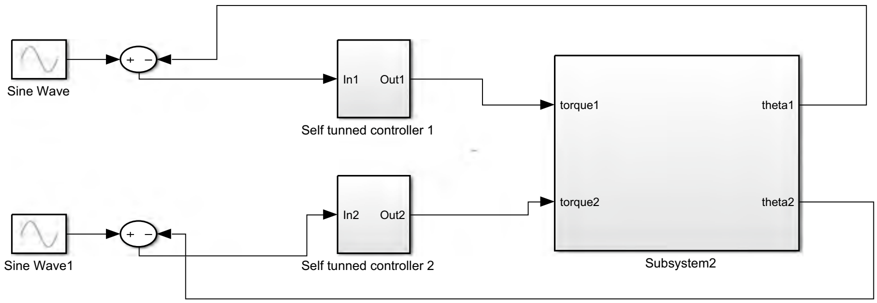


Fig. 7 Optimized self-tuned (FUZZY PID) model in SIMULINK

designed. Further, the fuzzy logic is being combined with PD controller and termed as fuzzy PD controller (Fig. 8).

3.3 Fuzzy PI Optimized Structured

This structure is the same as the above one the main difference here is the PD controller is being integrated which results in PI controller, and fuzzy logic is also being attached between PD and integrator. The whole structure as a sum give origin to fuzzy PI controller for optimization, again genetic algorithm is being used with the help of inbuilt optimtool function.

3.4 Optimized Self-Tuned Fuzzy PID (STFPID) structure

Here, two fuzzy layer are utilized, the output of fuzzy layer 1 is being multiplied by the product of gains and output of fuzzy layer 2, then fed to the plant. The fuzzy logic self-tuning structure is being represented below is termed as a self-tuned fuzzy PID controller (STFPID). This controller gives the lesser error value after optimizing the genetic algorithm.

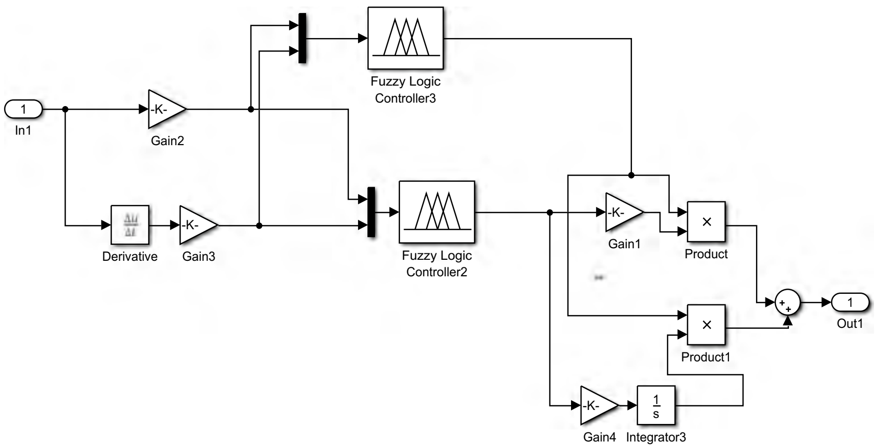


Fig. 8 Structure of self tuned fuzzy PID controller

4 Results and Discussions

For tracking the trajectory, the two-link robotic manipulator and the controller are optimized on its own with the help of a self-tuned fuzzy PID controller. The controller structure as discussed above, these controllers is working efficiently and fulfilling all their aims for which they have been designed. To check all this, a simulator is needed and MATLAB is used to compare each optimized controller. Figure 9 represents the tracking, the reference, and the output wave is perfectly matched. If it does not match, differentiate it between these two waves with the guidance of color-coding. The dynamic model is designed with the equations, and the controllers, which are being attached to it, are perfect and working precisely. Figure 10 describes a self-tuned fuzzy PID controller when operating through the genetic algorithm process as a simulation result, it shows the minimum error value. In comparison to all the other controllers which are being used, this controller contains many features. The self-tuned structure is allowing the controller to automatically tune itself depending upon the presence of non-linearities. The fuzzy has many advantages and can deal with multiple-links non-linearities, increased time, overshoot, and many more. The base

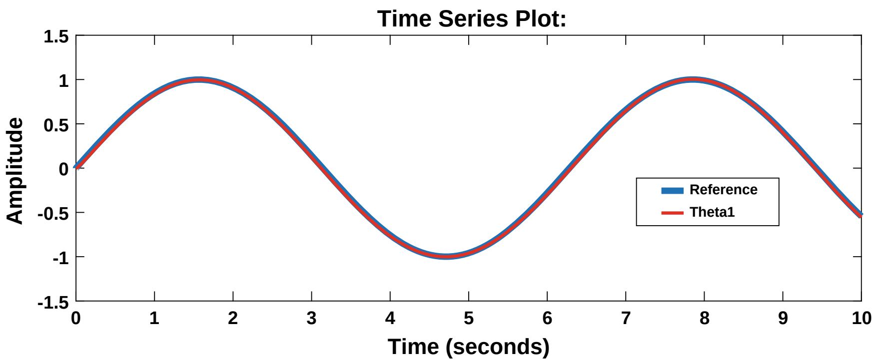


Fig. 9 Representing accurate tracking

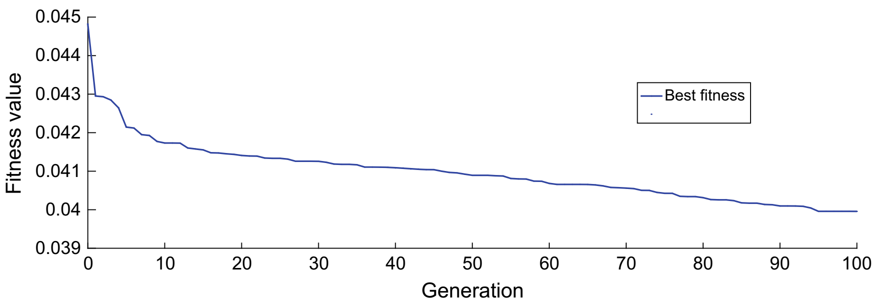


Fig. 10 Self-tuned fuzzy PID after optimization

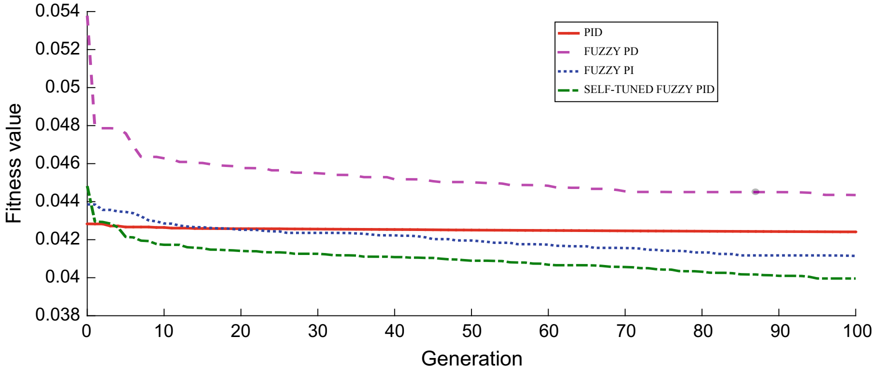


Fig. 11 IAE values of various controllers

for the fuzzy is PID in which all the three control parameters such as proportional, integral, and derivative are combined and serves in reducing error and providing stability to the system. Figure 11 showing the simulation result after working through a genetic algorithm all the five controllers are being individually simulated up to 100 generations. The self-tuned fuzzy PID controller has the least value of error as compared to any other. Table 2 represents the mathematical value of error for each controller in descending order and the self-tuned fuzzy logic has an error value of 0.03665 which is like next to perfection in the third graph all the controller's error. These are identified by separate graphs that are being differentiated with different color coding. This controller can work as an adaptable controller and can self-tuned and optimized itself according to the surrounding parameters and non-linearities.

5 Conclusion

In today's world of automation, the need for developing an adaptable controller is essential. This controller is working as a self-adaptable controller that can be tuned and adjusted on its own with the help of a self-tuned fuzzy PID controller and genetic algorithm. With the optimized simulation results, it is found that this controller has the least value of error when compared to other controllers that are also being designed and simulated. The proposed controller not only helps to control the two link manipulator but also has the application for controlling the underwater robot, to control an adaptable gripper, and many more.

Table 1 Parameters of the dynamic model

Parameters	Link 1	Link 2	Values
m_{11}	0.392924 kg	m_{22}	0.094403 kg
l_{c1}	0.104648 m	l_{c2}	0.081788 m
l_{11}	0.2032 m	l_{22}	0.1524 m
I_{1P}	0.0011411 kg m^2	I_{2P}	0.0020247 kg m^2
b_{1vp}	0.141231 N-m/radian/s	b_{2vp}	0.3530776 N-m/radian/s
m_{vp}	0.56699 kg		
g	/s^2		

Table 2 Tabulated form of IAE value in descending order

Various controller	Values of IAE
PID without optimization	0.2218
PID with optimization	0.0429
Fuzzy PD with optimization	0.04438
Fuzzy PI with optimization	0.04114
Self-tuned Fuzzy PID using GA	0.03665

References

1. Sharma R, Rana KPS, Kumar V (2014) Performance analysis of fractional order fuzzy PID controllers applied to a robotic Manipulator. Expert Syst Appl 41(9):725, 4274–4289:726
2. Kumar V, Rana KPS, Mishra JK, Nair SS (2016) A robust fractional order fuzzy P+ fuzzy I+ fuzzy D controller for nonlinear and uncertain system. Int J Autom Comput 14(4):474–488
3. Kumar V, Rana KPS (2017) Nonlinear adaptive fractional order fuzzy PID control of a 2-link planar rigid manipulator with payload. J Franklin Inst 354:993–1022
4. Kumar J, Kumar V, Rana KPS A fractional order Fuzzy PD+I controller for three link electrically driven rigid robotic manipulator system. J Intell Fuzzy Syst IOS Press, Netherlands (SCI Index, Impact factor-1.261)
5. Weile DS, Michielssen E (1997) Genetic Algorithm Optimization applied to electromagnetics, a review. IEEE Trans Antennas Propag 45(3):343–353
6. Kong Z, Jia W, Zhang G, Wang L (2015) Normal parameter reduction in soft set based on particle swarm optimization algorithm. Appl Math Model 39:4808–4820
7. Ghanbari A, Kazemi SMR, Mehmanpazir F, Nakhostin MM (2013) A cooperative ant colony optimization-genetic algorithm approach for construction of energy demand forecasting knowledge based expert systems. Knowl-Based Syst 39:194–206
8. Jagatheesan K, Anand B, Samanta S, Dey N, Ashour AS, Balas VE (2017) Design of a proportional-integral-derivative controller for an automatic generation control of multi-area power thermal systems using firefly algorithm, IEEE/CAA J Automatica Sinica, 1–14 (2017). https://doi.org/10.1109/JAS.2017.7510436
9. Yang XS, Deb (2009) Cuckoo search via Lévy Flights In: . Proceedings world congress on nature and biologically inspired computing, India, pp 210–214
10. Yang XS, Gandomi AH (2012) Bat-algorithm, a novel approach for global engineering optimization. Eng Comput 29(5):464–483
11. Ohtani Y, Yoshimura (1996) Fuzzy control of a manipulator using the concept of sliding mode. Int J Syst Sci 27(2):179–186

12. Hazzab A, Bousserhane IK, Zerbo M, Sicard P (2006) Real-time implementation of fuzzy gain scheduling of PI controller for induction motor machine control. Neural Process Lett 24:203–215
13. Sharma R, Bhasin S, Gaur P, Joshi D (2019) A switching-based collaborative fractional order fuzzy logic controllers for robotic manipulators. Appl Math Modell
14. Lin F (2007) Robust control design: an optimal control approach. John Wiley & Sons Ltd., England

Rule Based Part of Speech Tagger for Arabic Question Answering System

Samah Ali Al-azani and C. Namrata Mahender

Abstract Part of Speech (POS) Arabic wording is difficult to read in detail and its functionality affects many programs and activities in the Natural Language Processing (NLP) area. POS tagging is a process to assign POS such as a verb, adjective, adverb, noun in each word for any sentence. Farasa is an active and reliable text processing toolkit for Arabic documents. It is an assortment of Java libraries and CLIs for MSA.2. These incorporate a separate tool for Arabic text Diacritizer, segmentation/tokenization module, POS tagger, Named Entity Recognition (NER), and parsing. One of the limitations over the Farasa affects the post-processing results due to the presence of inappropriate tags. For our application on question answering system(QAS) correct POS, tagging is essential for better accuracy. The POS tagger is developed using a rule-based approach which is based on domain-specific. The corpus (database) on which the rule-based POS tagger is built is centered on the core subject of the Arabic language 4th standard textbook of Arabic Medium state board of Yemen. During the development of QAS, the POS tagger is a very essential stage in which the answers for the framed questions are obtained from the paragraphs of a given lesson. The present article provides insights into the complete process of linguistic rule-based POS tagger development for QAS. Sentence segmentation, word tokenization, to stemmer development which becomes an important part of proper morphological analysis is explained. As a result, the morphological analyzer is the input to the rule-based POS tagger. Ultimately, in this article, a comparison of marking based on our POS rule with Farasa is presented and for QAS, our rule-based POS tagger gave better results than Farasa.

Keywords Arabic language · Natural language processing (NLP) · Part of speech (POS) tagging · Morphological analyzer · Stemmer · Tokenization

---

S. A. Al-azani (✉) · C. Namrata Mahender
Department C.S. and I.T, Dr. Babasaheb Ambedkar, Marathawada University, Aurangabad, Maharashtra, India
e-mail: alazani183@gmail.com

C. Namrata Mahender
e-mail: nam.mah@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_58

1 Introduction

Natural language processing (NLP) is a category of artificial intelligence that assists computers to recognize, translate, and administer human language. It is utilized in numerous disciplines and incorporates computational linguistics, computer science, and so on. It also bridges the gap between computer understanding and human communication [1]. This methodology recognizes natural language texts and linguistic procedures namely, part of speech (POS) tagger, lexicon, and tokenization. These are applied to recreate inquiries into the right inquiry that separates the significant answers from a structured database [2]. The inquiries dealt with by this methodology are of Factoid type and have a profound semantic understanding [3].

1.1 Arabic Natural Language Processing

The Arabic language is a combination of many variations among different similarities that have a specific event, such as the standard official text of media and education throughout the Arabian World [4]. It is one of the top languages in the world. Its phenomenal script, distinguished style, and strong vocabulary confer a unique character and characteristic to the language. Arabic is the largest member of the Semitic language family. Nowadays, it is an official language in more than 20 countries and over 300 million native speakers [5]. Concerning other Semitic dialects, Arabic morphology was set up around the abstract idea of the root, three consonants articulating to significance, regardless of whether exact or ambiguous. The conventional derivational morphology is dependent on the root-and-pattern model concentrating on this abstract consonantal root. A pattern is an intermittent affix (or transfix) composed of vowels and non-revolutionary consonants embedded around spaces for the root consonants. To each pattern, conventional syntax relates a morphological classification or potentially inflectional highlights. These formalizations are applied by traditional grammar to depict both inflectional and derivational morphology. Moreover, the quantity of 'voweled stem canonic patterns' for action words and nouns is almost 10,000 [6]. Some problems in Arabic texts include considerable translations and translated labeled entity, its satire is usually contradicted texts on Arabic. Despite the fact that the methodology demonstrated to create high exactness, it has a few flaws. It requires building a tremendous corpus (dataset) and marking it manually by human experts. The system of manual commentary can be extremely troublesome in any event, for native speakers because of criticism and social references. It can also be costly and tedious. A constraint is that NLP tools designed for Western dialects are not effectively versatile to Arabic due to the specific highlights of the Arabic language [7].

2 Related Work

Many different researchers have focused on part of speech tagging in many languages like Hindi, English, Arabic, Chinese, Marathi, and others. Here some important related work in part of speech tagging techniques is discussed.

Singa et.al. (2012) has advanced part of speech tagger on based rule approach with a rate of accuracy 50, 77, 85% on lexicon data words (50-100-1000) in Manipuri language (Manipuri is mightily Speaker in Manipur, Bangladesh, Tripura, and Myanmar, Assam [8].

Zelalem (2013) has used a hybrid approach of HMM and rule-based tagger, on a collected dataset of 354 sentences with accuracy of 77.19, 61.88, and 80.47% in Kafi- **Noonoo** language (language in southwestern Ethiopia) [9].

Deepali et al. (2018) has designed the POS tagging base rule in the Marathi language, on a collection of 1364 words with an accuracy rate was 100% [10].

Aliwy et al. (2018) suggested a new approach using HMM and n-grams taggers for tagging Arabic words in a long sentence to collect 1000 datasets as documents and 526,321 as a separate token, this system gives an accuracy rate of 0.888, 0.925, and 0.957 [11].

Barud et al. (2019) developed parts of speech tagger for Awngi language using Hidden Markov Model (HMM), in a dataset 94,000 sentences were collected, total word of 188,760 and attained accuracy rate of 93.64 and 94.77% [12].

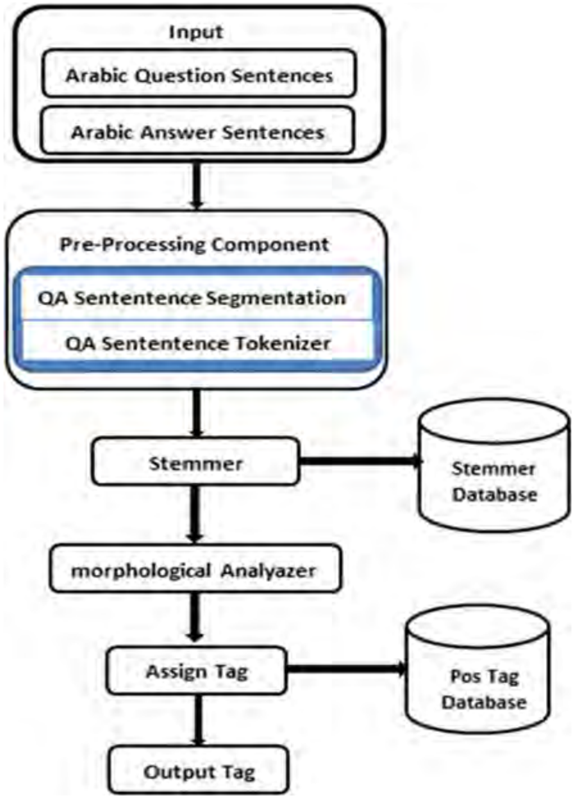
Hagos, 2020 designed a Ge'ez (Ethiopian language) POS tagging using a Hybrid approach, the data set collected was 15,154 words, 1,305 sentences, and the result accuracy rate was 77.87, 82.23, and 94.32% [13].A detailed description of building a POS tagger is discussed in further sections.

3 Proposed System

3.1 Data Collection

The proposed system considers 40 questions and 40 answers that are taken from five different lessons of the 4th standard textbook of the Arabic language as the raw input. The data is pre-processed, and a well-designed stemmer is developed for supporting the morphological analyzer. An output of a morphological analyzer is used as an input to the part of speech tagging. A linguistic rule-based POS tagger is developed, and ultimately, the tag data is provided by the system. This tagged data will be used in post-processing for better performance of the QA system. The detailed block diagram of the system proposed is shown in Fig. 1, and further stages have been discussed in detail with example.

Fig. 1 Proposed system



3.2 Pre-processing Component

3.2.1 Questions and Answers Sentence Segmentation

In the Arabic language, the structure is dissimilar for different sentence types. The proposed system consists of two segments namely, question and answer. The former is the question part, and the latter is the answer part. The input data is taken from the 4th standard textbook of Arabic language. The questions are framed using the textbook lesson and the number of word used are confined. The answers for the question are stored in the answer file and the answers are based on the lessons.

3.2.2 Questions and Answers Sentence Tokenization

Tokenization is the method of tokenizing or parting a string, text into a listing of tokens. One can consider token parts like a word is a token in a sentence, and a sentence is a token in a passage. This process of separating tokens is the input text.

Each word is separated from the sentence considering white space or symbols as one token and treat each word individually. Then POS needs to split the input text into tokens tagging. The tokenization of the question sentence is shown below:

' لماذا يعد البن اليمني من اجود انواع البن في العالم ؟ '

and word tokenization has split answer sentence into words as follows:

' يعد' ' البن' ' اليمني' ' من' ' اجود' ' انواع' ' البن' ' في' ' العالم' ' لما' ' يمتاز' ' بة
' ' من' ' نكهة' ' طيبة ' '.'

3.3 Stemming

Stemming is the operation of decreasing a word to its word stem that contains affixes to suffixes and prefixes or the roots of words known as the lemma. A rule-based steamer uses specific pre-defined rules according to language to mark another type of word used in its base. These language connected rules are created manually by language practitioners. Rule-based stemming methods are divided into three categories such as morphological, table lookup, and affix stripping. Arabic stemmer is a very different and difficult structure than other languages. Stemming is very necessary for natural language understanding and natural language processing. Arabic word language structure is a grammatical rule on the root, and pattern scheme, its deliberate as a root based language with more than 10,000 roots. Arabic words are commonly founded on three-dimensional roots: three consonants, which characterize the hidden significance of the word. Diverse long and short vowels, prefixes, and postfixes are added to that root to make the ultimate desired inflection of sense. These modifications follow designs that reflect across roots. Stemming is the only source to extract the root in the Arabic language. For example, the Arabic word ' المعلمات 'contains the following component in Tables 1 and 2.

Figure 2 depicts an example for code execution and the flow of stemming in the proposed model (Figs. 3 and 4).

3.4 Morphological Analyzer

The morphological analyzer of Arabic words is a procedure for each word of the input text to choose its root and pattern. The outcomes of the morphological analyzer can be used for further analysis[14]. The morphological analysis goals are to train the internal structure of a word. Words after molding are analyzed to check if they are sorted or not. When a stem word is produced, then the word root is formed

Table 1 Example of Arabic Affix

Word	root	prefix	Suffix	Infix
المعلمات	علم	ال	ات	ا

Fig. 2 Result for Arabic stemming from QA Sentences

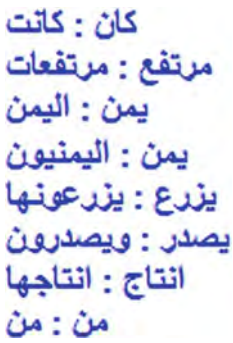


Fig. 3 Role of the morphological analysis of Arabic words

Fig. 4 Result of rule-based POS tagger for QA in the Arabic language

by combining substituted letters with the stem word. The morphological analyst is expected to produce root names for the given input document [2].

Table 2 Arabic prefixes

Words Example	Prefix
بالحديقة	ب
كدواء	ك
والاولاد	و
الزراعة	ال

Table 3 Rule-based POS tagging and Farasa POS tagging

Rule-based POS tagging	Farasa POS tagging	Word	NO
RB	NOUN	عند	1
IN	NOUN	الى	2
VB	NOUN	اطلقة	3
JJ	NOUN	مزدلة	4
IN	NOUN	او	5
JJ	NOUN	اجود	6
VB	NOUN	تقديم	7
JJ	NOUN	كرم	8
VBD	NOUN	نادت	9
IN	NOUN	لانة	10
JJ	NOUN	صعوبة	11
PP	NOUN	بجانب	12

3.5 *Tag Generation*

Initially, Farasa is attempted to obtain the tagged word, but the results were not that much appreciable. Table 3 shows the incorrect tags using Farasa.

A linguistic rule-based POS tagger is chosen to be developed due to the limitations over the Farasa.

3.6 *Rule Based POS Tagger*

The data is manually analyzed to find the context for each part of speech and based on the data a rule is developed.

(a) **Algorithm for POS Tagging System**

Step 1 Input the question and answer sentence segmentation.

Table 4 Proposed system data

Lessons	No. of (Q & A)	No. of words (Q & A)
L1	16	134
L2	16	100
L3	16	123
L4	16	135
L5	16	131
Total Words		623

Step 2 Start to tokenize the Q & A sentence into words.

Step 3 Generate a stemmer for all words to obtain the original words by using a morphological analyzer.

Step 4 Gather all the incorrect tags for all the words from the Farasa parser as a word by word and transfer to the database to produce the correct tag for all words.

Step 5 Allocate a suitable tag to append to its word.

(b) **Result**

In this article, the data were collected from the Arabic language among one of the subjects from the 4th standard Arabic medium state board pattern in Yemen. Five different lessons are considered in this work and formed 40 questions and 40 answers respectively. Table 4 illustrates the data in lessons.

The formula to calculate the accuracy and the performance in Farasa parser tagging and rule-based POS tagging is applied:

Accuracy = (No. of Correctly tagged/ Total No. Tagged in documents)*100.

Subsequently, the accuracy result in Farasa parser was 92.77%. and the proposed system accuracy is 100%.

4 Conclusion and Future Scope

Question Answering System comes beneath the purview of natural language processing, thus natural language understanding is an essential task. For the present work, the corpus (database) for the QAS is build based on the Arabic language subject of the 4th standard Arabic medium state board pattern in Yemen. POS tagger is a very important component of any QA system as it impacts the accuracy of question answers generated for the system. In thisarticle, the major inputs to the POS tagger is based on sentence splitting, word tokenization, and stemming which serves as the basis for better performance of POS tagging. Linguistic POS tagger rule-based design and implementation are presented in detail concerning the QA system. The results are compared with Farasa and determined that our rule-based system yielded better results. In the future scope, this tagged data will be employed to measure the results of the QA system.

References

1. Shaalan K, Siddiqui S, Alkhatib M, Monem AA Challenges in Arabic natural language processing , School of Informatics, University of Edinburgh1, UKFaculty of Computer and Information Sciences, Ain Shams University 4, Abbassia, 11566 Cairo, Egypt
2. Yao X (2014) Feature-driven question answering with natural language alignment , Johns Hopkins University, Doctor of Philosophy thesis 2014
3. AL-Taani A, Abu Al-Rub S (2009) A rule-based approach for tagging non-vocalized Arabic words. Int Arab J Inf Technol 6(3)
4. Habash NY (2010) Introduction to Arabic natural language processing, A Publication in the Morgan and Claypool Publishers series. ISBN: 9781598297966
5. https://arabicquick.com/an-introduction-to-the-arabic-language
6. Darwish (2002) Building a Shallow Arabic morphological analyzer in one day. In: Proceedings of the ACL workshop on computational approaches to semitic languages, Philadelphia, PA, pp 1–8
7. Ezzeld AM, Shaheen M (2012) Survery of Arabic question answering: challenges ,tasks , approaches , tools , and future trends. In: The 13th international Arab conference on Information technology ACT2012. ISSN.1812–0857
8. Raju Singha KH, Purkayastha BS, Singha KD (2012) Part of speech tagging in Manipuri: a rule-based approach. Int J Comput Appl 51(14): 0975–8887
9. Mekuria Z (2013) Design and development of part-of-speech tagger for Kafi-noonoo language, Master's thesis, Addis Ababa University, Addis Ababa
10. Deepali G, Naik Ramesh R, Namrata Mahender C (2018) Rule-based part-of-speech tagger for Marathi language, © 2018 IJSRST | vol 4, issue 5. Print ISSN: 2395–6011. Online ISSN: 2395–602X Themed Section: Science and Technology
11. Aliwy AH, Al_Raza DA (2018) Part of speech tagging in Arabic long sentence. Int J Eng Technol 7(3.27):125–128
12. Wubetu Barud Demi lie, Parts of Speech Tagger for Awngi Language. ISSN 2321 3361 © 2019 IJESC
13. Gebremedhin H, gebremeskel S (2020) Ge'ez POS tagger using hybrid approach. Int J Comput Sci Inf Technol Res 8(1):12–23
14. Awajan A (2018) A rule-based morphological analyzer of Arabic words. https://www.researchgate.net/publication/330006249

Change Detection in Land Use-Land Cover Using Convolutional Neural Network

Sahithi Samudrala, Mekala Pranvitha nand, Suhail Mohammad, and Radhesyam Vaddi

Abstract Change detection in satellite imagery is a concerning topic for researchers and scientists. Studying changes in satellite images gives an enormous amount of knowledge in studying the geography and ecosystem. The dataset used in this work is taken from the AVIRIS sensor and normalized to reduce the time of execution. This paper covers the important aspects of the classification of satellite imagery using convolutional neural networks. Changes are detected based on the classification maps in multi-temporal imagery. Experimental results show that the proposed method got accuracy on par with state-of the-art methods.

Keywords Satellite imagery · Convolutional neural networks

1 Introduction

The changes on the earth's surface from natural and human causes are rapid. Scientists and researchers are concerned about detecting the changes without knowing the scale of changes which is very hard and need to be addressed [1]. Many techniques had been developed for detecting changes on land cover which are optimal for both satellite imagery and also for digital imagery [2]. Satellite images are different from normal images. They are not to be compared with each other. We can adapt the techniques or algorithms used for normal imagery on satellite imagery. Normal images are three or four channels and are speeded over the visual spectrum. Satellite images are multiple channels and are ranged from infrared to ultraviolet in the electromagnetic spectrum

S. Samudrala (✉) · M. P. nand · S. Mohammad · R. Vaddi
Department of Information Technology V.R, Siddhartha Engineering College, Vijayawada, India
e-mail: sahithi.rocking25@gmail.com

M. P. nand
e-mail: mekalapranvithanand@gmail.com

S. Mohammad
e-mail: 8125527706ms@gmail.com

R. Vaddi
e-mail: syam.radhe@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_59

with minimal intervals. Digital images represent the intensity of the red–green–blue in the image, whereas satellite images are reflection indexes for each interval [3].

The push-broom passively was considered for scanned images [4]. The spectral images can be collected in multiple ways from satellite to airborne imaging through more complex spectroscopes. Spectroscopes scan land cover in the spatial perspective in all the intervals of the electromagnetic spectrum collecting spectral data.

The spectral data collected for each spatial pixel is labeled manually by surveying onsite, and respected ground truths of the spectral data are provided [5]. But it is not possible to survey land in a remote region manually, so we are using a deep neural network to learn the patterns of each spectral data in a remote region from the knowledge of spectral data which was provided with ground truths collected by manual surveys generate the classification maps aka ground truths [6].

The full article is organized as follows. Section II describes the project that needs to be changed for the detection of the dataset. Section III explains the CNN algorithm and Section IV depicts the methodology for the proposed system. Section V presents a complete analysis of the results. Finally, Section VI describes the conclusion and future work.

2 Dataset Description and Normalization

The dataset used in this work is from the AVIRIS sensor [7] with the following features.

(a) Each class is uniquely represented by an integer in the ground truth.
(b) They are 16 classes in the dataset
(c) The spatial resolution of the imagery is 145 × 145 pixels with a spectral resolution of 220 channels for each pixel.

The data consists of a wide range of variables with very high numeric values. Cleaning the outliner values and normalizing the entire data was done for reducing the computation stress. Reducing the scale can fasten the computation, and there is no loss in the pattern. The patterns are shown in Fig. 1 (Fig. 2).

3 Algorithm

There is a wide choice of selection in deep neural networks for supervised learning. The primary reason is to choose a convolutional neural network and its unique feature to detect new patterns. Recent studies prove that this is one of the best approaches in supervised learning mechanisms.

(A) *There are 12-layered convolutional neural networks (CNNs), three convolutional layers, and three max pooling layers that are accompanied by a fully* *connected neural network layer.* Figure 1*explains the individual layers—convolutional layer*.

There are three convolutional layers used with ReLU as activation function and with filters of size (200,100,100). This is using with a kernel size of (2,2) for the first conv-layer and (1,1) for the next two layers.

(B) *Pooling layer*

(C) *The max pooling layer is used with a kernel size of (2,2), and each layer of pooling uses the ReLU as an activation function—fully connected layer*.

The fully connected layer has four dense layers; the input for the first fully connected layer was flatten, and then return to the first layer, and each layer has ReLU as the activation function except the last output layer which delivers activation as softmax, which returns the probability in the range of zero to one.

4 Methodology

The change is detected in between two images which are taken from a satellite image [8]. Two land-cover images are taken from different timestamps 2013 and 2015. Figure 3 represents the methodology of the proposed model (Fig. 4).

(A) *The two classification maps such as classification map1*map *2 are obtained using the CNN classifier. A simple difference operation is performed on these two maps and recognized the change map with 97% accuracy* [9].*Training*

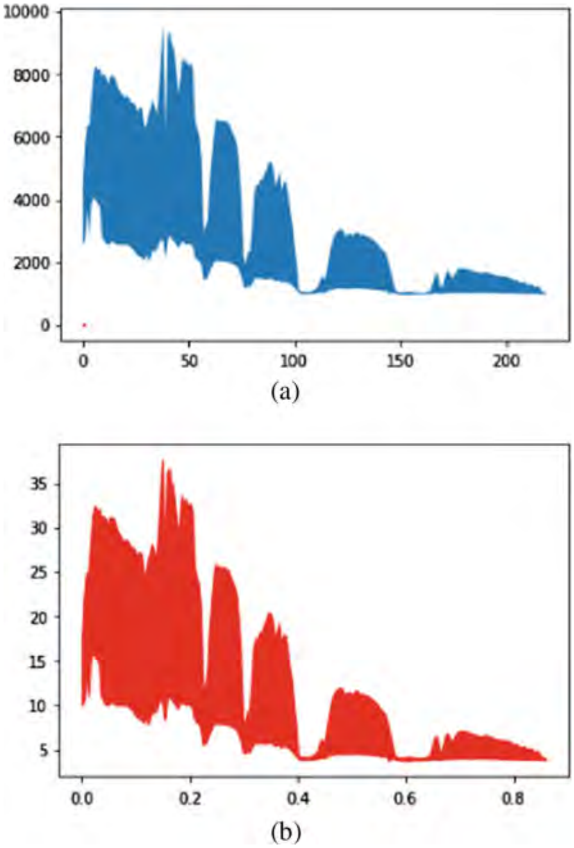
The training was performed on 8199 samples, and the test size was 2050. The model was trained for 20 epochs with a learning rate of 0.01. It took 14 s to complete each iteration, i.e., an average time of 5 min to get a training accuracy of 99% and a test accuracy of 99.66%. The learning curve is shown in Fig. 5 and the loss curve in Fig. 6.

5 Results

Figure 8 describes the results and the change map obtained from the two images that are shown in Fig. 7. Consider these two images they are inputs taken through satellite using Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor data which is temporally varied images that means considered in two different timestamps that are 2013 and 2015 from the area Hermiston located in Oregon United States. The width, height, and band of the channel (or) channel color are (390,240,242), respectively.

Both the images are classified using a classifier that is by adopting the convolutional neural networks (CNN) algorithm which produces two classification maps by undergoing the process of CNN; they are conv-layer, batch normalization, max pooling, dropout, flatten, and dense. Figure 7a, b describes the CNN process and the

Fig. 1 Scale graph of the data **a** before normalization **b** after normalization



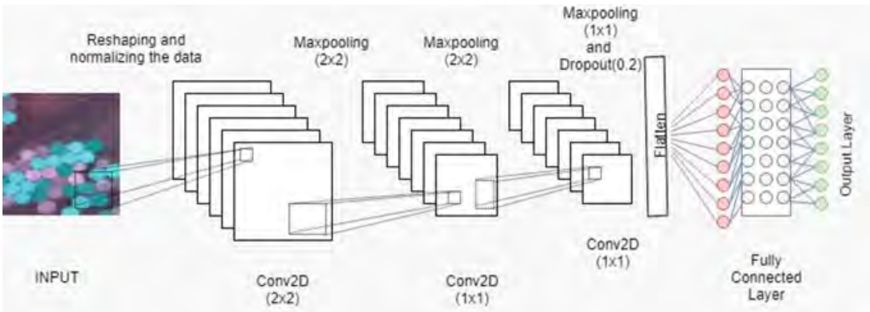


Fig. 2 Different layers of the convolutional neural network

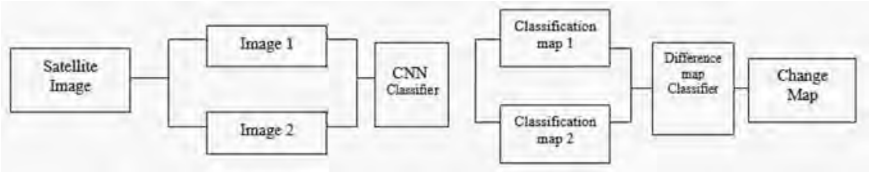


Fig. 3 Architecture diagram of the proposed change detection method

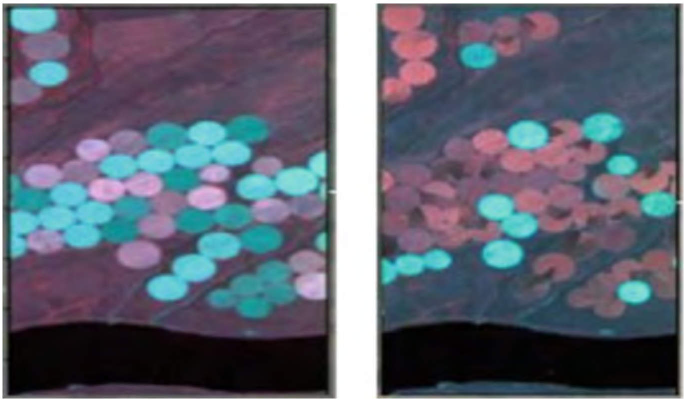


Fig. 4 Satellite images used for change detection

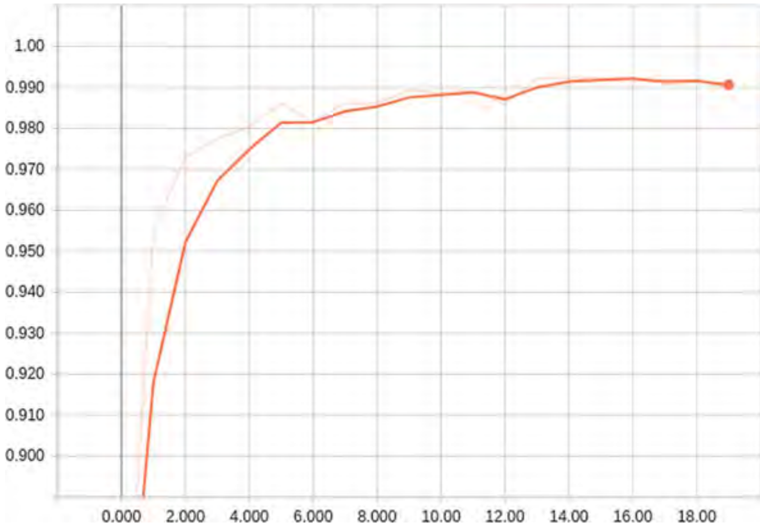


Fig. 5 Learning curve

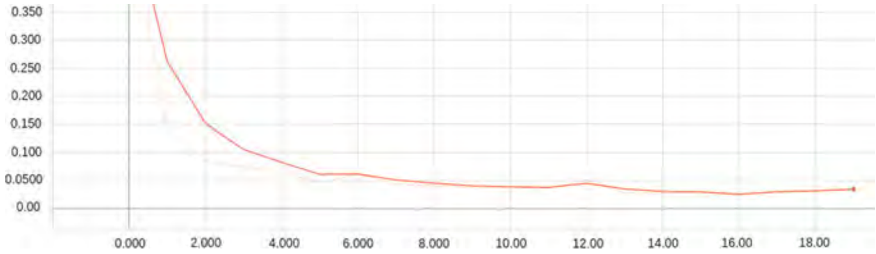


Fig. 6 Loss curve

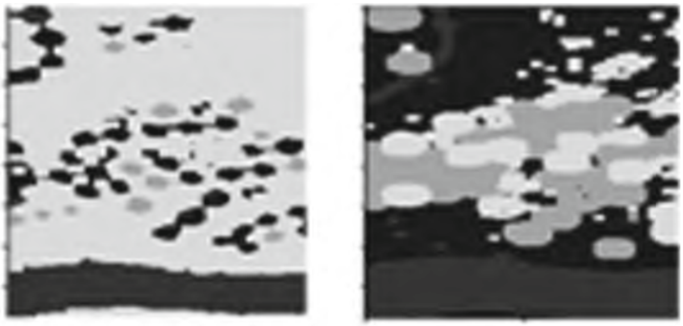


Fig. 7 Classification maps **a** 2013 **b** 2015

Fig. 8 Change map

two classification maps. The height and width of these images will be (390,200); no band channel is analyzed.

To generate the change map results, it is necessary to find the difference between the two classification maps taken from two different timestamps which are shown in

Table 1 Changes notified with number of pixels table-type styles

Class	Year 2013	Year 2015	Difference
1	31,456	8269	23,187
2	14,967	11,122	3845
7	16,239	2236	14,003
9	15,338	5373	−41,035

Fig. 7. In the figure, the yellow area indicates the changed area and purple indicates the unchanged area.

The following Table 1 illustrates the changes in the classes. The first column specifies the classes, and the remaining columns are the exact count of class labels with difference in the last column.

6 Conclusion and Future scope

The technique of using a deep convolutional neural network model for generating classification maps is optimal for this use case. The generated classification maps for two temporal images are parsed for generating change maps. This kind of approach is most beneficial for detecting the changes in the known classes. And this method has been widely used in applications such as agriculture, forestry, and mineralogy.

The change map obtained was strong enough to identify the total changes in the complete spatial resolution. The class-wise changes are attained from the convolution neural networks output layer. Further, the study can be made in developing a hybrid convolutional neural network, which can classify two-satellite imagery and identify the difference in the classification maps by using the dual pipeline method.

References

1. Lu D, Mausel P, Brondizioand E, Moran E (2004) Change detection techniques. Int J Remote Sens 25:2365–2407
2. Coppin PR, Bauer ME (1996) Digital change detection in forest ecosystems with remote sensing imagery. Remote Sens Rev 13:207–234
3. Singh A (1989) Digital change detection techniques using remotely-sensed data. Int J Remote Sens 10:898–1003
4. Blaschke T Towards a framework for change detection based on image objects. Göttinger GeographischeAbhandlungen, p 113
5. Zhang H, Gong M, Zhang P, Su L, Shi J (2016) Feature-level change detection using deep representation and feature change analysis for multispectral imagery. IEEE Geosci Remote Sens Lett 13
6. Doña C (2014) Empirical relationships for monitoring water quality of lakes and reservoirs through multispectral images. IEEE J Select Topics Appl Earth Observ Remote Sens 7(5)

7. Baumgardner MF, Biehl LL, Landgrebe DA (2015) 220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3. Purdue University Research Repository. https://doi.org/10.4231/R7RX991C
8. Huang C (2016) Surface water change detection using change vector analysis IEEE
9. Gong M,Zhan T, Zhang P, Miao Q (2017) Superpixel-based difference representation learning for change detection in multispectral remote sensing images. IEEE Trans Geosci Remote Sens

An Efficient Ant Colony-Based Self-adaptive Routing Strategy for IoT Network Model

K. M. Baalamurugan, D. Vinotha, and Premkumar Sivakumar

Abstract The intellectual device is known as the Internet of Things (IoT) not only offer services, but also facilitate the allocation of heterogeneous resources and diminish resource utilization based on service time. This is considered as a crucial fact toward the large-scale environments. To make the service effectual with the proper response, the request for the route has to be analyzed simultaneously and to offer better global solutions. Therefore, this work anticipates a self-adaptive ant colony optimizer (SA-ACO) for IoT environment. This SA-ACO model relies on the provisioning of a global solution with the self-adaptive characteristics and modeled to select the finest attributes. The node clustering is dependent on the population and to provide the functionality over multi-directional and the resourceful realization of self-adaptive route searching. The memory utilization is analyzed with the selection of nominal solutions to acquire promising results. To validate the competency of the anticipated model and to measure the effectiveness, this work can be analyzed over the real-time environment in future. The simulation is performed in the MATLAB environment that the anticipated model can acquire the finest solution with strong exploration functionality, and superior performance is evaluated with the comparison of the model with PSO and RR, respectively.

Keywords Internet of Things (IoT) · Self-adaptive ant colony optimizer (SA-ACO) · Artificial intelligence (AI)

1 Introduction

The advancements of the Internet of Things (IoT) have entered a progression of intelligence industries. Recently, the energy savings, configuration optimization,

K. M. Baalamurugan (✉)
School of Computing Science and Engineering, Galgotias University, Greater Noida, UP, India
e-mail: k.baalamurugan@galgotiasuniversity.edu.in

D. Vinotha · P. Sivakumar
Department of Computer Science and Engineering, Annamalai University, Chidambaram, TN, India
e-mail: vinothacse54@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_60

and environmental protection toward IoT resources have turn toward diverse issues that has to be resolved [1]. There are diverse artificial intelligence approaches for resource scheduling that are modeled for a certain application and not applicable for resolving IoT service crisis examined [2]. For provisioning the service, this is based on the complete IoT layout for service system. The resource optimization allocation problem based on service-oriented network collaborative model is extremely complex [3]. This comes under the typical NP-hard combinational optimization crisis. The fact is that how to reduce the resource utilization, and reduce the response time. Subsequently, how to multi-optimal services in numerous candidate set is chosen to fulfill the objectives [4]. Therefore, this is a challenging factor in multi-objective optimization crisis.

Various investigations have tried to resolve the multi-objective service selection problem in internet services. The author in [5], depicted that the approximation strategy for multi-objective driven service selection. The author in [6], provided research on multi-objective optimization to reach the QoS. This model anticipated a parent set model for QoS aware service composition. The author in [7], anticipated a model that assists decision making in predicting effectual, QoS optimization service with clustering. The prevailing research model is executed with an adaptive service composition acquired from the immune system. Moreover, the above-mentioned model is concentrated base don QoS.

The features with IoT services are heterogeneity, large-scale, dynamic, and unreliability in nature that shows differentiation in web services. An effectual confront to resolve the IoT service domain is the construction of effectual service selection procedures for optimal management of both QoS and energy. This crisis is crucial over large-scale IoT environments that comprise of a huge amount of distributed entities. The author in [8] depicts that IoT is a paradigm where real-world physical things that are associated with the services and internet provisioning via the computational devices. The three QoS scheduling approach for service-based IoT was anticipated by Sherubha and Mohanasundaram [9]. The sensing-based service model is considered to be constructed on the top of IoT services and infrastructure. The author in [10], allocated services to interface heterogeneous resources and generates an optimal solution for computation for a hard problem.

Moreover, to determine the coordination among the environment and population, some population are based on evolutionary procedure, co-evolutional methods are initiated toward the immune-optimization model, and superior outcomes are attained for resolving the combinatorial optimization crisis. The cooperative and competition model are two essential factors over the multi-objective optimization crisis. This model relies on the attainment of higher success in resolving single-objective optimization issues. The author in [11], provided an evolutionary algorithm for a multi-objective process which is competent in preserving the archive diversity by the huge operator and dynamic sharing. The author anticipated competitive evolutionary model. There are diverse sub-populations that are optimized for decision variables, respectively. The differences among the models are mapped in association with every sub-population and decision variables that are not centralized; however, it is demonstrated with the superior outcomes.

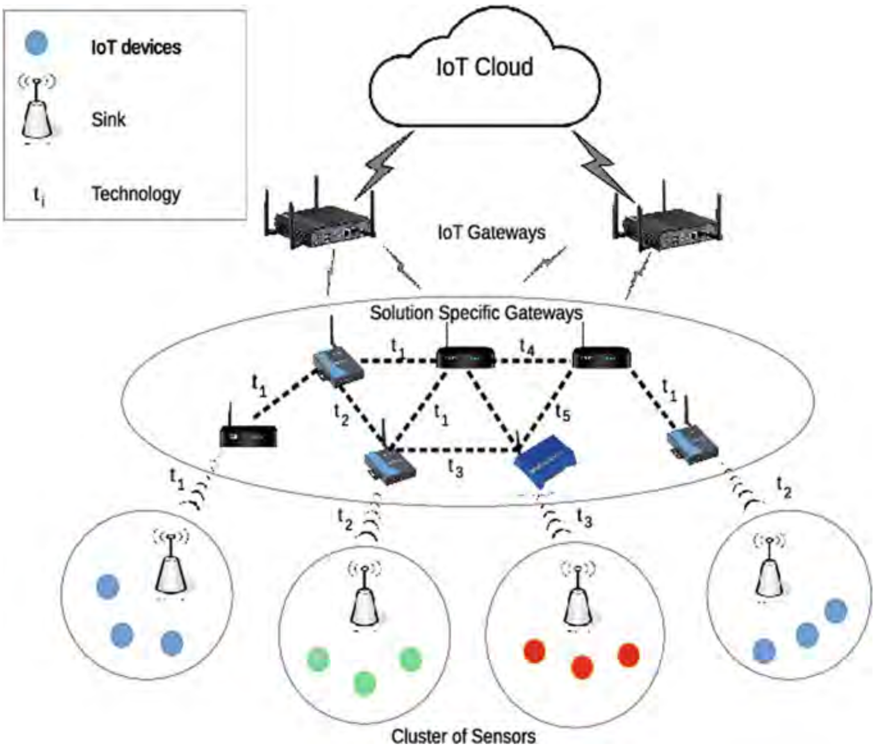


Fig. 1 Generic view of IoT-based routing model (*Source* Angelakis et al. [2])

The significant contributions of this work are:

1. With the inspiration from the meta-heuristic model, an adaptive approach is modeled and designed with sub-population to attain a global solution that assists effectual interactions between the populations. This is also based on global optimization solutions.
2. The self-adaptive nature of ant to gather the food and to reach the source is examined and the evolution of ant population, that is resourcefully resolved with the selection of non-dominated outcomes.
3. The statistical and the clustering model with the heuristic process, varying components are examined, that produces the operations more purposefully and directionally.

The sections of the paper are as follow: Sect. 2 depicts the background model of prevailing approaches, Sect. 3 is the anticipated self-adaptive ant colony optimization method for routing. Section 4 shows the numerical results and the validation, and Sect. 5 draws the conclusion and future research directions.

2 Related Works

Recently, evolutionary algorithms are turned to be a main-stream model for performing various researches in the soft computational field. The author in [11] provided a comprehensive analysis of the modern multi-objective evolutionary process. However, the multi-objective optimization process relies on the artificial immune system that comprises of multi-objective problem, constrained problem, artificial network model for the vector-based immune system, non-dominant immune model, and so on. However, based on the analysis, the author in [11] anticipated a genetic model and provides various strategies for choosing the vaccines and develop the immune operators. The author in [12] provided a biological notion with the vaccine usage for promoting exploration in search space.

In recent times, a novel immune model for MOO is anticipated by Sherubha and Mohanasundaram [12]. Here, the de-generation for predicting the clonal selection process for multi-modal optimization is examined. The novel multi-class clustering approach is based on the maximal margin-based clustering approach, and an evolutionary model is anticipated. However, this model concentrates on the local searching process. The local improvement operator with (convergence acceleration operator) was initiated, and the hill-climbing with sidestep is modeled for local searching, where a novel multi-objective optimization structure relies on non-dominated searching and local sorting was anticipated. The author [13] used a novel ranking approach termed as global margin ranking with position deployment of every individual in objective space to attain marginal dominance all through the population.

In [14], the anticipated novel co-evolutionary model that relies on an elite approach, where every elite individual are utilized for guiding the search function. The author in [14], modeled a structure termed as hyper-MOO. The sub-population size was modified based on the appropriate MOO performance. The author used multiple sub-population and utilized diverse statistical and clustering techniques to assist the new population generation and local searching. However, certain investigators concentrate on collaboratively decomposing multi-objective optimization issues. Additionally, it is inspired by the mammalian endocrine system with an artificial endocrine model for managing power that is modeled with robotic analysis. This mechanism is anticipated to control the cooperative evolution among the particles.

In [15], has utilized the evolutionary model and diverse populations for multi-objective optimization issues. With the immune-based model, there are diverse sub-population are evolved with diverse evolutionary approaches. The immune co-evolutionary model possess two phases that are modeled to acquire the optimal partitioning balance. The author in [16], provided an evolutionary model for dealing with the garmenting problem that initiates the distance affinity and dominance affinity. The author anticipated evolutionary methods for handling routing issues. The investigator used two diverse sub-populations that are co-operatively evolved with the evolution process to attain superior global optimality for radial basis function estimation with neural networks.

3 Methodology

This section discusses the self-adaptive ant colony optimization approach used for handling effectual routing over IoT environment. The flow diagram of SA-ACO is given in Fig. 2.

In general, ACO is a popular meta-heuristic approach that is inspired by the ant's ability to determine the shortest path among the food sources and the colonies. This foraging nature for collecting the food source is utilized to resolve the diverse optimization issues like the prediction of the shortest path to find the corresponding neighborhood for collecting the food. The ants pretend to communicate with one another for realizing the chemical signals termed as pheromones. While moving back to their location, the ant releases a certain amount of pheromone with path proportional to the

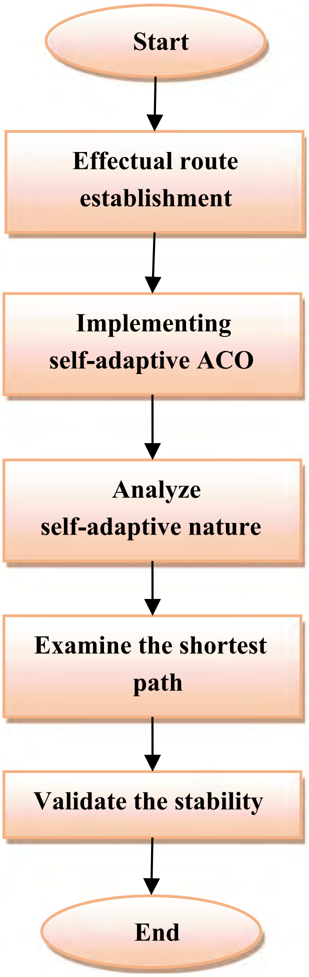


Fig. 2 Flow diagram of self-adaptive ACO

quality and quantity of sources. However, all the other ants possess superior probability for the successive paths with higher pheromone concentrations. Moreover, all ants have to follow the shortest path for the source to the home. Consider, 'm' ants and 'n' probable paths where every ant chooses the probable path based on the highest concentration over the pheromones with probable paths.

The anticipated self-adaptive ACO approach is considered to be an effectual model with heuristic nature for various NP-hard-based research issues that includes job-shop scheduling and traveling salesman problems. Moreover, the anticipated ACO model is used for scheduling issues over the cloud and includes scheduling over the VM on the cloud environment and the scheduling tasks for the VMs to attain the load balancing tasks in VM and reduces the response time during task functionality. Also, the ACO algorithm is utilized for scheduling various IoT tasks over the cloud. Added, this algorithm is utilized for providing the deadline-aware task scheduling process over the IoT-based computational infrastructure. The anticipated model concentrates in improving the profits of service providers while determining the fulfillment of deadline-based IoT task constraints.

While performing the route identification tasks with the help of connected nodes with an ultimate target to reduce the response time, the ant selected the node for improving the stability during task assignment with the probability which is expressed as in Eq. (1).

$$P_{ij}^{k}(t) = \frac{\left(\tau_{ij}(t)\right)^{\alpha}\left(\eta_{ij}(t)\right)^{\beta}}{\sum_{s}(\tau_{is}(t))^{\alpha}(\eta_{is}(t))^{\beta}} \tag{1}$$

Here, α and β are heuristic constants, $\alpha \geq 0$ is a heuristic factor that manages the consequences of pheromone quantity with $\beta \geq 1$ is a heuristic parameter that deals with the significance of route allocation. $\eta_{ij}(t)$ is a heuristic model that specifies the routing stability and is computed with Eq. (2).

$$\eta_{ij}(t) = \frac{\text{load}_j}{R_{ij}} \tag{2}$$

Here, R_{ij} is evaluated and load_jspecifies the load over the node and computed with the equation mentioned above. However, R_j increases when load_jreduces and $\eta_{ij}(t)$ reduces. As an outcome, the routing stability of the nodes over the network is evaluated. $\tau_{ij}^{k}(t)$ is pheromone trail quantity with allocating the task to nodes by route prediction with t time and $\tau_{ij}^{k}(t+1)$ is pheromone trail for allocating tasks for all ants during the iteration $t+1$ which is evaluated as in Eq. (3).

$$\tau_{ij}^{k}(t+1) = (1-\rho)\tau_{ij}^{k}(t) + \rho\Delta\tau_{ij}^{k}(t) \tag{3}$$

Here, $\Delta\tau_{ij}^{k}(t) = \frac{1}{R_{ij}}$ and ρ is constant that specifies the pheromone evaporation rate that stimulates the evaporation effect of pheromones during every step. In addition, the pheromone trail is updated globally with all ants with probable task determination over the complete iteration process. This global updation is expressed as in Eq. (4).

$$\tau_{ij}^{k}(t+1) = \left(1 - \rho_g\right)\tau_{ij}^{k}(t) + \rho_g \Delta\tau_{ij}^{k}(t) \tag{4}$$

Here, $\Delta\tau_{ij}^{k}(t) = \frac{1}{L_{\text{best}}}$, L_{best} is the finest route identified with ρ_g global evaporation rate. The initial pheromone value is computed with Eq. (5).

$$\tau_{ij}^{k}(0) = \frac{R_{ij}}{R_{\text{Average}}} \tag{5}$$

Algorithm

1. Parameter initialization
2. Compute R_{ij}
3. Initialize pheromone matrix
4. Compute $\eta_{ij}^{k}(t)$
5. While $iteration \leq N_{iteration}$ do
6. Position the ant to initiate the route randomly
7. For $ant = 1 to N_k$ ant = 1toN$_k$ do
8. For $S_i = 1$ s$_i$ = 1 to nodes
9. Compute P_{ij}
10. Add the stability route to the routing table
11. End for
12. Revise $\tau_{ij}^{k}(t+1)$
13. End for
14. Evaluate the prior solution and update the best]
15. If solution is best then
16. Revise $\tau_{ij}^{k}(t+1)$
17. End if
18. $iteration = iteration + 1$;
19. endwhileend while
20. Clear routing table

The anticipated SA-ACO model is utilized to determine the effectual process of measuring route stability over the IoT environment. The nodes have to fulfill the QoS constraints with appropriate response time by determining its self-adaptive nature. The load and the service time of the nodes are measured with the initialized parameters. The heuristic model determines the maximal amount of iterations and initializes the total amount of nodes over the network with the productivity rate of all sensors and service rates.

The successive steps are applied for measuring the response time of nodes with the provided equations. The initial pheromone and task allocation stability for all nodes are given in the equation above. With all iteration, the ant has to offload the computational measure of the nodes with probability evaluation. The node selection is done with the clustering method. This method evaluates the solution probability and the cumulative probabilities that are assessed in an ascending manner. The random numbers are produced with the cumulative probability. At last, the appropriate solutions are generated with a random number. This is validated as the effectual selection of resourceful solutions. When the ant completes the routing process over the connected nodes, the local pheromone is updated. With all iterations, the algorithm needs to revise the global pheromone. The iteration is completed until the maximal amount of iterations are attained.

4 Numerical Results

The MATLAB simulation environment is run over PC with Intel core Pentium process with 2.60 and 8 GB RAM. With this experimentation, various parameter settings are given below. The parameter settings are provided by performing numerous primary experimentations.

The experimentation is performed to measure the anticipated model's ability for stabilized routing establishment with average response time, routing imbalance, and node connectivity, respectively. The imbalance degree is expressed as in Eq. (6).

$$\text{Degree of imbalance} = \frac{Max\left(R_j\right) - Min\left(R_j\right)}{R_{\text{Average}}} \tag{6}$$

The response time-based SD is computed with the route distribution to establish the stability, which is expressed as in Eq. (7).

$$\text{SD} = \sqrt{\frac{\sum_j \left(R_j - R_{\text{average}}\right)^2}{N_{\text{nodes}}}} \tag{7}$$

From Tables 1, 2 and 3, it is observed that the metrics like nodes imbalance degree, average node connectivity, and route response are measured, respectively. The anticipated model is evaluated and compared with PSO and RR where the performance of SA-ACO is superior in contrast to other models. Similarly, Figs. 3, 4 and 5 depicts the graphical representation of the anticipated approach. This model works effectually and improves the stability rate of the network.

Table 1 Route Response

No of IoT nodes	RR	PSO	SA-ACO
250	89	78	75
500	90	82	77
750	93	83	79
1000	96	89	80
1250	98	93	81
1500	100	95	85
1750	108	98	87
2000	110	99	90

Table 2 Average node connectivity time

No of IoT nodes	RR	PSO	SA-ACO
250	92	89	81
500	98	90	82
750	105	92	83
1000	109	95	89
1250	115	100	93
1500	120	100	95
1750	128	105	98
2000	130	107	99

Table 3 Degree of route imbalance

No of IoT nodes	RR	PSO	SA-ACO
250	0.25	0.23	0.22
500	0.2	0.15	0.07
750	0.1	0.1	0.05
1000	0.09	0.1	0.05
1250	0.1	0.13	0.06
1500	0.5	0.15	0.08
1750	0.8	0.18	0.09
2000	0.18	0.25	0.2

5 Conclusion

This work concentrates on constructing a multi-objective model among the service and the multiple route request generated from the service provisioning environment. This work makes use of optimal request generation and imbalance encountered by the routing nodes, as the anticipated self-adaptive ant colony optimizer model utilizes the adaptive nature to deal with the routing issues over the heterogeneous environment.

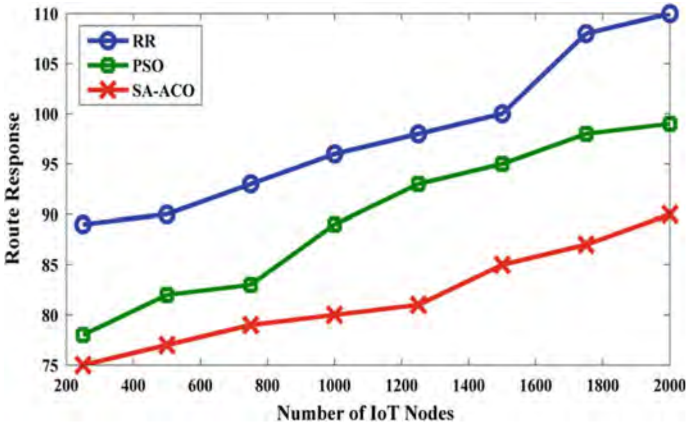


Fig. 3 Route response computation

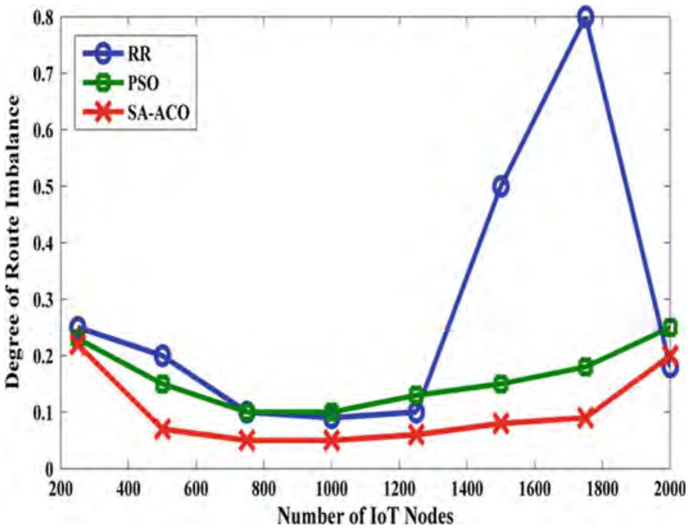


Fig. 4 Degree of route imbalance

This model is provided with a technique to deal with the superior population to identify the source and to map the route for the followers. Therefore, a stronger routing model is attained. The nodes are clustered to be adopted over the diverse network environment and assist the population to find the source in a resourceful manner. It is also provided to assist in route searching and to realize its characteristics. The simulation outcomes depict that the anticipated model can acquire the best, stronger ability and gives superior performance. In the future, hybrid optimization is applied to validate the routing performances

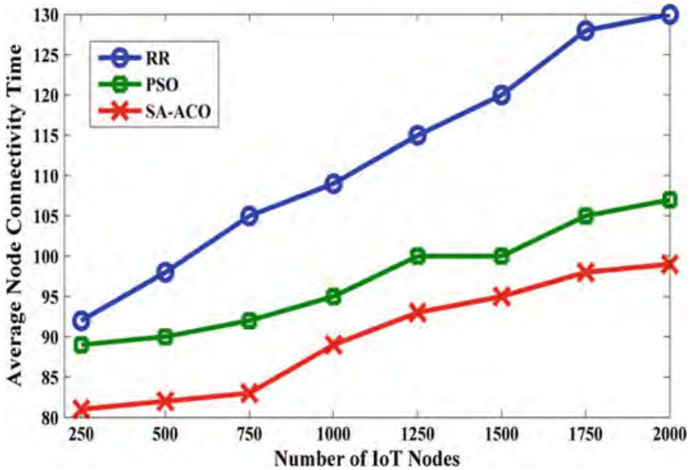


Fig. 5 Average node connectivity time

References

1. Branke J, Nguyen S, Pickardt CW, Zhang MJ (2016) Automated design of production scheduling heuristics: a review. IEEE Trans Evol Comput 20(1):110–124
2. Angelakis V, Avgouleas I, Pappas N, Fitzgerald E, Yuan D (2016) Allocation of heterogeneous resources of an IoT device to flexible services. IEEE Internet Things J 3(5):691–700
3. Chen Y, Huang JW, Lin C, Hu J (2015) A partial selection methodology for efficient QoS-aware service composition. IEEE Trans Serv Comput 8(3):384–397
4. Saravanan T, Nithya NS (2019) Modeling displacement and direction aware ad hoc on-demand distance vector routing standard for mobile ad hoc networks. Mob Netw Appl 24(6):1804–1813
5. Ding YS, Jin YL, Ren LH, Hao KR (2013) An intelligent self-organization scheme for the Internet of Things. IEEE Comput Intell Mag 8(3):41–53
6. Jin XN, Chun SJ, Jung J, Lee K-H (2017) A fast and scalable approach for IoT service selection based on a physical service model. Inf Syst Front 19(6):1357–1372
7. Saravanan T, Nithya NS (2018) Energy aware routing protocol using hybrid ANT-BEE colony optimization algorithm for cluster based routing. In: 4th IEEE ınternational conference on computing communication and automation (ICCCA), pp 1–6
8. Xu N, Ding YS, Ren LH, Hao KR (2018) Degeneration recognizing clonal selection algorithm for multimodal optimization. IEEE Trans Cybern 48(3):848–861
9. Sherubha P, Mohanasundaram N (2019) An efficient ıntrusion detection and authentication mechanism for detecting clone attack in wireless sensor networks. J Adv Res Dyn Control Syst 11(5)
10. Li L, Wang WL, Xu XL (2017) Multi-objective particle swarm optimization based on global margin ranking. Inf Sci 375:30–47
11. Wang JH, Zhang WW, Zhang J (2016) Cooperative differential evolution with multiple populations for multiobjective optimization. IEEE Trans Cybern 46(12):2848–2861
12. Sherubha P, Mohanasundaram N (2019) An efficient network threat detection and classification method using ANP-MVPS algorithm in wireless sensor networks. Int J Innov Technol Explor Eng (IJITEE) 8(11). ISSN: 2278-3075
13. Sherubha P, Mohanasundaram N, Graph based event measurement for analyzing distributed anomalies. In: Sensor networks, Sadana Academic Proceedings in Engineering Sciences, Springer (Accepted for publication)

14. Sherubha P, Mohanasundaram N, Sasirekha SP (2019) Clone attack detection using random forest and multi objective cuckoo search classification. In: IEEE ınternational conference on communication and signal processing, April 4–6, 2019, India
15. Shang RH, Dai KY, Jiao LC, Stolkin R (2016) Improved memetic algorithm based on route distance grouping for multi-objective large scale capacitated arc routing problems. IEEE Trans Cybern 46(4):1000–1013
16. Yao GS, Ding YS, Jin Y, Hao KR (2017) Endocrine-based coevolutionary multi-swarm for multi-objective workflow scheduling in a cloud system. Soft Comput 21(15):4309–4322

Optimizing Node Coverage and Lifespan of Wireless Body Area Network Using Hybrid Particle Swarm Optimization

S. Selvaraj and R. Rathipriya

Abstract Wireless body area network (WBAN) is one of the emerging wireless sensor networks for medical applications and treatments with the constrained power of numerous tiny sensors. Nowadays, many of these medical application researches focus on the low-power propagation sensing units in the mesh of the health monitoring system. This article expounds on the applicability of a discrete version of a popular benchmark swarm intelligence algorithm PSO called discrete particle swarm optimization (DPSO) and its hybrid version for energy-optimized WBAN using node coverage. DPSO-based WBAN model optimizes the node coverage for uninterrupted connectivity over the longest possible network life. The network simulator NS-2, for creating WBAN nodes, installation, connection, data transmission in the network environment is used. The simulation results showed that the proposed WBAN model has better performance in the terms of node coverage and network lifespan.

Keywords Body area sensor · Wireless sensor network (WSN) · Discrete particle swarm optimization (DPSO) · Genetic algorithm (GA)

1 Introduction

Recently, the WSN adopts any kind of ground-environment, likewise the sensor unit information with the real-time data to be used for the coverage, deployment, propagation [1, 2]. These sensor units, based on the applications it may have some prior knowledge. Generally, it consists of low-cost and low-power sensing units in small size and able to works broad range, and these applications are used as military, ground, monitoring, health devices, etc. [1–4].

Swarm intelligence is the meta-heuristic method and it gives an optimum solution to result in significant improvement in WBAN by optimizing network energy, coverage, and the level of lifespan increments [5, 6]. Figure 1 shows the WBAN structure using medical sensor data propagation. Table 1 describes the details like bandwidth and data rates of medical frequency bands.

S. Selvaraj · R. Rathipriya (✉)
Department of Computer Science, Periyar University, Salem, Tamil Nadu, India
e-mail: rathipriyar@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_61

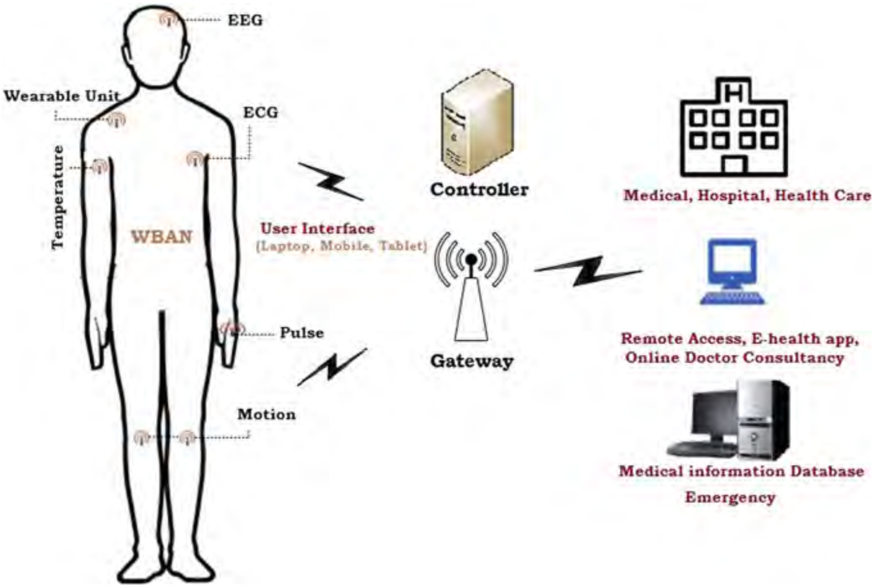


Fig. 1 Wireless body area network structure

Table 1 Worldwide medical frequency bands using bandwidth and data rates

Frequency band	Class	Bandwidth (kHz)	Channel	Data rate (kbps)
398–406	MICS (WW)	300	10	57–455
418–448	WMTS (Japan)	320	12	56–187
862–869	WMTS (Europe)	400	14	75–607

MICS Medical Implant communication service
WMTS Wireless medical telemetry service
ISM Industrial scientific medical

The full article is organized in the following manner. The introduction of WBAN is discussed in Sect. 1 and the review literature of WBAN, multi-target optimization, and swarm intelligence are explained in Sect. 2. Section 3 describes the methods and materials needed for the research work. The experimental result of the proposed work is elaborated in Sect. 4. Section 5 concludes the article with further enhancement.

2 Review Literature

This section provides the review literature needed for the research work. In [1–4], standard PSO formed as a baseline optimization method for performance testing of improvements to the technique as well as to represent PSO to a broader optimization

Table 2 Environment parameters for WBAN

Wireless technology	Illness	Parameters
WSN, AUR, Bluetooth	General illness	Cardiac output, pulse, the position of patient, meter, temperature, blood pressure, pulse-oximeter, breathing, oxygen consumption, galvanic skin reaction, ECG, movement and breath, proximity to other patients, blood glucose [1–5]
RFID, FOG Computing, GPS, Wi-Fi, WBAN	ICU patient monitoring, Dengue, chronic illness	ECG, oxygen levels, temperature, blood pressure, pulse [1–5]

community. In the second stage, the best classifier variants improved in terms of attribute type support and temporal complexity. These works defined and addressed two data-related issues that could affect the efficiency of particle swarm optimization: high-dimensional data sets, mixed attribute data, and proposed solution to each of these issues including recent improvements by a PSO algorithm, with the latest developments that helped to improve performance on standard measures to extend the first particle swarm optimization [5]. Table 2 showed the parameters used and the type of illness of the most commonly used wireless technology [9].

The solution transfers from one population to another through certain assessment procedures in the population-based method. GA uses a genetic system such as selection, mutation, and crossover, while for the assessment process, PSO uses swarm behavior such as updating location and speed [6].

3 Methods and Materials

This section describes the required methods and materials for the proposed DPSO and its hybridization for optimizing the node coverage and network lifespan. A combination of optimization algorithms is used in the many works to optimize the parameters for the numerous problems of various domains.

3.1 Genetic Algorithm

The genetic algorithm creates a paradigm for evolution simulating genetically determined Darwinians and the normal cycle of elimination. The transition procedure ensures that the entire population moves to the global optimum through better chromosomes. Mutation operations maintain the diversity of the population and prevent

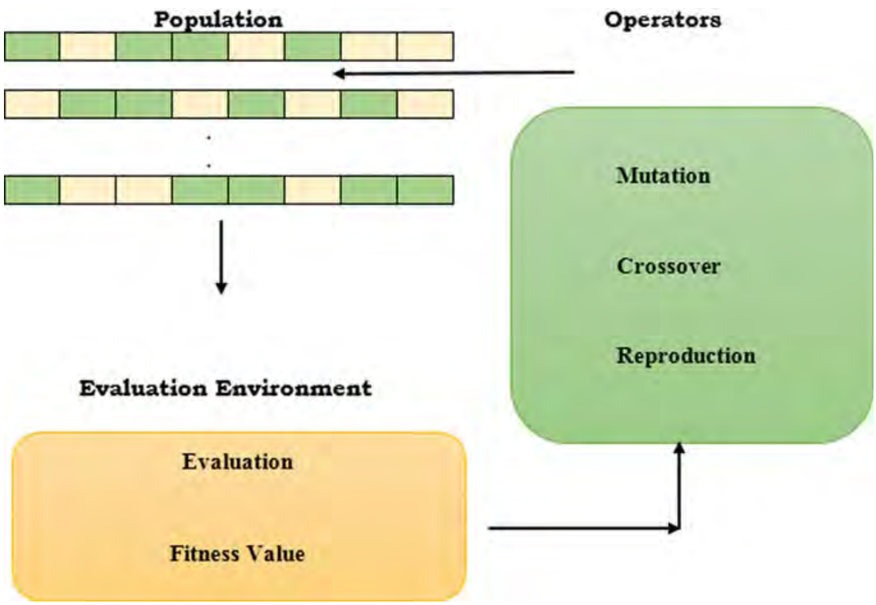


Fig. 2 Flow of GA

the population from dropping in the optimal location. Figure 2 shows the flow of GA [6, 7].

3.2 Discrete PSO

To replace the velocity concept, the study presented and modified several operators based on the swap operator and the swap series concept. It is known to be a special DPSO to solve discrete problems. DPSO begins the process of searching using a randomized group of particles and the flow presents in Fig. 3. Then, particles are used to defined pbest and gbest [8].

The critical difference of DPSO is that the particulate velocity and position. The variation of probabilities characterizes these, and the particles generated by an integer in [0, 1]. A particle flies, therefore, in a search space confined to zero and one. The velocity is limited, and the interval has [0, 1].

The sigmoid function transformation as $S\ (v_i\ (t+1))$ is shown in the Eq. 1 can be used to limit the speed.

$$S_{sig}(V_i(t+1)) = \frac{1}{1+e^{(V_i(t+1))}} \tag{1}$$

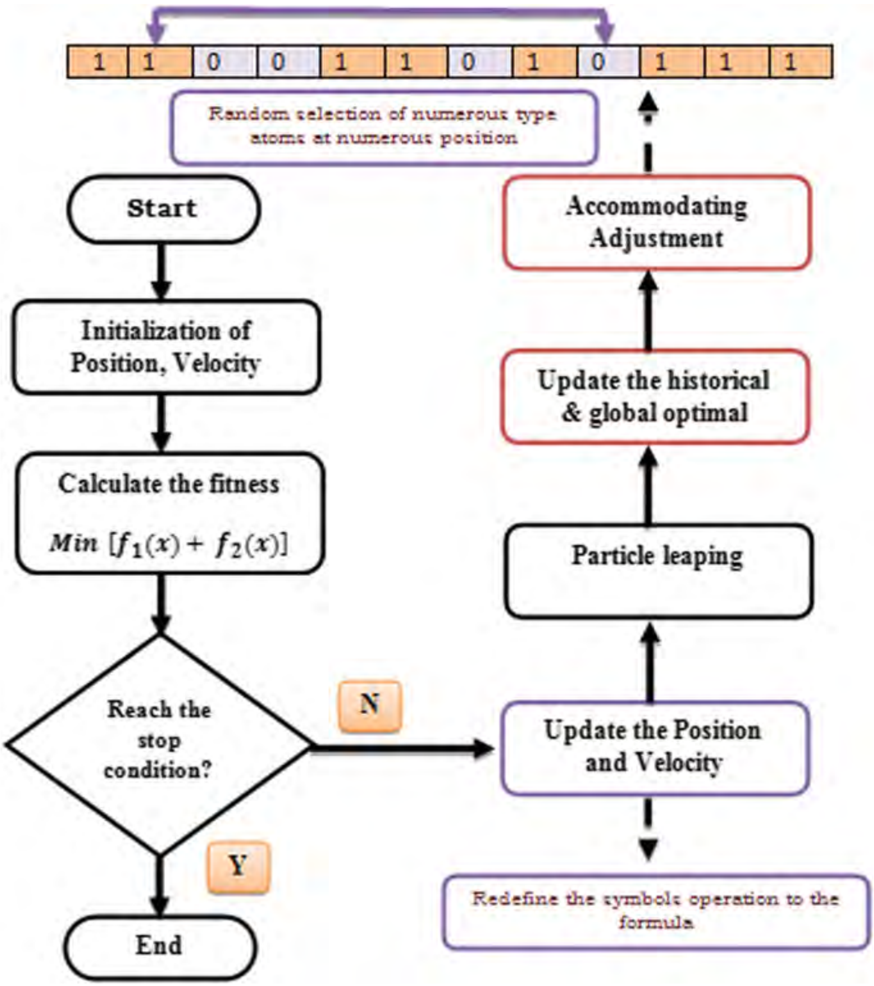


Fig. 3 Flow of DPSO

The updated position is determined using Eq. 2, where r_3 is assigned to a uniform random value and the range is [0, 1].

$$x_{iD} = \begin{cases} 1 \text{ if } r_3 < (v_i(t + 1)) \\ 0 \qquad \text{otherwise} \end{cases} \tag{2}$$

3.3 Simulated Annealing

An analogy to annealing of ideal crystals in thermodynamics is based on a simulation annealing (SA). The algorithm is designed to maximize the molecules' thermal movement at fixed temperature simulations. Thus, for instance, the term 'temperature' used to name a crucial control parameter is derived from thermodynamical SA. In SA operations, the temperature parameter is high continuously. Rapid cooling creates defects that do not achieve sufficient energy levels in the crystal structure [8–11]. The mechanism imitates this process while optimizing the parameters in SA. The possibility of such approval is dependent on temperature and should be negligible at low temperatures. A significant function of SA's approach is an opportunity to consider a momentarily bad option [12–15]. Figure 4 illustrates the complete workflow of the SA algorithm for better understanding.

4 HPSO-Based Energy Model for Extended WBAN Lifespan

The major drawback of the DPSO is that in high-dimensional space, it is easy to collapse into local optimum and has a poor convergence rate in the iterative method. To overcome the above-said problem, DPSO is hybridized with SA to develop a hybrid DPSO algorithm (namely HPSO) to avoid local optima. The HPSO-based energy efficiency model is proposed for WBAN. It is a population-based optimization method. Algorithm 1 describes the HPSO-based energy model for extended WBAN lifespan [13–15].

The WBAN used in this study has three different layers namely, the sensing layer split into an outdoor ground sensor, indoor/room sensor, and inside-wearable WBAN sensors [16–20]. Network Layer has the controller to sense and control all sensing units like a gateway of WBAN. This layer is connected to public, private, industry clouds, and storage. Finally, it has the interface layer responsible for data aggregation, visualizations, and the analytical process of body sensors.

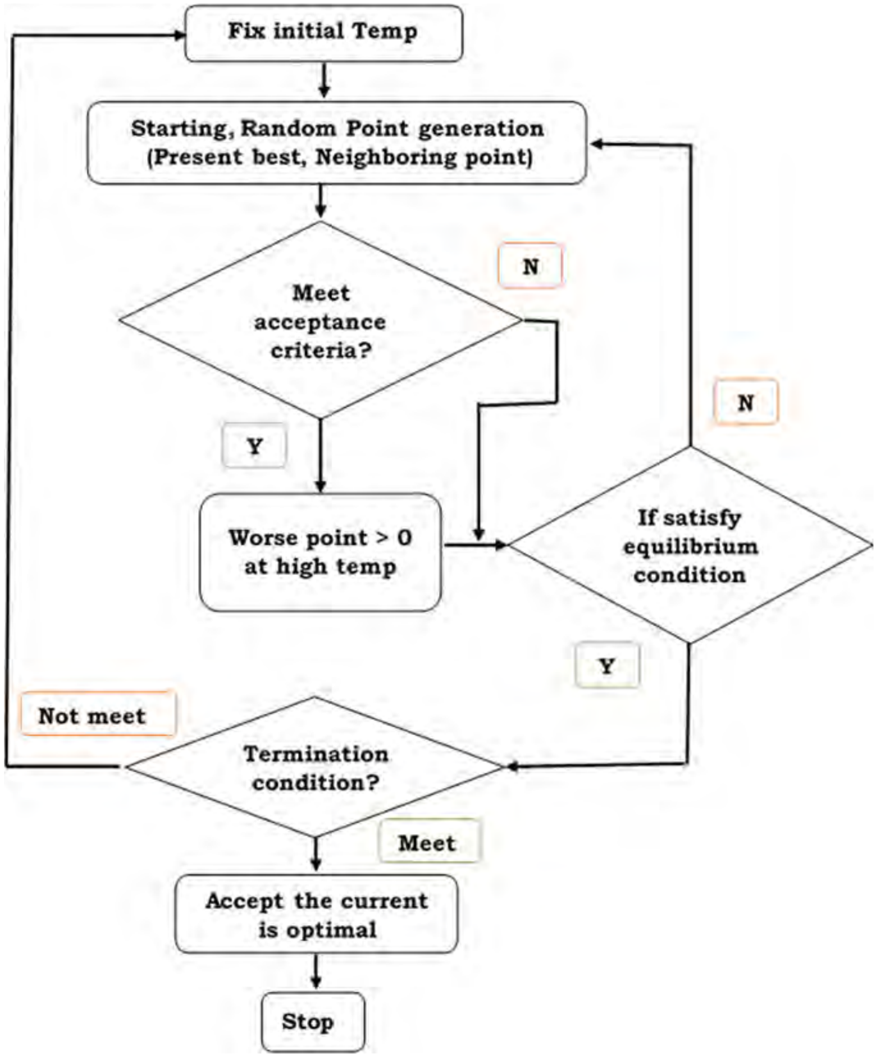


Fig. 4 Flow of SA

Algorithm 1: HPSO Based Energy Model for Extended WBAN Lifespan

Input: Set of initial active nodes, Energy status of the nodes in WBAN
Output: best, the globally optimal set of active nodes

```
Step 1. Set initial particles                 // Initialization
Step 2. Initial the parameters (v and p) of each particle
Step 3. Set pbest and gbest
Step 4. f=evaluate()
Step 5. Repeat
                for each candidate
                   SA(best(particle)
                   update position, pbest and gbest and
                f1=evaluation()
                 if f1 > f
                Update position, pbest and gbest
                end if
                End for
        SA(best(particle)
        Update best
        Update v
        Update p
        until facing the stopping criteria
Step 6. Return best as optimal global particle
```

The main objective of this proposed model is to identify the optimal global subset of an active node for communication in WBAN with minimal network operational energy for an extended network lifespan. HPSO is initialized with an initially random set of active nodes in the network. HPSO-WBAN energy model is used to increase the network lifespan of WBAN by managing the nodes' remaining energy in the network and the total number of nodes alive during the simulation for successful network communications.

4.1 Fitness Function

A multi-objective maximization fitness function evaluated in the HPSO hybrid algorithm optimizes the life cycle of the WBAN using an active set of nodes. It helps to ensure that the condition envisaged with each response is acceptable. To desire, network coverage is a maximum probability of points in the target region that can be sensitive and well-defined by Eq. 3. Table 3 illustrates the design goals and criteria taken for the study. Table 4 shows the network simulation setup of the proposed work.

$$\max f_A(x) = \frac{A_{TS}}{A}. \tag{3}$$

Table 3 Proposed design perception

Design goals	Criteria	Challenges
Propagation (sense to network)	Zig-bee, Wi-Fi, Bluetooth	Diversity, expense, protection, privacy, data acquisition, and truthfulness
Propagation (network to cloud)	CoAP/HTTP/FTP	
Energy source	Battery powered	
Latency	50 s for data	
Storage	Cloud (public, private, industry)	

Table 4 Simulation parameters

Parameters	Value
Area	500 × 500
No. of relay nodes	8
No. of sensor nodes	100
Transmission of relay nodes	50 m
Propagation	Two Ray Ground
Network interface type	Wireless Phy
Traffic type	CBR, FTP
IEEE 802.15.4 standard	Default values
Simulation time	1000 s
Initial energy	50 J
Energy threshold	20% of the initial energy
Sensors	EEG, ECG, pulse, motion, blood pressure, blood glucose

A_{TS} is the active sensors in the target region, and A is the region.

5 Results and Discussions

This study is to analyze the efficiency of the GA, DPSO, and hybrid for node coverage optimization. The estimation time is seen in Figs. 5, 6, and 7 for DPSO, GA, hybrid.

5.1 Packet Delivery Ratio

Figure 6 shows the packet delivery, and the *x*-axis indicates the network node level; the *y*-axis is the delivery ratio of the particular nodes in the region. The packet

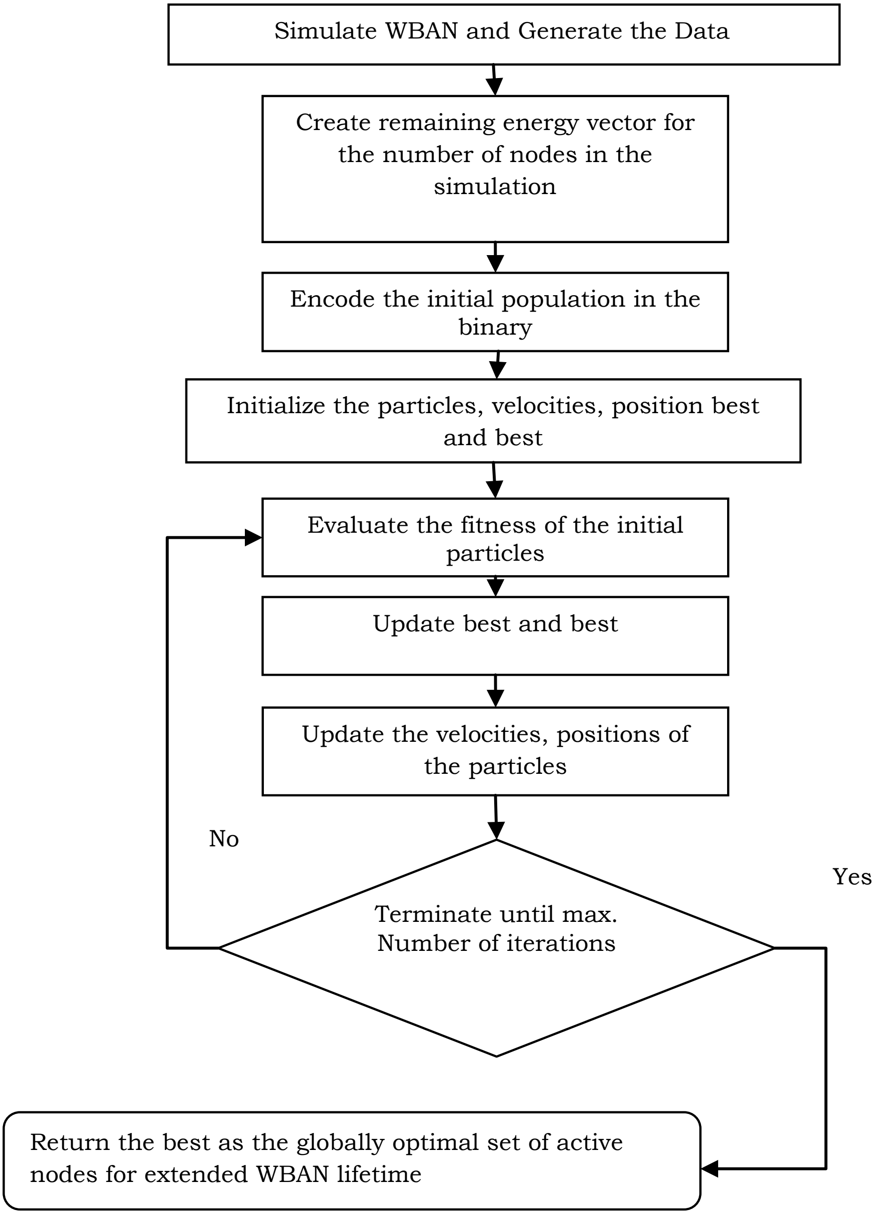


Fig. 5 A workflow of HPSO-based energy model for extended WBAN lifespan

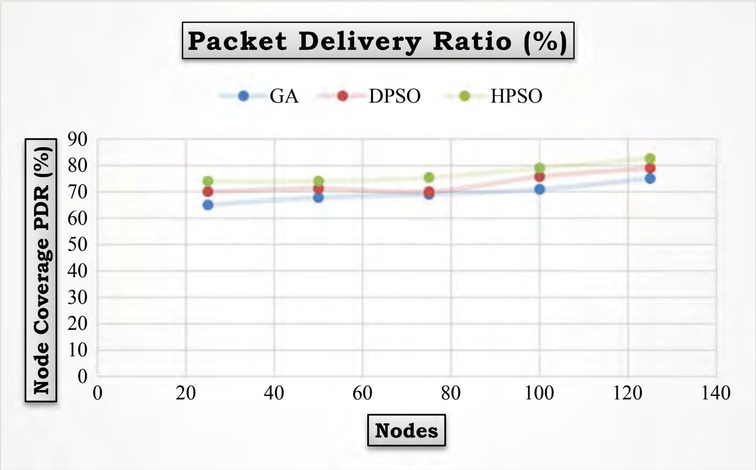


Fig. 6 Packet delivery ratio

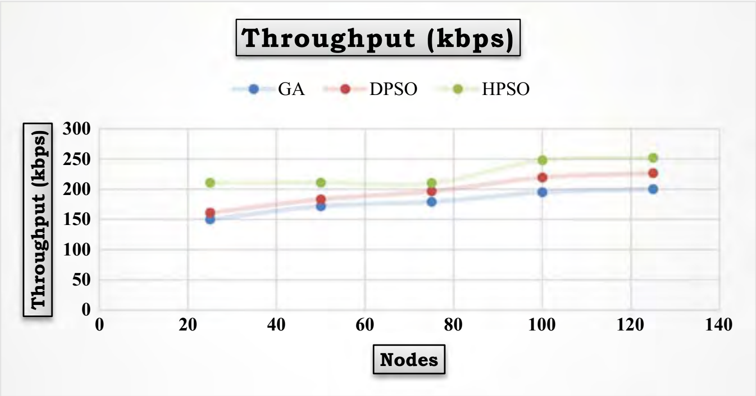


Fig. 7 Throughput

delivery ratio is 17, 24, and 31% increases in GA, DPSO, and HPSO. It needs a cumulative number of packages sent and some packets received to determine the packet distribution ratio. The packet rating is sustained at 90 percent, even though network size exceeds 100 nodes, with higher efficiency compared to GA, DPSO, and HPSO. Figure 6 represents the packet delivery ratio with a percentage.

$$\mathbf{Packetdeliveryratio} = \frac{\sum \mathrm{PacketsReceived}}{\sum \mathrm{PacketSent}}$$

5.2 *Throughput*

Figure 7 shows the throughput level of WBAN and results showed DPSO, hybrid PSO ratio of the throughput is high compare with the GA, DPSO. It is comprehensible from Fig. 7, the throughput value for 150 nodes in the simulated WBAN is 08 s. Similarly, the throughput value for 120 nodes in the simulated WBAN is 52 s. Therefore, the throughput value is openly relative to the network range.

5.3 *End-To-End Delay*

Figure 8 represents the end-to-end delay performance of GA, DPSO, and hybrid PSO for optimal node coverage in WBAN. *X*-axis represents the number of nodes, whereas the *y*-axis represents the end-to-end delay time in second. Hybrid BPSO-based node coverage algorithm has lower end-to-end delay value when compared to GA and BPSO similarly, the EED value for 100 nodes in the simulated WBAN is 57 s. Therefore, the end-to-end delay value is directly proportional to the network size.

$$\text{End - to - End Delay} = \text{Received time} - \text{Sent time}$$

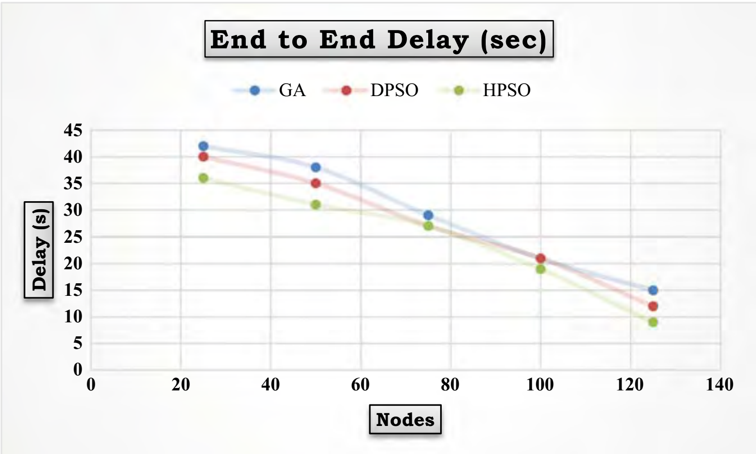


Fig. 8 End-to-end delay

6 Conclusion and Future Scope

In this work, the coverage optimization of nodes in the WBAN is accessed through QoS parameters such as packet transfer rate, end-to-end delay, and performance. The HPSO algorithm has eliminated the local convergence impact of HPSO and GA algorithms and increases the consistency between exchanging and filtering information to find the best solution in the search field. The performance of the HPSO and standard DPSO algorithm has been studied in the simulation environment for identifying the optimal node coverage nodes for seamless communications. From the simulation results, it can be seen that the HPSO-based node coverage algorithm for WBAN performs better than the DPSO. In the future, different hybrid mechanisms will be identified, proposed, and studied for the same environment.

References

1. Sarkar A, Maunder S (2015) Path loss estimation for a wireless sensor network for application in ship. Int J Comput Sci Mob Comput
2. Astudy, Haryono T (2016) Novel binary PSO algorithm based optimization of transmission expansion planning considering power losses. In: International conference on innovation in engineering and vocational education, IOP Publishing, IOP Conf. Series: Materials Science and Engineering, vol 128, p 012023. doi:https://doi.org/10.1088/1757-899X/128/1/012023
3. Dhamodharavadhani S (2015) A survey on clustering based routing protocols in Mobile ad hoc networks. In: 2015 international conference on soft-computing and networks security (ICSNS). doi: https://doi.org/10.1109/icsns.2015.7292426
4. Sivabalan S, Dhamodharavadhani S, Rathipriya R (2019) Opportunistic Forward routing using bee colony optimization. Int J Comput Sci Eng 7(5):1820–1827. https://doi.org/10.26438/ijcse/v7i5.18201827
5. Nguyen BH, Xue B, Andreae P, Zhang M, A new binary particle swarm optimization approach: momentum and dynamic balance between exploration and exploitation. IEEE Trans Cybern
6. Jamil F, Iqbal MA (2019) Adaptive thermal-aware routing protocol for wireless body area network. In: International conference on broadband and wireless computing, communication, and applications
7. Kaur HP (2015) Cost-based efficient routing for wireless body area networks. Int J Comput Sci Mob Comput
8. Dhamodharavadhani S, Rathipriya R (2020) Enhanced logistic regression (ELR) model for big data. In Garcia Marquez FP (ed) Handbook of Research on Big Data Clustering and Machine Learning, pp 152–176, IGI Global. http://doi.org/10.4018/978-1-7998-0106-1.ch008
9. Mohd Kaleem M (2014) Energy consumption using network stability and multi-hop protocol for link efficiency in wireless body area networks. J Comput Eng 16 113:120
10. Mohamad MS, Omatu S, Deris S, Yoshioka M, Abdullah A, Ibrahim Z (2013) An enhancement of binary particle swarm optimization for gene selection in classifying cancer classes. Algor Molecul Biol 8:1510
11. Romesh Singh M (2019) Development of efficient multi-hop protocols for wireless body area network. Int J Innov Technol Explor Eng
12. Sivabalan S, Dhamodharavadhani S, Rathipriya R (2020) Arbitrary walk with minimum length based route identification scheme in graph structure for opportunistic wireless sensor network. Swarm Intell Resour Manage Internet Things 2020:47–63. https://doi.org/10.1016/b978-0-12-818287-1.00006-1

13. Sivabalan S, Rathipriya R (2017). Slot scheduling Mac using energy efficiency in ad hoc wireless networks. In: 2017 international conference on inventive communication and computational technologies (ICICCT). doi: https://doi.org/10.1109/icicct.2017.7975234
14. Thangavel K, Bagyamani J, Rathipriya R (2012) Novel hybrid PSO-SA model for biclustering of expression data. Proc Eng 30:1048–1055. https://doi.org/10.1016/j.proeng.2012.01.962
15. Rathipriya R, Thangavel K (2012) A discrete artificial bees colony inspired biclustering algorithm. Int J Swarm Intel Res 3(1):30–42. https://doi.org/10.4018/jsir.2012010102
16. Dhamodharavadhani S, Rathipriya R (2020) Variable selection method for regression models using computational intelligence techniques. In Ganapathi P, Shanmugapriya D (ed) Handbook of research on machine and deep learning applications for cyber security (pp. 416–436). IGI Global. https://doi.org/10.4018/978-1-5225-9611-0%2Ech019
17. Dhamodharavadhani S, Rathipriya R (2021) Novel COVID-19 mortality rate prediction (MRP) model for India using regression model with optimized hyperparameter. J Case Inform Technol (JCIT), 23(4):1–12. https://doi.org/10.4018/JCIT.20211001.oa1
18. Selvaraj S, Rathipriya R (2019) Energy efficiency in wireless body area networks using path loss model. Int J Comput Sci Eng 7(5):1695–1700. https://doi.org/10.26438/ijcse/v7i5.16951700
19. Dhamodharavadhani S, Rathipriya R (2018) Region-wise rainfall prediction using mapreduce-based exponential smoothing techniques. In: Advances in intelligent systems and computing advances in big data and cloud computing. 229–239. https://doi.org/10.1007/978-981-13-1882-5_21
20. Dhamodharavadhani S, Rathipriya R, Chatterjee JM (2020). COVID-19 mortality rate prediction for India using statistical neural network models. Frontiers Public Health 8. https://doi.org/0.3389/fpubh.2020.00441

Collective Examinations of Documents on COVID-19 Peril Factors Through NLP

E. Laxmi Lydia, Jose Moses Gummadi, Chinmaya Ranjan Pattanaik, B. Prasad, CH. Usha Kumari, and Ravuri Daniel

Abstract The outbreak of the novel COVID-19 virus is identified across all experimental scientific tests that assist victims to fight against the pandemic situation. The problem seems to have a large number of scientific COVID-19 articles with different risk factors. The quick identification of documents allows the processing and interpretation of inevitable essential knowledge for investigators. This article provides a solution by creating an unsupervised framework for the interpretation of clinical trials over COVID-19 risk factors with a diverse range of articles related to vaccines and treatments from a large corpus of documents. It also provides practical informative knowledge regarding COVID-19 risk factors and helps researchers to enable any single author to obtain appropriate information. The present application uses artificial intelligence, natural language processing approaches, incorporated throughout

E. Laxmi Lydia (✉)
Department of Computer Science and Engineering, Vignan's Institute of Information Technology (A), Visakhapatnam, Andhra Pradesh, India
e-mail: elaxmi2002@yahoo.com

J. Moses Gummadi
Department of CSE, VFSTR (Deemed to be University), Guntur, India
e-mail: josemoses@gmail.com

C. Ranjan Pattanaik
Department of Computer Science and Engineering, Ajay Binay Institute of Technology, Cuttack, Odisha, India
e-mail: chinmaya.pattnaik@rediffmail.com

B. Prasad
Department of Information Technology, Vignan's Institute of Information Technology (A), Visakhapatnam, Andhra Pradesh, India
e-mail: arjunprasad.bode@gmail.com

CH. Usha Kumari
Department of ECE, GRIET, Gokaraju Rangaraju Institute of Engineering & Technology, Hyderabad, India

R. Daniel
Department of Computer Science and Engineering, Bapatla Engineering College (Autonomous), Bapatla, India
e-mail: danielravuri@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_62

the search engines, to search for keywords to classify categories with normalized linguistic data. The text data are instead parsed in phrases and thresholds the text with recognition of data frame components with relevant outcomes.

Keywords SARS-CoV-2 · COVID-19 · MERS-CoV · Cytotoxic chemotherapy · Natural language processing (NLP) · Artificial intelligence (AI)

1 Introduction

The newly updated incidence of coronavirus (2019-nCoV) in Wuhan has become a worldwide threat to humans. The world within 2 months after its arrival in December 2019, nearly 70,000 people was infected with such a newly emerging coronavirus that resulted in around 1700 deaths. In response to the increasing number of patients infected daily, this new virus is estimated to be a source of a major pandemic. Two such big coronavirus epidemics have previously been encountered in the globe. In 2002–2003, a new coronavirus named SARS broke out and spread into other countries, like Vietnam, Canada, mostly in China and Hong Kong. About 8000 cases have been filed initially with overall 774 deaths worldwide [1]. From then, SARS death cases were reduced. Nevertheless, among several bat species with the possibility to infect human cells, a new SARS-like coronavirus was found [2]. Throughout the Middle East Asian countries, after 10 years a new variety of coronavirus known as MERS was identified. A total of 1308 laboratory-confirmed cases, along with 449 individuals treated as female, were registered in Saudi Arabia between November 2012 and February 2016 [3]. 400 of the 449 female patients; 179 were reproductive (15–45 years) age details were available and 16 (i.e., 8.9 percent) of the 179 cases. Including its 1308 MERS-CoV confirmed cases, five MoH cases of pregnancy have been identified [4].

Such studies revealed that the Indian economy at a low rate, coronavirus infections occur more in India. One of the researcher Suresha revealed that out of 9 virus-infected cases occurred among the 1706 occurrences (1,04%), 4 HCoV-OC43 cases, 3 cases HCoV-NL63 were positive, and 2 HCoV229E were positive [5]. According to the AIIMS study, New Delhi demonstrated that 17.8% of infants suffer from coronavirus infection. Most of these clinical trials produced certain survey factors which might imply the potential propagation route or risk factors. A broad range of environmental viruses including the coronavirus, hantaviruses, lyssaviruses, Lassa virus, Rabies, Nipah, Henipavirus, Ebola virus, SARS coronavirus, and Marburg virus have been reported from bats [6]. Bats are also considered a natural host. SARS and MERS have been identified to be from a bat in early coronavirus epidemic situation. Nationally and internationally, there are over 1200 bat species of varying sizes, ranges, and ecosystems. In the Indian subcontinent, almost 128 bat species were described. The most unusual, formed and rendered bat (Kerivoula picta), rare bat of Salim Alis (Latidens salimalii); the Indian flying fox (Pteropus giganteus) was identified as unusual bats found in the world. Such bats move from higher Himalayan areas, northwestern deserts to tropical forests in the eastern region. Indian flying fox is the dominant species in India that have twelve types of flying foxes. Throughout the world, only three are popular the flying Indian fox, Rousettus leschenaultia, and Cynopterus sphinx. Certain habitats are uncommon and can only be found on peaks and islands of Andaman and Nicobar in the Thailand region [7].

The population of India is at greater or less risk with an outburst of coronavirus and got alerted to prevent the death cases due to coronavirus pandemic. Drastic action is expected to grow substantially to prevent any global epidemic of coronavirus. To establish the diseases of coronavirus spreading throughout this population, a new strain needs to be governed. In underlying mechanisms of virus transmission, research needs to be carried out. Providing essential insights into the sequence of transmission, multiple strains of coronavirus have different receptors and receptor biology. Strategies to recognize receptors for unidentified viruses and medicines intended to stimulate the interconnection between virus receptors. Academic research on the human biology and disease progression of coronavirus, the new drugs too should be pursued. COVID-19 was initiated in the development and manufacture of vaccines. GSK (Vaccine giant) and Jenner Institute, in Great Britain, intended to provide the same framework for the preparation of MERS-CoV flu shots. India is indeed one of the leading manufacturers of medicines, and several provinces use those manufactured vaccines. India often has a solid system for the short-term development of large-scale vaccinations. Resources that manufacture medicines use the formulation of the coronavirus which includes initial research papers as a preparation to prevent an unexpected occurrence to assess the efficacy and reliability of the vaccine in India.

2 Literature Survey

India is experiencing a rapid epidemiological transformation with an emerging occurrence of chronic sicknesses that has a population of nearly one billion with lower-middle-income countries (i.e., 68, 85 percent). In 2012 [8], cardiac disease and stroke have been among the top three major causes that contributed to the decline in premature global deaths, and hypertension, another of the strongest cardiovascular risk factors, gives rise to 45% of heart diseases [9]. In 1990 India, chronic diseases reported as 3.78 million (40.4%), and by 2020, this has crossed 7.63 million (66.7%). Hypertension has become a major health issue in most Indian regions. 10.8% of deaths and 4.6% of disabilities are estimated in India [10]. The higher efficacy of hypertension in India was related to measures like aging, current drinking, stress, and anxiety level and body mass of more than 90 cm in coastal Pondicherry [11].

Bartwal et al. [12] noticed that hypertension to be 41.7%, and hypertension is linked to age development, the household background of hypertension, increased salt consumption, a mixed diet, a rise in the thickness of the neck, body mass index, and waist–hip ratio. Interestingly, Kokiwar et al. [13] identified that alcohol consumption amongst rural communities in Central India hasn't been tied to hypertension while factors as the upper social class, sedentary exercise habits; misuse of tobacco, and diabetes were significantly correlated in hypertension [14].

The framework of World Health Organization (WHO) has been used [15] for the assessment of behavioral risk factors like drinking, eating habits, physical activity, smoking, and obesity. A weight of kilograms divided by height (kg/m^2) is derived from body mass index calculation. To represent the involvement of behavior classified by each risk factor was labeled as 0/1. A risk factor index has reportedly been managed to create by summarizing the factors. The spike glycoprotein or antibody with the foremost growth factor of viral infection is aimed at particularly strategies to fix Cov vaccines and medicines. Few persons demonstrated adequacy in vitro studies and expanded to parallelize animal or human studies, with insufficient use to combat COVID-19 infection [16].

Diagnosing COVID-19, coronavirus-caused disease (sars-cov-2) with several vaccines (MADs/RADs) has been interpreted scientifically. They have analyzed clinical results of how to use antiretroviral treatment for coronavirus detection and care [17]. The clinical results from SARS, MERS, COVID-19 patients trained to treat antiretroviral have been screened in primary studies. Two randomized clinical trials, 24 observational studies were administered from a previous sample of 433 names, including clinical evidence antiretroviral. Including its 21 laboratory experiments demonstrating patient outcomes, three trials have always been conducted on SARS patients, six were conducted on MERS patients, and 12 were performed on COVID-19 patients. 3 out of 361 patients undergoing lpv/r died in randomized trials; scientific proof remained poor. The beneficial role of lpv/r has been defined as post-exposure prophylaxis by three studies [18].

Articulating the current preclinical and clinical data from initial studies emphasize several significant modified features of SAR-COV-2 that further distinguish between SAR-COV and middle eastern coronavirus syndrome (MER-COV), such as large variability of infection assessment. Recent clinical trials have shown that a variety of such drugs are effective, such as favipiravir, an antiviral within that vast array that influences viral replication, and hydroxychloroquine, a converted antimalarial drug that impairs with the endosomal access system of the virus. They have anticipated that the global pandemic disaster will lead to more systematic pharmaceutical techniques to design provides a broad data analysis [19].

For the control of this immediate and lethal disease, no unique anti-virus medicines or vaccines were generated. Modern screening tests are also carried with underlying mechanisms, the therapeutic effect of TCM, and the reorganization of new natural anticoronaviral compounds [20].

The probability of developing extreme and even deadly respiratory diseases is relatively high for cancer and transplant patients with COVID-19, particularly as they may be treated with immune or immune-stimulant drugs. This investigation focuses on the impact of all these medicines on host immunity to COVID-19. Various approaches for immune-suppressing or -stimulating drugs use Ovid Medline. They are cytotoxic chemotherapy, low-dose steroids, tnfα, il-6, jak inhibitors, il-1 blockade, mycophenolate, tacrolimus, anti-cd20, and ctla4-ig. For performance, 89 research studies have been conducted [21]. Many epidemiological studies have proven that the efficacy of hypertension elevated among COVID-19 patients affected by COVID infection as a high-risk factor [22].

Based on the most frequent use of COVID-19 drugs chloroquine and hydroxychloroquine (HCQ), patients were identified to have improvement. Taking into account the low-risk factors, the extensive further association of other diseases shows cost-effectiveness and easy access in India. Researchers suggested that both prescribed drugs are worthy of an immediate clinical track trial, which can be carefully considered as observational medicines for clinical use [23].

Evidence of patient protection in the use of chloroquine and hydroxychloroquine is obligatory for many years. When experimenting with COVID-19, the main target of chloroquine and hydroxychloroquine under different viral conditions has been limited [24].

In a standard procedure, investigators may have randomized controlled monitoring precautions for the effects of chloroquine and hydroxychloroquine—medication or in conjunction with other medications—as eligibility criteria under several concurrent clinical evaluations. In case of randomized infections caused by another coronavirus, such as MERS-COV and SARS-COV, and unrandomized studies of COVID-19, randomized trials would be tested if no support for important findings appears specifically observed in randomized studies or if the proof is small or very low [25].

The only remedy to the sudden infectious disease epidemic can be repositioning of medications. A total of 21 objectives have been assessed toward structure-based virtual libraries from zinc drugs and existing dataset of natural products using selective ligand screening. A collection of 78 most often-used antiviral medicines has also been developed that effectively addresses those commercially available and drug trials for SARS-COV-2. Requirements and future drug targets for these drugs have been expected. This analysis includes different vaccines and aims for further in vitro and in vivo SARS-COV-2 treatments. The new developments in existing clinical trials of these drugs show potential drug repositioning strategies for the treatment of SARS-COV-2 infections [14].

3 Methodology

Following is the process which enables medical research group, communities, and policymakers to easily review recent studies and developments in a specified area of knowledge. Interactive search queries allow practitioners to search for COVID-19-related knowledge. The full-text scientific study is also integrated into the final output frame and can be clicked directly.

(a) Initially, a variety of abstracts were particularly been analyzed by COVID-19 and its variations. Due to the extremely huge scale of the information.
(b) SciSpacy on the abstracts used to perform tokenization and data analysis includes stemming, lemmatization, and stop-word elimination. ScispaCy is

perhaps a Python package that comprises SpaCy modeling for life sciences, research-based or pharmacological text processing.

(c) The manuscripts (abstracts) have been put into a TF-IDF model for TF-IDF estimation.

(d) This identified abstract was calculated against the cosine similarities of the dynamic user application.

(e) Finally, it filters the most related documents and reveals the ten most important documents.

A. **Procedure**

Install the SpaCy pipeline with a broad vocabulary and 600 k word vectors for biomedical data.

Step 1: Import packages such as NumPy, pandas, scisspacy, spacy, phrasematcher, CountVectorizer, TQDM.

Step 2: Load the metadata in the abstract of the article designers choose.

Step 3: Collect all related research papers relevant to COVID-19 and its derivatives.

Step 4: Load sci model and initiate tokenizer (to discard stop words and lemmatizing words).

Step 5: Perform training process through TF-IDF vectorizer over research papers.

Step 6: Characterize the similarity of cosine to procure top n documents.

Step 7: Classifies method to obtain the top ten search query documents and a process to respond to every question.

Step 8: Set methods to screen the data in the table.

4 Result Analysis

The following experimental results were carried out based on the Kaggle dataset, i.e., COVID-19 open research dataset challenge. This article provides a platform for extracting and classifying relevant information that could be enhanced in future using highly qualified interpretation in the absence of expert feedback. Data concerning risk factors, associated with COVID-19 such as smoking, additional lung disorders, bacterial infections, as well as other co-organisms, socio-cultural, and psychological factors to influence the actual impacts and similarities of the virus; preterm infants and nursing mothers to measure whether heterogeneous respiratory/viral infections create the virus quite easily transmitted are examined. Figure 1 shows the top n document abstracts related to COVID-19 risk factors. Figure 2 shows the top n document abstracts related to risk factors like smoking and pre-existing pulmonary disease; Fig. 3 shows the top n document abstracts related to COVID-19 heart risks.

```
SearchDocuments(['COVID-19 risk factors'])
```

	Title	publish_time	abstract	Score
40379	Multivariate Analysis of Factors Affecting COVID-19 Case and Death Rate in U.S. Counties: The Significant Effects of Black Race and Temperature	2020-04-22	Objectives: Coronavirus disease-19 (COVID-19) has spread rapidly around the world, and many risk factors including patient demographics, social determinants of health, environmental variables, underlying health conditions, and adherence to social distancing have been hypothesized to affect case and death rates. However, little has been done to account for the potential confounding effects of these factors. Using a large multivariate analysis, this study illuminates modulators of COVID-19 incidence and mortality in U.S. counties while controlling for risk factors across multiple domains. Methods: Data on COVID-19 and various risk factors in all U.S. counties was collected from publicly available data sources through April 14, 2020. Counties with at least 50 COVID-19 cases were included in case analyses and those with at least 10 deaths were included in mortality models. The 661 counties meeting inclusion criteria for number of cases were grouped into quartiles and comparisons of risk factors were made using t-tests between the highest and lowest quartiles. Similar comparisons for 217 counties were made for above average and below average deaths/100,000. Adjusted linear and logistic regression analyses were performed to evaluate the independent effects of factors that significantly impacted cases and deaths. Results: Univariate analyses demonstrated numerous significant differences between cohorts for both cases and deaths. Risk factors associated with increased cases and/or deaths per 100,000 included increased GDP per capita, decreased social distancing, increased age, increased percent Black, decreased percent Hispanic, decreased percent Asian, decreased health, increased poverty, increased diabetes, increased coronary heart disease, increased physical inactivity, increased alcohol consumption, increased tobacco use, and decreased access to primary care. Multivariate regression analyses demonstrated Black race is a risk factor for worse COVID-19 outcome independent of comorbidities, poverty, access to health care, and other mitigating factors. Lower daily temperatures was also an independent risk factor in case load but not deaths. Conclusions: U.S. counties with a higher proportion of Black residents are associated with increased COVID-19 cases	0.361

Fig. 1 Finding documents related to COVID-19 risk factors

```
SearchDocuments(['Risk factors such as Smoking, pre-existing pulmonary disease'])
```

	Title	publish_time	abstract	Score
40112	Smoking is Associated with COVID-19 Progression: A Meta-Analysis	2020-04-16	Objective: To determine the association between smoking and progression of COVID-19. Design: A meta-analysis of 12 published papers. Data Source: PubMed database was searched on April 6, 2020. Eligibility criteria and data analysis: We included studies reporting smoking behavior of COVID-19 patients and progression of disease. Search terms included smoking, smoker*, characteristics, risk factors, outcomes, and COVID-19, COVID, coronavirus, sar cov-2, sar cov 2. There were no language limitations. One author extracted information for each study, screened the abstract or the full text, with questions resolved through discussion among both authors. A random effects meta-analysis was applied. Main Outcome Measures: The study outcome was progression of COVID-19 among people who already had the disease. Results: We identified 12 papers with a total of 9,025 COVID-19 patients, 878 (9.7%) with severe disease and 495 with a history of smoking (5.5%). The meta-analysis showed a significant association between smoking and progression of COVID-19 (OR 2.25, 95% CI 1.49-3.39, p=0.001). Limitations in the 12 papers suggest that the actual risk of smoking may be higher. Conclusions: Smoking is a risk factor for progression of COVID-19, with smokers having higher odds of COVID-19 progression than never smokers. Physicians and public health professionals should collect data on smoking as part of clinical management and add smoking cessation to the list of practices to blunt the COVID-19 pandemic.	0.428
16516	Smoking Upregulates Angiotensin-Converting Enzyme-2 Receptor: A Potential	2020-03-20	The epicenter of the original outbreak in China has high male smoking rates of around 50%, and early reported death rates have an emphasis on older males, therefore the likelihood of smokers being overrepresented in fatalities is high. In Iran, China, Italy, and South Korea, female smoking rates are much lower than males. Fewer females have contracted the virus. If this analysis is correct, then Indonesia would be expected to begin experiencing high rates of Covid-19 because its male smoking rate is over 60% (Tobacco Atlas). Smokers are vulnerable to respiratory viruses. Smoking can upregulate angiotensin-converting enzyme-2 (ACE2) receptor, the known receptor for both the severe acute respiratory syndrome (SARS)-coronavirus (SARS-CoV) and the human respiratory	0.332

Fig. 2 Finding documents related to risk factors like smoking and pre-existing pulmonary disease

```
SearchDocuments(['risk factors such as heart risks'])
```

	Title	publish_time	abstract	Score
16508	Understanding coronavirus disease (COVID-19) risk perceptions among the public to enhance risk communication efforts: a practical approach for outbreaks, Finland, February 2020	2020-04-02	Understanding risk perceptions of the public is critical for risk communication. In February 2020, the Finnish Institute for Health and Welfare started collecting weekly qualitative data on coronavirus disease (COVID-19) risk perception that informs risk communication efforts. The process is based on thematic analysis of emails and social media messages from the public and identifies factors linked to appraisal of risk magnitude, which are developed into risk communication recommendations together with health and communication experts.	0.304
	Multivariate		Objectives: Coronavirus disease-19 (COVID-19) has spread rapidly around the world, and many risk factors including patient demographics, social determinants of health, environmental variables, underlying health conditions, and adherence to social distancing have been hypothesized to affect case and death rates. However, little has been done to account for the potential confounding effects of these factors. Using a large multivariate analysis, this study illuminates modulators of COVID-19 incidence and mortality in U.S. counties while controlling for risk factors across multiple domains. Methods: Data on COVID-19 and various risk factors in all U.S. counties was collected from publicly available data sources through April 14, 2020. Counties with at least 50 COVID-19 cases were included in case analyses and those with at least 10 deaths were included in mortality models. The 661 counties meeting inclusion criteria for number of cases were grouped into quartiles and comparisons of	

Fig. 3 Finding documents related to risk factors like heart risks

5 Conclusion

As new coronavirus empirical research expands swiftly, the medical analysis organization finds it extraordinarily difficult to follow up on current alerts. Extraction of efficient and relevant patient-centered COVID-19 risk factors information using artificial intelligence. This provides insights over the ongoing battle against this infectious disease from respective natural language processing tools. Consequently, the medical research communities are in great demand for these methods and make use of linguistic properties. This article follows a robust approach as a search process from a given input keyword. This creates and supports a well-defined framework as the platform includes extractive summaries (SciBert). In a fraction of the time, the researchers were able to read the article easily. Furthermore, this framework can turn into an expert classification framework with a little creativity, where specialists can asynchronously click on the correct fragment phrases to transform an unsupervised approach to a supervised learning function.

Acknowledgements This research article work does not have any funding support from any funding agencies.

References

1. World Health Organization (2020) Cumulative number of reported probable cases of severe acute respiratory syndrome (SARS). https://www.who.int/csr/sars/country/en
2. He JF et al (2004) Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. Science 303:1666–1669
3. World Health Organization—WHO (2019) Middle East respiratory syndrome coronavirus (MERS-CoV). https://www.who.int/emergencies/mers-cov/en
4. Alserehi H, Wali G, Alshukairi A, Alraddadi B (2016) Impact of middle east respiratory syndrome coronavirus (MERS-CoV) on pregnancy and perinatal outcome. BMC Infect Dis 16:105
5. Kumar P, Medigeshi GR, Mishra VS, Islam M, Randev S, Mukherjee A, Chaudhry R, Kapil A, Ram Jat K, Lodha R, Kabra SK (2017) Etiology of acute respiratory infections in infants: a prospective birth cohort study. Pediatr Infect Dis J. 36(1):25–30
6. Calisher CH, Childs JE, Field HE, Holmes KV, Schountz T (2006) Bats: important reservoir hosts of emerging viruses. Clin Microbiol Rev 19(3):531–545
7. Boro AR, Saikia PK, Saikia U (2018) New records of bats (Mammalia: Chiroptera) from Assam, northeastern India with a distribution list of bat fauna of the state. J Threatened Taxa 10(5):11606–11612
8. Kumar NP, Shankarego HS, Revathy R (2011) An assessment of preventable risk factors for chronic non-communicable diseases in an adult population. Asian J Epidemiol 4(1):9–16
9. Lim SS, Vos T, Flaxman AD et al (2012) Comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. 2e Lancet 380(9859):2224–2260
10. Institute for Health Metrics and Evaluation (2014) India high blood pressure. https://www.Healthmetricsandevaluation.org/search-gbd-data
11. Ganesh K, Naresh A, Bammigatti C (2015) Prevalence and risk factors of hypertension among male police personnel in urban Pondicherry, India. Kathmandu Univ Med J 12(4):242–246
12. Bartwal J, Awasthi S, Rawat CMS, Singh RK (2014) Prevalence of hypertension and its risk factors among individuals attending outpatient department of rural health training centre, Haldwani. Ind J Commun Health 26(1):76–81
13. Kokiwar PR, Gupta SS, Durge PM (2012) Prevalence of hypertension in a rural community of central India. J Assoc Physicians India 60:26–29
14. Rajkumar E, Romate J, Factors BR (2020) Hypertension knowledge, and hypertension in rural India, Hindawi. Int J Hypertens. 1–7 Article ID 8108202. https://doi.org/https://doi.org/10.1155/2020/8108202
15. World Health Organization (2005) 2e WHO step wise approach to chronic disease risk factor surveillance. World Health Organization, Geneva, Switzerland
16. Dhama K, Khan S, Tiwari R, Dadar M, Malik Y, Singh K, Chaicumpa W (2020) COVID-19, an emerging coronavirus infection: advances and prospects in designing and developing vaccines, immunotherapeutics and therapeutics. Hum Vaccines Immunother doi: https://doi.org/10.1080/21645515.2020.1735227
17. Chobanian AV, Bakris GL, Black HR et al (2003) Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. Hypertension 42(6):1206–1252
18. Ford N, Vitoria M, Rangaraj A, Norris SL, Calmy A, Doherty M (2020) Systematic review of the efficacy and safety of antiretroviral drugs against SARS, MERS or COVID-19: initial assessment. J Int AIDS Soc 23(4):e25489. https://doi.org/10.1002/jia2.25489
19. Tu YF, Chien CS, Yarmishyn AA et al (2020) A review of SARS-CoV-2 and the ongoing clinical trials. Int J Mol Sci 21(7):2657. doi:https://doi.org/10.3390/ijms21072657
20. Yang Y, Islam MS, Wang J, Li Y, Chen X (2020) Traditional Chinese medicine in the treatment of patients infected with 2019-new coronavirus (SARS-CoV-2): a review and perspective. Int J Biol Sci 16(10):1708–1717. Published 2020 Mar 15. doi:https://doi.org/10.7150/ijbs.45538

21. Russell B, Moss C, George G et al (2020) Associations between immune-suppressive and stimulating drugs and novel COVID-19-a systematic review of current evidence. Ecancermedicalscience 14:1022. Published 2020 Mar 27. doi:https://doi.org/10.3332/ecancer.2020.1022
22. Ruocco G, Feola M, Palazzuoli A (2020) Hypertension prevalence in human coronavirus disease: the role of ACE system in infection spread and severity. Int J Infect Dis 95:373–375
23. Kapoor KM, Kapoor A (2020) Role of chloroquine and hydroxychloroquine in the treatment of COVID-19 infection. Syst Lit Rev. medRxiv 2020.03.24.20042366; doi: https://doi.org/https://doi.org/10.1101/2020.03.24.20042366
24. Jeria RB, Reyes MX, Franco JV, Acuna MP, Torres Lopez LA, Rada G (2020) Chloroquine and hydroxychloroquine for the treatment of COVID-19: a living systematic review protocol. medRxiv 2020.04.03.20052530; doi: https://doi.org/https://doi.org/10.1101/2020.04.03.20052530
25. Canrong Wu, Liu Y, Yang Y, Peng Zhang Wu, Zhong YW, Wang Q, Yang Xu, Li M, Li X, Zheng M, Chen L, Li H (2020) Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. Acta Pharmaceutica Sinica B 10(5):766–788

Design and Development of Topic Identification Using Latent Dirichlet Allocation

P. Lakshmi Prasanna, S. Sandeep, V. Kantha Rao, and B. Sekhar Babu

Abstract Data storing and retrieving are the most important task in the present condition. Storing can be ended based on the topic that the document describes. Text mining generates documents from the collection of topics. To identify the topics, we have to categorize the documents; to classify, we are using topic modeling. Text mining technique is used for discovering latent semantic structure which is a fragment of topic modeling. Various research areas that make use of probabilistic modeling includes software engineering, political science, and medical science. A topic model is a probability-based model that discovers the major themes which are a group of documents. The main idea is to treat the documents as mixtures of topics in the topic model, and every topic is viewed as a probability distribution of the words. This research work aims to propose a model called topic modeling using LDA, and this model has been experimented on two datasets, where one is two news group dataset, and other is twenty news group dataset, and finally, all the results are tabulated.

1 Introduction

LDA is an unconventional probably based model for compilations of detached information and consequently more suitable for document data. An unsubstantiated accession was supposedly to be utilized for identifying and penetrating the bunch of terms in huge clusters of documents. This method imagines that each text is a combination of concepts; each and every word is accredited to each of the concepts that has limited prospect. This model also uncovers various topics that the texts correspond to and to what extent every individual concept is at hand in a document. LDA bears to discover the credibility dispersions over terms; then, it finds arrays of words that form together with definite possibility. Aforementioned arrays are marked as “topic.”

LDA is a Bayesian inference model designed by David et al. [1]. Every text is linked to a likelihood allocation more than topics, and further, topics are probability

P. Lakshmi Prasanna (✉) · S. Sandeep · V. Kantha Rao · B. Sekhar Babu
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India
e-mail: pprasanna@kluniversity.in

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_63

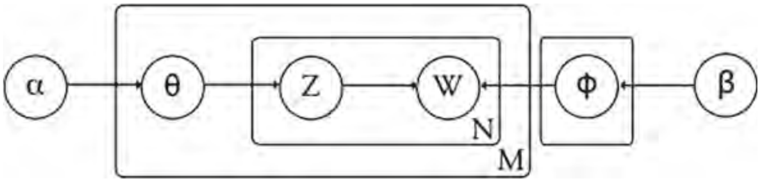


Fig. 1 Block diagram of LDA

segregation son terms. Every expected result of an arbitrary is changeable with the possibility that the happening will occur and come out of the distribution as an equation that links them. For example, in a sport coin flips for decision of who will go first and second as a deciding factor, there are two probable results out of it: heads or tails. Heads is personified as 1 and tails as 0. Here is an indiscriminate variable which is marked X. And dualistic expected reactions is delineated by X as 0 or X as 1. The probability allotment of X or $P(X = x)$ is: $P(X = 0) = 0.5$ $P(X = 1) = 0.5$. Figure 1 represents the block diagram of LDA.

2 LDA Criterion

To attain rest results, the below mentioned few parameters can be helpful.

1. Topics Number—Quantity of topics which have to be acquired through the body.
2. α and β are hyper parameters—α symbolizes document-topic density, and β symbolizes topic and token density. Depending upon the upper worth of α, texts are possessed of added topics and greater the β; themes are made up of bigger pack of terms [2].
3. Iterations no—The duration for algorithm to come together (or how do we sense that the algorithm has converged)?

1. The hidden topics digit in the corpus (K) needs to be explained previous to initiating the practice of the model.
2. Since the perplexity reduces very slowly, algorithm convergence needs a large number of iterations.
3. If the training information set is large, then the algorithm requires a long time for preparation.
4. Being a probabilistic model, LDA requires additional clarifications in preparing statistical inference. Hence, it does not toil on same small texts like sentence categorization or tweets.
5. For providing meaningful knowledge to advance excellence of topics, LDA does not take into account connection of concepts or words.

3 Topic

To identify what is a topic and how it can be interpreted has been a continuous point of discuss in the literature. Commonly, themes provide an outline to the subject matter present in the texts being considered as well as brief explanation of the contents of the dataset. Whether the statement "topics" is even the appropriate word to use for the groups of words generated by the topic model evaluation which has been under radar for decades. "Topic" might not constantly be apt word to be use for the divisions which are generated through topic modeling [3]. Some have recommended that "discourse" may be sophisticated, as subjects are not constantly amalgamated linguistically. For instance, mathematics, lineage, reasoning, science, theory, century, and epistemology shall not be well thought-out semantically coherent as they are seen as discourse of subjects found contained by philosophy. When every word has a different meaning, it would be intricate to give them an explicit topic [4].

3.1 Topic Models

Topic models are a class of algorithms which exploit co-occurrences of words in documents in order to discover hidden sets of words which explain the co-occurrence patterns and are referred to as topics. Probabilistic topic models explain pragmatic documents with an underlying, hidden probabilistic model. The observed documents are assumed to be random samples from this model. In a probabilistic topic model, each document is associated in the midst of a probability allocation over a set of ideas, and concepts get associated with a prospective disposal upon the set of words. Similar documents share a similar topic distribution. Topic models are often employed in regard to text mining errands, e.g., to have understanding and visualizing the content of large document corpora or for detecting relations between topics and other variables of interest [5]. Additionally, topic models can be employed as a mean for dimensionality reduction (documents are mapped to a lower dimensional topic space), as input for prediction tasks, in recommender systems (e.g., for predicting semantically related tags) or in information retrieval (e.g., to understand and disambiguate the topic of query terms).

3.2 Topic Modeling Using LDA

For finding topics, LDA algorithm was used by applying Bayesian belief networks theorem to calculate probability and topic identification of each term and each document. For feature reductions, LDA Algorithm was considered to reduce features and filtered top terms of each topic and store it in the filtered document term matrix.

Latent Dirichlet allocation is being primarily utilized for evaluating texts [2]. It presumes that here is N no. of topics on how texts will be produced, and each topic is corresponding to multinomial circulation on words in the terminology. A document $w_{\mathrm{d}} = \{w_{\mathrm{dt}}\}d_{\mathrm{tt}} = 1$ is accomplished by variety a concoction and these topics and fragmenting words from the mixture [1]. Figure 2 indicates the procedure of LDA Algorithm, and Table 1 represents the notations of LDA.

Applied the Bayesian probability to identify the model.

- Represent β as the total terms of the documents.
- Represent θ as specifying all the topics.
- Display all the terms in the documents with a minimum range of 10.
- Dividing the terms into topics.
- All terms are displayed in the topic wise.
- Display all the probability values with words.
- Based on these probabilities to create a word cloud.

1. For every topic n=1,2,3....N , Sketch a word fractionΦ_n~drichlet (β)

2. For every text d=1,2...D,Illustrate topic fraction θ_d ~drichlet (α)

3. For every phrase t=1,2.....d_t

   (i) Sketch a topic project p_{dt}~ categorical (θ_d)

   (ii) Sketch a term w dt~ catogorical (ΦP_{dn})

Fig. 2 Procedure of LDA Algorithm

Table 1 Notations of LDA

Symbol	Description
N	No. of topics
Y	No. of unique words in the vocabulary
D	No. of documents
DT	No. of words in the documents DT
θ_{d}	Proportion of topics specific to documents
On	Proportion of words specific to topic N
P_{dn}	Identify the topics of *n*th word in document d
W_{dt}	Identify the word in document d
αβ	Parameters of Dirichlet distribution

4 Procedure of LDA

See Fig. 2.

5 Notations of LDA Algorithm

Here, the first step depicts the number of topics and after that represents all word momentarily apportion toward the topics, and the procedure is completed absurdly, and from time to time, similar terms can be functional to various topics. The final step shows the updated adaptation of the topic assignment depended on their credibility as per the above criteria:

1. Primary criteria is about the length of prevalence is that tokens across the topics—it can be called as $P(w/t)$.
2. The another criteria is about the topic of the issues in the document $P(t/d)$.

As per the Bayesian belief network theorem $P(t/w) = P(t/d) * P(w/t)$. To calculate probability of each term. Figure 3, 4, 5, 6, and 7 represent the outcome of topic modeling using LDA.

Top terms per topics for 2 groups data (Figs. 3 and 4).

Top terms with probabilities for 20 news groups data (Figs. 5 and 6).

Filtered document term matrix (FDTM) (Fig. 7).

```
topStermsperTopic
      Topic 1            Topic 2
[1,]  "subject:"         "the"
[2,]  "message-id:"      "newsgroups:"
[3,]  "writes:"          "lines:"
[4,]  "references:"      "gmt"
[5,]  "path:"            "date:"
[6,]  "apr"              "from:"
[7,]  "can"              "1993"
[8,]  "organization:"    "re:"
[9,]  "article"          "organization:"
[10,] "one"              "people"
[11,] "re:"              "apr"
[12,] "just"             "article"
```

Fig. 3 Top terms per topics for 2 groups data

```
probabilities
          subject:            message-id:              writes:
      0.0172949546           0.0170884915           0.0135671864
       references:                  path:                  apr
      0.0122303777           0.0118613931           0.0114458514
               can          organization:              article
      0.0112066146           0.0088592080           0.0075160422
               one                    re:                 just
      0.0070988660           0.0067745911           0.0064550337
             from:                  date:                  car
      0.0063414342           0.0062106215           0.0061137902
       alt.atheism                  like                  know
      0.0060162284           0.0049348162           0.0048697613
              will     nntp-posting-host:                  see
      0.0047899506           0.0046996792           0.0045395942
```

Fig. 4 Top terms with probabilities for 2 groups data

6 Conclusion

Topic modeling began as of text mining method designed for disclosing latent semantic structure within a compilation of texts. In text mining, each archive is produced from anthology of topics. It relies upon probabilistic modeling that has a huge assortment of relevance such as image detection, semantic understanding, and automatic music improvisation recognition. In this chapter to proposed topic modeling employing latent Dirichlet allocation [LDA], the LDA works backward to learn the topic illustration in all texts and the word allotment to every topic. The main focus of this paper is on LDA algorithms, and the outcomes will be displayed in 20 news group dataset and 2 groups dataset.

Topic: 8
Words: 0.015*"govern" + 0.008*"money" + 0.007*"militia" + 0.006*"cost" + 0.006*"stratus" + 0.006*"navi" + 0.005*"spend" + 0.005*"henri" + 0.005*"libertarian

Topic: 9
Words: 0.008*"medic" + 0.008*"netcom" + 0.008*"isra" + 0.007*"israel" + 0.007*"bank" + 0.007*"pitt" + 0.007*"diseas" + 0.006*"research" + 0.006*"harvard" +

Topic: 10
Words: 0.011*"govern" + 0.009*"drug" + 0.007*"legal" + 0.006*"polic" + 0.006*"court" + 0.006*"public" + 0.005*"countri" + 0.005*"detector" + 0.005*"radar" +

Topic: 11
Words: 0.011*"weapon" + 0.011*"gun" + 0.009*"firearm" + 0.009*"crime" + 0.007*"control" + 0.006*"crimin" + 0.006*"kill" + 0.006*"colorado" + 0.006*"carri" +

Topic: 12
Words: 0.032*"window" + 0.030*"file" + 0.017*"program" + 0.011*"imag" + 0.009*"version" + 0.007*"entri" + 0.007*"display" + 0.007*"color" + 0.006*"format" +

Topic: 13
Words: 0.018*"christian" + 0.012*"jesus" + 0.008*"bibl" + 0.007*"church" + 0.006*"word" + 0.006*"religion" + 0.006*"life" + 0.006*"christ" + 0.005*"truth" +

Topic: 14
Words: 0.030*"game" + 0.026*"team" + 0.017*"play" + 0.011*"season" + 0.009*"hockey" + 0.009*"score" + 0.009*"player" + 0.007*"leagu" + 0.006*"goal" + 0.006*"

Topic: 15
Words: 0.010*"server" + 0.008*"softwar" + 0.008*"motif" + 0.008*"avail" + 0.007*"graphic" + 0.007*"type" + 0.006*"applic" + 0.006*"keyboard" + 0.006*"suppor

Topic: 16
Words: 0.017*"exist" + 0.011*"atheist" + 0.011*"israel" + 0.009*"atheism" + 0.008*"scienc" + 0.006*"appear" + 0.006*"alaska" + 0.006*"isra" + 0.006*"book" +

Topic: 17
Words: 0.035*"nasa" + 0.016*"columbia" + 0.012*"center" + 0.010*"research" + 0.009*"andrew" + 0.008*"gari" + 0.007*"scienc" + 0.007*"american" + 0.006*"euro

Topic: 18
Words: 0.016*"wire" + 0.013*"player" + 0.007*"roger" + 0.007*"grind" + 0.006*"basebal" + 0.005*"outlet" + 0.005*"play" + 0.005*"circuit" + 0.004*"stat" + 0.

Topic: 19
Words: 0.014*"cwru" + 0.013*"cleveland" + 0.013*"ohio" + 0.011*"freenet" + 0.011*"john" + 0.010*"list" + 0.008*"western" + 0.007*"magnus" + 0.006*"michael"

Fig. 5 Top terms with probabilities for 20 news groups data

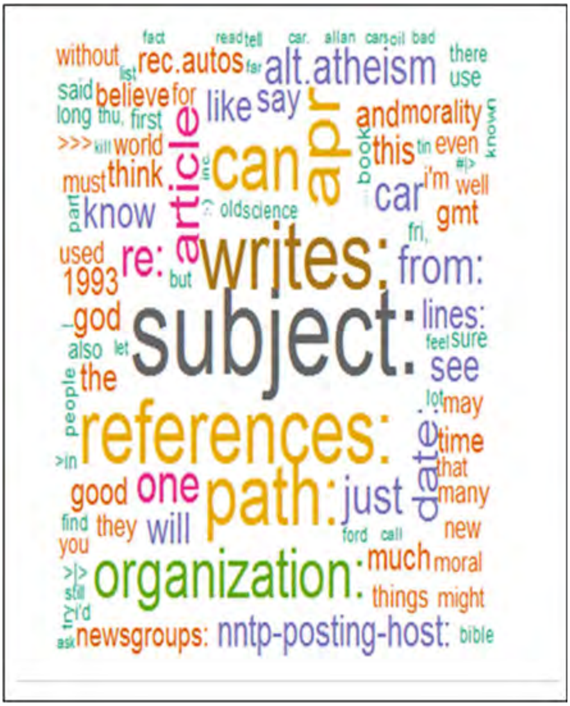
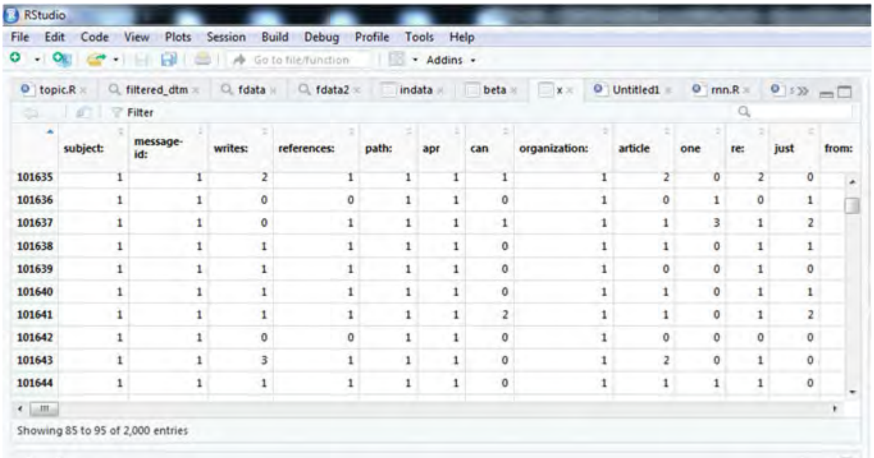


Fig. 6 News groups dataset of word cloud



	subject:	message-id:	writes:	references:	path:	apr	can	organization:	article	one	re:	just	from:
101635	1	1	2	1	1	1	1	1	2	0	2	0	
101636	1	1	0	0	1	1	0	1	0	1	0	1	
101637	1	1	0	1	1	1	1	1	1	3	1	2	
101638	1	1	1	1	1	1	0	1	1	0	1	1	
101639	1	1	1	1	1	1	0	1	0	0	1	0	
101640	1	1	1	1	1	1	0	1	1	0	1	1	
101641	1	1	1	1	1	1	2	1	1	0	1	2	
101642	1	1	0	0	1	1	0	1	0	0	0	0	
101643	1	1	3	1	1	1	0	1	2	0	1	0	
101644	1	1	1	1	1	1	0	1	1	1	1	0	

Fig. 7 Filtered document term m0atrix for filtered features

References

1. BleiDM, Andrew Y (2003) Latent Drichlent allocation. J Mach Learn Res
2. Tong Z, Zhang H (2016) A text mining research based on Lda topic modelling. In: The sixth international conference on computer science, engineering and information technology
3. DayaSagar KV, Shyam Krishna C, Lalith Kumar G, Surya Teja P, Charless Babu G (2018) A method for finding threated web sites through crime data mining and sentiment analysis. Int J EngTechnol (UAE) 7(2):62–65
4. Wallach HM (2008) Structured topic models for language, PhD thesis
5. Roose H, Roose W, Daenekindt S (2018) Trends in contemporary art discourse: using topic models to analyze 25 years of professional art criticism. CulturSociol 12:303–324
6. Kousar A, Subrahmanyam K (2019) Feature selection, optimization and clustering strategies of text documents. Int J Electr Comput Eng 9(2):1313–1320
7. BleiDM (2012) Surveying a suite of algorithms that offer a solution to managing large document archives. Commun ACM
8. Bastani1 K, Namavari1 H, Shaffer J (2016) Latent Dirichlet Allocation (LDA) for topic modeling of the CFPB consumer complaints. IEEE
9. Kaur PC, Ghorpade T, RamraoAdik V (2017) Extraction of unigram and bigram topic list by using Latent Dirichlet Markov allocation and sentiment classification. In: 2017 international conference on energy, communication, data analytics and soft computing
10. PotharajuSP, Sreedevi M, AndeVK, TirandasuRK (2019) Data mining approach for accelerating the classification accuracy of cardiotocography. ClinEpidemiol Global Health
11. Poornima BK, Deenadayalan D, Kangaiammal A (2017) Text preprocessing on extracted text from audio/video using R. Int J Comput Intell Inform 6(4)
12. Sajid A, Jan S et al (2017) Automatic topic modeling for single document short texts. In: International conference on frontiers of information technology (FIT).
13. Sleeman J, Halem M, Finin T (2017) Discovering scientific influence using cross-domain dynamic topic modeling. In: 2017 IEEE international conference on big data (big data)
14. Sharma N, Yalla P (2017) Classifying natural language text as controlled and uncontrolled for UML diagrams. Int J Adv Comput Sci Appl
15. Sapul MSC, Aung TH, Jiamthapthaksin R (2017) Trending topic discovery of Twitter Tweets using clustering and topic modeling algorithms. In: 2017 14th international joint conference on computer science and software engineering (JCSSE)
16. Lakshmi Prasanna P, Rajeswara Rao D (2017) Literature survey on text classification: a review.J Adv Res Dyn Control Syst 9(12):2270–2280

Hand Gesture Controlled Robot

N. H. Prasad, A. Mariyan Richard, and B. N. Lakshmi Narayan

Abstract Due to the recent advancements in digital technologies, every human being has started to believe more in technology, where internet of things [IoT] assists people to remain more accurate in their work. So, IoT controlled car is a robot, which needs to be controlled by using human gestures. The user must wear a gesture device during which the sensor can be included. The hand movement will be recorded by the sensor for a specific direction, and it can end in deploying robotic motion within the respective directions. The robot and gesture instrument are connected wirelessly with help of radio waves. Wireless communication can help to interact with the robot in a more user-friendly way. In this, it commands the car by using accelerometer sensors that are connected to a hand glove. The sensors are intended to exchange the remote, which is generally utilized to run the car. This will allow the user to regulate actions, i.e., forward, backward, leftward, and rightward movements, while using an equivalent accelerometer sensor to regulate the throttle of the car.

Keywords Hand gestures · Accelerometer · Microcontroller · RF transceiver

1 Introduction

Nowadays, robotics is emerging as a ubiquitous paradigm field of technology. Robot is operated with the help of different system generate programs. An automatic robot does not require any human-driven control. Robot takes its own decision by sensing its environment. Most of the machine driven robots are automatic as they possess excess speed and great accuracy in their work. But some projects demand semi-automatic system or human controlled robots. Most of the control systems are based on gesture recognition or with voice and motion control. Only a small transmitting device should be worn in the palm, where it includes an accelerometer. This will

N. H. Prasad · A. Mariyan Richard (✉) · B. N. Lakshmi Narayan
Department of MCA, NMIT, Bengaluru, India
e-mail: mariyanrich01@gmail.com

B. N. Lakshmi Narayan
e-mail: narayan614@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_64

transmit an appropriate command to the robot and that it can perform any desired action. The data is then incorporated by a MCU and eventually our motor driver IC to regulate the motors.

The proposed research work is all about a robotic car, which is measured using hand gestures, i.e., the handling and controlling of the car depends on the gesture of user. In this project, gestures are captured by using accelerometer and it is computed by software namely, microcontroller software, and therefore, the specifications are forwarded to MCU and encoder circuit, where it is further processed by Nrf24L01 transceiver [1]. In the transceiver section, the Nrf24L01 transceiver is used to send and receive data by using radio waves and process it with MCU for providing those specifications to the robotic vehicles, which acts accordingly to the gesture.

1.1 Proposed Work

In our project, the hand gesture system includes two parts, namely transmitter and receiver. The transmitter part includes accelerometer, one RF-transmitter-module, and Arduino-nano board. The receiver part includes one RF-receiver-module, wheels, and motor driver IC's. The proposed project require two separate 5 V power supply, which can be applied to both the sections. This research work has three axis accelerometer but only two axis; i.e., *x*-axis and *y*-axis are used. The particular analog value is then converted to digital value. Then, the digital values are collected and processed by the Arduino nano and forwarded to the RF transmitter. Further, it will be received by the receiver and sends to the acceptor to stimulate the motor in a specific direction. When there is a tilt in the palm, the robotic car moves accordingly [2]. Hand gesture-based scars are controlled by using hand instead of buttons or joystick. Here, only hand movement is important to manage the robot (Fig. 1).

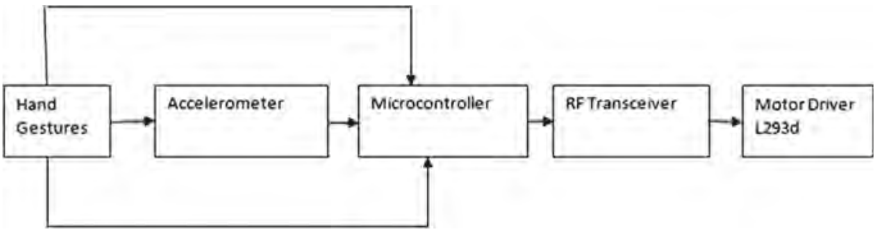


Fig. 1 Process of human-hand gestures

1.2 Block Diagram

1.2.1 Proposed Algorithm

1. Sensed the movement from hand to microcontroller by hand gesture.
2. MCU receives the data and will give the robot instructions.
3. Submit instructions to IC encoder.
4. Encoded data is transmitted over the transmitter.
5. The receiver receives the encoded data in the receivers end.
6. Receiver must forward the encoded data to the decoder.
7. Decoder decodes the data and then sends it to the driver.
8. In all actions, the motor driver drives the engine by following orders and gestures.
9. The robot eventually goes along with the movements.

2 Literature Survey

We described and discussed current work regarding gesture control robot in this segment. Hand-held interface for robot navigation. Using his or her hand movements, a robot can be operated by the user. To record the hand trajectories of a user, a 3-axis accelerometer is adopted. The data from the trajectory is transmitted wirelessly to a computer via an RF module. The trajectories obtained are divided into six control commands used to operate a robot. The classifier adopts the algorithm of dynamic time-warping to classify the hand trajectories. The current study also has drawbacks that the findings of simulation indicate that the classifier could only reach the right rate of 92.2%.

P. Davis et al. Throughout his work on visual gesture recognition, al shows the technique of using a model- based approach to understand human-hand gestures. This employs a small state machine to shape four qualitatively distinct aspects of a general gesture. In certain pictures, fingertips are traced to measure the translating trajectories. Instead the trajectories are practiced to get the gesture's start and stop position. Management is represented as a vector chart, and then matched to save gesture vector models using vector displacement-based table lookup. Results showing seven movements found using pictures examined at 4 Hz on a SPARC-1No different hardware. The seven movements are acting agents for left, right, up, down, catch, rotate, and avoid acts [3].

Rogalla et al. in his paper the use of gesture and speech control to order a robot assistant still requires classical user interfaces to provide advice to a mobile robot assistant. Oral or motion orders achieve a more natural way of directing. In this article, we present new approaches and improvements to existing methods in use in our lab. Our goal is to communicate with a robot to facilitate robust performance of simple tasks using natural and direct communication techniques. In this paper, we explain the framework for understanding the robot's vision and voice. Then, we display robot control to select the appropriate robot response to solve basic manipulations [4].

In his paper, hand gesture recognition with depth images**, Jesus Suarez** et.al presents a review of the literature on the use of depth for hand tracking and recognition of gestures. P.M. In his paper, GESTURE CONTROL ROBOT, Singh et.al presumes that human–machine intercommunication is shifting away from mouse and pen today and is turning the mechanical environment into a widespread and much more cooperative one. With each day, the passage between computers and humans is reduced with the advent of new technology to reduce living standards. Maneuvers have played a critical role in depreciating this gulf. A dogmatic study of "Human–Machine Interaction" using gestures9 has been acted upon in this article. The architecture of the machine is divided into two parts namely: part of the accelerometer and part of the robot [4].

Rahul Ranjan Singh presents a wirelessly controlled robot on his paper, wireless Controlled robot movement device designed to prepare microcontroller, which can be operated human gestures by simple. Within the transmitter device, the robot may move forward, backwards, left, and right. The sensor will admit the hand movement in a particular direction resulting in the robot's motion in the particular direction. Half of the accelerometer and half of the robot. The wireless receiver module takes on the wave signal received by the joystick wireless transmitter module. The module then decodes the wave signal and sends it over UART to the ATmega328P microcontroller. UART looks for Universal asynchronous transmitter receivers. It is commonly used as a protocol for microcontroller interaction. The microcontroller then transmits the motor controlling signal according to the agreed wireless signal to the motor driver IC L298.

2.1 Proposed Methodology

Above figure shows that the gesture movements/direction is sensed from hand to microcontroller. So, the particular encoded values are used to encode the values. And are collected and processed through the nano board to the RF transmitter. Then, in the receiver end, the received data is transmitted to the decoder and stimulate motors to a particular direction. The robot moves in four directions namely backward, forward, right, and left. When there is tilt/bend in the palm, the robotic car moves accordingly. Hand gesture controlled cars are controlled by using hand instead of buttons or joystick. Here, only hand movement is important to manage the robot [5] (Fig. 2).

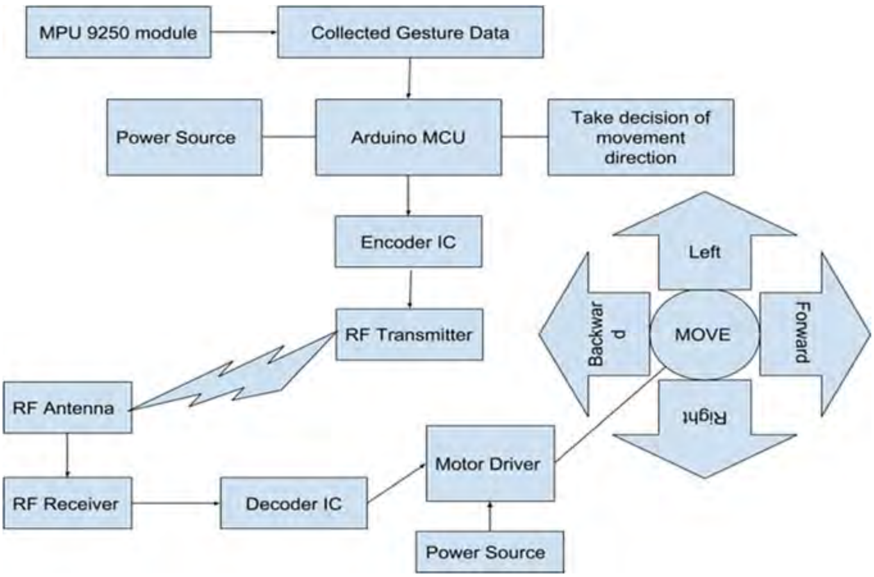
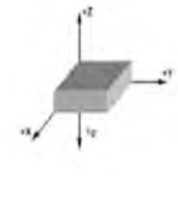
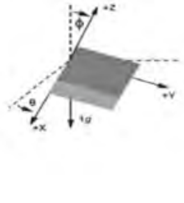
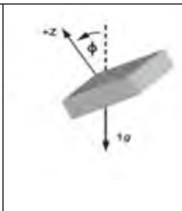
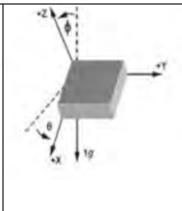
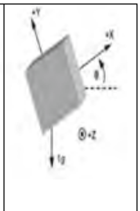
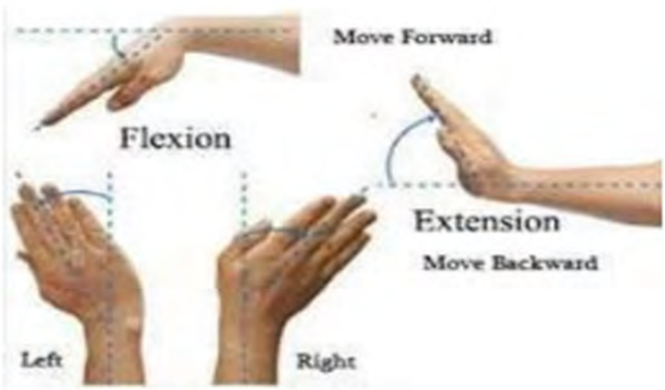


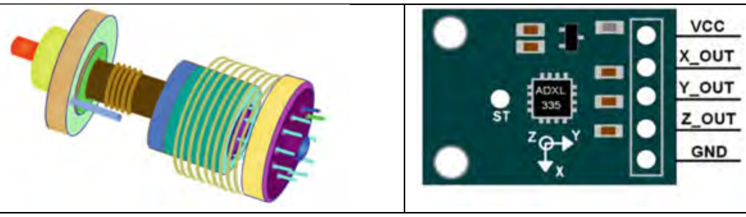
Fig. 2 Methods and process of hand gesture using Arduino

				
Stop condition	Forward Motion	Backward motion	Right Motion	Left Motion



Hand Gestures are

Accelerometer is an instrument that is used to measure the vibration or acceleration of the movement of a hand, where the force caused by vibration or change in motion that causes the mass to squeeze the piezoelectric material, which produces an electrical charge that is proportional to the force exerted upon it.

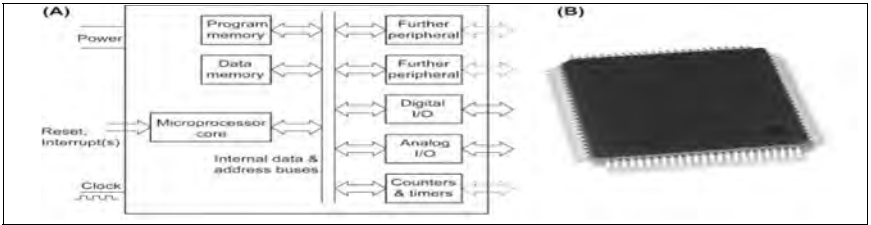


Accelerometer

2.2 *Components*

2.2.1 MCU Arduino Nano Board

Arduino-nano board is also called as MCU, and it is created by Arduino.cc. This Arduino-nano is used in Arduino-UNO board and designed with microcontroller like Atmega328. It is used all over world with variety of applications and specifications. An Arduino board also includes Arduino Mega, Arduino Pro Mini, Arduino UNO, Arduino YUN, Arduino Lily pad and Arduino Leonardo, etc. [6].



Microcontroller diagram

2.2.2 Module NRF24L01+RF

It is used all over the world for its wide variety of applications that require wireless control. It is also called as single chip radio transceiver for the worldwide with its respective 2.4–2.5 GHz ISM band. In this, every module transmits and receives data from each other. These modules are cheap/low cost and you can use them with any MCU. This transceiver consists of a completely integrated frequency synthesizer, modulator, demodulator crystal oscillator, and a protocol engine [7]. The RF module is used to transmit and receive the data between two devices through radio signals. It is also used in embedded system for wireless communication.

2.2.3 L298N Motor Driver

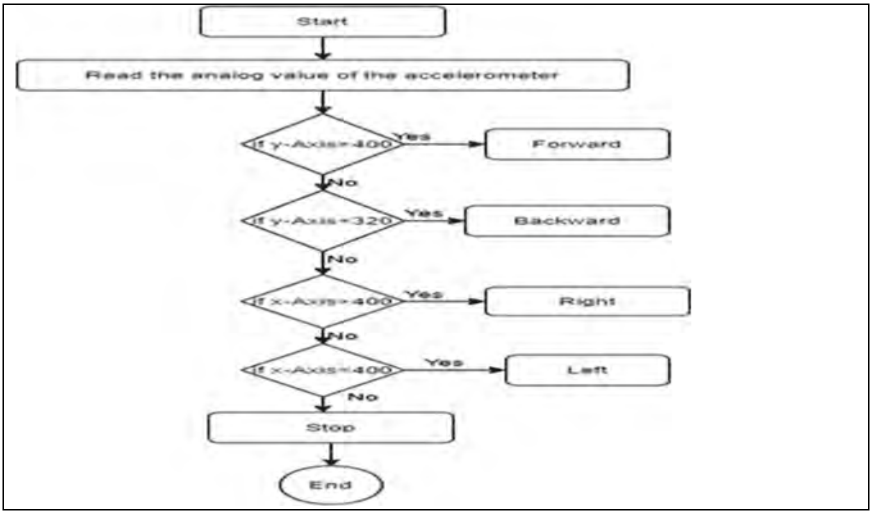
The L298N motor driver is used to permit speed and direction control to two DC motors at the same time. It is also called as dual H bridge motor driver. Which poses 2 terminal blocks for two motors namely motor A and motor B. This in turn has high voltages between DC motors.

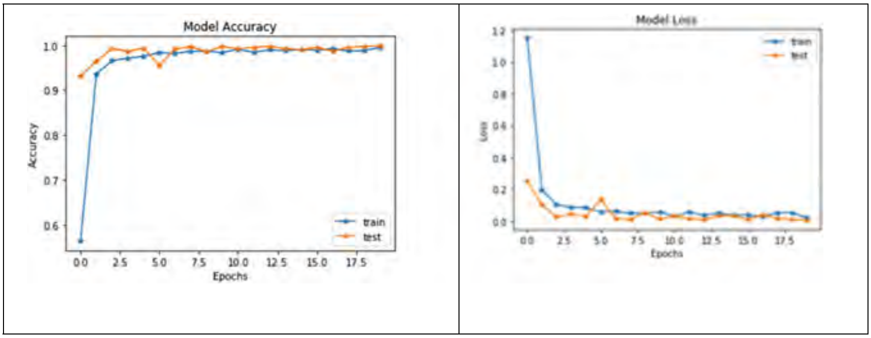
2.2.4 AdXl335

ADXL335 is a three axis accelerometer. This reads the X, Y, and Z axis of accelerometer values as analog voltages. It can measure the tilted values of accelerometer due to gravity. The accelerometer is used to sense the amount of speed and in which direction the device is moving.

2.2.5 TT Gear Motor

TT gear motor is also called as TT DC gearbox motor with great ratios in it. It is low cost plastic gearbox motor, and it is used to make our projects move. This TT DC gearbox motor provides gear ratio of 1:48, which can provide nearly double the speed of the blue metal TT DC motor that has 1:90 gear ratios. Overall it is used to control the speed via PWM from Arduino or any MCU [8].





3 Conclusion

This paper was successfully implemented and tested. The testing was done for test cases, and it was found that it worked well for all conditions. That follows my hand gesture and works according to the data which is transmitted via hand wirelessly. This device measures the gestures and transmits the particular instructions to the robot to maneuver consistently with the user.

4 Future Enhancements

The proposed system is applicable in dangerous environment where cameras are attached to the robot often and may be viewed by the user who is in his station. In medical field, where miniature robots are developed which will help doctors for efficient surgery operations, and it will be used. For more methodical response, threshold values are often employed to detect gesture and advanced features like finger counts that provide distinct functional commands can be used. We can add video camera on the internet for live streaming. We can add bomb and metal detectors and can dispatch to a place, which is harmful for an individual to travel. This sort of hand gesture system is often developed for whole body and can be utilized in military operations.

References

1. Taipei T (2010) The 2010 IEEE/RSJ international conference on intelligent robots and systems. October 1822
2. Kun, Miller WT (1996) Adaptive dynamic balance of abipedrobotusing neural networks. In: IEEE conference on robotics and automation, pp 240–245
3. Zhang S, Wang C, Chan S-C, Wei X, Check HH (2015) IEEE Sens J 15(5)
4. Flikkema P (2014) Department of Electrical Engineering and Computer Science, Northern Arizona University Flagstaff, Arizona
5. Wu Y, Lim J, Yang M-H (2013) A benchmark. In: Proceedings of IEEE conference computer vision and pattern recognition: online object tracking, pp 2411–2418
6. Zhabelova G, Vyatkin V: Multiagent smart grid automation architecture based on IEC 61850/61499 intelligent logical nodes. IEEE Trans Indus Electr 59(5):2351–2362
7. Suarez J, Murphy RR (2012) Hand gesture recognition with depth images: a review. 9–13 Sept 2012. doi: https://doi.org/10.1109/ROMAN.2012.6343787
8. Mansuri R, Vakale S, Shinde A, Patel T (2013) Hand gesture control robot vehicle. IJECT 4(SPL-2):77–80
9. Aswath S, Krishna Tilak C, Suresh A, Udupa G (2014) Human gesture recognition for real—time control of humanoid robot. Int J AdvMechAutomobEng (IJAMAE) 1(1):96–100
10. Jain M, Aditi AL, Fahad Khan M, Maurya A (2012) Wireless gesture control robot: an analysis. Int J Adv Res ComputCommunEng 1(10)
11. ShivrajSBN, Sumathi S (2014) Hand gesture based direction control of Robocar using Arduino microcontroller. Int J Recent TechnolEng (IJRTE) 3(3):32–35
12. Somkuwar V, SandeepkakdeRG, Design and implementation of gesture controlled robot using flex sensor and accelerometer
13. Baron G, Czekalski P, Golenia M, Tokarz K (2015) Gesture and voice driven mobile tribot robot using Kinect sensor. IEEE 33–34
14. Thi-Lan L, Minh-Quoc N, Thi-Thanh-Mai N (2013) Human posture recognition using human skeleton provided by Kinect. IEEE 340–345
15. Matthew T (2010) Recognition hand gesture with Microsoft's Kinect. In: Paper written for CS228 winter 2010
16. Jamie S, Andrew F, Mat C, Toby S, Mark F, Richard M, Alex K, Andrew B. Real-time human pose recognition in parts from single depth images. Microsoft Research Cambridge & Xbox Incubation

17. Lee J, Hironori T, Hitoshi Y, Akihiro K, Yasue M (2013) A robust gesture recognition based on depth data. In: The 19th Korean-Japan joint workshop on frontiers of computer vision, pp 127–131
18. Von Marina S (2013) A introduction to depth sensor for gesture recognition. Written for State University of New York at Stony Brook, pp 1–11
19. Mantecón T, del Blanco CR, Jaureguizar F, García N (2016) Hand gesture recognition using infrared imagery provided by leap motion controller. In: International of conference on advanced concepts for intelligent vision systems, ACIVS 2016, Lecce, Italy, pp 47–57, 24–27 Oct 2016. doi: https://doi.org/10.1007/978-3-319-48680-2_5
20. https://www.researchgate.net/publication/_Gesture_Controlled_Robot_using_Arduino_and_Android
21. https://www.ijitee.org/wp-content/uploads/papers/v9i2/B7185129219.pdf
22. https://www.elprocus.com/h-bridge-motor-control-circuit-using-l293d-ic/
23. www.ijsr.netvol 4, Issue 3, March 2015

Interdependence in Artificial Intelligence to Empower Worldwide COVID-19 Sensitivity

E. Laxmi Lydia, Jose Moses Gummadi, Chinmaya Ranjan Pattanaik, A. Krishna Mohan, G. Jaya Suma, and Ravuri Daniel

Abstract Researchers from different disciplines are striving to leverage a solution for COVID-19 with a unique commitment of scientific collaborations and with cognitive technologies, and highly flexible learning processes are required to maintain the transmission of knowledge, prototype, and code by integrating the application areas to a specific culture and cross-border cooperation. The research experts in the artificial intelligence (AI) and machine learning (ML) domain were tracked and predicted with real-time data observed throughout the world regarding the pandemic situation and timely assessment of the distributed COVID-19 patient information. The considered physiological features followed by clinical tests of patients with COVID-19 offer very simple access to subsequent data transformation, which was relevant but complicated. This paper works on in-depth exploratory data analysis (EDA) prediction analysis over the global medical database of COVID-19 will be available for benefiting future artificial predictive, analytical, and biomedical research, which includes additional COVID-19 approaches associated with pandemics.

E. Laxmi Lydia (✉)
Department of Computer Science and Engineering, Vignan's Institute of Information Technology (A), Visakhapatnam, Andhra Pradesh, India
e-mail: elaxmi2002@yahoo.com

J. Moses Gummadi
Department of CSE, VFSTR (Deemed to be University), Guntur, India
e-mail: josemoses@gmail.com

C. Ranjan Pattanaik
Department of Computer Science and Engineering, Ajay Binay Institute of Technology, Cuttack, Odisha, India
e-mail: chinmaya.pattnaik@rediffmail.com

A. Krishna Mohan
Department of Computer Science and Engineering, JNTUK, Kakinada, Andhra Pradesh, India

G. Jaya Suma
JNTUK-UCEV, Vizianagaram, India

R. Daniel
Department of Computer Science and Engineering, Bapatla Engineering College (Autonomous), Bapatla, India
e-mail: danielravuri@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_65

Keywords COVID-19 · SARA-CoV-2 · Exploratory data analysis (EDA) · Artificial intelligence (AI) and machine learning (ML)

1 Introduction

Since 24th April of 2020, 8,23,626 citizens with 40,598 confirmed deaths, and almost every person in the country is contaminated from SARA-CoV-2 infection [1]. Technological innovations are essential to the same battle against the epidemic; despite, there is a need to disrupt infection rates and the increased scope including its flu epidemic. ML and artificial intelligence (AI) interventions which include some 200 relevant papers including medical journals from 1st January to 24th April 2020 are already used in COVID-19 categories. Nevertheless, large-scale statistics, as well as prototype presenting, organization confirmation, and changes in local environments, are required for AI technologies dealing with COVID-19 to provide a massive impact. It further calls for international cooperation and equality, and indeed the participation of several other involved people, each with medical professionals.

Scientists and professionals are mostly from the regional and personal information sciences, even from Africa and Latin America, such as the African global network sciences, which should be active to maintain a positive effect on any AI program performed at the world scale. To do anything, technologies will not negotiate data confidentiality and security should be developed in front of the slow internet demand. For instance, medical diagnostic devices without any kind of integration. All these applications create the possibility to endorse cutting-edge AI virtualization as well as other environment-friendly challenges to the concerned use of quantitative services.

It is still essential to clarify the involvement of AI in presenting meaningful solutions to the identified epidemic. Interestingly, a worldwide research initiative must be constructed to initiate the measures against this pandemic but instead in the future epidemic without laying anybody behind it. One can fairly assume which in the comment on this thread-Coronavirus era, the world would become more interactive now than before, but also that AI has become more and more one of several drivers of modern civilization [2]. This virus outbreak highlights the urgent need to implement principles of AI effectively [3] by relevant parties. 'Solidarity' is indeed the motto of the world health organization's international treatment program to guide select an acceptable cure against COVID-19 [4]. Confidence in each neighborhood inspires healthy distance, which is crucial to prevent the transmission of the infection within populations. Likewise, the presumption of solidarity could perhaps facilitate the development and promotion of imaginative and socially responsible AI applications in combating COVID-19 and now the worldwide environmental sustainability objectives [5].

AI systems from investigation towards control, some of the issues regarding AI systems is that it does not understand the problem or even how best to make a concerted effort. Further collaboration around researchers and the AI society are important. AI organization, which includes public servants, medical practitioners, and the first emergency workers, has already been seeking and thus should keep seeking support and assistance from consultants. Individuals could indeed encourage signal methodologies which are specialized software applicants. Recurrent or time-consuming activities encompass figuring designs in fabulously wealthy pictures, video, audio, or clinical research information. For example, tomography tests or activities requiring their aggregation of huge datasets from places such as diagnosis monitoring or professional network.

2 Literature Survey

An evaluation of new researchers at both the AI/COVID-19 interplay by several of us [6] provides a mechanism in which interdisciplinary investigation can be categorized on three factors: molecular, medical, and environmental (pharmacology and Infodemics). Biomedical implementations comprise simulation of protein structure [7], studies on viral nucleic acid [8], modifying medicine [9]. But instead, the revelation of medical products [10] most such developments use a variety of AI mechanisms together with specification and extraction of molecular biological medical databases, deep learning systems to forecast protein sequence attributes or protein–ligand binding inclinations, and its use of document learning models to genetics. Prediction of perspectives for genetic information including the use of clinical applications strengthened learning.

Health strategies to reduce treatment plans vary throughout the assessment to medical evaluation, rehabilitation, and assessment of the overall result. Deep learning frameworks would assist COVID-19 image-based diagnostics via trends in ultrasound images and imaging scans [11].

Natural-in-the-loop AI algorithms were established to reduce the waiting time for the examination of radiologists. Smartwatches, digital devices, numerous different wearable technologies, as well as other predictive maintenance software applications that enable physicians to validate patients from afar to save time, and safety precautions can effectively have been used to evaluate disease [12]. Each combination of data references which including medical history and medical imaging can help in assessing the consequences of patients [13] and could be utilized to show conventional healthcare necessities as with increased population for primary care unit rooms. Automation also might be useful for telecom and several other activities such as sanitizing and sterilization of surgical equipment.

AI can modify surveillance systems and observational data simulations besides epidemiology [14]. Specifically, AI is used to recognize and optimize treatment strategies in government health policies, such as preventing, social separation, and reconsideration and to enhance the traditional method of epidemiology by resembling components not widely recognized for diagnostic computations of transmission [15, 16]. Unattended clustering strategies and asset-scoring heuristics can further help determine similitude among both provinces and predict whether each location might need more resources, to incorporate multiple sources of information.

Besides, Infodemics and computer vision is being used to resolve misconceptions and deception through the management of the current knowledge overflow, causing outrage and making it very difficult to locate trustworthy sources [17, 18]. Analyzing audiences (e.g., media platforms, TV, radio) while optimizing assumption checks [19] could be helped with modeling techniques. For example, social media research provides some insights throughout environmental changes and distractions all over the infection and therefore its social; and cultural consequences [20]. The on-going work is to recognize the increase and dissemination of inciting hatred and defensive response on organizations and populations experiencing discrimination which could result in the actual act of violence in medicine [21]. Mostly in COVID-19 address, web portals and virtual agents could be able to propagate trustworthy feedback in scale [22] which needs frameworks for disseminating notifications to facts.

The performance appraisal method for the emerging global epidemic includes teamwork and priority [23], which focuses on either the key unsatisfied conditions that consider organizational reality. Some of the approaches must be feasible and focused on implementing processes [24] most of which are proven and must resist irreversibly damaging the patient care with innovative technologies that might not necessarily lead to improving efficiency. Moreover, prospective and current reforms are needed to meet the particular needs and circumstances of regions of the world with different levels of economic development. Patient-centered care service offerings ought not to neglect effective capitalist structures to maintain proper, health and safety minimizing risk, and damage potential. Priority given to combat COVID-19 is thorough the evaluation of alternative solutions remains mandated, which could be quickly pursued against risking.

To stimulate innovation, manufacturing, and high availability of resources for COVID-19 [25] progress in a way to achieve the advanced health based on information technology. Cooperation with the public health players, government sector participants, and other collaborators are also underway. Collaboration and cooperation among both government agencies and multilateral organizations will indeed facilitate the channeling of ground research that would help introduce technologies in relatively weak economic policy and institutional arrangements countries. Comparing the integrity, therapeutic reliability, and production adequacy of healthcare-related application areas using AI would allow for formulating an effective process of making decisions yet would procedure permits with guidelines [26, 27].

Combating COVID-19 worldwide Infodemics as a scientific concept should be viewed at the same time as the outbreak actually, as the shift in behavior is important to just the global epidemic reaction. News and mass communication stay significant, and health insurance premiums of deception and distortion transmission have to be quantified. Infodemic detection strategies could also be used to facilitate the immediate traditions of facts to the local community, communication, and context-specific information and intervention. AI methods and tools for filling individuals' and politicians' health information gaps should be used for a society-wide response based on evidence and science. The regional and global community should therefore convey and broaden professional exercise, establish guidelines, enabling collaborations, and focus on providing guidance and technical help to the authorities and the appropriate national policymakers to maintain international peace and security to address Infodemics productively while guaranteeing its essential right to communicate.

The global epidemic promotes xenophobia, hatred, and discrimination, raising a pervasive, and likely long-term social justice threat [28]. Changing the nature of violence and promoting, it will help to develop more successful solutions and change the result.

3 Methodology

3.1 Communicating Medical Information as Well as a Framework

Investigations are necessary for smart devices, dozens of data-sharing projects throughout COVID_19, covering the international, national, and local thresholds, are occurring right mostly in three dimensions. Such metrics usually involve genetic sequence [29], genomic analyses [30]. Protein compositions, medical evidence for patients, and medical imaging details for the occurrences of pharmacological information [31]. The ultra-fragmentation of intelligent exchange of data efforts is a major obstacle because it may lead to progress that would be confined to individual programs and entire communities. The improvement and distribution of modern innovations might be accelerated by traceability interventions for statistics, template, and reliability requirements. During this point, international data retention strategies, which are free, inclusive, and compatible, and checked, will repair damaged and foster collaboration among different inhabitants of the area geographies [32].

Fully accessible scientific knowledge will speed up the proliferation of awareness through intersectional AI collaborations around state lines and service delivery of nationwide health organizations. Timely identification, authorization, and impact analysis on nutrition allow possible hazards incorporating open source information (for instance, Epidemic Intelligence Open Source (EIOS) [33] network's location information application). This same social network for health protection recognizes governments, international agencies, and academic institutions which either interdependently evaluates or accepts precise data on epidemics events under that same ideology of cooperation and not rivalry for automatic recognition. The EIOS service published that the very first post on 31st December 2019, documenting an unexplained outbreak of respiratory infections in Wuhan.

Common services and interconnectivity among both repositories can allow for organized intervention and decision-making either at regional, national, and municipal stages of development mostly from the epidemiological aspect. Acknowledging the epidemiological mechanisms and vulnerability characteristics of various influenza populations, while focusing on the multiple moments, may include consideration of national health capability, public policy strategies, ecosystem processes, and social impacts of COVID-19.

Limitations must be conquered, due to the special; technical, architectural conditions, absence of records; characteristics of authentication and explanation; security concerns about personal information and confidential information, and connectivity needs. The collaboration of word embedding and proven predictive analytics might also speed up approaches' adaptive response to different societies. Examples of commonly accessible simulations comprise visual analysis models, prediction of patients' outcomes, filtering misconceptions, and intelligence based on variations of dissemination across virtual media or distilling expertise statistics by broad research publications. Artificial intelligence performance management software packages that reflect the reality of acceptable, interpersonal, therapeutic, law are needed to achieve real-world for open predictive analytics.

3.2 Digital Collaboration and Regional Priorities

Additionally, advanced analytics for tackling age-old issues were introduced. However, it does not need to denote developers should have the possibility to build such application areas. Any AI framework for countering COVID-19 should be checked to make sure it follows ethical standards and values human dignity in particular. Besides, specialists are constantly faced with suggestions about protecting basic rights, including the right to data protection, even in compliance with national security and rationality-by-design standards to create, and deploy AI-driven remedies. When implementing either of those advanced analytics on a global level, there is a need to certify that they will always not infringe violate global democratic rights obligations, which include anti-discrimination responsibilities, surveillance prevention, and editorial sources protection. Decision-makers might also help ensure that AI-friendly approaches are central to beliefs such as diversity and affordability. AI medicare beneficiaries will also encourage equal for balanced access to the global healthcare service system to support participating countries' national health obligations.

The complex nature of both the epidemic calls for systemic approaches, but organizational change is also important to consider biases and contexts. In locations with lower system implementation of chronic disorders with relevant picture configurations, including infectious diseases like tuberculosis and HIV, a system besides diagnosing COVID-19 pneumonia will have to be radically different training. Similarly, separate socioeconomic, cultural, and contextual possibilities than some outlined by academic literature—mainly developed for China or western countries—should indeed be put into consideration by statistical measures customized to developing countries, island countries, humanitarian assistance, or unstable countries. Chatbot also needs analytical patterns for language acquisition to advise primary prevention and sometimes large volumes of education information available for very few hundred from out 7000 linguistic groups today.

Table 1 Data collection based on the attributes

Data	Attributes
Age data	Age group, total cases, percentage
Hospital data	State, NumPrimaryHealthCenters_HMIS, NumCommunityHealthCenters_HMIS, NumSubDistrictHospitals_HMIS, NumDistrictHospitals_HMIS, TotalPublicHealthFacilities_HMIS, NumPublicBeds_HMIS, NumRuralHospitals_NHP18, NumRuralBeds_NHP18, NumUrbanHospitals_NHP18, NumUrbanBeds_NHP18
Testing lab	Lab, address, Pincode, city, state, type
Covid data	Date,Time,State/UnionTerritory, ConfirmedIndianNational, ConfirmedForeignNational, Cured, Deaths, Confirmed
World data	Province/State, Country/Region, Lat, Long, Date, Confirmed, Deaths, Recovered, Active, WHO Region
Making predictions (India)	Province/State, Country/Region, Lat, Long, Date, Confirmed, Deaths, Recovered, Active, WHO Region

4 Result Analysis

Initially load the data using the Kaggle dataset, then normalize data, statewide index the data using set_index, attributes are listed to analyze the data, handles the missing data. The following are the tables that show the considered data with considered attributes for COVID-19 worldwide and India using the prophet model for better predictions. For exploratory data analysis, this paper performs forecasting of data using prophet model libraries. The use of the Prophet.make_future_dataframe will extend the data frame depending on the days specified.

Table 1 describes the information of obtained COVID-19 attribute data for future predictions Table 2 describes the information of obtained COVID-19 Indian attribute data for future predictions using the prophet model. Table 3 describes the information of obtained COVID-19 Worldwide attribute data for future predictions using the prophet model. Figure 1 describes the plot diagram for overall active confirmed, recovered, and death cases in India. Figure 2 describes the plot diagram for predicted active confirmed, recovered, and death cases in India.

5 Conclusion

Several of the proposals and services proposed have still been appropriately designed to either be operational, despite the multiple implementations that offer different development, testing, and delivery probabilities. Consequently, for patients and the automation society, it is necessary to recognize the increasing advancements, which

Table 2 Indian COVID-19 data using a Prediction model

COVID-19 India	Attributes	Model
Confirmed Cases in India	Date,Confirmed	model = Prophet()
Making predictions	ds(date), trend, yhat_lower, yhat_upper, trend_lower, trend_upper, additive_terms, additive_terms_lower, additive_terms_upper, weekly,weekly_lower, weekly_upper, multiplicative_terms, multiplicative_terms_lower, multiplicative_terms_upper, yhat	forecast_india_conf = model.predict(future)
Recovered cases in India	ds(date), trend, yhat_lower, yhat_upper, trend_lower, trend_upper, additive_terms, additive_terms_lower, additive_terms_upper, weekly, weekly_lower, weekly_upper, multiplicative_terms, multiplicative_terms_lower, multiplicative_terms_upper, yhat	forecast_india_recover = model.predict(future)
Deaths in India	ds(date), trend, yhat_lower, yhat_upper, trend_lower, trend_upper, additive_terms, additive_terms_lower, additive_terms_upper, weekly, weekly_lower, weekly_upper, multiplicative_terms, multiplicative_terms_lower, multiplicative_terms_upper, yhat	forecast_india_death = model.predict(future)

will help to react shortly, improve mid-term, and plan for potential infectious diseases by implementing artificial intelligence predictions. This article describes the artificial intelligence requirements as well as the categorization of evolutionary COVID-19 data, artificial intelligence-based communication for medical information, control, and digital collaboration by concerning the regional priorities. Results for COVID-19 were tested and predicted by using exploratory data analysis and prediction analysis through a global database of COVID-19 medical data.

Table 3 Worldwide COVID-19 data using a prediction model

COVID-19 World	Attributes	Model
Making predictions (world)	Province/State, country/Region, lat, long, date, confirmed, deaths, recovered, active, WHO region	model = Prophet()
Confirmed cases in world	ds(date), trend, yhat_lower, yhat_upper, trend_lower, trend_upper, additive_terms, additive_terms_lower, additive_terms_upper, weekly, weekly_lower, weekly_upper, multiplicative_terms, multiplicative_terms_lower, multiplicative_terms_upper, yhat	forecast_world_conf = model.predict(future)
Recovered cases in world	ds(date), trend, yhat_lower, yhat_upper, trend_lower, trend_upper, additive_terms, additive_terms_lower, additive_terms_upper, weekly, weekly_lower, weekly_upper, multiplicative_terms, multiplicative_terms_lower, multiplicative_terms_upper, yhat	forecast_world_recover = model.predict(future)
Deaths in world	ds(date), trend, yhat_lower, yhat_upper, trend_lower, trend_upper, additive_terms, additive_terms_lower, additive_terms_upper, weekly, weekly_lower, weekly_upper, multiplicative_terms, multiplicative_terms_lower, multiplicative_terms_upper, yhat	forecast_world_death = model.predict(future)

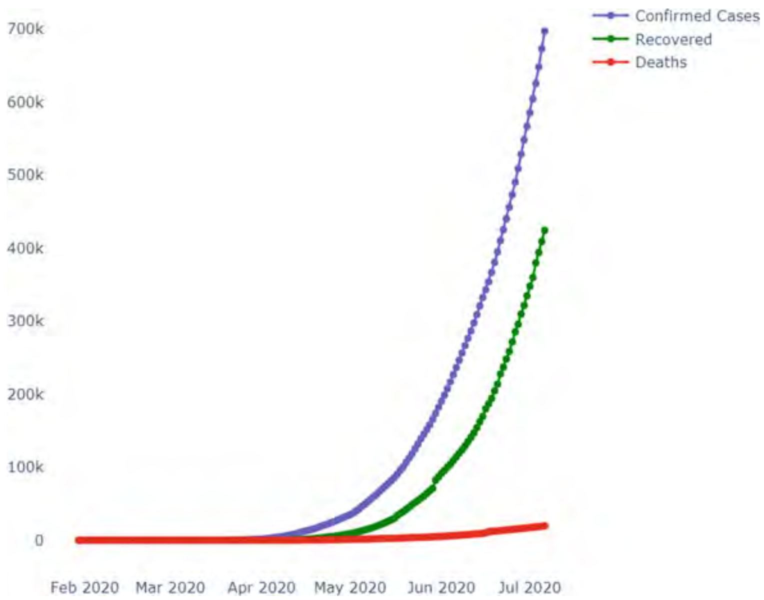


Fig. 1 Plot for total cases in India

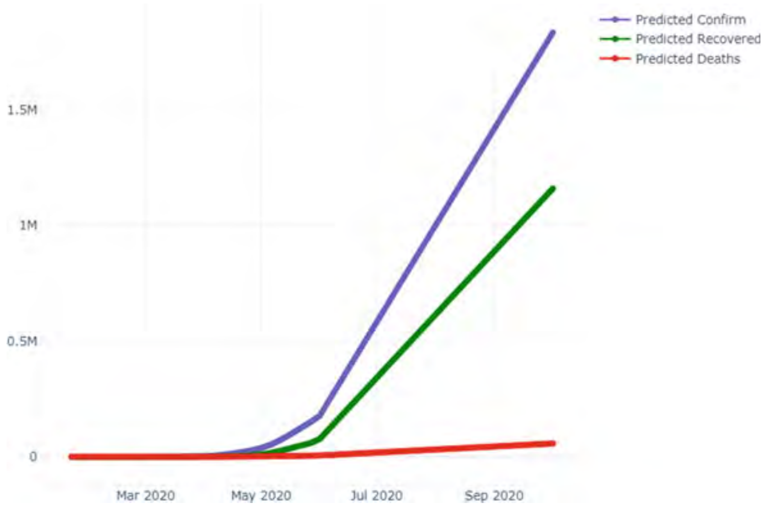


Fig. 2 Plot for predicted cases in India

References

1. WHO (2020) Coronavirus disease (COVID-2019) situation reports. https://www.who.int/emergencies/disease/novel-coronavirus-2019/situation-reports
2. UN Secretary General's High-level Panel on Digital Cooperation (2019) The age of digital interdependence
3. Jobin A, Lenca M, Vayena E (2019) Nat Mach Intell 1:389–399
4. WHO (2020) Solidarity" clinical trial for COVID-19 treatments. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov/solidarity-clinical-trial-for-covid-19-treatments
5. Vinuesa R (2020) Nat Commun 11:233
6. Bullock J, Luccioni A, Hoffmann Pham K, Lam CS, Luengo-Oroz M (2020) Preprint at https://arxiv.org/abs/2003.11336
7. Senior AW et al (2020) Nature 577:706–710
8. Lopez-Rincon A et al (2020). Preprint at. https://doi.org/10.1101/2020.03.13.990242
9. Richardson P et al (2020) Lancet 395:E30–E31
10. Zhavoronkov A et al (2020) Preprint at https://doi.org/https://doi.org/10.26434/chemrxiv.11829102.v2
11. Shi F et al (2020). IEEE Rev Biomed Eng. https://doi.org/10.1109/RBME.2020.2987975
12. Radin J, Wineinger NE, Topol EJ, Steinhubl SR (2020) Lancet dig. Health 2:E85–E93
13. Shi W, Peng X (2020). Preprint at. https://doi.org/10.2139/ssrn.3546089
14. AI-qanessMAA, Ewees AA, Fan H, AbdEI Aziz MJ (2020) Clin Med 9:674
15. Nuria O et al (2020). Sci Adv. https://doi.org/10.1126/sciadv.abc0764
16. Dandekar R, Barbastathis G (2020) Preprint at https://arxiv.org/abs/2003.09403
17. WHO (2020) Infodemic management: a key component of the COVID-19 global response. https://apps.who.int/iris/handle/10665/331775 (2020)
18. WHO (2020) Coronavirus Disease 2019 (COVID-19)—Situation Report 86
19. WHO (2020) Coronavirus Disease 2019 (COVID-19)—Situation Report 100
20. Gallotti R et al (2020). Preprint at. https://doi.org/10.1101/2020.04.08.20057968
21. Velasquez N et al (2020) Preprint at https://arxiv.org/abs/2004.00673
22. Sundareswaran V, Firth-Butterfield K (2020) Chatbots provide millions with COVID-19 information everyday, but they can be improved- here's how. World Economic Forum https://www.weforum.org/agenda/2020/04/chatbots-covid-19-governance-improved-here-s-how/
23. WHO (2020) Digital health. https://www.who.int/health-topics/digital-health
24. Topol EJ (2019) Nat Med 25:44–56
25. WHO (2020) Access to COVID-19 Tools (ACT) Accelerator
26. The CONSORT-AI and SPIRIT-AI Steering Group (2019) Nat Med 25:1467–1468
27. Collins GS, Moons KGM (2019) Lancet 393:1577–1579
28. UN Human Rights (2020) Press briefing note on Americas/prison conditions. https://shar.es/aHIzht
29. Cohen J (2020) Science. https://www.sciencemag.org/news/2020/01/chinese-researchers-reveal-draft-genome-virus-implicated-wuhan-pneumonia-outbreak
30. Hadfield J et al (2018) Bioinformatics 34:4121–4123
31. Zastrow M (2020). Nature. https://doi.org/10.1038/d41586-020-00740-y
32. Teran J (2020) UN Department of Economic and Social Affairs. https://covid-19-response.unstatshub.org/open-data/hdx-making-covid19-data-accessible/
33. WHO (2020) Epidemic intelligence from open sources. https://www.who.int/eios
34. GitHub (2020) UNGlobalPulse/covid19-literature-search. https://github.com/UNGlobalPulse/covid-19-literature-search
35. Wang LL et al (2020) Preprint at https://arxiv.org/abs/2004.10706

Design VLSI Architecture for 2_D DWT Using NEDA and KSA Technique

Satyendra Tripathi, Bharat Mishra, and Ashutosh Kumar Singh

Abstract Image compression emanating utilizations of data compression on digital images. The Discrete Wavelet Transform (DWT) is getting familiar with an extreme move in image processing. To eliminate the drawback in the JPEG standard and rising areas of mobile and Internet communications, the new JPEG2000 standard has been created based on the principles of DWT. With the prevalence of the 2_D DWT, the technology has seen predictable improvement. All things considered, there is still a need for more robust and efficient compression technology using DWT. Hence, a 2_D DWT is implementing with a multiplier-less NEDA technique. NEDA is consisting of a buffer, ROM, and adder. There are many types of adders used in digit circuits but Kogge Stone Adder (KSA) is used in the proposed method. KSA is a very efficient adder and consists of the XOR gate and HA. Proposed Schemes are simulated Xilinx software with the VHSIC Hardware Description Language (VHDL) platform and calculates parameter i.e. speed and frequency.

Keywords 2_D discrete wavelet transform (DWT) · New efficient distributive arithmetic (NEDA) · Read only memory (ROM) · Half adder (HA) · Kogge stone adder (KSA)

1 Introduction

A portion of its highlights are; versatile time-recurrence windows, lower associating mutilation for signal handling [1, 2]. As a result of these highlights 1_D DWT [3, 4] and 2_D DWT [5] can be utilized in different applications, for example, numerical examination [6], analysis of various signals, image coding, pattern recognition, watermarking. It can also be used in Biomedicine [7]. Lots of calculations and plans have been prompted during the most recent three decades to make an appropriate

S. Tripathi (✉) · B. Mishra
MGCGV Chitrakoot, Chitrakoot, India
e-mail: satyendra.mgcgv16@gmail.com

A. K. Singh
IIIT-Allahabad, Prayagraj, Allahabad, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4_66

equipment execution of 1_D DWT [8] and 2_D DWT. DWT gives a proper computing technique for the representation of a wide sample of signals. DWT gives sufficient information about the synthesis and testing of the original signal with a reduction in time.

The important guideline is the end of monotonous signs and information that can reduce the methodology. From a logical point of view, adjust the pixel framework of 2-D [9] size, into a scientific, numerically uncorrelated informational index. The change in the information can be seen previously, and then the picture is transmitted. Later when the picture will be needed, it is straightforwardly decompressed to revamp the ordinal one. There may be a slight difference of information from the unique picture; however, the measure of data is the same. The attention on picture pressure was in presence for years now. At first, the examination takes a shot at the picture pressure method that was centered on diminishing the span of video for broadcasting. The utilization of picture pressure in such a situation, won't diminish the measure of the mixed media record, and also exceeds the speed of transmission to get successful and productive usage of transmission of data.

In any case, locale and speed are large conflicting necessities to improve speed comes to fruition generally in greater areas. Earlier, control usage was helping conversely to improve the area and speed. Regardless, control is being given more noteworthiness as zone and speed because of sublime advancement of minimized and remote handheld blended-media contraptions. The usage of power is the most essential factor for these devices [10].

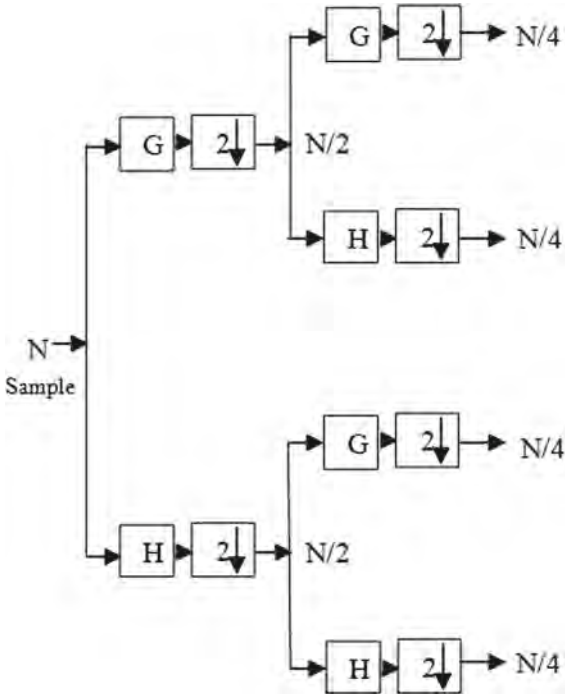
2-D DWT is used to many applications like medical image, tomography, speech and voice, pattern recognition, biomedicine, and computer graphics. 2-D DWT is a technique that divided the information in the multiresolution frequency domain. 2-D DWT is divided information in four parts i.e. normal, normal distinction, contrast normal and contrast distinction is present Fig. 1. G means the average of two samples and H means the difference between two samples.

2 Literature Review

Samit et al. [1], abridge the circuit intricacy and enhance the execution are proposed; DWT is applied to the periodic and non-periodic signal that used to speech and video signal. This article designed 9/7 DWT based on DA and KSA. The drawback of this paper is KSA does not properly work in all the binary bits and sources of info, which profoundly expands the circuit unpredictability. Proposed procedures depend on rapid region productive 2-D DWT utilizing 9/7 channel based altered MDA technique also, KSA. MDA strategy is connected to the bottommost and eminent pass channel of the DWT. This method comprises a viper, move to enroll, and handout of a multiplier. Plan and result are actualized in Xilinx programming and confirmed resistor exchange level and waveform.

Biswas et al. [2], proposed a programmable one-dimensional discrete wavelet bundle change processor. Contrasted and existing designs, the proposed processor can

Fig. 1 2-D DWT



complete both wavelet bits of information to be deserted are found. In any case, the picture pressure innovation has turned into a mainstream decision for dealing with the very mind-boggling spatial determination of the current picture catching sensors. The fantastic video transmission for TV and media has additionally lifted the utilization of picture pressure. In addition, picture pressure attempts an important association in various basic and incalculable applications, for example, video conferencing, TV, biomedical imaging, remote detecting (utilizes satellite symbolism for atmosphere and earthly data total), FAX, space investigation, records handling, and transmission, controlling remote vehicle for military, squander administration, and so on.

Mamatha et al. [6], the previously mentioned application particularly demonstrates the necessity for sufficient capacity, significant data transfer capacity for transmission, and delayed transmission period for a picture, sound, and video signals. In this present circumstance, the innovation that can be gotten to and executed under these conditions is the picture pressure. Picture pressure gives an answer for every one of the difficulties looked in the above zones. It lessens the measure of sight and sound information required for capacity and transmission. It packs the information, at that point stores it, or transmits it. At the point, when the information is required to be perused, at that point the picture decompression replicates the first data. For example, when the pressure proportion is 32:1, at that point the capacity zone, exchange speed, data transfer capacity utilization, and transmission postpone requirements can be limited by a factor of 32, with no misfortune in quality. The portrayal of the photo is first changed from the given spatial space to a substitute area with the assistance of well-known flag changing and coding procedures.

Martina et al. [7], the existing method consumed more memory is used to storing the intermediate computational multiplier based DWT. The DWT needs to process gigantic measures of information at high speed, making it unsuitable for real-time image compression with better performance. The objective of the research work is to infer effective designs for the equipment execution of the multiplier-less DWT, compared to the existing architectures.

Kumar et al. [11], the optimization factors are the speed, low area, and hardware multifaceted nature required, with the size of the given information picture and the necessary degrees of deterioration. The MDA based DWT design for the 9/7 wavelet channel coefficient multiplier-less engineering is enhanced, to accomplish better speed and higher equipment usage, by utilizing a solitary clock to focus on tasks updates. In the cutting edge mechanical industry, it is very difficult to go over a field that isn't impacted by the advanced picture preparing. It assumes a vital part of all specialized areas, somehow. Keeping in mind the end goal to keep up a conservative portrayal, just a few applications are depicted that are identified with the proposed work. By and large, the fields that make utilization of advanced picture handling can be divided into morphology, criminology, photography, microscopy, biomedical imaging, remote detecting, transportation, explore sciences, military application, and numerous others.

3 Proposed Methodology

Multiplier-less NEDA and KSA technique are shown in Fig. 2. In this figure, input morsel is passed through shift registers and all input–output is added symmetrically with the help of adder. The LPS and HPS input are passed through NEDA and KSA technique and the output becomes Y_{HPS} and Y_{LPS}.

3.1 Example of Multiplier-Less HPS

If it takes the HPS coefficients such as g_0, g_1, g_2, and g_3 multiply by v_1, v_2, v_3, and v_4 then multiplier-less 1-D DWT [12, 13] HPS output is

$$Y_{\text{HPS}} = \begin{bmatrix} g_0 & g_1 & g_2 & g_3 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$$

where,

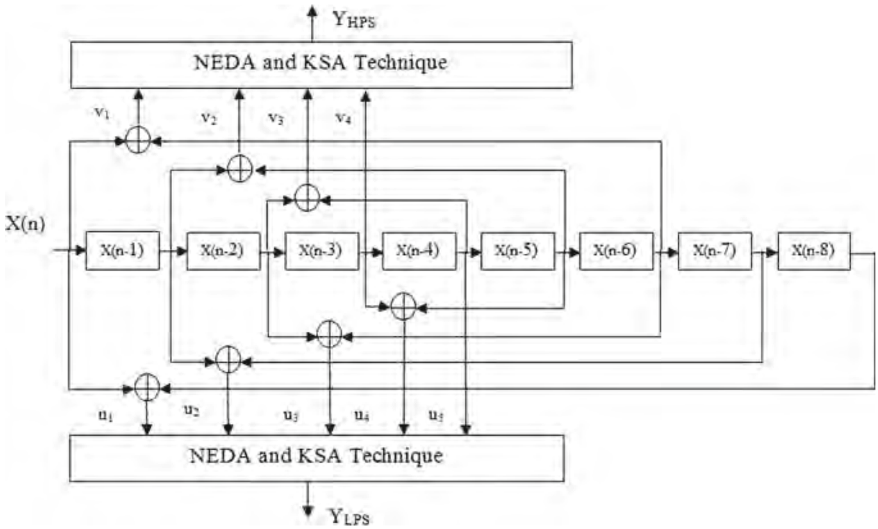


Fig. 2 Proposed 1-D DWT using NEDA and KSA technique

$v_1 = X(n) + X(n-6)$

$v_2 = X(n-1) + X(n-5)$

$v_3 = X(n-2) + X(n-4)$

$v_4 = X(n-3).$

$$Y_{\mathrm{LPS}} = \begin{bmatrix} 71 & -38 & -4 & 6 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$$

So,

$$Y_{\mathrm{HPS}} = \begin{bmatrix} 0\,1000111 & 11011010 & 11111100 & 00000110 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$$

All the HPS coefficient arranges down to up is below:

$$Y_{\mathrm{H}} = \begin{bmatrix} 1\,0\,0\,0 \\ 1\,1\,0\,1 \\ 1\,0\,1\,1 \\ 0\,1\,1\,0 \\ 0\,1\,1\,0 \\ 0\,0\,1\,0 \\ 1\,1\,1\,0 \\ 0\,1\,1\,0 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$$

All rows pass through look up table and replace the LPS coefficient to input

$$Y_{\mathrm{H}} = \begin{bmatrix} 1\,0\,0\,0 \\ 1\,1\,0\,1 \\ 1\,0\,1\,1 \\ 0\,1\,1\,0 \\ 0\,1\,1\,0 \\ 0\,0\,1\,0 \\ 1\,1\,1\,0 \\ 0\,1\,1\,0 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_1 + v_2 + v_4 \\ v_1 + v_3 + v_4 \\ v_2 + v_3 \\ v_2 + v_3 \\ v_3 \\ v_1 + v_2 + v_3 \\ v_2 + v_3 \end{bmatrix}$$

Let $v_1 = 4$, $v_2 = 4$, $v_3 = 4$ and $v_4 = 2$ and put above equation and last row value is represented by 2's complement value.

Then

$$\begin{aligned} &KK_1 = v_1 = 0100,\ \ KK_2 = v_2 + v_2 + v_4 = 1010,\ \ KK_3 = v_1 + v_3 + v_4 = 1010, \\ &KK_4 = v_2 + v_3 = 1000,\ \ KK_5 = v_2 + v_3 = 1000,\ \ KK_6 = v_3 = 1000 \\ &KK_7 = v_1 + v_2 + v_3 = 1100,\ \ KK_8 = v_2 + v_3 = \text{not } (1000) + \ ''0001'' = 1000 \end{aligned}$$

All KK_1 to KK_8 value passed through sign extension block then

$$\begin{aligned} &KK_1 = v_1 = 00100, KK_2 = v_2 + v_2 + v_4 = 01010, KK_3 = v_1 + v_3 + v_4 = 01010, \\ &KK_4 = v_2 + v_3 = 01000, KK_5 = v_2 + v_3 = 01000, KK_6 = v_3 = 001000, \\ &KK_7 = v_1 + v_2 + v_3 = 01100, KK_8 = v_2 + v_3 = not(01000) + \ ''0001'' = 11000 \end{aligned}$$

KK_1 is left shift one bit and add KK_2 and store output YY_1

$$\begin{aligned} &= 0'00100 \\ &\ \ + 01010 \\ YY_1 &= 011000 \end{aligned}$$

YY_1 is left shift one bit and add KK_3 and store output YY_2

$$\begin{aligned} &= 0'011000 \\ &+ 01010 \\ YY_2 &= 1000000 \end{aligned}$$

YY_2 is left shift one bit and add KK_4 and store output YY_3

$$\begin{aligned} &= 0'1000000 \\ &+ 01000 \\ YY_3 &= 10000000 \end{aligned}$$

YY_3 is left shift one bit and add KK_5 and store output YY_4

$$\begin{aligned} &= 0'10000000 \\ &+ 01000 \\ YY_4 &= 100000000 \end{aligned}$$

YY_4 is left shift one bit and add KK_6 and store output YY_5

$$\begin{aligned} &= 0'100000000 \\ &+ 00100 \\ YY_5 &= 0110000000 \end{aligned}$$

YY_5 is left shift one bit and add K_7 and store output YY_6

$$\begin{aligned} &= 0'11100000000 \\ &+ 01100 \\ YY_6 &= 100100000000 \end{aligned}$$

YY_6 is left shift one bit and add KK_8 and store output YY_7

$$\begin{aligned} &= 0'10010000000 \\ &+ 11000 \\ Y_7 &= 1000010000000 \end{aligned}$$

Final output $Y_{\mathrm{HPS}} = YY_7 = (0000010000000)_2$ (Carry Reject).

4 Simulation Result

Figure 3 is the view of the technology of first level DWT. Here ‘e’ is the input of 4-bit, clk is the clock signal applied and here two output Comes each of 12-bit.one is ‘yh’ means the output for the eminent pass filter and the other is; yl’ means the output for the bottommost pass filter.

Figure 4 shows the RTL view of first level DWT. This view includes the shift registers, KSA, D-flip flops, and all its components. First, the input passes through D-flip flops, and then symmetrically addition is performed through KSA and then after using the NEDA technique, the final output comes.

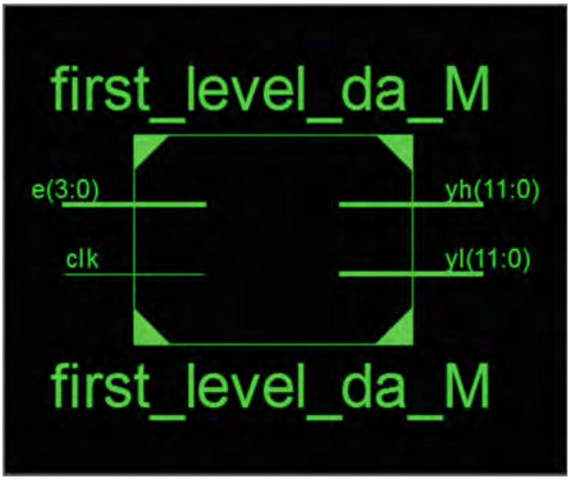
Figure 5 shows the view of the technology of second level DWT. Here the input is of 4-bit and the two outputs are of 19 bit. One is ‘yh1’ which is the output for the eminent pass filter and the other is ‘yl1’ which is the output for the bottommost pass filter.

Figure 6 shows the RTL view of 2-D DWT. It has all the components of 2-D DWT. It contains all the shift registers, D-flip flops, BK adder [14]. This RTL schematic depends on the view of technology.

Figure 7 shows the waveform of the 2_D DWT. Here the input is given as ‘0011’ and the output finally comes for both the filters. ‘yh’ is ‘1111101000000000000’ for eminent pass filter output and ‘yl’ is ‘0000011000000000000’ for the bottommost pass filter output.

In this Table 1, when it comes to 1_D and 2_D DWT in the case of ‘a number of slices’ the proposed delineation is 56.6% exceptional than the previous delineation, in the case of ‘number of flip flops’ the proposed delineation is 33.3% exceptional than the previous delineation, in case of ‘number of LUTs’ the proposed delineation is 29.2% exceptional than the previous delineation, in case of ‘maximum delay’ the proposed delineation is 25.19% exceptional than the proposed delineation. Similarly,

Fig. 3 VTS for 1-D DWT



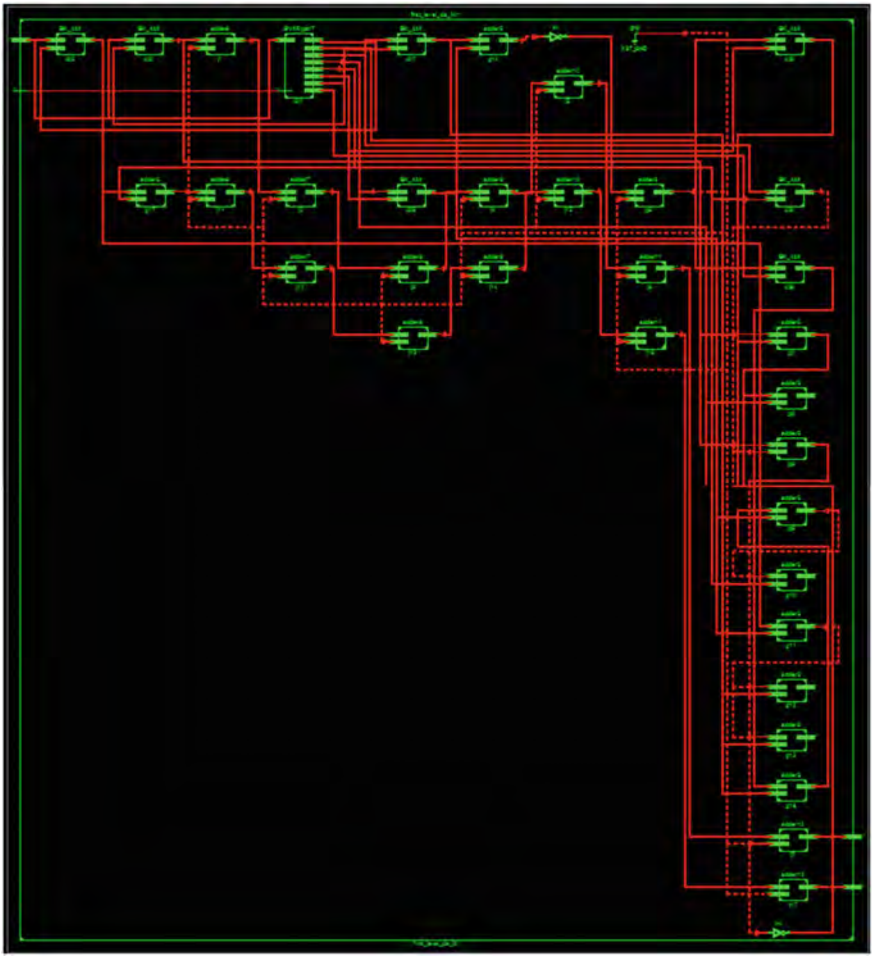


Fig. 4 RTL for 1-D DWT

when it comes to 2-D DWT, the proposed delineation is much better than the previous delineation.

5 Conclusion

The Discrete Wavelet Transform is being designed using the New Efficient Distributive Arithmetic technique and Kogge Stone Adder. This transform breaks the signal into a mutually orthogonal set of wavelets. It is the implementation of a wavelet transform in which it uses some discrete set of wavelets and translations after obeying

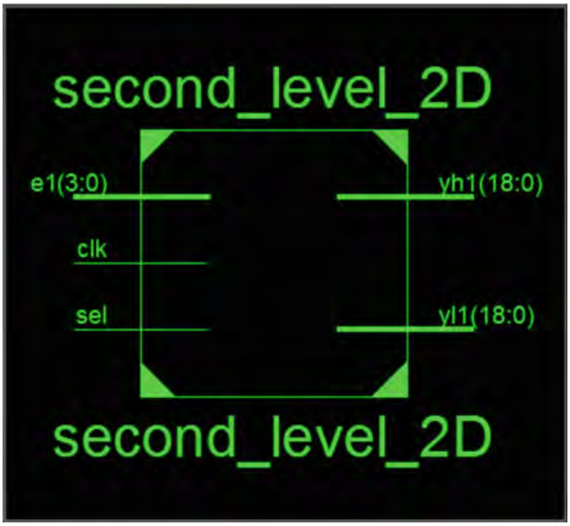


Fig. 5 VTS for 2-D DWT

some rules. The DWT provides complete information for analysis and synthesis of signals and also reduces the time. When it comes to 1_D and 2_D DWT in the case of 'a number of slices' the proposed delineation is 56.6% exceptional than the previous delineation, in the case of 'number of flip flops' the proposed delineation is 33.3% exceptional than the previous delineation, in case of 'number of LUTs' the proposed delineation is 29.2% exceptional than the previous delineation, in case of 'maximum delay' the proposed delineation is 25.19% exceptional than the previous delineation.

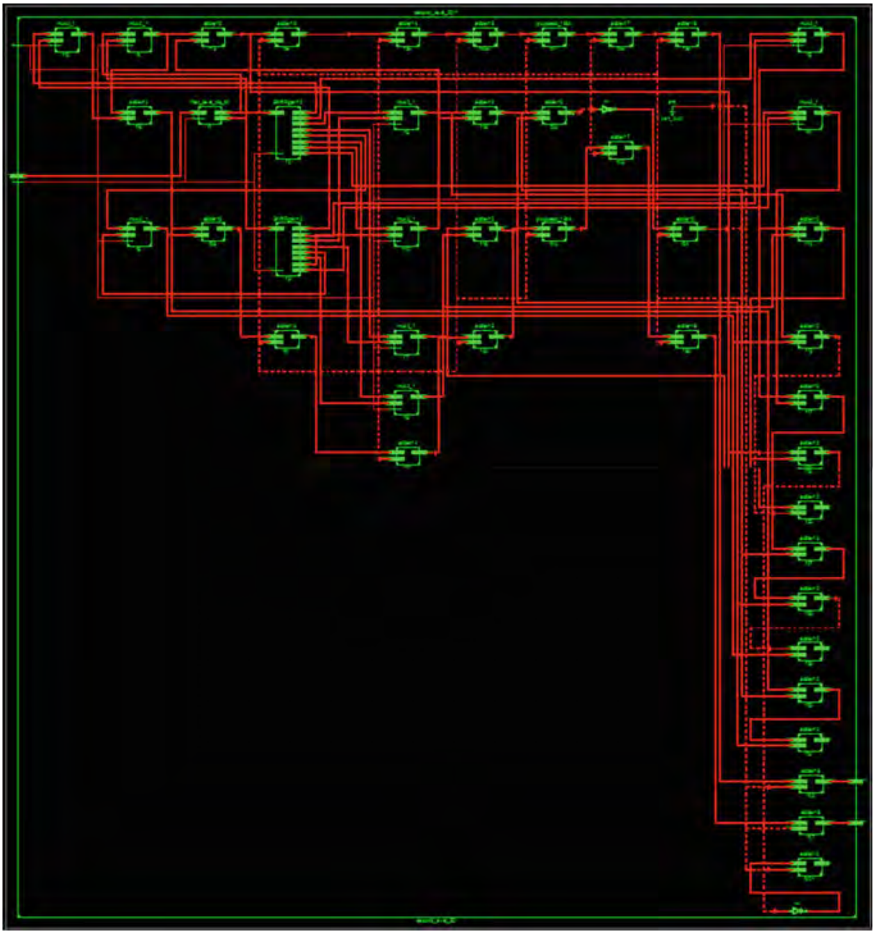


Fig. 6 RTL for 2-D DWT

Fig. 7 VHDL Test-bench in 2-D DWT

Table 1 Comparison of result with previous 2-D DWT implementation

Parameter	1-D DWT		2-D DWT	
	Previous design	Proposed design	Previous design	Proposed design
Number of slice	346	150	752	514
Number of slice flip flop	48	32	454	224
Number of LUTs	359	254	1393	899
MCPD	23.146 ns	17.316 ns	28.998 ns	

References

1. Dubey SK, Kourav AK, Sharma S (2017) High speed 2-D discrete wavelet transform using distributed arithmetic and Kogge Stone adder technique. In: International conference on communication and signal processing, April 6–8, India
2. Biswas R, Malreddy SR, Banerjee S (2017) A high precision-low area unified architecture for lossy and lossless 3D multi-level discrete wavelet transform. IEEE Trans Circ Syst Video Technol 45(5):1–11
3. Cao X, Xie Q, Peng C, Wang Q, Yu D (2006) An efficient VLSI implementation of distributed architecture for DWT. In: Proceedings of IEEE workshop on multimedia and signal process, pp 364–367
4. Tewari G, Sardar S, Babu KA (2011)High-Speed & Memory Efficient 2-D DWT on Xilinx Spartan3A DSP using scalable polyphase structure with DA for JPEG2000 Standard.978-1-4244-8679-3/11/$26.00 ©2011 IEEE
5. Martina M, Masera G (2006) Low-complexity, efficient 9/7 wavelet filters VLSI implementation. IEEE Trans Circuits Syst II, Expr Brief 53(11):1289–1293
6. Mamatha I, Tripathi S, Sudarshan TSB (2016) Pipelined architecture for filter bank based 1-D DWT. In: International conference on signal processing and integrated networks (SPIN), pp 47–52
7. Martina M, Masera G, Ruo Roch M, Piccinini G (2015) Result-biased distributed-arithmetic-based filter architectures for approximately computing the DWT. In: IEEE transactions on circuits and systems—I: regular papers, vol 62, No 8
8. Mohanty BK, Meher PK (2009) Efficient multiplierless designs for 1-D DWT using 9/7 filters based on distributed arithmetic. ISIC 2009
9. Mohanty BK, Meher PK (2011) Memory efficient modular VLSI architecture for high-throughput and low-latency implementation of multilevel lifting 2-D DWT. IEEE Trans Signal Proces 59(5)
10. Alam M, Rahman CA, Jullian G (2003) Efficient distributed arithmetic based DWT architectures for multimedia applications. In: Proceedings of IEEE workshop on SoC for real-time applications, pp 333–336
11. Mohanty BK, Meher PK (2013) Memory-efficient high-speed convolution-based generic structure for multilevel 2-D DWT. IEEE Trans Circ Syst Video Technol 23(2):353–363
12. Zhao X, Vi Y, Erdogan AT, Arslan T (2000) A high-efficiency reconfigurable 2-D discrete wavelet transform engine for JPEG 2000 implementation on next generation digital cameras. 978-1-4244-6683-2/10/$26.00 © 2010 IEEE
13. Baviskar A, Ashtekar S, Chintawar A, Baviskar J, Mulla A (2014) Performance analysis of sub-band replacement DWT based image compression technique. 978-1-4799-5364-6/14/$31.00 © 2014 IEEE
14. Ramkumar B, Kittur HM (2012) Low-power and area-efficient carry select adder. IEEE Trans Very Large Scale Integr (VLSI) Syst 20(2)
15. Sahoo R, Roy S, Chaudhuri SS (2014) Haar wavelet transform image compression using run length encoding. In: International conference on communication and signal processing, April 3–5, 2014, India

Author Index

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 733,
https://doi.org/10.1007/978-981-33-4909-4

Advances in Intelligent Systems and Computing 1318

S. Smys
João Manuel R. S. Tavares
Robert Bestak
Fuqian Shi *Editors*

Computational Vision and Bio-Inspired Computing

ICCVBIC 2020

Springer

Advances in Intelligent Systems and Computing

Volume 1318

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within “Advances in Intelligent Systems and Computing” are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Indexed by DBLP, EI Compendex, INSPEC, WTI Frankfurt eG, zbMATH, Japanese Science and Technology Agency (JST).
All books published in the series are submitted for consideration in Web of Science.

More information about this series at http://www.springer.com/series/11156

S. Smys · João Manuel R. S. Tavares ·
Robert Bestak · Fuqian Shi
Editors

Computational Vision and Bio-Inspired Computing

ICCVBIC 2020

Springer

Editors
S. Smys
Department of Computer Science
and Engineering
RVS Technical Campus
Coimbatore, Tamil Nadu, India

Robert Bestak
Czech Technical University in Prague
Prague, Czech Republic

João Manuel R. S. Tavares
Departamento de Engenharia Mecanica
SDI
Faculty of Engineering
Universidade do Porto
Porto, Portugal

Fuqian Shi
College of Graduate Studies
University of Central Florida
Orlando, FL, USA

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-981-33-6861-3 ISBN 978-981-33-6862-0 (eBook)
https://doi.org/10.1007/978-981-33-6862-0

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

We are honored to dedicate the proceedings of ICCVBIC 2020 to all the participants and editors of ICCVBIC 2020.

Preface

It is with deep satisfaction that I write this preface to the proceedings of the ICCVBIC 2020 held in RVS Technical Campus, Coimbatore, Tamil Nadu, November 19–20, 2020.

This conference was bringing together researchers, academics and professionals from all over the world, experts in *Computational Vision and Bio-Inspired Computing.*

This conference particularly encouraged the interaction of research students and developing academics with the more established academic community in an informal setting to present and to discuss new and current works. The papers contributed the most recent scientific knowledge known in the field of computational vision, soft computing, fuzzy, image processing and bio-inspired computing. Their contributions helped to make the conference as outstanding as it has been. The Local Organizing Committee members and their helpers have put much effort into ensuring the success of the day-to-day operation of the meeting.

We hope that this program will further stimulate research in computational vision, soft computing, fuzzy, image processing and bio-inspired computing and provide practitioners with better techniques, algorithms and tools for deployment. We feel honored and privileged to serve the best recent developments to you through this exciting program.

We thank all the authors and participants for their contributions.

Coimbatore, India

Dr. S. Smys
Conference Chair
ICCVBIC 2020

Acknowledgements

ICCVBIC 2020 would like to acknowledge the excellent work of our conference organizing committee and keynote speakers for their presentation on November 19–20, 2020. The organizers also wish to acknowledge publicly the valuable services provided by the reviewers.

On behalf of the editors, organizers, authors and readers of this conference, we wish to thank the keynote speakers and the reviewers for their time, hard work and dedication to this conference. The organizers wish to acknowledge Dr. Smys, Dr. João Manuel R. S. Tavares, Dr. Robert Bestak and Dr. Fuqian Shi for the discussion, suggestion and cooperation to organize the keynote speakers of this conference. The organizers also wish to acknowledge the speakers and participants who attended this conference. Many thanks are given for all the persons who helped and supported this conference. ICCVBIC 2020 would like to acknowledge the contribution made to the organization by its many volunteers. The members have contributed their time, energy and knowledge at a local, regional and international levels.

We also thank all the Chair Persons and conference committee members for their support.

About the Conference

This conference proceedings volume contains the written versions of most of the contributions presented during the conference of ICCVBIC 2020. The conference provided a setting for discussing recent developments in a wide variety of topics including computational vision, fuzzy, image processing and bio-inspired computing. The conference has been a good opportunity for participants coming from various destinations to present and discuss topics in their respective research areas.

ICCVBIC 2020 conference tends to collect the latest research results and applications on *Computational Vision and Bio-Inspired Computing*. It includes a selection of 63 papers from 227 papers submitted to the conference from universities and industries all over the world. All of the accepted papers were subjected to strict peer reviewing by 2–4 expert referees. The papers have been selected for this volume because of the quality and the relevance to this conference.

ICCVBIC 2020 would like to express our sincere appreciation to all the authors for their contributions to this book. We would like to extend our thanks to all the referees for their constructive comments on all papers; especially, we would like to thank the organizing committee for their hard working. Finally, we would like to thank Springer publications for producing this volume.

Dr. S. Smys
Conference Chair
ICCVBIC 2020

Contents

About the Editors

Dr. S. Smys received his M.E. and Ph.D. degrees all in Wireless Communication and Networking from Anna University and Karunya University, India. His main area of research activity is localization and routing architecture in wireless networks. He serves as Associate Editor of Computers and Electrical Engineering (C&EE) Journal, Elsevier, and Guest Editor of MONET Journal, Springer. He has served as a reviewer for IET, Springer, Inderscience and Elsevier journals. He has published many research articles in refereed journals and IEEE conferences. He has been General chair, Session Chair, TPC Chair and Panelist in several conferences. He is a member of IEEE and a senior member of IACSIT wireless research group. He has been serving as Organizing Chair and Program Chair of several international conferences, and in the Program Committees of several international conferences. Currently, he is working as Professor in the Department of CSE at RVS Technical Campus, Coimbatore, India.

João Manuel R. S. Tavares graduated in Mechanical Engineering at the Universidade do Porto, Portugal, in 1992. He also earned his M.Sc. degree and Ph.D. degree in Electrical and Computer Engineering from the Universidade do Porto in 1995 and 2001 and attained his Habilitation in Mechanical Engineering in 2015. He is a senior researcher at the Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial (INEGI) and Associate Professor at the Department of Mechanical Engineering (DEMec) of the Faculdade de Engenharia da Universidade do Porto (FEUP). His main research areas include computational vision, medical imaging, computational mechanics, scientific visualization, human–computer interaction and new product development.

Robert Bestak obtained Ph.D. degree in Computer Science from ENST Paris, France (2003) and M.Sc. degree in Telecommunications from Czech Technical University in Prague, CTU, Czech Republic (1999). Since 2004, he has been an Assistant Professor at the Department of Telecommunication Engineering, Faculty of Electrical Engineering, CTU. He participated in several national, EU, and third party research projects. He is the Czech representative in the IFIP TC6 organization, and chair of the working group TC6 WG6.8. He annually serves as Steering and Technical

Program Committee member of numerous IEEE/IFIP conferences (Networking, WMNC, NGMAST, etc.) and he is member of editorial board of several international journals (Computers & Electrical Engineering, Electronic Commerce Research Journal, etc.). His research interests include 5G networks, spectrum management and big data in mobile networks.

Dr. Fuqian Shi currently working as Graduate Faculty Scholar, College of Graduate Studies at the University of Central Florida, USA. He has published more papers in standard journal and conferences. He has acted as a committee member in many international conferences and editorial board in refereed journals. He acted as a reviewer in many reputed journals. He is a member of professional bodies like IEEE, IEEE Computer Society, Society, Man and Cybernetics, Computational Information System, ACM, Zhejiang Provincial Industrial Design Association and Information Management Association of Wenzhou City. His research interest includes computer networks, computer programming, computer graphics, image processing, data structure, operating system and medical informatics.

Smart Surveillance System by Face Recognition and Tracking Using Machine Learning Techniques

D. K. Niranjan and N. Rakesh

Abstract With the recent technological advancements, the theft and other such unnecessary activities can be controlled. At present, many traditional security systems are available to just record the situation, but the unnecessary activities will be already taken. The main objective of this research work is to control and stop the theft by tracking the burglar face or the stranger face image and send the notification to the owners. This system will not only alert the owners, and it can also give access to the door lock with the face recognition feature. In this way, the proposed system replaces the traditional security system with the feature like RFID, PIN method, etc. The accuracy of the proposed system is high, and it works in all the situations, where it can be operated with both battery and main power system. The system has its own database to recognize the person, and the person with registered facial image can only be given access to the door. The developed system has a low powered device with high accuracy system. Moreover, it also helps to reduce the unnecessary activities.

Keywords Face tracking · Smart door lock · Alert notification with image

1 Introduction

Nowadays, the culprits are more, and at least one case is registered on theft and other unnecessary activities. The other unnecessary activities will be controlled by deploying the security system and monitor the activities like crime and kidnapping. To stop these activities, a security system is much needed for a home, where each and every one comes at different time and requires access to the door. The solutions like keyless entry and theft control can be done through the facial recognition.

D. K. Niranjan (✉)
Department of Computer Science and Engineering, Amrita School of Engineering, Bengaluru, India

N. Rakesh
Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham, Bangalore, India
e-mail: n_rakesh@blr.amrita.edu

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*, Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_1

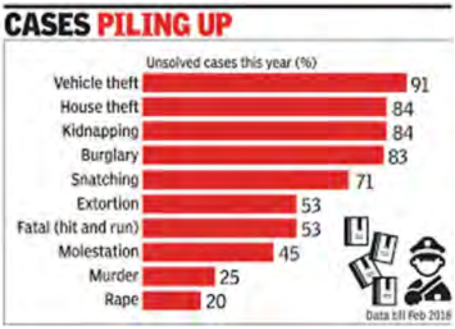


Fig. 1 Cases registered but not solved

To access the auto-lock door system, the proposed image recognition systems will automatically recognize the stored facial images. The face detection will be done by registering the faces to the system, where each and every person of the household should register the faces to access the system. By implementing the proposed system, there will be no need for any keys and cards to access it. Whenever a person wants to enter the house, a camera will be used to detect and recognize the face for taking the necessary action. If the unknown person comes in contact with the known person of the family member, then the system will be normal and unlock the system. If the unknown person tries to enter the house, then the system captures the face image and sends it to the mail notification of the registered person. The whole world is scared of these types of activities, and further to control these activities, this remains as one of the types of solution. This system will majorly help the old age people and children, where it will make them safe when there is no one at the home.

Figure 1 shows that the number of cases registered but not solved, where there is a greater number of cases registered are due to theft and kidnapping. To control these activities, the owners are required to adopt toward the latest technologies for maintaining safety. The surveillance system helps for the safety side, wherein with the proposed system, the face will be detected and starts recognition, if it is matched, then the system starts other process such as opening the door. If the system data is not matched, it sends the notification to the user. The process of recognition of face is done through the Haar cascade method. The face recognition remains as the complex process because some of them wear spectacles, and even in that situation, the system should recognize the face with no lag in time. This system works normally even in the night time, and also, the system uses its own light whenever it is dark to leverage better efficiency. After sending a mail notification, the system records intruder activities in a special folder. This system works even when the power is turned off by the intruder with a battery backup for 3 h.

The system captures the intruder and tracks the activities, where the camera moves along with them in that circumstances of area. The mail notification will be sent to the owners as well as to the nearby police station to stop the activities. If the person tries to hit the door, then the system buzzer beeps louder so that it can be heard to nearly 50 m of radius. If the user car is moved by the unknown person, it sends the alert mail; if the car is moved along with the family member, then it will not send the alert. If the unknown person tries to do mischievous activities to the registered family members or some harassment to the family members, then the system alerts the user.

This paper is organized as follows. Section 2 discusses the background work in the proposed research domain. In Sect. 3, the proposed system is explained. In Sect. 4, hardware description is provided. Section 5 describes the set-up and result of the system. Finally, Sect. 6 concludes the proposed research work and discuses about the future research scope.

2 Background Work

In the [1], it explains the attendance system that has been proposed through facial recognition technique, which is a real-time system, and in this system, they analyze the attendance of students and employee with check-in and check-out timings to maintain the status of students and employee. The unknown person data is sent to the mobile application. In this system, the main drawback is the entry point and to the exit point. If the person exits at the entry point, it is failed. They have used OpenCV and Haar cascade method to recognize the face. They have done only for seven persons of only 30 training data for each person. The accuracy level of the system is less as they have only 30 training data, where the dataset is less.

In the [2], this paper explains that the accessing of door is done by using the face recognition. They have used PIR sensor, when the person comes near to the door, the system recognizes and opens the door. The drawback of this paper is that, when the person comes near the system starts working until then the system will remain in the sleep mode. If power is off, then the system will not work because the system needs power to work.

In the [3], this paper explains that the system is only designed for sensitive area, where the authorized persons will only get the permission for entering the area, as the paper says that there will be fail in the recognition. The fail in the recognition represents that there is only less trained data. Out of 100 people, two persons recognition is failed. It is only used in small area. It aims to restrict the unknown person.

In the [4], this paper explains that the system controls the access permission. They have used principal component analysis for face recognition. In this, they have used GUI with the Arduino board for facial recognition. In this method, there is no training data, so the facial recognition expression will be changed every time for the same person, but storing of the person dataset will also be limited. For any facial recognition, there should be more dataset on the same person. Even the accuracy level of the person is also remaining less.

In the [5], this paper explains that the system has implemented a pin made-based smart door lock system. If the entered pin is wrong, then the system will send the alert message to the owner. In this method, the security designed is high. The system uses ultrasonic sensor, when the person comes near to the door, then only the system comes to the activation. If the unknown person enters the wrong pin, then the system sends message but not the intruder person.

3 Proposed System

The main motivation for this research work is to leverage new security devices for personal care [6]. The system is designed for the safety purpose, where the system works based on the facial recognition concept by using the Haar cascade method. If the registered face person wants to enter the home, they need to just see the camera once then the system detects and recognizes the person to open the door [7, 8]. If the person face is not registered, the system will consider it as the unknown person and sends the emergency alert message to the registered owner as well as to the nearest police station mail. If the main power is off, the system will work on the battery power for around 3 h. If the unknown person does the mischievous activities, the device will send the captured image to the registered mail id. Then, the system will start tracing the unknown person activities by recording and store in the separate folder for future purpose. If the person does the mischievous activities like damaging the door or the misbehaving, the system will buzzer at load such that it covers around 50 m radius. The block diagram is shown in the Fig. 2.

The proposed system requires to register the faces in the system at first stage [9]. The registered faces are stored in the dataset folder. In our system, we have registered ten persons data, each person's data is trained, and it is of 1000 images for best accuracy. In our system, we have used Haar cascade method for detecting the face. In our system for registering the faces, the device needs to run registering the face, and the system starts video format of around 4 min to capture all the images of

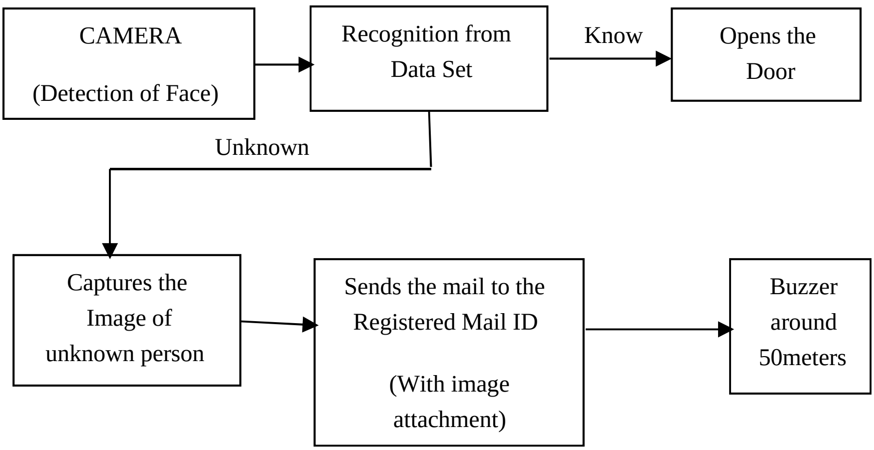


Fig. 2 Block diagram of process of the system

1000 images of each person [10]. The system converts the image of the registered person in to the gray scale image. These images are stored based on their IDs. The dataset stored in the database for recognition of the face. The process of registering the image in dataset is shown in Fig. 3.

The system consists of the data image stored in the database. Then, the system starts working as the surveillance system, where it detects the registered person based on the dataset stored in the device. Whenever the person comes, the device detects and recognizes, and it will give access to open the door to the registered system. If the person is not registered, then the system sends notification as well as it will give buzzer sound so that the action will be taken at the earliest and will be stopped. The process of recognition is shown in the Fig. 4.

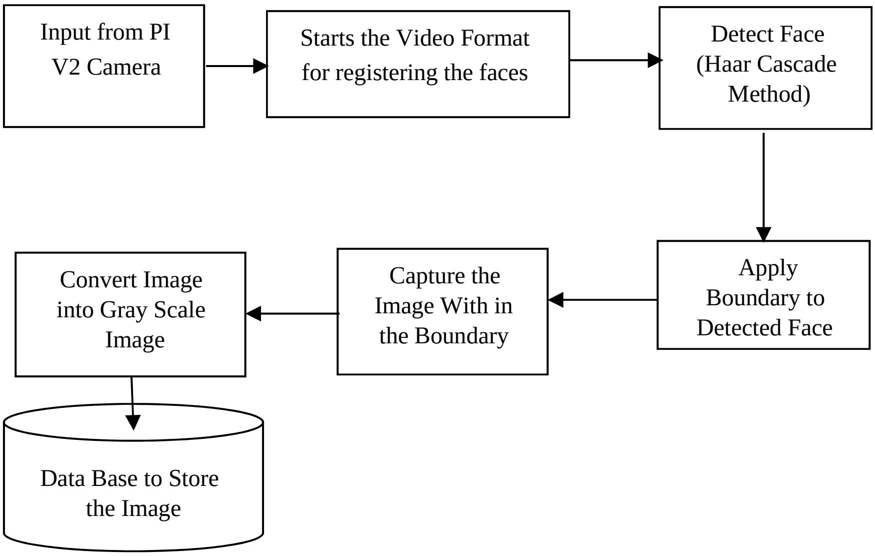


Fig. 3 Block diagram for registering the face

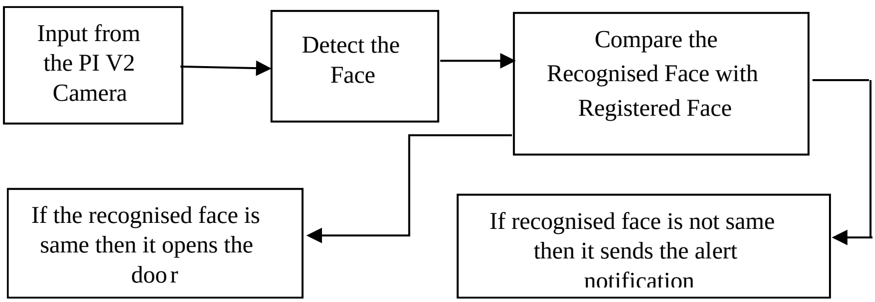
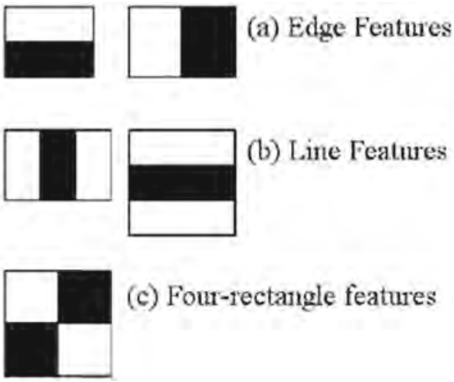


Fig. 4 Process of face recognition and access door

Fig. 5 Haar cascade feature



In this module, we recognize and detect the input images. This module is connected to the outer side of the door, where the captured image converted into Haar feature-based cascade classifiers. Comparision and Matching the face with this feature extraction image with the database [11]. The application-specific unit which consists of door lock security system, it is associated with Door lock system module of authentication module and it starts functioning according to results of the module to perform door lock open or close operation based on Face Recognition. If face is matched then give access to open the door, If the compared face is not matched with the stored data then it sends the email notification to the user.

In our proposed system, we are using Haar cascade approach for best efficiency. We have used OpenCV platform to recognize the face. In our approach, we have used local binary patterns histogram (LBPH) algorithm, based on the Haar cascade method. The system needs positive images which mean the images of faces, the negative images means the image without face. This algorithm is used to train the classifier. The Haar cascade feature is used which is shown in Fig. 5.

The Haar cascade is of three features, edge, line and four-rectangle features [12, 13]. This feature is where it recognizes only the face, and background is removed off, of the non-facial segment. This feature is used to train the classifier of the face where it detects parts of the face in detail. This classifier is used because it gives 95% of accuracy in the result, compared to other method.

The proposed home security system is a machine learning-based approach where a cascade function is trained from a lot of positive and negative images. It is then used to detect objects in other images. A procedural view of how this person detection works is shown in the Fig. 6, flow chart.

The mischievous condition is that, if the unknown person tries to access the door, or by damaging the door, or damaging the items under the camera visible side, then it acts as the mischievous section. If the culprits turns-off the main power supply, the system will run by the battery power, where it runs for around 3 h and it stores in the database as well as in the cloud.

If the culprits damage the system, then the data in the device will be uploaded to the cloud base for future assistance. Each and every part of the activities will be

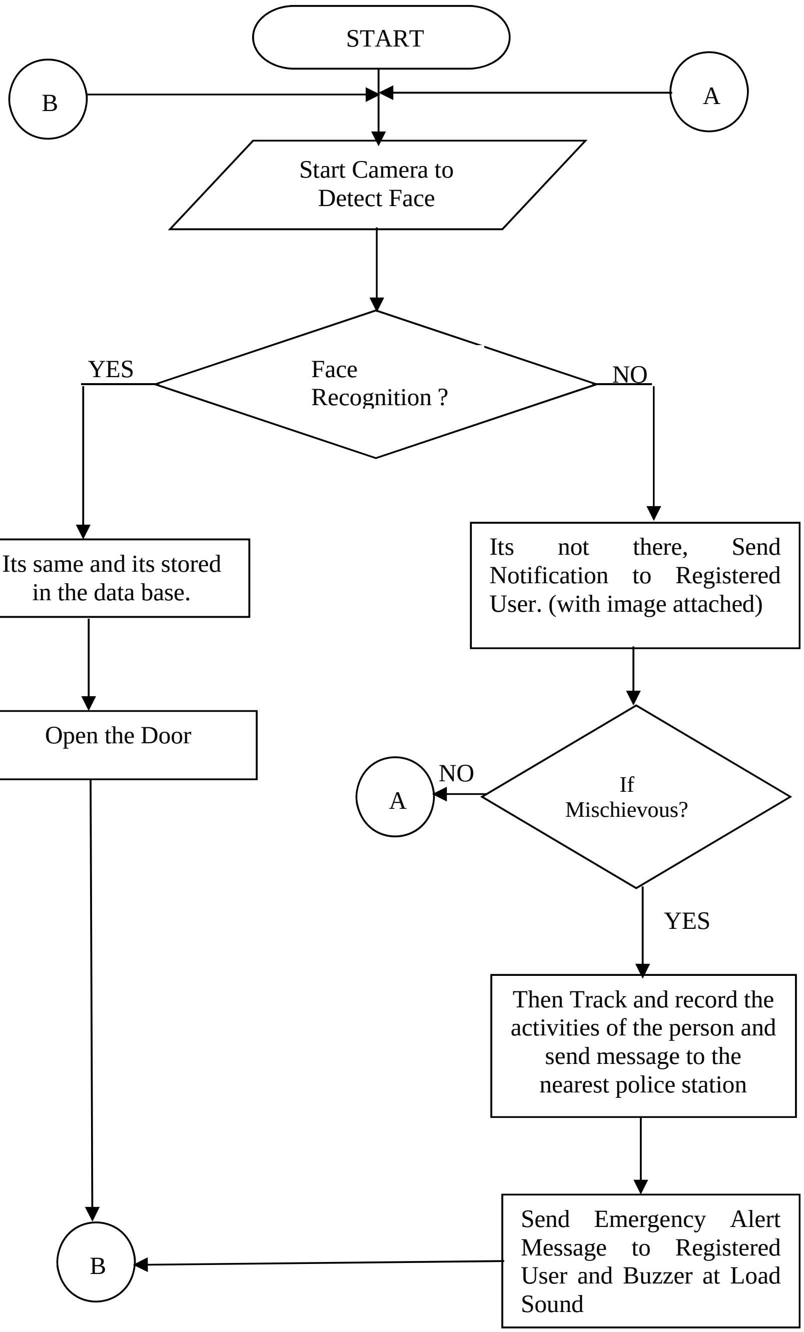


Fig. 6 Flow chart of the proposed system

recorded by tracking the culprits. These activities are traced based on the IR sensor placed. As the IR sensor is interfaced with the servo motor, it is rotated along the horizontal axis as well as the vertical axis. In our proposed system, it automatically turns on the lights when it is dark, for obtaining a best accuracy.

The Haar cascade is considered as the best method for facial recognition compared to other methods, where it gives 95% accuracy of the face. This system will track the person, depending on the condition, as the person can be passing the road or the courier person coming, then the system captures the person face and store, but other conditions it alerts the system. This system will only send mail notification to the person with image. This device alerts in case of high security breaking, where it alerts by emergency alarm and light. To avoid the high risk, the system can be used under these areas, home, bank lockers, ATM, government services, etc…

4 Hardware Description

In our system, we have used Raspberry Pi 3B+ for securing the whole process and the recognizing the faces at the low cost and low powered system [14]. It consists of in-built WIFI so that we can send mails to the registered mail IDs, where it consists of 1.4 GHz speed processor which is of 64-bit. It consists of CSI camera port which is in-built for the PIV2 camera, and for display purpose, it consists of HDMI port or the display port for PI display screen. The memory slot of the device can be expanded as the user requirements. The module is as shown in the Fig. 7.

The proposed system has used 5MP camera for best results, where the pi camera is connected to its CSI port by the flex cable, this camera only takes the picture and video of the activities not the sound of the activities. It has capable of taking at the 1080p clarity video or the picture so that it is most effective (Fig. 8).

The proposed system has used IR sensor for tracing the person. If the person moves, the sensor detects the position, and IR sensor is interfaced with the servo motors for tracking the unregistered person. IR sensor detects the person and

Fig. 7 Raspberry Pi 3B+

Fig. 8 PI V2 camera (5MP)

measures the distance, and then, the system changes the camera angle to cover the area. The IR sensor is shown as in the Fig. 9.

The servo motor is used to access the door and control the camera angle. These are done by interfaced with IR sensor, as the IR distance is calculated and sent to the motor, to pulse width modulation. As the PWM value is varied, the rotation angle is also varied and attains the position of the activities. The servo motor is as shown in the Fig. 10.

In our model, we have used LED light which is interfaced with the LDR, to turn ON/OFF the lights when needed.

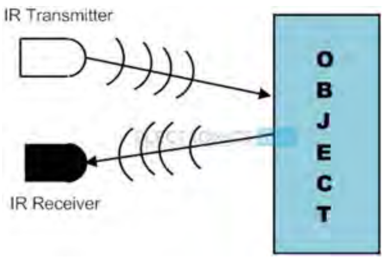


Fig. 9 IR sensor

Fig. 10 Servo motor (SG-90)

5 Results

The proposed system consists of Pi camera connected to the camera serial interface (CSI) to the Raspberry Pi, to register the faces to the dataset. In our model, we have OpenCV concept with the LBPH algorithm with the Haar cascade feature extraction, to recognize the face of the person. We have collected ten persons faces to be train the system, where 1000 images of face are trained of each person; these images are registered by their IDs. The 1000 images are registered by taking the video format by the system. Each video is of 4 min, in that, the person facial expression is changed every second. In these 4 min of video, the one person only face is registered and stored in dataset with ID. Whenever they see the camera, the smart door lock will be opened. The trained data of image is shown in the Fig. 11.

The Fig. 12 shows the face recognized, it compares with the dataset of images which has been trained. If the recognized images match with the dataset images, then door opens, else the mail notification will be sent with the image of the person.

Fig. 13 shows the email notification which has been sent to the registered mail id of the person. The mail consists of the "EMERGENCY" message, and the person image which has been captured will be sent to both user as well as to the nearest police station, depending on the situation. If any mischievous activities are carried out, then it sends to the nearest police station to stop the activities and take the necessary action. It is shown in Fig. 13.

The proposed model is shown in the Fig. 14, the system consists of sensors and actuators as shown in the figure. Servo motor tracks the face and to access the door. The buzzer is used to alert, for the radius of 50-meter if any miscellaneous activities happens.

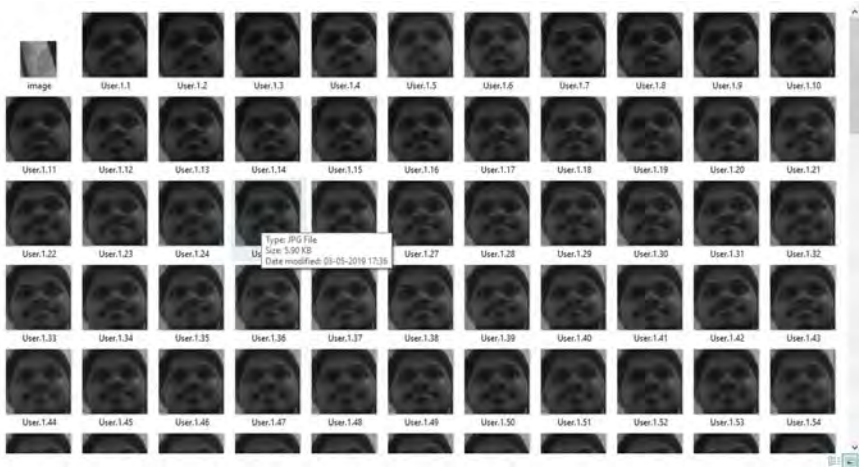


Fig. 11 1000 set of image are registered of one ID

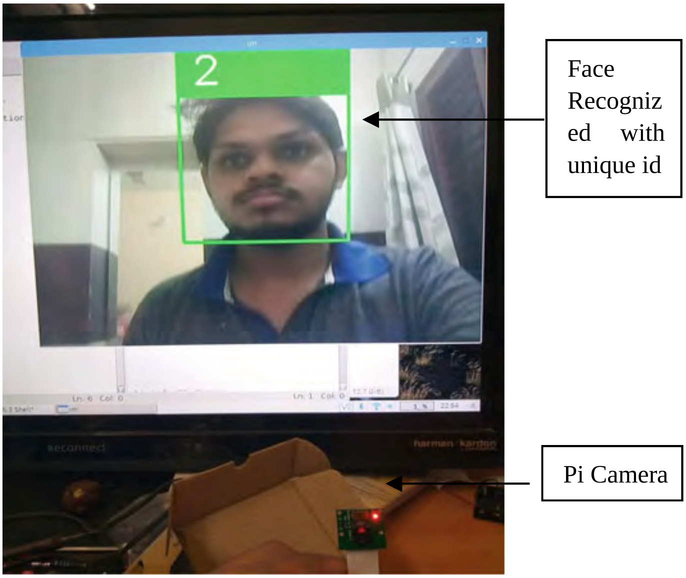


Fig. 12 Face recognized with their unique ID (door will be opened)

Fig. 13 Face not-recognized and emergency mail is sent

The proposed designed model is as shown in Figs. 14 and 15. Fig. 15a shows the night vision of the security system. The proposed system will give the best accuracy and efficiency of the system.

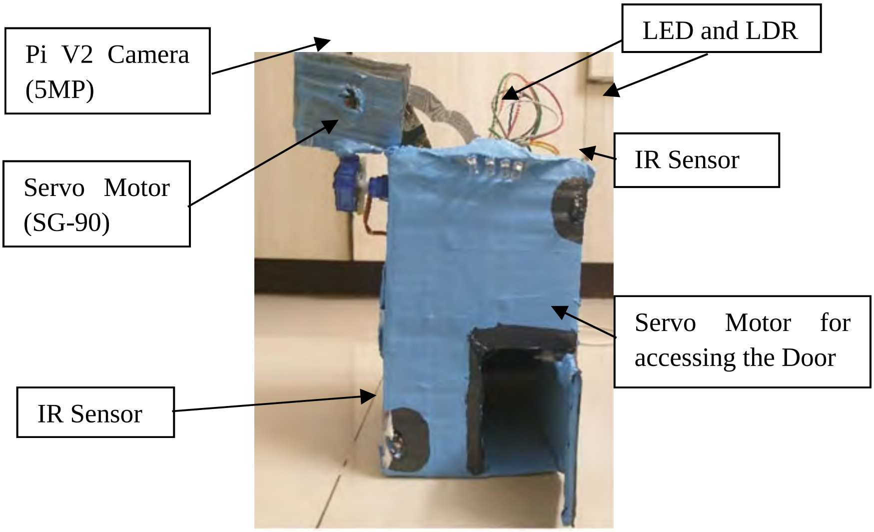


Fig. 14 Proposed model

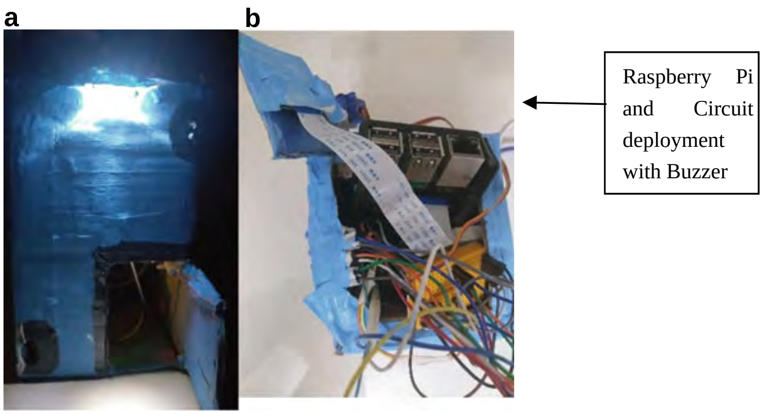


Fig. 15 Image (**A**) is at the night time, the lights are turned ON for recognition of face. Image (**B**) is of the circuit deployment of the system

6 Conclusion and Future Scope

6.1 Conclusion

The surveillance system is essentially required to enhance the safety of individuals, as the theft and other unnecessary activities are increasing in recent times. Some devices are existing, but it can only accommodate for limited person with limited database for training the classifier. In this system, the database storage accommodates 1000 images of each person, and the memory required for storing the database is too minimal, further it hardly takes around 300 MB for one person. This device will help all type of age group, so they will not worry about the biometric and password. The device will automatically move at the required angle, it detects and recognizes the person to open the door, and this process will take 2–3 s. The calibration of this device is performed more accurately for better efficiency. If an unregistered person is detected, the system sends alert message to the owner. If any mischievous activities occur, the system sends emergency message to the user as well as to the nearest police station. This method is faster and user-friendly compared to the traditional method, and in traditional method, it only implements smart lock system, when the person comes near to the door. But in the proposed method, it detects the person, who enters the secured area. Traditional method is usually a time consuming method, and if the power goes off, then the system does not work. But, in our system, it runs on the battery and does the efficient work with low power. This device can deployed in home, ATM, bank lockers, and in Government services.

6.2 Future Scope

It has been planned to elaborate the design, by sending SMS and implementing GSM, creating application for mobile device to easily access and add its own Internet data network and renewable energy.

References

1. M. Srivastava, A. Kumar, A. Dixit, Real time attendance system using face recognition technique, in *International Conference on Power Electronics & IoT Application in Renewable Energy and its control (PARC)* (2020)
2. A. Nag, J.N. Nikhilendra, M. Kalmath, IoT based door access control using face recognition, in *International Conference for Convergence in Technology (I2CT)* (2018)
3. D.A. Chowdhry, A. Hussain, M.Z.U. Rehman, F. Ahmad, Smart security system for sensitive area using face recognition, in *IEEE Conference on Sustainable Utilization and Development in Engineering and Technology* (2013)
4. N. Bakshi, V. Prabhu, Face recognition system for access control using principal component analysis, in *International Conference on Intelligent Communication and Computational Techniques (ICCT)* (2017)
5. Kyungil University, Gyeongsan, Gamasil-gil, Security and usability improvement on a digital door lock system based on internet of things. Int. J. Sec. Appl. **9**(8), 45–54 (2015)
6. V. Kamesh, M. Karthick, K. Kavin, M. Velusamy, R. Vidhya, Real time fraud anomaly detection in e-banking using data mining algorithm. South Asian J. Eng. Technol. **8**(1), 144–148 (2019)
7. V.S. Sureshkumar, D. Joseph Paul, N. Arunagiri, T. Bhuvaneshwaran, S. Gopalakrishnan, Optimal performance and security of data through FS-drops methodology. Int. J. Innov. Res. Eng. Sci. Technol. **3**(5), 1–7 (2017)

8. J.D. Irawan, E. Adriantantri, A. Farid, RFID and IOT for attendance management system. MATEC Web Conf (ICESTI) **164**, 01020 (2018)
9. M. Vijayakumar, E. Prabhakar, A hybrid combined under—over sampling method for class imbalanced datasets. Int. J. Res. Adv. Develop. (IJRAD) **02**(05), 27–33 (2018)
10. N. Rakesh, Performance analysis of anomaly detection of different IoT datasets using cloud micro services, in International Conference on Inventive Computation Technologies (ICICT 2016)
11. S. Sreedharan, N. Rakesh, Securitization of smart home network using dynamic authentication, in International conference on Computer Networks and Inventive Communication Technologies (ICCNCT) (2018)
12. R.S. Anuradha, R. Bharathi, K. Karthika, S. Kirithika, S. Venkatasubramanian, Optimized door locking and unlocking using IoT for physically challenged people, Issue 3 Harlow. Int. J. Innov. Res. Comput. Commun. Eng. **4** (2016). ISSN(Online): 2320–9801 ISSN (Print): 2320–9798.
13. N. Majgaonkar, R. Hodekar, P. Bandagale, Automatic door locking system. Int. J. Eng. Develop. Res IJEDR **4**. ISSN: 2321–9939.
14. L.S. Oliveira, D.L. Borges, F.B. Vidal, L. Chang, A fast eye localization and verification method to improve face matching in surveillance videos, in IEEE International Conference on Systems, Man, and Cybernetics (SMC) (2012), pp. 840–845

Object-Based Neural Model in Multicore Environments with Improved Biological Plausibility

R. Krishnan and A. Murugan

Abstract It is generally known that the computational neuroscience is usually considered as a mathematics-driven model. Computational neuroscience maps the functions of the neural populations to computer-based mathematical functions. Presently, numerous efforts have been taken to provide neural models that mimic the biological functions of neurons. This endeavor suggests the need for deploying comprehensive object-based basic neural models in the place of mathematical models that has been implemented by just using object-oriented programming languages. This paper describes the object-based model (OBM) and how it is different from the models using object-oriented languages. The consistent premise 'Everything is not learning' is set and is justified by the biological supports produced here. This paper also discusses about the alternate methods adopted by various researchers to prove that only a complete object-based design of neurons will have convergence toward the objective that mimic the actions of brain. This paper analyzes the need for different programming paradigms and concludes with a suggestion that the adaptive programming language 'Go' is a language that suits the implementation of such a model.

Keywords Brain mimicking · Intelligence · Artificial neural network · Object-based model (OBM) · Go language · Concurrency

1 Introduction

"Even with a million processors, we can only approach 1% of the scale of the human brain, and that is with a lot of simplifying assumptions," says Steve Furber [1]. This bold and truthful proclamation by the top scientist strengthened the following proposal as an endeavor toward a fresh approach in mimicking the brain.

The further part of this paper is divided into five parts where the first part pins the boundaries so that the terminologies and jargon around computational neuroscience

R. Krishnan (✉) · A. Murugan
PG & Research Department of Computer Science, Dr. Ambedkar Government Arts College (Affiliated to University of Madras), Vyasarpadi, Chennai 600039, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*, Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_2

do not lead to misinterpretations about what the brain mimicking is considered. The second part sets the objective for the research work. The third part analyzes the available models and the need and possibilities for change. The fourth part explains the requirements and structure of an OBM. The fifth part of the paper selects 'Go' language for the implementations of this model and also describes the proposal of OBM.

1.1 Artificial Intelligence and Thinking

Artificial intelligence (AI) is defined by computational scientists in many ways from making the machine intelligent to copying expertise. Still, it is wobbling in their minds according to their perceptual changes and the process of setting goals. Before adding another perception, it was imperative to arrive at a sensible definition of intelligence. Interestingly intelligence is also a thing that possesses endless definitions by psychologists and scientists [2]. In the work [3], there has been an effort to give a precise definition of intelligence. 'Intelligence measures an agent's ability to achieve goals in a wide range of environments' is the definition S. Legg and M. Hutter have arrived at. They themself call it an informal definition. In their future work [3], they formulate a formal measure for machine intelligence using the most accepted and most general framework, reinforcement learning. Their emphatic conclusion that IBM's DeepBlue chess supercomputer will earn low universal intelligence measure because it is specific to a single environment raises the following question. If intelligence is a measure, then what is measured? Intelligence measures the ability of the agent in achieving the goal. The goal is set by the environment and achieving it itself will not yield one a high intelligence measure. For a sample from a human intelligence point of view, losing a game to motivate a child or to avoid the expenditure of the winner's party may yield higher intelligence measures. Even though this is sidelined as 'humanness' in contrast to 'artificial intelligence,' this work considers this point to be the part of the agent's ability enhancement. If intelligence is the measure of ability then what is the measured ability of the brain? If the machine thinks and decides then the measure of it will be intelligence, where a machine having the highest measure will be called an AI machine. Knowing the consequences of 'thinking machine,' which was an effort to build a dream machine, the proposed work as of now is only an academic endeavor in analyzing the ability of the brain and evaluating the possibilities of its machine implementations with the available technologies and programming paradigms. Figure 1 illustrates a perception of the process of thinking, and the keywords extracted out of it are parallel/concurrent and non-deterministic.

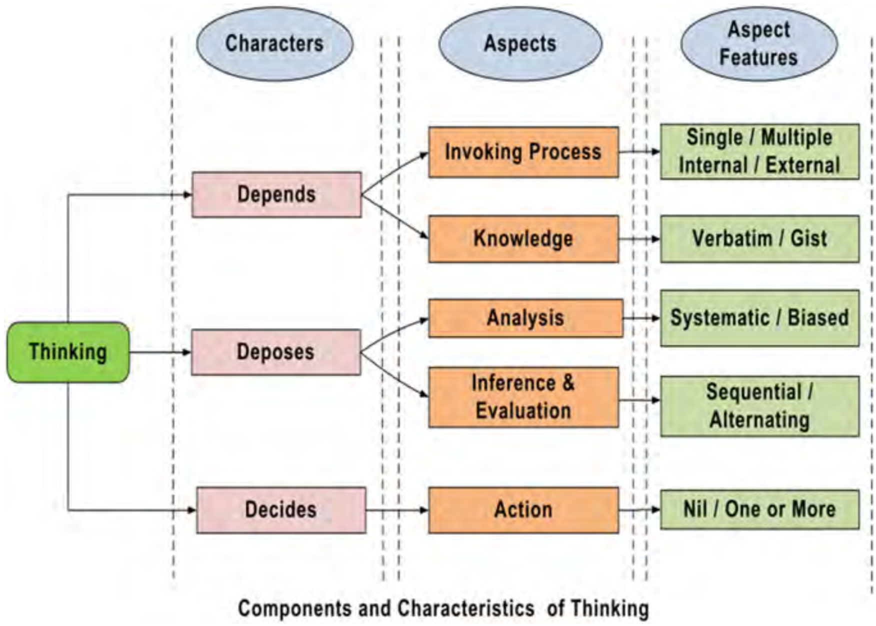


Fig. 1 Components and characteristics of thinking

2 The Prime Objectives of This Research

The prime objectives of the whole work are proposed as follows:

i. To have mimicking as a prime objective rather than utility-based short-term goals.
AI (or) brain mimicking is the ultimate goals in the field of computer science or information technology. When psychologists and biologists probe the behavior and the structure of the brain to ascertain what is intelligence, computational scientists concurrently endeavor to implement the already found truths about the brain or to implement imaginative perception about the brain with the nomenclature AI. To achieve the universal goal, the approach is to first set short-term goal-based or utility-based objectives. This work is to find other avenues or approaches toward brain mimicking rather than emphasizing only on utility factors. It also becomes necessary to emphasize that the factors that have been established by the computational scientist in their endeavor will be utilized with caution and proper reasoning, keeping in mind that our objective is to deviate from utility-based models toward functional mimicking of the brain.

ii. To have a different functional locus of a brain like 'Thinking' (a process of the brain) as the objective apart from usual learning to move closer toward mimicking.

It is not certain that a person reacts identically toward identical events. Similarly, it is also not certain that two persons react identically to an event. This work presumes that the differentiating factor between any two reactions is a place where the root of intelligence is hiding. The probing method and the avenue of the brain have to be different for finding that root of intelligence.

iii. To have 'emotion' (a temporal status of the brain) as an addition to the objective with the computer-based implementations of cognition and decision making. Psychologists have studied the interplay between emotion, cognition and decision making. The work [4] describes, a happy person decides on the basis of his experience (positive) and in contrast the same person who when sad thinks deep (pessimistic) and decides on the basis of an exhaustive analysis of pros and cons from all the data and knowledge possessed by him. He hints these two approaches, the first one as top-down and the second one as bottom-up. Therefore, the programming paradigms suitable to the aspects defined by psychologists are to be searched for. Applying them with the same decision-making problem of our previous works [5, 6] and achieving feasible results which may be optimum or sometimes sub-optimal according to the emotional constraints will be a different approach. This may move the brain mimicking to the emotion-based temporal behavior of the brains.

iv. To have logical decision making combined with mathematical decision making in the set objectives.
All neural network or support vector machines (SVM) [7, 8] or deep learning (DL) [9, 10] are loaded with mathematical algorithms to map the brain functions producing similar outputs of a brain to any given input using the massive amount of dataset. In contrast here, the objective is set to mingle 'logical decision making' (non-mathematical decision making) like first-order logic, resolution principles, etc.

v. To have implementations as 'multicore centric' to mimic the concurrent cum cooperative functioning of a neural population.
The learning algorithm uses concurrent processing to perform independent mathematical operations concurrently, where this work is to build a model that mimics the cooperative functioning of neurons in spite of being concurrent.

vi. To have uncertainty depicted in the proposed models.
The models developed should exhibit uncertainty within a finite range maybe in speed or may even be in decision making.

vii. To have a modern programming paradigm utilized.

The models developed should imbibe the suitable characteristics of any new programming paradigms. By doing so, the work may have suggestions in modifications in programming language implementations.

It is ascertained that all the above objectives are an endeavor to extend any available work in a new direction. This paper picks the above objectives to an initial extent and analyzes the possibilities of moving closer to brain mimicking.

3 Traditional Versus Modern Artificial Neural Networks

It is important to emphasize that whatever the second-generation artificial neural networks (ANNs) have achieved is through mathematical modeling. Such emphasis will not brush aside the quality of duplicating the brain functions by those neural models. It will only comprehend that such methods have limitations in their aspirations of mimicking a brain. This is also accepted by the nanoscientists who were trying to build a neuro-morphic computer as they describe their effort will even be a baby step toward the objective of mimicking the brain [11]. It is evident that any computer-based model is not going to be completely biological in nature, but they can virtualize the biological elements.

Computational neuroscientists also rightly describe the mathematical models that try to mathematically implement or correlate a brain function. These mathematical models consider the outputs generated by a human brain for the inputs at a particular situation and match a mathematical function to produce similar outputs for the same inputs. These models have achieved only mathematical implementations of a specific brain function under a specific condition but not in the way the brain does it. Therefore, such models are considered as 'weak' models of the brain even though their success according to their utility factor has made us find an alternative word to describe them as 'Engineered AI' [12].

3.1 *Influencing Factors of Third-Generation ANNs*

The analysis of second-generation ANNs created the perception for the formation of the hypothesis of the proposed model. The objective is to try to compute outputs to match the outputs of a human brain, mimicking the brain's functional flow. The emphasis is that the main objective of the proposed ANN is to imitate the human brain, in contrast to the mathematical models that have been achieved to find solutions for specific problem situations. This work scrutinizes the methods used and also unused, in other words, revisit the probabilities of finding alternative models, which may enable better brain mimicking capabilities to the proposed neural networks. It becomes important to contrast from the available definition of the artificial neuron itself. Therefore, a new model has to be proposed for an artificial neuron. This should also eliminate the emphasis only on mathematical calculation.

The search for support toward the set objective confirmed that the objective set is a prospective one and also is supportive of the proposal of providing suitable modifications to the basic computational model itself. Particularly, the work [13] was of immense support to this endeavor of looking back where the authors say looking back should not be considered as a failure of any endeavor; in contrast, it may provide broader future prospects. In the same work [13], the nature of spiking neurons was described to be 'sparsely connected and spiking on discrete timings,' is supportive of the work [6] in which it occupies a part. The third generation spiking neurons are also perceived as a small step toward the biologically plausible model. Further study into spiking neurons confirmed that perception where the work [14] necessitates the inclusion of gene/protein parameters into the functioning of neural models. This work gave an insight into the biologically plausible modifications needed in the computational neural models. In particular, the following two questions raised by the authors [14] gave the idea of the need to modify the basic perception of an artificial neuron.

(a) How to link the gene/protein functionalities to the neuronal behavior? and
(b) Which gene/protein should be included in the model and how to match the timing of gene transcription and translation into proteins according to the spiking speed of the neuron?

It is observed that the tone of the questioning of the above work [14] was the prime factor that enlightened the required perceptional change by the computational model proposers to be more biologically centered. It is also observed that the author's terminology, 'neural parameters,' for gene and protein selection, sounds more design-centered rather than biological perception.

It was decided to think about neurons as a functional unit that gets external inputs in contrast to the traditional way of defining it as a programmatic function with specified parameters. It is perceived that the genetic code determines the functioning of each neuron, rather than a gene pattern being used as data as suggested in the work [14].

4 Need for an Object-Based Model

The following discussion, in specific terms re-discussion, is of high importance to evaluate, elucidate, and emulate the current models of ANNs. This paved the way for finding the scope of new probabilities and designing a new model, respectively. The first task of evaluating the current models gave the conclusion that being mathematical in nature the probability of them becoming a brain mimic is low. That is why it was decided that any proposed model has to be adaptive in nature as a biological neuron to become a nearer mimic of the brain and has to have a lesser emphasis on mathematical calculations. It is observable that the neurons depicted by most of the ANNs in computational neuroscience are numeric data with temporally varying numeric values. A quick review confirmed that it was the case when the structured programming approach was used and a network was considered as a matrix. Surprisingly even the advent of a new programming paradigm called Object-oriented programming has not modified the matrix representation of ANNs to object-based models. The object-oriented paradigm has given the ease in coding reusability but not facilitated any change in neural models. Further probing into available works divulged spiking neurons [15, 16], nano-inductors cum memristors [17, 18], spinnaker, and modern concurrent programming languages. Each one of them has contributed to the object-based model proposed in this paper. Spiking neurons have sustained the idea

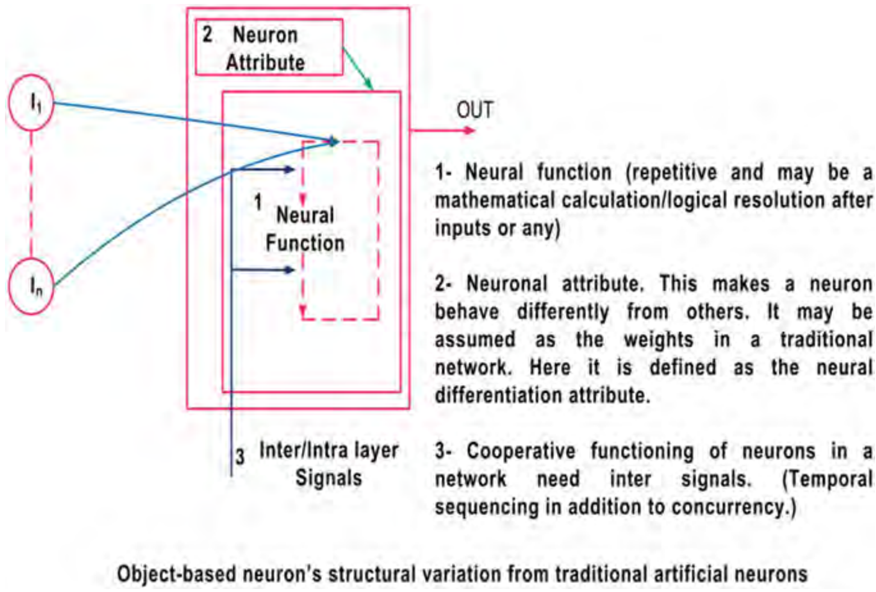


Fig. 2 Object-based neuron's structural variation from traditional artificial neurons

of the need for more biological plausibility and gave the idea of intercommunicating computational objects. Nano-inductors cum memristors [17] supported the need for tiny objects working in synergy. Spinnaker and nano-inductors being two contrasting approaches, where the former one being the interconnection of massively capable millions of processors and the latter wanting to be a massive collection of tiny objects and both being still valued to single-digit percentage in mimicking the brain, opined the search for further ANN models. Both of the above endeavors being designing physical object building approaches directed the search toward paradigm shifts in architecture and programming in computational science. The scope was found to be attached in the area of multicore servers using concurrent languages. Figure 2 illustrates the structural variations proposed for this OBM. The proposed structure is believed to provide diversified functioning of neurons in a network.

4.1 Object-Based Neuron Should Be a Firmware

The probability of achieving a firmware model for a neuron and further into a neural network in mimicking a brain is perceived to be high in comparison to the functional hardware models. The functional hardware models [19] will increase the speed, reduce the size to nano-units, and will confirm a precise inter-communication but will lack in dynamicity in comparison to firmware models. It is trivial for the firmware models to depend on computational hardware available from time to time.

The improved computational multicore environments of the current generation have facilitated this work.

4.2 Object-Based Model Should Be Spiking

Spiking neurons are said [20] to be more biologically plausible models of neural networks. A spike is a discrete event, and such networks concentrate on membrane potential and their response to a quantum of the spike received by it. The huge amount of endeavors in this area of research are making the spiking neurons [21] a wanted area for researchers. The physical model [22, 23] building of a neuron is done through building physical prototypes (nano-inductors, memristors) [24, 25], and virtual modeling is also done through software simulation tools [26]. This work which is purely the software computational model accommodates only the term spike into it. This model does not endeavor toward any physical model. This model perceives an artificial model of a neuron has to be a running process or a routine. A routine that imitates a neuron has to be as tiny as possible and should be alive all the time but should be active only when required. That is, it should be alive but idle until a spike makes it active.

5 Object-Based Model (OBM) Should be Functional in Nature

A neuron is a member of a family of billions of neurons and is individually a functional unit. For simulating a neuron as a computation model, the easiest term that qualifies for such a model is a 'function.' If you see a 'function' as a neuron, immediately the suitability is questioned because all functions are not considered as separate processes from the invoking process. There comes the immediate two alternate terms which can be running units within a computer are 'threads' and 'p-threads.' The p-threads of UNIX and the threads of JAVA which look to be a suitable alternative get denied due to inter-communication and context switching overheads. Their dependency on the operating system (OS) and lack of application-level control are also the reasons for finding suitable alternatives to them. The third alternative programming language term 'object' which evolved has individual properties and common functionalities that may coincide with the definition of the biological neuron. This also gets denied to duplicate a biological neuron because data members of the object are individual, whereas the methods are not like threads, which are alive always. They are called-functions whenever necessary, and even, they are shared by objects belonging to the same type. In other words, for multiple instances of a class, the methods are not loaded independently and all the implementations use shared segment concepts.

After knocking into the closed doors of these structural, object-oriented, and functional programming, the endeavor turned to visit the modern trends in a programming language as hinted earlier. The search toward the types of threads in concurrent processing which are non-preemptive in nature leads to the doors of the 'Go' language.

5.1 Search for a Modern Programming Paradigm

The objectives set for the OBM must pass the feasibility test with regard to contemporary hardware architecture available and also programming paradigms that can suit the model. The availability of multicore architecture immediately satisfies the proposal of multiple neural objects occupying multiple cores concurrently. The vital part of the proposal relies on the search and find conducted for a suitable programming paradigm which will make the model feasible. The criteria set for the desired programming paradigm are as follows,

1. The programming language or environment must have the feature of looking back and changing the basics according to the architectural change.
2. The language should support multiple processes and communicate easily [27].
3. More than inter-process communication the environment should provide control over process execution and hierarchy.
4. The application or network must embed execution control rather than an intermediate environment (a virtual machine) other than the operating system hindering the liberty of the process directly exploiting the services of the operating system.
5. The language should be mechanically sympathetic [27] to the preferred multicore architecture.

The go language is an ace up one's sleeve for this paper. Tables 1 and 2. illustrate the same.

Table 2 illustrates the neural traits and Go language features as the basis for the proposed neural model.

Figure 3. illustrates the neural network functioning of the proposed model in Go language.

6 Conclusion

This paper repurposes available technologies, programming paradigms, and ANN models for the objective of mimicking the brain. The proposal 'object-based model' for a neuron is to give a change in perception of an artificial neuron as a running routine/thread. The paper analyzes the reasoning behind the need for change. The paper also describes the possibilities for such a model with the modern programming

Table 1 Correlation of neural traits and Go language

Neural trait	"Go" language feature	Remarks
A neuron is a functional unit performing a part of a larger function	Goroutine	Goroutine is a lightweight processing unit which can perform a part of a larger application
Neurons are anonymous	Goroutines are anonymous	i. Goroutines need not have a name ii. Functions having names can become Goroutines of multiple instances still will not have any ID iii. The external factors like parameters and messages make similar routines function differently
Neurons cannot be controlled individually [28]	Goroutines are scheduled cooperatively	This work prefers co-operative scheduling followed by Go language before the Go version 1.14. The Goroutines will not be preempted by the Go scheduler when it is running through an OS thread, similar to a neural process that is initiated will not be controllable externally
Neuron's functions are temporal [29]	Go has time-based libraries	Goroutines respond to time library functions like a ticker
Neurons respond to signals and messages	Go has channels, 'sync' and workgroups	Goroutines intercommunications are very handy and have minimal process overheads compared to threads of familiar languages like 'Java' and 'Python'

(continued)

Table 1 (continued)

Neural trait	"Go" language feature	Remarks
Neurons work concurrently in groups	Goroutines suit concurrency	i. Goroutine design is a multicore concurrency centric ii. The performance of Goroutine in a multicore environment makes its choice automatic for process-centric neural implementations iii. The improved concurrency makes it suitable to work in non-deterministic situations

Table 2 Neural character and model structure mapping

Neural traits	Goroutine structure

Neuron perform tasks indefinitely between cell birth and death	Goroutines should be performing its task within an infinite loop
	func neuron() {
	for {
	// Neural Task
	}
	}

Neurons are initiated by the external task	Main invokes many instances of neurons
	func main() {
	go neuron()
	go neuron()
	}

Neurons perform short tasks and wait for other neurons	neuron() blocks for signals within infinite loop
	func neuron() {
	for {
	// channel blocks and Synchronous wait
	}
	}

Neurons when alive can be, active or inactive	Goroutines are invoked and can be in running or blocked state
Neurons wait for signals and messages. They use signals for state change and messages for process control	Go language supports goroutine blocking for channels and sync. broadcast(). Channels can be used for messages and broadcast() be used for signaling. (Here the word state is used for routine states within an application, therefore, phase change is used for making changes within the routine)
Inter Neuronal communication can be one to one or many to one or one to many	i. Neural implementation will use channels for one to one and many to one communications ii. Sync.broadcast() can be used for one to many
The neural performance is massively parallel but controlled parallelism is evident in temporal operations	i. Millions of Goroutines can be invoked and the tool of wait groups in Go language can be used for sequencing neural functions ii. The controlling functions will be done by separate Goroutines

language Go. The challenge to the success of the model relies on the problem selections, concurrency, and non-deterministic approach of the model to reach the goal. The future works can include real-time applications like human behavior mapping to support medical diagnostics or crime detection or guided marketing, etc.

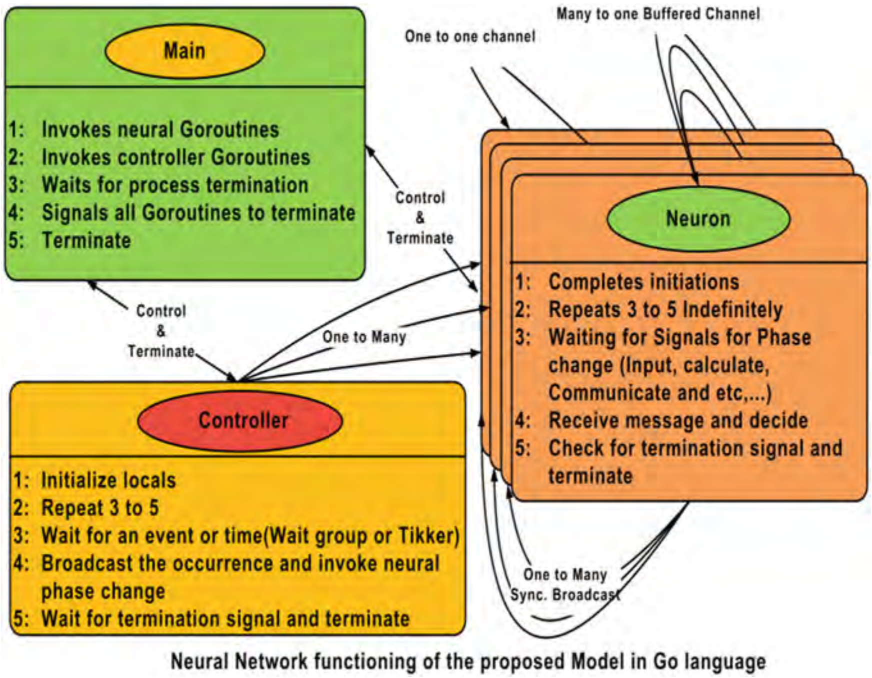


Fig. 3 Neural network functioning of the proposed model in Go language

References

1. *A New Supercomputer Is the World's Fastest Brain-Mimicking Machine*, https://www.scientificamerican.com/article/a-new-supercomputer-is-the-worlds-fastest-brain-mimicking-machine/
2. S. Legg, M. Hutter, A collection of definitions of intelligence. Frontiers Artif. Intell. Appl. **157**, 17 (2007)
3. S. Legg, M. Hutter, *A Formal Measure of Machine Intelligence. arXiv preprint cs/0605024* (2006)
4. N. Schwarz, Emotion, cognition, and decision making. Cogn. Emot. **14**(4), 433–440 (2000)
5. K. Shyamala, P. Chanthini, R. Krishnan, A. Murugan, Artificial neural network model adopting combinatorial inhibition process in multiple solution problems. Int. J. Eng. Technol. **7**(3.4), 167–173 (2018)
6. K. Shyamala, P. Chanthini, R. Krishnan, A. Murugan, Adoption of combinatorial graph for inhibitory process in optimization problems. Int. J. Appl. Eng. Res. **13**(13), 11261–11266 (2018)
7. M. Saber, A. El Rharras, R. Saadane, H.K. Aroussi, M. Wahbi, Artificial neural networks, support vector machine and energy detection for spectrum sensing based on real signals. Int. J. Commun. Netw. Inf. Sec. **11**(1), 52–60 (2019)
8. D. Thukaram, H.P. Khincha, H.P. Vijaynarasimha, Artificial neural network and support vector machine approach for locating faults in radial distribution systems. IEEE Trans. Power Deliv. **20**(2), 710–721 (2005)
9. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*. MIT press (2016)
10. S. Jürgen,Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015)

11. *Researchers Develop Device that Mimics Brain Cells Used for Human Vision*, https://phys.org/news/2020-02-device-mimics-brain-cells-human.html
12. *Beyond Deep Learning—3rd Generation Neural Nets*, https://www.datasciencecentral.com/profiles/blogs/beyond-deep-learning-3rd-generation-neural-nets
13. *Spiking Neural Networks, The Next Generation of Machine Learning*, https://towardsdatascience.com/spiking-neural-networks-the-next-generation-of-machine-learning-84e167f4eb2b
14. N. Kasabov, L. Benuskova, S.G. Wysoski, A computational neurogenetic model of a spiking neuron, in *Proceedings. IEEE International Joint Conference on Neural Networks*, vol. 1 (IEEE, 2005)
15. S. Ghosh-Dastidar, H. Adeli, T*hird Generation Neural Networks: Spiking Neural Networks. Advances in Computational Intelligence.* Springer Berlin, Heidelberg (2009), 167–178
16. P.A. Merolla, J.V. Arthur, R. Alvarez-Icaza, A.S. Cassidy, J. Sawada, F. Akopyan, B. Brezzo, A million spiking-neuron integrated circuit with a scalable communication network and interface. Science **345**(6197), 668–673 (2014)
17. *Scientists Want to Mimic the Human Brain and They've Made a Breakthrough* https://www.weforum.org/agenda/2016/10/scientists-want-to-mimic-the-human-brain-and-they-ve-made-a-breakthrough/
18. J. Grollier, D. Querlioz, M.D. Stiles, pintronic nanodevices for bioinspired computing. Proc. IEEE **104**(10), 2024–2039 (2016)
19. A. Baddeley, Working memory: theories, models, and controversies. Annu. Rev. Psychol. **63**, 1–29 (2012)
20. Y. Hao, X. Huang, M. Dong, B. Xu, A biologically plausible supervised learning method for spiking neural networks using the symmetric STDP rule. Neural Netw **121**, 387–395 (2020)
21. J. Choi, M. Ahn, J.T. Kim, Implementation of hardware model for spiking neural network, in *Proceedings on the International Conference on Artificial Intelligence (ICAI). The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp)* (2015), p. 700
22. D. Sarkar, J. Tao, W. Wang, Q. Lin, M. Yeung, C. Ren, R. Kapadia, Mimicking biological synaptic functionality with an indium phosphide synaptic device on silicon for scalable neuromorphic computing. ACS Nano **12**(2), 1656–1663 (2018)
23. A. Trafton, *Mimicking the Brain in Silicon* https://news.mit.edu/2011/brain-chip-1115
24. V.K. Sangwan, D. Jariwala, I.S. Kim, K.S. Chen, T.J. Marks, L.J. Lauhon, M.C. Hersam, Gate-tunable memristive phenomena mediated by grain boundaries in single-layer MoS2. Nat. Nanotechnol. **10**(5), 403–406 (2015)
25. Y. Babacan, F. Kaçar, K. Gürkan, A spiking and bursting neuron circuit based on memristor. Neurocomputing **203**, 86–91 (2016)
26. R. Brette, M. Rudolph, T. Carnevale, M. Hines, D. Beeman, J.M. Bower, M. Zirpe, Simulation of networks of spiking neurons: a review of tools and strategies. J. Comput. Neurosci. **23**(3), 349–398 (2007)
27. *Scheduling In Go: Part I—OS Scheduler*, https://www.ardanlabs.com/blog/2018/08/scheduling-in-go-part1.html
28. *Brain Basics: The Life and Death of a Neuron*, https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Life-and-Death-Neuron
29. A.M. Rossi, V.M. Fernandes, C. Desplan, Timing temporal transitions during brain development. Curr. Opin. Neurobiol. **42**, 84–92 (2017)

Advancement in Classification of X-Ray Images Using Radial Basis Function with Support of Canny Edge Detection Model

C. M. A. K. Zeelan Basha, T. Sai Teja, T. Ravi Teja, C. Harshita, and M. Rohith Sri Sai

Abstract With the recent technological innovations, medical image processing plays an important role in offering better diagnosis and treatment. In this work, an X-ray image has been used to identify various orthopedic- and radiology-based mussel disorders. For the preprocessing stage, mean, median, wiener filters are used for noise removal, canny edge segmentation is proposed for image acquisition, and radial basic function (RBFNN) machine learning optimization is developed for classification of disorders, respectively. To classify the X-ray image, the above methods are applied to various body parts such as head, neck, skull, palm, and spine. The results compete with present methods and achieve 97.82% accuracy for the classification of X-ray image diseases.

Keywords X-ray image · RBFNN · Mean · Median · Weiner filters · Canny edge segmentation

1 Introduction

X-ray images are utilized to form the imaging of body parts and muscles. They help physicians to identify specific diseases that affect their patients. X-rays are most often used to detect the broken bones as cuts and fractures clearly display in X-ray pictures, enabling doctors to rule out sprains and pains. These images are shaded with black and white because various tissues and bones can absorb different rate of radiations and calcium in hard materials like bones that absorb more. Such that bones shown in X-rays look like the white color. Flesh and fat soft tissues can engage fewer X-rays so it looks like a gray color. This chromatic visualization is shown in Fig. 1 clearly.

C. M. A. K. Zeelan Basha (✉) · T. Ravi Teja · C. Harshita · M. Rohith Sri Sai
Department Of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India

T. Sai Teja
Department Of Electronics and Communication Engineering, Maulana Azad National Institute of Technology, Bhopal, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*, Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_3

Fig. 1 X-ray image

The figure above explained has X-ray model, any X-ray image differentiates as like here, i.e., tissues and flesh absorb photons from X-rays. Bones absorb fewer photons compared to tissues, and this entire interest of imaging is detected by using film or image detectors. The X-rays are two-dimensional raw images, which are projections from radiation phenomena. In this work, the main focus will be given to the automatic detection of brain-related fractures. The fractures and tumors have classified the X-ray images in an effective manner. In general, the X-ray-based disorders can handle the area of interest and structures. An automatic X-ray skull image classification is always remaining as a challenging task. This can be explained as below:

1. Intensity values and the structure of interest.
2. Individual dimensions.

The relationship between X-ray images and the detection of disorders with respect to hard materials has been given a prior task. This work has presented the X-ray images of skull classification by using RBFNN classification and image acquisition with the help of segmentation and filters. A dataset of 500 X-ray images is collected from KIMS hospital, Hyderabad, out of which 300 are normal and 100 are fractured cases.

Organization of the paper: Sect. 1 explains about the general introduction of X-ray images and image acquisition methods, and Sect. 2 describes literature survey related to skull X-ray image; in Sect. 3, the methodology has been explained, and Sects. 4 and 5 briefly explains about results and conclusion of the paper.

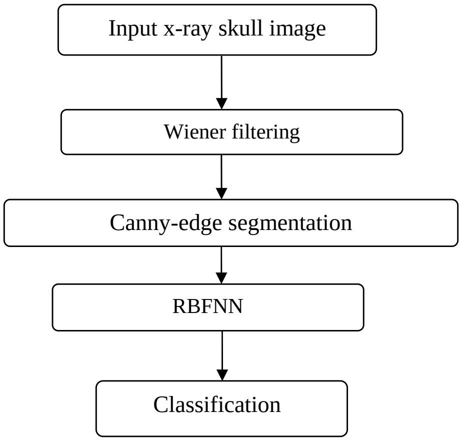
2 Literature Survey

Earlier articles have used CAD structures to encounter ILD in chest radiographs through texture assessment. For instance, the CAD device of the Kun Rossman Laboratory in Chicago has divided the lung into more than one area of the hobby and analyzed the lungs "ROI to decide whether or not or not there were any abnormalities [1]. Then, pretrained NNs have been used to classify suspicious regions to be detected. This gadget can assist medical doctors in beautifying the accuracy of interstitial lesion detection. Plans et al. developed a flexible scheme for CAD of ILD [2]. This technique can discover a selection of pathological competencies of interstitial lung tissue that is totally based on an energetic contour set of rules that would choose the lung vicinity. The area is then divided into 40 special areas of interest [3]. Then, a-dimensional Daubechies wavelet redecorate is completed at the ROI to calculate the texture diploma. However, with the high-quality application of deep reading in the detection of lung ailments and there can be a little literature on the detection of interstitial lung disease within the absence of a massive chest X-ray dataset on ILD [4]. Most of the literature has used CT datasets to locate ILD. Other diseases in chest X-rays are similar to pulmonary nodules, tuberculosis, and ILD [5–7]; there are extraordinary ailments that can be detected, including cardiomegaly, pneumonia, pulmonary edema, and emphysema. There is a good deal less literature on those sicknesses, and a brief communication is given here. Detecting cardiomegaly commonly calls for reading the coronary heart length and calculating the cardiothoracic ratio (CTR) and growing a cardiac tumor screening device. "Candemir et al." used 1D-CTR, 2D-CTR, and CTAR as functions, and they used SVM to categorize 250 cardiomegaly photographs and 250 regular pics, acquiring accuracy of 76.5% [8, 9]. Islam et al. used multiple CNNs to discover cardiomegaly. The community becomes as it should be adjusted on 560 photograph samples and mounted on a hundred snapshots and that they acquired the maximum accuracy of 93%, which is 17% elements higher than within the literature [10–12]. Pneumonia and pulmonary edema may be categorized with the aid of extracting texture abilities. Parveen et al. used an FCM clustering algorithm to find pneumonia. The outcomes confirmed that the lung location of the chest becomes low in black or dark gray. When an affected person has pneumonia, the lungs are full of water or sputum. Thus, there can be more absorbed radiation, and the lung areas will be white or mild gray. This approach can assist medical doctors to come across the degree of infection without troubles and appropriately. "Kumar et al." used a device reading set of regulations to carry out texture analysis of chest X-rays.

3 Methodology

For the preprocessing stage, image acquisition methods like wiener filter are selected. Various filters like mean and median have various difficulties at the time of image

Fig. 2 Block diagram

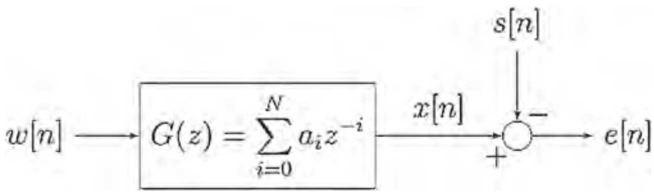


noise reduction. For segmentation, canny edge-based segmentation is used, this method can help to perform a brief study on the selected skull image information. After that radial bases, neural networks have been applied for classification.

Figure 2 explains about proposed X-ray image classification with deep learning mechanism. In this, various steps are used to identify the problems in the skull.

3.1 Wiener Filtering

In X-ray image analysis, wiener filters are very useful for noise eliminations and image de-noising is a significant task in medical image processing. Various filters like median and mean are available, but these have more limitations. So, wiener filter can handle presented limitations,



(1)

The above Eq. (1) explains about wiener noise removal function; this can handle the noise in the X-ray images with efficient manner (Fig. 3).

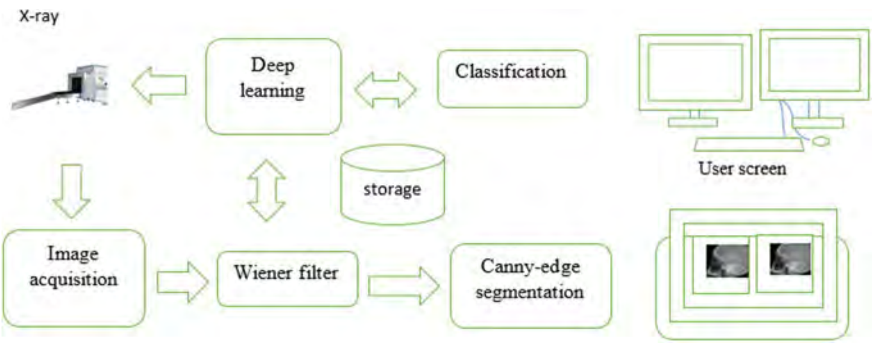


Fig. 3 X-ray image process

3.2 Canny Segmentation

In this segmentation, we are using edge-based detection for multi-stage operations; it is a technique to extract useful structural information from images. This computer vision requires the coefficients like OSTU segmentation method. This canny segmentation can handle the criteria has illustrated below.

1. Edge-based low error rate detection, this catches the image edges.
2. Edge point detection which is the center of localized edges.
3. Image noise has been removed and marked the space where falls edges presented.

Edges are detected clearly by using canny segmentation, due to which the fracture could be easily detected. On the whole, the accuracy in detecting will increase.

$$H_{ij} = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(i-(k+1))^2 + (j-(k+1))^2}{2\sigma^2}\right); \quad 1 \le i, \quad j \le (2k+1) \tag{2}$$

In this, H represents impulse response of the canny edge segmentation method.

$$B = \frac{1}{159}\begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 2 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix} * A.. \tag{3}$$

B is the output of segmentation process, and A is the input.

$$\begin{aligned} d(x, y) &= \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \\ w(x, y) &= \exp\left(-\frac{\sqrt{d(x,y)}}{2h^2}\right) \end{aligned} \tag{4}$$

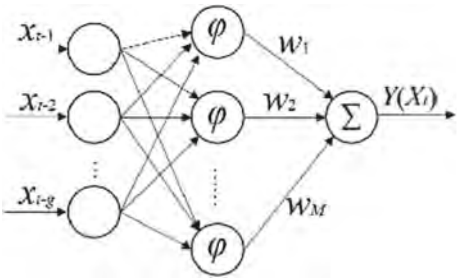


Fig. 4 RBFNN deep learning model

The above Eq. (4) is useful for weight and distance calculation with respect to adaptive ness in the segmentation process.

3.3 RBFNN

Radial basis function is a deep learning process; in this, we classify the selected skull X-ray image. This classification method competes with SVM and KNN models; RBF community learning requires the dedication of the RBF centers and the weights. The selection of the RBF facilities is maximum vital to RBF network implementation. The facilities can be positioned on a random subset of all of the education examples or decided by using clustering or via a getting to know the procedure. One can also use all of the information elements as centers inside the beginning and then selectively dispose of facilities the use of the-NN class scheme [27] (Fig. 4).

X = input factor and C = weighted sum.

For some RBFs which include the Gaussian, it is also crucial to determine the smoothness parameter. Existing RBF community studying algorithms are specially derived for the Gaussian RBF community and can be changed for this reason at the same time as distinctive RBFs are used.

Mathematical computations of RBFNN:

$$\mathbf{P} = \mathbf{P}_j - \frac{\mathbf{P}_j \mathbf{h}_j \mathbf{h}_j^{\mathrm{T}} \boldsymbol{P}_j}{\lambda_j + \mathbf{h}_j^{\mathrm{T}} \mathbf{P}_j \mathbf{h}_j}, \tag{5}$$

where P is the projection matrix which is used to crop the analysis of linear networks that would be useful.

$$\begin{aligned} \mathbf{A}_m^{-1} &= \left(\mathbf{H}_m^{\mathrm{T}} \mathbf{H}_m + \lambda \mathbf{U}_m^{\mathrm{T}} \mathbf{U}_m\right)^{-1} \\ &= \mathbf{U}_m^{-1} \left(\widetilde{\mathbf{H}}_m^{\mathrm{T}} \widetilde{\mathbf{H}}_m + \lambda \mathbf{I}_m\right)^{-1} \left(\mathbf{U}_m^{\mathrm{T}}\right)^{-1} \end{aligned}$$

$$= \mathbf{U}_m^{-1} \begin{bmatrix} \frac{1}{\lambda + \tilde{\mathbf{h}}_1^{\mathrm{T}} \tilde{\mathbf{h}}_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda + \tilde{\mathbf{h}}_2^{\mathrm{T}} \tilde{\mathbf{h}}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda + \tilde{\mathbf{h}}_1^{\mathrm{T}} \tilde{\mathbf{h}}_1} \end{bmatrix} \left(\mathbf{U}_m^{\mathrm{T}}\right)^{-}$$

$$= \mathbf{U}_m^{-1} \widetilde{\mathbf{A}}^{-1} \left(\mathbf{U}_m^{\mathrm{T}}\right)^{-1}. \tag{6}$$

where A^{-1} represents the variance matrix.

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}. \tag{7}$$

$$\mathbf{H} = \begin{bmatrix} H_{11} & H_{12} & \ldots & H_{1m} \\ H_{21} & H_{22} & \ldots & H_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ H_{p1} & H_{p2} & \ldots & H_{\mathrm{pm}} \end{bmatrix}. \tag{8}$$

H represents the design matrix, respectively.

$$\text{Sensitivity} = \frac{\text{tp}}{\text{tp} + \text{fn}} \tag{9}$$

$$\text{Specificity} = \frac{\text{tn}}{\text{tn} + \text{fp}} \tag{10}$$

$$\text{Accuracy} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{fp} + \text{tn} + \text{fn}} \tag{11}$$

Equations (5)–(11) explains about mathematical modeling of RBFNN and performance metrics of proposed X-ray image calculations (Fig. 5).

4 Results and discussion

In this work, skull image has been taken as input applying the image acquisition methods like wiener filter and segmentation for classification we are applying RBFNN results are simulated using MATLAB 2015b.

Tables 1 and 2 explain about performance metrics of proposed method, and these are compared with the existed methods (Figs. 6, 7, and 8).

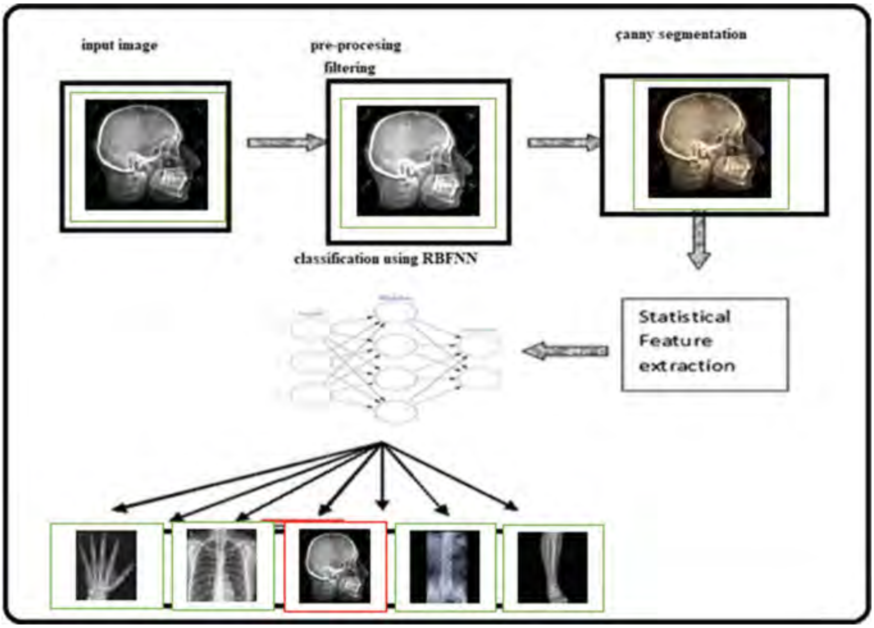


Fig. 5 Architecture of classification of X-rays

Table 1 Confusion metrics parameters

Actual	Predicted	
	Positive	Negative
Positive	T_p = True positive	F_n = False negative
Negative	F_p = False negative	T_n = True positive

Table 2 De-noising PSNR value

Filter	PSNR	References
Wavelet	18.51923	[11]
Median	20.2567	[13]
High boost	23.1678	[12]
Proposed wiener	31.52	Present method

Tables 3, 4, and 5 explain about various methods which are already implemented; these are compared with the proposed RBFNN model (Table 6; Fig. 9).

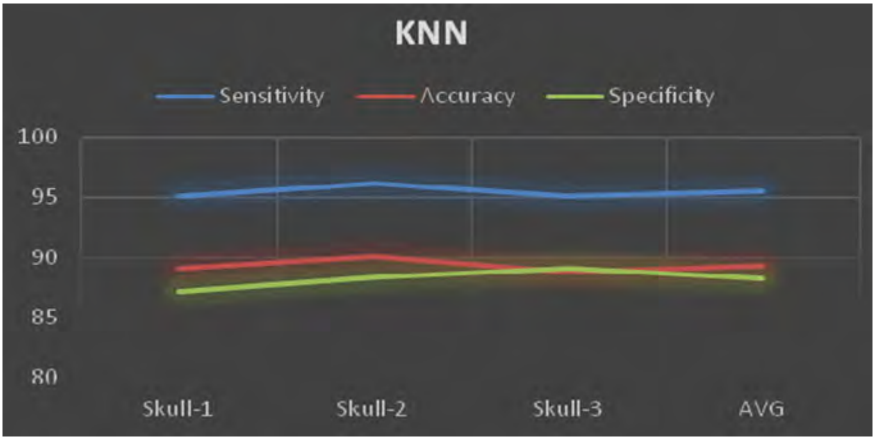


Fig. 6 KNN model

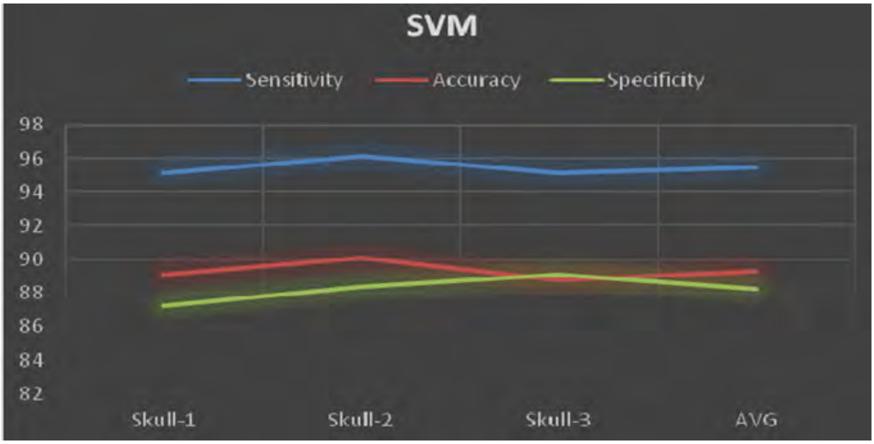


Fig. 7 SVM model

5 Conclusion

In this investigation, a skull X-ray image denoising, segmentation and contrast adjustment have been deployed to improve the quality of X-ray images. After that, statistical analysis is performed based on SVM, KNN, and RBFNN methods. In this, RBFNN model achieves better performance and compete with the present computed-aided design. This work is very helpful for finding the X-ray image diagnosis process in various medical laboratories with orientation. This is helpful for radiologists and

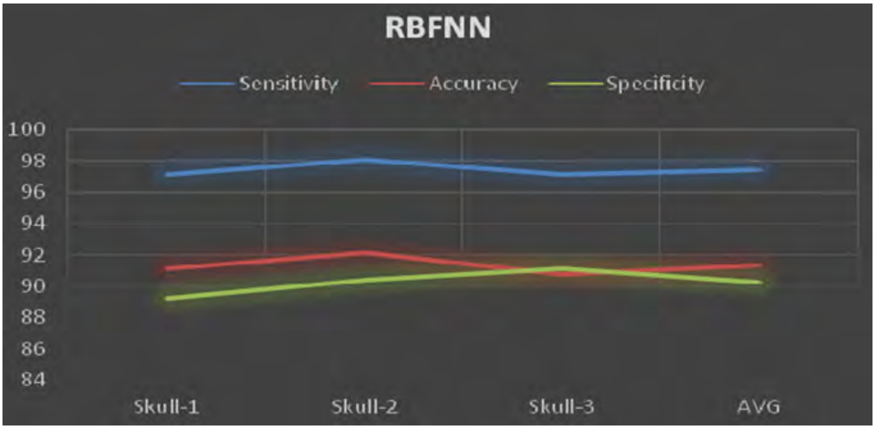


Fig. 8 RBFNN model

Table 3 KNN-classification

Type	Sensitivity	Accuracy	Specificity
Skull-1	95.23	89.12	87.23
Skull-2	96.12	90.12	88.42
Skull-3	95.12	88.78	89.13
AVG	95.49	89.34	88.26

Table 4 SVM classification

Type	Sensitivity	Accuracy	Specificity
Skull-1	96.23	90.12	88.23
Skull-2	97.12	91.12	89.42
Skull-3	96.12	89.78	90.13
AVG	96.49	90.34	89.26

Table 5 RBFNN

Type	Sensitivity	Accuracy	Specificity
Skull-1	97.23	91.12	89.23
Skull-2	98.12	92.12	90.42
Skull-3	97.12	90.78	91.13
AVG	97.49	91.34	90.26

Table 6 Comparison of work

Type	Sensitivity	Accuracy	Specificity
KNN	95.49	89.34	88.26
SVM	96.49	90.34	89.26
RBFNN	97.49	91.34	90.26

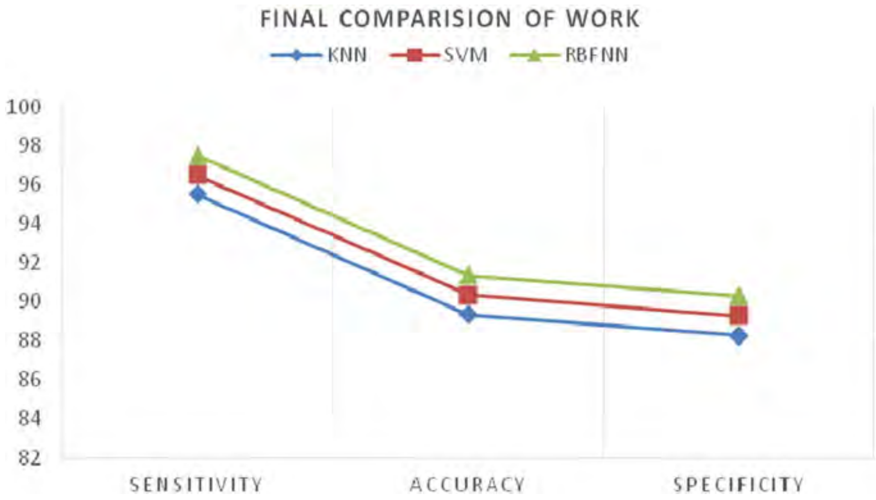


Fig. 9 RBFNN versus remaining models

doctors to perform accurate and real diagnosis. The performance metrics like accuracy has been achieved by 90%, sensitivity 97%, and specificity 91%, respectively, which is a better achievement when compared to the present research work.

References

1. M. Loog, B. van Ginneken, M. Nielsen, Detection of interstitial lung disease in PA chest radiographs, in *Medical Imaging 2004: Physics of Medical Imaging*, vol. 5368 (SPIE, San Diego, 2004). https://doi.org/10.1117/12.535307
2. H. Abe, H. Macmahon, J. Shiraishi, Q. Li, R. Engelmann, K. Doi, Computer-aided diagnosis in chest radiology. Semin. Ultrasound CT MRI **25**(5), 432–437 (2004)
3. M.T. Islam, M.A. Aowal, A.T. Minhaz, K. Ashraf, *Abnormality Detection and Localization in Chest X-rays Using Deep Convolutional Neural Net*works. arXiv preprint arXiv:170509850 (2017)
4. A. Kumar, W. Yen-Yu, L. Kai-Che, I.C. Tsai, H. Ching-Chun, H. Nguyen, Distinguishing normal and pulmonary edema chest x-ray using Gabor flter and SVM, in 2014 IEEE International Symposium on Bioelectronics and Bioinformatics, ed by C. Li (IEEE ISBB, Taiwan, 2014), 1–4. https://doi.org/10.1109/isbb.2014.6820918
5. U. Avni, H. Greenspan, E. Konen, M. Sharon, J. Goldberger, X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words. IEEE Trans. Med. Imaging. **30**(3), 733–746 (2011)
6. N.M. Noor, O.M. Rijal, A. Yunus, A.A. Mahayiddin, C.P. Gan, E.L. Ong, et al., Texture-based statistical detection and discrimination of some respiratory diseases using chest radiograph, in *Advances in Medical Diagnostic Technology* (Springer, Singapore, 2014). pp. 75–9
7. V. Bindhu, BIomedical image analysis using semantic segmentation. J. Innov. Image Process. (JIIP) **1**(02), 91–101 (2019)
8. C.Z. Basha, K.M. Sricharan, C.K. Dheeraj, R. Ramya Sri, A study on wavelet transform using image analysis. Int. J. Eng. Technol. (UAE), **7**(2), 94–96 (2018)

9. C.M.A.K. Zeelan Basha, T. Maruthi Padmaja, G.N. Balaji, Automatic X-ray image classification system. in *Smart Innovation, Systems and Technologies*, vol. 78 (Springer Science and Business Media, Deutschland GmbH), pp. 43–52
10. C.M A.K. Zeelan Basha, Maruthi Padmaja, and G.N.Balaji,"Computer Aided Fracture Detection System" Journal of Medical Imaging and Health Informatics Vol. 8, 526–531, 2018.
11. C.M.A.K. Zeelan Basha, T. Maruthi Padmaja, G.N. Balaji, EAI endorsed transactions on pervasive health and technology. **5**(18), 1–6 (2019)
12. C.Z. Basha, M.R.K. Reddy, K.H.S. Nikhil, P.S.M. Venkatesh, A.V. Asish, Enhanced computer aided bone fracture detection employing x-ray images by Harris Corner technique, in *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)* (Erode, India, 2020), pp. 991–995.

Brain Tumour Three-Class Classification on MRI Scans Using Transfer Learning and Data Augmentation

C. A. Ancy and Maya L. Pai

Abstract Accurate classification is a prerequisite for brain tumour diagnosis. The proposed method is a modified computer-aided detection (CAD) technique used for leveraging automatic classification in brain magnetic resonance imaging (MRI), where the model has trained a pipeline of convolutional neural networks (CNNs) using transfer learning (TL) on ResNet 50 with PyTorch. The proposed method employs benchmarked datasets from figshare database, where data augmentation (DA) is applied to increase the number of datasets that can further increase the training efficiency. Thus, the retrained model can classify the tumour images into three classes, i.e., glioma, meningioma, and pituitary tumours. Classification accuracy was tested by comparing the accuracy matrices, loss matrices, and confusion matrix and found to be 99%. The proposed model is the first of its kind that employs both DA and TL on the ResNet 50 model for performing a three-class classification on brain tumour, and results reveal that it outperforms all other existing methods.

Keywords Computer-aided detection · Magnetic resonance ımaging · Convolutional neural network · Transfer learning · ResNet 50 · Data augmentation

1 Introduction

The use of machine learning techniques for classification has changed the facets of CAD systems [1–3]. Brain tumour has become one of the perilous diseases across the globe. Generally, tumour is caused due to the anomalous development of cells all over the skull. Computer vision and deep learning techniques will widely employ the trained CNN models to facilitate diagnosis and treatment triage [4]. Many different imaging techniques exist, and this will give information about size, shape, location, and type of brain tumours such as magnetic resonance imaging (MRI), magnetic resonance spectroscopy (MRS), computed tomography (CT), single-photon emission

C. A. Ancy (✉) · M. L. Pai
Department of Computer Science and IT, School of Arts and Sciences, Kochi, India

Amrita Vishwa Vidyapeetham, Coimbatore, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*, Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_4

computed tomography (SPECT), and positron emission tomography (PET). Figure 1 depicts the different brain imaging techniques used by the radiologists. Here, many factors such as angle, lighting, and resolution can disrupt the classification results. Among different imaging techniques available, MRI images are widely employed due to their strong resolution to soft tissues and they can also give detailed description on the brain images. MRI scans are of different types, they are FLAIR, T1-weighted, T2-weighted, etc. The proposed work uses T1-weighted contrast-enhanced magnetic resonance images (CE-MRI) for the work. T1-weighted images use all three image planes, i.e., axial, sagittal, and coronal views. It uses the gadolinium-chelate injection method to highlight the contrast in the images. Figure 2 shows the images of different types of MRI scans.

Among all imaging techniques available, early detection plays a very important role in the diagnosis and treatment outcomes of the disease. According to the World Health Organization (WHO) standards, neurologists have classified brain tumours into more than 120 different types [5]. The proposed method deals with multiclass

Fig. 1 Different brain imaging techniques

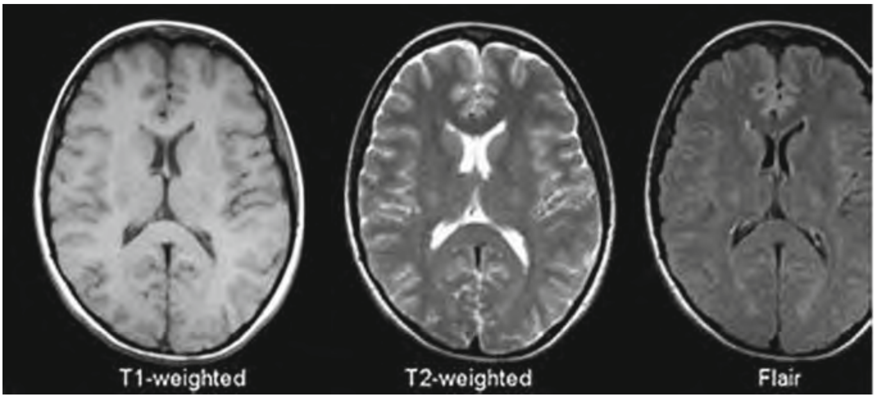


Fig. 2 Different types of MRI scans

classification, where the data is taken from publicly available dataset figshare. Among the existing 120 different types of tumours, the dataset contains 2D images of three kinds of brain tumours: glioma, meningioma, and pituitary tumours, i.e., three-class classification.

CAD method was first evolved using different traditional machine learning algorithms, which takes a long time since it involves different steps like image preprocessing, segmentation, feature extraction, and classification. Despite these drawbacks, many such CAD models outperformed the scenario [6]. There accuracy depends on the handcrafted features which they obtained and efficient modelling of machine learning algorithms employed. It was after 2014 that much of the works in brain tumour detection was done using deep learning (DL) techniques. DL methods do not require handcrafted feature extraction, whereas they have the capability to self-learn from the datasets.

Among different methods in DL, CNN models are widely employed and have achieved good results nowadays due to several reasons: (1) they are so powerful for complex image segmentation and classification problems; (2) availability of a large number of labelled training data in medical imaging; (3) availability of different transfer learning techniques to get knowledge; and (4) availability of powerful graphics processing units (GPU). The proposed work also addresses hardware restrictions as medical image datasets would be heavy, and it requires better hardware specifications for storing, training, and deploying such heavy datasets to ensure maximum accuracy.

Figure 3 shows a midline post-contrast sagittal T1-weighted MRI image, where different parts of the brain are marked. The types of tumours are classified based on the nature of the growth, type of spreading, and also the location in which the growth appears.

Figure 4 depicts the three different views (sagittal, axial, and coronal) in which MRI scanning is performed to obtain T1-weighted images. And Fig. 5 shows the three prominent brain tumour types, which are taken from the figshare database and considered for the three-class classification in the present work. The tumour portions are marked with red shading.

The structure of the paper is as follows. state of the art is described in Sect. 2, and the architecture of proposed method and experimental setup is discussed in Sect. 3. Section 4 gives the test results and comparisons with brief discussions. Section 5 gives the conclusion, and Sect. 6 discusses the outlook for future scope.

2 State Of The Art

The research work using CNN particularly in the field of brain tumour detection has increased since 2014. According to the statistics [7], the number of works published per year as in Google Scholar containing the keywords CNN and the brain tumour was 19 in the year 2014, 47 in 2015, 137 in 2016, 468 in 2017, 977 in 2018, increased to 1790 in 2019 and 351 in 2020. From all these studies, it was observed that the

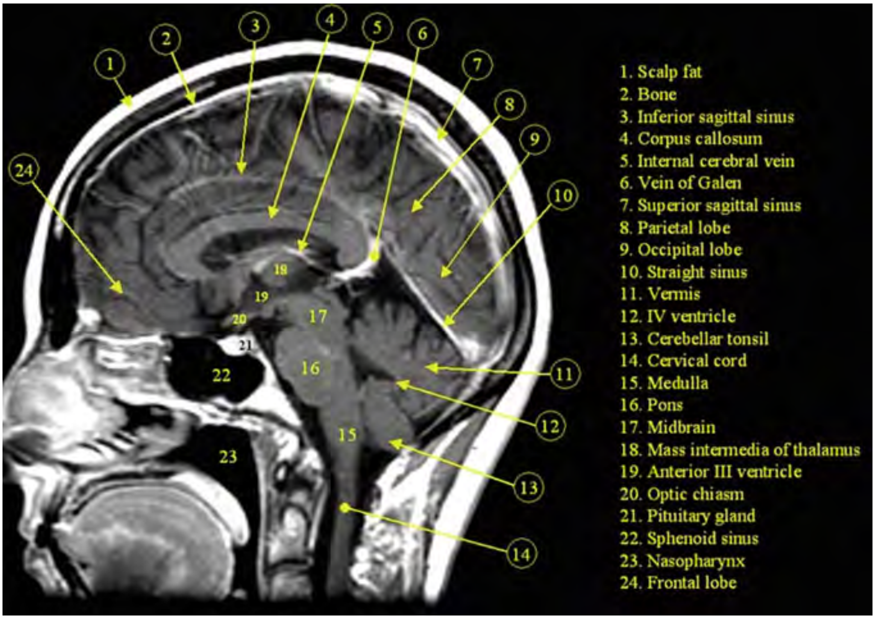


Fig. 3 Parts of brain

Fig. 4 Sagittal, axial, and coronal views of MRI scans

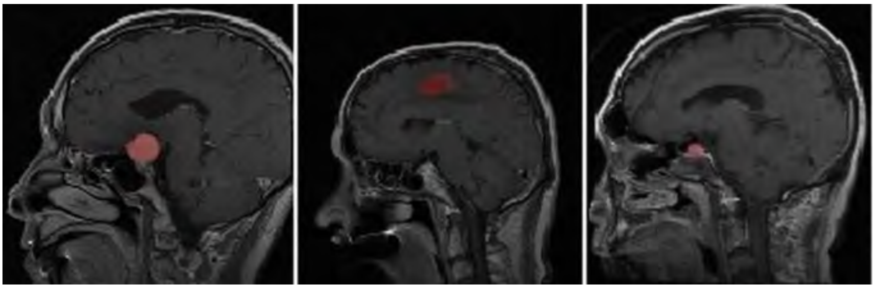


Fig. 5 MRI samples of meningioma, glioma, and pituitary tumour types

availability of brain tumour images was a major problem in many of the works. Due to this problem, overfitting can occur to CNN architectures. In 2018, Mohsan et al. [8] proposed a method that combines discrete wavelet transform (DWT) and principal component analysis (PCA) and got good accuracy for four class brain tumour classification. Another work which came in 2018 was by Naceur et al. [9], where he used fully automatic brain tumour segmentation using end-to-end incremental deep neural networks in MRI images. In 2019, Khaled [10] came up with a review article where he describes different works that happened in the brain tumour detection domain using different CNN models. Another method came up in 2019, by Deepak et al. [11], where they addressed the three-class brain tumour classification problem using CNN features via transfer learning. They employed deep transfer learning and uses a pretrained GoogleNet to extract features from brain MRI images and obtained a classification accuracy of 98%. Another work in 2019, by Vijayakumar [12], proposed a model of capsule neural network that can work well even with a small number of datasets, unlike the CNN model. In 2019, Hossam et.al [13] came up with a multiclass classifier where the model used two publicly available datasets and obtained appreciable overall accuracy. In 2020, Maglogiannis et al. [7] came up with a work that uses MRI images and the transfer learning for brain tumour classification. Their CAD systems employ transfer learning for feature extraction and use it in nine deep pretrained convolutional neural network (CNN) architectures.

The proposed method employs the data augmentation technique to avoid insufficient data problems and overfitting. Using pretrained CNN models along with transfer learning and adjusting the learning and hyperparameters can achieve good results, and it works much faster and simpler than models working with randomly initialized weights [14, 15]. The proposed method uses three-class tumour classification in brain MRI by data augmentation and transfer learning using the ResNet 50 model, which is one of the outperformed models.

3 Architecture of Proposed Method and Experimental Setup

3.1 Proposed Method

In the proposed method, we employ the pretrained architecture of ResNet 50, and the extracted features are used by deep transfer learning to produce knowledge. We use CE-MRI images from the figshare database for classification (glioma, meningioma, and pituitary). To avoid the problem of the small number of training datasets and the problem of overfitting, our classification systems use data augmentation techniques to improve the number of datasets and deep transfer learning for feature extraction. The proposed method is the first of its kind which employs both DA and TL for the three-class classification of brain MRI. ResNet is one of the outperformed pretrained models which is widely employed in this area because of its simpler structure and less complexity. ResNet employs residual learning methods for better training of networks and thus reduces the error caused by increasing depth. It has many models available, where we employ the ResNet 50 model as it can give better results. Thus, after fine-tuning and learning, the retrained model could accurately classify the tumour into three types. Classification accuracy and various performance metrics are computed. Good results are obtained in case of accuracy, small training time, and less computation cost compared to the related work. Figure 6 shows the generic model of our proposed system, and Fig. 7 depicts the architecture of the pretrained model of ResNet 50.

In the retrained model, instead of training the model from the scratch, the learned features are transferred from the pretrained model. The pretrained model employed in the work is ResNet 50, which is already a built model for training and contains many fully connected layers for convolution and pooling, and is used as the starting point. It learns and passes the knowledge to the retrained model. By using the pretrained model, we can directly use the weights and architecture obtained, as it is previously learned on large datasets, and apply directly to our problem using transfer learning. ResNet 50 pretrained model is directly available in the Keras library and is trained using the ImageNet dataset. ResNet 50 is a residual network variant and uses 50 layers. The retrained model is created by fine tuning ResNet 50 which is already been trained. The architecture of the pretrained model can be employed, and weights can be initialized randomly and train the model according to our dataset again. Images acquired from the benchmarked publicly available figshare database undergo

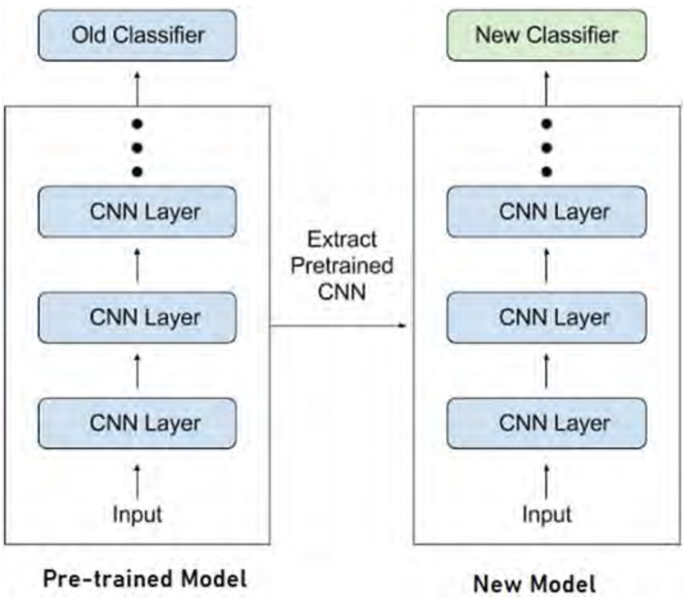


Fig. 6 Generic architecture of the new model

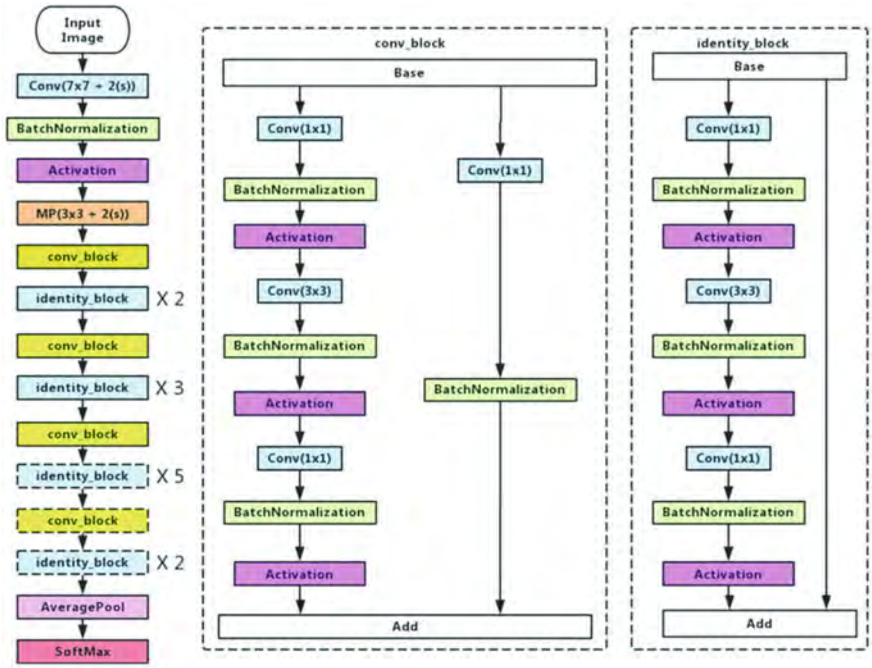


Fig. 7 Architecture of ResNet 50

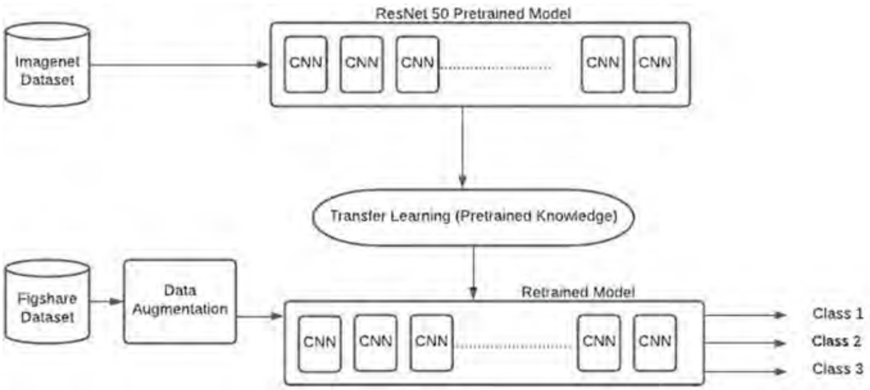


Fig. 8 Proposed model

preprocessing to remove noise and data augmentation to multiply in number. Thus, the retrained model uses the modified data for training and the knowledge from transfer learning. Retrained models are used to speed up learning and to avoid the need of training large datasets again. While using the pretrained model, it is important to retrain the upper layers as the higher-level features will be class-specific.

Figure 8 shows the diagram of the proposed model of our work. The objective of the proposed model is to bring up a fully automated CAD which can increase the overall accuracy, reduce the overfitting, and speed up the training time.

3.2 Experimental Setup

Running and training the DL models with large and real datasets has been the major hindrance for researchers, as it requires a huge computational power. We implemented the proposed classification model using python and GPU-computing libraries (CUDA, OpenCL) in Google Colaboratory (Colab) which is a free research project of Google. It is a cloud-based Jupyter notebook which gives the users a fully fledged runtime for DNN architectures and hence requires no setup and also contains many ML libraries preinstalled. We can also add our libraries and can store the code in Google drive so that it can be worked from anywhere. It poses a robust GPU with free-of-charge, hence widely employed in CNN training as it is more secure and durable.

3.3 Dataset Used and Preprocessing Employed

The datasets from the public database figshare [16] are used in the proposed work. It contains a 3064 T1-weighted CE-MRI modality for the first gate check, which is collected from 233 patients of three different hospitals in China during 2005–2010. The dataset contains 2D images of three kinds of tumour images: meningioma (708 images), glioma (1426 images), and pituitary tumour (930 images), which are in (.mat) data format and are of three different views: axial, sagittal, and coronal. Before feeding the data to augmentation, RoI-based preprocessing step for data cleaning is performed, to remove the small amount of Salt and Pepper and Gaussian noise which are present in the MRI scans. The data size of images is 512×512 pixels , whereas each pixel size is 49 mm $\times$ 49 mm. Grey images are first normalized and converted into RGB format, which is represented by $m \times n \times 3$ array with values ranging from [0, 1]. It is then resized to $224 \times 224 \times 3$ (For ResNet 50 model). After the reduction, the dataset is divided into 70% (2145 images) training and the remaining 30% (919 images) for validation using the Pareto principle [17]. The splitting of datasets has also been done using the other popular ratios like 80–20, 75–25, still, 70–30 gave the highest overall accuracy.

3.4 Data Augmentation

Data augmentation is a method that is applied on-the-fly during training to avoid the lack of data problem which causes overfitting and increases generalization performance by enhancing the training dataset itself. It also solves the problem of class imbalance, through oversampling the minority class. In the case of data insufficiency, the model can memorize the details of the training set, but can't generalize it to the validation set. Deep learning works better with optimum datasets. To obtain the desired accuracy, we need to address the problem of lack of satisfactory amounts of data. We applied eight different augmentation techniques to extend the available data. This session briefs about the transformations applied in the data augmentation. They are edge detection, sharpening, gaussian blur, and emboss for the noise invariance and skewness, flipping, rotation, and shears for geometric transformations invariance. The original dataset contains 3064 MRI images. After augmentation, it got increased by a factor of 8, thus the modified datasets become large. It was observed that augmentation techniques increased the individual accuracy of data samples from three classes and also the overall accuracy of the system. Accuracy of class glioma, meningioma, pituitary, and overall accuracy of the model before augmentation was 95%, 96%, 95%, and 95%, respectively, while after augmentation was 99.46%, 98.10%, 99.84%, and 99%, respectively. Thus by applying these techniques, considerable amount of high-quality abundant data was obtained in the training phase.

3.5 Deep Transfer Learning

Transfer learning is an optimization technique in machine learning which is primarily employed for improving performance and speed up the training process. We use a pretrained model approach for our work, which is the most commonly employed model in DL problems. It uses a model that is trained for a particular task and repurposes it to use for another related task as the inductive transfer. In the proposed work, the pretrained ResNet 50 model was retrained using the transfer learning technique where the data augmented training set from the figshare database was used. We use stochastic gradient descent (SGD) with 0.9 momenta, and it gives a decreasing learning rate for training and thus we get closer to the desired parameters. The learning rate was chosen to be 10^{-4} for the work. The value of the learning rate is suitably chosen as, if it is too small, optimization takes a lot of time whereas, if it is too large, the optimization may overshoot. Minibatch size is set to 128 images as a very large batch size can adversely affect the model quality. Finally, fine-tuning is done to lightly adjust the weights of the model. To perform the transfer learning, we train for 50 epochs and the networks are validated at appropriate iterations.

4 Test Results and Comparisons with Brief Discussions

In this section, evaluation criteria and performance matrices computed are described and the results of the proposed method are presented. The proposed classifier classifies the output into three classes: glioma, meningioma, and pituitary tumours. Figures 9, 10, and 11 show the MRI sample images of the final outputs obtained from the three-class classifier. The model could classify with a minimum number of incorrect predictions.

4.1 Performance Matrices and Evaluation

The performance of the given model was computed using various evaluation indices such as accuracy matrices, loss matrices, and confusion matrix. Training and validation accuracy and loss of the proposed model with 70% of training samples are given

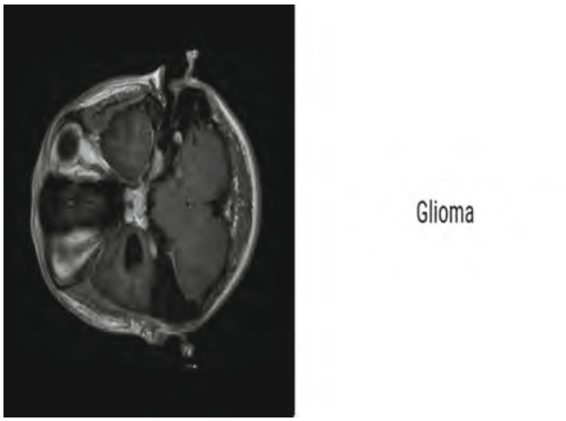


Fig. 9 Image classified as glioma

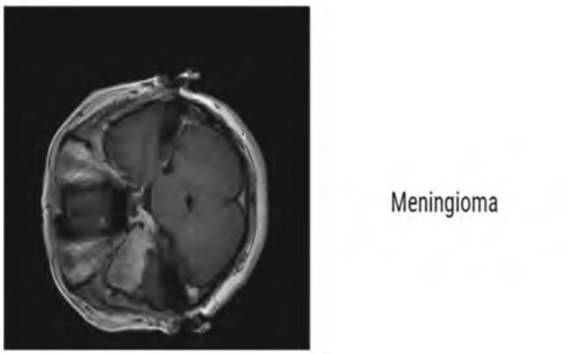


Fig. 10 Image classified as meningioma

Fig. 11 Image classified as pituitary

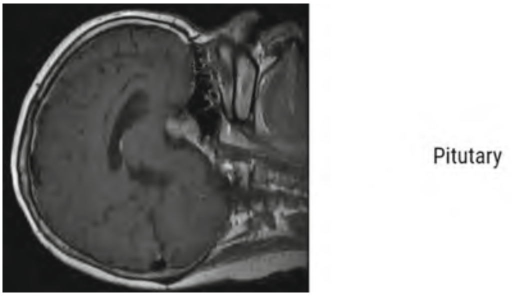
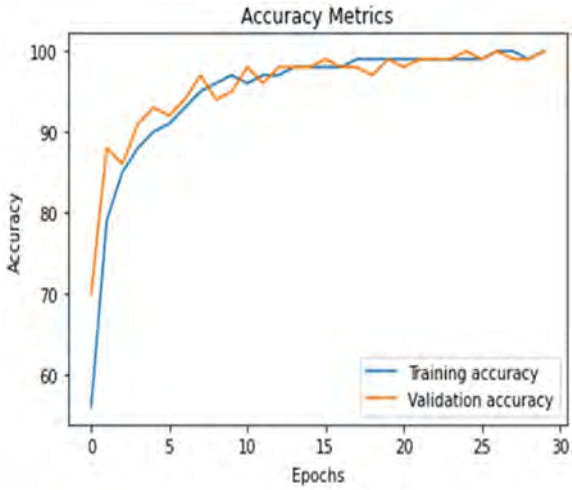


Fig. 12 Training and validation accuracy



below. Classification accuracy is defined as the ratio of the number of correctly classified samples with the total number of data samples. Accuracy matrices for the training and validation are depicted in Fig. 12. It gives the overall accuracy of the proposed model. It shows how the proposed model's prediction was correct compared to the true data. Good value in accuracy metrics reveals that the model is classified well. With the obtained results, it was revealed that the overall accuracy of the proposed model was 99%. Figure 13 shows the lose metrics obtained. Loss is measured not as a percentage, but it is a number calculated on training and validation dataset which gives how bad the prediction was on a single sample. It is the sum of errors of each sample in the validation and training datasets. A decline in the loss metrics reveals that the model's prediction was good which in turn proves that the proposed model was accurate.

Figure 14 shows the confusion matrix obtained by using the Scikit-learn library where the *x*-axis denotes the true label and the *y*-axis denotes the predicted label. Table 1 shows the obtained values of the confusion matrix. It is used for quantifying the performance of the classifier. This table gives a summary of the prediction results

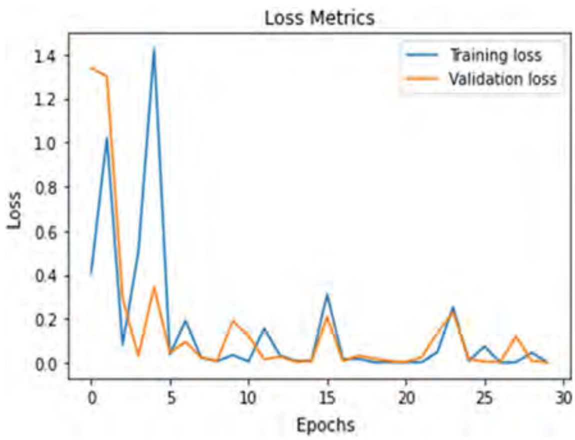


Fig. 13 Training and validation loss

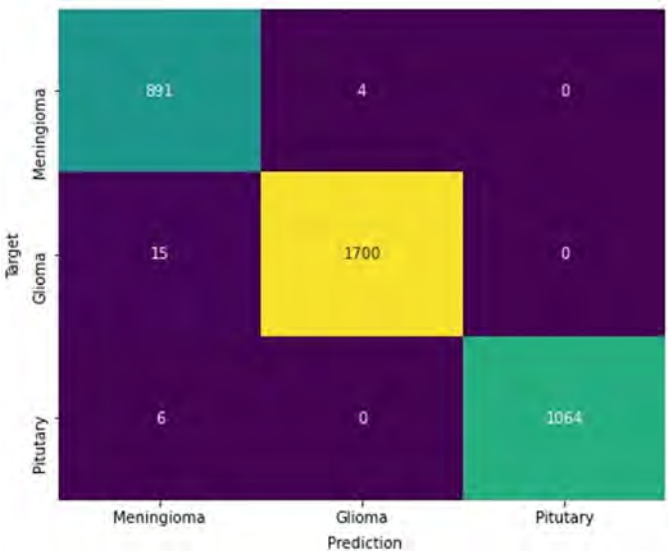


Fig. 14 Confusion matrix

Table 1 Confusion matrix for the proposed work obtained after training with 70% of train data

Predicted				
		M	G	P
Actual	M	891	4	0
	G	15	1700	0
	P	6	0	1064

P, M, G denotes Pituitary, Meningioma, and Glioma images respectively of the three-class classification problem. Here the diagonal values give the correct predictions, i.e., 891, 1700, 1064 for classes meningioma, glioma, and pituitary. From these values, accuracy, precision, recall, specificity, and F1 score can be computed using the true-positive, true-negative, false-positive, and false-negative cases.

4.2 Test Results

Accuracy, precision, recall, specificity, and F1 score of individual samples from these three classes were also found out to know the individual class performance. Those can be computed from the values obtained from the confusion matrix. Table 2 gives the accuracy, precision, recall, specificity, and F1 score values of the individual class predictions of class glioma (G), meningioma (M), and pituitary (P), computed using the formula (1) to (5). Table 2 shows the performance matrices values like accuracy, precision, recall, specificity, and F1 score obtained for the individual classes. From the results of Table 2, it was revealed that individual classification efficiency was pretty good for class glioma and pituitary. For class meningioma, it was slightly week.

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{FN} + \text{TN}) \tag{1}$$

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{2}$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{3}$$

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FP}) \tag{4}$$

$$\text{F1 Score} = 2\text{TP}/(\text{TP} + \text{FP}) \tag{5}$$

Table 2 Performance matrix values for three-class classifier outputs

	Accuracy	Precision	Recall	Specificity	F1 score
G	99.46	99.13	99.77	99.24	0.99
M	98.10	99.55	97.7	99.24	0.98
P	99.84	99.43	1	1	0.99

P, M, G denotes Pituitary, Meningioma, and Glioma images respectively

Table 3 Comparison of the overall accuracy of the model with existing works

Author(s)	Methodology	Training data (%)	Accuracy (%)
Pashaei [18]	CNN ELM	70	93.68
Afshar [19]	CapsNet	70	90.89
Deepak and Ameer[11]	CNN & transfer learning	5	97.1
Swati et al. [20]	Transfer learning and fine tuning	75	94.82
Proposed method	Data augmentation and transfer learning using ResNet 50	70	99

4.3 Comparison with Existing Works

The overall accuracy of the model is compared with other existing works too. Table 3 shows the comparison value of the overall accuracy of the proposed method with other existing related works. The proposed method took 50 epochs to train, 70% training data to obtain an overall accuracy of 99%. The table displays the methodology employed in each state of the art, along with the percentage of data used for training and the accuracy obtained. Results revealed that the proposed method outperformed all other existing methods.

5 Conclusion

An accurate and fully automated three-class classifier is proposed for classifying brain tumours, which takes CE-MRI benchmarked datasets from the figshare database. The problem with insufficient data samples and overfitting is addressed by data augmentation, which significantly improved the image samples for training and did the necessary preprocessing. The proposed system was pretrained with ResNet 50 architecture and transfer learning for hyperparameter tuning which eases the classification with simpler architecture and less time taken. The proposed method employs the combination of both DA and TL from a pretrained network, which is the first of its kind and test results and comparisons revealed that the proposed model outperformed all the state of the art. The work outperformed in the case of simpler architecture employed, a smaller number of epochs taken for training, reduced time consumption, lower number of outfits, and with a better classification of 99%. The limitations of the work include lower classification accuracy for meningioma tumours.

6 Outlook for Future Scope

The work can be extended to perform further. First, the classification result of meningioma was a little week, and appropriate tuning of the transfer learning model may solve this issue. Secondly, different image fusion techniques can also be tried in the preprocessing phase to improve upon the accuracy of deep neural networks. Third, extending the same model with other existing models such as GoogleNet, AlexNet, SENet, VGG models, and their performances can be compared. Fourth, images from different modalities can be tried with models such as X-rays, PET, and CT types. Fifth, the number of epochs taken and its effect on the classification results can also be studied further.

References

1. M.S. Suchithra, M.L. Pai, Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters. Inform. Process. Agricul. (2019). https://doi.org/10.1016/j.inpa.2019.05.003
2. P. Aswathi Anand, M.L. Pai, Artificial neural network model for identifying early readmission of diabetic patients. Int. J. Innov. Technol. Explor. Eng. (IJITEE). **8**(6) (2019)
3. K.S. Varsha, M.L. Pai, Rainfall prediction using fuzzy c-mean clustering and fuzzy rule-based classification. Int. J. Pure. Appl. Mathe. **119**, 597–605 (2018)
4. S. Shakya, Analysis of artificial intelligence based image classification techniques. J. Innov. Image Process. (JIIP) **2**(01), 44–54 (2020)
5. N.J. Tustison, K.L. Shrinidhi, M. Wintermark, C.R. Durst, B.M. Kandel, J.C. Gee, M.C. Grossman, B.B. Avants, Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with ANTsR. Neuroinformatics **13**(2), 209–225 (2014). https://doi.org/10.1007/s12021-014-9245-2
6. E.I. Zacharaki, S. Wang, S. Chawla, D.S. Yoo, R. Wolf, E.R. Melhem, C. Davatzikos, Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. Magn. Reson. Med. **62**(6), 1609–1618 (2009). https://doi.org/10.1002/mrm.22147
7. R. Chelghoum, A. Ikhlef, A. Hameurlaine, S. Jacquir, Transfer learning using convolutional neural network architectures for brain tumor classification from MRI images. IFIP Adv. Inform. Commun. Technol. 189–200 (2020). https://doi.org/10.1007/978-3-030-49161-1_17
8. H. Mohsen, E.-S. El-Dahshan, E.-S. El-Horbaty, A.-B. Salem, Classification using deep learning neural networks for brain tumors. Fut. Comput. Inf. J. **3**(1), 68–71 (2018). https://doi.org/10.1016/j.fcij.2017.12.001
9. M.R. Naceur, S. Rachida, A. Akil, K. Rostom, Fully automatic brain tumor segmentation using end-to-end incremental deep neural networks in MRI images. Comput. Methods Programs Biomed. **166**, 39–49 (2018). https://doi.org/10.1016/j.cmpb.2018.09.007
10. M.K. Abd-Ellah, A.I. Awad, A.A.M. Khalaf, H.F.A. Hamed, A review on brain tumor diagnosis from MRI images: practical implications, key achievements, and lessons learned. Magn. Reson. Imaging **61**, 300–318 (2019). https://doi.org/10.1016/j.mri.2019.05.028
11. S. Deepak, P.M. Ameer, Brain tumor classification using deep CNN features via transfer learning. Comput. Biol. Med. **111**, 103–345 (2019). https://doi.org/10.1016/j.compbiomed.2019.103345
12. T. Vijaykumar, Classification of brain cancer type using machine learning. J. Artif. Intell. Caps. Netw. **2**, 105–113 (2019)

13. H.H. Sultan, N.M. Salem, W. Al-Atabany, Multi-classification of brain tumor images using deep neural network. IEEE Access **7**, 69215–69225 (2019). https://doi.org/10.1109/access.2019.2919122
14. M.I. Sharif, J.P. Li, M.A. Khan, M.A. Saleem, Active deep neural network features selection for segmentation and recognition of brain tumors using MRI images. Pattern Recogn. Lett. **129**, 181–189 (2020). https://doi.org/10.1016/j.patrec.2019.11.019
15. J. Bernal, K. Kushibar, D.S. Asfaw, S. Valverde, A. Oliver, R. Martí, X. Lladó, Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. Artif. Intell. Med. **95**(April), 64–81 (2019). https://doi.org/10.1016/j.artmed.2018.08.008
16. Figshare brain tumor dataset, https://doi.org/10.6084/m9.figshare.1512427.v5
17. A. Sarah, I. Abdelaziz, M. Ammar, H. Hesham, An enhanced deep learning approach for brain cancer MRI images classification using residual networks. Artif. Intell. Med. **102**, 101779 (2020). https://doi.org/10.1016/j.artmed.2019.101779
18. A. Pashaei, H. Sajedi, N. Jazayeri, Brain tumor classification via convolutional neural network and extreme learning machines, in IEEE 8th International Conference on Computer and Knowledge Engineering (ICCKE, 2018), pp. 314–319.
19. P. Afshar, K.N. Plataniotis, A. Mohammadi, Capsule networks for brain tumor classification based on MRI images and course tumor boundaries, in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP, 2019), pp. 1368–1372.
20. Z.N. Swati, Q.Z. Khan, M. Kabir, F. Ali, Z. Ali, S. Ahmed, Lu. Jianfeng, Brain tumor classification for MR images using transfer learning and fine-tuning. Comput. Med. Imaging Graph. **75**(July), 34–46 (2019). https://doi.org/10.1016/j.compmedimag.2019.05.001

Assessing the Statistical Significance of Pairwise Gapped Global Sequence Alignment of DNA Nucleotides Using Monte Carlo Techniques

Rajashree Chaurasia and Udayan Ghose

Abstract Generally, global pairwise alignments are used to infer homology or other evolutionary relationships between any two sequences. The significance of such sequence alignments is vital to determine whether an alignment algorithm is generating the said alignment as evidence of homology or by random chance. Gauging the statistical significance of a sequence alignment obtained through the application of a global pairwise alignment algorithm is a difficult task, and research in this direction has only provided us with nebulous solutions. Moreover, the case of nucleotide alignments with gaps has been scarcely explored. Very little literature exists on the statistical significance of gapped global alignments employing affine gap penalties. This manuscript aims to provide insights into how the statistical significance of gapped global pairwise alignments that may be inferred using Monte Carlo techniques.

Keywords Global pairwise alignment · Scoring matrix · Statistical significance · Gapped alignments · Affine gap penalty · Monte Carlo method · Extreme value distribution

1 Background

Global pairwise alignments are generally used to measure the evolutionary relatedness or homology of two sequences over their entire lengths. The most popular pairwise global alignment algorithm is the Needleman–Wunsch algorithm [19]. However, the standard Needleman–Wunsch works well for minimal length sequences (of the

R. Chaurasia (✉)
Guru Nanak Dev Institute of Technology, Directorate of Training and Technical Education, Government of NCT of Delhi, Delhi, India
e-mail: rajashree.14416490019@ipu.ac.in; rajashree.chaurasia@gmail.com

R. Chaurasia · U. Ghose
University School of Information, Communication & Technology, Guru Gobind Singh Indraprastha University, New Delhi, India
e-mail: udayan@ipu.ac.in

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_5

order of a few hundred nucleotides in a reasonable time) only. Several improvements over this basic algorithm have been devised in [5, 10, 12, 18, 23] to alleviate the problem of scaling up to longer sequences, the most notable among these being the Myers-Miller algorithm for linear space alignment [18]. Many online tools are now available to implement both the basic Needleman–Wunsch algorithms for small length sequences and a faster variant based on the improvisations. Two of such tools are widely used when compared to the others found at [4, 21]. The optimal global alignment scores obtained using these algorithms are not considered as a sufficient proof for their shared homology. Further, it is required to measure their statistical significance to be sure that such an optimal alignment could not be obtained by coincidence. The statistical significance of a global pairwise optimal alignment can be obtained from the optimal alignment score [7]. However, the alignment and the corresponding optimal score are dependent on two parameters, viz. the substitution matrix used and the gap penalty employed. For amino acid sequences, standard substitution matrices (also known as scoring matrices) like Point Accepted Mutation (PAM) [6] and BLOcks SUbstitution Matrix (BLOSUM) [9] are applied. Nevertheless, there are no standard matrices for nucleotide sequences. Scoring matrices for nucleotides consist of match and mismatch scores among the nucleotide bases. Further, gap penalties are either linear or affine. Linear gap penalties impose a penalty proportional to gap length. Affine gap penalties, on the other hand, impose distinct penalties for introducing a gap (higher) and widening an open gap (lower). The extension penalty is taken proportional to gap length after the introduction of a gap. In practice, affine gap penalties are much more widely used than simple linear penalties as they combine the benefits of both a constant gap penalty and a linear penalty. More complex gap penalty models like convex gap penalties and profile-based gap penalties also exist, though they are not widely employed in global alignment.

The significance of optimal alignment scores can be evaluated by its *p*-value that conveys the probability of the best possible alignment with the given optimal score that have occurred with random coincidence. In other words, if the optimal score is higher than all scores obtained in a random model, it can be said that the optimal alignment given by the alignment algorithm is statistically significant, and its *p*-value will remain close to zero. The choice of the random model determines the *p*-value estimation process. There are a few techniques for generating the random model in question. For instance, a random sequence generator function can be used to create independent random sequences of a specified length without providing a template sequences input. Some random sequence generators preserve the G–C (Guanine–Cytosine) content or composition of a template which, is a real sequence from a sequence database [8, 24, 26, 27]. In contrast, others are based on more complex Markov models that are organism-specific or models that preserve dinucleotide frequencies and codon usage [3]. Independent random sequences of similar length can also be taken from real sequence databases.

It is a recognized fact that the distribution of a large group of independent identically distributed random variables follows the Gaussian distribution. It is further substantiated that, for local sequence alignments containing gaps, the distribution of the maximum of such random variables tends towards the extreme value distribution [13]. For local alignments of gapped sequences, some empirical evidence exists [1, 29] that the distribution of scores tends to follow the extreme value distribution (EVD), some studies further pointing out the Gumbel-type EVD as an approximate distribution for scores [11, 16, 17, 20, 22]. However, for global pairwise gapped nucleotide sequence alignments, no theoretical results for optimal score distributions are known even for the simplest random models and fixed sequence-specific alignment. Reich et al. [25] have studied the score distributions for global alignments of nucleotides using a static scoring model and zero-gap penalties. The authors used the *Z*-score, Monte Carlo techniques and a random model that generated independent sequences with individual nucleotide base frequencies around 25%. They found that the score distributions were close to the distributions of sequences randomly retrieved from a sequence database. However, the case for gapped global nucleotide alignments has not been considered in their work. In another study, Altschul and Erickson [3] used a random shuffling generator that preserved dinucleotide and codon usage and used the *p*-value for estimation of score distributions to demonstrate that the score distribution is marginally non-normal. However, the tail behaviour of the distribution was not expounded in their study.

Monte Carlo techniques employ statistical analysis using continual random sampling and are generally used in resolving problems that can be construed in nomenclatures of probability. This study has generated the null model of alignment scores through random shuffling of the query sequence. By continuously sampling the query sequence randomly, the proposed research work has inferred a probabilistic explanation of the statistical significance of the alignment scores. The main concentration is given on affine gapped pairwise nucleotide global alignments and attempts to provide a distribution for the score significance of such alignments.

It is worthwhile to note, however, that assessing the significance of alignment can only point us in the direction of a target exhibiting an exciting pattern and should not be taken as valid proof of biological significance [2, 15, 28]. Experimental studies in a wet-lab can make the only confirmation of biological relevance.

Section 1 of this paper gives a brief background of the research done in this direction, followed by Sect. 2 that details the methodology used in the study. Section 3 discusses the results obtained, and Sect. 4 concludes the paper with comments on laying the groundwork for further extensions to the study.

2 Methods

The algorithm for the global pairwise alignment of DNA sequences employed in our study is the Myers-Miller linear space alignment algorithm implemented in the EMBOSS 6.5.0.0 package for Windows (see ftp://emboss.open-bio.org/pub/EMBOSS/windows/, free and open-source). The EMBOSS program 'stretcher' implements the improvisation over standard Needleman–Wunsch and takes many parameters, some of which are the two sequences to be aligned, the scoring matrix file, the gap open penalty, and the gap extension penalty. The user manual and documentation for this function can be found online at several websites (e.g. see https://bioinfo.ccs.usherbrooke.ca/cgi-bin/emboss/help/stretcher). The fixed-parameter set model is employed for alignment, wherein the standard scoring matrix with symmetric match and mismatch scores and default gap penalties as used by the online version of this tool (see https://www.ebi.ac.uk/Tools/psa/emboss_stretcher/) is applied for the pairwise global alignment. In a fixed-parameter set model, the model parameters are static quantities. This means that these values are not drawn from a random distribution. In our study, the model parameters, as described in section one, are the substitution matrix and affine gap penalties. However, the significance of alignment scores is studied by using the Monte Carlo methods, where the model parameters are required to remain fixed so that they do not bias the alignment score distributions unnecessarily.

The scoring matrix used in this study specifies a match score of '5' and a symmetric and uniform mismatch score of '−4'. These are the default scores used by the EMBOSS 'stretcher' program for bases A, T, G, and C (see the 'EDNAFULL' datafile section at https://bioinfo.ccs.usherbrooke.ca/cgi-bin/emboss/help/stretcher). The gap open penalty (the default value for nucleotides as per the EMBOSS 'stretcher' program) is set to 16, and the gap extension penalty (the default value for nucleotides as per the EMBOSS 'stretcher' program) is set to 4.

Specific sequences of varied lengths are carefully selected from the National Center for Biotechnology Information (NCBI) database (see https://www.ncbi.nlm.nih.gov/nuccore) followed by a Basic Local Alignment Search Tool for Nucleotides (BLASTN) search [4] for highly similar sequences. The sequence pairs thus selected vary in length from less than 1 Kb (where Kb implies kilo basepair) up to 10 Kb. Each of these sequence pairs is globally aligned using the 'stretcher' command, and the optimal alignment score is noted along with other details. A brief representation of the selected sequence pairs is given in Table 1.

In order to generate the random or null model, this research work has utilized the EMBOSS 'shuffleseq' (see https://emboss.bioinformatics.nl/cgi-bin/emboss/shuffleseq) program that shuffles a template nucleotide sequence and generates random sequences based on the composition of the template. Each of these randomly generated sequences preserves the composition of the second sequence that was used as an input to the global alignment algorithm. The sequences from the null model are then aligned globally to the first sequence used as an input to the 'stretcher' program. Hereafter, the first sequence used as input in aligning via EMBOSS 'stretcher' will be referred to as the 'target' sequence, and the second sequence that is used to generate the random model will be referred to as the 'query' sequence. Three sets of random sequences are generated for each sequence pair. The first set contains 200 randomly shuffled sequences, the second set comprises 500 transposed sequences, and the third set holds 1000 such sequences.

Lastly, the scores of alignments are remaining fit for a specific sequence pair to three types of extreme value distributions (EVDs) by estimating their shape, location, and scaling parameters. The p-value is then calculated from these parameter estimates using the cumulative distribution function [14] of the said EVDs. The three types

Table 1 A brief representation of selected sequence pairs for alignment

S.No.	Pair	Sequence length (in base pairs, bp)	Alignment length	Alignment score	Percent identity	Percent similarity	Percent gaps
1	NC_000079.6: 23,763,668 −23,764,412 (Mus musculus H1.1 linker histone)	745 bp	765 bp	819	60.0	60.0	11.1
	NC_030679.2: c135757942–135,758,641 (Xenopus tropicalis H1.3 linker histone)	700 bp					
2	NM_000523 (Homo sapiens homeobox D13)	1008 bp	1032 bp	3678	84.3	84.3	3.5
	NM_008275 (Mus musculus homeobox D13)	1020 bp					
3	KC978991.1 (Faba bean necrotic stunt alphasatellite 1 isolate)	1001 bp	1038 bp	1449	62.9	62.9	5.2
	NC_038958.1 (Pea yellow dwarf alphasatellite 1 isolate)	1021 bp					
4	NM_000522 (Homo sapiens homeobox A13)	1167 bp	1167 bp	4743	90.4	90.4	1.4
	NM_008264 (Mus musculus homeobox A13)	1151 bp					
5	AF531299.1 (Homo sapiens histone H1	1620 bp	1629 bp	-1665	35.1	35.1	54.8
	NC_000079.6: 23,763,668 −23,764,412 (Mus musculus H1.1 linker histone)	745 bp					
6	NC_038298.1 (Bayou virus nucleocapsid)	1958 bp	2015 bp	5434	76.3	76.3	3.8
	KX066124.1 (Muleshoe hantavirus strain HV segment S)	1996 bp					
7	NC_000006.12: 31,575,565–31,578,336 (Homo sapiens tumor necrosis factor)	2772 bp	2963 bp	6004	68.7	68.7	18.4
	M64087.1 (Equus caballus tumor necrosis factor-alpha)	2610 bp					

(continued)

Table 1 (continued)

S.No.	Pair	Sequence length (in base pairs, bp)	Alignment length	Alignment score	Percent identity	Percent similarity	Percent gaps
8	NC_026662.1 (Simian torque teno virus 31 isolate)	3907 bp	4004 bp	46	50.6	50.6	19.5
	NC_014480.2 (Torque teno virus 2)	3322 bp					
9	NC_026662.1 (Simian torque teno virus 31 isolate)	3907 bp	4084 bp	587	50.8	50.8	8.7
	MH649256.1 (Anelloviridae sp. Isolate)	3904 bp					
10	NC_004764.2 (Budgerigar fledgling disease virus-1)	4981 bp	4981 bp	24,635	99.4	99.4	0.0
	AB453162.1 (Budgerigar fledgling disease polyomavirus strain APV4)	4981 bp					
11	NC_048296.1 (Bird's-foot trefoil enamovirus 1 isolate)	5736 bp	5835 bp	12,135	69.2	69.2	3.5
	KY985463.1 (Alfalfa enamovirus 2 isolate)	5729 bp					
12	U31789.1 (Human papillomavirus type 48)	7100 bp	7257 bp	12,975	65.7	65.7	3.2
	U31790.1 (Human papillomavirus type 50)	7184 bp					
13	NC_001362.1 (Friend murine leukemia virus)	8323 bp	8368 bp	29,594	84.1	84.1	1.1
	AB187565.1 (Murine leukemia virus graffi)	8319 bp					
14	M10060.1 (Human T-lymphotropic virus 2)	8952 bp	9039 bp	23,497	74.3	74.3	3.0
	Y14570.1 (Simian T-lymphotropic virus 2)	8855 bp					
15	NC_001802.1 (Human immunodeficiency virus 1)	9181 bp	9611 bp	39,052	89.9	89.9	4.8
	KT284376.1 (Human immunodeficiency virus 1 isolate from the USA)	9579 bp					

The selected sequences are taken from the NCBI nucleotide database (reference numbers are indicated). Note that sequence pair 10 has no gaps (ungapped alignment of same size sequences), and sequence pair 5 aligns sequences of different lengths. Percent identity and percent similarity in the case of nucleotides are equal and imply a ratio of the number of matching bases to the total alignment length. Percent gaps give the ratio of the number of gaps in the alignment to the total length of the alignment of extreme value distributions that have been used in our study are the Generalized EVD, the Gumbel distribution (also known as the Type I EVD), and the Fréchet distribution (also known as the Type II EVD). Generalized EVD is a distribution that combines Type I, Type II, and Type III EVDs. Type III EVD is also known as the Weibull distribution and is not included in this study. All processing of data and its analysis has been programmed using the Python programming language.

3 Results and Discussion

For each sequence pair listed in Table 1, nine curve plots are generated. Every sequence pair is fit to three types of EVDs for each of the three random sequence sets (denoted in the plots as RN = 200, RN = 500, and RN = 1000). The corresponding *p*-values are also saved to an excel sheet carrying other summary data, some of which are included in Table 1.

Figure 1 shows all the nine plots for sequence pair 1. In each graphical plot, the optimal scores obtained by aligning every randomly generated sequence with the

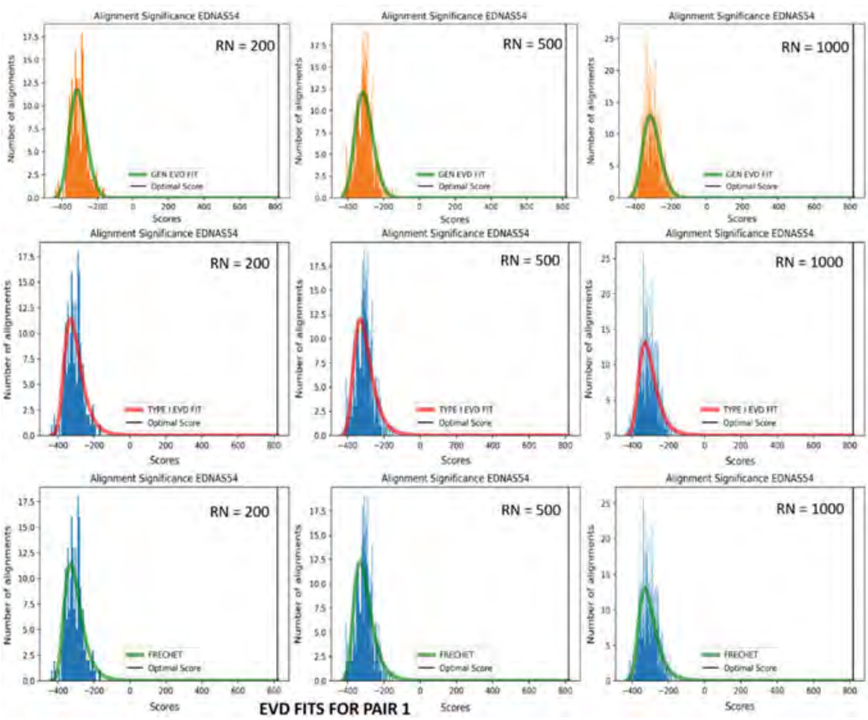


Fig. 1 Curve plots for EVD fits for sequence pair 1. EDNAS54 signifies the substitution matrix file used for EMBOSS 'stretcher' program. RN indicates the random model size

query sequence are plotted against the number of alignments achieving that optimal score. The histogram of scores so obtained is then fit to a generalized EVD, a Gumbel-type EVD, and a Fréchet-type EVD. From Fig. 1, it is clearly evident that all three types of EVDs fit the data sufficiently. This means that the alignment score obtained for sequence pair one is statistically significant. Further, it may be inferred that for this sequence pair, the generalized EVD as well as type I and type II EVDs fit the data well.

Table 2 shows the *p*-value estimates for each of these nine cases, for all sequence pairs. It can be discerned from this table that for sequences having alignment lengths up to 9 kb, and random model size of 200 randomly shuffled sequences, the data fit the generalized, Gumbel- as well as Fréchet-type EVDs. This suggests that the alignment scores obtained for all these pairs of sequences are highly statistically significant and closely follow the EVD distributions considered in this study. In other words, the probability of this random model score distribution being more significant than the optimal alignment score becomes extremely small. For random model sizes of 500 and 1000, only the last two pairs of sequences show that the score distribution follows the Gumbel-type EVD, whereas the remaining pairs fit all types of EVDs considered here, sufficiently. For brevity, the plots for a few notable cases are presented in the figures that follow.

Since *p*-values less than 0.05 are considered statistically significant, all the estimates in Table 2 that are not in boldface are highly statistically significant. Further, a *p*-value of 0.0 is an indication of the null hypothesis (that the optimal alignment score was obtained by mere chance) being an impossibility. High alignment scores are generated as the sequence lengths increase, where sequence similarity is reasonable while aligning. As the optimal alignment scores increase, the random model optimal alignment scores shift farther away from the optimal maximum global pairwise score. For high alignment scores (around 30,000), the generalized EVD and Type II EVD (Fréchet) do not fit the data consistently. However, the Gumbel-type I EVD fits all cases satisfactorily regardless of the percent identity/similarity, percentage, gaps, alignment length, or sequence lengths. This may be so because the Gumbel-type EVD is known to hold for a wide variety of underlying densities [7], including the Normal or Gaussian distribution (which is the case for the null model here).

Note that in Fig. 2, the target sequence is approximately twice in length than the query. Due to the low sequence similarity and high gap percentage, the optimal score is negative. Nevertheless, the maximal random scores are less than the optimal alignment score by a considerable distance, and the score is highly significant in the sense that it is almost nearly impossible that it could have been obtained by fluke. It is also worth noting that sequence pair 10 has the highest similarity of 99.4% among all other sequence pairs, and the optimal alignment contains no gaps. Figure 3 shows the curve plots for this case, and it is clear that the alignment score is highly significant in all cases, and the score distribution follows the EVD distribution faithfully. Further, looking at curve plots of sequence pair 15 in Fig. 4 tells us that the scores follow the Gumbel-type EVD sufficiently well, while the other two EVDs may or may not fit reliably.

Table 2 *P*-values estimates for each sequence pair listed in Table 1

Pair number (refer Table 1)	RN = 200 (random model size = 200)			RN = 500 (random model size = 500)			RN = 1000 (random model size = 1000)		
	Gen EVD	TYPE I EVD	TYPE II EVD	Gen EVD	TYPE I EVD	TYPE II EVD	Gen EVD	TYPE I EVD	TYPE II EVD
1	0.0	5.85409E−12	5.86E−12	0.0	8.41682E−12	3.95695E−12	0.0	2.93221E−12	2.92155E−12
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.22045E−16
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	1.18E−09	4.9E−13	4.63E−13	0.0	7.30527E−14	7.69385E−14	0.0	2.25375E−14	1.63203E−14
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	3.83E−13	3.81E−13	0.0	7.29306E−13	7.29417E−13	0.0	5E−13	5.02E−13
9	0.0	4.97E−13	4.96E−13	0.0	2.24809E−12	7.77167E−12	0.0	1.51201E−12	6.46594E−13
10	1.79E−10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14	0.0	0.0	0.0	0.057849298	0.0	0.072291831	0.094655538	0.0	0.08820396
15	0.075242874	0.0	0.0	0.0	0.0	0.0	0.090770228	0.0	0.075284757

The numbers in boldface indicate high *p*-values, and the corresponding result is not significant. Gen EVD = Generalized extreme value distribution, Type I EVD = Gumbel EVD, and Type II EVD = Fréchet EVD

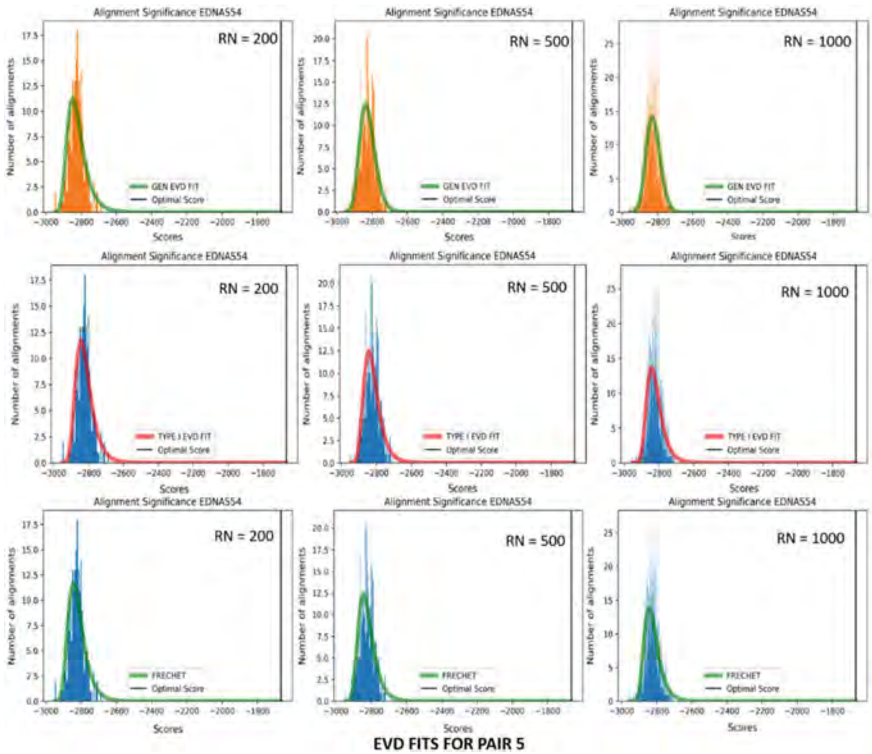


Fig. 2 Curve plots for EVD fits for sequence pair 5. Pair 5 aligns two sequences of different lengths (the target being 1620 bp and the query being 745 bp long). Further, the optimal alignment score so obtained is negative, the percent similarity/identity is the least (35.1%), and the percentage gaps is the highest (54.8%) among the pairs of sequences selected for the study

4 Conclusion and Future Work

The proposed research work has attempted to provide empirical evidence of the statistical significance of optimal alignment scores acquired through global pairwise alignment with gaps. Further, this research work has considered the fixed-parameter model wherein only a particular set of a predetermined scoring matrix, and default gap penalties are applied to all the sequence pairs for each random model. By applying the repeated random sampling technique (Monte Carlo methods), it has been attempted to assess the statistical significance of gapped global pairwise alignments for the case of nucleotides.

The optimal alignment scores generated for small length sequences of the range 1–10 kb are found to be highly significant. Further, the statistical significance of such alignments tends to follow the extreme value distribution, as is evident from the *p*-value estimates obtained. However, as the alignment lengths increase beyond 9 kb (and as the alignment scores increase as a consequence of moderate to high

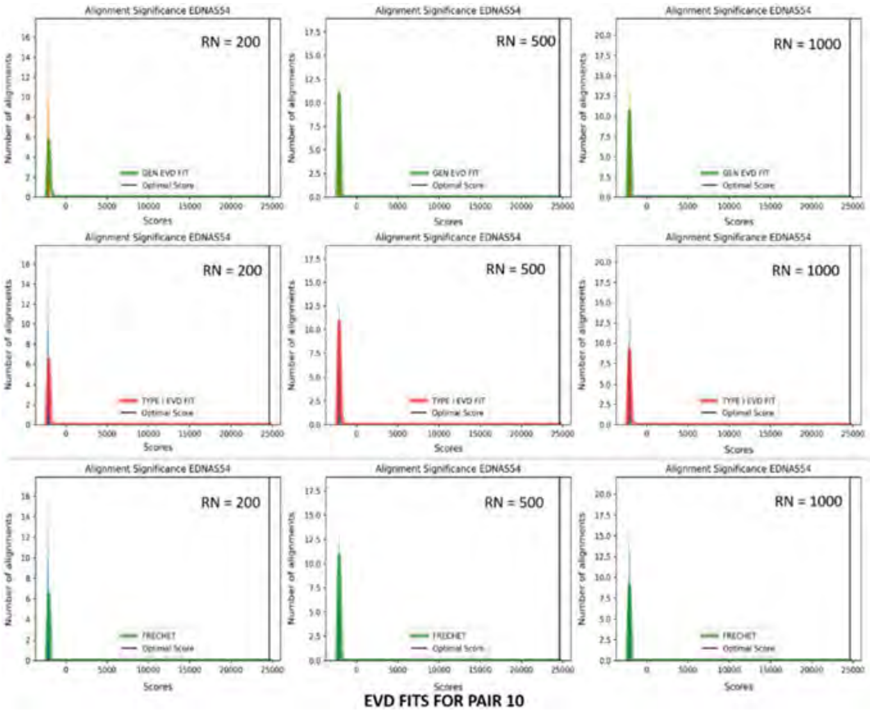


Fig. 3 Curve plots for sequence pair 10. The histogram is barely visible in this case owing to the very high globally optimal score obtained compared to the optimal scores of the null model

similarity), it is observed that the alignment scores closely abide by the Gumbel-type EVD reliably well. Therefore, it is reasonable to conclude that for longer sequences, the statistical significance of scores may be explained by the Type I extreme value distribution. Further, looking at all the plots shown in the figures in section three, it can be seen that the distributions are positively skewed with a right tail. However, these tails are thinner than the tails of the Gaussian distribution. The *p*-value calculated in this study considers the right tailed behaviour of the distribution and the very low *p*-values (that are not in boldface) clearly show that the probability of obtaining an optimal alignment score by chance will be almost negligible. In cases where the *p*-value turns out to be zero, this probability also reduces to zero, indicating that arriving at a score as extreme as the optimal, by fluke, is impossible.

Regardless, further studies are necessary that scale up well to moderate-length and large sequences. In order to do away with the runtime overhead of generating and aligning the randomly shuffled sequences with the target to obtain the null model, other techniques that utilize a precompiled random model may be devised. Further, the results may be validated against a random sampling of real sequences of similar lengths as the target alignment. Complex random models employed by Altshul and

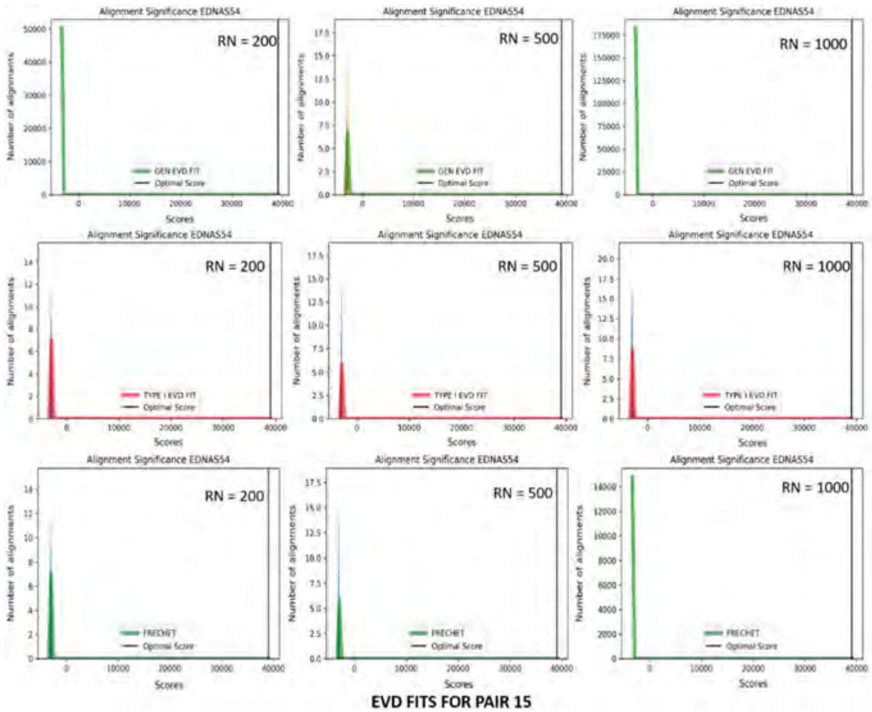


Fig. 4 Curve plots for sequence pair 15. For a random model size of 200 and 1000, the generalized EVD fails to fit the alignment score histogram. Likewise, the Fréchet EVD distribution does not fit the data for the null model size of 1000 shuffled sequences

Erickson [3] can also be explored to strengthen further the empirical analysis for global gapped pairwise nucleotide alignment significance.

This study can be extended to include a multiple fixed-parameter set model in conjunction with multiple null models, and the results obtained may then be validated against a real sequence database search to provide robust empirical evidence of the statistical significance of alignment. Work in this direction is currently underway.

Conflicts of Interest The authors declare no conflicts of interest.

References

1. S.F. Altschul, W. Gish, Local alignment statistics. Methods Enzymol. **266**, 460–480 (1996). https://doi.org/10.1016/s0076-6879(96)66029-7
2. S.F. Altschul, M.S. Boguski, W. Gish, J.C. Wootton, Issues in searching molecular sequence databases. Nat. Genet. **6**(2), 119–129 (1994). https://doi.org/10.1038/ng0294-119

3. S.F. Altschul, B.W. Erickson, Significance of nucleotide sequence alignments: a method for random sequence permutation that preserves dinucleotide and codon usage. Mol. Biol. Evol. **2**(6), 526–538 (1985). https://doi.org/10.1093/oxfordjournals.molbev.a040370
4. BLAST Global Alignment (n.d.) Needleman-Wunsch Global Align Nucleotide Sequences. https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=blastn&BLAST_SPEC=GlobalAln&LINK_LOC=BlastHomeLink
5. A. Chakraborty, S. Bandyopadhyay, FOGSAA: Fast Optimal Global Sequence Alignment Algorithm, 3 (2013). https://doi.org/10.1038/srep01746
6. M.O. Dayhoff, R. Schwartz, R.C. Orcutt, A model of Evolutionary Change in Proteins, in Atlas of Protein Sequence and Structure vol. 5, supplement third. Nat. Biomed. Res. (1978), pp. 345–358. ISBN 978–0–912466–07–1
7. R. Durbin, S. Eddy, A. Krogh, G. Mitchison, *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids* (Cambridge University Press, Cambridge, 1998). https://doi.org/10.1017/CBO9780511790492
8. EMBOSS Shuffleseq (n.d.). EMBOSS Shuffleseq Tool. https://emboss.bioinformatics.nl/cgi-bin/emboss/shuffleseq
9. S. Henikoff, J.G. Henikoff, Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. U.S.A. **89**(22), 10915–10919 (1992). https://doi.org/10.1073/pnas.89.22.10915
10. D.S. Hirschberg, A linear space algorithm for computing maximal common subsequences. **18**, 341–343 (1975). https://doi.org/10.1145/360825.360861
11. X. Huang, D.L. Brutlag, Dynamic use of multiple parameter sets in sequence alignment. Nucleic Acids Res. **35**(2), 678–686 (2007). https://doi.org/10.1093/nar/gkl1063
12. T. Kahveci, V. Ramaswamy, H. Tao, T. Li, *Approximate Global Alignment of Sequences. Approximate Global Alignment of Sequences* (n.d.) https://doi.org/10.1109/bibe.2005.13
13. S. Karlin, S.F. Altschul, Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proc. Natl. Acad. Sci. U.S.A. **87**(6), 2264–2268 (1990). https://doi.org/10.1073/pnas.87.6.2264
14. MATLAB Help Center (n.d.) Assessing the significance of an alignment. https://www.mathworks.com/help/bioinfo/examples/assessing-the-significance-of-an-alignment.html
15. A.Y. Mitrophanov, M. Borodovsky, Statistical significance in biological sequence analysis. Brief. Bioinform. **7**(1), 2–24 (2006). https://doi.org/10.1093/bib/bbk001
16. R. Mott, Accurate formula for P-values of gapped local sequence and profile alignments. J. Mol. Biol. **300**(3), 649–659 (2000). https://doi.org/10.1006/jmbi.2000.3875
17. R. Mott, Maximum-likelihood estimation of the statistical distribution of Smith-Waterman local sequence similarity scores. Bltn. Mathcal. Biol. **54**, 59–75 (1992). https://doi.org/10.1007/BF02458620
18. E.W. Myers, W. Miller, Optimal alignments in linear space. Bioinformatics **4**, 11–17 (1988)
19. S.B. Needleman, C.D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48**, 443–453 (1970). https://doi.org/10.1016/0022-2836(70)90057-4
20. R. Olsen, R. Bundschuh, T. Hwa, Rapid assessment of extremal statistics for gapped local alignment. in *Proceedings. International Conference on Intelligent Systems for Molecular Biology* (1999) 211–222.
21. Pairwise Sequence Alignment (n.d.) EMBL-EBI Pairwise Sequence Alignment Tools. https://www.ebi.ac.uk/Tools/psa/
22. W. Pearson, Empirical statistical estimates for sequence similarity searches. J. Mol. Biol. **276**(1), 71–84 (1998)
23. D.R. Powell, L. Allison, T.I. Dix, A versatile divide and conquer technique for optimal string alignment. **70**, 127–139 (1999). https://doi.org/10.1016/s0020-0190(99)00053-8
24. Random DNA Sequence Generator (n.d.). https://www.faculty.ucr.edu/~mmaduro/random.htm
25. J.G. Reich, H. Drabsch, A. Däumler, On the statistical assessment of similarities in DNA sequences. Nucleic Acids Res. **12**(13), 5529–5543 (1984). https://doi.org/10.1093/nar/12.13.5529
26. RSAT (n.d.). Random Sequence Web Tool. https://rsat.sb-roscoff.fr/random-seq_form.cgi

27. SMS (n.d.). Sequence Manipulation Suite. https://www.bioinformatics.org/sms2/random_dna.html
28. M.S. Waterman, *Mathematical Methods for DNA Sequences* (CRC Press Inc., United States, 1989).
29. M.S. Waterman, M. Vingron, Rapid and accurate estimates of statistical significance for sequence database searches. Proc. Natl. Acad. Sci. U.S.A. **91**(11), 4625–4628 (1994). https://doi.org/10.1073/pnas.91.11.4625

Principal Integrant Analysis Based Liver Disease Prediction Using Machine Learning

M. Shyamala Devi, Kamma Rahul, Ambati Aaryani Chowdary, Jampani Sai Monisha Chowday, and Satheesh Manubolu

Abstract Human liver is considered as the most essentially functioning body organ to aid the process of food digestion and remove the toxic substances. The later diagnosis of liver disease will lead to life-threatening condition that damages the life of an individual. Machine learning can be used to completely analyze the features and predict the severity of liver disease. This paper attempts to analyze each clinical features that influence the target with the following contributions. Firstly, the liver dataset from UCI machine repository is subjected with the data processing and cleansing. Secondly, the ANOVA test is applied to verify the features with PR(>F) < 0.05 that highly influences the target. Thirdly, the data is applied to principal component analysis and then fitted to all the classifiers to analyze the performance metrics. Fourth, the data is applied to linear discriminant analysis and then fitted to all the classifiers to analyze the performance. Experimental results shows that "total proteins" do not contribute to target, and passive aggressive classifier is found to provide the accuracy of 71% before and after feature scaling for both the principal component analysis and linear discriminant analysis.

Keywords Machine learning · Classification · Accuracy and feature scaling

1 Introduction

People are facing several disease due to the evolving contemporary lifestyle, food habits and adverse change in the environmental condition. Predicting the disease in early stage may help the patient to fight the critical condition. Several attempts are made to diagnose and predict the disease using machine learning models. Data mining finds hidden pattern information in the huge amount of medical data. [1] The disease can be predicted based on the patient symptoms and KNN, and CNN can be used for the accurate prediction of the diseases. This paper [2–4] develops medical

M. Shyamala Devi (✉) · K. Rahul · A. A. Chowdary · J. S. M. Chowday · S. Manubolu
Department of Computer Science & Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India
e-mail: shyamaladevim@veltech.edu.in

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*, Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_6

cost prediction along with the statistical machine learning model. The coefficient of determination in the prediction model (0.42) was higher than that of a traditional linear model (0.25). The review on [5–8] predictive models to perform chronic disease diagnosis is analyzed in this paper. It is analyzed that about 45% of studies have used SVM models, 23% of the studies have used K-nearest neighbor and Naïve Bayes models, 18% of studies have applied logistic regression, and 14% of studies have applied random forest for disease prediction [9, 10].

1.1 Disadvantages of the Existing Model

The coefficient of determination in the prediction model (0.42) was higher than that of a traditional linear model (0.25). The review on [11, 12] predictive models that help to perform chronic disease diagnosis is analyzed in this paper. It is analyzed that about 45% of studies have used SVM models, 23% of the studies have used K-nearest neighbor [13] and Naïve Bayes models, 18% of studies have applied logistic regression, and 14% of studies have applied random forest for the disease prediction [14]. In the existing models, the random forest classifiers and KNN attempt to predict the liver disease. The review shows that no attempts have been made to analyze the performance of all the classifiers with and without feature scaling [15]. Various diseases are diagnosed using neural networks to increase the speed of decision making, and it can lower the false positive rates. The accuracy of the machine learning algorithm will mainly depend on the quality of the dataset.

2 Overall Architecture

2.1 Dataset Preparation

The Indian liver patient dataset from UCI machine repository has been utilized in this research work [https://archive.ics.uci.edu/ml/machine-learning-databases/00225/] with this link. The overall workflow is shown in Fig. 1. The paper's contributions are given below.

(i) Firstly, the liver dataset is subjected to the data processing and cleansing.
(ii) Secondly, the ANOVA test is applied to verify the features with PR(>F) < 0.05 that highly influence the target.
(iii) Thirdly, the data is applied to principal component analysis and then fitted to all the classifiers before and after feature scaling to analyze the performance.
(iv) Fourth, the data is applied to linear discriminant analysis and then fitted to all the classifiers and after feature scaling to analyze the performance.

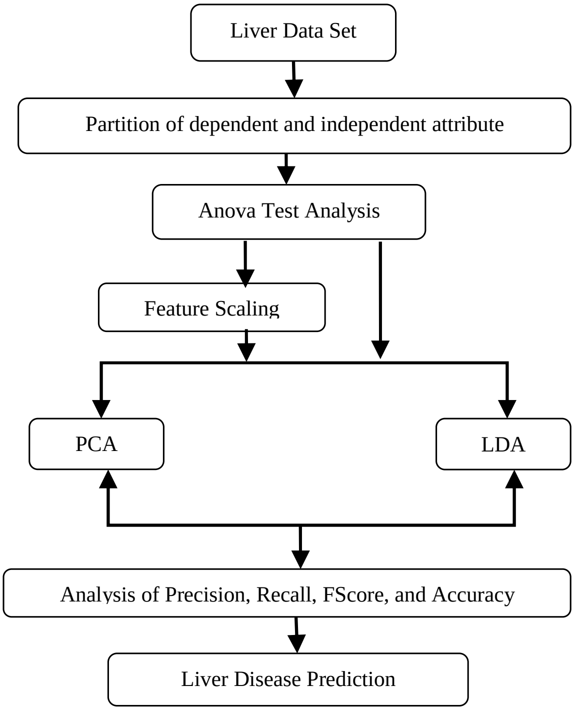


Fig. 1 Overall architecture flow

3 Feature Analysis

3.1 ANOVA Test Analysis

ANOVA test is applied to dataset features, and results show that "total proteins" have value of PR(>F) > 0.05 and do not contribute to target as shown in Table 1.

ANOVA test is used to analyze the features of the dataset by comparing both the null and alternate hypothesis. If the *P* value associated with the Fstatistic is less than 0.05, then the existence of that feature will highly influence the target.

3.2 Data Exploratory Analysis

The correlation of each feature in the dataset is analyzed and is shown in Figs. 2 and 3.

The correlation quantifies the relationship of the features present in the dataset. It shows the association level of two features and their contribution toward predicting the target. The associative level of two features can either be strong or weak. Here the pair (Gender, Total_Bilirubin) is a highly correlated pair toward target prediction.

Table 1 ANOVA test analysis with the dataset features

Features	sum_sq	df	F	PR(>F)
Age	2.248032	1	11.17143	0.00088
Gender	0.8094	1	3.973363	0.04669
Total_Bilirubin	5.778375	1	29.60928	7.80E-08
Direct_Bilirubin	7.213982	1	37.43959	1.73E-09
Alkaline_Phosphotase	4.072429	1	20.55844	0.000007
Alamine_Aminotransferase	3.182228	1	15.94122	0.000074
Aspartate_Aminotransferase	2.750741	1	13.72864	0.000231
Total_Protiens	0.146043	1	0.712934	0.398819
Albumin	3.103722	1	15.53743	0.000091

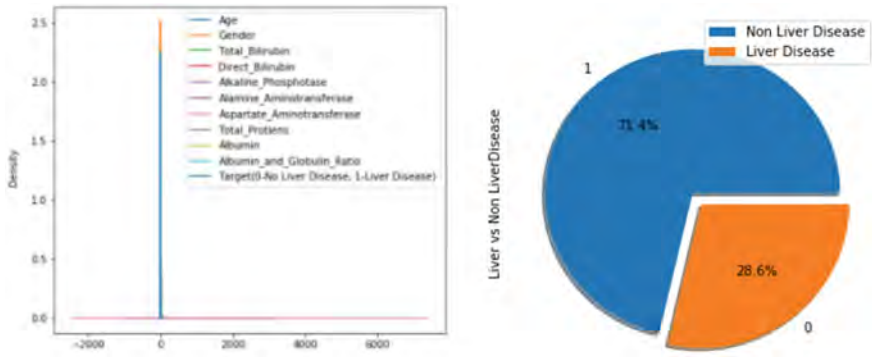


Fig. 2 Dataset ınformation and kernel density estimation plot

3.3 *Dimensionality Reduction with PCA and LDA*

Principal component analysis is a linear dimensionality reduction that uses singular value decomposition of features that reduces to minimized dataset. PCA decomposes multivariate dataset into a multiple sets of orthogonal components that forms the highest variance. The linear discriminant analysis fits the dataset into its conditional densities using Bayes' rule within the linear decision boundary. İt assumes that entire class agrees to same covariance matrix and fits the Gaussian density to each class in the dataset. The dataset is turned into reduced dataset by projecting discriminative features using transform method. This paper uses both PCA and LDA to perform dimensionality reduction. The relationship between component and cumulative variance with PCA is shown in Fig. 4.

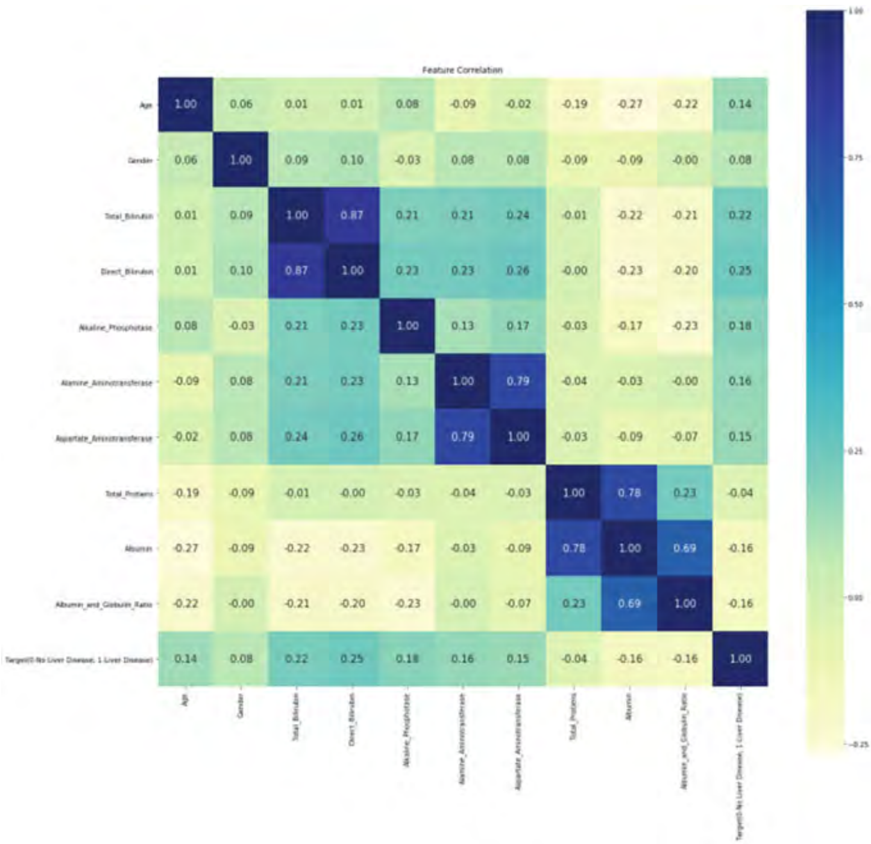


Fig. 3 Dataset correlation ınformation

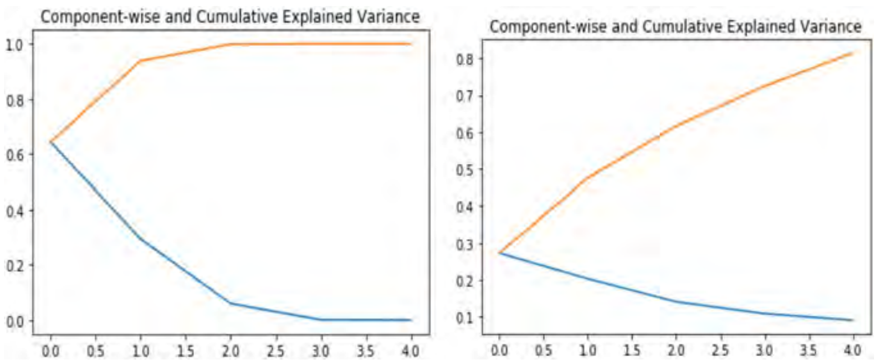


Fig. 4 PCA analysis—five components (left) before (right) after feature scaling

4 Results and Discussion

4.1 Implementation Setup

The Indian liver dataset extracted from the UCI machine learning repository is used for implementation. The dataset consists of the features, namely age, gender, total bilirubin, direct bilirubin, alkaline phosphatase, alamine aminotransferase, aspartate aminotransferase, total proteins, albumin and target. The data preprocessing is done by label encoding the values of gender and missing values. The data cleansing is done after performing ANOVA test. ANOVA test is applied to dataset features, and results show that the Fstatistic value of "total proteins" greater than 0.05 and its existence does not influence the target. So the feature "total proteins" is eliminated before processing the data for classifier to perform the purpose of data cleansing. The correlation of all the features toward the target is done, and dimensionality reduction is applied. The dataset is split with 80:20 for training and testing dataset.

4.2 PCA Analysis Before and After Feature Scaling

Feature scaling is done to normalize the values of independent variables in the dataset. It greatly affects the performance of any algorithm toward the target prediction. So we have implemented the dimensionality reduction before and after feature scaling in order to view the deviation in the performance metrics. Principal component analysis with five components is applied in the dataset, the five-component PCA reduced dataset is fitted with the classifiers like logistic regression, KNN, kernel SVM, decision tree, random forest, gradient boosting, AdaBoost, ridge, RidgeCV, SGD, passive aggressive and bagging classifier, and the performance metrics are shown in Figs. 5 and 6.

4.3 LDA Analysis Before and After Feature Scaling

Linear discriminant analysis with five components is applied in the dataset, the five-component LDA reduced dataset is fitted with the classifiers like logistic regression, KNN, kernel SVM, decision tree, random forest, gradient boosting, AdaBoost, ridge, RidgeCV, SGD, passive aggressive and bagging classifier, and the performance metrics are shown in Figs. 7 and 8.

dflogPCA_BeforeFeatureScaling - DataFrame

Index	Classifer	Precision	Recall	FScore	Accuracy
0	Logistic Regression	0.661059	0.683761	0.614966	0.683761
1	KNeighborsClassifier	0.657262	0.675214	0.661115	0.675214
2	Kernel SVM Classifier	0.781609	0.675214	0.552749	0.675214
3	Gaussian Naive Bayes Classifier	0.794853	0.589744	0.580059	0.589744
4	Decision Tree Classifier	0.634681	0.649573	0.639823	0.649573
5	Extra Tree Classifier	0.707721	0.717949	0.710101	0.717949
6	Random Forest Classifier	0.723649	0.735043	0.722019	0.735043
7	Gradient Boosting Classifier	0.707071	0.717949	0.682724	0.717949
8	Ada Boost Classifier	0.704639	0.717949	0.687845	0.717949
9	Ridge Classifier	0.444444	0.666667	0.533333	0.666667
10	RidgeCV Classifier	0.444444	0.666667	0.533333	0.666667
11	SGD Classifier	0.785179	0.555556	0.537539	0.555556
12	Passive Aggressive Classifier	0.466377	0.521368	0.488725	0.521368
13	Bagging Classifier	0.737979	0.74359	0.739865	0.74359

Fig. 5 PCA dataset with classifier performance metrics analysis before feature scaling

dflogPCA_AfterFeatureScaling - DataFrame

Index	Classifer	Precision	Recall	FScore	Accuracy
0	Logistic Regression	0.673077	0.692308	0.637363	0.692308
1	KNeighborsClassifier	0.60592	0.641026	0.611403	0.641026
2	Kernel SVM Classifier	0.613043	0.666667	0.548212	0.666667
3	Gaussian Naive Bayes Classifier	0.724034	0.623932	0.631393	0.623932
4	Decision Tree Classifier	0.684858	0.700855	0.68615	0.700855
5	Extra Tree Classifier	0.657262	0.675214	0.661115	0.675214
6	Random Forest Classifier	0.669412	0.692308	0.650883	0.692308
7	Gradient Boosting Classifier	0.634021	0.666667	0.63109	0.666667
8	Ada Boost Classifier	0.672174	0.692308	0.671324	0.692308
9	Ridge Classifier	0.444444	0.666667	0.533333	0.666667
10	RidgeCV Classifier	0.444444	0.666667	0.533333	0.666667
11	SGD Classifier	0.718084	0.675214	0.684296	0.675214
12	Passive Aggressive Classifier	0.705856	0.709402	0.707417	0.709402
13	Bagging Classifier	0.672174	0.692308	0.671324	0.692308

Fig. 6 PCA dataset with classifier performance metrics analysis after feature scaling

dflogLDA_BeforeFeatureScaling - DataFrame

Index	Classifer	Precision	Recall	FScore	Accuracy
0	Logistic Regression	0.628713	0.666667	0.618385	0.666667
1	KNeighborsClassifier	0.568486	0.598291	0.578544	0.598291
2	Kernel SVM Classifier	0.444444	0.666667	0.533333	0.666667
3	Gaussian Naive Bayes Classifier	0.444444	0.666667	0.533333	0.666667
4	Decision Tree Classifier	0.580982	0.606838	0.589771	0.606838
5	Extra Tree Classifier	0.542125	0.581197	0.555424	0.581197
6	Random Forest Classifier	0.580982	0.606838	0.589771	0.606838
7	Gradient Boosting Classifier	0.656147	0.683761	0.637955	0.683761
8	Ada Boost Classifier	0.576923	0.641026	0.576923	0.641026
9	Ridge Classifier	0.444444	0.666667	0.533333	0.666667
10	RidgeCV Classifier	0.444444	0.666667	0.533333	0.666667
11	SGD Classifier	0.785507	0.683761	0.57138	0.683761
12	Passive Aggressive Classifier	0.674463	0.683761	0.677841	0.683761
13	Bagging Classifier	0.567892	0.589744	0.576307	0.589744

Fig. 7 LDA dataset with classifier performance metrics analysis after feature scaling

dflogLDA_AFTERFeatureScaling - DataFrame

Index	Classifer	Precision	Recall	FScore	Accuracy
0	Logistic Regression	0.628713	0.666667	0.618385	0.666667
1	KNeighborsClassifier	0.568486	0.598291	0.578544	0.598291
2	Kernel SVM Classifier	0.444444	0.666667	0.533333	0.666667
3	Gaussian Naive Bayes Classifier	0.444444	0.666667	0.533333	0.666667
4	Decision Tree Classifier	0.580982	0.606838	0.589771	0.606838
5	Extra Tree Classifier	0.542125	0.581197	0.555424	0.581197
6	Random Forest Classifier	0.580982	0.606838	0.589771	0.606838
7	Gradient Boosting Classifier	0.656147	0.683761	0.637955	0.683761
8	Ada Boost Classifier	0.576923	0.641026	0.576923	0.641026
9	Ridge Classifier	0.444444	0.666667	0.533333	0.666667
10	RidgeCV Classifier	0.444444	0.666667	0.533333	0.666667
11	SGD Classifier	0.754902	0.666667	0.67464	0.666667
12	Passive Aggressive Classifier	0.677984	0.683761	0.680419	0.683761
13	Bagging Classifier	0.609668	0.623932	0.615226	0.623932

Fig. 8 LDA dataset with classifier performance metrics analysis after feature scaling

5 Conclusion

This paper attempts to explore the feature analysis of the dataset by interpreting the relationship of feature with each other. The data correlation matrix is extracted to identify the highly correlated dataset pair toward the target prediction. The dimensionality reduction of the dataset is done with principal component analysis and linear discriminant analysis. The PCA reduced dataset and LDA reduced dataset are fitted with the classifiers, namely logistic regression, KNN, kernel SVM, decision tree, random forest, gradient boosting, AdaBoost, ridge, RidgeCV, SGD, passive aggressive and bagging classifier before and after feature scaling. Experimental result shows that the PCA reduced dataset with random forest has 74% accuracy before feature scaling, and passive aggressive classifier has 71% accuracy after feature scaling. Experimental result shows that LDA reduced dataset with gradient boosting and passive aggressive classifier has 68% accuracy before and after feature scaling. It is observed that passive aggressive classifier is projecting the nominal accuracy for predicting the target variable.

References

1. D. Dahiwade, G. Patle, E. Meshram, Designing disease prediction model using machine learning approach, in *3rd International Conference on Computing Methodologies and Communication (ICCMC)* (Erode, 2019), pp. 1211–1215
2. T. Takeshima, S. Keino, R. Aoki, T. Matsui, K. Iwasaki, Development of medical cost prediction model based on statistical machine learning using health ınsurance claims data. Res. Meth. Model. Meth. **21**, 97 (2018)
3. G. Battineni, G.G. Sagaro, N. Chinatalapudi, F. Amenta, Applications of machine learning predictive models in the chronic disease diagnosis. J. Pers. Med. **10**(2), 21 (2020). https://doi.org/10.3390/jpm10020021
4. M. Chen, Y. Hao, K. Hwang, L. Wang, Disease prediction by machine learning over big data from healthcare communities. IEEE Access **5**, 8869–8879 (2017). https://doi.org/10.1109/ACCESS.2017.2694446
5. K. Shailaja, B. Seetharamulu, M.A. Jabbar, Machine learning in healthcare: a review, in *Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (Coimbatore, 2018), pp. 910–914. https://doi.org/10.1109/ICECA.2018.8474918
6. R. Bhardwaj, A.R. Nambiar, D. Dutta, A study of machine learning in healthcare, in *IEEE 41st Annual Computer Software and Applications Conference (COMPSAC),* (Turin, 2017), pp. 236–241, https://doi.org/10.1109/COMPSAC.2017.164
7. T. Le Nguyen, T.H. Do, Artificial ıntelligence in healthcare: a new technology benefit for both patients and doctors, in *International Conference on Management of Engineering and Technology* (Portland, 2019), pp.1–15, https://doi.org/10.23919/PICMET.2019.8893884
8. B. Nithya, V. Ilango, Predictive analytics in health care using machine learning tools and techniques. in *International Conference on Intelligent Computing and Control Systems (ICICCS)* (Madueai, 2017), pp. 492–499, https://doi.org/10.1109/ICCONS.2017.8250771
9. T. Vijayakumar, Neural network analysis for tumor investigation and cancer prediction. J. Electron. **1**(02), 89–98 (2019)
10. S. Shakya, Analysis of artificial intelligence based image classification techniques. J. Innov. Image Process. (JIIP) **2**(01), 44–54 (2020)

11. Koresh, H. James Deva, Computer vision on based traffic sign sensıng for smart transport. J. Innov. Image Process. (JIIP) **1**(01), 11–19
12. A. Bashar, Survey on evolving deep learning neural network architectures. J. Artif. Intell. **1**(02), 73–82
13. M. Duraipandian, Performance evaluation of routing algorithm for manet based on the machine learning techniques. J. Trends Comput. Sci. Smart Technol. (TCSST) **1**(01), 25–38 (2019)
14. S. Subarna, Analysis of artificial ıntelligence based ımage classification techniques. J. Innov. Image Process. 44–54 (2020)
15. V. Suma, Computer vision for human-machine interaction-review. J. Trends Comput. Sci. Smart Technol. **1**(02), 131–139 (2019)

Classification of Indian Classical Dance 3D Point Cloud Data Using Geometric Deep Learning

Ashwini Dayanand Naik and M. Supriya

Abstract Indian classical dances have many unique postures that require to be identified and classified correctly. Though many classification techniques exist for two-dimensional dance images, there is a need to classify the three-dimensional images as it is still on the evolving side. Geometric deep learning is one of the growing fields in machine learning and deep learning. It enables to learn from complex type of data represented in the form of graphs and 3D objects (manifolds). Deep learning algorithms like convolution neural networks (CNN) and recurrent neural networks (RNN) have achieved higher performance on the broad range of problems. Using these algorithms, one can also classify the images. Deep learning algorithm works well for Euclidean data such as points, lines, and planes. CNN cannot be implemented on the non-Euclidean data such as graphs and 3D object (manifolds), and thus, neural network architecture that can learn from non-Euclidean data is required. In the proposed work, implementation of geometric deep learning is done on 3D image data represented as point cloud. PointNet architecture will work efficiently with point cloud data. This architecture has been used to classify Indian classical dance point cloud data into five dance forms, namely Bharatanatyam, Odissi, Kathak, Kathakali, and Yakshagana.

Keywords Geometric deep learning · Point cloud data · Indian classical dance forms · Euclidean and non-Euclidean data

1 Introduction

Amalgamation of many cultures can be found in India, and it includes language, dance, music, religion, and food. Indian classical dance is one of the ancient arts,

A. D. Naik (✉) · M. Supriya
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India

M. Supriya
e-mail: m_supriya@blr.amrita.edu

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_7

which has unique features such as hand mudra, leg postures, and costumes each of which are different for each classical dance form. According to natya shastra, one of the ancient literatures on dance estimates the presence of dance in India between 500 BCE and 500 CE. There are 11 classical dance forms in India, namely Bharatanatyam and Bhagavatha Mela from Tamil Nadu, Kathakali from Kerala, Manipuri from Manipur, Kathak from Uttar Pradesh, Kuchipudi from Andhra Pradesh, Odissi from Odisha, Mohiniyattam from Kerala, Sattriya from Assam, Yakshagana from Karnataka and Chhau from Eastern India. Each Indian classical dance expresses cultural richness of each state. The person who is not aware of dance cannot identify the difference in dance postures and hand mudras. If one is interested to learn the basics and technicalities of Indian classical dance forms, they can find it from many online sites who offer such training. Such type of learning to understand the basics can now happen through many digitized platforms too. The proposed work focusses on one such approach to enable the learning of digital images represented in 3D format, which brings out the third dimension to the standard 2D images and enables a higher learning accuracy.

For classifying the images, machine learning [1] and deep learning [2] are the appropriate approaches. Many approaches have been proposed in the literature to classify Indian classical dance images. The main focus of such models is to consider two-dimensional image data which is fed as input to the convolution neural network (CNN) [3]. Current work includes classifying Indian classical dance forms represented as three-dimensional point cloud data using geometric deep learning. Convolution neural network is only suitable for image data which is represented as Euclidean data. Convolution is done by sliding the filter over the input at every location providing a highly accurate system. If the image is curved, it becomes as a 3D object. Convolving around 3D shapes by vector like filter is not possible. So, the geometric deep learning is the solution for the problem. Here, the non-Euclidean data such as 3D object point cloud data is considered as the input. Representation of set of points in 3D by x, y, z location is called point cloud. Here, dance form 3D point cloud data is fed to the network to classify the point cloud with label. To classify the point cloud data, PointNet architecture is used. PointNet architecture directly takes the 3D point cloud data without transforming them into other forms. There are other implementations, which makes use of mesh, volumetric and RGB(D) data. Advantage of using point cloud data among these is that it makes use of raw sensor data and it is canonical feature.

2 Related Works

In recent years, many research works are performed in the field of image processing, machine learning, and deep learning. Ankita et. al. have made an attempt to recognize Indian classical dance from the videos [4]. Training has been done using support vector machine (SVM). The work has been trained on the dataset consisting of 211 videos that yield an accuracy of 75.83%. CNN architecture has been proposed to classify Bharatanatyam mudra and poses in the work proposed in [5]. Here, the dataset consists of 200 different mudras and provides an accuracy of 89.92%. Pose descriptor such as histogram of oriented optical flow (HOOF) [6] has also been used to represent the dance video frames. Online dictionary learning technique is used to learn pose basis using SVM which classifies the videos into three different dance forms-Bharatanatyam, Kathak, and Kathakali. The accuracy of 86.67% has been achieved by this approach.

Many studies have been conducted on action recognition on three-dimensional objects using geometric-based methods. The work by Yui et al. has introduced the special form of the data such as graphs and manifolds [7] which is also called as non-Euclidean data. Human manifolds are used for action recognition. Discriminating information can be extracted between actions by manifold charting. Singular value decomposition (SVD) is used for factorizing the data tensors which are projected on a tangent space. After computing the tangent bundle's intrinsic distance, actions can be classified. Implementation of geometric deep learning on non-Euclidean data is briefly explained in the work proposed in [8]. Deep learning does well at extracting features from Euclidean data. Since graphs and manifolds come under non-Euclidean data, normal convolution approach cannot be implemented, but by generalizing the convolution operation, it can be achieved. Hence, the proposed model uses deep learning for non-Euclidean domains by making use of spatial-domain feature.

Three-dimensional surface data of an object can be represented using point clouds with x-, y-, z-coordinates. There are few works which have implemented deep learning on point clouds. Instead of normal CNN convolution technique, edge convolution is involved in point cloud world [9]. Edge convolution (EdgeConv) has many properties, such as incorporating local neighborhood information and learning global shape properties. Here, the architecture is tested on the ModelNet40 and S3DIS 3D datasets. Graph convolution neural network (GCNN) for 3D point cloud classification is implemented in the work proposed in [10]. GCNN is the extension of traditional CNN, where in GCNN data used is in graph form. In this work, graph-CNN is used for classifying 3D point cloud data of an object. Model implemented the combination of localized graph convolutions with two types of graph pooling methods (to downsize the graph). Graph-CNN can explore the point cloud data. This model achieves competitive performance on the 3D object classification on ModelNet dataset. Ernesto Brau et.al. have implemented deep CNN to estimate the 3D human pose which learns from 2D joint annotations [11]. This architecture uses traditional way, but the output layer will project the predicted 3D joints on the 2D. It enforces pose constraints using an independently trained network that learns a prior distribution over 3D poses. An approach is made to recognize the human action using human skeletal information [12] which considers the skeletal angular representation. The scale of the actor is identified and also correlation among different body parts is maintained.

A neural network using the point clouds has been developed by Charles et al. in [13]. This model called PointNet has been designed such that it provides a unified architecture for variety of applications such as classification of objects, semantic parsing, and segmentation. Main challenge addressed in this work is to feed the unordered point cloud data to the network which has the features such as permutation invariance and rotation invariance. PointNet is tested on ModelNet40 [14, 15] dataset. Literature summary is shown in the Table 1.

From the survey, it could be concluded that most of the work uses two-dimensional Indian classical dance image data. There is need of three-dimensional data to be trained in the emerging future applications, and hence, geometric deep learning is the possible option. Considering three-dimensional point cloud data with PointNet architecture is more effective as there is no need of conversion of data to other form. Such a model can be used to recognize the dance images and their varied forms more precisely yielding high classification accuracy.

3 Proposed Methodology

As described in the previous section, the model with PointNet architecture uses the 3D point cloud data of Indian classical dance forms and classifies into five labels, namely Bharatanatyam, Odissi, Kathak, Kathakali, and Yakshagana. Existing PointNet architecture is tested on ModelNet dataset. This dataset holds Object File Format (OFF) format files which basically carries 3D point cloud data of different categories. Since ICD point cloud data is not available, it has to be generated following certain procedure. Images are extracted from the internet. Point cloud dataset has to be generated by using 3D builder and MeshLab software. 3D builder is used to change the 2D image into 3D Polygon File Format (PLY) format. 3D data is required in the OFF Object File Format (OFF) format, and PLY data has to be converted into OFF format using MeshLab. OFF data of all the five dance forms are fed to the PointNet architecture to classify them into different dance labels. As the 3D point cloud data of each dance form will have large points, it is better to downsize the point data to lesser points to increase the performance.

In this work, each dance point cloud data is fixed to 2048 points. PointNet architecture is used in the proposed work to train and classify 3D point cloud data of Indian classical dance (ICD) images.

The Indian classical dance point cloud dataset used in this proposed model has five dance forms each consisting of 40 point cloud OFF files, and each folder is split into train and test data folder in 80:20 ratio. Here, training dataset contains a 30 OFF files and testing dataset contains 10 OFF files. Figure 1 shows few samples from the ICD point cloud data in object file format (OFF).

PointNet architecture is fed with ICD point cloud dataset. Workflow of the proposed model is presented in Figs. 2. and 3.

As described earlier, 3D point cloud data with x-, y-, z-coordinates (where z is the depth information) is fed to the PointNet model. Basically, PointNet consists of multilayer perceptron (MLP), max pooling layer and fully connected layers with ReLu and batch normalization. Figure 4 shows the basic architecture of PointNet.

Architecture resolves two challenges. First one is, unordered pointset is considered as input which means model must be invariant to the $N!$ Permutations. N orderless

Table 1 Literature summary

Papers	Methods used	Accuracy achieved	Observations
Reference [4]	DCNN, Optical Flow, SVM classifier	Dataset: ICD Video frames Bharathanatyam-65.78%, Manipuri-71.42%	Suitable for 2D image data and Accuracy is less
Reference [5]	CNN	Dataset: ICD offline and online Images 89.92%	Suitable for 2D image data
Reference [6]	pose descriptor based on histogram of oriented optical flow (HOOF), SVM classifier	Dataset: ICD 86.67%	Suitable for 2D image data
Reference [7]	Examine a standard manifold charting and some alternative chartings on special manifolds, particularly, the special orthogonal group, Stiefel manifolds, and Grassmann manifolds	Dataset: Cambridge gesture, the UMD Keck body gesture, and the UCF sport datasets	–
Reference [8]	Proposed a unified framework allowing to generalize CNN architectures to non-Euclidean domains (graphs and manifolds)	Dataset: Cora,PubMed MoNet-81.69% for Cora, 78.81% for PubMed	–
Reference [9]	Point Cloud Segmentation using proposed Neural Network, EdgeConv used	Dataset: ShapeNet 84.1%	Point cloud data are converted to graph forms before it is fed to the network
Reference [10]	Graph convolutional neural networks, graph signal processing, 3D point cloud, Supervised Learning, pointGCN	Dataset: ModelNet40 89.51%	Point cloud data are converted to graph forms before it is fed to the network
Reference [11]	Deep CNN for 3D Human pose estimation	Dataset: Human3.6 M Walking:70.2% Jogging:79.7%	3D human pose is learned from 2D joint annotations

(continued)

Table 1 (continued)

Papers	Methods used	Accuracy achieved	Observations
Reference [12]	Dynamic Time Warping (DTW) as a template matching solution is applied to do the action classification task, Skeleton tracking	Dataset: Taiji dataset 80%	3D skeletal joint information is fed to the network. Model is not considering the surface data
Reference [13]	PointNet architecture uses 3D point cloud data for 3D Classification and Segmentation	Dataset used:ModelNet 89.2%	Unordered 3D point cloud data which has got permutation invariance and rotation invariance property is fed to the network

Fig. 1 ICD point cloud dataset

points each represented by *D*-dimensional vector are called a point cloud data. The second challenge is, the model must be invariant under geometric transformation. Hence, the input is aligned by transformer network (T-Net) which transforms the input data points. Multilayer perceptron is shared across each input point. Now, each point is embedded to 64-dimensional space. Feature transformer is done, and the points are converted to 1024 embedding space. Next step is to aggregate all the points in larger dimensional space which is done with the help of max pooling function.

4 Implementation

Below are the steps involved in implementing ICD 3D point cloud classification using PointNet architecture:

1. Download the images of five different dance forms from the Internet.

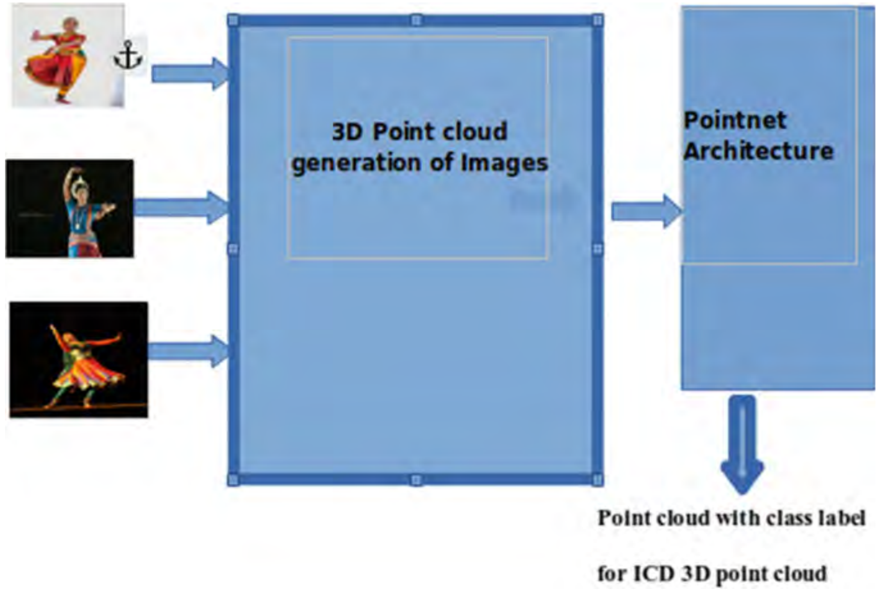


Fig. 2 PointNet architecture implementation on ICD point cloud dataset

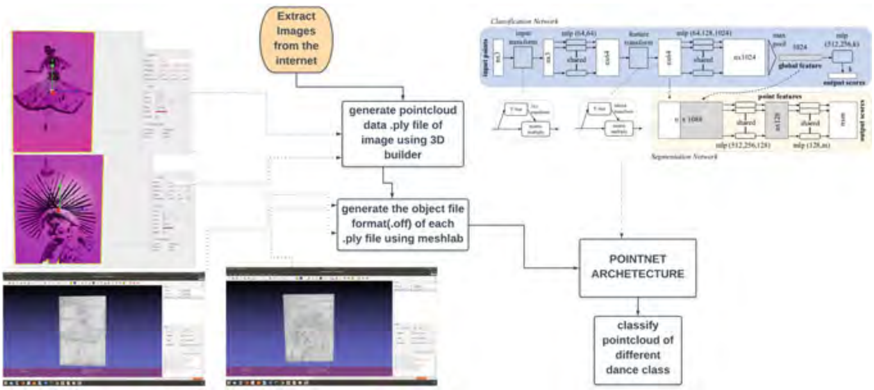


Fig. 3 3D data format conversion and classifying ICD point cloud data

2. Convert 2D image to 3D point cloud data (.ply file format); it has vertices and faces with x-, y-, z-coordinates values. 2D image can be converted to 3D ply file using 3D builder software. Data format of Polygon File Format (PLY) of ICD is shown in Figs. 5 and 6.
3. Using MeshLab, convert .ply to .off file format. Figure 7 shows the OFF file in MeshLab after conversion. Figure 8 shows the OFF file vertices and faces.

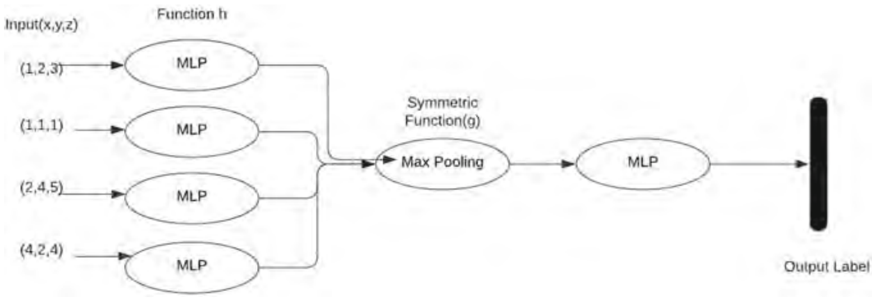


Fig. 4 Basic architecture of PointNet

Fig. 5 Sample PLY file of Yakshagana with 9012 vertices and 19,724 faces

Fig. 6 Ply file showing *x*, *y*, *z* coordinates

	x	y	z
0	0.000000	84.581993	10.159999
1	126.745995	84.581993	10.159999
2	126.745995	83.311996	10.159999
3	122.173996	82.804001	10.159999
4	0.000000	82.295998	10.159999
...	...	...	...
32121	92.455994	0.000000	0.000000
32122	98.551994	0.000000	0.000000
32123	106.679993	0.000000	0.000000
32124	114.807999	0.000000	0.000000
32125	126.491997	0.000000	0.000000

32126 rows × 3 columns

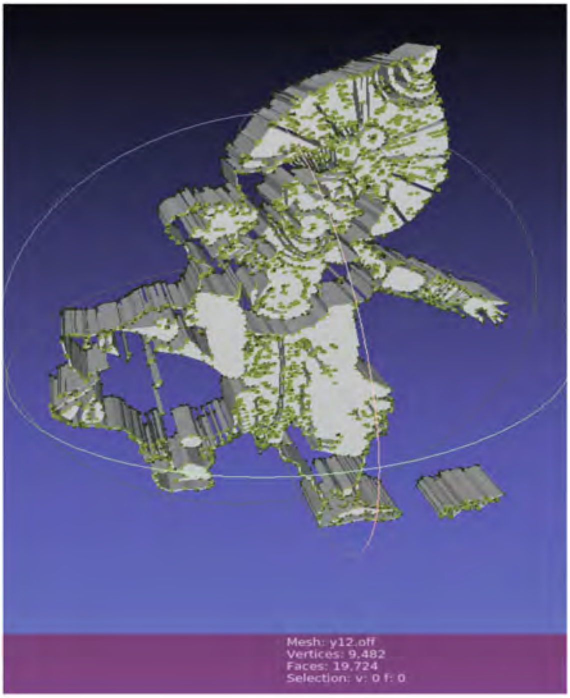


Fig. 7 Sample OFF file of Yakshagana with 9012 vertices and 19,724 faces

4. Data folder contains all the five dance classes with train and test folder containing OFF files of each dance forms.
5. Each OFF file contains large number of points. Only 2048 locations or points are to be sampled. Each point cloud data is converted to numpy array.
6. Set the number of points to sample to 2048 and batch size to 16.
7. Now feed the OFF data to the PointNet architecture.
8. Now next step is to build the model. Each convolution and fully connected layer consist of convolution, dense, batch normalization, ReLU activation. As PointNet consists of two core components. One is multilayer perceptron (MLP). Other one is transformer network for input and features transformation
9. Train the model with learning rate 0.001 and test the result. Table 2 shows the training and validation result for number of epochs.
10. Now visualize the result. Figure 9. shows the prediction result.

5 Conclusion and Future Work

The proposed work uses geometric deep learning approach and PointNet architecture to classify ICD 3D point cloud data. The proposed work achieved the training accuracy of 96.64% and validation accuracy of 73% for 250 epochs. Model can classify

Fig. 8 OFF file

```
OFF
7895 15786 0
0 58.674 13.22832
54.86399 58.674 13.22832
0 56.896 13.22832
19.05 56.642 13.22832
54.86399 56.642 13.22832
18.796 56.388 13.22832
19.05 56.388 4.40944
19.304 56.388 13.22832
18.796 55.88 13.22832
19.05 55.88 4.40944
19.304 55.88 13.22832
18.796 55.626 4.40944
19.05 55.626 13.22832
19.304 55.626 4.40944
18.288 55.372 13.22832
18.542 55.372 4.40944
19.05 55.372 4.40944
18.034 55.118 13.22832
18.288 55.118 4.40944
18.542 55.118 13.22832
18.796 55.118 4.40944
19.812 55.118 4.40944
```

Table 2 Training and validation result of ICD point cloud data

Epochs	Train_loss	Valid_loss	sparse_categorical_accuracy	val_sparse_categorical_accuracy
1	2.0038	2.0868	0.4497	0.3933
20	2.1127	88.0075	0.4295	0.2921
50	1.3189	1.8251	0.7248	0.6966
100	1.1898	1.8205	0.7517	0.5843
150	0.8927	1.2007	0.8725	0.7865
200	0.7566	2.0571	0.8993	0.7191
250	0.6963	1.7391	0.9664	0.7303

more accurately by using the surface data which is captured by 3D Light Detection and Ranging (LiDAR) laser scanners. This approach can be extended to classify all the dance forms available in India. In the present digital era, this approach will help the students and the dance gurus to provide dance education in a more precise way, even though they are not able to visit the dance center. However, the training set has to be increased with more of the dance images with varied postures. The training and classification can also be performed on the varied postures and their attires too. The

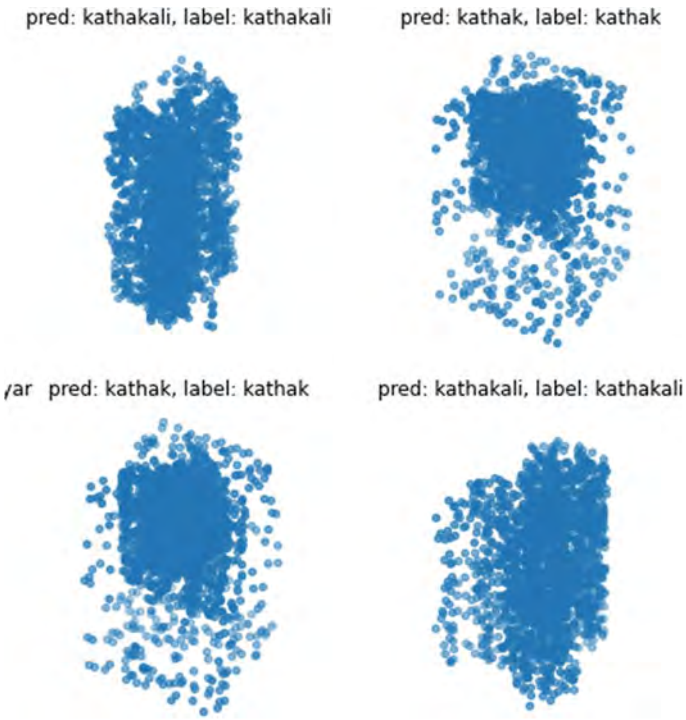


Fig. 9 Snapshot of visualized result of ICD 3D point clouds predictions and the label

model can also be fine-tuned to yield a higher validation accuracy with increasing the number of epochs or adding more images for the training and testing.

References

1. S. Pandey., M. Supriya, A. Shrivastava, Data classification using machine learning approach, in *3rd İnternational Symposium on İntelligent System Technologies and Application*, vol. 683 (2018), pp. 112–122
2. C. Sudarshana Tamuly, Jyotsna, J. Amudha,. Deep learning model for ımage classification, in *International Conference On Computational Vision and Bio Inspired Computing* (2019)
3. C. Sudarshana Tamuly, Jyotsna, J. Amudha, Effective spam image classification using CNN and transfer learning.in *3rd International Conference on Computational Vision and Bio Inspired Computing* (2019)
4. B. Ankita, B. Riya, S. Goutam, S. Pushkar, R. Balasubramanian, Indian dance form recognition from videos. in *13th International Conference on Signal-Image Technology & Internet-Based systems* (2017)
5. P.V.V. Kishore, K.V.V. Kumar, E. Kiran Kumar, AS.C.S. Sastry, M. Teja Kiran, D. Anil Kumar, M.V.D. Prasad. *Indian Classical Dance Action Identification and Classification with Convolutional Neural Networks* (Hindawi, 2018)

6. S. Soumitra, P. Pulak, C. Bhabatosh, Indian Classical Dance classification by learning dance pose bases. in *IEEE Workshop on the Applications of Computer Vision* (2012)
7. Y.M. Lui. Tangent bundles on special manifolds for action recgnition. IEEE Trans. Circu. Syst. Video Technol. **22**(6), 930–942 (2012)
8. F. Monti, D. Boscaini, J. Masci, E. Rodola, J. Svoboda, M. Bronstein, Geometric deep learning on graphs and manifolds using mixture model CNNs. in *IEEE Conference on Computer Visiona and Pattern Recognition* (2017), pp. 5115–5124
9. Y. Wang, Y. Sun, Z. Lui, S.E. Sharma, *Dynamic Graph CNN for Learning on Point Clouds.* (Researchgate, 2018)
10. Y. Zhang, M. Rabbat, A Graph-CNN for 3D point cloud classification. in *IEEE International Conference on Acoustics,Speech and Signal Processing* (2018)
11. E. Brau, J. Hao, 3D Human Pose Estimation via Deep Learning from 2D Annotations. in *IEEE fourth International Conference on 3D Vision* 16544819 (2016).
12. H.-M. Zhu, C.-M. Pun, Human action recognition with skeletal ınformation from depth camera. in *IEEE International Conference on Information and Automation* (2013)
13. C.R. Qi, H. Su, K. Mo, L.J. Guibas, Pointnet: deep learning on point sets for 3D classification and segmentation. in *IEEE Conference on Computer Vision and Pattern Recognition* (2017)
14. Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, 3D ShapeNets: a deep representation for volumetric shapes. in *IEEE Conference on Computer Vision and Pattern Recognition* (2015)
15. ModelNet Dataset, Available: https://modelnet.cs.princeton.edu/

Fire Detection by Parallel Classification of Fire and Smoke Using Convolutional Neural Network

A. Robert Singh, Suganya Athisayamani, S. Sankara Narayanan, and S. Dhanasekaran

Abstract Fire detection is considered as a part of remote surveillance in domestic, industrial and the areas that are not approachable by human like deep forests. In this paper, convolutional neural network (CNN) is used to detect fire by classifying both fire and smoke in videos. A sequence of 2D convolutional layers and max pool layers is used to convert the video frames into feature maps with lower rank. The neural network is trained with the videos containing both fire and smoke. The videos with either fire or smoke or both are tested for fire detection with the FIRESENSE and other such open-source databases. The results show that the proposed method can classify the fire, smoke and fire with smoke with a recognition rate of up to 94%, 95% and 93%, respectively.

Keywords CNN · Fire detection · Smoke detection · Classification

1 Introduction

Vision processing is an emerging area that provides lot of efficient real-time applications [1, 2]. Fire detection is considered as an open problem that is important to address the challenges associated with timely detection and control. Such an example is the recent Australian bush fire [3]. Today the deep forests are also monitored by using surveillance camera. Similarly, the industries like offshore oil rig are also vulnerable to fire accidents that can lead to severe damage for lives and assets. In these occasions, the fast and accurate fire detection will be more helpful.

A. Robert Singh · S. Dhanasekaran
School of Computing, Kalasalingam Academy of Research and Education, Anand Nagar, Srivilliputhur, Tamil Nadu, India

S. Athisayamani (✉)
School of Computing, Sastra Deemed to be University, Thanjavur, Tamil Nadu, India

S. Sankara Narayanan
Department of Computer Science and Engineering, VelTech Rangarajan Dr. Sahunthala R&D Institute of Science and Technology, Chennai, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_8

Table 1 Sample frames for the three classes

Class 1: Fire only	Class 2: Smoke only	Class 3: Fire and Smoke
		
		

There are three ways available for analysing and detecting the fire: hardware-based sensor systems, basic image processing methods and vision processing with evolutionary methods. In general, the vision processing methods for fire detection are using two ways: localizing fire and localizing smoke, where most of the fire accidents start with smoke. In this case, the fire localization methods fail to detect the fire. On the other hand, the fire due to electric and oil leakages are starting with a small amount of fire and blast to a big size. In this case, the smoke detecting methods fail to perform the task. These problems can be addressed by localizing both fire and smoke. This paper considers the important features for performing feature extraction of fire and smoke which are RGB intensity values, regional entropy, dynamicity and irregularity. In this paper, these features are used to classify the fire and smoke video frames. A parallel classification method is proposed to find the three labels like fire, smoke and normal video frames. In this paper, CNN is designed to classify three classes of videos with fire only, smoke only and fire with smoke. Some sample video frames for the three classes are shown in Table.1 taken from the FIRESENSE database [4].

2 Related Works

A good number of literature are available for fire detection using hardware sensing and image processing methods as well as evolutionary methods. The hardware sensing methods is analysing the sensor data. Chen et al. [5] proposed a fire detection method with smoke and gas sensor. This will detect the fire if the amount of smoke and the gases like CO_2 and CO crosses a threshold value. This will take more time in the case of high flame with less smoke. Tang et al. [6] have proposed another hardware-based remote alarm system to alert about fire accidents in remote area by using a GSM unit. This method fails to detect the smoke. Ono et al. [7] proposed a method to find fire in expressway tunnel. The fire flame region is automatically segmented and used for training a neural network. The method proposes an optimal location for placing the CCTV camera to cover a maximum area for fire surveillance. This method uses image processing methodologies to detect fire and depending on the result, the concerned officials are notified.

Celik et al. [8] proposed a color image processing method to segment the fire areas in the image. This method used YCbCr color model to detect the fire pixels in the video. The Cb and Cr components are used to filter the pixels using thresholding. The method recorded 31% false alarm which is a high rate for emergency remote surveillance. Marbach et al. [9] proposed an image processing based fire detection method in video frames. In this paper, the pixels in the video frames are divided into three groups: active pixels, saturated pixels and fire pixels. These pixels are used to find the fire pattern using luminance and chrominance sensitivity. Dhanujalakshmi et al. [10] proposed a remote fire detection method using Raspberry Pi by image processing method. This method uses a local thresholding to localize the fire pixels. The segmentation of fire in the video frame is proposed for notifying through WhatsApp.

Li et al. [11] proposed a CNN-based fire detection on images. In this paper, the CNN architectures like faster-RCNN, R-FCN, single short multibox detector (SSD) and YoLo v3 are proposed for fire detection. Except SSD, the other architectures can detect either fire or smoke.

All the above discussed methods are well designed to detect either smoke or fire. In this paper, the problem of detecting both fire and smoke is addressed by having three classes in CNN-based classification that can simultaneously detect the smoke and fire.

3 Proposed CNN for Fire and Smoke Detection

The preprocessing consists of segmentation of background and foreground in the video frames. The robust orthonormal subspace learning method [12] is used for background subtraction. Let V be the 4D matrix that denotes the video frames, B be the background matrix and F be the foreground matrix that are relatively lower-order than V. The foreground and background images are calculated with the objective of minimal coefficient factor as given in Eq. (1).

$$\min_{S,x,M} x_{r-1} + \alpha M_1 \tag{1}$$

subject to,

$$S.x + M = F \quad \text{and} \quad B = S.x$$

where, $S = \{S_1, S_2, \ldots S_k\} \in \mathbb{R}$ is the orthonormal subspace, $x = \{x_1, x_2, \ldots x_k\} \in \mathbb{R}$, k is the dimension of subspace. This is the first step in the training and testing

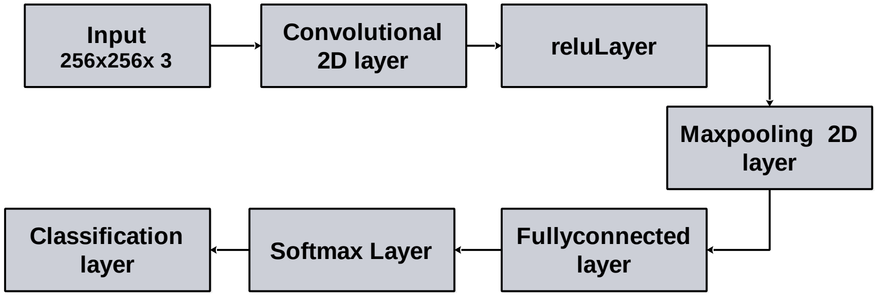


Fig. 1 Architecture of proposed CNN

processes. The proposed architecture has seven different layers as shown in Fig. 1. The input layer resizes the incoming frame into 256 × 256 × 3 sized frame.

The convolutional layer is obtained by mapping two pixels from the input image into one pixel. This pixel is obtained by weighted average of every two neighbor pixels. This function is given in Eq. (2).

$$f(x, y) = w \times \frac{I(x, y) + I(x, y + 1)}{2} \tag{2}$$

where $f(x, y)$ is the pixel in the convolutional layer, w is the weight and $I(x, y)$ is the pixel in input layer in gray scale. Thus, the output of convolutional layer will be of size 128 × 128 pixels. The dimensions of the sample in each convolutional layer are shown in Fig. 2.

Feature maps are calculated by mapping the dynamic features through temporal direction. The feature maps identified in the first 2D convolutional layer are shown in Fig. 3. These features are obtained with different weight values ranging between [0, 1]. The same weight values are applied in further convolutional layers also.

ReLU layer is the linear rectifier layer, which is a part of convolutional layer. This enhances the feature map generated by the convolutional layer. In this paper, white balance in the greyscale image is applied. This assigns positive value to the white components and negative value to the black components. Thus, the fire features and smoke features are highlighted. The output of ReLU layer with the first 2D convolutional layer is shown in Fig. 4.

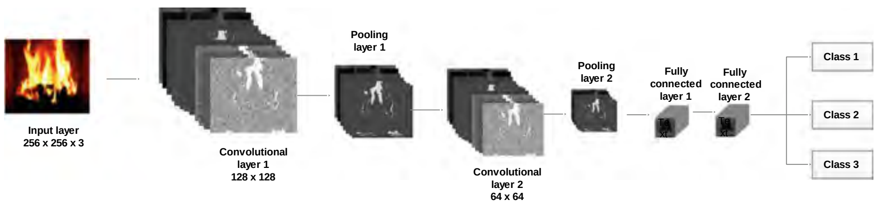


Fig. 2 Conversion of the input image through convolutional layers and max pool layers

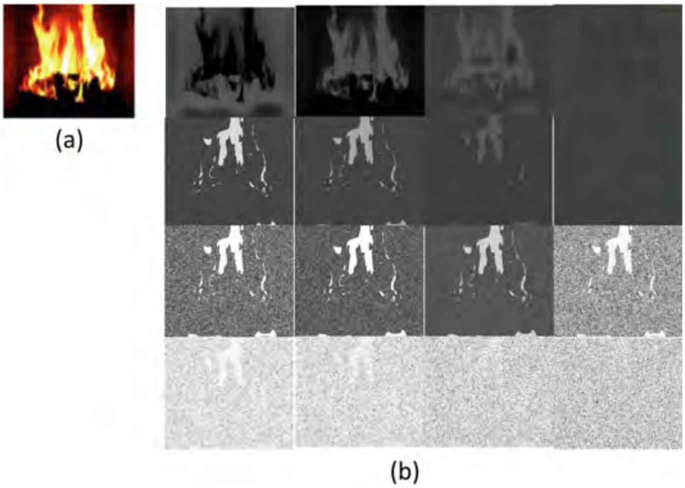


Fig. 3 **a** Input video frame, **b** feature maps in convolutional layer 1

Fig. 4 **a** Original video frame **b** One of the output feature map of ReLU layer

Each convolution layer is followed by a max pooling layer. Here, each frame is divided into number of sub-blocks. The maximum of each sub-block is used to replace it as a single value as given in Eq. (3). For a block of size $n \times n$, the max pooled value is the maximum convoluted pixel value (g).

$$(x, y) = \max_{\substack{1 \le i \le n \\ 1 \le j \le n}} g(i, j) \tag{3}$$

Thus, the set of convolutional layer, ReLU layer and max pool layer is repeatedly applied until a 1-D feature map is obtained. The final layer is the fully connected layer that applies flattening on the matrix to convert into 1-D vector. The order of the output of the fully connected layer is equal to the columns equal to the number of columns obtained at the final max pooling layer.

4 Training and Testing

The proposed architecture is further referred as smokiFi-CNN throughout the experiment. Three different datasets [4, 12, 13] are used for training and testing the smokiFi-CNN. The confusion matrix obtained for the classification is given in Fig. 5.

In overall, 80% of the images in all the three datasets were used for training and 20% of the images were used to test the CNN. The logarithmic scale of learning rate is compared with the loss, and the result shows that the local maxima is the point where the CNN learning starts and the global minima is the point when the CNN start to over fit as shown in Fig. 6. Here, the local maxima is 0.91, and global minima is 0.21.

Number of epochs is an important factor for training and testing that represents the total number of times the dataset is trained forward and backward. The accuracy and loss obtained in both training and testing (validating) are compared against different number of epochs.

Fig. 5 Confusion matrix for smokiFi-CNN

		Truth data		
		Class 1	Class 2	Class 3
Classifier results	Class 1	40	2	5
	Class 2	3	36	3
	Class 3	2	2	39

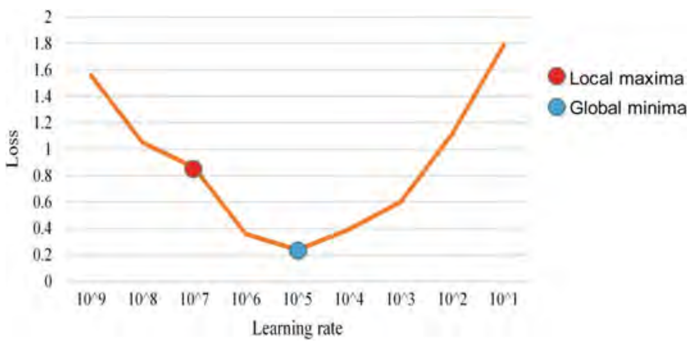


Fig. 6 Comparison of learning rate versus loss

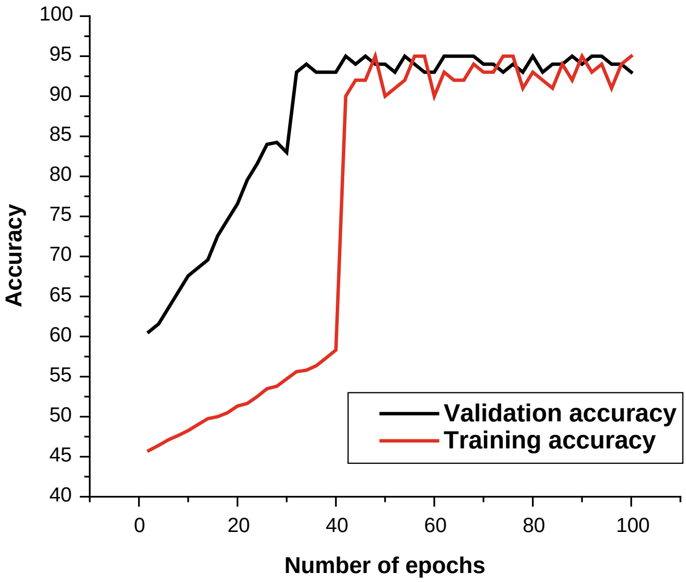


Fig. 7 Comparison of number of epochs and accuracy

Figure 7 shows the comparison of number of epochs and accuracy. The training starts with no previous knowledge, and so, the training accuracy for 2 epochs is 45 where it improves along with the increase in number of epochs. The accuracy stabilizes within a short range after the learning is reaching the converging stage. The validation accuracy starts from a higher level say 60 and increments along with the number of epochs.

Figure 8 compares the number of epochs and the loss. The value of loss is inversely propositional to the accuracy. Hence, the loss is reducing when the number of epochs is increasing.

5 Result Analysis and Discussion

The proposed architecture is implemented using MATLAB 2020, and the combined datasets [4, 12] and [13] are used for training and testing. The video frames from 59 videos of the three classes, namely fire only, smoke only and fire with smoke, are labeled as 1, 2 and 3, respectively. The proposed method is compared with the state-of-the-art methods as shown in Table 2. The result shows that the proposed method is giving accuracy more than majority of the state-of-the-art methods for classifying fire. The two methods [14, 15] have higher accuracy than the proposed method. They classify only two categories say, fire and normal. The occurrence of smoke and combination of smoke and fire is not considered in these two methods.

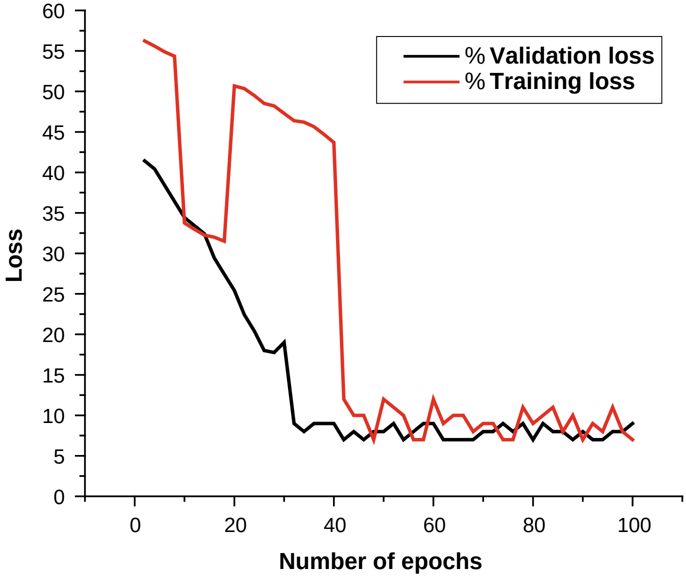


Fig. 8 Comparison of number of epochs and loss

Table 2 Comparison of classification performance with state-of-the-art methods

Fire and smoke detection method	False positive (%)	False negative (%)	Accuracy (%)
Deep CNN [14]	8.87	2.12	94.5
CE + MV + SV [16]	11.67	0	93.55
De Lascio et al. [17]	13.33	0	92.86
Habibugle et al. [18]	5.88	14.29	90.32
Rafiee et al. (YUV color) [19]	17.65	7.14	74.20
Celik et al. [20]	29.41	0	83.87
Chen et al. [21]	11.76	14.29	87.1
Arpit Jadon et al. [15]	1.23	2.25	96.52
Khan Muhammad et al. [14]	0	0.14	95.86
SmokiFi-CNN	1.1	2.1	94.74

Figure 9 is the visual representation for the output of classification for the video frames from the FIRESENSE database. Each output shows that there are classification values for three classes. From the results, it is evident that the percentage of classification for the fire, smoke and combination of both is more appropriate for the contents of the frame. Other than accuracy, the important metrics for classification evaluation are Precision, Recall and F1 Score. They are calculated as given in Eq. (4) through (6).

Fig. 9 Result of classification using CNN

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}} \tag{4}$$

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}} \tag{5}$$

Table 3 Comparison of training methods with the state-of-the-art methods

Metrics	GLNRGB [22]	ALEXRGB [22]	VGGNET CNN [23]	SmokiFi CNN
False positive	0.54	0.68	1.50	1.56
False negative	0.99	1.01	0.99	0.96
True positive	0.61	0.72	0.92	0.89
Precision	0.53	0.51	0.38	0.36
Recall	0.38	0.42	0.48	0.48
$F1$-Score	0.45	0.46	0.42	0.41

$$F1 \text{ score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

These metrics are analysed for the proposed method against the state-of-the-art methods as shown in Table 3. The results show that the proposed method achieves minimum precision and maximum recall values that lead to a minimum F1 score value than the existing methods.

6 Conclusion

This paper proposes a CNN architecture to classify the videos into three classes say fire only, smoke only and fire with smoke. A reasonable large number of videos are used for training the architecture. The testing results show that the proposed architecture can classify the fire and smoke images with maximum accuracy and minimum $F1$-score. In future, the architecture will be improved by adding more number of layers to achieve a better accuracy in classifying the videos with both fire and smoke.

References

1. A.R. Singh, A. Suganya, Efficient tool for face detection and face recognition in color group photos, in *3rd International Conference on Electronics Computer Technology* (Kanyakumari, 2011), pp. 263–265
2. A. Robertsingh G. Sathana, S. Sathya Sheela, Remote theft identification using raspberry Pi system based on motion detection. SSRG Int. J. Comput. Sci. Eng. (SSRG-IJCSE), **4**(4), 21–23 (2017)
3. E. Nick, B. Andy, Z. Naaman, *How Big are the Fires Burning in Australia? (Interactive map)* (Guardian Australia, 2020)
4. N. Grammalidis, K. Dimitropoulos, E. Cetin, *FIRESENSE Database of Videos for Flame and Smoke Detection* (Zenodo, 2017)
5. S.-J. Chen, D.C. Hovde, K.A. Peterson, A.W. Marshall, Fire detection using smoke and gas sensors. Fire Saf. J. **42**(8), 507–515 (2007)

6. Z. Tang, W. Shuai, L. Jun, Remote alarm monitor system based on GSM and ARM. Procedia Eng. **15**, 65–69 (2011)
7. T. Ono, H. Ishii, K. Kawamura et al., Application of neural network to analyses of CCD colour TV-camera image for the detection of car fires in expressway tunnels. Fire Saf. J. **41**(4), 279–284 (2006)
8. T. Celik, K.-K. Ma, Computer vision based fire detection in color images, in *Proceedings of the IEEE Conference on Soft Computing on Industrial Applications (SMCia '08)* (2008), pp. 258–263
9. G. Marbach, M. Loepfe, T. Brupbacher, An image processing technique for fire detection in video images. Fire Saf. J. **41**(4), 285–289 (2006)
10. R. Dhanujalakshmi, B. Divya, C. Divya@sandhiya, A. Robertsingh, Image processing based fire detection system using rasperry Pi system. SSRG Int. J. Comput. Sci. Eng. 4(4), 18–20 (2017)
11. P. Li, W. Zhao, Image fire detection algorithms based on convolutional neural networks. Case Stud. Thermal Eng. **19** (2020).
12. X. Shu, F. Porikli, N. Ahuja, Robust orthonormal subspace learning: efficient recovery of corrupted low-rank matrices, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 3874–3881
13. D.Y. Chino, L.P. Avalhais, J.F. Rodrigues, A.J. Traina, BoWFire: detection of fire in still images by integrating pixel color and texture analysis, in *28th SIBGRAPI Conference on Graphics, Patterns and Images* (2015), pp. 95–102
14. K. Muhammad, J. Ahmad, Z. Lv, P. Bellavista, P. Yang, S.W. Baik, Efficient deep cnn-based fire detection and localization in video surveillance applications. IEEE Trans. Syst. Man Cybern. Syst. **99**, 1–16 (2019)
15. A. Jadon, M. Omama, A. Varshney, M.S. Ansari, R. Sharma, FireNet: *A Specialized Lightweight Fire & Smoke Detection Model for Real-Time IoT Applications.* arXiv, arXiv:1905.11922 (2019).
16. P. Foggia, A. Saggese, M. Vento, Real-time fire detection for video-surveillance applications using a combination of experts based on color. Shape Motion IEEE Tran. Circuits Syst. Video Technol. **25**(9), 1545–1556 (2015)
17. R. Di Lascio, A. Greco, A. Saggese, M. Vento, Improving fire detection reliability by a combination of video analytics, in *Proceedings of the International Conference Image Analysis and Recognition, Vilamoura* (Springer, Cham, Switzerland, 2014)
18. Y.H. Habiboǧlu, O. Günay, A.E. Çetin, Covariance matrix-based fire and flame detection method in video. Mach. Vis. Appl. **23**, 1103–1113 (2012)
19. A. Rafiee, R. Dianat, M. Jamshidi, R. Tavakoli, S. Abbaspour, Fide and smoke detection using wavelet analysis and disorder characteristics, in *Proceedings of the 2011 3rd International Conference on Computer Research and Development (ICCRD)* (2011), pp. 262–265
20. T. Celik, H. Demirel, H. Ozkaramanli, M. Uyguroglu, Fire detection using statistical color model in video sequences. J. Vis. Commun. Image Represent. **18**, 176–185 (2007)
21. T.H. Chen, P.H. Wu, Y.C. Chiou, An early fire-detection method based on image processing, in *Proceedings of the International Conference on Image Processing (ICIP)* (2014), pp. 1707–1710
22. A. Leibetseder, M.J. Primus, S. Petscharnig, K. Schoeffmann, Real-time image-based smoke detection in endoscopic videos, in *Proceedings of the on Thematic Workshops of ACM Multimedia* (2017), pp. 296–304
23. P. Matlani, M. Shrivastava, Hybrid Deep VGG-NET convolutional classifier for video smoke detection. Comput. Model. Eng. Sci. **119**, 427–458 (2019)

A Split Key Unique Sudoku Steganography (SKUSS)-Based Reversible High Embedded Data Hiding Technique

Utsav Kumar Malviya and Vivek Singh Rathore

Abstract Confidential data transfer or secure communication is always remaining as an essential need in this era of modern communication. This research work is an implementation of a new method named Split Key Unique Sudoku Steganography (SKUSS). It is a reversible data hiding with high embedding capacity, where high embedding signifies data bits per pixels of the cover image that are more with maintaining the naturalness of the cover image. For the implementation of SKUSS first, a unique Sudoku has been generated with a user-defined 64-bit key and a 64-bit public key, then that Sudoku is used for hiding data digits (ASCII code) into cover image pixels. The data can be only recovered with that same Sudoku with the correct public and user keys. This method is based on a hybrid algorithm that applies the techniques of Sudoku and Steganography to offer different security features to the images transmitted between entities on the Internet. Based on the proposed method, the authenticity and integrity of the transmitted images can be verified either in the spatial domain or in the encrypted domain or both domains. The work is implemented on MATLAB design and simulation tools.

Keywords Public–private key · Sudoku · Steganography · Bitmap image file · Split Key Unique Sudoku Steganography (SKUSS) · Bit per pixels

1 Introduction

Applications where data security is a big concern than bandwidth, steganography can be used over encryption. Steganography is a data hiding technique for secure data communication; Steganography is less suspicious than encryption. There are certain issues with steganography like network monitoring systems that will not mark the steganography files, which have secrete data in it. Hence, if someone tries to steal secrete data, they cover unmark files with different files and can send it simple email [1]. Another problem is that lots of data have to be transmitted which arises

U. K. Malviya (✉) · V. S. Rathore
Government Engineering College, Bilaspur, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_10

suspiciousness to intruders. Also, the same method cannot be used for all types of cover image [2]. In steganography, the time for hiding data should be low enough so it does not disturb the communication [3]. The issue is to maintain the balance between robustness, imperceptibility, and extent as increasing one-factor reverse affects others. If payload size increases, it bargains with the imperceptibility [4]. So, the major problem is to maintain balance among parameters.

Sudoku-based data hiding resolves the issues up to certain levels define in [1–4]. Steganography method based on Sudoku was developed in 2008 [5]. Chang's method significantly improves embedding capacity by using Sudoku. Because Sudoku can have a large number of possible solutions in comparison with other steganography methods like LSB hiding. In 2015, Nguyen [1] presented a Sudoku-based reversible data hiding method, reversible signifies that at the receiver end, both the secrete data and cover image get recovered. Nguyen's [1] uses the Sudoku scheme for data hiding which allows embedding more data bits also a good quality of cover image at the receiver side. Recovery of stego image was the advantage of Nguyen's method [1] work in comparison with Chang's [4] method. Later, in 2017, for obscure, security Sudoku-based dual cover hiding technique [3] presented. Jan et al. [5] and Chang et al. [4] methods use a single cover image where the cover image is divided into two areas: first embedding area in which secrete data is hidden and the second non-embedding area which has a location map. Sarvan's method [6] resolves the issue of location map in the same image and uses a separate image for the location map. The use of dual image allows very high security; however, the bandwidth requirements also get double. Later, in 2018, Cheng [2] develop a method for data hiding in an image using Sudoku and use division arithmetic and generalized exploiting modification direction (DA-GEMD) for embedding the data into the cover image where pixels are selected using Sudoku. Chang's [5] and Nguyen's [7] methods use Sudoku and interpolation of the cover image used for location map and hide the data bits into pixels newly generated using interpolation, the issue with the interpolation technique to maintain the naturalness of the original image hence DA-GEMD [2, 8] based location map is better in keeping naturalness of cover image, Steganography using Sudoku and DA-GEMD are not reversible.

This work defines a new method SKUSS for data hiding using Sudoku-based encoding to choose the digit replacement of cover image with message data, which resolve the general steganography issues as was define in [1–4]. The data hiding method by [5, 7, 6] and [9] was not reversible and proposed SKUSS has an advantage as it is reversible data hiding method. The proposed SKUSS uses a single image for data hiding; hence, it requires less bandwidth then [6, 10]. Nguyen's [1] data hiding method use Sudoku as a key which makes a key itself big data (648-bit key), proposed SKUSS use a 64 bit and generate unsolved Sudoku with the split key method then solve the Sudoku with Pramberg's [11] method of Sudoku solver, This technique of the proposed work reduces the Key size and payload. Because only 64 bits key same cover image can be used for data and location map with high naturalness of cover image. This work uses Jorgen Pramberg [11] method for solving Sudoku; Jorgen [11] states that the generated Sudoku puzzle should only have one solution. Table 1 below shows the literature summary with the outcome of a few researcher's work.

Table 1 Literature summary

Authors	year	Method	Outcomes
Chang et al. [4]	2008	Data hiding in cover Sudoku solution is used to guide cover pixels	1.5 bpp hidden in test Image obtain PSNR of 44.5
Kumar [3]	2017	Dual cover image-based data hiding scheme-based where Sudoku used as pixel finder	0.06688 bpp hidden in test Image obtain PSNR of 59.27
Cheng et al. [2]	2017	Division arithmetic and generalized exploiting modification direction (DA-GEMD) used for data hiding and Sudoku used for finding pixel to be changed	1.28 bpp hidden in Lena Image obtain PSNR of 45.41
Nguyen et al. [1]	2015	Develop a new reversible data hiding scheme using Sudoku-based encoding and LSB substitution	1.2 bpp hidden in Lena Image obtain PSNR of 41.03

2 Methodology

The concept of Sudoku-based data hiding is that not to make any changes in the original cover image pixels even in the cover image only encoded according to the secrete data. Encoding of the cover image instead of direct data bits hiding gives high robustness, imperceptibility, and extent.

Figure 1 explains the process flow of the data hiding process using the proposed SKUSS method. This work has four stages: first, create an unsolved unique Sudoku with given 64-bit key using proposed split key method; second, solve the key-based newly developed Sudoku with Jorgen's Sudoku solver [11]; third, interpolate the cover image segments of size 3×3 into the size of 6×6, and last hide the message data bits into an interpolated cover image with decoding method of solved Sudoku. Proposed Split Key Unique Sudoku Steganography (SKUSS) uses two split keys: one 8 BCD digit (32 bit) public Key [12] provided to the authorized user and one 8 BCD digit (32 bit) private key for an authorized service provider, with a combination of both the keys develop the final 16 BCD digit key (64 bit). The final 16 digit Sudoku key is used to develop unsolved Sudoku. This method of developing unsolved Sudoku using split public and private keys makes the proposed method robust against intrusion.

A key-based Sudoku is used for locating the data position and the interpolated pixel of cover modified accordingly. Because the interpolated pixels are slightly changing as per data digits [13], and data digits are not changing the original cover image pixels, the naturalness of cover image will be sustained with high PSNR at receiver and also the cover image can be extract at the receiver side and the reversible steganography will be achieved.

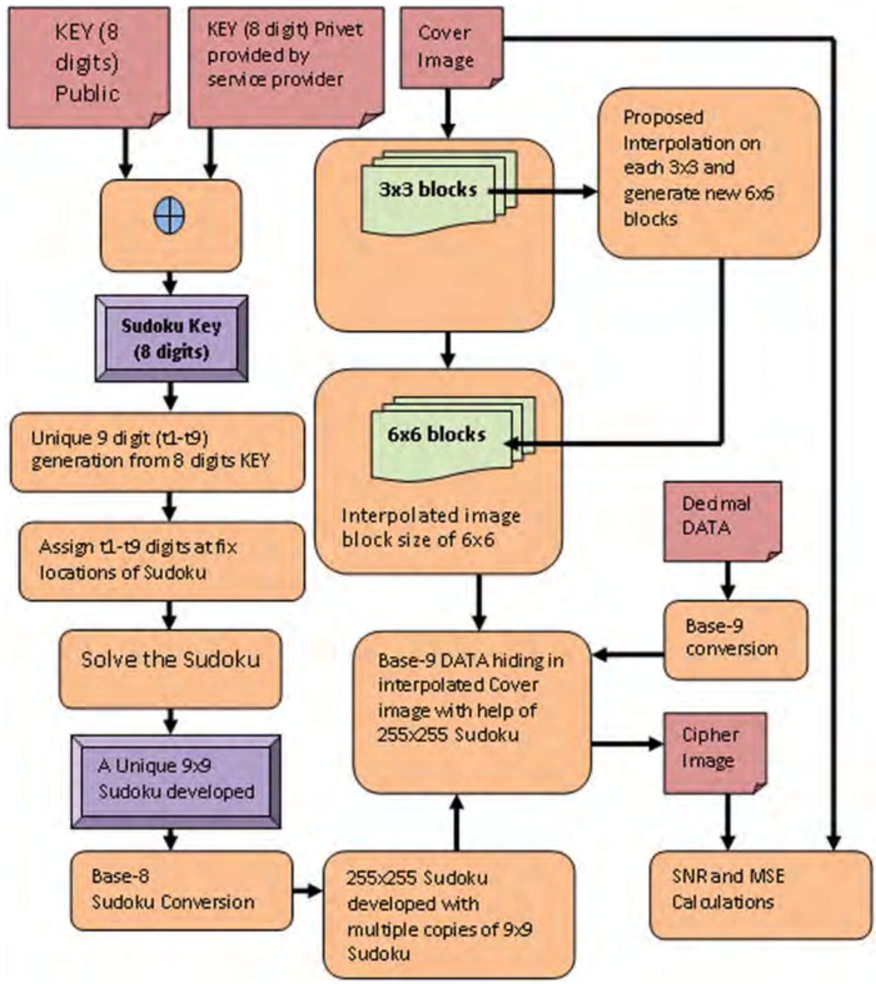


Fig. 1 Process flow of SKUSS: data embedding

2.1 Algorithm

Step1 : At transmitter side, 8 digit public key (Eqs. 1 and 8) digit private key (Eq. 2) combines using XOR logic operation and produce 8 digit Sudoku key (Eq. 3) which further converted into base 9 (Eq. 4). Base 9 conversion required because Sudoku can have 9 digits only. With the 8 digit Sudoku key, nine different BCD numbers (t_1–t_9) were developed using process shown in Eq. 6. Notice that the numbers t_1–t_9 are not same and are random. Using fixed Sudoku format (Eq. 5) and space for t_1–t_9, the unsolved Sudoku is created which is unique because it was developed with the help of public and private keys. It may be noticed that for different keys, the different unsolved Sudoku's will be created.

$$\text{Key} = K_1\ K_2\ K_3\ K_4\ K_5\ K\ K_6\ K_7\ K_8 \tag{1}$$

$$\text{NK} = K_2\ K_3\ K_4\ K_5\ K\ K_6\ K_7\ K_8\ K_1 \tag{2}$$

where K_1,$K_{2,}$ … K_8 are the eight digits of the public key. Logical XOR operation between NK with private key (PK) generates the main Key MK

$$\text{MK} = \text{NK} \oplus \text{PK} \tag{3}$$

$$\text{MK}_r = \{(10 - r) - \text{MK}\} \tag{4}$$

Sud is a 9 × 9 matrix of BCD digits where the position of t_1,t_2… t_9 are fixed and unknown.

$$\text{Sud} = \begin{matrix} t_1 & \text{UN} & \text{UN} & \text{UN} & \text{UN} & t_2 & \text{UN} & t_3 & \text{UN} \\ \text{UN} & t_4 & \text{UN} & \text{UN} & t_5 & \text{UN} & \text{UN} & \text{UN} & t_6 \\ \text{UN} & \text{UN} & t_7 & t_8 & \text{UN} & \text{UN} & t_9 & \text{UN} & \text{UN} \\ \text{UN} & \text{UN} & t_1 & t_2 & \text{UN} & \text{UN} & t_3 & \text{UN} & \text{UN} \\ \text{UN} & t_5 & \text{UN} & \text{UN} & t_4 & \text{UN} & \text{UN} & \text{UN} & t_7 \\ t_6 & \text{UN} & \text{UN} & \text{UN} & \text{UN} & t_8 & \text{UN} & \text{UN} & \text{UN} \\ t_9 & \text{UN} & \text{UN} & \text{UN} & \text{UN} & \text{UN} & \text{UN} & t_1 & \text{UN} \\ \text{UN} & t_2 & \text{UN} & \text{UN} & \text{UN} & \text{UN} & \text{UN} & \text{UN} & t_3 \\ \text{UN} & \text{UN} & t_4 & \text{UN} & \text{UN} & t_5 & \text{UN} & \text{UN} & \text{UN} \end{matrix} \tag{5}$$

Process below is used to find the unknown variable of Eq. (5) with the value of MK from Eq. (4)

$$\begin{array}{l} \text{For i} \\ = 1{:}\,23 \\ \quad \text{For j} = 1{:}\,8 \\ \qquad \text{if } \left(\text{MK(j)}_{\text{dp}} \neq (\text{t}_{1:\text{i}-1}, \text{t}_{1:\text{i}-2}, \text{t}_{1:\text{i}-3} \ldots\ldots \text{t}_1 \right) \\ \qquad\qquad \text{t}_\text{i} = \text{MK(j)}_{\text{dp}} \\ \qquad \text{elseif } \left(\text{MK(j)}_{\text{d(p+1)}} \neq (\text{t}_{1:\text{i}-1}, \text{t}_{1:\text{i}-2}, \text{t}_{1:\text{i}-3} \ldots\ldots \text{t}_1 \right) \\ \qquad\qquad \text{t}_\text{i} = \text{MK(j)}_{\text{d(p+1)}} \\ \qquad \text{elseif } \left(\text{MK(j)}_{\text{d(p+2)}} \neq (\text{t}_{1:\text{i}-1}, \text{t}_{1:\text{i}-2}, \text{t}_{1:\text{i}-3} \ldots\ldots \text{t}_1 \right) \\ \qquad\qquad \text{t}_\text{i} = \text{MK(j)}_{\text{d(p+2)}} \\ \qquad\qquad \vdots \\ \qquad\qquad \vdots \\ \qquad\qquad \text{else} \\ \qquad\qquad \text{t}_\text{i} = \text{MK(j)}_{\text{d(p+7)}} \\ \qquad \text{end} \\ \text{end} \end{array} \tag{6}$$

Step 2: The unknown BCD digits in Sudoku matrix ‘sud’ of Eq. (5) is solved using Jorgen’s Sudoku solver. The solved Sudoku is represented with variable ‘Sud_1’ below in Eq. 7. This solved Sudoku is derived by the 8 digits public and 8 private key. Different keys will always produce new unique solved Sudoku. This Sudoku also has good avalanche means one digit input key changes cause changes of 75–90% digits of developing Sudoku. It behaves like a chaotic system. Convert Sudoku (Eq. 7) into base 9 Sudoku (Eq. 8). Next creates multiple copies of solved base 9 Sudoku (9 × 9) so the final matrix has 255 × 255 digits shown in Fig. 2.

$$\mathrm{Sud}_1 = \begin{matrix} t_1 & U_1 & U_2 & U_3 & U_4 & t_2 & U_5 & t_3 & U_6 \\ U_7 & t_4 & U_8 & U_9 & t_5 & U_{10} & U_{11} & U_{12} & t_6 \\ U_{13} & U_{14} & t_7 & t_8 & U_{15} & U_{16} & t_9 & U_{17} & U_{18} \\ U_{19} & U_{20} & t_{10} & t_{11} & U_{21} & U_{22} & t_{12} & U_{23} & U_{24} \\ U_{25} & t_{13} & U_{26} & U_{27} & t_{14} & U_{28} & U_{29} & U_{30} & t_{15} \\ t_{16} & U_{31} & U_{32} & U_{33} & U_{34} & t_{17} & U_{35} & U_{36} & U_{37} \\ t_{18} & U_{38} & U_{39} & U_{40} & U_{41} & U_{42} & U_{43} & t_{19} & U_{44} \\ U_{45} & t_{20} & U_{46} & U_{47} & U_{48} & U_{49} & U_{50} & U_{51} & t_{21} \\ U_{52} & U_{53} & t_{22} & U_{54} & U_{55} & U_{56} & t_{23} & U_{57} & U_{58} \end{matrix} \tag{7}$$

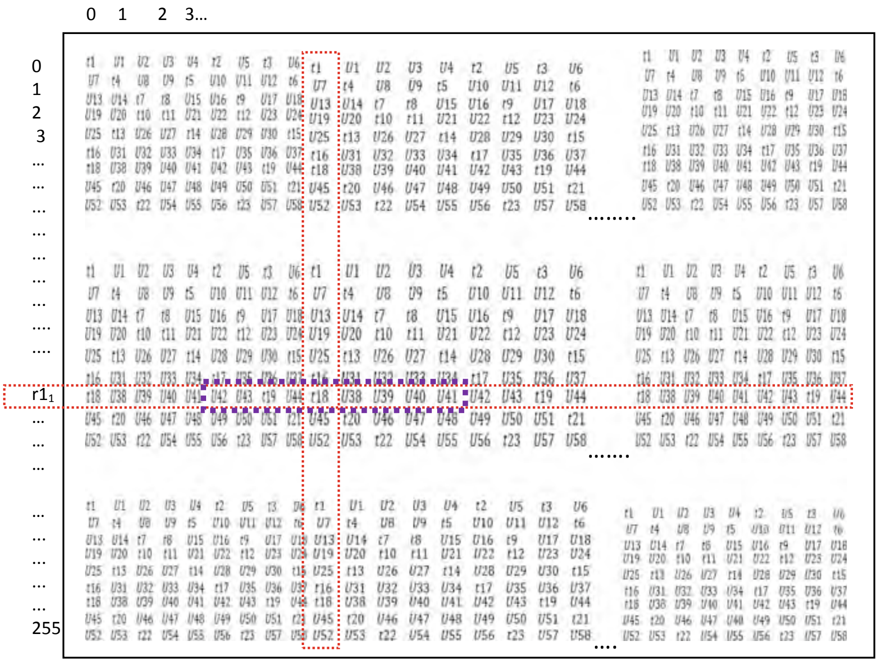


Fig. 2. 255 × 255 unique Sudoku (SubM)

Convert Sudoku (Eq. 7) into base 9 Sudoku by subtraction each element of Sud_1 (Eq. 7) by one. Base 9 conversion is required because Sudoku can have only 9 digits.

$$\mathrm{Sud\,M} = \mathrm{Sud}_1 - 1 \tag{8}$$

Step 3: Isolate 3 × 3 segments of cover image (Eq. 9). Perform interpolation using the formula (Eqs. 10, 11 and 12) and convert cover image segment 3 × 3 into 6 × 6 segment and to maintain naturalness of cover image, four pixels of cover image are used to develop new interpolated cover image pixel. 6 × 6 interpolated segment of cover image is shown in Eq. 13.

Let 'img' is one 3 × 3 module of cover image

$$img = \begin{matrix} r1_1 & r1_2 & r1_3 \\ r2_1 & r2_2 & r2_3 \\ r3_1 & r3_2 & r3_3 \end{matrix} \tag{9}$$

$$rx_{ni} = \left[rx_i + \left(\frac{rx_i + rx_{i+1}}{2} \right) \right] \Big/ 2 \tag{10}$$

$$cx_{ni} = \left[ri_x + \left(\frac{ri_x + r(i+1)_x}{2} \right) \right] \Big/ 2 \tag{11}$$

$$rcx_{ni} = \left\{ \left[rx_{ni} + \left(\frac{rx_{ni} + rx_{n(i+1)}}{2} \right) \right] \Big/ 2 + \left[cx_{ni} + \left(\frac{cx_{ni} + cx_{n(i+1)}}{2} \right) \right] \Big/ 2 \right\} \Big/ 2 \tag{12}$$

where x is constant

$$\mathrm{img1} = \begin{matrix} r1_1 & r1_{n1} & r1_2 & r1_{n2} & r1_3 & r1_{n3} \\ c1_{n1} & rc1_{n1} & c1_{n2} & rc1_{n2} & c1_{n3} & rc1_{n3} \\ r2_1 & r2_{n1} & r2_2 & r2_{n2} & r2_3 & r2_{n3} \\ c2_{n1} & rc2_{n1} & c2_{n2} & rc2_{n2} & c2_{n3} & rc2_{n3} \\ r3_1 & r3_{n1} & r3_2 & r3_{n2} & r3_3 & r3_{n3} \\ c3_{n1} & rc3_{n1} & c3_{n2} & rc3_{n2} & c3_{n3} & rc3_{n3} \end{matrix} \tag{13}$$

img1 in Eq. (13) is the 6 × 6 block of interpolated of the cover image.

Step 4: This step explains Sudoku-based hiding of data digits into cover image. The unique public-private key-based solved Sudoku multiple copies matrix 255 × 255 as explain in step-1 shown in Fig. 2. Let input ASCII message data is D_i and ND_i is base 9 converted message data where i is the digit position. From the interpolated cover image 'img1' (Eq. 13), find the values of ($r1_1$ and $r1_{n1}$) and in Fig. 2, Sudoku selects the position correspond to $(r1_1, r1_{n1})$ then selects nine values (4 forward, 4 backward, and from position $(r1_1, r1_{n1})$. For example, let $(r1_1, r1_{n1})$ selected element in Fig. 2 is t18, then chosen nine digits will be as $S_i = [U42\ U43\ t19\ U44\ t18\ U38\ U39\ U40\ U41]$, As these nine digits are part of a Sudoku, hence all 9 digits will be different. Now, for first digit data ND_1 find its position (p) of ND_1 in the S_i and then in interpolated cover image img1 (Eq. 13) replace the value at $(r1_1, r1_{n1})$ with $(r1_1, p)$, After first data digit encoding again in interpolated cover image selects ($r1_1$ and $c1_{n1}$) and then finds position $(r1_1, c1_{n1})$ in Fig. 2 Sudoku matrix SudM, then select second data digit ND_2 similarly select nine values (4 forward, 4 backward and from position $(r1_1, c1_{n1})$ and again find ND_2 digit position (p) in S_i and replace interpolated image $(r1_1, c1_{n1})$ value with $(r1_1, p)$, keep doing the process for all data digits will produce stego-image which is Sudoku-based encoded version of cover image as shown in Eq. 14.

$$g1 = \begin{matrix} r1_1 & Y_{n1} & r1_2 & Y_{n2} & r1_3 & Y_{n3} \\ Y_{n4} & rc1_{n1} & Y_{n5} & rc1_{n2} & Y_{n6} & rc1_{n3} \\ r2_1 & Y_{n7} & r2_2 & Y_{n8} & r2_3 & Y_{n9} \\ Y_{n10} & rc2_{n1} & Y_{n11} & rc2_{n2} & Y_{n12} & rc2_{n3} \\ r3_1 & Y_{n13} & r3_2 & Y_{n14} & r3_3 & Y_{n15} \\ Y_{n16} & rc3_{n1} & Y_{n17} & rc3_{n2} & Y_{n18} & rc3_{n3} \end{matrix} \tag{14}$$

The whole idea of SKUSS is that here not making any changes in the original information pixels of the original cover image, new generated pixels values after interpolated pixel get modified and small changes in this interpolated pixels do not significantly affect quality of the image. Also the cover image pixels are just finding location in the 255 × 255 Sudoku matrix and as per the data digit and that location pixel of cover gets modified, it may be noted that the change in pixel can be ±8 only (Fig. 3).

At the receiver side with the same 16 BCD digits (64 key), public and private keys, same Sudoku generated as already explained in Step-1 and Step-2 of algorithm. At the receiver side, reverse encoding of 255 × 255 Sudoku matrix on stego image reconstructs the base-9 of original data which can be further converted into base-10 for original ASCII message data. Also simple decimation of the stego-image at receiver side reconstructed the original cover image. As both the data and the cover image reconstructed at the receiver side, the proposed SKUSS method is a reversible data hiding method.

3 Results and Discussion

The work is implemented using MATLAB and tested for a different combination of cover images and data sizes. One example of simulation is explained here with eight-digit public key 52,642,847 and eight-digit private key 12,345,678, the unsolved Sudokum and then its solved Sudoku developed and shown in Fig. 4.

Data conversion into Base-9 can be explained with an example that let the data is 'Ram' its base-10 ASCII is {082,097,109} and its base-9 is {101,117,131}. The base-9 data digits is used to modify interpolated pixels values.

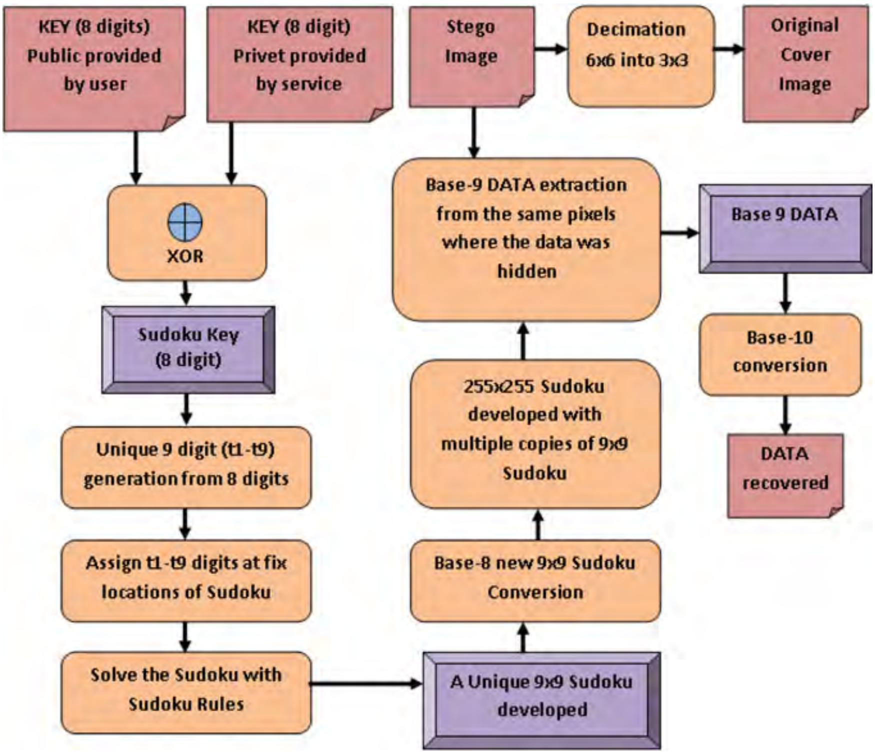


Fig. 3 Process flow of SKUSS: Data Extraction

sudoku =

6	0	0	0	0	5	0	8	0
0	3	0	0	4	0	0	0	1
0	0	7	9	0	0	2	0	0
0	0	6	5	0	0	8	0	0
0	4	0	0	3	0	0	0	7
1	0	0	0	0	9	0	0	0
2	0	0	0	0	0	0	6	0
0	5	0	0	0	0	0	0	8
0	0	3	0	0	0	4	0	0

(a)

sud =

6	2	1	3	7	5	9	8	4
5	3	9	8	4	2	6	7	1
4	8	7	9	6	1	2	3	5
3	9	6	5	1	7	8	4	2
8	4	5	2	3	6	1	9	7
1	7	2	4	8	9	3	5	6
2	1	8	7	9	4	5	6	3
9	5	4	6	2	3	7	1	8
7	6	3	1	5	8	4	2	9

(b)

Fig. 4 **a** Unsolved Sudoku developed with a combination of 8 digits public key and 8 digit private key **b** Sudoku after the solution

Figure 5 below shows the original Lena image, interpolated Lena image, and final stego Lena image. The image format taken is a bit map file.

Bits per pixel is the number of bits that can be hidden inside a pixel. Proposed method is not a direct data hiding method; here, the interpolated cover image is just changed according to base-9 data digits and Sudoku encoding and develop stego image. The capacity of data which can be hidden inside the proposed work can be

Fig. 5 **a** Original 512 × 512 pixels Lena image **b** interpolated Lena image of 1024 × 1024 pixels **c** stego Lena image of size 1024 × 1024 pixels

explained with an test image of Lena with 512 × 512 × 3 pixels. Total 170 × 170 × 3 = 86,700 numbers of 3 × 3 blocks can be developed with the test image and after interpolation, 3 × 3 block is converted into 6 × 6 block. Hence, the interpolated image will have{(170 × 170 × 6 × 6) + (170 × 6 × 2) + (170 × 2 × 6) + (2 × 2)} × 3 = 3,133,452 pixels. In one 6 × 6 block of interpolated image total 18, base-9 digits can be encoded hence total 86,700 × 18 × 8 = 12,484,800 maximum data bits can be hidden in interpolated cover image.

$$\text{Maximum BPP} = \frac{\text{maximum numbers of data bits Decoded at reciver}}{\text{Pixels in cover image}} = \frac{12,484,800}{3,133,452} = 3.98$$

Maximum BPP can be achieved in this work is 3.98; hence, up to 49.75% of cover image size message data can be hided with proposed SKUSS method however with maximum BPP (i.e., 3.98) the PSNR obtained is only 28.87 which is very low and cannot be considered as maintained image naturalness. Table 2 shown below is the results obtained for different sizes of message data hidden and recovered message with the proposed method. PSNR and MSE are computed between the cover and stego image. It may be noted that for cover image of size 512 × 512 × 3, total 1,024,000 data digits can fully be recovered in the proposed method; if need to hide bigger data (i.e., more than 1,024,000 data digits), then the size of the cover image must be increase.

Table 3 shown below shows the analysis between embedding capacity/bit per pixels and also comparison with Nguyen's [1] data hiding method. The experimental results are shown in Table 2 for analyzing the test image of Lena with size of 512 × 512.

From Table 3 and Fig. 6, it may be observed that proposed work can hide 1.2 bits per pixel for the image naturalness PSNR 47.82; however, Nguyen [1] can hide only

Table 2 MSE and PSNR observed for different message size for Lena's cover image of 512 × 512

Number of message data digits hided	Between cover and stego image		Bits per pixel	Message recovered
	PSNR	MSE		
2000	74.31	0.0024	0.00051	YES
4000	70.79	0.0055	0.01021	YES
8000	65.01	0.0207	0.02042	YES
16,000	62.15	0.0399	0.04084	YES
32,000	58.65	0.0894	0.08169	YES
64,000	55.84	0.1708	0.16339	YES
128,000	53.83	0.2713	0.32679	YES
256,000	51.35	0.4803	0.6535	YES
512,000	47.18	1.2545	1.30718	YES
1,024,000	43.19	3.1440	2.61436	YES

Table 3 Analysis between BPP and PSNR and comparison with Nguyen [1] data hiding scheme

Embedded capacity (BPP)	PSNR observed in SKUSS (proposed method)	PSNR observed in (Nguyen [1])
0.5	52.07	47.81
0.8	50.98	44.65
1	50.09	42.57
1.2	47.82	41.09

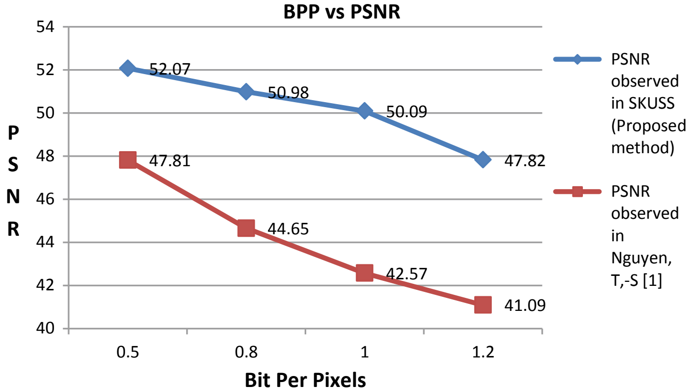


Fig. 6 Comparison between proposed SKUSS data hiding and Nguyen [1] data hiding scheme

Table 4 Comparative results

Methods and observe results		Proposed work observe results
Chang et al. [4]	1.5 bpp hidden and obtain PSNR of 44.5	1.5 bpp hidden and obtain PSNR of 45.08
Kumar [3]	0.06688 bpp and obtain PSNR of 59.27	0.06 bpp and obtain PSNR of 60.11
Cheng et al. [2]	1.28 bpp hidden and obtain PSNR of 45.41	1.3 bpp hidden and obtain PSNR of 47.18
Nguyen et al. [1]	1.2 bpp hidden and obtain PSNR of 41.03	1.2 bpp hidden and obtain PSNR of 47.82

0.5 bits per pixels for image naturalness PSNR 47.81. This proves that this work is having high embedding capacity.

Form Table 4, it can be observed that the embedding capacity of proposed work is better than other methods.

4 Conclusion

This work shows a new method of data hiding named Split Key Unique Sudoku Steganography (SKUSS), where two split keys one 8 digits public key and another 8 digit private key combine and a 16-digit steganography key developed. Unsolved Sudoku is designed with the help of Steganography key and that Sudoku is solved with the fast Sudoku solver algorithm. With the help of solved Sudoku, the message data digits modifies the interpolated pixels of cover image and produces Stego image in Bit map format. Here, no direct message data is hidden only pixels modifies according to message. Hence robustness, naturalness imperceptibility, and extent are maintained significantly. As only newly generated pixels of the cover image get modify according to data the reconstruction of the original image at the receiver side became easy with simple decimation of stego-image and this work achieves reversible data hiding. High PSNR and low MSE are achieved because of the same. The work is a design and simulated using MATLAB tool. Proposed SKUSS can hide 1.2 bits per pixel for the image naturalness PSNR 47.82 though Nguyen [1] can hide only 0.5 bits per pixels for image naturalness PSNR 47.81; hence, it may be concluded that proposed work is having high embedding capacities. This work is tested with bit map file format only and in near future it can be implemented with the other image file formats.

References

1. T.S. Nguyen, C.C. Chang, A reversible data hiding scheme based on the Sudoku technique. Displays **39**, 109–116 (2015). https://doi.org/10.1016/j.displa.2015.10.003
2. J.C. Cheng, W.C. Kuo, S.R. Su, Data-Hiding based on sudoku and generalized exploiting modification direction. J Electron Sci Technol **16**(2), 123–128 (2018)
3. M.V.S. Kumar, E. Mamatha, C.R. Reddy, V. Mukesh, R.D. Reddy, Data hiding with the dual based reversible image using the Sudoku technique, in *IEEE International Conference on Advances in Computing, Communications, and Informatics (ICACCI)* (2017), pp. 2166–2172. https://doi.org/10.1109/ICACCI.2017.8126166
4. C. Chang, Y. Chou, T.D. Kieu, An information hiding scheme using sudoku, in*IEEE International Conference on Innovative Computing Information and Control* (2008), pp. 17–21. https://doi.org/10.1109/ICICIC.2008.149
5. S. Jana, A.K. Maji, R.K. Pal, A novel SPN-based video steganography scheme using Sudoku puzzle for secured data hiding. Innov. Syst. Softw. Eng. **15**, 65–73 (2019). https://doi.org/10.1007/s11334-019-00324-8
6. P. Jorgen, Sudoku Solver and Generator (2010)
7. C. Chang, T. Nguyen, Y. Liu, A reversible data hiding scheme for image interpolation based on reference matrix, in IEEE *International Workshop on Biometrics and Forensics (IWBF)* (2017), pp. 1–6. https://doi.org/10.1109/IWBF.2017.7935098
8. C. Chang, C. Li, Reversible data hiding in JPEG images based on adjustable padding, in *5th International Workshop on Biometrics and Forensics (IWBF)*, pp. 1–6 (2017) https://doi.org/10.1109/IWBF.2017.7935083
9. Y. Lin, C. Wang, W. Chen, F. Lin, W. Lin, A novel data hiding algorithm for high dynamic range images. IEEE Trans. Multimedia **19**(1):196–211 (2017). https://doi.org/10.1109/TMM.2016.2605499
10. S. Dixit, A. Gaikwad, S. Gaikwad, S.A. Shanwad, Public key cryptography based lossless and reversible data hiding in encrypted images. Int. J. Eng. Sci. Comput. **6**(4), 75–79 (2016). https://doi.org/10.4010/2016.822
11. S. Chakraborty, S.K. Bandyopadhyay, Steganography method based on data embedding by sudoku solution matrix. Int. J. Eng. Sci. Invent. **2**(7), 36–42 (2011)
12. R. Kohias, U. Maurer, Reasoning about public-key certification: on bindings between entities and public keys. IEEE J. Sel. Areas Commun. **18**(4), 551–560 (2000). https://doi.org/10.1109/49.839931
13. V.M. Manikandan, V. Masilamani, Reversible data hiding scheme during encryption using machine learning. Int. Conf. Robot. Smart Manuf. Procedia Comput. Sci. **133**, 348–356 (2018)
14. A. Haj, A.H. Nabi, Digital image security based on data hiding and cryptography. in *International Conference on Information Management* (ICIM), pp. 437–440 (2017). https://doi.org/10.1109/INFOMAN.2017.7950423
15. S. Rawal, Advanced encryption standard (AES) and it's working. Int. Res. J. Eng. Technol. 3(8), 125–129 (2014)

Identification of Insomnia Based on Discrete Wavelet Transform Using Time Domain and Nonlinear Features

P. Mamta and S. V. A. V. Prasad

Abstract Insomnia is a type of sleep disorder that affects both the psychological and mental state. Conventionally, clinicians diagnose insomnia with a clinical interview that is subjective and suffer from personal error decision. The aim of this study is to identify insomnia subjects from normal subjects by using an electroencephalogram (EEG) signal, which is considered from a publicly available CAP sleep database. The EEG signal is decomposed by applying discrete wavelet transform (DWT) to obtain different brain wave patterns namely beta, alpha, theta, and delta waves. The time domain and nonlinear features are extracted from the Fp2–F4 EEG channel to form a feature vector. The performance of the four different classification techniques comprising k-nearest neighbor (KNN), support vector machine (SVM), ensemble, and decision tree (DT) is evaluated by employing fivefold cross-validation. The DT classifier achieved a classification accuracy of 85%. Also, the results demonstrate the feasibility of prefrontal channel EEG (Fp2–F4) for the identification of insomnia.

Keywords Insomnia · Discrete wavelet transform · EEG · Time domain features · Nonlinear features

1 Introduction

Insomnia is related to sleep dispossession, which has the tendency to affect both the psychological and mental state. Mental issues are more prone to insomnia or any other sleep disorders. Individuals with insomnia may have a ten-fold risk in the development of psychiatric illness or depression compared to the individuals, who are good sleepers. Sleep issues are most common in patients with bipolar disorder, anxiety, and depression. About one-third of depressed individuals have insomnia symptoms. Approximately 10% of older patients and 40% of young depressed patients suffer

P. Mamta (✉)
Department of EEE, G.Narayanamma Institue of Technology and Science, Hyderabad, India

S. V. A. V. Prasad
Department of EEE, Lingayas Vidyapeeth, Faridabad, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_11

from hyper-insomnia. Traditionally, clinicians treat patients with psychiatric disorders or depression also view insomnia as a significant symptom. Generally, the diagnosis of insomnia is based on the insomnia severity index questionnaire [ISI], where physicians ask some sleep-related questions such as daytime sleepiness and awake. Besides, physicians also diagnose insomnia through objective measures using PSG recording from the subjects complaining of insomnia. Subjects spend overnight or two days at the sleep center to record PSG, where it includes EEG, EOG, and EMG signals.

EEG is a non-invasive diagnostic instrument that is used to record the potential difference of a signal by placing a 10–20 electrode system on the brain scalp. It is prevalently used in the diagnosis of depression, insomnia, epilepsy, and many other psychiatric anomalies. EEG was claimed as a non-invasive procedure to study cognitive response [1, 2] and other disorders like insomnia, epilepsy, and depression [3–5]. Among all the types of cognitive data, EEG indicates emotional human brain activity in real time. Nauta et al. highlighted that the prefrontal cortex plays a crucial role in various aspects of the cognitive process [6]. In [7] demonstrated the EEG signal classification of healthy, epilepsy groups, groups of epileptic syndrome during seizure using a wavelet-based neural network classifier. In [8], the alertness level is distinguished by the error backpropagation neural network classifier and uses the power spectral density of DWT as input features. Furthermore, it is studied that the wavelet transform decomposes the EEG signal into frequency sub-bands through which statistical features were extracted [9]. In a study, [10] has analyzed normal and insomnia subjects from ECG and EEG signals and extracted linear and nonlinear features such as largest Lyapunov exponent, sample entropy, and correlation dimension from the signals. The execution of wavelet transform (WT) has proven to be an optional tool for the Fourier transform (FT) to the study of non-stationary EEG signals [11, 12]. The author [13] demonstrated graph spectral theory using a hypnogram. Logistic regression was applied for the identification of insomnia and obtained accuracy, sensitivity, and specificity of values 81%, 87%, and 75%, respectively. In the study [14], nonlinear features of the EEG signal are used for the classification of insomnia subjects from healthy subjects and achieved the classification accuracy of 83% by using a support vector machine.

At present, many of the researchers have reported their studies on automatic sleep stage classification [15] and sleep disorder [16] using different machine learning algorithms. But, there are very few studies focusing on the identification of insomnia. Observational studies have demonstrated that insomnia is highly prevalent in depression patients and often, insomnia is considered as a core symptom of depression [17]. This paper remains as the first step toward a more profound understanding of the relation among insomnia, the prefrontal cortex, and depression. To the best of the author's knowledge, this is the first study to report that the prefrontal region and channel Fp2–F4 are feasible for the identification of insomnia using nonlinear and time domain features considered as a core symptom of depression. The proposed method used in this work provides a significant performance increase in comparison with others.

In this paper, the single-channel (Fp2–F4) EEG signal based on the prefrontal cortex region is analyzed from insomnia patients and healthy patients. The signals are decomposed using DWT, which includes four brain wave patterns such as beta (β), alpha (α), theta (θ), and delta (δ). Then, time domain and nonlinear features are computed from the obtained brain wave patterns. Further, computed features are applied to four different classification techniques namely SVM, DT, KNN, and ensemble using a fivefold cross-validation method to access the classification accuracy. Herewith, the study aims to identify the insomnia subjects from the healthy subjects based on discrete wavelet transform (DWT) using single-channel EEG (Fp2–F4). In this work, it is shown that the methodology used in our study could identify insomnia using the prefrontal cortex region.

This paper is organized as follows: Sect. 2 about EEG data collection and Sect. 3 presents an overview of the methodologies employed in our work. In Sect. 4, results and discussions are elaborated and the conclusion of the study is presented in Sect. 5.

2 EEG Data Collection

In this study, the EEG data is obtained from the publicly available datasets on the physionet named CAP Sleep database version1.0.0. Dataset is constructed by the sleep disorder center, the ospedate Maggiore of Parma, Italy [18]. It has 108 subjects of polysomnographic (PSG) recordings comprising of three EEG channels (F3 or F4, O1 or O2, and C4 or C3), EEG bipolar channels, two EOG channels, EMG signals, and ECG signals.

In our proposed work, the Fp2–F4 EEG channel is selected for insomnia identification and the experiment is carried out on five healthy subjects and seven insomnia patients with a sampling rate of 512 Hz. The number of samples drawn from data is 51200 for the first 10 s and it followed by 1 min with six epochs. Also, the EEG signal is employed by IIR notch filter to remove the power line interference (PLI) noise of 50 Hz. The EEG sample data segment of healthy subjects and insomnia patient is shown in Fig. 1.

3 Methodology

The flow diagram of the insomnia identification based on single-channel EEG (Fp2–F4) is shown in Fig. 2, which comprises of three main parts (i) discrete wavelet transform (DWT), (ii) feature extraction methods, and (iii) classification.

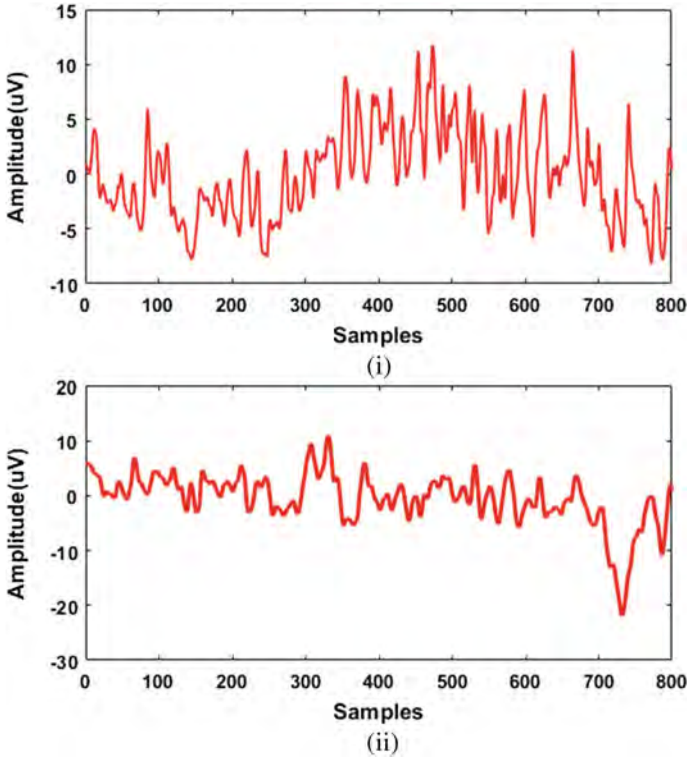


Fig. 1 Sample EEG signal of (i) insomnia subject and (ii) healthy subject

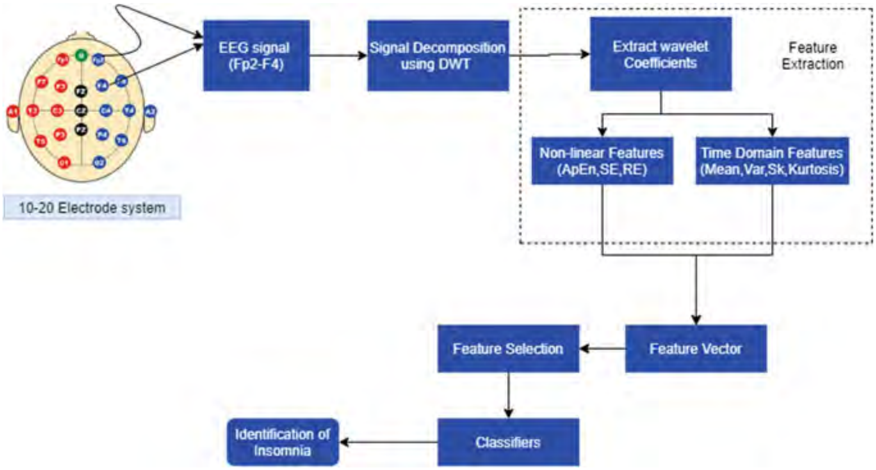


Fig. 2 Flow diagram for the insomnia identification

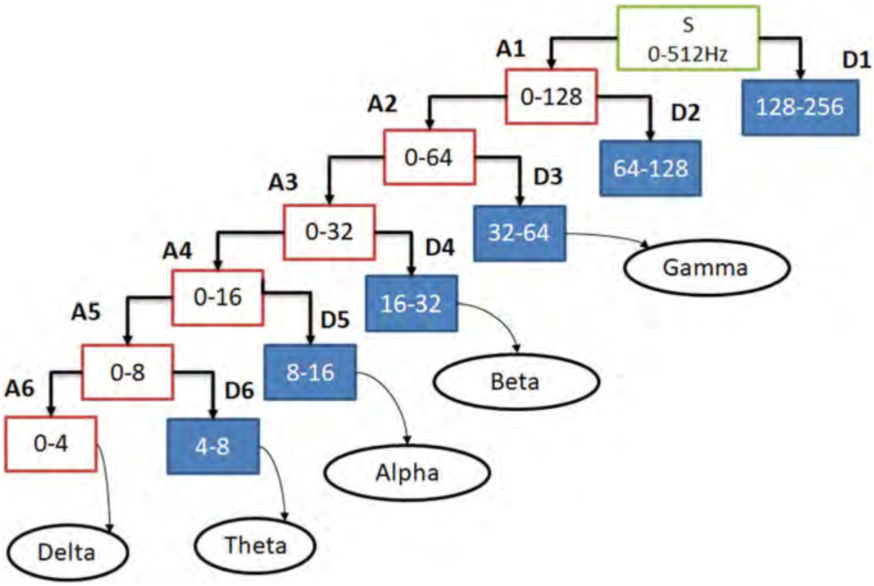


Fig. 3 The six-level decomposition of EEG signal using DWT

3.1 Discrete Wavelet Transform

Discrete wavelet transform (DWT) divides the EEG signal into two: detail and approximate coefficients at different frequencies. The detail coefficient is the high-frequency component, high-pass filter, and the approximate coefficient is the low-frequency component, low-pass filter. In this work, discrete wavelet transform (DWT) is employed to EEG signal of a healthy subject and insomnia patient to obtain mainly four different brain wave patterns comprising beta, alpha, theta, and delta waves. The EEG signal (S) with a sampling frequency of 512 Hz is decomposed applying six-level multiresolution decompositions using mother wavelet as Daubechies 4(Db4). According to the basic theory of the Nyquist sampling theorem [19], the corresponding frequency patterns are obtained as illustrated in Fig. 3. The detail coefficients at levels 4, 5, and 6 (D4, D5, and D6) represent the three brain wave patterns which include β (16–32 Hz), α (8–16 Hz), and θ (4–8 Hz) waves, respectively. The approximate coefficient at level 6 (A6) represents the δ wave (0–4 Hz).

3.2 Feature Extraction

This study has utilized two feature extraction methods, which are divided into (1) time domain method and (2) nonlinear approach. Statistical features and nonlinear

Table 1 Features extracted from Fp2–F4 channel EEG

Features	Label
f1	Mean, variance, kurtosis, and skewness
f2	Approximate entropy (ApEn)
f3	Shannon entropy (SE)
f4	Renyi's entropy (RE)

features are computed from the single-channel (Fp2–F4) EEG full-wave, β, α, θ, and δ waves. The extracted features are labeled and listed in Table 1.

3.2.1 Time Domain Method

In the time domain method, the statistical parameters comprising mean, variance, skewness, and kurtosis are determined. These parameters are mostly known for their ability to describe the statistical moments of the EEG signal and are known as linear features. The expression to calculate the mean, variance, skewness, and kurtosis as follows:

$$\text{Mean}(\mu) = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{1}$$

$$\text{Variance}(\vartheta) = \sigma^2 \tag{2}$$

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} /y_i - \mu/^2} \tag{3}$$

where σ is standard deviation

$$\text{skewness}(\psi) = \frac{E(y-\mu)^3}{\sigma^3} \tag{4}$$

$$\text{kurtosis}(k) = \frac{E(y-\mu)^4}{\sigma^4} \tag{5}$$

where y_i is the sample and N is the number of samples.

3.2.2 Nonlinear Methods

Nonlinear methods can capture the chaos behavior and sudden changes in the EEG signal caused by biological events developing in the brain. In this study, the nonlinear methods used are approximate entropy (ApEn), Shannon entropy (SE), and Renyi's entropy (RE). The extracted features are listed in Table 1.

Approximate Entropy

ApEn is introduced by Pincus to address the complexity and irregularity of the time series [20]. ApEn is a statistical method employed to quantize the abnormality of the signal. The more signal is complex, the higher the ApEn values. The following expressions compute approximate entropy (ApEn).

Let us consider the initial signal be $y(1)$, $y(2)$, ..., $y(N)$, where N is the number of samples of the signal.

$$Y(i) = [y(i), y(i+1), y(i+2), \ldots, y(i+m-1)], 1 \leq i \leq N-m+1$$

where $Y(i)$ represents the time series of the signal and m is the embedding dimensions. Approximate entropy (ApEn) is calculated by

$$ApEn(m, r, N) = \psi^m(r) - \psi^{m+1}(r) \tag{6}$$

where

$$\psi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln(C_i^m(r)) \tag{7}$$

To compute $C_i^m(r)$ the distance between the two elements namely $Y(i)$ and $Y(j)$ are compared with a threshold value (r). In this work, the threshold value is chosen as $r = 0.25 \times$ sd (standard deviation of the signal).

And: $C_i^m(r)$ for each i,$i = 1, 2, 3,\ldots, N - m$ is defined as:

$$C_i^m(r) = \frac{\text{number of } d[Y(i), Y(j)] \leq r}{N-m+1} \tag{8}$$

where $C_i^m(r)$ represents a correlation integral for a given time series, it measures the information generated in a chaotic system with a threshold value r, for m dimensions.

Shannon Entropy

Shannon entropy evaluates the irregularity of the signal [21]. The Shannon entropy of Y as an EEG signal is defined as:

$$\mathrm{SE}(Y) = -\sum_{i=1}^{n} P_i \log_b P_i \tag{9}$$

where P_i is the probability distribution, and b is the logarithmic base, i.e., bits $b = 2$.

Renyi's Entropy
Renyi's entropy (RE) is a measure used for computing the spectral complexity of the time series [22]. This work has considered α value and logarithm base as 2. It is defined as:

$$\mathrm{RE}(\alpha) = \frac{\alpha}{1-\alpha}\log\left(\sum_{i=1}^{n} P_i^{\alpha}\right), \alpha \geq 0 \quad \text{and} \quad \alpha \neq 1 \tag{10}$$

where quantity P_i is the probability distribution of the EEG signal with n bins and α is the Renyi's entropy order.

3.3 Classification Techniques

The proposed research work has considered four different classifiers which include k-nearest neighbor (KNN) [23], support vector machine (SVM), ensemble [24], and decision tree (DT) [25]. These classifiers are used to evaluate and compare the classification performance by employing fivefold cross-validation. And, also to demonstrate best suitable classifier for the identification of insomnia, SVM has been employed in the field of depression differentiation [26]. KNN is used in the field of medical informatics, including the diagnosis of stress and epilepsy [27]. The classification performance of the classifiers evaluated by calculating sensitivity (S_e), specificity (S_p), and accuracy (A_{cc}) is defined as follows [28]:

$$A_{\mathrm{cc}} = \text{ No. of correct decisions/ total number of cases} \tag{11}$$

$$S_e = \text{ True negative/ False positive } + \text{True negative} \tag{12}$$

$$S_p = \text{ True positive/ False negative} + \text{True positive} \tag{13}$$

4 Results and Discussion

This section presents a detailed evaluation of our results. In this study, all data preprocessing and parameter analysis have been implemented using MATLAB software. The features f1, f2, f3, and f4 are extracted from full-wave EEG signal and brainwave patterns comprising β, α, θ, and δ waves that are obtained by employing DWT with six-level decomposition using Fp2–F4 channel EEG signal. The extracted features form a feature set. Further, PCA is utilized to decrease the dimensionality of the

Table 2 Classification performance of features with different classifiers

Features	Classifiers	Accuracy (%)	Sensitivity (%)	Specificity (%)
f1	DT	78.3	83	73.3
	SVM	65	43.3	86.7
	KNN	78.3	83.3	73.3
	Ensemble	78.3	83.3	73.3
f2	DT	60	53.3	66.7
	SVM	63.3	63.3	63
	KNN	73	83.3	63.3
	Ensemble	63.3	76.6	50
f3	DT	76.7	83	70
	SVM	68.3	73	63
	KNN	66.7	67	67
	Ensemble	75	83	67
f4	DT	63.3	77	50
	SVM	66.7	53	80
	KNN	68.3	73	63
	Ensemble	78.3	83	73

feature set and it also improves the performance of the classifiers with a high dimension of feature set. The reduced feature set is fed to four different classifiers KNN, SVM, ensemble, and DT to identify the insomnia patient's EEG signal from the normal subject by employing a fivefold cross-validation technique. Classification performance is obtained for individual feature f1, f2, f3, and f4 by applying four different classifiers, which are given in Table 2, and Fig. 4 depicts the comparison of different classifiers based on features.

In this study, it showed that f1 and f4 had obtained a classification accuracy of 78.3%, and f3 has achieved a classification accuracy of 76.7% less compared to f1 and f4. The accuracy obtained by f2 is low compared to f1, f3, and f4 with a value of 73%. But the individual feature is not providing good accuracy. So, the selected features are combined and applied to four different classifiers to evaluate the classification performance. The selected combined features are f3–f4, f2–f3–f4, f1–f2–f3–f4, and f1–f2 are listed in Table 3. The features f3–f4 and f2–f3–f4 have not shown accuracy improvement compared to the individual feature sets, whereas the features f1–f2–f3–f4 and f1–f2 have achieved an accuracy of the value of 80% and 85%, respectively, as shown in Fig. 5.

The study demonstrates that the choosing feature combination for the identification of insomnia is reasonable. It also indicates that the Fp2–F4 EEG channel is suitable and efficient in identifying the EEG signal of insomnia patients. Our outcomes are compared with the previously reported finding on the identification of insomnia, as shown in Table 4 and the classification accuracy of the proposed method is compared with the others based on channel selection as depicted in Fig. 6. Though

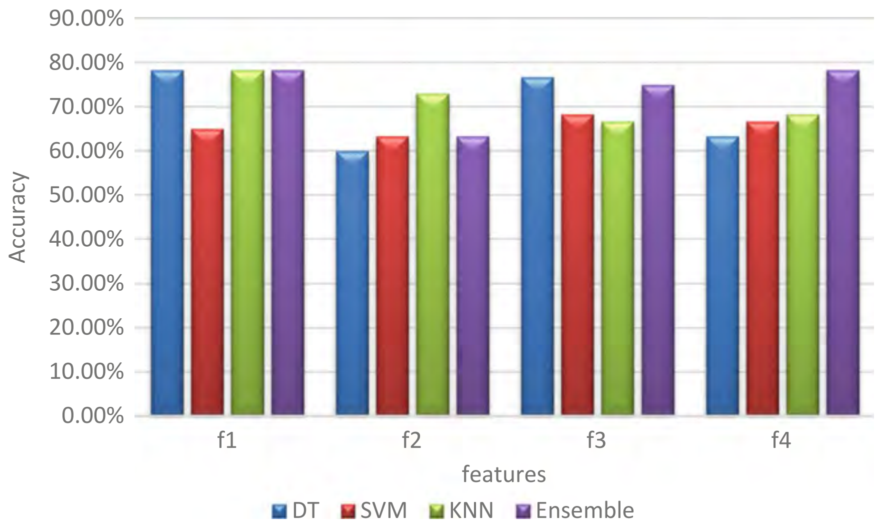


Fig. 4 Classification accuracy based on individual features

Table 3 Classification performance of selected feature combination with different classifiers

Features	Classifiers	Accuracy (%)	Sensitivity (%)	Specificity (%)
f3–f4	DT	76.7	77	77
	SVM	68.3	73	63
	KNN	76.7	63	70
	Ensemble	75	70	83
f2–f3–f4	DT	73.3	70	66.7
	SVM	68.3	66.7	70
	KNN	71.7	73.3	70
	Ensemble	78.30	80	66.7
f1–f2–f3–f4	DT	80	70	90
	SVM	65	43.3	86.7
	KNN	78.3	76.7	80
	Ensemble	78.3	76.7	80
f1–f2	DT	85	76.7	93.3
	SVM	63	43.3	83.3
	KNN	81.7	80	83.3
	Ensemble	81.7	80	83

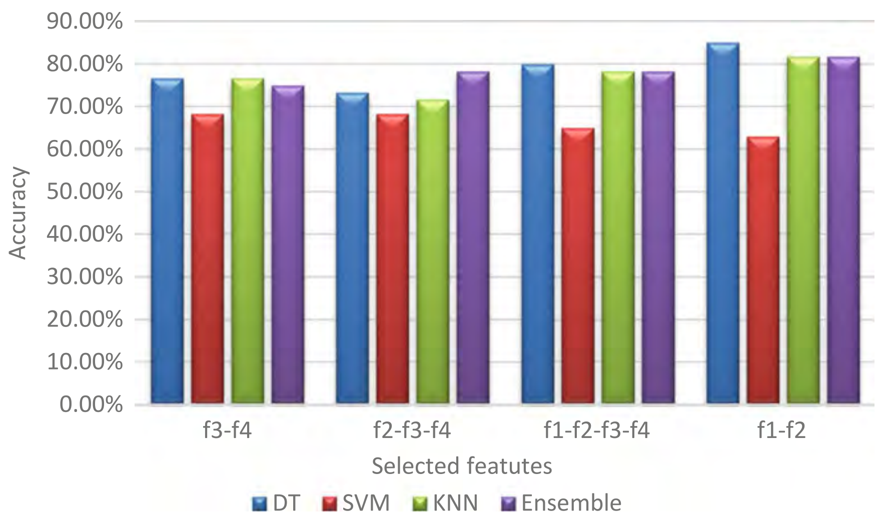


Fig. 5 Classification accuracy based on selected features

Table 4 Comparison of study

References	Channel	Features	Validation	Classifiers	Accuracy (%)
[13]	C4–A1	Linear features	Leave-one-out	Logistic regression	81
[14]	C3–A2	Nonlinear features	50–50%	SVM	83
Our study	Fp2–F4	DWT—linear and nonlinear features	Fivefold cross-validation	DT	85

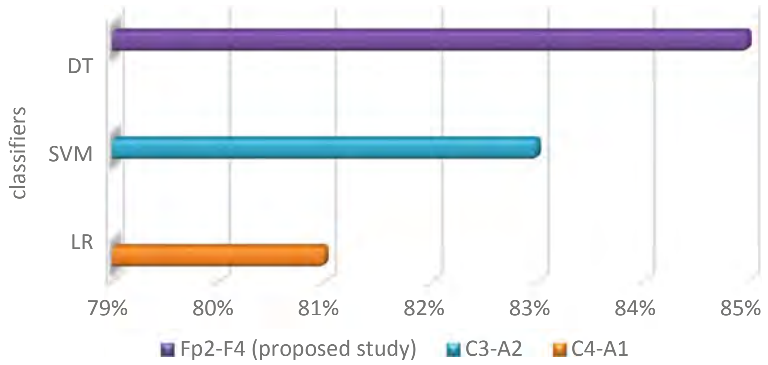


Fig. 6 Comparison of classification accuracy based on channel selection

the referred studies showed good classification accuracy, two main differences are to be highlighted in our work. Firstly, the channel selection of EEG. Secondly, extracted both linear features from time domain analysis and nonlinear features that represent the nonlinearity of the signal.

5 Conclusion

Insomnia is most common in people suffering from depression, stress, and anxiety. In this work, the EEG signal of insomnia and healthy subjects is measured from the Fp2–F4 channel which represents the prefrontal region. The time domain and nonlinear features are computed from detail and approximate coefficients by using DWT. The selected features are combined and fed to four different classifiers in which DT demonstrates the highest classification accuracy of about 85%. Overall, the identification of insomnia using the DWT technique produces good results with feature combination of time domain features and ApEn. The results also suggest that the channel Fp2–F4 is feasible to identify and analyze the insomnia EEG signal. The overall results are much satisfactory and fare well in comparison with others.

References

1. A.S. Gevins, G.M. Zeitlin, C.D. Yingling, et al., EEG patterns during cognitive tasks. I. Methodology and analysis of complex behaviors. Electroencephal. Clin. Neurophys. **47** 693–703 (1979)
2. F. Fan, Y. Li, Y. Qiu, Y. Zh., Use of ANN and complexity measures in cognitive EEG discrimination, in *27th IEEE Annual Conference on Engineering in Medicine and Biology*, 4638–4641, Shanghai, China (2005)
3. R.R. Rosa, M.H. Bonnet, Reported chronic insomnia is independent of poor sleep as measured by electroencephalography. Psychos. Med. **62**(4), 474–482 (2000)
4. H. Adeli, S. Ghosh-Dastidar, N. Dadmehr, A waveletchaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy. IEEE Trans. Biomed. Eng. **54**(2), 205–211 (2007)
5. R. Tibodeau, R.S. Jorgensen, S. Kim, Depression, anxiety, and resting frontal EEG asymmetry: a meta-analytic review. J. Abnormal Psychol. **115**(4), 715–729 (2006)
6. W.J.H. Nauta, Te problem of the frontal lobe: a reinterpretation. J. Psychiatric Res. **8**(3–4), 167–187 (1971)
7. I. Omerhodzic, S. Avdakovic, A. Nuhanovic, K. Dizdarevic, Energy distribution of EEG signals:EEG signal wavelet-neural network classifier. Int. J. Biol. Life Sci. **6**(4), 210–215 (2010)
8. M. Kemal Kiymik, M. Akin, A. Subasi, Automatic recognition of alertness level by using wavelet transform and artificial neural network. J. Neurosci. Methods **139**, 231–240 (2004)
9. A. Subasi, Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. Expert Syst. Appl. 701–711 (2005)
10. H. Abdullah, T. Penzel, D. Cvetkovic, Detection of Insomnia from EEG and ECG, in *IFMBE Proceedings. 15th International Conference on Biomedical Engineering*, vol. 43, pp. 687–690 (2014)

11. I. Clark, R. Biscay, M. Echeverria, T. Virues, Multiresolution decomposition of nonstationary EEG signals: a preliminary study. Comput. Biol. Med. **25**(4), 373–382 (1995)
12. D.P. Subha, P.K. Joseph, U.R. Acharya, C.M. Lim, EEG signal analysis: a survey. J. Med. Syst. **34**(2), 195–212 (2010)
13. R. Chaparro-Vargas, B. Ahmed, N. Wessel, T. Penzel, D. Cvetkovic, Insomnia characterization: from hypnogram to graph spectral theory. IEEE Trans. Biomed. Eng. **63**(10), 2211–2219 (2016)
14. H. Abdullah, C.R. Patti, C. Dissanyaka, T. Penzel, D. Cvetkovic, Support vector machine classification of EEG nonlinear features for primary insomnia, in *Proceedings of the International Conference for Innovation in Biomedical Engineering Life Sciences*, pp. 161–164 (2018)
15. K. Chen, C. Zhang, J. Ma, G. Wang, J. Zhang, Sleep staging from single-channel EEG with multi-scale feature and contextual information. Sleep Breath. **23**(4), 1159–1167 (2019)
16. S. Fallmann, L. Chen, Computational sleep behavior analysis: a survey. IEEE Access **7**, 142,421–142,440 (2019)
17. S. Kaya, C. McCabe, What role does the prefrontal cortex play in the processing of negative and positive stimuli in adolescent depression? Brain Sci. **9**, 104 (2019)
18. A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.Ch. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet.: components of a new research resource for complex physiologic signals 215–220 (2003)
19. H. Ocak, Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. Expert Syst. Appl. **36**(2), 2027–2036 (2009)
20. S.M. Pincus, Approximate entropy as a measure of system complexity. Proc. Natl. Acad. Sci. USA **88**, 2297–2301 (1991)
21. C.E. Shannon, A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379–423 (1948)
22. P. Grassberger, T. Schreiber, C. Schaffrath, Nonlinear time sequence analysis. Int. J. Bifurc. Chaos **1**, 521–547 (1991)
23. S. Shakya, Analysis of artificial ıntelligence based ımage classification techniques. J. Innov. Image Process. (JIIP) **2**(01), 44–54 (2020)
24. C. Kuo, G. Chen, A short-time Insomnia detection system based on sleep EOG with RCMSE analysis. IEEE Access **8**, 69763–69773 (2020)
25. S.F. Liang, Y.H. Shih, P.Y. Chen, C.E. Kuo, Development of a human-computer collaborative sleep scoring system for polysomnography recordings. PLoS ONE **14**(7) (2019)
26. O. Faust, P.C.A. Ang, S.D. Puthankattil, P.K. Joseph, Depression diagnosis support system based on eeg signal entropies. J. Mech. Med. Biol. **14**(3) (2014)
27. J.S. Wang, C.W. Lin, Y.T.C. Yang, A k-nearest-neighbor classifer with heart rate variability feature-based transformation algorithm for driving stress recognition. Neurocomputing **116**, 136–143 (2013)
28. A. Baratloo, M. Hosseini, A. Negida, G. El Ashal, Part 1: simple definition and calculation of accuracy, sensitivity and specificity. Emergency **3**(2), 48–49 (2015)

Transfer Learning Techniques for Skin Cancer Classification

Mirya Robin, Jisha John, and Aswathy Ravikumar

Abstract Increase in usage of cosmetics, pollution, and radiations will always result in skin-related diseases. Moreover, skin cancer has also become a common disease. There are many features available to help in the process of identification of skin cancer. Deep learning algorithms are successfully applied to identify the skin cancer. Various transfer learning strategies can also be applied. This paper attempts to compare various transfer learning approaches and the degree of accuracy by using these models. The analysis can be made by training the system with the details obtained from images available in the database and the current image is then tested to find whether it is malignant or not. The dataset images obtained from Kaggle dataset have been used for this study. Pre-trained models like InceptionV3, ResNet50, and MobileNet are used along with the extracted pre-trained weights in this research work.

Keywords Skin cancer · Deep learning · MobileNet · ResNet50 · Prediction

1 Introduction

Skin cancer is considered as a rapidly spreading disease in this era due to the prevailing atmospheric conditions. There are approximately 5.4 million cases in USA alone each year. According to recent reports, during 2008–2018, there has been a 53% increase in new melanoma cases annually. There is also an expected rise in the mortality rate of this disease in the next decade. If the disease is diagnosed in the later stages, the rate of survival is less than 14% but if the skin cancer is detected

M. Robin (✉) · J. John · A. Ravikumar
Department of Computer Science and Engineering, Mar Baselios College of Engineering and Technology, Kerala, India

J. John
e-mail: jisha.john@mbcet.ac.in

A. Ravikumar
e-mail: aswathy.ravikumar@mbcet.ac.in

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_12

at early stages, the rate of survival is approximately 97%. Hence, early detection of skin cancer is lifesaving. Dermatologists usually perform diagnosis manually. It was observed that the skin specialist follow a series of traditional steps for diagnosis which includes the observation with naked eye of suspected lesions, then dermoscopy (magnifying lesions microscopically) and finally conduct the biopsy. The above steps require time and sometimes there are chances patient may move to later stages and the accurate diagnosis is highly dependent on the skill of the clinician. Even the best dermatologist may have only 80% accuracy in correctly diagnosing cancer. The availability of skilled dermatologist is also lacking globally in public healthcare. Hence, in order to diagnose skin cancer in a fast and accurate manner and to solve the above problems, extensive computer image analysis algorithms have been developed. Most of the algorithmic solutions are parametric, and require data to be normally distributed. The data available is diverse and its nature cannot be controlled and this leads to the inefficiency of these methods to diagnose the cancer. But the non-parametric solutions lack the requirement of the data to be in normal distribution form. In this work, various pre-trained deep learning models are used to train and analyze skin cancer. In this paper, we address different pre-trained learning methods for developing deep neural network model for image classification to identify the skin cancer based on skin lesions [1].

A. Skin Cancer

Melanoma: Melanoma is a malignancy of melanocytes. Melanocytes are special cells in the skin located in its outer epidermis. Melanoma can be observed by human eye as it develops in the epidermis. The harmful effects can be reduced by the early detection and treatment. Excision operation can cure melanoma if detected at an early stage but the main threat faced in the treatment is the increased rate of false negative of malignant melanoma. Figures 1 and 2 [2] show benign and malignant images from the dataset. Female patients usually have melanoma on the lower limbs and male patients have them on the back but it can also be found on other organs containing cells such as the mouth and eye which is very rare [3].

Basal-Cell Carcinoma: There are at least 3 to 4 million cases of basal-cell carcinoma (BCC) in the USA annually. It arises from the deepest layer of the epidermis. They usually look like red patches or open sores. The cases of spread of BCC are very rare but people who have had BCCs are prone to develop it again in their lifetime.

2 Literature Survey

There are numerous researches being done for skin cancer based on analysis of images. In 2018, the International Skin Imaging Collaboration (ISIC) hosted a challenge contest in skin cancer detection which became the defacto standard. Mobile apps were also devised which could detect skin cancer. Different classification algorithms have been used to obtain better classification accuracy. Convolutional neural network (CNN) structure was introduced by Fukushima [4] and later Le-Cunn [5]

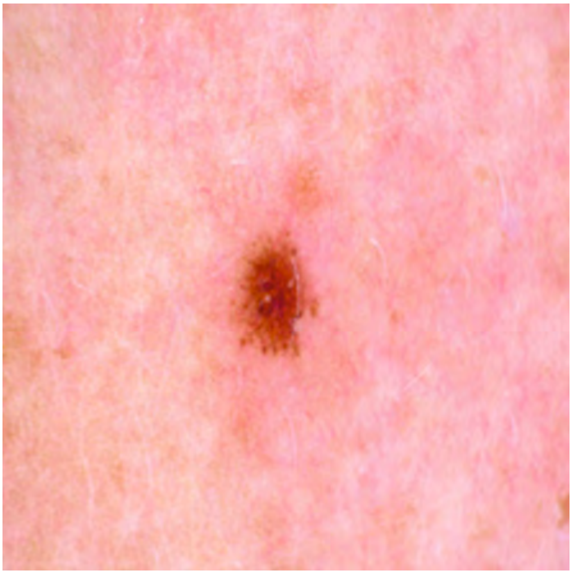


Fig. 1 Benign image [2]

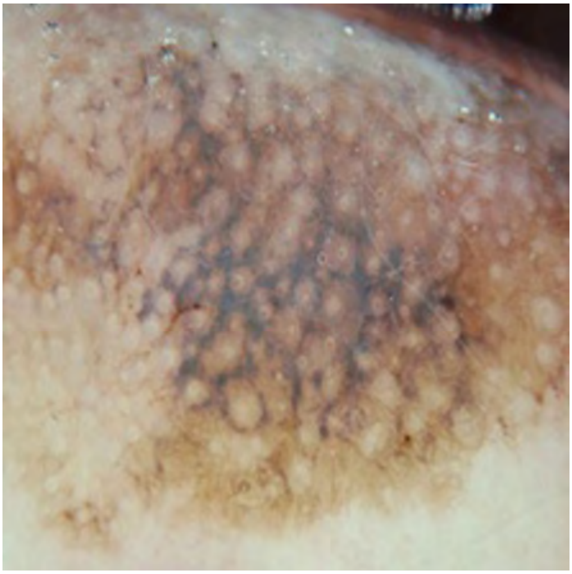


Fig. 2 Malignant image [2]

which led to a boom in this analysis. CNNs basically mimic the human visual cognition system and are considered to be the best method for image classification. The pre-trained GoogLeNet, Inception V3, CNN model came from Esteva et al. which was a major breakthrough. They used 129,450 clinical skin cancer images including 3,374 dermatoscopic images [1]. The classification accuracy reported was 85.5%.

In 2018, Haenssle et al. [6] utilized a deep convolutional neural network to classify a binary diagnostic category of dermatoscopy melanocytic images, and reported 86.6% sensitivity and specificity for classification. A multiclass classification using ECOC SVM and deep learning CNN was developed by Dorj et al. [7]. The approach was to use ECOC SVM with pre-trained AlexNet Deep Learning CNN and classify multi-class data. An average accuracy of 95.1% is reported in this work [1] (Table 1).

3 Proposed System

The proposed system is based on the transfer learning concept for the classification of the skin cancer lesions as malignant or benign. Here, the model is built based on the information obtained from the pre-trained models which are trained on huge datasets and the patterns are saved. The transfer learning helps to make use of the previously learned information. These pre-trained models have been extensively trained on big data sets and they are used widely to solve problems with similar dataset. In this work mainly three pre-trained models ResNet50, MobileNet and InceptionV3 weights for the classification.

The dataset used for the work is from Kaggle with half of the skin lesion images are pathology confirmed. The dataset consists of 3000 lesion images, which includes 1799 benign images and 1500 malignant images. Image resizing refers to the scaling of images. Scaling comes handy in many image processing. It helps in reducing the number of pixels from an image and that has several advantages, for example, it can reduce the time of training of a neural network as more is the number of pixels in an image more is the number of input nodes that in turn increases the complexity of the model. It also helps in zooming in images. The image needs to be resized at times as per the standard size requirements.

In image processing application like image recognition and classification, the most commonly used neural network is convolutional neural network (CNN). The CNN is mainly used highly demanding fields like self-driving cars, medical diagnosis, robotics, etc. CNN is the supervised ANN trained using the data set which is labeled. CNN is used to automatically learn the relation between classes and the hidden feature relations. The main components are the hidden layers and fully connected layers. The main steps are convolution, pooling, and the final fully connected layer.

The proposed model consists of the following steps: The data collection, image resizing, the models are setup and the trained models are used for the classification and finally the testing phase and evaluation of the model as shown in Fig. 3. The transfer learning is highly effective due to the simple way to implement it to the specific problem, the better performance can be obtained with less training time, and the need of labeled data is not relevant in this case and is versatile.

Figure 4 shows sample skin lesion benign images from the Kaggle dataset. The Kaggle dataset is split based on the 80–20 rule into training and testing phase. In the

Table 1 Comparison of existing techniques

Method	Advantage	Disadvantage
Image-wise supervised learning (ISL) and multi-scale super pixel-based cellular automata (MSCA) for segmentation of skin lesion [1]	• More capable of segmenting skin lesions of varying contrast and size • Better segmentation performance • Compared to other methods more reliable and accurate	• Lower segmentation accuracy, due to the larger super pixels responsible for more discriminative features that are crucial for lesion detection • Small part of lesion area is represented by smaller super pixel scales
K-mean and PSO technique for detection of skin cancer [2]	• The quality of image and increase in detection accuracy is obtained by the making use of the k-mean with PSO • Enhanced detection and quality of the image obtained	• When dataset is large or complex performance of PSO clustering method is found to degrade • Prior knowledge of the initial selection of number of clusters in k-means is required
Skin lesion segmentation using the artificial bee colony algorithm [3]	• Compared to other methods it is faster, simple for implementation, flexible, and have fewer parameter • Higher performance	• Fails to get accurate segmentation • Accuracy is less due to lack of contrast between skin lesion and background
Computer-aided diagnosis of melanoma Skin cancer [4]	• User friendly and robust tool for automated diagnostics of the skin cancer • This tool is more useful for the rural areas where the experts in the medical field may not be available • It increases the diagnosis accuracy as well as the speed	• System includes challenges in input data collection, preprocessing, processing, and system assessment
Convolutional neural network (CNN) [5]	• CNNs are considered as the best image classification and analysis model • CNN is incorporated into deep learning for improved accuracy • Automated feature extraction is provided • The number of steps is drastically reduced by using convolution on patches of adjacent pixels	• Reproducibility is troublesome due to usage of non-public datasets for training and additionally testing

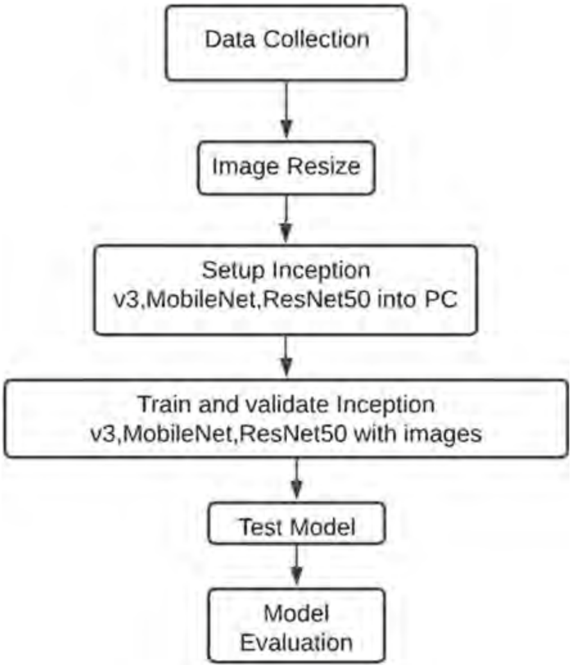


Fig. 3 Proposed method

second step, the image augmentation is done using the zooming, shearing, and flipping operations to increase the dataset. The third step is the loading of the pertained models and in order to avoid overfitting the dropout and normalization (batch normalization) is applied. The neural network is having two dense layers with 64 and 2 neurons, respectively, with softmax activation function on the final layer. The loss function used here is binary cross entropy. The model was trained with 20 epochs by varying the learning rate and batch size.

4 Result and Analysis

The transfer learning concept is used here for the classification of the skin. The main advantage of transfer learning is there is always a starting point to begin rather than building the model from the initial scratch. In this method, the previously learned models are leveraged and these models are already trained on a big dataset and the weights are freezed for easy progress. In this work, mainly the three pertained networks are used the MobileNet, Inception v3, and ResNet.

The first step was to analyze all the images in the training set and calculate the bottle neck which refers to the previous layer of the last fully connected deep neural network layer.

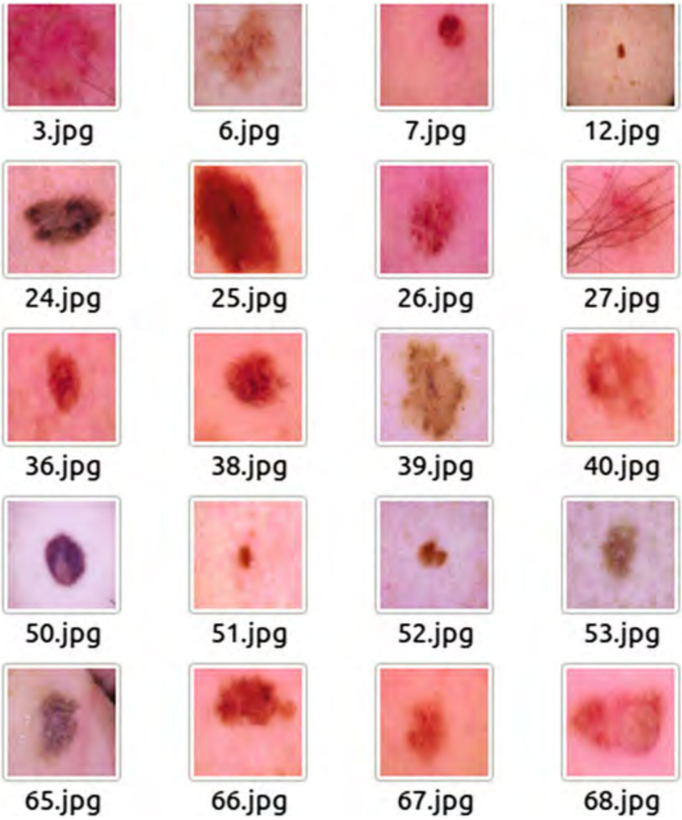


Fig. 4 Sample skin lesion benign image

A. *Inception-v3 Model*

A 42-layer deep neural net with learning rate 0.0001 and Adam optimizer used for both asymmetric and symmetric basic building blocks which consist of convolutional layers, pooling layers with max pool, average pool functions, dropout layers, and the final fully connected neural network layer is an Inception V3 model. Figure 5 shows the accuracy loss graph and Fig. 6 shows the confusion matrix when the skin lesion dataset was trained and classified by using the Inception V3 model.

B. *ResNet Model*

Residual network (ResNet) model is a pre-trained neural net used in image and vision-based problems. The ImageNet challenge was won by the ResNet. The ResNet have 50 layers and the input image size is 224 × 224 pixels. This is basically a CNN with multiple layers stacked which helps to analyze the low-, medium-, and high-level features. Figure 7 shows the accuracy and loss in training using Resnet50. Figure 8

Fig. 5 Accuracy and loss graph

Fig. 6 Confusion matrix

Confusion Matrix	TP	FP	FN	TN
Inception v3	20	18	10	12

Fig. 7 Accuracy and loss graph

Fig. 8 Confusion matrix

Confusion Matrix	TP	FP	FN	TN
ResNet	26	28	4	2

shows the confusion matrix when the skin lesion dataset was trained and classified by using the ResNet model.

C. *MobileNet Model*

MobileNet was proposed by Google and the main integral part of this network is the filter that is used for depthwise separation which helps to implement the pointwise collection. And in this model, it is applied to each input channel separately. Both depthwise and pointwise convolutions are applied to the training dataset and the optimum parameter are considered and the lightweight neural net is developed. MobileNet showed a good learning rate and accuracy on the training dataset and Fig. 9 shows the accuracy and loss in training using MobileNet. Figure 10 shows the confusion matrix when the skin lesion dataset was trained and classified by using the MobileNet model (Table 2).

This paper has compared three different models for classification of benign and malignant images in the given skin cancer images. Inception, ResNet, and MobileNet

Fig. 9 Accuracy and loss graph

Fig. 10 Confusion matrix

Confusion Matrix	TP	FP	FN	TN
MobileNet	22	15	8	15

Table 2 Comparison of models

Models	Sensitivity	Specificity	Precision	Negative predicitive value	False positive rate	False negative rate	Accuracy	$F1$ score
Inception v3	0.6667	0.4000	0.5263	0.5455	0.6000	0.3333	0.5333	0.5882
MobileNet	0.7333	0.5000	0.5946	0.6522	0.5000	0.5667	0.7177	0.6567
ResNet50	0.8667	0.0667	0.4815	0.3333	0.9333	0.1333	0.6967	0.6190

are the convolutional neural networks used for an image classification task. All of the models showed a similar and statistically significant performance. Among the CNN models discussed in this paper, ResNet model achieved the best results on our dataset with a sensitivity of 86.67%, followed by MobileNet with 73.37% and and Inception v3 with 66.67%. On the other hand, MobileNet has achieved 71.77% of accuracy, ResNet has achieved 69.67%, and Inception v3 has achieved 53.33%.

5 Conclusion and Future Scope

In this work, the skin cancer was classified based on skin lesions confirmed by pathology. They are trained by using the transfer learning-based models to classify it either as malignant or benign cancer. From the results, it is clear that the output is better when compared to the dermatologist suggestions. The better results can be obtained by fine tuning the network. In this work, the preprocessing method was not applied since the dataset was highly unbalanced but in the future, the preprocessing steps can be included to make the system more accurate. The model can be easily available through Web-based platform or even as API for assisting the dermatologists. In the future, more dermoscopy images can be added to the training dataset to make it better and more efficient. The dataset of different age groups and categories can be included to make the model more diverse in nature. Metadata of the images can be included for prediction to make it even more efficient. In the future, the personalized system can be developed based on the patient's medical history and other personal information.

References

1. M.A. Kadampur, S. Al Riyaee, Skin cancer detection: applying a deep learning based model driven architecture in the cloud for classifying dermal cell images, in *Informatics in Medicine Unlocked* (2020)
2. Kaggle Dataset, Skin Cancer: Malignant vs. Benign Processed Skin Cancer pictures of the ISIC Archive
3. J. Lemon, S. Kockara, T. Halic, M. Mete, Density-based parallel skin lesion border detection with webCL. BMC Bioinform. (2015)
4. K. Fukushima, Neocognitron: a hierachical neural network capable of visual pattern recognition. Neural Networks **1**(2), 119–130 (1988)
5. Y. LeCun, B.E. Boser, J.S. Denker, D. Henderson, R.E. Howard, W.E. Hubbard, L.D. Jackel, Handwritten digit recognition with a back-propagation network, in *Advances in Neural Information Processing Systems*, pp. 396–404 (1990)
6. H.A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, A. Ben Hadj Hassen, L. Thomas, A. Enk, L. Uhlmann, Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists
7. U.-O. Dorj, K.-K. Lee, J.-Y. Choi, M.J.M.T. Lee, The skin cancer classification using deep convolutional neural networks. Applications, pp. 1–16 (2018)

8. A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in *Advances in Neural Information Processing Systems*, pp. 1097–1105 (2012)

Particle Swarm Optimization Based on Random Walk

Rajesh Misra and Kumar Sankar Ray

Abstract Particle swarm optimization (PSO) undergoes several modifications since it was first proposed by J. Kennedy and R. Eberhart in 1995. There are many variants of PSO till date. These variants often confuse researchers about the improvement, applicability and novelty of those variants. As a result, in 2007 Daniel Bratton and James Kennedy defined a standard for particle swarm optimization which is the extension of the original algorithm with latest improvement. This latest improvement of Particle Swarm Optimization often known as SPSO when tested against baseline benchmark functions it fails to achieve required performance in non-separable and asymmetrical functions. Here, a new algorithm has been proposed to modify the canonical PSO algorithm by introducing the concept of random walk and particle reinitialization. Inherent nature of random walk balances the "exploration" and "exploitation" properties in search area and successfully handles the problem of trapping in local and global optima, which most of the algorithm fails to handle. This newly proposed variant clearly outperforms other well-known algorithms which belong to the category where velocity term is eliminated. Also, it achieves better performance when tested with same benchmark function where SPSO fails to perform well.

Keywords Random walk · Constrained biased random walk · Canonical PSO · Gaussian distribution

1 Introduction

Among the metaheuristic algorithms, PSO is most promising method. Particle swarm optimization is mostly influenced by the concept of the particles' interaction and their quick movements from one place to another in a large search region. Particle

R. Misra (✉) · K. S. Ray
Electronics and Communication Science Unit, Indian Statistical Institute, Kolkata, India

K. S. Ray
e-mail: ksray@isical.ac.in

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_13

position and velocity are two important components of PSO algorithm. As per social behavior of particles, velocities are changed and the particles get new positions by applying the velocity on current position. In 1998, Y. Shi and R. Eberhart proposed canonical PSO by adding *inertia weight* to the original PSO [1, 2], P_{best} and G_{best} are clearly explained in their papers. A very common limitation imposed by almost every researcher is that PSO weather it is canonical or original or some other variants stuck into local optima. Though it is a matter of debate whether researchers use proper topology as it is rightly pointed out by Blackwell and Kennedy in 2015 [3]. Though original PSO has two components velocity and position, in 2003, Kennedy proposed a new idea by eliminating velocity term named as bare bone PSO (BBPSO) [4]. BBPSO becomes much simpler method and is very easy to understand while keeping the same benefit intact with original PSO. Lot of variations over original or canonical PSO are developed over the last three decades, so much that it creates confusion which approaches to use and where. To overcome this critical issue, Bratton and Kennedy proposed a standard for particle swarm optimization, known as SPSO [5]. In 2011, SPSO got major improvement with adaptive random topology and rotational invariance, known as SPSO11 by Clerc [6]. It is recommended that any new PSO variant should be better than existing SPSO11 in terms of experiment on a difficult enough non-biased benchmark. In 2013, Mauricio Zambrano-Bigiarini et al. set a baseline for future PSO improvement by conducting experiment on a complex benchmark test suite [7]. When SPSO11 tested against baseline benchmark functions though it shows very good performance in unimodal and some of the multimodal functions, but for non-separable and asymmetrical functions it fails to achieve required performance. So at this juncture, SPSO11 which is the latest PSO according to the baseline benchmark function performance is employed. There is very less number of research focus on BBPSO than original/canonical (classical PSO). Our main focus is to develop a new algorithm which is “velocity free” and well performed with the latest benchmark defined in [7].

Shakya et al. [8] combine discrete and binary version of PSO by proposing bi-velocity PSO algorithm for solving multicast routing problems. Haoxiang et al. [9] use modified ACO method to solve routing protocol for vehicular network.

Our approach is to apply random walk in PSO particle movement. First, particle positions are initialized randomly over the search space. Then, rather than calculating velocity of each particle, probability functions between all particles are computed. Our chosen probability function is influenced by the work done by Blanchard and Volchenkov [10] which is based on biased random walks on undirected graphs. Utilizing this probability, a random coin toss is performed to decide which way corresponding particle should move. This random coin toss procedure is the basis of randomization of random walk method. In our algorithm, this random coin toss operation is like, if it is HEAD then “exploit,” if it is TAIL then “explore.” Thus, in both the cases, the benefit of searching the entire search space has been achieved for building more solution region (exploration) and if any particle finds a promising solution, then move toward that solution (exploitation). Most of the well-known algorithm fails to find the balance between “exploration” and “exploitation,” and they trap into local optima and perform poorly, whereas our algorithm handles this property efficiently. When an experiment is performed on various multimodal and unimodal functions as suggested in [7], it gives us extremely satisfactory results which not only outperform other variants PSO where velocity is not used and very much comparable to SPSO11 performance.

Paper organization is as follows: Simple random walk and biased random walk are discussed in Sect. 2. The proposed random walk PSO algorithm is detailed in Sect. 3. Experimental results are shown in Sect. 4. Section 5 draws conclusion.

2 Random Walk

2.1 Simple Random Walk

Karl Pearson first introduced the phrase random walk in 1905 [11]. Here, particle movement is based on random choice and expected progression of the particle is root-mean-square distance after n steps [12]. The simple random walk approach is discussed in [13].

Let us algorithmically explain simple random walk as follows.

Algorithm 1 Simple Random Walk

```
1: procedure SRandomWalk(Particle)
2: Toss ← Compute a Random Function    # like Tossing an Fair coin.
3:         if Toss = HEAD then
4:             Move Left
5:         else
6:             Move Right
7:         end if
8: end procedure
```

In the above procedure, particle movement is uniform where the random process (i.e., the coin tossing) is unbiased. There is also another version of random walk called biased random walk which is explained below.

2.2 Biased Random Walk

In simple random walk, movement of particles is influenced by equal probability. After n random steps, particle is close to the starting position. But according to the biased random walk, particle jumps from the current position to the next position with unequal probability. Let us take an example and consider a biased coin tossed and 80% chances coming head and 20% tail. If head indicates left direction movement and tail indicates right direction movement, then it is obvious that particle moves left more than right. After n number of steps, particles' expected position will be more left direction than right.

Algorithmically biased random walk [13] is explained as follows.

Algorithm 2 Biased Random Walk
1: **procedure** BiasedRandomWalk(Particle)
2: Toss ← Compute a Random Function # like Tossing an unfair coin.
3: **if** Toss = HEAD **then**
4: Move Left with uniform displacement
5: **else**
6: Move Right with uniform displacement
7: **end if**
8: **end procedure**

The above biased random walk procedure is based on an uneven coin tossing that outcomes with either maximum time HEAD or TAIL. So, the particles' movement will be more left or more right. As a result, the particle will not roam around its starting position. Particles basically move between regions. In every possible step, movement is uniform.

Another version of random walk also exists, called constrained biased random walk [13], algorithmically which is as follows.

Algorithm 3 Constrained Biased Random Walk
1: **procedure** ConstrainedBiasedRandomWalk(Particle)
2: Toss ← Compute a Random Function # like Tossing an unfair coin.
3: **if** Toss = HEAD **then**
4: Move Left with Non-uniform displacement # like +Δ or -Δ
5: **else**
6: Move Right with Non-uniform displacement # like +Δ or -Δ
7: **end if**
8: **end procedure**

3 Proposed Random Walk-Based PSO

3.1 General Idea

Our proposed approach is based on the idea of performing constrained biased random walk of particles over PSO topology. Let us first discuss the general framework of our approach by building of particle swarm optimization graph. This is discussed in Sect. 3.2. Then, probability of all particles is computed; but for that it has been required to calculate attribute values and weight parameters. Attribute value α_i is computed for each particle i. Calculation of α_i is explained in Sect. 3.3. Next, each particle i computes weight parameter A_{ij} with respect to all other particles j which is discussed in Sect. 3.4. Once A_{ij} is computed for each particle, finally probability function P_{ij}^{α} is computed for every particle connected to all other particles using α_i and A_{ij}. Probability function (P_{ij}^{α}) computation is discussed in Sect. 3.5. Once the probability computation has been done, each particle moves to the direction based on random coin toss operation. By direction it means a specific particle chosen as a target particle (P_{targ}) and our intended particle P_i jumps to the nearest location of P_{targ}. The procedure of choosing P_{targ} is by utilizing probability (P_{ij}^{α}) as mentioned earlier. A random number r, like a random coin, is picked up, if $r \leq \min$ (P_{ij}^{α}) particle. Then randomly, any particle is chosen as possible target particle P_{targ}. Otherwise, highest probability particle will be chosen as P_{targ}. After P_{targ} is chosen, a displacement vector k^d is calculated and added to the current position of the particle. Algorithmically, it has been summarized in Algorithm 4.

Algorithm 4 Random Walk PSO

procedure RANDOMWALKPSO(PARTICLE N)
    Construct PSO graph
    do
        Compute α_j
        for j **do**
            Calculate A_{ij}
            Calculate probability P_{ij}^{α} for all particles
            $\theta_{max} \leftarrow Max(P_{ij}^{\alpha})$ # [assigning Maximum probability]
            $\theta_{min} \leftarrow Max(P_{ij}^{\alpha})$ # [assigning Minimum probability]
            r >0 and r< 1
            if $r < \theta_{max}$ **then**
                $P_{trag} \leftarrow randomly\ pick\ any\ particle.$
            else
                $P_{trag} \leftarrow particle\ with \theta_{max}\ probability.$
                Compute k^d based on P_{trag} # *[See Equation - (5)]*
                $P_j \leftarrow P_j + k^d$ # Update new position.
            end if
        end for
    while minimum error or maximum iterations criteria is not attained
end procedure

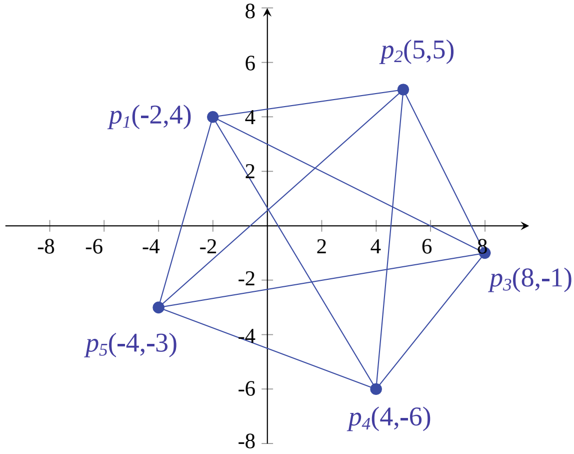


Fig. 1 PSO graph with 5 randomly initialized nodes

3.2 PSO Graph Construction G(P, E)

In particle swarm optimization graph $G\,(P, E)$, each particle is counted as node $p_i \in P$ and the Euclidean distance from p_j to p_i is an edge e_j that belongs to E. Entire PSO graph construction is discussed in [13]. The weight of each edge of the PSO graph is calculated as Euclidean distance between $p_i(x_i, y_i)$ and $p_j(x_j, y_j)$ as follows:

$$\left[p_i(x_i, y_i), p_j\left(x_j, y_j\right)\right] = \sqrt{\left(x_j - x_i\right)^2 + \left(y_j - y_i\right)^2}$$

Let us explain this PSO graph construction with a simple example. Imagine there are 5 particles randomly distributed in search domain where 2D is considered for simplicity. Suppose the particles are—$P_1(-2, 4)$, $P_2(5, 5)$, $P_3(8,-1)$, $P_4(4,-6)$, $P_5(-4,-3)$ as shown in Fig. 1. Each particle is connected with other particles by an edge.

3.3 Attribute Computation (α_i)

The purpose of computation of attribute values is that it is used in probability computation in later phase. A concept of rank has been used in attribute computation. A detailed discussion of attribute computation is available in [13]. Algorithmically, attribute computation method is as follows.

Algorithm 5 Attribute Computation of all nodes of the PSO graph G

```
procedure AttributeComputation (PSO graph G(P,E), rank (1:N))
  D [rank is a integer vector containing integer values from 1 to n.]
    for node ← 1to n do
        fitness(node) ←Compute fitness of all the nodes.
    end for
    Sort the fitness vector according to nodes fitness values.  #[From Highest fitness value to Lowest.]
        for j ← 1 to fitness(node) do
                α(node_j) ← rank(n)   #[node_j holds highest fitness value and rank(n) is the maximum rank from rank vector.]
            n ← n − 1                 #[rank(n) now holds next highest rank.]
            #[In next iteration node_{j+1} will indicate next highest fitness node and so on]
        end for
end procedure
```

There are various different optimization functions. For simplicity, an example where the particles are close to (0,0) coordinate is considered as best fitness achieving particles, as shown in Fig. 2. This fitness consideration is just for an example. As per Fig. 2, p1 achieves maximum rank $\alpha = 5$; similarly, other particles receive their respective ranks from rank vector *rank*(5) as there are 5 particles in our example.

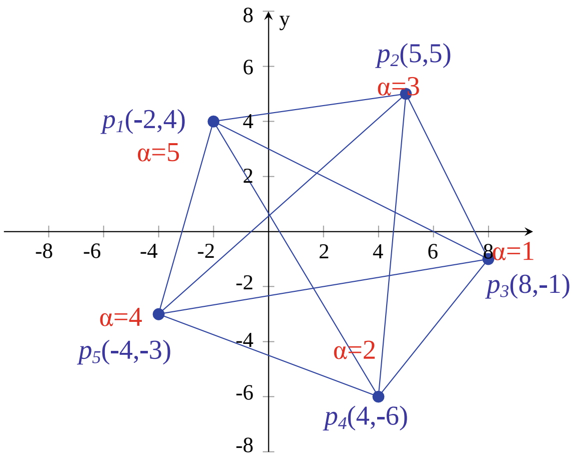
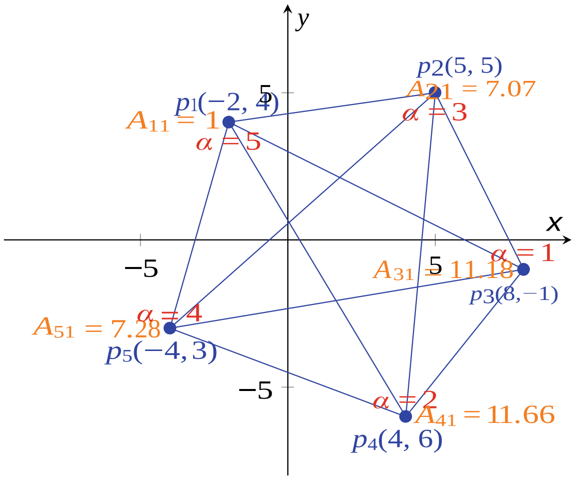


Fig. 2 Same graph with attribute value α

Fig. 3 PSO graph with A_{ij} [13]



3.4 Weight Parameter Calculation (A_{ij})

Weight parameter computation algorithm is provided in Algorithm 6. Detailed discussion is available in [13] (Fig. 3).

Algorithm 6 Weight Parameter Computation (Aij) on Attribute PSO Graph G_α(P,E)

procedure WEIGHTPARAMCAL (ATTRIBUTED PSO GRAPH G_ α (P,E))
  for $j \leftarrow$ 1to n **do**
    for $i \leftarrow$ 1to n **do**
      if j==1 **then**
        $A_{ij} \leftarrow 1$
      else
        $A_{ij} \leftarrow \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$
      end if
    end for
  end for
end procedure

3.5 Probability Computation (P^{α}_{ij})

Now all the parameters are available for the computation of probability, for each possible path from node p_j to p_i. This probability guides us where will be the position

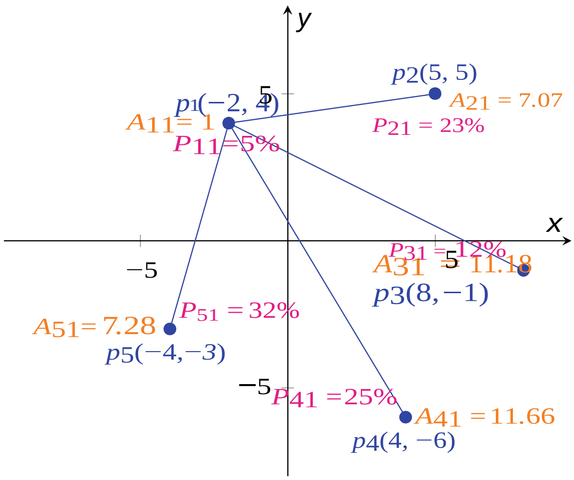


Fig. 4 Similar graph as stated above is considered for probability calculation P_{ij}^{α}

of the particle p_j. Probability computation function P_{ij}^{α} from node j to node i is as follows

$$P_{ij}^{\alpha} = \frac{\alpha_i A_{ij}}{\sum_k^{\infty} A_{kj}}. \tag{1}$$

Inspiration of using this above function comes from the work done by Blanchard et al. on "*Fair and Biased Random Walks on Undirected Graphs and Related Entropies.*" This probability function is effectively utilized in various graphs such as network graph and social graph where biased random walk can be applied. One of the notable works in this area is by Gómez-Gardeñes and Latora [14].

Figure 4 shows the graph after calculation of probability function. Once the calculation is done, each path from P_1 assigned a probability value which reflects the possibility of selecting that path. The end particle of the path becomes its target particle P_{targ} probability calculation of the above graph is as follows

$$P_{11} = \frac{5 \times 1}{1 \times 11.18 + 2 \times 11.66 + 3 \times 7.07 + 4 \times 7.28 + 5 \times 1} = \frac{5}{89.83} = 0.05$$

So, $P_{11} = 5\%$.

Similarly for other 4 nodes, we get

$$P_{21} = \frac{3 \times 7.07}{89.83} = \frac{21.21}{89.83} = 0.23 = 23\%$$
$$P_{31} = \frac{1 \times 11.18}{89.83} = \frac{11.18}{89.83} = 0.12 = 12\%$$
$$P_{41} = \frac{2 \times 11.66}{89.83} = \frac{23.32}{89.83} = 0.25 = 25\%$$

$$P_{51} = \frac{4 \times 7.28}{89.83} = \frac{29.12}{89.83} = 0.32 = 32\%$$

Based on the above calculations, particle p_5 gets highest probability and p_1 gets lowest probability.

3.6 Target Particle Selection (P_{targ})

Before going into much detail about p_{targ} selection, let us first discuss what happens in canonical PSO. Velocity and position in canonical PSO are calculated as follows

$$V_i(t) = w * V_i(t-1) + C_1 * R_1(P_{LB}(t) - X_i(t-1)) + C_2 * R_2(P_{GB}(t) - X_i(t-1)) \quad (2)$$

$$X_i(t+1) = X(t) + V_i(t+1) \quad (3)$$

where $V_i(t+1)$ is the velocity of particle i at time $(t+1)$ and $X_i(t+1)$ is the position of the particle i at time $(t+1)$. The inertia weight (w) is introduced by Shi and Eberhart in "*Parameter selection in particle swarm optimization.*" This variant of PSO is called canonical PSO. Inertia weight w tries to balance the exploration and exploitation capability of the particles. But yet, it fails to find balance in the position so that particle should not be trapped in local/global optima.

Now the question is it can get rid of velocity and still find the next position. The answer is given by Kennedy in his work velocity-free bare bone PSO (BBPSO). The next position of the particle in BBPSO is determined by the Gaussian distribution of the particles in the search space. But the Gaussian version is not as good as canonical version. Bare bone PSO suffers from premature convergence or converges to a point which is neither global nor local optimum. This is where our p_{targ} comes into play. Particle pi probabilistically finds a strong candidate among other particles which gives the best solution at that iteration. It is marked as p_{targ} and tries to jump toward that direction.

The next question is to choose ptarg and why it is appropriate. In our (α_i) calculation, the particle which is nearest to solution got the maximum value. In p_{ij} computation, $(\alpha_i \times A_{ij})$ receive maximum values that reflects which particle to select. That is why the particle holding highest probability value became p_{targ}. p_{targ} always gives us the local search ability that has more *exploitive* power. For solving separable function this *Exploitive* nature help particles to conver towards the solution very fast.

According to the random walk concept, a random process like picking a random card, coin toss or rolling dice decides the path of the particle. Similarly, a random number is chosen in our approach to decide the path selection of the moving particle p_i.

If p_i failed to select p_{targ}, then p_i randomly selects any other particle as p_{targ} and tries to jump to that position. By this approach, p_i explores more into the search region and successfully avoids trapping in local optima. This flexibility of random

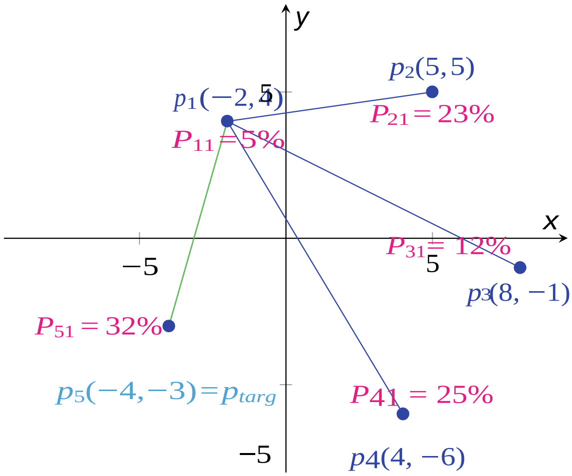


Fig. 5 Node p_5 on the other side of the green color edge is selected as P_{trag}

walk helps the particle to solve multimodal and complex optimization functions, as these functions need more exploration rather than exploitation. In Fig. 5, the p_{targ} particle has been chosen.

3.7 New Position Computation (P_{new})

After p_{targ}, p_i jumps toward that direction; but jump depends on random coin toss. The new position of the particle p_i is computed as follows

$$p^d_{\mathrm{new}} = p^d_i + k^d \tag{4}$$

where p^d_{new} is the new position of the particle p^d_i and k^d is the step count. d is the number of dimension which is based on application. k^d is calculated as follows

$$P(s,n) = p^{(n+s)/2}(1-p)^{(n-s)/2}\binom{n}{(n+s)/2} \tag{5}$$

s indicates no. of steps, and n is no. of coin toss. As the probability distribution $P(s, n)$ is no longer an even function of s, it has been expected that the particle will tend to drift toward p_{targ} at a steady rate of $(2p - 1)$ steps per coin toss. Applying the simplified Stirling formula to the above biased random walk result, it is found that

$$\log P(s,n) \approx -\left(\frac{n}{2}\right)\left[\left(1+\frac{s}{n}\right)\log\frac{1+s/n}{2p} + (1-s/n)\log\frac{1-s/n}{2(1-p)}\right] \tag{6}$$

By simplifying the above equation, we get

$$P(s,n) = \frac{e^{-(s-\overline{s})^2)(8p(1-p)n)}}{(8\pi p(1-p)n)^{1/2}} \tag{7}$$

So finally, k^d will be

$$k^d = \frac{e^{-(s-\overline{s})^2)(8p(1-p)))}}{(8\pi p(1-p)(P(s,n)))^{1/2}} \tag{8}$$

3.8 Particle Reinitialization

Though our algorithm perfectly balance local and global optimal search but as a failsafe approach a reinitialization method is incorporated. This method is simple, and a counter is kept to check whether all particles are trapped inside local optima. If no further improvement is seen, the particles are reinitialized. But here the trick is that particles are not reinitialized for the entire search space rather than the search region which is the “so-far-best-achieved” region. As all the particles already search the entire search area, there is no point to begin from the same region all over again. Our experimental results show the enormous benefits from this approach.

4 Experimental Setting

4.1 Benchmark Functions

The performance of our random walk PSO (RWPSO) algorithm is benchmarked against a set of functions defined in CEC 2013, named as *“Problem definitions and evaluation criteria for the CEC-2013 special session and competition on real-parameter optimization”* [15]. This set has been chosen because this is the latest benchmark functions for optimization and this set is tested with SPSO11, where author mentions a baseline for future PSO improvements. Three different sets of PSO variants—SPSO11, BBPSO and RWPSO (our proposed algorithm)—are considered for test comparison.

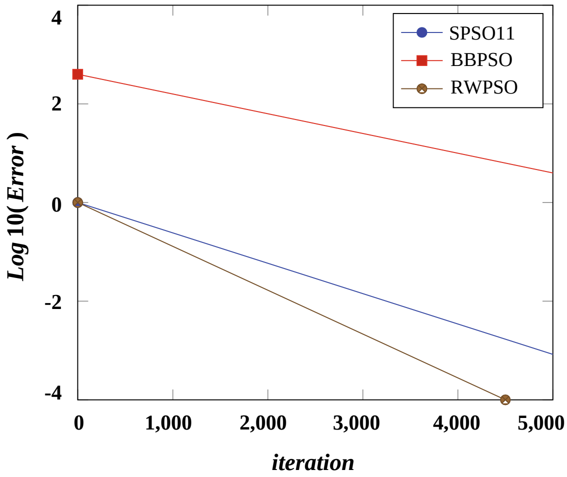


Fig. 6 Sphere function

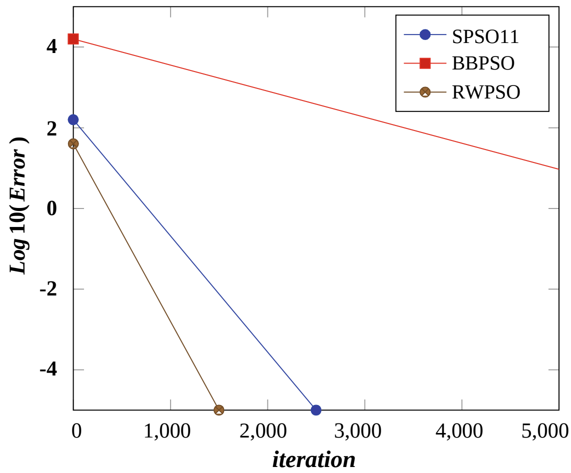


Fig. 7 Rotated discus function

Five functions are considered from [15]. They are "**sphere function**"**and** "**rotated discus function**" from unimodal function set, "**rotated Rosenbrock's function**" **and** "**Rastrigin's function**" from basic multimodal function set and "**composition function 1 (n = 5, Rotated)**" from composition function set. For detailed definitions of all these functions, please follow [15].

In our work, population size is maintained as 40. All these experimental details and experimental settings are taken from [15].

In Figs. 6, 7, 8, 9 and 10, 5 graphs are shown for 5 different optimization functions. Vertical axis of those graphs represents error, and horizontal axis represents iteration. It should be noted that this error parameter is based on Log10. Basically, this graph represents the function error values $[f(x) - f(x^*)]$ with respect to no. of iterations. For example, in sphere function graph (Fig. 6), SPSO11 algorithm takes more than 5000 iterations to get error value 0 and BBPSO takes lot more iterations than SPSO11 but still cannot reach error value 0, whereas RWPSO takes approximately 4500 iterations to reach 0 error values. These graphs help identify which algorithm

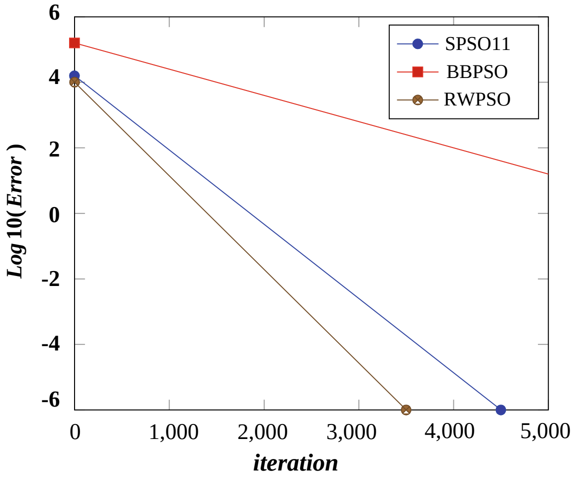


Fig. 8 Rotated Rosenbrock's function

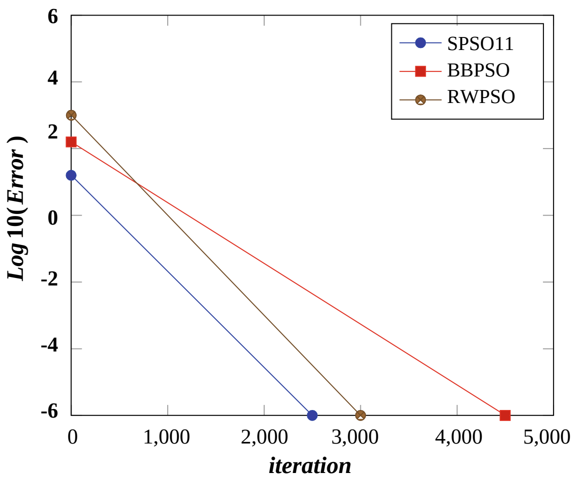


Fig. 9 Rastrigin's function

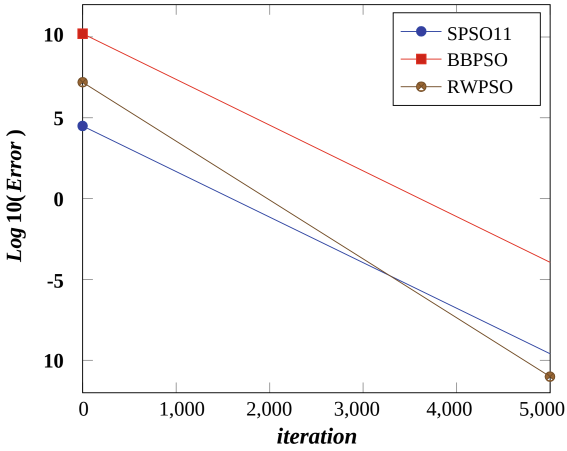


Fig. 10 Composition function 1 ($n = 5$, Rotated)

convergence is faster than others. By observing those 5 graphs, it is concluded that our proposed algorithm RWPSO is way better than BBPSO and very much competitive with SPSO11 in terms of convergence.

In Table 1, all 5 test function results are summarized. In first row, sphere function runs with 10, 20 and 30 dimensions. Three algorithms show their mean and standard deviation values for each dimension after 51 runs. If it has been observed closely, it is seen that RWPSO clearly outperforms BBPSO and is very close to SPSO11. In consecutive rows, other results of the functions have been shown.

5 Conclusion

This paper uses random walk approach with random toss coin operation and a particle reinitialization method, which prove the strength of the algorithm RWPSO of finding solution region in much faster way. Not only that, this algorithm proves its capability to reach optimal solution in various difficult functions. Here, the results of 5 optimization functions are presented from 3 categories. More functions are tested from [2], which will be reported elsewhere.

Table 1 Summary statistics for the 10, 30 and 50 dimensional case in fifty-one times experiments

Function category	Test function	Dimension	SPSO11 [6]	BBPSO [4]	RWPSO
			Mean ± Std	Mean ± Std	Mean ± Std
Unimodal functions	Sphere function	10	−1.400E+03±0.00E+00	5.20E+03 ± 2.67E+03	−1.400E+03 ± 0.00E+00
		30	−1.400E+03 ± 1.875E−13	6.19E+05 ± 1.23E+02	−1.400E+03 ± 0.78E−15
		50	−1.400E+03 ± 3.183E−13	9.41E+05 ± 5.28E+06	−1.400E+03 ± 2.78E−08
Unimodal functions	Rotated discus function	10	−1.100E+03 ± 4.556E+03	5.20E+03 ± 2.67E+03	−1.630E+03 ± 4.122E+03
		30	−1.100E+03 ± 6.702E+03	5.39E+05 ± 3.43E+04	−1.630E+03 ± 3.78E+05
		50	−1.100E+03 ± 8.717E+03	6.27E+06 ± 3.23E+06	−1.633E+03 ± 5.26E+05
Multimodal functions	Rotated Rosenbrock's function	10	−9.000E+02 ± 4.974E+00	−14.32E+03 ± 5.67E+01	−10.521E+03 ± 4.114E+01
		30	−9.000E+02 ± 2.825E+01	−15.16E+04 ± 5.13E+02	−10.931E+03 ± 3.25E+05
		50	−9.000E+02 ± 2.405E+01	−15.23E+04 ± 5.23E+06	−10.284E+03 ± 3.66E+04
Multimodal functions	Rastrigin's function	10	−4.000E+02 ± 5.658E+00	−6.62E+01 ± 7.24E+01	−4.125E+02 ± 4.213E+00
		30	−4.000E+02 ± 2.740E+01	−6.162E+02 ± 5.13E+01	−4.931E+01 ± 2.25E+01
		50	−4.000E+02 ± 4.183E+01	−8.23E+03 ± 5.73E+04	−4.42E+02 ± 3.89E+01
Composition function	Composition function 1 ($n = 5$, Rotated)	10	7.000E+02 ± 3.042E+02	12.73E+01 ± 2.66E+02	8.332E+02 ± 1.342E+00
		30	7.000E+02 ± 6.796E+01	−8.342E+02 ± 2.56E+01	−8.13E+01 ± 5.25E+01
		50	7.000E+02 ± 0.000E+00	12.23E+03 ± 5.73E+04	7.82E+01 ± 1.89E+00

References

1. Y. Shi, R. Eberhart, A modified particle swarm optimizer, in *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360).* Anchorage, AK, USA, pp. 69–73, (1998). https://doi.org/10.1109/ICEC.1998.699146;P. May, H.C. Ehrlich, T. Steinke, ZIB structure prediction pipeline: composing a complex biological workflow through web services, in Euro-Par 2006. LNCS, vol. 4128, ed. by W.E. Nagel, W.V. Walter, W. Lehner (Springer, Heidelberg, 2006), pp. 1148–1158
2. J. Kennedy, R. Eberhart, Particle swarm optimization, *Proceedings of ICNN'95 - International Conference on Neural Networks,* vol. 4. Perth, WA, Australia, 1995, pp. 1942–1948. https://doi.org/10.1109/ICNN.1995.488968
3. T. Blackwell, J. Kennedy, Impact of communication topology in particle swarm optimization. IEEE Trans. Evol. Comput. **23**(4), 689–702 (2019). https://doi.org/10.1109/TEVC.2018.2880894
4. J. Kennedy, Bare bones particle swarms, *Proceedings of the 2003 IEEE Swarm Intelligence Symposium. SIS'03 (Cat. No.03EX706)*, Indianapolis, IN, USA, 2003, pp. 80–87. https://doi.org/10.1109/SIS.2003.1202251
5. D. Bratton, J. Kennedy, Defining a standard for particle swarm optimization, in *2007 IEEE Swarm Intelligence Symposium*, Honolulu, HI, 2007, pp. 120–127. https://doi.org/10.1109/SIS.2007.368035
6. M. Clerc, Standard particle swarm optimisation (2012). ffhal-00764996f
7. M. Zambrano-Bigiarini, M. Clerc, R. Rojas, Standard particle swarm optimisation 2011 at CEC-2013: a baseline for future PSO improvements, in *2013 IEEE Congress on Evolutionary Computation, Cancun*, 2013, pp. 2337–2344. https://doi.org/10.1109/CEC.2013.6557848
8. S. Shakya, L.N. Pulchowk, A novel bi-velocity particle swarm optimization scheme for multicast routing problem. IRO J. Sustain. Wireless Syst. **02**, 50–58 (2020)
9. W. Haoxiang, S. Smys, Qos enhanced routing protocols for vehicular network using soft computing technique. J. Soft Comput. Paradigm (JSCP) **1**(2), 91–102 (2019)
10. P. Blanchard, D. Volchenkov, Fair and biased random walks on undirected graphs and related entropies, in *Towards an Information Theory of Complex Networks*, ed. by M. Dehmer, F. Emmert-Streib, A. Mehler (2011)
11. K. Pearson, The problem of the random walk. Nature, p. 294 (1905). https://doi.org/10.1038/072294b0-1905
12. E.W. Weisstein, Random walk–1-dimensional. https://mathworld.wolfram.com/RandomWalk1-Dimensional.html (2002)
13. R. Misra, K.S. Ray, A modification of particle swarm optimization using random walk. arXiv:1711.10401v2[cs.AI] (2017), https://arxiv.org/abs/1711.10401v2
14. J. Gómez-Gardeñes, V. Latora, Entropy rate of diffusion processes on complex networks. Phys. Rev. E pp. 2339–2348 (2008) https://doi.org/10.1103/PhysRevE.78.065102
15. J.J. Liang, B.-Y. Qu, P.N. Suganthan, A.G. Hernández-Diaz, Problem definitions and evaluation criteria for the CEC 2013 special session and competition on real-parameter optimization, Technical Report 201212, Computational Intelligence Laboratory. Zhengzhou University, Zhengzhou China and Technical Report, Nanyang Technological University, Singapore, Tech. Rep., Jan 2013, https://www.ntu.edu.sg/home/EPNSugan/indexfiles/CEC2013/CEC2013.htm. Last accessed 12 Feb 2013

Signal Processing Algorithms Based on Evolutionary Optimization Techniques in the BCI: A Review

Ravichander Janapati, Vishwas Dalal, N. Govardhan, and Rakesh Sengupta

Abstract Brain–computer interfaces (BCIs) collect, analyze and transform brain signals into commands that are linked to desirable tasks involved by target equipment. The feature extraction process is used to deploy an alternate solution interpretation of signal obtained, which made way a collection of BCI actions more effectively. Pre-processing phase involving re-referencing of electrodes, deterioration, normalization, size reduction and removal of artifacts, etc. is often used before feature extraction. The classification of features was examined in this paper, and evolutionary technique is applied to BCI. The implications of the paper can be helpful for academicians, researchers and scientists in this domain to quickly understand the previous work in this field. These algorithms can be used in various applications such as the classification of motor imagery tasks, filter banks, etc.

Keywords Brain–computer interface · Feature extraction · Evolutionary technique · Motor imagery · Classification · Signal processing algorithms

1 Introduction

The most rapidly growing theme in neuroimaging is brain–computer interface (BCI) [1]. Their purpose is to translate for different purposes, including in automation, communications and entertainment, biological signals derived from different areas of the brain [2, 3]. It has the most common use of amyotrophic lateral sclerosis (ALS), the impairment of leg and so on for the recovery of people affected by paralysis [4, 5].The basic module for the BCI includes: signal pretreatment, extraction and

R. Janapati (✉)
ECE Department, S R Engineering College, Warangal, India
e-mail: ravi_chander_j@srecwarangal.ac.in

V. Dalal · R. Sengupta
Cognitive Science Department, S R Engineering College, Warangal, India

N. Govardhan
ECE Department, SRIT(W), Warangal, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_14

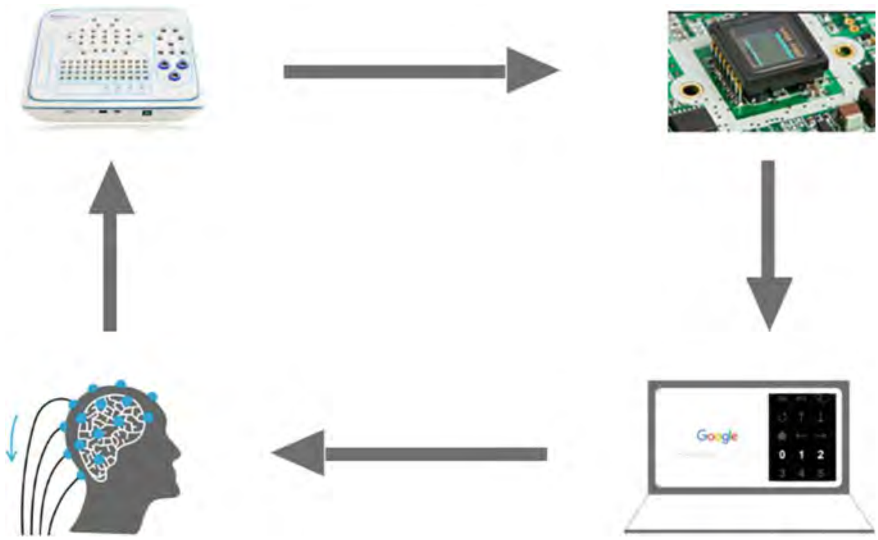


Fig. 1 BCI-based Web interface navigation

classification. The following steps are taken. The classified results produce the control signals needed to drive an aid. High dimensionalism of features and the collection of specific features are the key focus of BCI study, so they can be discriminated against as strongly as possible [6]. It is also found that the exactness of the classifier is significantly reduced due to having a large number of irrelevant data in the form factor. Another module along with the after the module, based on feature selection, is introduced throughout this context. PCA [7], SVD [8] and ICA [9] are widely used feature selection techniques. MI EEG signals detection with SVM and PSO. In traditional classifiers, the nuclear specifications are typically selected as per the verifiable evidence, overlooking the significance of refining the function to enhance the output of a classifier. A PSO algorithm has been applied to boost this by choosing the best kernel and limitation variables and thereby enhancing classification accuracy. Combination of PSO and SVM used to diagnose and other sleep disorders [10] and to refine the SVM for EMG signals classification [11]. Figure 1 shows the BCI interface.

2 Literature Survey

To survey the literature regarding EEG-based BCIs, we explore IEEE, Web of Science, Scopus, Science direct and Google scholar. Among the many articles that met our search criteria, we first excluded those that had nothing to do with BCI research. The articles were screened, based on title, abstract and number of citations to only include studies involving evolutionary-based BCI for signal processing applications. This screening exercise has yielded 21 articles as listed in Table 1, which presents the diverse evolutionary-based BCI for signal processing applications. Initially, it is investigated from reputed journals for the last decade, i.e., 2010–2020, by considering parameters such as algorithms used, signal processing function and applications. Table 1 shows the evolutionary algorithms used in BCI signal processing applications.

3 Literature Review and Discussion

This study discusses the revised version of the M-PSO PSO algorithm in order to improve the EEG and ECoG signal feature collection. Four BIAs and function extraction are included in the study to search for relevant features providing information on BCI systems. This technique was intended to increase the precision of classification. Therefore, a current optimal solution was developed that uses the SF to control parameters and to determine reliable metrics for the best aspects. Designers introduced the SVM classification and the proposed algorithm to compare our outcomes.

For all aspects, the majority of the EEG or ECoG datasets exceed the algorithm. Our findings show that it is possible to use another proposed algorithm to pick fewer electrodes that are substantially helpful in building a BCI system [12]. Based on these characteristics, the gain and bandwidth of the filter are built and adjusted while integrating harmonic SSVEP responses as well. Not only does this technique improve accuracy, but it also increases the number of commands obtainable by facilitating the use of stimulus frequencies to evoke poor SSVEP responses. The findings indicate the ability of bio-inspired architecture to be expanded to include additional SSVEP features (e.g., time-domain waveform) for future BCIs based on SSVEP [13].

This article provides a novel technique based on a clustering algorithm for feature selection. The proposed strategy is validated on a dataset using power spectral density as the function and artificial bee colony as the clustering algorithm. The results thus obtained have validated our argument that when the dataset is condensed to partly of its unique size, enclose typically the pertinent features, an increase in accuracy is observed. Around the same time, the computational complexity was also reduced [14]. Selecting the EEG channels used to establish the training predictor has an effect on the output of the classifier. Current findings on the actual dataset of both metaheuristic techniques when the classifier is a Bayesian predictor. We equate those outputs significantly with a random set of EEG networks. This method greatly improves the precision of the training indicator, per the empirical findings [15] with bio-inspired optimization for feature selection and classifier enhancement to boost the precision rate of the MI-BCI, and a new CSP\AM-BA-SVM approach is suggested. Optimal various time selections for each topic are added to the proposed method. The parameters are taken from EEG signal with the popular spatial pattern (CSP). By using a one-vs-one approach, binary CSP is extended to multi-class issues. This method presents the application of the bat optimization algorithm (BA) and the hybrid

Table 1 Evolutionary-based algorithms used in BCI

Reference No.	Algorithm	Signal processing function	Application
[12]	Ant colony optimization (ACO), genetic algorithm (GA), cuckoo search algorithm (CSA) and modified particle swarm optimization (M-PSO)	Feature selection	SVM classifier
[13]	Artificial bee colony (ABC) cluster algorithm	Feature selection for motor imagery EEG data	Classification
[14]	Bio-inspired filter banks (BIFB)	Frequency detection method	Filter
[15]	Genetic algorithm and simulating annealing	Choose the input features of a classifier	classifier
[16]	Hybrid asttractor metagene (AM) algorithm along with the bat optimization algorithm (BA)	Choose the majority discriminant CSP features and optimize SVM parameters	Classifier optimization
[17]	Bio-inspired filter banks (BIFBs)	Feature extraction stage	Filter banks
[7]	A modified genetic algorithm (GA) wrapped around a support vector machine (SVM) classifier	Classification	Classification
[13]	Artificial bee colony (ABC) cluster algorithm	Reduce the features	Classification
[18]	Hybrid GA-PSO-based K-means clustering technique	Feature selection, blind source separation	Classification of two class motor imagery tasks
[19]	Quantum behaved particle swarm optimization	Feature extraction	Classification
[20]	And multilevel hybrid PSO-Bayesian linear discriminant analysis	Channel selection and feature selection	Classification
[21]	CPSO-based TWSVM classifier combined with CSP	Classification	Classification of MI electroencephalography (EEG)
[22]	PSO by linear discriminant analysis	Selecting optimal channels	Classification accuracy
[23]	Classic algorithm and the culling algorithm	Reduce the huge space of features extracted from raw electroencephalography	Feature selection

(continued)

Table 1 (continued)

Reference No.	Algorithm	Signal processing function	Application
[24]	Modified particle swarm optimization (PSO)	Feature selection	Classification accuracy
[25]	PSO-RBFN	Feature selection	Classifier
[26]	Incremental quantum particle swarm optimization (IQPSO) algorithm	Classification	Classification
[27]	Particle swarm optimization (PSO)-based neural network (NN)	Projected the features into a neural network	IoT applications
[28]	Ring topology-based particle swarm optimization (RTPSO) algorithm	Feature extraction	Classifiers
[29]	Choquet integral along with PSO algorithm	Classifier	For MI recognition

attractor metagene (AM) algorithm to pick the most discriminating CSP attributes and optimize SVM specifications [16].

This article developed a new SSVEP detection framework that includes advantage of SSVEPs' implicit genetic features. In the extracting features phase, the BIFBs collect frequency specificity, topic specificity and harmonic SSVEP responses and boost class separability. The proposed approach is tested on two online available test sets and exceeds many known detection algorithm. The BIFBs particularly show potential in the high frequency band, wherever SNR is small. This approach therefore not only improves the ITR of an SSVEP-based BCI, but may also boost comfort conditions owing to little visual exhaustion. The outcome shows the potential of bio-inspired design, as well as further SSVEP features will be extended to include the results [17]. Blind source isolation of the EEG signals in specific contributes to improved classification precision leading to spectral energy conversion. In addition, the use of specific features subsets instead of the full set of features can be helpful even to advanced classifiers like a support vector machine. When searching for functional subsets exacerbates the risk that the classifier will fit into the tests used to train the BCI, a range of different methods exist to minimize the risk and can be evaluated during relevant feature searches. The selection of features is showing promise inquiry for signal processing in BCIs, because the subject-specific features could be used off-line to obtain optimal online output [7].

This research aims to enhance classification accuracy and reduces the redundant and irrelevant features of a dataset. In order to minimize attributes and achieve their subsequent precision, they implement artificial bee colony (ABC) cluster optimizations [13]. K-means clustering methods based on hybrid GA-PSO was employed to differentiate between two groups of motor imaging (MI) activities. The proposed hybrid GA-PSO clustering centered on the segmentation of the K-means cluster has been developed to perform better the genetic algorithm (GA) and particulate swarm optimization (PPSO)-centered k-media, which implies the grouping of methods, in terms of accuracy and execution time. That performance appraisal on TFRs is derived, and the definition of event-related synchronization (ERD) and ERD (ERD) is established [7].

This develop a new evolved optimization algorithms for optimizing the discovery channels, incorporating popular spatial patterns for the collection of functions and supporting the identification of vector machines for the optimum quantum behaved particle swarm. Experimental findings demonstrated that, relative to the common spatial pattern method with maximum channels in raw datasets, the new binary quantum-completed particle swarm optimization method surpassed all other three main spatial pattern methodologies: the substantial decrease in the error classification and the number of channels. A motor imagery-based brain–computer system can be enhanced by the current technique considerably [13]. Propose an effective channel and selection, channel and feature-based signal processing architecture focused on particle swarm optimization (PSO). For the optimization and classification of characteristics, modified Stockwell transforms have been used, and multimodal PSO-Bayesian Linear discriminatory analyses implemented. The findings demonstrate that perhaps the channel selection system will speed up convergence to the optimal global level and minimize time training. As the proposed framework can boost classification efficiency, decrease the number of features effectively and significantly reduce test time, it can act as a guidance for relevant BCI application system analysis in real time [18].

In this report, we suggested a new CPSO-based classification system for TWSVM in conjunction with CSP feature extraction technique for the identification of MI events. In order to encourage the signal–noise ratio, the adaptive reduction method has been adopted. For multi-channel EEG signal generation, CSP has been used. For classification outcome of the MI BCI system, the specification of the TWSVM variable is very critical. In addition, the classification TWSVM surpassed the fastest total length of the CPU. In CPSO SVM, the analysis showed a slight change in the PSO TWSVM system again. An extensive BCI database analysis revealed that perhaps the CPSO TWSVM classification performs superior overall exposed to different machine learning approaches, which have been commonly use for established MI recognition research [19].

Authors propose the enhancement of the CSP PSO optimization technique. The whole research investigates the selection by the linear discriminant analysis of optimal channels amongst all channels and compares distinction correctness among CSP and CSP with PSO. This article suggests an optimum channel selection approach by BPSO rather than using all channels. Researchers used a smaller number of channels with greater precision as demonstrated above with the proposed procedure. In addition, developers are changing the fitness function because every time the calculation takes place the same system enhancement configuration [20]. The GA for functional choice in the BCI domain has been established with smelting entities, so it performs function removal in an unregulated way. This mode is extremely significant in the BCI domain because readjustment (i.e., to be carried out after a few thousand activities) is not controlled. Even then, BCI systems are not limited to the realm of potential optimization implementations. In reality, the optimization can be used in any investigation in which it is appropriate choose from a much bigger main number of characteristics a limited number of genes [21].

In order to categorize topic emotional states based on EEG signals researchers presented an enhanced algorithm and merged such feature selection to develop an online brain–computer (BCI) online emotion detection application. In particular, various functional attributes have been identified from the time domain, from the frequency domain and from the time frequency field. Instead, a multi-level linearly decreasing inertia weight (MLDW) modified particle swarm optimization (PSO) approach for selecting features was used. The MLDW method can also be used to optimize the inertia weight reduction method effectively. After this, the forms of emotions are listed by the vector cluster [22]. This study addresses an analysis of the implementations of different classification algorithms including the selected classification PSO-RBFN. Three well-known clustering algorithms are comparable to one another by their ability to minimize the loss function, to measure the effects of clustering algorithms on RBFN classification results. Better clusters acquired with the PSO algorithm can now be inferred by helping RBFN enhance its classification efficiency. In relation to two other classifiers, the PSO-RBFN can meet or exceed FFSVC and IFFSVC for most datasets. Whenever the PSO classification is applied to the two other classification models. the extendability and the easy implementation and learning of a PSO-RBFN classifier to deal with a supervised learning problem allow it desirable for just a PSO-RBFN classifier in near real time for an EEG classification application [23].

In this research introduced using the incremental IQPSO method for the gradual classification of the EEG data stream for the very first stage. IQPSO constructs an effective classifier as a set of laws, based on conceptual symbolic depiction of experience and increased comprehension. The suggested algorithm profits concurrently from gradual training and quantum-oriented improvements. Compared IQPSO 's output with ten other classifiers that were used on two EEG datasets for BCI systems. The findings indicate that even in addition to performing linear classification, specificity and retrieval IQPSO is superior to other classificators. In addition, it indicates appropriate online learning time consumption [24]. It is proposed an advanced neural network (NN), which provides a clear—headed among BCI devices and IoT, based upon particle swarm optimization (PSO) to resolve above well-mentioned issues. Researchers built a system architecture in the tests, which collected specific EEG data using the particle swarm optimization and then transformed the attributes into a neural network training system. The experimental results showed that the PSO-based NN methodology is feasible for the precision of 98.9% in the classification of the engine images (MI) activities [25].

This article suggests an R3PSO methodology with two classifier approaches for measuring performance that will be utilized in an EEG motor imaging classification to determine parameters. In organize to evaluate the effectiveness of the study classification models, FFNN, SS-SVM, SMO-SVM and QP-SVM are being used as classifieds and as a tenfold CV and holdout process. The process of extraction of the function uses certain wavelet coefficient statistical techniques [26]. Throughout this research, they establish an advanced and efficient BCI framework for the purposes of MI detection that uses SBCSP, LDA and the fuse method. In addition, they achieved the maximum AUC while using quadratic additive with PSO as a nuclear fission level prediction model. By using a portable EEG collector, Mindo 4S, and the based algorithms, functional BCIs could be built in real-world situations [27]. The results of three mental task classifications indicate greater FPSOCM-ANN precision in multiclass (SVM, LDA and linear perceptron) classification systems in both classes and time windows. The findings of these three mental tasks are unique. While the classification performance of clinicians is decreased, this is enhanced by the time periods, the better precision at a window period of 7 s [28].

4 Conclusion

This paper has examined some of the more recent studies that employ EA-based methods to enhance the analysis of EEG signal. The paper has concentrated on the observations such as (1) dimensional reduction by feature or electrode reduction, (2) promote classifier learning process and (3) benefits of EA-based methods in traditional approach assembly. Such a hybrid system will take advantage of features based on EA to optimize the ensemble's required parameters and boost the overall performance. To the best of our knowledge, in EEG and BCI research, such a device has not been used yet. This paper offers an investigation of the EA-based methods such as decomposition, filtering process to minimize the sizes of the function or electrode collection. This is a multi-objective topic in which it is beneficial to minimize the size of sets of features or electrodes while preserving the precision of the classification.

References

1. R. Abiri et al., A comprehensive review of EEG-based brain–computer interface paradigms. J. Neural Eng. **16**(1), 011001 (2019)
2. S. Bhattacharyya et al.,Interval type-2 fuzzy logic based multiclass ANFIS algorithm for real-time EEG based movement control of a robot arm. Rob. Autonomous Syst. **68**, 104–115 (2015)
3. K. Takahashi, T. Nakauke, M. Hashimoto, Remarks on hands-free manipulation system using bio-potential signals from simple brain-computer interface, in *2006 IEEE International Conference on Systems, Man and Cybernetics*, vol. 2 (IEEE, New York, 2006)
4. V. Suma, Computer vision for human-machine interaction-review. J. Trends Comput. Sci. Smart Technol. (TCSST) **1**(02), 131–139 (2019)
5. V. Bindhu, An enhanced safety system for auto mode E-vehicles through mind wave feedback. J. Inform. Technol. **2**(03), 144–150 (2020)
6. J. Atkinson, D. Campos, Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. Expert Syst. Appl. **47**, 35–41 (2016)

7. D.A. Peterson et al., Feature selection and blind source separation in an EEG-based brain-computer interface. EURASIP J. Adv. Sign. Process. **19**, 218613 (2005)
8. A.F. Cabrera, D. Farina, K. Dremstrup, Comparison of feature selection and classification methods for a brain–computer interface driven by non-motor imagery. Med. Biol. Eng. Comput. **48**(2), 123–132 (2010)
9. K. Li et al., Single trial independent component analysis for P300 BCI system, in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (IEEE, New York, 2009)
10. Y. Maali, A. Al-Jumaily, A novel partially connected cooperative parallel PSO-SVM algorithm: study based on sleep Apnea detection, in *Proceedings of the IEEE Congress on Evolutionary Computation (CEC '12)* (IEEE, Brisbane, Australia, June 2012), pp. 1–8
11. A. Subasi, Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders. Comput. Biol. Med. **43**(5), 576–586 (2013)
12. O.P. Idowu, P. Fang, G. Li, Bio-inspired algorithms for optimal feature selection in motor imagery-based brain-computer interface, *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (IEEE, New York, 2020)
13. P. Rakshit et al., Artificial bee colony based feature selection for motor imagery EEG data, in *Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012)* (Springer, India, 2013)
14. A.F. Demir, H. Arslan, I. Uysal, Bio-inspired filter banks for SSVEP-based brain-computer interfaces, in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)* (IEEE, New York, 2016)
15. S. Basterrech et al., Nature-inspired algorithms for selecting EEG sources for motor imagery based BCI, in *International Conference on Artificial Intelligence and Soft Computing* (Springer, Cham, 2015)
16. S. Selim et al., A CSP\AM-BA-SVM approach for motor imagery BCI system. IEEE Access **6**, 49192–49208 (2018)
17. A.F. Demir, H. Arslan, I. Uysal, Bio-inspired filter banks for frequency recognition of SSVEP-based brain–computer interfaces. IEEE Access **7**, 160295–160303 (2019)
18. P. Tiwari, S. Ghosh, R.K. Sinha, Classification of two class motor imagery tasks using hybrid GA-PSO based-means clustering. Comput. Intell. Neurosci. **2015** (2015)
19. L. Zhang, Q. Wei, Channel selection in motor imaginary-based brain-computer interfaces: a particle swarm optimization algorithm. J. Integrat. Neurosci. **18**(2), 141–152 (2019)
20. Y. Qi et al., Channel and feature selection for a motor imagery-based BCI system using multilevel particle swarm optimization. Comput. Intell. Neurosci. **2020** (2020)
21. L. Duan et al., Recognition of motor imagery tasks for BCI using CSP and chaotic PSO twin SVM. J. China Univer. Posts Telecommun. **24**(3), 83–90 (2017)
22. J.-Y. Kim et al., A binary PSO-based optimal EEG channel selection method for a motor imagery based BCI system, in *International Conference on Hybrid Information Technology* (Springer, Berlin, Heidelberg, 2012)
23. I. Rejer, Genetic algorithms for feature selection for brain–computer interface. Int. J. Pattern Recognit. Artif. Intell. **29**(05), 1559008 (2015)
24. Z. Li et al., Enhancing BCI-based emotion recognition using an improved particle swarm optimization for feature selection. Sensors **20**(11), 3028 (2020)
25. E. Cinar, F. Sahin, New classification techniques for electroencephalogram (EEG) signals and a real-time EEG control of a robot. Neural Comput. Appl. **22**(1), 29–39 (2013)
26. K. Hassani, W.-S. Lee, An incremental framework for classification of EEG signals using quantum particle swarm optimization, in *2014 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)* (IEEE, New York, 2014)
27. O.P. Idowu et al., Efficient classification of motor imagery using particle swarm optimization-based neural network for IoT applications, in *2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT* (IEEE, New York, 2020)

28. H. Mirvaziri, Z.S. Mobarakeh, Improvement of EEG-based motor imagery classification using ring topology-based particle swarm optimization. Biomed. Sign. Process. Control **32**, 69–75 (2017)
29. T.-Y. Hsieh et al., Developing a novel multi-fusion brain-computer interface (BCI) system with particle swarm optimization for motor imagery task, in *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (IEEE, New York, 2015)
30. S.-L. Wu et al., Fuzzy integral with particle swarm optimization for a motor-imagery-based brain–computer interface. IEEE Trans. Fuzzy Syst. **25**(1), 21–28 (2016)

Cancelation of 50 and 60 Hz Power-Line Interference from Electrocardiogram Using Square-Root Cubature Kalman Filter

Roshan M. Bodile and T. V. K. Hanumantha Rao

Abstract Heart is the most vital organ in the human body, generating a systematic time-varying signal due to its electrical activity is called as an electrocardiogram (ECG). However, the acquisition process of ECG adds sufficient amount of unwanted artifacts in the clean ECG. Out of these artifacts, power-line interference (PLI) commonly corrupts the ECG signal. Therefore, in this paper, the dynamical model joint with a square-root cubature Kalman filter (SR-CKF) is proposed to remove 50 and 60 Hz PLI from ECG. The SR-CKF is tested on an arrhythmia database with an input signal to noise ratio (SNR) of −10 to 10 dB. Results obtained after denoising show that the SR-CKF performs better in terms of keeping original content or diagnostic information of the ECG, less mean square error (MSE), and achieving higher SNR output compared to the discrete wavelet transform (DWT) and notch filter.

Keywords Electrocardiogram · Dynamical model · SR-CKF · Discrete wavelet transform · Power-line interference · Notch filter

1 Introduction

An electrical activity produced at atria is followed by ventricles that generate PQRST waves, and these sequences of waves are vital for clinical purposes. The heart activity can be fast, slow, abnormal, or normal, and hence, the ECG signal is more important clinically. The detection of any abnormalities in rhythm or heart rate at an early stage can save a person's life. The heart signal is weak, and during acquisition, noise contamination like electromyogram, motion noise, baseline wander, power-line interference (PLI), etc. masks the clean ECG signal. Out of these artifacts, the

R. M. Bodile (✉) · T. V. K. Hanumantha Rao
Department of ECE, National Institute of Technology, Warangal, India

T. V. K. Hanumantha Rao
e-mail: tvkhrao75@nitw.ac.in

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_15

PLI commonly corrupts the ECG signal. So, to extract the original morphology of the ECG, the elimination of PLI present in the ECG recordings is essential.

An overview of the literature that relates to the work is presented here. In early work, the digital filtering methodology [1] was used to remove PLI by designing a bandpass filter with a cut-off frequency of 50 Hz. Similarly, to eliminate PLI, the use of fewer taps in finite impulse response (FIR) filter [2, 3] is suggested. The FIR technique presented claims that it is more effective than a traditional FIR filter for attenuating PLI. The notch filter [4, 5] is another classical method for removing PLI from noisy ECG. However, the design of high quality factor is challenging and also susceptible to the artifact and can quickly become stable to unstable. Hence, the adaptive notch filters are desirable, which can adapt according to selective frequency and noise or signal contents. In the literature, various adaptive notch filters are suggested that can manipulate the quality factor adaptively. Apart from these methodologies, the techniques based on the decomposition [6–9] such as DWT, packet wavelet, etc. are famous for removing 60 Hz PLI. However, these methods are not suitable for high noise contamination situations. Apart from these techniques, the DWT with adaptive dual thresholding is also suggested by researchers [10]. The gradient-based adaptive filters with and without references [11–13] are used to eliminate PLI from the ECG signal. These adaptive filters require some samples to follow the original morphology of the ECG after starting transient. According to the electricity standard, the variation in the PLI frequency is trivial, and hence, the PLI variation in ECG processing is also negligible. Therefore, the methods based on tracking and cancelation of PLI using Kalman filter, smoother, and extended Kalman filters [14–19] are found more flexible and stable. The Kalman filter with the dynamical model is more adaptable to eliminate either 60 or 50 Hz PLI and also its harmonics in a boisterous environment.

This paper has both 60 and 50 Hz PLI frequency that corrupts the ECG signal in range -10 to 10 dB input SNR. The dynamical model in [14] is more flexible and does not require any prior assumptions, and it merely depends on the quality factor, which is the ratio of noise covariances. Therefore, this model is combined with the SR-CKF [20], which is a numerically, more stable method. This work is also focused on high noise contamination that masks the ambulatory arrhythmia ECG recordings.

The remainder of the paper is organized as follows: In Sect. 2, the dynamical model is briefly described and the proposed method with time update and measurement steps. Data details, quantitative assessment, and 50 and 60 Hz filtering results are provided in Sect. 3. The last section gives the overall conclusion of the work.

2 Proposed Method

2.1 Dynamical Model

The single-tone PLI noise contamination can be assumed as some sinusoids with random phase and amplitude:

$$x_n = A\cos(2\pi * n * f_0 / f_s + \phi) \tag{1}$$

where A, n, f_0, f_s, and ϕ are amplitude, time index, PLI noise, sampling frequency, and phase, respectively. After adding the model error parameter (η_n) and some trigonometric manipulation [14] in Eq. (1), it is expressed as:

$$x_{n+1} + x_{n-1} = 2 * \cos(2\pi * f_0 / f_s)x_n + \eta_n. \tag{2}$$

It is seen that PLI noise does not have any rapid fluctuations in phase and amplitude. However, the addition η_n makes the model more flexible; hence, it is desired to add η_n in the SR-CKF approach.

The clean ECG corrupted by PLI noise can be assumed as a concoction of PLI, clean ECG, and other unwanted noise or signal; that is,

$$y_n = w_n + x_n \tag{3}$$

where quantities x_n and w_n represent PLI and zero mean arbitrary term, indicating all signals and noises apart from PLI. However, it neglects the fact that w_n it can contain biosignals and other noises, as this work is considering only 50 and 60 Hz PLI. The tracking of PLI noise is possible when Eqs. (2) and (3) are converted into the state-space form [14] to apply the SR-CKF approach, and it is given as:

$$\begin{cases} x_{n+1} = Cx_n + d\eta_n \\ y_n = a^T x_n + w_n \end{cases} \tag{4}$$

where $C = \begin{bmatrix} 2\cos(2 * \pi * f_0 / f_s) & -1 \\ 1 & 0 \end{bmatrix}$, $x_n = \begin{bmatrix} x_n \\ x_{n-1} \end{bmatrix}$, $d = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, and $a = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

The above model in Eq. (4) is now ready to apply on noisy ECG using the SR-CKF approach. After tracking PLI, the filtered ECG signal can be obtained by merely subtracting this tracked PLI from noisy ECG. The block diagram of the ECG denoising scheme is shown in Fig. 1.

2.2 Square-Root Cubature Kalman Filter

Let us assume multi-dimensional integral [20] of the following form

$$I(f) = \int_{R^n} f(x) * e^{(-x^T x)} \mathrm{d}x \tag{5}$$

where $f(x)$ and R^n be arbitrary function and domain of integration. Integral in Eq. (5) is hard to calculate in a general form. Therefore, using the cubature rule, it is

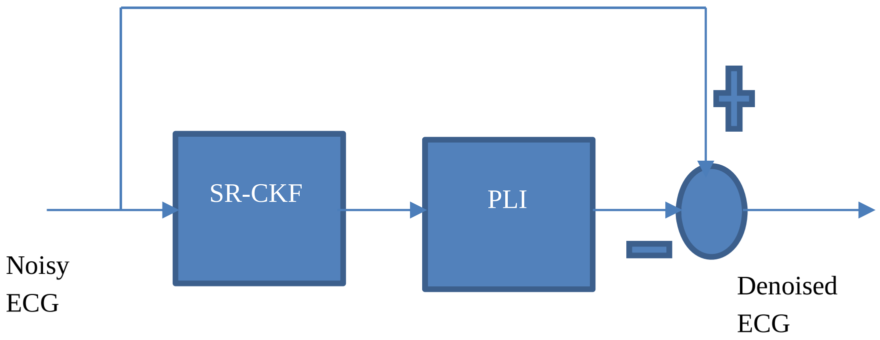


Fig. 1 ECG denoising scheme

approximated into points and corresponding weights [21]

$$I(f) = \sum_{j=1}^{m} \omega_j f\left(\zeta_j\right) \tag{6}$$

where ω_j and ζ_j are the jth cubature points and their corresponding weights. Here, the third-degree cubature rule is considered, and hence, total support points are equal to (2*state dimension). Figure 2 shows the cubature points and their corresponding weights for the two-dimension system. After this approximation and calculation, the steps required for the SR-CKF [20–22] are:

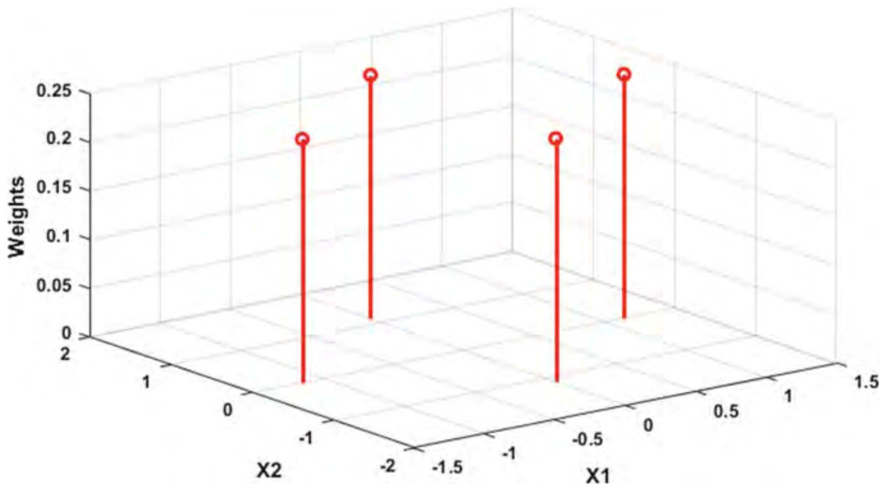


Fig. 2 Cubature points and their corresponding weights

Time update

- Compute the cubature points (CP).

$$\mathrm{X}^{j}(k-1|k-1) = S(k-1|k-1)J(j) + \overline{x}(k-1|k-1), \tag{7}$$

$j = 1, 2, \ldots, 2*$ state dimension.

where $\begin{cases} \sqrt{n}[1]_j,\ j = 1, 2, ..., \text{state dimension} \\ \sqrt{n}[1]_{j-n},\ j = 1, 2, ..., 2 * \text{state dimension} \end{cases}$

- After calculation of CP, evaluate the propagated CP state dimension.

$$\mathrm{X}^{j\cdot}(k-1|k-1) = g\big(k-1, \mathrm{X}^{j}(k-1|k-1)\big),\ j = 1, 2, \ldots, 2 * \text{state dimension} \tag{8}$$

- Let Tri $(\cdot)$ is triangularization algorithm, now compute the prior state and its square-root of the covariance matrix:

$$\overline{x}(k|k-1) = \frac{1}{m} * \sum_{j=1}^{m} \mathrm{X}^{j\cdot}(k-1|k-1), m = 2* \text{ state dimension}, \tag{9}$$

$$S(k|k-1) = Tri\big(\big[\chi^{\cdot}(k|k-1), S_Q(k-1)\big]\big), \tag{10}$$

where $\chi^{\cdot}(k|k-1) = \frac{1}{\sqrt{m}}\begin{bmatrix} \mathrm{X}^{1\cdot}(k|k-1) - \overline{x}(k|k-1)... \\ \mathrm{X}^{m\cdot}(k|k-1) - \overline{x}(k|k-1) \end{bmatrix}, j = 1, \ldots, m.$

Measurement

- Estimate CP.

$$\mathrm{X}^{j}(k-1|k-1) = S(k-1|k-1)J(j) + \overline{x}(k-1|k-1), j = 1, \ldots, m. \tag{11}$$

- After calculation of CP, evaluate the propagated CP.

$$\mathrm{X}^{j\cdot\cdot}(k-1|k-1) = \gamma\big(k, \mathrm{X}^{j}(k-1|k-1)\big),\ j = 1, 2, \ldots, 2 * \text{ state dimension} \tag{12}$$

- Now, compute the prior measurement and its square-root of the covariance matrix:

$$\overline{z}(k|k-1) = \frac{1}{m} * \sum_{j=1}^{m} \mathrm{X}^{j\cdot\cdot}(k-1|k-1), m = 2*, \text{ state dimension}, \tag{13}$$

$$S_{zz}(k|k-1) = Tri([Z(k|k-1), S_R(k-1)]), \tag{14}$$

where $Z(k|k-1) = \frac{1}{\sqrt{m}}\begin{bmatrix} \mathrm{X}^{1\cdot\cdot}(k|k-1) - \overline{z}(k|k-1)\ldots \\ \mathrm{X}^{m\cdot}(k|k-1) - \overline{z}(k|k-1) \end{bmatrix}, j = 1, \ldots, m.,$

- Calculate the cross-covariance matrix.

$$S_{xz}(k|k-1) = X(k|k-1)Z^T(k|k-1), \tag{15}$$

where $X(k|k-1) = \frac{1}{\sqrt{m}}\begin{bmatrix} \mathrm{X}^{1\cdot}(k|k-1) - \overline{x}(k|k-1)\ldots \\ \mathrm{X}^{m\cdot}(k|k-1) - \overline{x}(k|k-1) \end{bmatrix}, j = 1, \ldots, m.,$

- Now, estimate the Kalman gain.

$$G(k) = \left(S_{xz}(k|k-1)/S_{zz}^T(k|k-1)\right)/S_{zz}(k|k-1). \tag{16}$$

- Finally, calculate the posterior state and its square-root of the covariance matrix.

$$\overline{x}(k|k) = \overline{x}(k|k-1) + G(k)(z(k) - \overline{z}(k|k-1)), \tag{17}$$

$$S(k|k-1) = Tri([X(k|k-1) - G(k)Z(k|k-1), G(k)S_R(k-1)]) \tag{18}$$

3 Results and Discussion

This section summarizes and discusses the main findings of the work. The performance of the proposed approach is tested on publicly available the MIT-BIH arrhythmia database [23], and ECG records are considered as −101, 103, and 217, respectively. The sampling frequency of these ECGs is 360 Hz, and 3600 samples are assessed for filtering and comparing purposes. The PLI noises present in the ECG can either 50 or 60 Hz as it varies from country to country. Therefore, this work considered both 50 and 60 Hz PLI noise. The performance of the SR-CKF scheme is measured using output SNR [24],

$$\text{Output SNR} = \frac{\sum_{l=1}^{L} s^2(l)}{\sum_{l=1}^{L} \{s(l) - \hat{s}(l)\}^2}, \tag{19}$$

and MSE

$$\mathrm{MSE}_l = \frac{1}{L}\sum_{l=1}^{L}\left(s(l) - \hat{s}(l)\right)^2 \tag{20}$$

where quantities $s(l)$ and $\hat{s}(l)$ are clean ECG and filtered ECG, respectively. Here, the results of the proposed method are compared with a notch filter (infinite impulse response) and DWT methods. The notch filter and DWT-based techniques are very competitive while removing PLI under noisy conditions. Hence, the SR-CKF framework is compared with the DWT and notch filter techniques. The proposed work and comparative methods simulations are performed in MATLAB.

3.1 60 Hz PLI Denoising Results

The filtering results are compared quantitatively using output SNR and qualitatively using reconstructed signal after denoising. A visual representation of this denoising result can be seen here in Fig. 3. Figure 3a is a clean ECG; after adding 60 Hz PLI noise to ECG record 217 at -10 dB input SNR, the noisy signal is depicted in Fig. 3b. All the filtering results are displayed in Fig. 3c–e. The DWT method at low SNR clearly shows the presence of 60 Hz PLI noise, even though DWT is a powerful denoising tool. On the other hand, the second comparative method is the notch filter removes the 60 Hz PLI noise from the noisy ECG. From Fig. 3d, it is also observed that there is a reduction in R-peaks amplitude for given samples. The filtering result produced by the proposed method is depicted in Fig. 3e. It is observed from Fig. 3e that the proposed method is removed 60 Hz PLI noise from ECG, and also it followed the morphology of the ECG signal very closely. Results obtained after denoising show that this method performs better in keeping original content or diagnostic information of the ECG compared to the DWT and notch filter. To further examine this result, output SNR values for each method were analyzed quantitatively. From Table 1, it is observed that the comparative methods yield less output SNR compared to the proposed approach. Quantitatively, the SR-CKF yields average output SNR of 23.67–23.80 dB, the notch filter provides average output SNR of 15.79–15.81 dB, and the DWT gives average output SNR of 12.60–22.69 dB. Moreover, the MSE performance also indicates that the SR-CKF achieved a minuscule difference between the estimated and clean ECG signal and shown in Fig. 4.

3.2 50 Hz PLI Denoising Results

In the previous section, 60 Hz PLI noise removal results are shown. Here, it considers the removal of 50 Hz noise from noisy ECG. Figure 5a is a clean ECG; after adding 50 Hz PLI noise to ECG record 103 at 0 dB input SNR, the noisy signal is depicted in Fig. 5b. Based on Fig. 5 and Table 2 results, the following observations are made. The

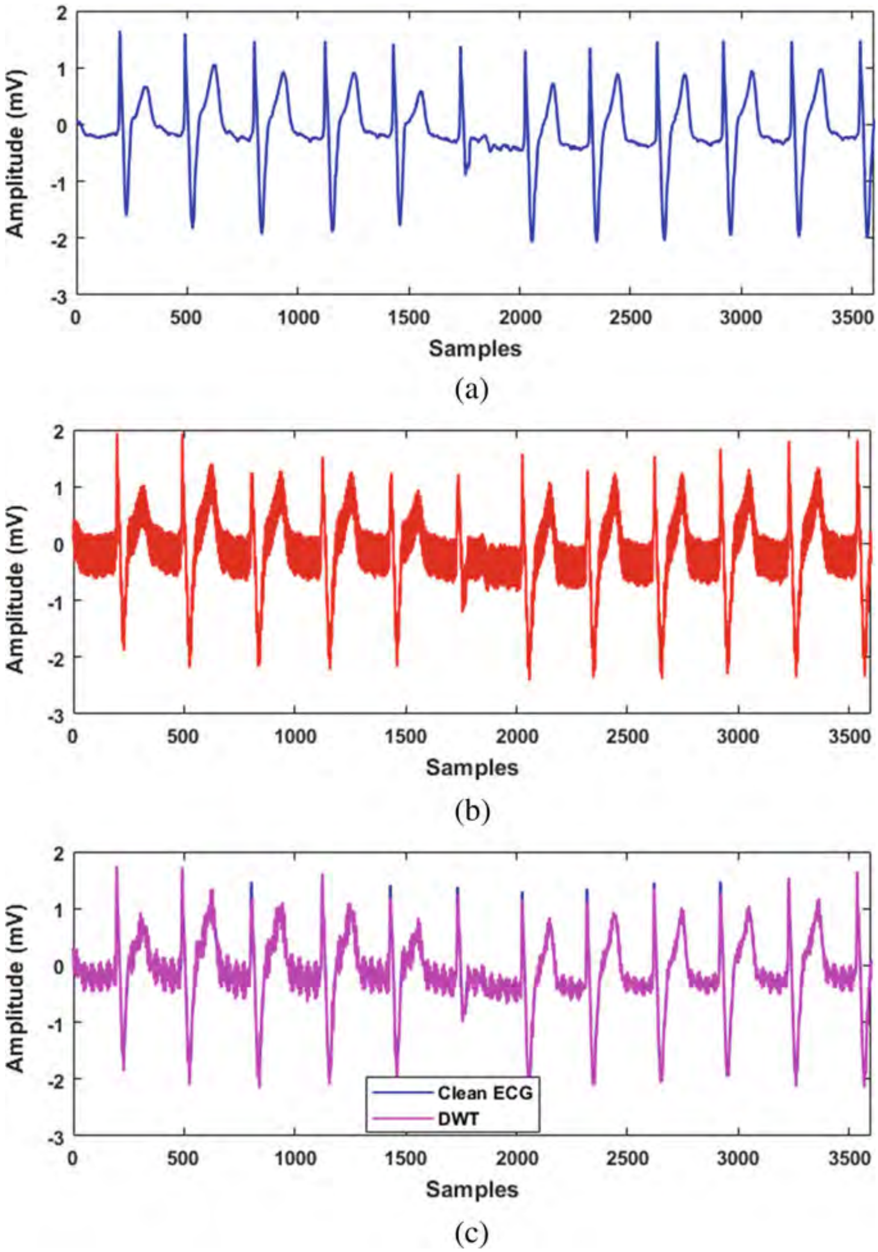
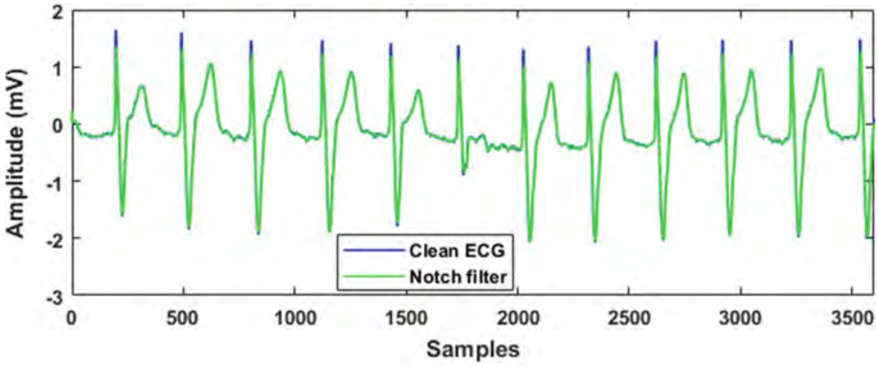
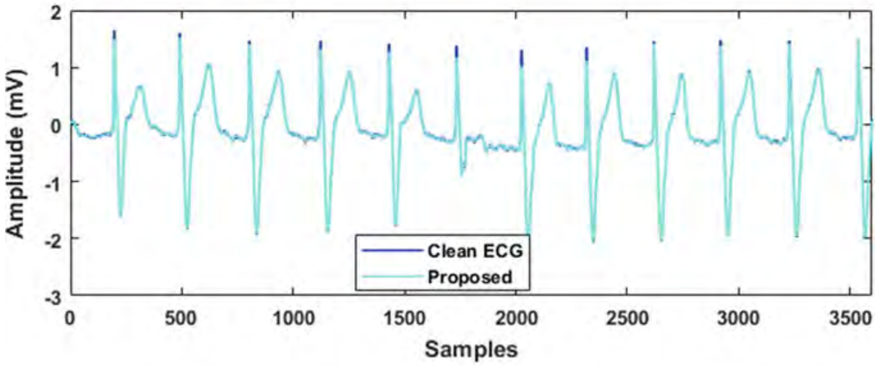


Fig. 3 60 Hz removal denoising results for ECG 217 at −10 dB. **a** Clean ECG, **b** noisy ECG, **c** DWT, **d** notch filter, and **e** proposed



(d)



(e)

Fig. 3 (continued)

DWT method in Fig. 5c at 0 dB SNR clearly shows the presence of 50 Hz PLI noise; though the fact that PLI is mild still it can alter the diagnostic information in the ECG. On the other hand, the notch filter method nearly removes the 50 Hz PLI noise from the noisy ECG. From Fig. 5d, it is also observed that there is a reduction in R-peaks amplitude for given ECG record 103. We can see from Fig. 5e that the SR-CKF is removed 60 Hz PLI noise from ECG, and also it followed the morphology of the ECG signal very closely. Likewise, results obtained after denoising show that the SR-CKF method performs better in keeping original content or diagnostic information of the ECG compared to the DWT and notch filter. Finally, quantitative results are compared in terms of the output SNR. The SR-CKF technique exhibits excellent performance in almost all cases. Quantitatively, the SR-CKF yields average output SNR of 21.85–21.86 dB, the notch filter provides average output SNR of 14.62–14.64 dB, and the DWT gives average output SNR of 10.19–22.06 dB. Moreover, the MSE performance also indicates that the SR-CKF achieved a minuscule difference between the estimated and clean ECG signal and shown in Fig. 6.

Table 1 Output SNR for ECG signals 101, 103, and 217

Output SNR				
Input SNR	101	103	217	Avg.
DWT				
−10	12.64	11.30	13.87	12.60
−5	15.63	15.35	18.92	16.64
0	19.09	18.36	21.98	19.81
5	20.57	20.15	22.67	21.13
10	22.48	21.99	23.60	22.69
Notch filter				
−10	16.14	12.90	18.32	15.79
−5	16.15	12.91	18.33	15.80
0	16.16	12.91	18.33	15.80
5	16.17	12.92	18.33	15.81
10	16.17	12.92	18.33	15.81
Proposed				
−10	23.10	23.74	24.17	23.67
−5	23.27	23.76	24.21	23.75
0	23.34	23.77	24.23	23.78
5	23.37	23.77	24.24	23.79
10	23.39	23.76	24.25	23.80

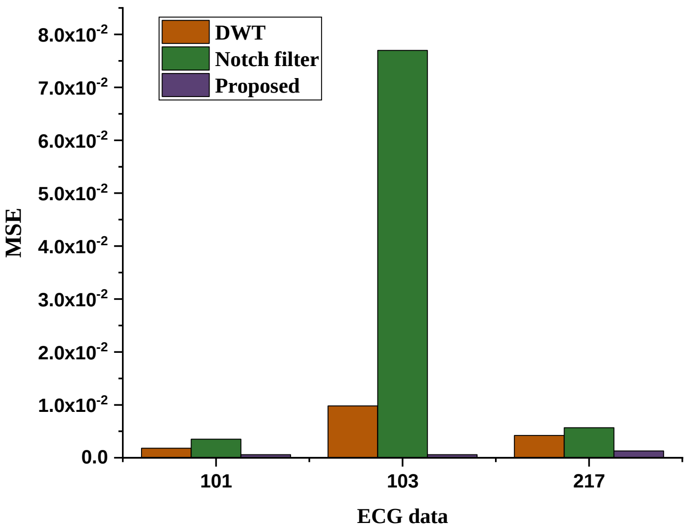
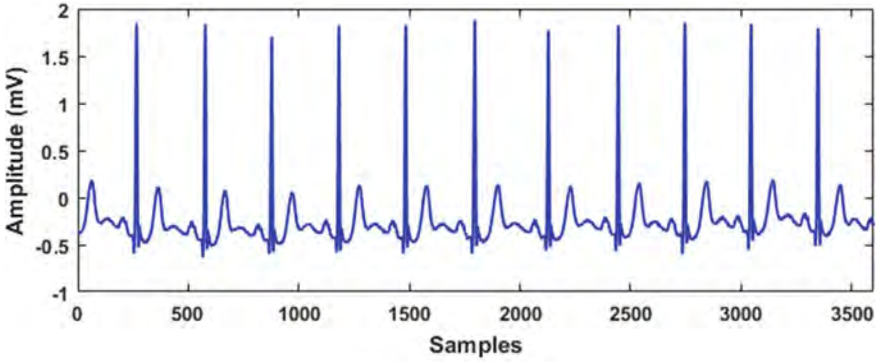
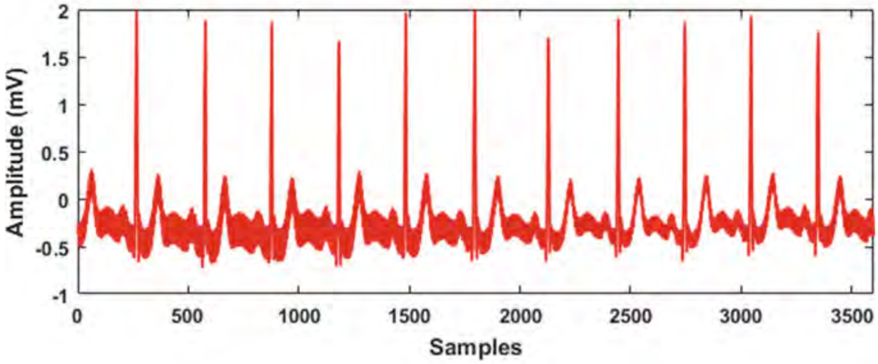


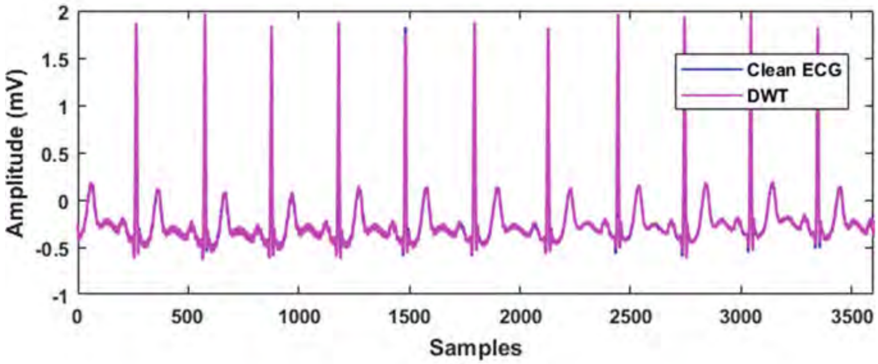
Fig. 4 Average MSE for different ECGs (in case of 60 Hz PLI removal condition)



(a)

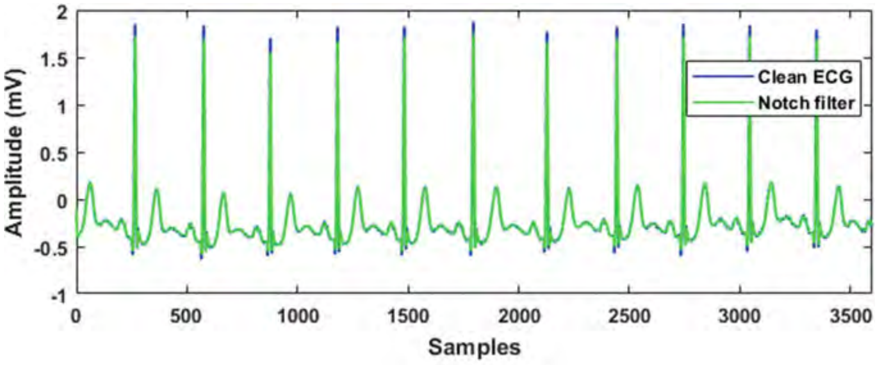


(b)

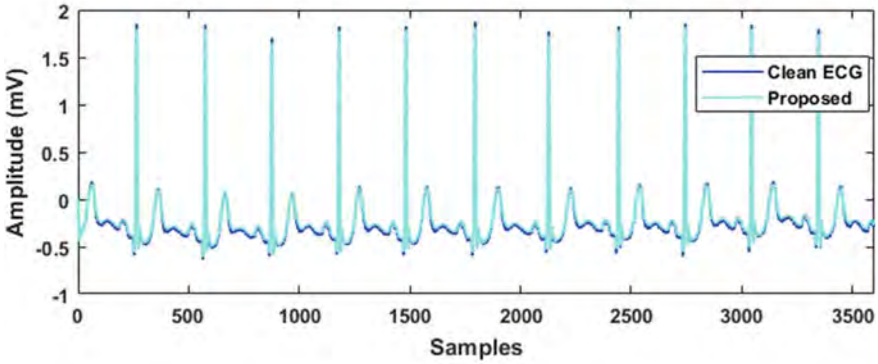


(c)

Fig. 5 50 Hz removal denoising results for ECG 103 at 0 dB. **a** Clean ECG, **b** noisy ECG, **c** DWT, **d** notch filter, and **e** proposed



(d)



(e)

Fig. 5 (continued)

4 Conclusion

In this paper, the dynamical model joint with the SR-CKF is proposed for eliminating 50 and 60 Hz PLI from ECG. The dynamical model is based on PLI data (50 Hz or 60 Hz), and the SR-CKF tracks PLI to remove from noisy ECG to preserve clinical information available in the ECG signal. The SR-CKF is tested on an arrhythmia database (ECG records—101, 103, and 217) with an input SNR of -10 to 10 dB. Results obtained after denoising show that the SR-CKF performs better in terms of keeping original content or diagnostic information of the ECG, minimum MSE, and achieving higher output SNR (for 50 Hz 21.85–21.86 dB and for 60 Hz 23.67–23.80 dB) compared to the DWT and notch filter.

Table 2 Output SNR for ECG signals 101, 103, and 217

Output SNR				
Input SNR	101	103	217	Avg
DWT				
−10	5.08	12.58	12.92	10.19
−5	10.60	16.97	15.37	14.31
0	13.46	20.75	17.16	17.12
5	18.70	21.30	20.37	20.12
10	21.71	23.18	21.29	22.06
Notch filter				
−10	14.94	11.62	17.29	14.62
−5	14.96	11.63	17.30	14.63
0	14.97	11.63	17.30	14.63
5	14.97	11.64	17.30	14.64
10	14.97	11.64	17.30	14.64
Proposed				
−10	21.53	22.06	21.95	21.85
−5	21.60	22.10	21.90	21.87
0	21.62	22.12	21.87	21.87
5	21.62	22.13	21.85	21.87
10	21.62	22.13	21.84	21.86

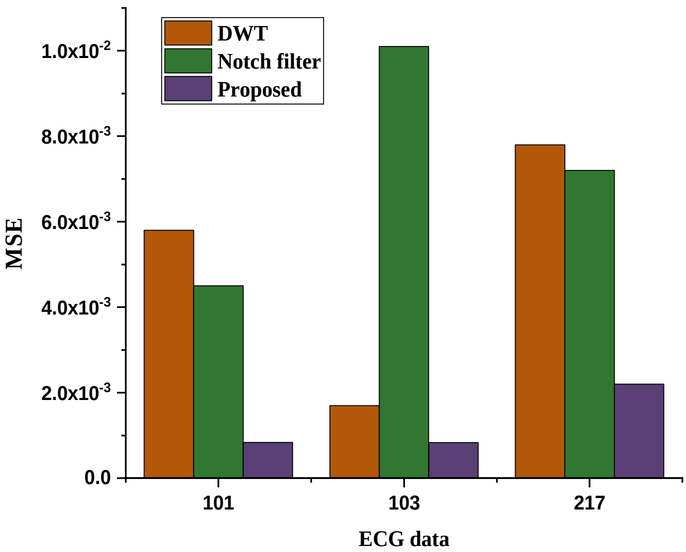


Fig. 6 Average MSE for different ECGs (in case of 50 Hz PLI removal condition)

References

1. M. Kunt, H. Rey, A. Ligtenberg, Preprocessing of electrocardiograms by digital techniques. Sign. Process **4**, 215–222 (1982)
2. J.A. Van Alste, T.S. Schilder, Removal of base-line wander and power-line interference from the ECG by an efficient FIR filter with a reduced number of taps. IEEE Trans. Biomed. Eng. **12**, 1052–1060 (1985)
3. R. Warlar, C. Eswaran, Integer coefficient bandpass filter for the simultaneous removal of baseline wander, 50 and 100 Hz interference from the ECG. Med. Biol. Eng. Comput. **29**, 333–336 (1991)
4. P. Tichavsky, A. Nehorai, Comparative study of four adaptive frequency trackers. IEEE Trans. Sign. Process. **45**, 1473–1484 (1997)
5. M. Sedlacek, J. Blaska, Low uncertainty power-line frequency estimation for distorted and noisy harmonic signals. Measurement **35**, 97–107 (2004)
6. S. Poornachandra, N. Kumaravel, A novel method for the elimination of power line frequency in ECG signal using hyper shrinkage function. Digit. Sign. Proc. **18**, 116–126 (2008)
7. Z. German-Sallo, ECG signal baseline wander removal using wavelet analysis, in *International Conference on Advancements of Medicine and Health Care through Technology* (Springer, Berlin, Heidelberg, 2011), pp. 190–193
8. B. El, O. Latif, R.K. Elmansouri et al., ECG signal performance de-noising assessment based on threshold tuning of dual tree wavelet transform. Biomed. Eng. Online **16**, 1–18 (2017)
9. D.L. Donoho, De-noising by soft thresholding. IEEE Trans. Inf. Theory **41**, 613–627 (1995)
10. W. Jenkal, R. Latif, A.D. Ahmed Toumanari, O. El B'charri, F.M.R. Maoulainine, An efficient algorithm of ECG signal denoising using the adaptive dual threshold filter and the discrete wavelet transform. Biocybern. Biomed. Eng. **36**, 499–508 (2016)
11. N. Razzaq, S.A.A. Sheikh, M. Salman, T. Zaidi, An intelligent adaptive filter for elimination of power line interference from high resolution electrocardiogram. IEEE Access **4**, 1676–1688 (2016)
12. Q. Wang, X. Gu, J. Lin, Adaptive notch filter design under multiple identical bandwidths. AEU—Int. J. Electron. Commun. **82**, 202–210 (2017)
13. J. Lin, X. Sun, J. Wu, S.C. Chan, W. Xu, Removal of power line interference in EEG signals with spike noise based on robust adaptive filter, in *IEEE Region 10 Conference, TENCON*; Singapore, pp. 2707–2710 (2016)
14. R. Sameni, A linear Kalman notch filter for power-line ınterference cancellation, in *The 16th CSI International Symposium on Artificial Intelligence and Signal Processing*, pp. 604–610 (2012)
15. P. Dash, R. Jena, G. Panda, A. Routray, An extended complex Kalman filter for frequency measurement of distorted signals. IEEE Trans. Instrum. Meas. **49**, 746–753 (2000)
16. A. Routray, A. Pradhan, K. Rao, A novel Kalman filter for frequency estimation of distorted signals in power systems. IEEE Trans. Instrum. Meas. **51**, 469–479 (2002)
17. L.D. Avendano-Valencia, L.E. Avenda, J. Ferrero, C.G. Castellanos-Dominguez, Improvement of an extended kalman filter power line interference suppressor for ECG signals, in *Computers in Cardiology*, pp. 553–556 (2007)
18. L.D. Avendano-Valencia et al., Reduction of power line interference on ecg signals using Kalman filtering and delta operator (2007)
19. G. Warmerdam, R. Vullings, L. Schmitt, J. Van Laar, J. Bergmans, A fixed-lag Kalman smoother to filter power line interference in electrocardiogram recordings. IEEE Trans. Biomed. Eng. **64**, 1852–1861 (2017)
20. I. Arasaratnam, S. Haykin, Cubature Kalman filters. IEEE Trans. Autom. Control **54**, 1254–1268 (2009)
21. D. Jianmin, S. Hui, L. Dan, H. Yu, Square root cubature Kalman filter-Kalman filter algorithm for ıntelligent vehicle position estimate. Proc. Eng. **137**, 267–276 (2016)
22. L. Xi, Q. Hua, Z. Jihong, Y. Pengcheng, Maximum correntropy square-root cubature Kalman filter with application to SINS/GPS integrated systems. ISA Trans. **80**, 195–202 (2018)

23. A.L. Goldberger, L.A.N. Amaral, L. Glass et al., PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation **101** 215–220 (2000)
24. K. Kærgaard, S.H. Jensen, S. Puthusserypady, A comprehensive performance analysis of EEMD-BLMS and DWT-NN hybrid algorithms for ECG denoising. Biomed. Signal Process. Control **25**, 178–187 (2016)

A Comprehensive Study on the Arithmetic Operations in DNA Computing

V. Sudha and K. S. Easwarakumar

Abstract The computer has become a part of human life, and fast computation is vital. Addition, subtraction, multiplication, and division are the fundamental mathematical operations. Most of the arithmetic operations performed on a computer are realized using these basic operations. Among the four, addition and subtraction operations form the basis, as the other two operations can be realized by using these procedures. This paper focuses on the algorithms proposed for implementing addition and subtraction operations in DNA computing.

Keywords DNA computing · Arithmetic operations · Representation · Reusability · Biological operations

1 Introduction

Computing is the process of performing certain operations with the help of a computer. DNA computing is a form of molecular computing, replacing silicon with DNA. DNA computing makes DNA the computational medium, providing solutions for problems that cannot be solved through normal architecture. Adleman [1] initiated DNA computing by solving an instance of the Hamiltonian path problem (HPP). He got this idea by mapping the finite control in the turing machine to the polymerase enzyme. The polymerase enzyme could read a base and find its complement which is much like the finite control that reads a character and replaces it with some other character. With this exciting thought, he tried solving an instance of the HPP, and succeeded in it. This opened the path for DNA computing. Adleman felt that the Watson–Crick pairing [2], polymerase, ligases, nucleases, gel-electrophoresis, and

V. Sudha (✉)
Kumaraguru College of Technology, Coimbatore, India
e-mail: sudha.v.cse@kct.ac.in

K. S. Easwarakumar
Anna University, Chennai, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_16

DNA synthesis [3] are the essential tools for computing a problem in DNA. Adleman used linear self-assembly for solving the problem.

In mathematics, all the advanced concepts are built upon the basic arithmetic operations. The arithmetic operations deal with the numbers of all types. It includes the basic operations such as addition, subtraction, multiplication, and division. The computer has become a part of human life, and fast computation is vital. Addition, subtraction, multiplication, and division are the fundamental mathematical operations. Most of the arithmetic operations performed on a computer are realized using these basic operations. Among the four, addition and subtraction operations form the basis, as the other two operations can be realized by using these procedures.

The success of a computing model depends on the implementation of the basic operations. In any model of computing, arithmetic operations are the basic operations that need to be implemented efficiently. When arithmetic operations need to be performed on given operands, these operands must be represented in a form acceptable by the respective systems. Hence, the following parameters must be taken into consideration for proposing an algorithm for performing arithmetic operations. They are representation of the operands, logic used for performing the operation, name of the biological operations used for realizing the above logic, etc. This paper discusses various existing algorithms proposed for performing basic arithmetic operations.

2 Existing System

DNA computing is an emerging type of molecular computing. Hence, it is important to find an efficient algorithm to address this problem. In DNA computing, a problem is solved by developing an algorithm and proved either experimentally or theoretically. There are many algorithms proposed in the literature for realizing these operations. Also, there are few insertion–deletion systems [4] for the same. In this section, we discuss the methodology, pros, and cons of these algorithms.

2.1 Number Representation

In DNA computing, the pieces of information must be represented in terms of four base values: Adenine (A), Cytosine (C), Guanine (G), and Thymine (T). Hence, to implement any operation in DNA computing, the inputs must be encoded using DNA bases. The different representations used in the literature are discussed in this subsection.

Guarnieri et al. [5] proposed a procedure for the addition of two-bit numbers. A bit can take two possible values: 0 and 1. Procedures are proposed to represent these two possible values. Following this, the encoding of a binary number is the encoding of each bit value in the number.

Gupta et al. [6] proposed the logic for performing arithmetic and logic operations. In the proposed method, the DNA strands are designed in such a way that the output of one operation can be the input of another operation. As operations such as NAND, NOR, and OR are basic for all arithmetic operations, it is implemented first. These gates take two inputs and produce one output. The first number is known as input, while the second is known as the operand. To perform the above-stated operation in DNA computing, the bit values 0s and 1s are represented as nucleotides. An instance of the encoding is given below for reference.

input strand: 1-AU; 0-UA

operand strand: 1: AT or TP; 0: TA or PT

i.e. 0 and 1 in the operand strand are given two possible representations.

Frank Qiu and Mi [7] used insertion, deletion, and substitution operations for performing Boolean operations. To apply these operations on the input, it must first be encoded using DNA bases. It is known that, for any Boolean operation, two operands are passed as an input. Among the two operands, at least one of the operands must be encoded using DNA bases. Let $X = X_n X_{n-1}, \ldots, X1$ and $Y = Y_n Y_{n-1}, \ldots, Y1$ be the two operands. The operand X is represented in DNA form using the following notation.

$$L_n - X_n - R_n - L_{n-1} - X_{n-1} - R_{n-1} - \ldots - L_1 - X_1 - R_1$$

where L_i and R_i represent the left and right locators for the bit X_i.

Barua and Misra [8] proposed the following methodology for representing a binary number in DNA computing. Given a binary number, all positions with bit value 1 are initially collected. This results in a collection of some decimal numbers. Next, each of the decimal numbers in the collection is given a unique DNA representation. Thus, a binary number is represented by a set of DNA strands in DNA computing.

Fujiwara et al. [9] used DNA operations such as merge, copy, detect, separator, cleavage, annealing, and denaturation for performing logic and arithmetic operations with DNA strands. They encoded each binary number using the following representation:

$$\begin{aligned} X = {} & A_0, A_1, \ldots, A_{n-1}, B_0, B_1, \ldots, B_{m-1}, C_0, C_1, D_0, D_1, 1, 0,], \\ & A_0, A_1, \ldots, A_{n-1}, B_0, B_1, \ldots, B_{m-1}, C_0, C_1, D_0, D_1, 1, 0,] \end{aligned}$$

In the above representation, $A_0, A_1, \ldots, A_{n-1}$ represents addresses of binary numbers, $B_0, B_1, \ldots, B_{m-1}$ represents the bit value, and C_0, C_1 and D_0, D_1 are the specified symbols cut by a cleavage operation.

For instance, if the address and bit position of a bit are i and j, then the corresponding single-strand representation for the above is

$$S_{i,j} = D_1 A_i B_j C_0 C_1 V \; D_0$$

where the value of $V \in \{0, 1\}$.

Brun [10] represented numbers as a set of tiles. To add two binary numbers, tiles are designed for all possible combinations. For instance, when a bit of a binary number needs to be added, then the tile can be designed for the combinations (0,0), (0,1), (1,0) and (1,1). The tiles are designed in such a way that there exists a unique matching for each tile, and also, the number of the matching side should be equal to or greater than τ, where τ represents the temperature in the tiling system. A similar procedure is used for implementing the multiplication operation.

Wang and Huang [11] used two types of DNA strands for number representation—one for representing the given input decimal number, and the other for the output strand. The decimal number, say *n*, is represented by a single DNA strand containing n segments. For instance, the decimal number 5 can be represented by the DNA strand containing five fragments namely 1, 2, 3, 4, and 5; however, the output strand is designed with log2n fragments. A molecular beacon is used for implementing the multiplication operation.

Both the face and place value of the numbers are important for performing a valid arithmetic operation. Sudha and Easwarakumar [12] used both values in their representation. In this representation, given number of any base value, it is directly converted into a string made up of DNA bases. This representation enables us to perform arithmetic operations easily.

2.2 Operation Implementation

Frank Qiu and Mi [7] implemented Boolean operations for binary numbers by designing a DNA strand for each bit present in the input. The designed DNA strand includes a complementary strand that contains the result. For instance, for performing the AND operation, the strand for the bit Xi is designed as shown below,

$$\overline{L_j} - \overline{X_i} - \overline{X_i \wedge Y_i} - \overline{R_i}$$

In the above representation, the bit *Xi* can take the value either 0 or 1. Depending upon the value it takes, the DNA strand looks as shown below,

1. $X_i = 0 : \underline{\overline{L_i}} - \underline{0} - \overline{0} \wedge Y_i - \underline{\overline{R}}_i$
2. $X_j = 0 : L_j - 1 - 1 \wedge Y_j - R_j$

When the strands in the above manner are placed in a test tube, annealing takes place, and which results in a strand with a looping structure. Since a strand with a looping structure is the required one, the remaining strands are removed from the test tube, which gives the necessary result. Like the above AND operation, other Boolean and binary operations can be implemented.

To add two numbers, say *a* and *b*, Fujiwara et al. [9] applied the following operations over the strands designed using the above representation.

1. For each $j(0 \leq j \leq m - 1)$, compute $x_j = aj \oplus bj$, and $y_j = a_j \wedge b_j$.
2. For each $j(0 \leq j \leq m - 1)$, compute $p_j = x_j \wedge y_j$.
3. For each $j(1 \leq j \leq m - 1)$, set $c_j = 1$ if $y_{i-1} = 1$ or there exists $k(<j)$ such that $p_{j-1} = p_{j-2} = ... = p_{k+1} = 1$ and $y_k = 1$, otherwise set $c_j = 0$.
4. For each $j(1 \leq j \leq m - 1)$, set $s_j = x_j \oplus cj$.

Barua and Misra [8] implemented the arithmetic operations using the following logic in a recursive manner. Let X and Y be the two operands to be added. Then, if any one of the operands X or Y is null, the result is the other operand. Else, if both the operands are present, the sum is obtained by performing an exclusive-or operation on each bit of the input.

Once the tiles are designed as per the above requirement, the desired result is obtained by the process called self-assembly. In self-assembly, a tile attaches with another tile only when the input side of a tile matches with the output side of another tile. Brun [10] applied self-assembly to perform the necessary operation.

It is known that, in single bit multiplication, the output is 1 only if both the input bits are 1. Wang and Huang [11] used the molecular beacon for realizing the multiplication operation. In this method, the probe sequence is designed in such a way that the first half and second half contain the encoding of the multiplier and the multiplicand bits, respectively. Now, the complement of the DNA sequence coded for bit value 1 is floated in the test tube. This complemented sequence attaches to the beacon with a bit value 1. Thus, the beacon containing both the multiplicand and multiplier as 1 becomes a linear double-chain structure.

Guarnieri et al. [5] proposed a procedure for the addition of two-bit numbers. After designing the DNA strands, they can bond with the respective strands. To the obtained result, bit values are added in the next position until all the four bits are added. The result from the final double-stranded DNA is encoded.

Nobuyuki et al. [13] proposed a restriction enzyme-free method for the Boolean matrix multiplication operation. This method represents Boolean matrices using a directed graph. The row and column names of the Boolean matrix form the vertex. An edge is marked between two vertices in the graph if the corresponding Boolean matrix entry is 1. To implement the above process in DNA computing, a unique DNA strand is designed for each vertex present in the graph. The edges of the graph are designed by taking the complement of the DNA sequences obtained by concatenating one-half of the DNA sequences of first vertex, and the other half from the second vertex. The hybridization of the above-constructed structure leads to the creation of various paths, from which the valid paths are filtered.

Oliver [14] also used the procedure described in Nobuyuki et al. [13], but he used nearly 100 restriction enzymes to filter the result. The arithmetic operations in Sudha and Easwarakumar [12] are implemented using the insertion and deletion operation defined in SInsDelP system.

2.3 Usage of the Result for the Next Iteration

Using the method proposed by Frank Qiu and Mi [7], only two numbers can be processed at a time. When the process needs to be expanded for more than two numbers, the process is repeated sequentially until all the inputs are processed. The output generated from each iteration can be used as such for the next iteration. The number of biological manipulations needed is $O(N)$.

Fujiwara et al. [9] used a different format for representing the result. Hence, the procedure used for number representation cannot be passed directly as input for the next iteration. Also, the operation used for transforming the output into the reusable format is not error resistant. This drawback is overcome by the representation proposed in Xiao et al. [15]. It requires $O(1)$ step to perform the operation.

The output DNA strands in the procedure proposed by Barua and Misra [8] is like the input strand. Hence, without any modification, the output strand can be used for further operations or iterations. This procedure requires the average of $O(\log n)$ and $O((\log n)^2)$ steps for addition and multiplication operations.

Though the procedure in Guarnieri et al. [5] can be extended to any number of bits, finding a unique encoding for each bit value in a bit position is difficult. Also, the operations are applied sequentially over the bits. The encoding used for representing the resulting strand is another disadvantage of this method. As the encoding of output strand is different from the input, the recursive application of the proposed procedure is restricted.

Since the strands in Sudha and Easwarakumar [12] uses both face and place value for designing the DNA strand, the output strand can be directly used for further iterations without any modifications.

2.4 Bio-operations

Fujiwara et al. [9] make use of the operations: merge, copy, detect, separate, cleavage, anneal, and denaturation for performing arithmetic operations.

Using the tile assembly system, Yuriy Brun proposed a procedure for performing addition and multiplication. He has illustrated the concept with the help of an example starting from $S8$. Here, the value 8 represents the maximum possible values for three inputs. He designed tiles for all possible outputs.

Jing Wang and Yourui Huang used a molecular beacon for implementing multiplication operations. To enable the binary multiplication operation, the first decimal number is converted to the binary form, and any of the available algorithms are applied. The given binary numbers on which we need to apply the operation are encoded using the above encoding schemes. In the following example, the procedure for the NAND operation is discussed in detail.

Example 1 Let 1001 and 0101 be the two operands of the NAND operation. Between the two, we take 1001 as the input and 0101 as the operand. Then, the DNA encoding for these two binary numbers using the above encoding scheme is 1001 = AUUAUAAU. As there are two different encodings for 0s and 1s, we get 16 possibilities. Among the 16 possibilities, only one is the complement of the input strand and, that is the answer to the operation. Among the 16 possibilities, only TAATPTTP is the exact complement of the input strand.

PCR site-specific mutagenesis and replicative transposition are the biological operations used in [12] for realizing the arithmetic operations.

3 Findings from the Above Study

In this paper, a study is done on the algorithms proposed for arithmetic operations. From the study, the following points are inferred: a system that directly converts the given input number into an equivalent value in terms of DNA bases and a reusable representation. Though there are few algorithms for addressing the above issues, still the length of the string remains an issue.

Thus, we are indeed of a system that directly converts the given number to a string made up of DNA bases with easily handleable length and eases the application of arithmetic operation.

References

1. L.M. Adleman, Molecular computation of solutions to combinatorial problems. Science **266**(5187), 1021–1024 (1996)
2. H.J. Cleaves, WatsonCrick pairing, in *Encyclopedia of Astrobiology*, ed. by M. Gargaud et al. (Springer, Berlin, Heidelberg, 2011)
3. J. Read, S. Brenner, *Encyclopedia of Genetics* (Academic Press, Cambridge, 2001), pp. 577–577
4. V. Sudha, K.S. Easwarakumar, A comprehensive study of insertion-deletion system in DNA computing. Int. J. Recent Technol. Eng. **1**, 1211–1213 (2019)
5. F. Guarnieri, M. Fliss, C. Bancroft, Making DNA add. Science (New York, N.Y.) **273**(5272), 220–223 (1996)
6. V. Gupta, S. Parthasarathy, M.J. Zaki, Arithmetic and logic operations with DNA, in *Proceedings of the 3rd DIMACS Workshop on DNA-Based Computers* (Philadelphia, USA), pp. 212–220 (1997)
7. Z. Frank Qiu, L. Mi, Arithmetic and logic operations for DNA computers, in *Parallel and Distributed Computing and Networks (PDCN'98)*, pp. 481–486 (1998)
8. R. Barua, J. Misra, Binary arithmetic for DNA computers, in *DNA Computing. DNA 2002. Lecture Notes in Computer Science*, vol 2568, ed. by M. Hagiya, A. Ohuchi (Springer, Berlin, Heidelberg, 2002), pp. 124–132
9. A. Fujiwara, K. Matsumoto, W. Chen, Addressable procedures for logic and arithmetic operations with DNA strands, in *Proceedings of Parallel and Distributing Processing Symposium*, vol 2568 (Springer, Berlin, Heidelberg, 2003), pp. 162–172
10. Y. Brun, Arithmetic computation in the title assembly model: addition and multiplication. Theor. Comput. Sci. **378**(1), 17–31 (2007)

11. J. Wang, Y. Huang, Decimal integer multiplication based on molecular beacons. J. Networks **8**(12), 2762–2768 (2007)
12. V. Sudha, K.S. Easwarakumar, An algorithm for arithmetic operations: an application in SInsDelP system. Appl. Math. Inform. Sci. **13**(s1), 173–182 (2019)
13. K. Nobuyuki, Z. Ibrahim, Y. Tsuboi, O. Ono, Matrix multiplication by DNA computing, in *The Eleventh International Symposium on Artificial Life and Robotics(AROB 2006)*, pp. 476–479 (2006)
14. J. Oliver, *Computation with DNA: Matrix multiplication* (Academic Press, Cambridge, 1998), pp. 113–122
15. D. Xiao, W. Li, J. Yu, X. Zhang, Z. Zhang, L. He, Procedures for a dynamical system on {0,1}n with DNA molecules. Bio Syst. **84**(3), 207–216 (2006)

Fuzzy C-means for Diabetic Retinopathy Lesion Segmentation

Shalini and Sasikala

Abstract The paper aims to segment the lesion of non-proliferative diabetic retinopathy (NPDR) which occurred due to diabetes. The risk of loss of sight can be decreased by 95% with the timely diagnosis of the NPDR disease. The proposed work aimed to segment the NPDR lesion called 'hard exudate'. Firstly, the fundus input image undergoes resizing by applying bi-cubic interpolation area method, and then the resized image is preprocessed with single channel extraction and median filter. Further, the NPDR lesion is segmented using k-means and fuzzy C-means (FCM) algorithms. By comparing the results of both segmentation algorithms, FCM shows the better result. The executions of the methods are evaluated using mean-squared error, structural similarity index measure, sensitivity, specificity and accuracy. The proposed FCM method for segmenting the 'hard exudate' lesion has achieved a better result of 95.05% accuracy.

Keywords Diabetic retinopathy · Retinal · Hard exudate · Binary thresholding · K-means · Fuzzy C-means · Lesion of interest

1 Introduction

The disease which damages the **retina** of human eyes owing to **diabetes** is called as '**diabetic retinopathy**'. **The prolonged glucose** obstructs supply of oxygen which starves the **retinal nerves** leading to retinal diseases. Thus, a retinal complication occurred due to diabetes is termed as diabetic retinopathy. Such a damage in retinal tissue results in cloudy or blurred vision and even leads to blindness.

This blindness due to **DR** is caused among working adults specifically. Almost people with diabetes possibly will get this retinal disease. The protracted diabetes is more likely to develop the DR disease. If the disease is left untreated earlier, the condition becomes worse causing vision loss that is irreversible. Even though numerous medical disorders can cause retinopathy (retinal disease), the most serious

Shalini (✉) · Sasikala
Department of Computer Science, IDE, University of Madras, Chennai, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*, Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_17

causes which doctors see in retina are hypertension and diabetes. Among these two cases, complication of diabetes is more critical since they are symptomless, and it should be identified as earlier as possible with a complete dilated eye exam twice in a year.

And this conventional eye exam is done manually which becomes a huge and complicated task as the number of patients suffering with the disease is increasing rapidly. The global statistics countersigned by World Health Organization (WHO), among 7.9 billion of current population 39.8 million were blind in which 4.8% were affected by DR. So an automated system needs to be emphasized to manage this scenario to avoid further complications [1].

The DR disease is categorized into two classes, namely **non-proliferative diabetic retinopathy** (NPDR) and **proliferative diabetic retinopathy** (PDR). The 'NPDR' is the first kind of the disease which deteriorates the walls of the retinal blood vessels in producing small swellings called micro-aneurysms (MA) stick out from the blood vessel also drips the fluid called hard exudates (HEXU) and leaks blood called as hemorrhages (HEMO) into the retina, consequently the bigger retinal blood vessels begin to widen and become irregular in diameter that leads to partial retinal mutilation. This type can proliferate from mild to severe level, also blockage of further blood vessels results in bulge of nerve fibers in the retina and leads to brain-related complications [1].

The second type is 'PDR' which is the advanced type of the DR disease which can progress from severe to more severe stage. In this type, abnormal blood vessels grow in the retina also the spoiled blood vessel blocks off and discharges the gelatin substance that occupies the focal point of the eye.

In the long run, retina occurs to get isolated from back of the eye. Additionally, the pressure produced in the eyeball obstructs the fluid flow in the eye that results in glaucoma by damaging optic nerve that carries vision from the eye to brain [1].

2 Literature Review

Shalini et al. [2] have presented a survey study on recognition of DR disease which gives a review on various algorithms and techniques that are instigated for detecting the lesions caused by diabetic retinopathy disease and also categorizing its severity stages while attaining improved accurateness in the process.

Ranamuka et al. [3] have proposed the algorithm that identifies hard exudate feature in DR with the morphological and fuzzy logic operations. Initially, the retinal fundus image undergoes HSI space conversion, and then median filter removes noise, contrast limited adaptive histogram equalization enhances the contrast followed by, and the biggest spherical region called optic disk is eradicated. Further, hard exudate lesion is detected and classified using fuzzy operation.

Franklin et al. [4] have proposed an effective system for detecting the hard exudate by utilizing the artificial neural networks. Initially, RGB fundus image undergoes into laboratory color space, and CLAHE operates on enhancement to handle non-uniform illumination. Attributes like color, size, shape, edge strength and texture are extracted. At last, the candidate region is categorized as hard exudate based on the extracted features using a multi-layer perceptron neural network.

Lachure et al. [5] have implemented a system to categorize the input image into abnormal and normal by identifying the exudate feature. At first, the input image undergoes color translation and filtering. The optic disk is removed using channel separation and histogram equalization. Finally, the statistical features like area, entropy, correlation, energy, contrast homogeneity, standard deviation, mean, skewness, kurtosis are extracted and based on these features, the classification of exudates are done.

Elbalaoui and Fakir [6] have proposed a paper on exudate detection. In the first stage, a feature called optic disk is removed using curvature scale space, and then the image undergoes grayscale conversion and for segmenting the exudates using mean-shift algorithm and maximum entropy thresholding algorithm with 92.24% accuracy.

Pal et al. [7] have proposed a procedure on detecting abnormal features in fundus images that are affected by DR. The algorithm carries preprocessing of image using green histogram equalization, and then region of interests is spotted using sobel and prewitt filters, channel extraction and CLA equalization along with morphological operations.

Hosanna Princye et al. [8] have presented a paper on hard exudate identification with methods like channel extraction, histogram equalization, circular Hough transform and FCM algorithms. The presented work has acquired 92.08% accuracy.

Dileep et al. [9] have presented a system for distinguishing the hard exudates occurred regions using preprocessing algorithms like HSI conversion, median filter, CLA histogram equalization then K-means and FCM algorithm for segmentation with 90% exactness of detection.

Sopharak et al. [10] have proposed the hard exudates identification in non-dilated diabetic retinopathy input images using FCM algorithm, and the work gives about 85.6% accuracy.

3 Methodology

The methodology depicted in Fig. 1 carries three phases like fundus image resizing, preprocessing the resized image and segmenting the lesion of interest (LOI) called 'hard exudates'. K-means and FCM algorithms are applied for segmentation. As a final point, the outcomes of the algorithms are compared using validation measures.

The retinal fundus image is standardized with bi-cubic interpolation area method [11]. This method resizes the image with zero padding and translated by 0.5 pixels in positive horizontal direction, by using the weighted average of four translated pixel values for each output pixel value [11].

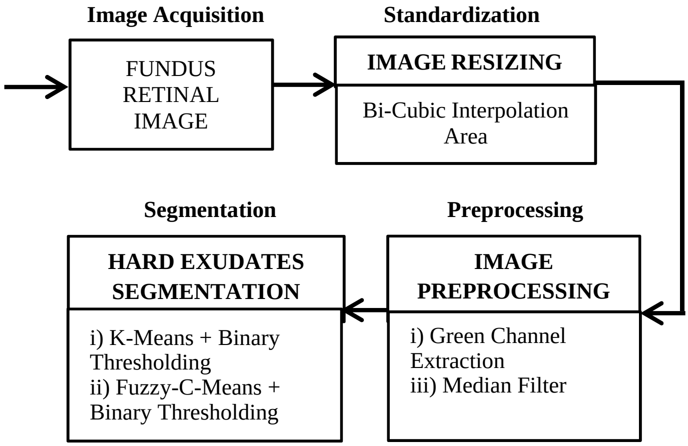


Fig. 1 Segmentation of hard exudates

$$R(a, b) = \sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij} x^i y^j \tag{1}$$

Now the standardized image undergoes preprocessing for channel extraction to isolate the green (G) band of the RGB fundus image, which projects the DR features more prominent than the blue (B) and red (R) bands.

$$\begin{aligned} & c^2 = c = c^T \\ & \text{i.e. } c^2(i \;\; j \;\; k) = p(i \;\; 0 \;\; k) = p(i \;\; 0 \;\; k) \end{aligned} \tag{2}$$

where c represents image, and i, j, k are R, G and B channels.

The median filtering is a technique [12] which enhances the DR lesion region by removing the noise while preserving the boundary.

$$M[x, y] = \text{median}\{x[i, j], (i, j) \in \omega \tag{3}$$

where ω represents the neighbor pixel.

To get the lesion of interest from preprocessed image, the segmentation process is done using three different algorithms as follows:

(a) Binary Thresholding

Binary thresholding is image segmentation method which creates binary images and has resultant only two values (0, 1), where 'zero' indicates dark pixel and 'one' denotes bright pixel [13]. This method fixes a threshold and compares pixel with the fixed threshold; where the lesser values are swapped with black pixel, and the values greater values are swapped with white pixel. Thus, the method classifies the pixels into groups, by setting an upper and lower bound to each group.

Algorithm:
Step i: Choose a random pixel value in an image $A_{(i,j)}$.
Step ii: Fix a constant threshold value 'TH'.
Step iii: If the value of chosen pixel is lesser, the pixel is set to 'zero'.

$$A_{(i,j)} < TH = 0 \tag{4}$$

Step iv: If the value of chosen pixel is greater, the pixel is set to 'one'.

$$A_{(i,j)} > TH = 1 \tag{5}$$

Step v: Repeat Step iii and Step iv, until the pixels of the whole image is checked with threshold (TH) value.

(a) *K-means*

K-means algorithm [14] is an iterative method that is used to panel an image into clusters and assigns data point to cluster whose center is adjacent. The center refers the average data points in the group. K-means algorithm performs faster and runs on large datasets.

Algorithm:
Let $D = \{d_1, d_2, d_3, \ldots, d_n\}$ be the set of data points and $C = \{c_1, c_2, \ldots, c_c\}$ be the set of centers.
Step i: Select one cluster center randomly.
Step ii: Compute the distance between the data point and cluster center.
Step iii: Assign the value to the cluster center which has a minimum distance from the data point.
Step iii: Select another cluster center.
where, 'c_i' represents the data points in each cluster.
Step v: Compute the distance between the data point and cluster center.
Step vi: Do the step iv until all the data points are assigned with cluster.

(c) *Fuzzy-C-Means*

The FCM algorithm assigns membership to data point on the basis of distance between the cluster center and the data point. More the data is neighboring, more is its membership directed for the particular cluster center. Evidently, summation of membership of data point equals '1'. The membership and cluster centers are updated with the formula [15]. It gives better results than K-means algorithm.

Algorithm:
Let $F = \{f_1, f_2, f_3, \ldots, f_n\}$ which refers the set of data points and $C = \{c_1, c_2, c_3, \ldots, c_c\}$ refers the set of centers.
Step i: Choose a random 'c' cluster center.
Step ii: Calculate the fuzzy membership μ_{ab} using:

$$\mu_{ab} = 1 \Big/ \textstyle\sum_{k=1}^{c} (d_{ab}/d_{ac})^{(2lm-1)} \tag{7}$$

Step iii: Calculate the fuzzy centers '*cb*' using:

$$C_b = \left(\sum_{a=1}^{n} (\mu_{ab})^m x_a\right) / \left(\sum_{a=1}^{n} (\mu_{ab})^m\right), \forall_{b=1,2,\ldots,c} \tag{8}$$

Step iv: Repeat step ii and iii until the minimum 'R' value is attained.
(or)

$$\left\| U^{(t+)} - U^{(t)} \right\| < \beta \tag{9}$$

where 'k' denotes the iteration step.
'β' denotes the termination condition between [0, 1].
'U' = $(\mu_{ab})_{n*c}$' denotes the fuzzy membership matrix.
'R' denotes the objective function.

4 Results and Discussion

The dataset for this segmentation method was obtained from online repository called DIARETDB1. The database contains 95 input images for the experiment with the resolution of 93 × 71 in 24 bit depth in which, 5 are unaffected by NPDR and 90 are NPDR affected. And these retinal input images are rescaled with 512 × 512 dimensions as depicted in Fig. 2 to maintain uniformity. This resizing helps in proper visibility of images on screens of different devices.

The hard exudate is the bright lesion formed by the eye disease called DR. This type of lesion exists in the first type of DR disease called NPDR. It is the sign of leaking blood vessels which drop out pale, fatty deposits on the retina which is termed as protein lipids. Segmentation of exudates is an essential factor in DR diagnosis to obstruct the disease severity. This segmentation process undergoes binarization which needs green channel extraction that gives good results for bright pixels so bright lesion detection becomes easier followed by, the unwanted noise is removed by applying the median filter. Then CLA histogram equalization method makes the regions of hard exudate more projective.

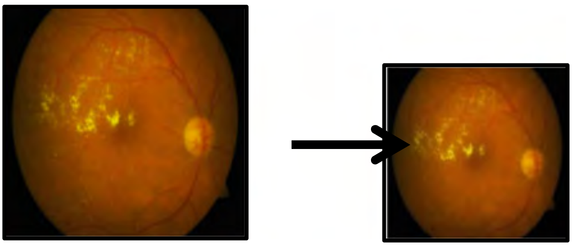


Fig. 2 Fundus input image, Resized image

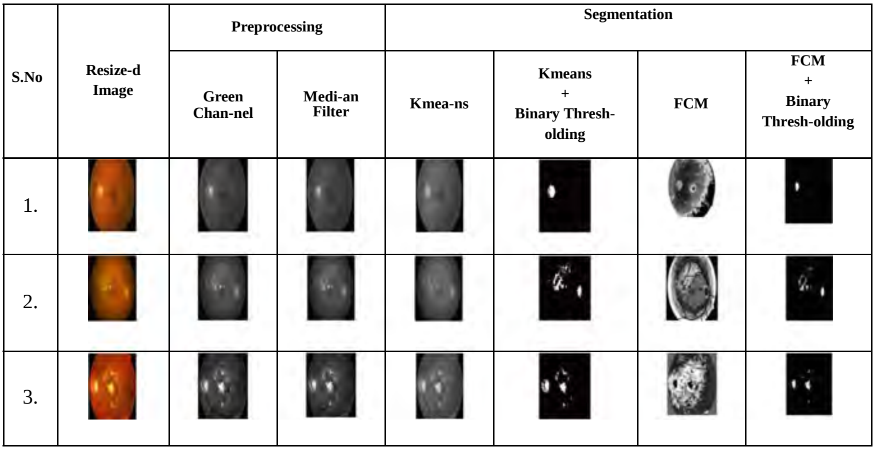


Fig. 3 Exudates segmentation

Further, the hard exudate is segmented using two different segmentation algorithms as shown in Fig. 3. First algorithm is called K-means where the set of data points that are similar in an image are clustered. The grouping is done by assigning each data point whose center is the nearest; the result is purely determined by the initial random assignments of data points and then binary thresholding is done, which segments the hard exudate lesion from the grouped image depending on the actual value. Unlike conventional thresholding algorithms, K-means decreases intra-cluster variance but the algorithm has local minimum and yields dissimilar results in altered executions. The second algorithm is called fuzzy C-means which also decreases intra-cluster variance and has local minimum. Unlike K-means and other segmentation algorithms, in FCM, the results depend on the initial choice of weights so the algorithm yields same results in altered executions. So among these two methods, the FCM method has given better results than the K-means method and achieved accuracy of 95.05%.

5 Performance Analysis

The evaluation of the proposed segmentation techniques is done using certain validation measures like mean-squared error (mse), structural similarity index (ssi), sensitivity (Sn) and specificity (Sp) and accuracy (Acc) are calculated.

Mean-squared error value validates the prominence of the segmented image, where the values must non-negative and closer to zero.

$$\frac{1}{n}\sum_{i=1}^{n}\left(M_i - M_i^{\wedge}\right)^2 \tag{10}$$

Here, n is data points, M_i is observed value, and $M_i^{\wedge}$ is predicted values.

Structural similarity index measure is a cluster validation method for envisaging the superficial quality of segmented image. It is used for measuring the similarity between two images based on luminance, contrast and structure. And this measure improves on traditional measure 'mean-squared error'.

$$S(A, B) = [l(A, B)^{\alpha}.c(A, B)^{\beta}.s(A, B)^{\gamma} \tag{11}$$

Here, setting the weights of α, β, γ to 1.

Sensitivity evaluates the percent of positive rates. (e.g., the count of retinal images which are properly recognized as NPDR affected).

$$\text{Sensitivity} = \frac{\text{tp}}{\text{tp} + \text{tn}} \times 100 \tag{12}$$

Specificity evaluates the percent of negative rates (e.g., the count of retinal images which are properly recognized as NPDR unaffected).

$$\text{Specificity} = \frac{\text{tp}}{\text{fp} + \text{tn}} \times 100 \tag{13}$$

Accuracy is evaluated as the sum of positive and negative rates (the count of NPDR affected and unaffected retinal images) dividable by the count of images in the dataset.

$$\text{Accuracy} = \frac{\text{tp + tn}}{\text{tp + fp + tn + fn}} \times 100 \tag{14}$$

where tp, fp, tn and fn represent true positive, false positive, true negative and false negative.

The validation of the proposed segmentation algorithms is evaluated with mse and ssim which are formulated in Eqs. (10) and (11). The values obtained are 0.52 and 0.15 for K-means, 0.52 and 0.15 for FCM.

The performance analysis of hard exudates segmentation using K-means with binary thresholding algorithm gives 78.07% sensitivity, 67.57% specificity, 83.21% accuracy and FCM with binary thresholding gives 97.92% sensitivity, 28.57% specificity, 95.05% accuracy which is formulated in Eqs. (12), (13) and (14). And the evaluated results are presented in Table 1 and Fig. 4, which shows that proposed FCM method gives the better result.

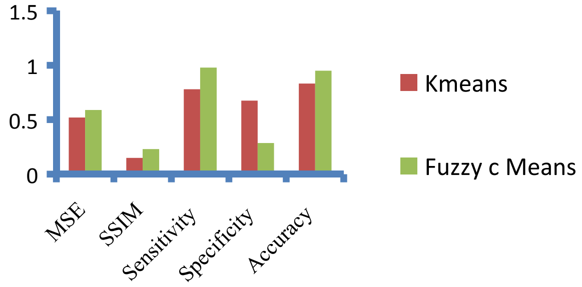


Fig. 4 Comparison chart of exudates segmentation

Table 1 Performance measures of exudates segmentation

S. No	Method	mse	ssim	Sn (%)	Sp (%)	Acc (%)
1	K-means + Binary thresholding	0.52	0.15	78.07	67.57	83.21
2	Fuzzy C-means + Binary thresholding	0.59	0.23	97.92	28.57	95.05

6 Conclusion

The segmentation of hard exudates has been executed with the FCM method. The implementation is done by resizing the image for standardization. Then the resized images are preprocessed using channel extraction and median filter which enhances the foreground from the background. For segmenting the hard exudates, two algorithms, namely K-means and FCM, have been applied. Among these two algorithms, FCM has performed better with an accuracy of 95.05%. In upcoming prospect, segmentation of DR can be done by reducing false positives to achieve higher precision. Also, the proposed FCM method depends on the initial choice of weights. Hence, different initializations lead to different results. So, there is necessity of intending, a statistically formalized method with maximization.

References

1. A. Pattanashetty, S. Nandyal, Diabetic retinopathy detection using image processing: a survey. Int. J. Comput. Sci. Network, pp. 661–666 (2016)
2. R. Shalini, S. Sasikala, A survey on detection of diabetic retinopathy, pp. 626–630 (2018). https://doi.org/10.1109/I-SMAC.2018.8653694
3. N.G. Ranamuka, R. Gayan, N. Meegama, Detection of hard exudates from diabetic retinopathy images using fuzzy logic. IET Image Process, pp. 121–130 (2012)
4. S.W. Franklin, S.E. Rajan, Diagnosis of diabetic retinopathy by employing image processing technique to detect exudates in retinal images. IET Image Process, pp. 601–609 (2013)
5. J.S. Lachure, A.V. Deorankar, S. Lachure, Automatic diabetic retinopathy using morphological operations. Int. J. Comput. Appl., pp. 22–24 (2015)
6. A. Elbalaoui, M. Fakir, Exudates detection in fundus images using meanshift segmentation and adaptive thresholding, in *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* (2018)

7. A.L. Pal, S. Prabhu, N. Sampathila, Detection of abnormal features in digital fundus image using morphological approach for classification of diabetic retinopathy. Int. J. Innov. Res. Comput. Commun. Eng. pp. 901–909 (2015)
8. P. Hosanna Princye, V. Vijayakumari, Detection of exudates and feature extraction of retinal images using fuzzy clustering method. IET publications, pp. 388–394
9. J. Dileep, P. Manohar, Automatic detection of exudate in diabetic retinopathy using K-clustering algorithm. Int. J. Recent Innov. Trends Comput. Commun., pp. 2878–2882 (2015)
10. A. Sopharak, B. Uyyanonvara, S. Barman, Automatic exudate detection from non-dilated diabetic retinopathy retinal images using fuzzy C-means clustering. Sensors open access publications pp. 2148–2161 (2009). www.mdpi.com/journal/sensors
11. https://in.mathworks.com/help/vision/ug/interpolation-methods.html
12. https://www.sciencedirect.com/topics/engineering/median-filtering
13. ttps://en.wikipedia.org/wiki/Thresholding_(image_processing)
14. https://sites.google.com/site/dataclusteringalgorithms/k-means-clustering-algorithm
15. https://sites.google.com/site/dataclusteringalgorithms/fuzzy-c-means-clustering-algorithm

A Secured System for Tele Cardiovascular Disease Monitoring

Azmi Shawkat Abdulbaqi, Saif Al-din M. Najim, Shokhan M. Al-barizinji, and Ismail Yusuf Panessai

Abstract Electrocardiogram (ECG) signals play an indispensable role in interpreting the heart's effectiveness in the form of electrosignals to diagnose different types of cardiac problems. These vital signals should also be transmitted safely to avoid any interruptions in the data loss or noise that may lead to illness detection. As ECG signals are observed with a higher-dimensional scale, this should be compressed for leveraging accurate control and transportation. A method of lossless compression called Huffman-based discrete cosine transform (DCT) is performed in this manuscript for achieving the efficient transmission of ECG data. DCT and inverse discrete cosine transform (IDCT) are suggested for improving data privacy and lowering the data complexity. This manuscript concentrates on achieving a high level of accuracy ratio in the reconstruction upon compression and transportation of the original data (OrDa) unaccompanied any failure in the lowest computational time. During the first stages, preprocessing and sampling are performed to eliminate the sounds and transmission of OrDa. The DCT-based Huffman quantization approach achieved great performance measures as "distortion percentage"(PRD), "signal-to-noise ration" (SNR), "quality score" (QS), and "compression ration" (CR) when comparing with current approaches in different data transformations.

Keywords DCT · Electrocardiogram (ECG) · Huffman coding (HufCod)

A. S. Abdulbaqi (✉) · S. A. M. Najim · S. M. Al-barizinji
College of Computer Science & Information Technology, University of Anbar, Ramadi, Iraq
e-mail: azmi_msc@uoanbar.edu.iq

S. A. M. Najim
e-mail: saifaddin.r@uoanbar.edu.iq

S. M. Al-barizinji
e-mail: shokhan.albarizinji@uoanbar.edu.iq

I. Y. Panessai
Faculty of Arts, Computing and Creative Industry, UPSI, Tanjong Malim, Malaysia
e-mail: ismailyusuf@fskik.upsi.edu.my

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_18

1 Introduction

All over the world, ECG signals are utilized to track the heart activity and analyze different heart diseases. It demonstrates the findings that change the heart activity. ECG signals can be utilized to diagnose heart problems, infarcts, and contagions. ECG measures the activity of the heart muscle. Such signals take the broad space required to store and transmit ECG signals. In order to rectify this issue, ECG data compression (DaCo) presents a significant role in efficiency and requires minimum amounts of storage space. Based on the recommended smart e-health system, IoT is utilized to provide this scheme [1]. IoT device collects and processes health data such as “Blood-Pressure”, “Oxygen” and levels of “Blood Sugar”, “Weight”, and “ECGs”. In such IoT apps, the management of huge-data sensors is an instance of the major difficulties. Information protection for IoT is now extremely complex and challenging, as menaces are increasingly harder to identify [2]. Different techniques are utilized to compress ECG signals like, wavelet compression, filters band compression, predictions of ANN, compression of matching pursuit, and compression based on clustering. The wavelet compression output signal removes lowest significant variables and keeps full space memory. The filters-band approach works best efficiently from the other transformation of the wavelet. However, it is not accurate, as there are opportunities to filter necessary values. Yet those methods of compression have augmented the time/costs of consumption [3]. The optimized lossless data decomposition approach to improve the reliability of transmission of ECG signals was utilized. The recommended algorithm works-effectively for signals compression unaccompanied information loss. The lossless algorithm is tested, and it was found to have achieved a high rate of CR if compared to other methods. This manuscript inspects both the DCT and the IDCT. Figure 1 displays the initial ECG signal and is a method of electrophysiological control for the heart’s electric activity [4, 5] (Fig. 2).

2 System Goals

The main goals of this manuscript are to improve the highest CR of the inputting ECG signals not accompanied by information loss, to determine the security and privacy of knowledge on the IoMT app, and to improve the decomposition representation of the lossless ECG signal.

3 Manuscript Organization

The first section presents the ECG signals compression apps. The second section explains the various lossless ECG signals literature. The third section includes the recommended methodology. The fourth section explained the measured performance

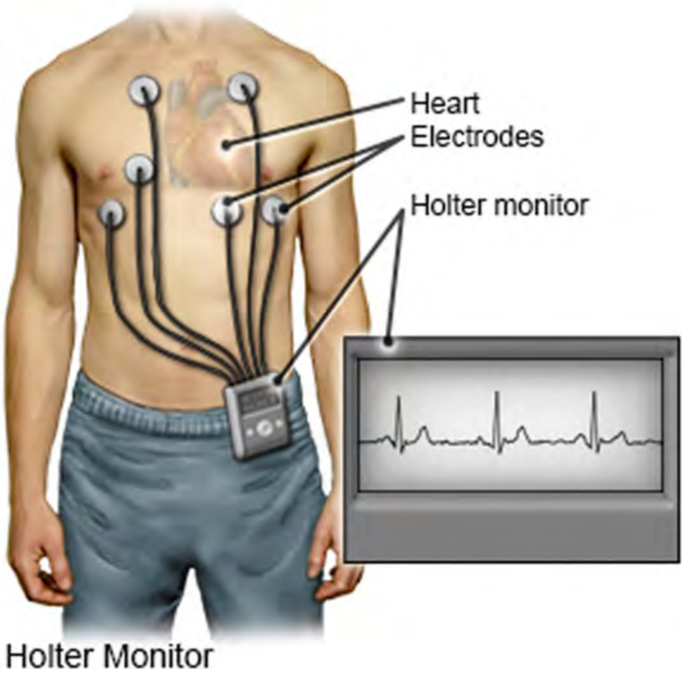


Fig. 1 Electrocardiogram system with Holter machine [6]

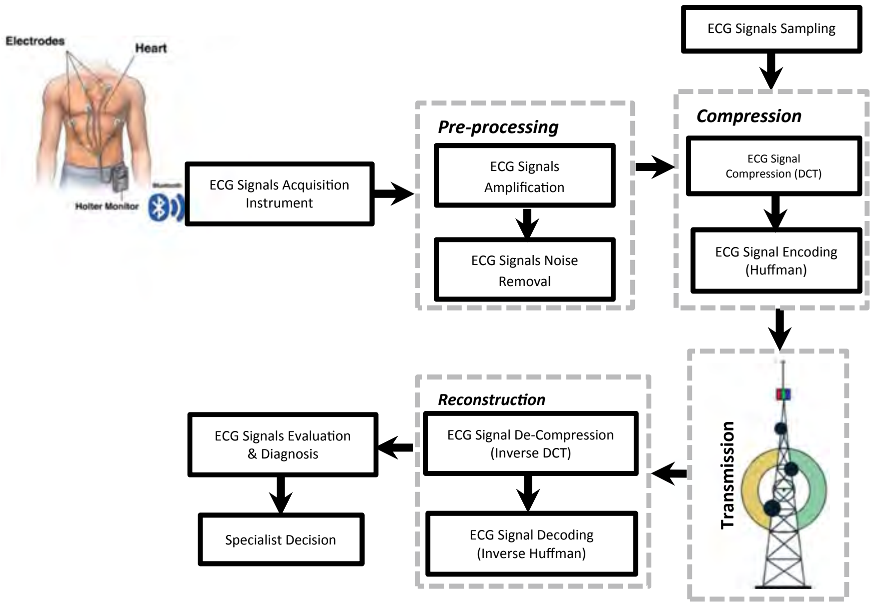


Fig. 2 Recommended system

analysis. The last section mentioned to the recommended strategy and covering of the recommended method.

4 Literature Review

The developments in the area of communication and electronics engineering lead to the utilization of big data (BiDa) with approximately all apps. The DaCo method acts a critical role in the development of information technologies that assist to manage the BiDa in a smart way. Manuscript [3] shows a new algorithm of ECG signals compression based on the adaptive Fourier decomposition (AFD). In any decomposition phase, it employs the Nevanlenna factorization and the highest choice of selection to obtain a rapid convergence ratio with more fidelity. That algorithm was practiced to the ECG signals and to gain the best findings. Manuscript [7] has developed an active procedure for QRS complexes detection based on the simplest analytical of the proposed temporary structure of ECG signals. In this procedure, the ECGs are prepared within some phases including detection the features, analysis the features, and remove the noises. The set of features gathered includes most of the ECGs information with needing to low data storage capacity, which develops a lossy ECG signals compressed form. Manuscript [8] concentrates on the prediction of adaptive linear associated with the adaptive coding scheme family called Golomb-Rice coding (GRC) to perform the compression of lossless data. The study does not utilize any RAM for data storing. VLSI has been implemented utilizing TSMC 0.18 μm technique. It indicates a general improvement in output parameters as compared to other applications. Manuscript [5] is an original method for ECG signals compression on little-power ECGs sensor nodes is improved to take advantage of insufficient ECG signals. The successful compression efficiency of the technique suggested offers a feasible solution for the design of wireless ECG sensor nodes to attain ultra-low power consumption. The recommended compression efficiency algorithm provides a solution for the architecture of a WSN ECG node to perform ultra-low power consumption. A DaCo novel and an IoT power reduction plan allowed by WSN are recommended.

Manuscript [9] introduced system that utilizes a lossless and lossy approaches with compressed data to permit a mixed transmission mode, to support adaptive data rate selection, and to save wireless broadcast power. The applied approach to an ECG, the data is compressed first utilizing a method of loss compression with a higher CR. Use entropy coding, the error rate between the OrDa and the DeDa is stored, which enables lossless recovery of the specific information when required. Manuscript [10] describes an effective algorithm to compress the data for tele-cardiac-patients monitoring inside the agricultural zone, based on DCT and two mixed encodings approaches. This method presents an attractive CR with a little PRD values.

5 The Proposed Work

5.1 *Pre-processing*

Preprocessing is critical in the ECG dataset; the signals are obtained without accurate representation from the heart and certain signals go missing (lost). The ECG data is extremely noisy and ECG signals are low. So we need to separate the desirable signals from the ECG data [9]. This method focuses on the elimination of the undesirable signal for further analysis. Here, the signals values are stowed in the numerical format [11].

5.2 *Sampling*

The sampling principal goal is to decrease the work time, additionally the cost. The data sampling is determined with different measurements to improve the recommended method implementation. A sampling of data is selected for data efficiently transmission unaccompanied by any data losing [12].

5.3 *DCT*

DCT is the transforming approach; it transforms a signal of TS into simple elements of the frequency. The DCT is utilized for the decrease of ECG dataset and the efficient extraction of functions [13]. The DCT exits with coupled signals providing a high energy compaction. The coefficients of transformation are zero or minimal and maximum coefficients are strong. The much more important information in ECG is compressed into coefficients first [14, 15]. To calculate DCT for the coefficients, asymmetric properties are utilized first, and then the asymmetric properties of those coefficients, and $Y(u)$ is utilized in Eq. (1). DCT coefficients can be represented by [16]:

$$Y(u) = \sqrt{\frac{2}{N}.\alpha(U).\sum_{x=0}^{N-1} f(x).\cos\left(\frac{\pi.(2x+1).u}{2N}\right)} \tag{1}$$

when $u = 0, 1, 2, 3, \ldots, N - 1$.

IDCT is determined utilizing the symmetric and asymmetric properties of the coefficients; $\alpha(U)$ is shown Eq. (2). Here, $\alpha(U)$ is given by [17]:

$$\alpha(U) = \begin{pmatrix} \frac{1}{2^{1/2}} & u = 0 \\ 1 & u > 0 \end{pmatrix} \tag{2}$$

when $u = 0, 1, 2, 3, \ldots, N - 1$.

5.3.1 DCT Coefficients

The standard ECG dataset is composed of ECG signals. DCT matrix to compress signals and then the coefficient of the DCT is given was utilized [18].

5.4 *Compression*

5.4.1 Lossless Compression

Utilization of the compression approach of lossless, the process of obtaining a high CR be available. Without this, the highest CR cannot be achieved [19]. The main benefits of lossless compression are not lost any information, and allow getting the main file when decompress is achieved [20].

5.5 *Huffman Coding (HufCod)*

HufCod is utilized to attain 20–90% compression effectiveness. The compression method involves encoding the message of the bit in binary bits format. After the encoding process, the decoding process will be done by tracking Huffman tree from the tree's root according to the specified sequences [21]. The key benefit is to operate at low complexity computational by replacing each character with relative frequency of the character-based a variable-length symbol [22].

5.5.1 ECG DaCo

In HufCod, the combination of data must quantize the overall data file in this process [23].

5.5.2 Huffman Encoding

The number of data values has been reduced through HuCo. The symbol sequence is the input to this algorithm and coding and it requires encoded. Tracing the symbols is an essential step and every symbol location includes number indexing [24]. Encoded relies on the frequencies of the letters and creates an illegible data format [25].

5.5.3 Huffman Decoding

The decoding is processed by a fixed sequence and the symbols are assigned. It substituted every character by that character's relative frequency in the text and minimizing the code's average length based variable-length code based on [26].

5.6 Applying IDCT

The IDCT receives the input as transformative coefficients $Y(u)$ and transforms it into TS $f(x)$. The IDCT applies to any set of coefficients. IDCT retransforms inside TS. DCT and IDCT provide very intense computational work [27, 28].

$$Y(u) = \sqrt{\frac{2}{N}.\alpha(U).\cos\left(\frac{\pi.(2x+1).u}{2N}\right)} \tag{2}$$

6 Analysis of Performance

This section discusses the detailed performance review of CR, PSNR, and MSE.

6.1 Comparison Enabling Dataset Description

The ECG data collection was taken from MIT-BIH Arrhythmia database, and the source ("https://archive.physionet.org/cgi-bin/atm/ATM") is utilized to implement our proposed compression algorithms [29]. The recording of the ECG was performed in different subjects. ECG dataset compression was utilized to solve restrictions related to security risk problems and limitations [30].

6.2 Compression Ration (CR)

The ratio of DaCo is known as compression strength. The measure is utilizing the DCT algorithm to reduce the data size. This also measures dataset complexity [31].

$$\text{Compression Ration (CR)} = \frac{\text{Uncompressed Signal Size}}{\text{Compressed Signal Size}}$$

Table 3 indicates the contrast of the compression degree between the current methods and the recommended methods to the respective CR. We can conclude from Table 3 that the recommended method resulted in the highest value of 25.74 than other current CR values of 11.6, 14.9, 14.3, 15.1, 5.65, 16, and 7.8.

6.3 Signal-To-Noise Ratio (SNR) and Distortion Percentage (PRD)

SNR calculates the ratio of overall positive signal power to the power of distorted noise [32].

$$\text{SNR} = 10xLog\left(\frac{\sum_0^{N-1}(X(n) - \text{mean}(X))}{\sum_0^{N-1}(X(n) - Y(n))^2}\right) \quad (4)$$

PRD is an intermediate sum of distortion between the source and distorted the waveforms of ECG. PRD can be represented as follows by the below Eq. [33]:

$$\text{PRD} = \sqrt{\frac{\sum_{n=1}^{N}(x(n) - x'(n))^2}{\sum_{n=1}^{N} x^2(n)}} \times 100 \quad (5)$$

x_i refers to the original ECG signals and x_i^Y refer to reconstructed signals.

6.4 The Score of the Quality (QS)

This is providing from before percentage of CR with PRD. QS defined as a critical functioning monitoring measured that helps with the consideration of compensating for errors which occur in reconstruction to select the correct compression operation. For instances, when a loss type of compression operation is performed, the increased QS values may imply the robustness of the compression method taken up [34].

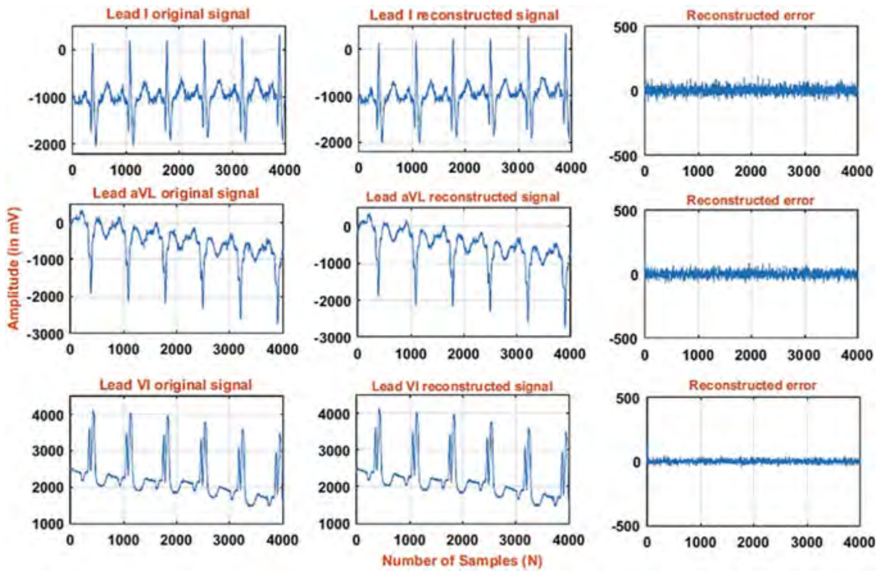


Fig. 3 Original and reconstructed ECG signals based on the recommended method with corresponding error signals

$$QS = \frac{CR}{PRD} \tag{6}$$

The findings are shown in Fig. 4 in graphical format to know the exact efficacy rate obtained by the proposed approach. This clearly shows that when compared with current methods, the recommended system gives a high CR and consistency resulting in a simple data transfer rate, and when comparing with another methods, PSNR values are equal to zero. It insists that the recommended approach has a highly optimized encryption and decryption framework with a powerful firewall and robust decryption system [35, 36] (Fig. 3).

From the tabulation, it can be concluded that the PRD values provided by system proposed were higher with a value of 1.91 than the values produced by current methods such as 5.3, 5.83, 2.43, 2.5, 3.63, 1.973, 2.29, and 1.911, respectively.

6.5 *Mean Squared Error (MSE)*

MSE calculates the error as average, which is the variation in the calculated values and the estimated value suggested [37]. MSE also tests estimate efficiency and offers an accepted response. The formula of Eq. (7) aids in measuring MSE.

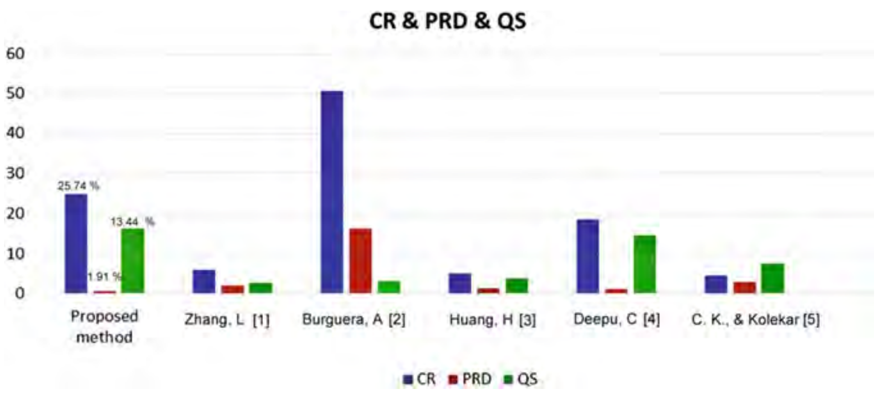


Fig. 4 Comparison between the prosed method and the literatures based on CR, PRD, and QS

Table 1 A brief of overall performance

Performance summary of the recommended technique based on Rec # 117	
Performance metrics	Realized values
CR	25.74
QS	13.44
SNR	52.78
PRD	1.91
MSE	0.2

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i)^2 \tag{7}$$

RMS value for various current methods is shown from Table 7 and suggested. RMS is proved that the suggested approach gives as MSE the value of 0.2 which is the lowest error value as compared with another methods. The lowest error value improves model accuracy and provides added reliability [38].

Based on Table 2, optimum values were obtained by the proposed method, namely CR = 25.74%; PRD = 1.91%; and QS = 13.44% (Table 1).

7 Conclusion

In order to improve CR in the e-health systems, a lossless compression method was introduced. The simplistic recommended approach reduces the need for memory space and the bandwidth of the network. The DCT function is clearly defined by the CR. Security and privacy are the two most important transactions on which IoMT

Table 2 Comparison of recommended algorithm performance with an actual lossless algorithm

ECG records No.	CRs (%)	PRDs (%)	QSs (%)	SNRs (dB)
100	19.85	3.28	6.05	52.77
101	17.95	3.60	4.99	55.70
102	21.35	4.18	5.11	51.65
103	19.33	4.09	4.72	59.75
104	21.16	5.30	3.99	53.58
105	20.02	4.57	4.38	57.78
106	18.66	4.92	3.79	57.48
107	21.94	5.58	3.93	56.99
108	23.18	4.71	4.92	49.96
109	22.36	5.01	4.47	56.16
111	21.17	5.58	3.79	52.58
112	22.91	2.07	11.08	47.31
113	20.81	4.53	4.60	60.23
114	22.48	4.17	5.39	51.09
115	22.25	3.13	7.11	55.33
116	19.47	2.89	6.75	57.75
117	25.74	1.91	13.44	52.78
118	23.83	3.31	7.19	48.85
119	26.01	4.41	5.89	51.00
121	32.47	2.36	13.73	53.14
122	21.71	2.39	9.10	54.93
123	24.38	2.69	9.06	50.92
124	27.12	2.20	12.34	57.97
200	22.64	7.64	2.96	50.04
201	16.04	3.71	4.33	60.18
202	24.73	4.72	5.24	58.11
203	19.37	7.14	2.71	51.88
205	19.68	3.04	6.48	53.05
207	28.23	6.15	4.59	54.75
208	21.36	7.83	2.73	50.64
209	15.38	4.61	3.33	56.70
210	21.87	5.92	3.70	53.28
212	16.77	5.52	3.04	55.51
213	15.61	4.04	3.86	60.41
214	24.15	5.74	4.21	55.14
215	20.18	8.65	2.33	46.92

(continued)

Table 2 (continued)

ECG records No.	CRs (%)	PRDs (%)	QSs (%)	SNRs (dB)
217	22.48	4.95	4.54	59.55
219	23.35	4.80	4.86	49.33
220	20.80	3.20	6.49	53.22
221	22.14	5.36	4.13	54.94
222	21.66	5.49	3.95	51.70
223	21.00	2.73	7.70	61.04
228	27.12	8.41	3.22	47.73
230	18.52	5.28	3.51	56.13
231	19.61	4.89	4.01	56.48
232	17.09	5.24	3.26	44.62
233	19.74	6.68	2.95	53.37
234	19.10	4.51	4.24	59.62
Average	21.56	4.65	5.38	54.17

apps depend. A high accuracy was obtained, as confirmed by the evaluation results, as well as a high security, utilizing the HufCod approach. This approach contains EncDec. In section of finding, different measurements with CR, MSE, and SNR with various extant techniques were compared. The recommended future work includes the development of apps utilized in the real-time compression processing in order to obtain the least computation time.

References

1. M. Elgendi, Less is more in biosignal analysis: compressed data could open the door to faster and better diagnosis. Diseases (2018)
2. S. Kalaivani, C. Tharini, Analysis and modification of rice Golomb coding lossless compression algorithm for wireless sensor networks. J. Theor. Appl. Inform. Technol. **96**(12), 3802–3814 (2018)
3. C. Tan, L. Zhang, H.-T. Wu, A novelBlaschke unwinding adaptiveFourier-decomposition-based signal compression algorithm with application on ECG signals. IEEE J. Biomed. Health Inform. **23**(2), 672–682 (2019)
4. A. Burguera, Fast QRS detection and ECG compression based on signal structural analysis (2019)
5. H. Huang, S. Hu, Y. Sun, ECG signal compression for low-power sensor nodes using sparse frequency spectrum features, in *IEEE Biomedical Circuits and Systems Conference (BioCAS)* (2018)
6. https://www.drugs.com/cg/heart-palpitations-in-adolescents.html
7. A. Burguera, Fast QRS detection and ECG compression based on signal structural analysis. IEEE J. Biomed. Health Inform. **23**(1), 123–131 (2019)
8. A.S. Abdulbaqi, I.Y. Panessai, Designing and ımplementation of a biomedical module for vital signals measurements based on embedded system. Int. J. Adv. Sci. Technol. (IJAST) **29**(3), 3866–3877 (2020)

9. C.J. Deepu, C.-H. Heng, Y. Lian, A hybrid data compression scheme for power reduction in wireless sensors for IoT. IEEE Trans. Biomed. Circuits Syst. **11**(2), 245–254 (2017)
10. C.K. Jha, M.H. Kolekar, ECG data compression algorithm for telemonitoring of cardiac patients. Int. J. Telemed. Clin. Pract. **2**(1), 31–41 (2017)
11. A.S. Abdulbaqi et al., *Recruitment Internet of Things For Medical Condition Assessment: Electrocardiogram Signal Surveillance*, *Special Issue, AUS Journal*, (Institute of Architecture and Urbanism, University of Austral de Chile, 2019), pp. 434–440
12. T.-H. Tsai, W.-T. Kuo, An efficient ECG lossless compression system for embedded platforms with telemedicine applications. IEEE (2018)
13. A.E. Hassanien, M. Kilany, E.H. Houssein, Combining Support Vector Machine and Elephant Herding Optimization for Cardiac Arrhythmias. arXiv:1806.08242v1[eee.SP], June 20, 2018
14. J. Dogra, M. Sood, S. Jain, N. Prashar, Segmentation of magnetic resonance images of brain using thresholding techniques,in *4th IEEE International Conference on signal processing and control (ISPCC 2017)*, Jaypee University of Information technology, Waknaghat, Solan, H.P, India, pp. 311–315, September 21–23, 2017
15. N. Prashar, S. Jain, M. Sood, J. Dogra, Review of biomedical system for high performance applications ,4th *IEEE International Conference on signal processing and control (ISPCC 2017)*, Jaypee University of Information technology, Waknaghat, Solan, H.P, India, pp 300–304, September 21–23, 2017
16. A. Dhiman, A. Singh, S. Dubey, S. Jain, Design of lead II ECG waveform and classification performance for morphological features using different classifiers on lead II. Res. J. Pharmaceut. Biol. Chem. Sci. (RJPBCS) **7**(4), 1226–1231 (2016)
17. B. Pandey, R.B. Mishra, An integrated intelligent computing method for the detection and interpretation of ECG based cardiac diseases. Int. J. Knowl. Eng. SoftData Paradigms **2**, 182–203 (2010)
18. A.S. Abdulbaqi, S.A.M. Najim, R.H. Mahdi, Robust multichannel EEG signals compression model based on hybridization technique. Int. J. Eng. Technol. **7**(4), 3402–3405 (2018)
19. S. Kalaivani, I. Shahnaz, S.R. Shirin, C. Tharini, Real-time ECG acquisition and detection of anomalies, in *Artificial İntelligence and Evolutionary Computations in Engineering Systems*, ed. S.S. Dash, M.A. Bhaskar, B.K. Panigrahi, S. Das (Springer, Berlin, 2016)
20. J. Uthayakumar, T. Venkattaraman, P. Dhayachelvan, A survey on data compression techniques: from the perspective of data quality, coding schemes, data types, and applications. J. King Saud Univ.- Comput. Inform. Sci. (2018)
21. R. Gupta, S. Singh, K. Garg, S. Jain, Indigenous design of electronic circuit for electrocardiograph . Int. J. Innov. Res. Sci. Eng. Technol. **3**(5), 12138–12145 (2014)
22. C.C. Chiu, T.H. Lin, B.Y. Liau, Using correlation coefficient in ECG waveforms for arrhythmia detection. Biomed. Eng. Appl. Basis Commun. **17**, 147–152 (2005)
23. S. Jain, Classification of protein kinase B using discrete wavelet transform. Int. J. Inform. Technol. **10**(2), 211–216 (2018)
24. N. Alajlan, Y. Bazi, F. Melgani, S. Malek, M.A. Bencherif, Detection of premature ventricular contraction arrhythmias in electrocardiogram signals with kernel methods. SIViP **8**(5), 931–942 (2014)
25. Y. Hirai, T. Matsuoka, S. Tani, S. Isami, K. Tatsumi, M. Ueda, T. Kamata, A biomedical sensor system with stochastic A/D conversion and error correction by machine learning. IEEE Access **7**, 21990–22001 (2019)
26. Ö. Yildirim, A novel wavelet sequence based on a deep bidirectional LSTM network model for ECG signal classification. Comput. Biol. Med. **96**, 189–202 (2018)
27. A. Diker, D. Avci, E. Avci, M. Gedikpinar, A new technique for ECG signal classification genetic algorithm Wavelet Kernel extreme learning machine. Optik **180**, 46–55 (2019)
28. J. Zhang, Z. Gu, Z.L. Yu, Y. Li, Energy-efficient ECG compression on wireless biosensors via minimal coherence sensing and weighted l1 minimization reconstruction. IEEE J. Biomed. Health Inform. **19**(2), 520–528 (2015)
29. A. Singh, S. Dandapat, Block sparsity-based joint compressed sensing recovery of multichannel ECG signals. Healthcare Technol. Lett. **4**(2), 50–56 (2017)

30. A. Singh, S. Dandapat, Exploiting multi-scale signal information in joint compressed sensing recovery of multi-channel ECG signals. Biomed. Signal Process. Control **29**, 53–66 (2016)
31. H. Mamaghanian, G. Ansaloni, D. Atienza, P. Vandergheynst, Power-efficient joint compressed sensing of multi-lead ECG signals, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 2014, pp. 4409–4412 (2014)
32. S. Kumar, B. Deka, S. Datta, Block-sparsity based compressed sensing for multichannel ECG reconstruction, in *Pattern Recognition and Machine Intelligence. PReMI 2019. Lecture Notes in Computer Science*, vol. 11942, ed. by B. Deka, P. Maji, S. Mitra, D. Bhattacharyya, P. Bora, S. Pal (Springer, Cham, 2019)
33. S. Eftekharifar, T.Y. Rezaii, S. Beheshti, S. Daneshvar, Block sparse multi-lead ECG compression exploiting between-lead collaboration. IET Sig. Process. (2018)
34. A. Sharma, A. Polley, S.B. Lee, S. Narayanan, W. Li, T. Sculley, S. Ramaswamy, A Sub-60-μ A multimodal smart biosensing SoCwith>80-dB SNR, 35μAphotoplethysmography signal chain. IEEE J. Solid-State Circuits **52**(4), 1021–1033 (2017)
35. Z. Zhang, J. Li, Q. Zhang, K. Wu, N. Ning, Q. Yu, A dynamic tracking algorithm based SAR ADC in bio-related applications. IEEE Access **6**, 62166–62173 (2018)
36. M.K. Adimulam, M.B. Srinivas, A 1.0 V, 9.84 fJ/c-s FOM reconfigurable hybrid SAR-sigma delta ADC for signal processing applications. Analog Integr. Circ. Sig. Process **99**(2), 261–276 (2019)
37. X. Zhang, Y. Lian, A 300-mV 220-NW Event-driven ADC with real-time QRS detection for wearable ECG sensors. IEEE Trans. Biomed. Circuits Syst. **8**(6), 834–843 (2014)
38. Y. Hou, J. Qu, Z. Tian, M. Atef, K. Yousef, Y. Lian, G. Wang, A 61-NW level-crossing ADC with adaptive sampling for biomedical applications. IEEE Trans. Circuits Syst. II Express Briefs **66**(1), 56–60 (2019)

Anomaly Detection in Real-Time Surveillance Videos Using Deep Learning

Aswathy K. Cherian and E. Poovammal

Abstract The real-time events are fast and occurring at highly dynamical moments. Hence, the important challenges are identifying the anomaly incidents properly. The specified methods and techniques are to be quick in identification for control and other measures of the events. In the proposed method, the anomalies are detected from the surveillance videos using the multiple instance learning and ID3 for extracting the features. The extracted features are then used as input to a deep neural network where the classification of the videos to anomalous and normal videos is done. The investigated dataset is with 128 hours of videos with ten percent of different realistic anomaly videos. The AUC of the proposed approach is 81. The proposed approach is most beneficial for the real-world anomaly recognition in surveillance videos.

Keywords Video surveillance · Anomaly detection · Multiple instance learning · I3D algorithm · Deep learning

1 Introduction

The monitoring of the real-world events simultaneously at many places are really becoming challenging due to higher volume of data. Recently, image processing has been gained momentum to provide societal society. The automatic and faster detection is essential to trace the events without human errors. Wei et al. [1] used a convolutional neural network (CNN) to carry out the scenes and objects to account for video anomalies' spatial and temporal aspects. Abnormal vehicle behavior of live stream videos was detected using developed application and evaluation by Wang et al. [2]. Ahmadi et al. [3] used a sequential topic modeling for the traffic scene abnormalities through sparse topical coding. An optimized CNN and genetic algorithm

A. K. Cherian (✉) · E. Poovammal
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai, India
e-mail: aswathyc@srmist.edu.in

E. Poovammal
e-mail: poovamme@srmist.edu.in

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_19

were used for anomaly detection in surveillance videos, and the presented ML technique was helpful to detect the video anomaly [4, 5]. A multistage pipeline approach is used to detect the video anomaly with an incrementally trained Bayesian CNN. Mehta et al. [6] built a deep learning model for anomaly detection of gun and fire violence and the detection rate with 45 fps. They showed the accuracy of the model around 87–90%. Ullah et al. [7] used a pre-trained CNN to extract the spatiotemporal features from the series of frames using a multilayer and bidirectional memory. Pawar and Attar [8] investigated the frame-level identification of anomaly detection using a two-dimensional convolutional auto-encoder and radial basis function with two real-world datasets. Tang et al. [9] integrated reconstruction and prediction for detecting anomalies and demonstrated at 30 fps. Iqbal et al. [10] used canny edge detection with Hough transform for high-resolution cloud videos to anomaly detection. Kisan et al. [11] developed a model using a change in the motion vector to detect real-world anomalies. CNN methods are used to detect anomaly due to weather and galaxy visual changes using ML [12, 13]. There are plenty of approaches that are used to detect the anomaly in the surveillance videos as per the literature. Recently, [14, 15] adopted auto-encoders that developed a model for normal video classification. The anomalies are determined using a reconstruction loss. In the proposed model, the prediction of both normal and anomaly video is made even though the training data is weakly labeled.

2 Dataset

Our proposed method uses the videos from the Sultani [16]. This is one of the largest real-time dataset available for anomaly detection. It contains around 1900 real-time surveillance videos. These video addresses are different scenarios of public life. The total length of the video is around 128 h with 13 different types of events. The events that are covered in this dataset include vandalism, shoplifting, fighting, explosion, assault, arrest, arson, burglary, stealing, robbery, fighting, abuse, and accidents. A sample screenshot of the video of anomalies is shown in Fig. 1. The figure represents around 14 anomalies with four frames of each from the dataset. The first set shows example of abuse followed by arson. The accident anomaly depicts a car hit by a bus. The last frames represent a normal video with no anomalies. It represents frames of a mall with people moving in and out. A total dataset is divided into two sets, one for training and the other for testing. The training consists of around 800 normal images and 810 testing images, whereas the testing contains 150 and 140 normal and anomalous videos, respectively. These videos are unlabeled raw data [17]. The number of videos in each event distributed for training and testing is given in Table 1.

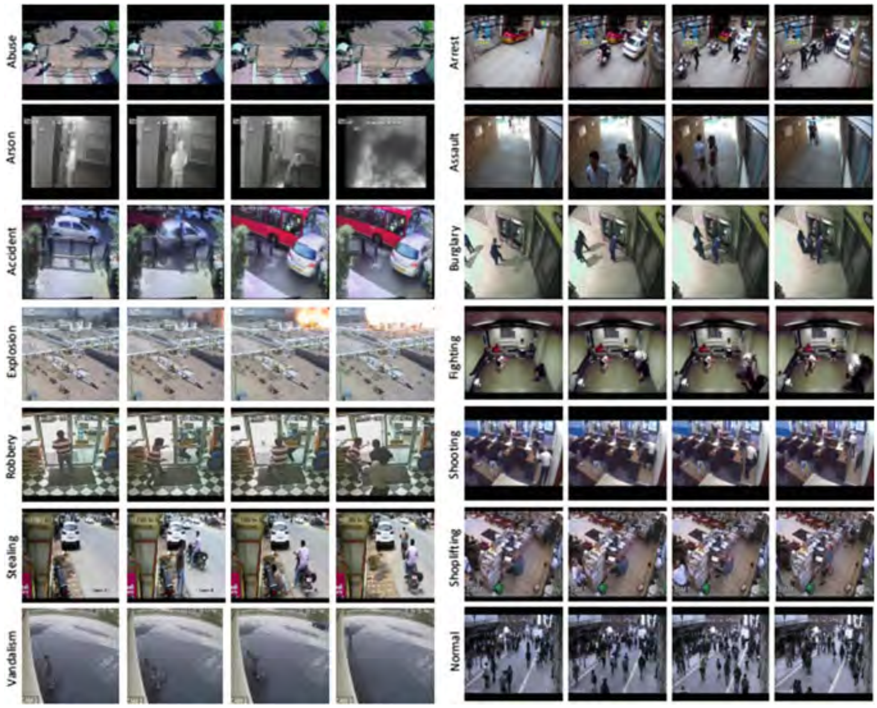


Fig. 1 Anomalies from the training and testing videos (sample sets) [16]

Table 1 Distribution of videos for testing and training of anomalies

Anomaly	Training set	Testing set
Abuse	48	50
Fighting	45	50
Arrest	45	50
Accidents	127	150
Arson	41	50
Robbery	145	150
Assault	47	50
Shooting	27	50
Burglary	87	100
Shoplifting	29	50
Explosion	29	50
Stealing	95	100
Vandalism	45	50

3 Proposed Research Methodology

The diagrammatic representation of the proposed method is shown in Fig. 2. The input video (set) is first divided into fixed number of segments, say 32 during the training. The segments also known as instances are separated into positive and negative set based on the MIL ranking loss function. The anomalous (positive) and normal (negative) is used to absorb features using the I3D network. Later, a classifier is used to predict the videos as anomalous or normal.

3.1 Multiple Instance Learning

The videos of the given dataset are raw and unlabeled. Labeling of this huge dataset is tedious, time consuming, and human impossible. For this purpose, we adapt a methodology called the multiple instance learning. Here, we propose a deep MIL framework where each video is treated as set and small segments of the video are treated as instance of these videos. MIL considers that all videos do not have accurate temporal annotations. This framework unlike the state-of-the-art frameworks do not require the exact location of anomaly in videos, instead it only requires a video level labeling indicating the presence of anomaly. The video that does not contain anomaly is labeled/grouped as negative and the video that contains anomaly is labeled as positive. The negative set is represented as S_m with $n_1, n_2, \ldots, n_m$ as negative instance and P_m represents the positive instances with $p_1, p_2, \ldots, p_m$. These instances are segments of the video. Every set contains instances of the video which are non-overlapping segments of the training video. Normally, the number of instances from each video is 32. In each set, the optimized objective function with respect to maximum scored instance is represented as Eq. 1:

$$\min_{\mathbf{w}} \left[\frac{1}{z} \sum_{j=1}^{z} \max\left(0, 1 - Y_{B_j}(\max_{i \in B_j}(\mathbf{w}.\phi(x_i)) - b)\right) \right] + \|\mathbf{w}\|^2 \tag{1}$$

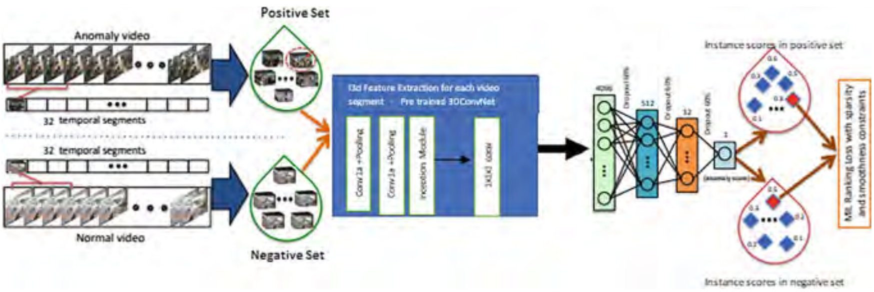


Fig. 2 Schematic representation of the proposed detection of anomalies [16]

where the total number of sets is given by z, the classifier to be learned is w, the set level label is denoted by Y_{Bj}, and $\phi(x)$ denotes feature representation of an image patch or a video segment.

3.2 Deep MIL Ranking Model

The score of anomalous videos is assumed to be higher than the normal video considering anomaly detection to be a regression model. This makes the value of ranking loss function to be higher for anomalous video segments, that is $f(V_n) < f(V_a)$ where V_a and V_n represent anomalous and normal video segments and $f(V_a)$, $f(V_n)$ are their corresponding ranking scores. As the annotations are only labeled during the training of the dataset, the above assumption is not suited. In such situations, we used the multiple instance ranking objective function:

$$\max\ i \in B_a\ \ f(V_{ia}) > \max\ i \in B_n\ \ f(V_{\text{in}}), \tag{2}$$

where max is taken over all video segments in each set. The ranking is enforced only on two instances that are having maximum score in positive and negative sets. The segment in the negative set has the greatest anomaly score. This represents the video to be similar to an anomalous segment, but in real, it is a normal instance. There occurs a drawback of false alarm in anomaly detection which is occurred when the negative instance is considered as a hard instance. By using Eq. 2, we are pushing the positive instances and negative instances far apart in terms of anomaly score.

3.3 I3D Algorithm

Feature extraction and classification is done by using the I3D algorithm and a three-layered FC neural network. Before extracting features all videos in the dataset is resized/trimmed to 240 × 320 pixels at a frame rate of 30 frames per pixel. The C3D network utilizes a single 3D network, whereas I3D network uses two 3D networks which works similar to a pre-trained 2D network in 3D. The architecture of I3D is shown in Fig. 3. C3D causes overfitting due to the unavailability of huge video dataset and the complicated structure of the 3D convolutional kernels. The proposed method utilizes the I3D (Fig. 3) [18] is the inception 3D CNN. In ID3 structure, the 2D convolutional kernels are inflated/expanded to 3D [19], that is, all the square filters (N2) are converted to cubic filters (N3) [20, 21]. The beginning of the network is found that with asymmetric filters for max-pooling, this helps to maintain time while pooling over the spatial dimension. It also analyzes the spatial information at various positions, and an average result is produced. I3D concentrates to grow the network wider instead of deeper. The overfitting of the model is avoided by training the model on the video dataset. This improves the accuracy of the model. The 1 ×

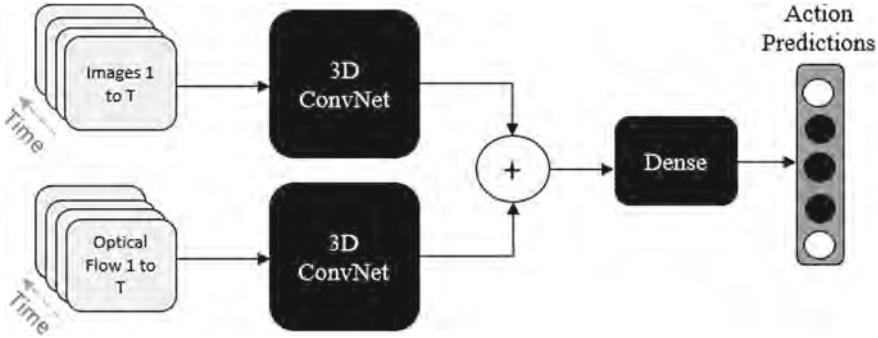


Fig. 3 Architecture of I3D

1 x 1 convolution reduces the input channel from the 3 × 3 x 3, thus reducing the computational cost.

3.4 Classification

The features extracted with I3D algorithm is fed to a three-layered fully connected neural network. The first fully connected layer has 512 neurons, and the activation function used is sigmoid [22]. The second layer contains 32 units followed by 1 unit. This is because of 60% dropout in each layer [23]. In the last layer, rectified linear unit (ReLU) is used as the activation function, and the initial learning rate recording was 0.0001 with Adagrad optimizer.

4 Discussion

As per the anomaly detection described in the literature, our approach is validated with the proposed model, and the comparative results are discussed here. Table 2

Table 2 Comparison of the present results with literature

Method	AUC
Binary classifier	50.0
Sultani et al. [16]	75.02
Hasan et al. [15]	50.6
Lu et al. [24]	65.51
The proposed method	81.16

Table 3 Accuracy recognition of C3D and I3D (feature extraction techniques)

Method	C3D	I3D
Accuracy	23.0	25.9

provides the area under the curve (AUC) of the proposed method with other state-of-the-art methods. Table 3 shows the quantitative accuracy result of the I3D and C3D feature extraction techniques.

5 Conclusions and Future Works

The proposed approach is found to be deterministic to real-time situations. Thus, the investigated algorithm is most beneficial for the anomaly detection of the surveillance videos. Monitoring the real-world actions is certainly challenging to quantify as per space and time. The presented method is observed with an AUC of 81. The results are helpful for the classification of real-world anomalies in surveillance videos. The future work of this paper would be to introduce an algorithm which would extract features from the videos with more precision, which in turn improves the accuracy at which the videos are classified. Also, the classifier does not provide the correct output in case of blurred or noisy images. Fine-tuning of the parameter of the classifier is expected in this respect.

References

1. H. Wei, K. Li, H. Li, Y. Lyu, X. Hu, Detecting video anomaly with a stacked convolutional LSTM framework. Lect Notes Comput Sci 2019;11754 LNCS, pp. 330–342
2. C. Wang, A. Musaev, P. Sheinidashtegol , T. Atkison, Towards detection of abnormal vehicle behavior using traffic cameras. Lect Notes Comput Sci 2019;11514 LNCS, pp. 125–136
3. P. Ahmadi, E.P. Moradian, I. Gholampour, Sequential topic modeling for efficient analysis of traffic scenes, in *9th International Symposium on Telecommunication: With Emphasis on Information and Communication Technology*, IST 2018 (2019)
4. D. Thakur, R. Kaur, An optimized CNN based real world anomaly detection in surveillance videos. Int. J. Innov. Technol. Explor. Eng. **8**(9 Special Issue), 465–473 (2019)
5. A. Joshi, V.P. Namboodiri, Unsupervised synthesis of anomalies in videos: transforming the normal, in *Proceedings of the International Joint Conference on Neural Networks* (2019)
6. P. Mehta, A. Kumar, S. Bhattacharjee, Fire and gun violence based anomaly detection system using deep neural networks, in *Proceedings of the International Conference on Electronics and Sustainable Communication Systems, ICESC 2020* (2020)
7. W. Ullah, A. Ullah, I.U. Haq, K. Muhammad, M. Sajjad, S.W. Baik, CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. Multimedia Tools Appl. (2020)
8. K. Pawar, V. Attar, Deep learning-based intelligent surveillance model for detection of anomalous activities from videos. Int. J. Comput. Vis. Rob. **10**(4), 289–311 (2020)

9. Y. Tang, L. Zhao, S. Zhang, C. Gong, G. Li, J. Yang, Integrating prediction and reconstruction for anomaly detection. Pattern Recogn. Lett. **129**, 123–130 (2020)
10. B. Iqbal, W. Iqbal, N. Khan, A. Mahmood, A. Erradi, Canny edge detection and Hough transform for high resolution video streams using Hadoop and Spark. Cluster Comput. **23**(1), 397–408 (2020)
11. S. Kisan, B. Sahu, A. Jena, S.N. Mohanty, Detection of violence in videos using hybrid machine learning techniques. Int. J. Adv. Sci. Technol. **29**(3), 5386–5392 (2020)
12. A.K. Cherian, A. Rai, V. Jain, Flight trajectory prediction for air traffic management. J. Crit. Rev. **7**(6), 412–416 (2020)
13. A.K. Cherian, P. Kumar, P.S.K. Reddy, E. Poovammal, Detecting bars in galaxies using convolutional neural networks. J. Crit. Rev. **7**(6), 189–194 (2020)
14. D. Xu, E. Ricci, Y. Yan, J. Song, N. Sebe, Learning deep representations of appearance and motion for anomalous event detection, in *BMVC* (2015)
15. M. Hasan, J. Choi, J. Neumann, A.K. Roy-Chowdhury, L.S. Davis, Learning temporal regularity in video sequences, in *CVPR* (June 2016)
16. W. Sultani, C. Chen, M. Shah, Real-world anomaly detection in surveillance videos, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2018)
17. https://www.dropbox.com/sh/75v5ehq4cdg5g5g/AABvnJSwZI7zXb8_myBA0CLHa?dl=0
18. J. Duchi, E. Hazan, Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization. J. Mach. Learn. Res. (2011)
19. J. Carreira, A. Zisserman, Quo vadis, action recognition? a new model and the kinetics dataset, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6299–6308
20. S. Xie, C. Sun, J. Huang, Z. Tu, K. Murphy, Rethinking spatiotemporal feature learning: speed-accuracy trade-offs in video classification, in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 305–321
21. X. Wang et al., I3D-LSTM: a new model for human action recognition, in *IOP Conference Series: Materials Science and Engineering* (2019)
22. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. (2014)
23. V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, in *Proceedings of the 27 th International Conference on Machine Learning*, Haifa, Israel (2010)
24. C. Lu, J. Shi, and J. Jia. Abnormal event detection at 150 fps in matlab, in *ICCV* (2013)

Convolutional Neural Network-Based Approach for Potholes Detection on Indian Roads

Noviya Balasubramanian, J. Dharneeshkar, Varshini Balamurugan, A. R. Poornima, Muktha Rajan, and R. Karthika

Abstract In developing countries like India, soaring count of potholes on roads is a cardinal responsibility, as accidents can occur due to their presence. Thus, it is imperative to detect them to ensure the safety of people. Many research works have been carried out to easily detect them but most of them are uneconomical. Using deep learning algorithms to detect potholes has become popular in recent times. Convolutional neural networks are very effective in identifying objects, and hence, this approach has been adopted. A 4000 image dataset is created using data augmentation techniques on the already existing 1500 image dataset. The created dataset is trained by using FasterR-CNN, SSD, YOLOv3 tiny, and YOLOv4 tiny algorithms. Since, there is always a trade-off between latency and accuracy in object detection algorithms, different architectures are employed. The final results are compared based on mAP. SSD with MobileNetv2 and YOLOv4 tiny outperforms other methods with a precision of 76% and 76.4%, respectively.

Keywords Pothole detection · R-CNN · SSD · Data augmentation · YOLO tiny · Image processing

1 Introduction

Potholes on the roads are a noticeable huge constructional catastrophe in the plane of a road, which is caused by traffic and bad weather. It has now become a common nuisance on our Indian roadways. Not only do they make roads look rundown and unsightly, but also they pose a threat to the safety of people who travel on these roads. It not only causes damage to vehicles but also causes accidents and severe injuries to anyone who is involved. Potholes are dangerous for pedestrians, bicyclists, and road workers as well. Anyone who uses the road could be injured. According to most

N. Balasubramanian · J. Dharneeshkar · V. Balamurugan · A. R. Poornima · M. Rajan · R. Karthika (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: r_karthika@cb.amrita.edu

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*, Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_20

recent figures by a few state governments, around 30 deaths happen each day on the streets because of potholes [1]. The death rate in 2017 has increased by more than 50% from the toll in 2016, i.e. to 3597 cases annually [2].

To curtail the number of accidents and other related losses, it is crucial to accurately detect these potholes and repair them. It is not preferred to detect potholes manually since it charges an exorbitant rate and is time-consuming. Therefore, a plethora of research was carried out to evolve a technology that can expose potholes, which would be a milestone in improving the effectiveness of the survey and quality of pavement by prior inspection and speedy response.

One of many reasons for such road accidents is human error since they are unable to see potholes and take impulsive decisions. To overcome this, advanced driver-assistance systems (ADAS) aid drivers in detecting hindrances and potential risks beforehand. ADAS is also responsible for regulating the control, balance, and mobility of the vehicle during precarious situations. ADAS has improved car and road safety through a protected cautious human–machine interface. Technologies like alerting the driver about potential risks, applying safeguards, and even taking control of vehicles during snag have been fabricated to improve safety properties and to get clear steer of road fatalities [3].

Vibration-based methods, three-dimensional (3D) reconstruction-based methods, and vision-based methods are prevailing methods for pothole detection [2]. Each of them has its own downside like high cost or low accuracy. While comparing these three, vision-based methods are superior to other methods in all aspects. The 3D reconstruction needs cost-intensive laser scanners, whereas the vibration-based method is undependable on certain vibration producing surfaces [4]. Besides, the detection of potholes is comparatively bit challenging because of its arbitrary shape and complex geometry.

Our paper proposes a vision-based detection method based on convolutional neural network (CNN) using single-shot multibox detector (SSD), FasterR-CNN, YOLOv3 tiny, and YOLOv4 tiny algorithms for the same. The paper is organized as follows: Sect. 2 describes related works in object detection, Sect. 3 introduces the methodology which explains four different algorithms, and the process carried out for data augmentation. In Sect. 4, experiments are conducted to evaluate the proposed methods of object detection and the results are analysed. Finally, in Sect. 5, we present the conclusion and the future work.

2 Related Work

In recent times, a surfeit of research has been carried out to evolve a technology that can expose potholes using deep learning methods. CNN is one of the cardinal deep learning methods of object detection. Krizhevsky et al. [5] trained a deep neural network with ImageNet architecture and the results conveyed that CNN is efficacious for large datasets using supervised learning. Thus, the convolutional neural network (CNN) approach is adopted for detecting the objects. Jiménez [2] used

Table 1 Comparison of existing methodologies

Method	Description
CNN	Efficacious for large dataset using supervised learning. But, inference time is high
FR-CNN	Gives reasonable precision but it is not preferred for real-time usage for object detection
YOLO	Efficient for real-time usage and faster than traditional CNN algorithms. The speed comes with a trade-off, i.e. accuracy
SSD	Higher accuracy compared to YOLO. Performance is poor for minuscule objects

pothole and non-pothole thermal images as input to CNN-based ResNet models for detecting potholes. To subdue inference time in CNN, a more substantial gain is obtained with the inception of a region-based convolutional neural network (R-CNN) [6]. Following the emergence of R-CNN, plenty of enhanced models have been propounded, including FasterR-CNN, SSD, YOLO, etc. The structure of the FasterR-CNN model comprises two modules. The first, being the regional proposal network (RPN), is a completely convolutional network that fosters object proposals that will be passed onto the next module. Fast R-CNN detector distinguishes proposals from previous modules and reverts a bounding box around the object [7]. Chebrolu et al. [8] used a deep learning model, to execute and spot pedestrians at all times during a day. The work was done using FasterR-CNN. The technique achieved a reasonable precision. Gokul et al. [9] compared the performance of FasterR-CNN and YOLO for spotting traffic lights. The results showed that FasterR-CNN with Inceptionv2 and ResNet-101 outperformed YOLO. But, for real-time usage, YOLO performed better. YOLO has more advantages compared to other object detection algorithms. Unlike other models, YOLO uses the complete image for training and directly enhances prediction performance [10]. An altered DL model for spotting vehicles on roads during blockage times was implemented by Haritha and Thangavel [11]. They have also carried out the comparison between the implemented model with YOLOv2 and also with multiscale ConvNets. Dharneeshkar et al. [12] generated a dataset and trained it to detect potholes using different versions of YOLO. This model detected potholes with reasonable accuracy. YOLOv3 tiny showed the highest precision of 0.76 compared to other models. The precision values of YOLOv3, YOLOv2, and YOLOv3 tiny are 0.69, 0.46, and 0.76, respectively. Though YOLO is faster than FasterR-CNN, SSD runs faster than YOLO and is also more accurate. Liu et al. [13] used a single-shot multibox detector which is a rapid single-shot object detector for various classes with a feed-forward neural network to detect objects in the images. The main motive of SSD is to use midget convolutional filters put in to feature maps for forecasting scores of each category and box offset for a predetermined set of default bounding boxes. The results showed that SSD has good accuracy in contrast to other object detection algorithms. Silvister et al. [14] compare SSD and deep neural network (DNN) and conclude that SSD gives fast real-time detection with better accuracy due to its tolerance in aspect ratio, unlike other DNN.

Each algorithm has its own merits and demerits as shown in Table 1. CNN is efficient for large dataset but is slow. FR-CNN is faster than CNN but is not preferred for real-time usage. YOLO is fast and it can be used for real-time usage as it detects potholes with a good accuracy than the traditional CNN algorithms. SSD gives better accuracy than other algorithms with a good speed. To overcome the demerits of each algorithm, we have trained it using FR-CNN, YOLO and SSD. The proposed work detects potholes with a good accuracy.

3 Methodology

In order to carry out pothole detection, an available dataset [12] of 1500 images of potholes is augmented and apportioned into two sets: 80% for training (train set) and 20% for testing (test set). The training set is used as an input to a few proposed CNN algorithms, namely FasterR-CNN, SSD, YOLOv3 tiny, and YOLOv4 tiny for training and validated with test set. Figure 1 presents a schematic flow of the proposed experiment.

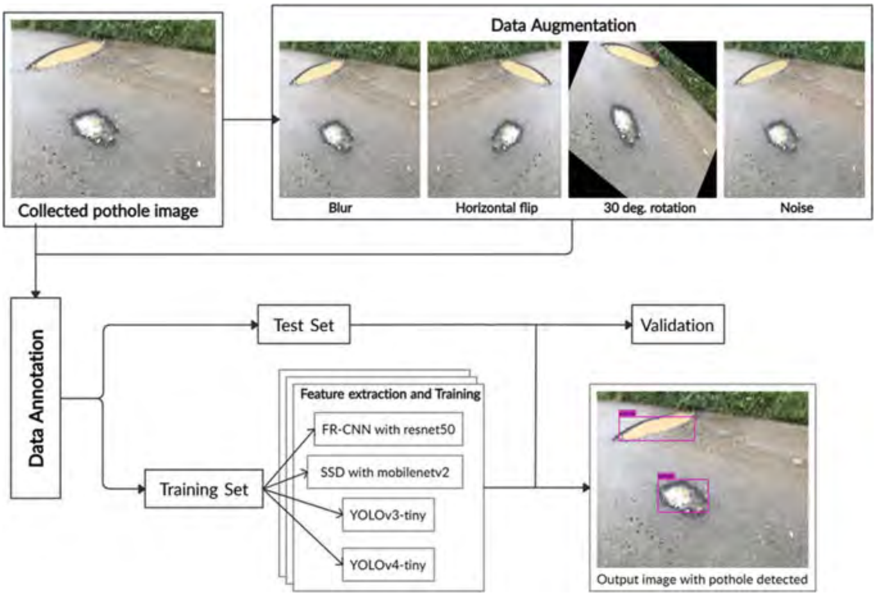


Fig. 1 Block schematic of proposed experimentation process

4 Dataset

Potholes in our country are unique and are different from potholes that are seen elsewhere in the world. Because of this, a dataset of 1500 images has been created [12]. The authors created this dataset by driving in and around Coimbatore. The dataset is a miscellany of pothole images at different angles, distances, lighting, and climatic conditions, which were captured from the dashboard of a car. The images are of good standard. The images are resized to 1024 * 768. Some images are shown in Fig. 2. With these dataset, additionally data augmentation is performed on the 1500 image dataset to increase the size of the dataset and annotation is performed using LabelImg.

Data Augmentation: Data augmentation is a technique used to expand the amount of dataset by newly creating data from existing data. It helps us to increase the size of the dataset and introduce variety in the dataset without collecting new data. The neural network treats these images as distinct images. Data augmentation helps

Fig. 2 Images from the [12] dataset

reduce overfitting [2]. Hence, a few amendments have been made with the available data [12] to obtain more data. Various augmentation techniques have been carried out to convert the dataset into a larger dataset to make it better in training and also to prevent overfitting.

The data augmentation carried out include:

(i) **Rotating the image** (Fig. 3a): Image of a pothole after the rotation will look like a pothole and will look as if the image was taken from a different gradient [2]. We have rotated the image clockwise by 30°.

(ii) **Horizontal Flipping** (Fig. 3b): It leaves the dimensions of the layer and the pixel information unchanged. Horizontal flip is used for images that appear almost identical if flipped horizontally.

(iii) **Adding random noise** (Fig. 3c): Gaussian noise is added to the image for data augmentation. Gaussian noise has zero mean. It has data points in all frequencies and this distorts the high-frequency features. Adding the right quantity of noise improves the learning capability [15].

(a) (b)

(c) (d)

Fig. 3 Some images after data augmentation

(iv) **Blurring the image** (Fig. 3d): Blurring of an image involves taking neighbouring pixels and averaging them [2]. This reduces detail and makes the image blur. The dataset size increases and the training of the model become better.

(v) **Translation of image**: Translation of the image means moving the image in a particular direction. It is very useful as objects can be found anywhere in the image. This makes the convolutional neural network to search everywhere and makes the training better.

By carrying out the above data augmentation techniques, the 1500 image dataset is converted into 4000 image datasets and object detection algorithms have been used to detect the potholes.

4.1 Proposed CNN Models

The train set after splitting is then passed to different CNN-based feature extraction architectures and applied four different algorithms, FR-CNN, SSD, YOLOv3-tiny, YOLOv4-tiny, and trained.

Faster Region-based Convolutional Neural Network (FR-CNN): A FasterR-CNN object detection network consists of a feature extraction network followed by two subnetworks. Usually, a pre-trained CNN networks such as ResNet-50 act as the feature extractor. The first among the two subnetworks that follow the feature extraction network is the region proposal network (RPN), which generates proposals to indicate the area of the image in which the object is likely to exist. The next subnetwork is trained to predict the class to which the object belongs as depicted in Fig. 4.

Here, FasterR-CNN with ResNet50 is used in which ResNet refers to residual neural networks. Residual networks are preferred because they use skip connections to skip over layers that help prevent the problem of disappearing gradients. This process is done until the weights from the previous layer are learned [7].

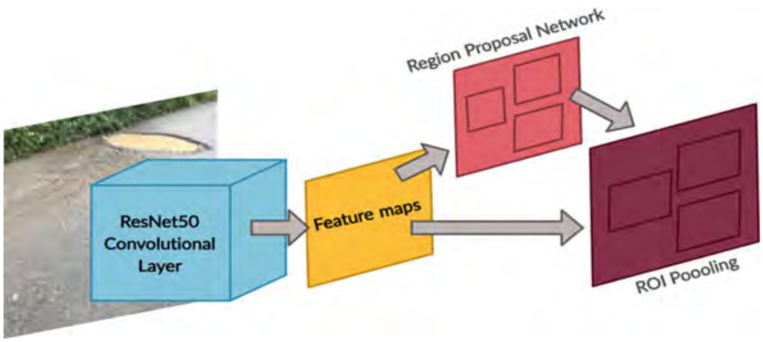


Fig. 4 FasterR-CNN architecture [7]

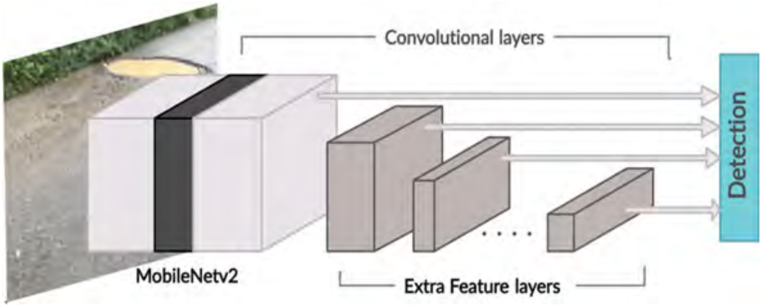


Fig. 5 SSD architecture [13]

Single-Shot Multibox Detector (SSD): As shown in Fig. 5, SSD approach is a simple, single deep neural network, which eliminates all the pre-processing techniques and encapsulates both localization and detection tasks in a single forward sweep of the network. This makes SSD facile to train and simple to consolidate into systems that need detection. The base convolution is derived from an existing image classification architecture, MobileNet-version2, a lightweight deep convolutional neural network that uses depth and point-wise separable convolution that will provide lower-level feature maps. The below mentioned formula is used for computing the scale of the default boxes for all feature maps. Scale (minimum), i.e. the lowest layer, has a value of 0.2 and Scale (maximum), i.e. the highest layer has a value of 0.9. fm feature maps are used for prognostication.

$$\text{Scale}_k = \text{Scale}_{\min} + \frac{\text{Scale}_{\max} - \text{Scale}_{\min}}{\text{fm} - 1}(k - 1), k \in [1, \text{fm}] \quad (1)$$

Therefore, SSD with MobileNetv2 carries out less computation than typical convolution (VGG-16) with only a small reduction in accuracy. During auxiliary convolution, the top of the base network incorporates predictions from a collection of default bounding boxes with respective feature map cells. Prediction convolution is to locate and identify objects in these feature maps by generating a set of fixed-size bounding boxes over different scales to match the shape of the object and scores for the existence of each object category in the respective default box [13].

You Only Look Once (YOLO): YOLO [16, 17] is an object detection system in real time that recognizes various objects in a single enclosure, as illustrated in Fig. 6. Moreover, it identifies objects more rapidly and precisely than other recognition systems. FasterR-CNN has proved to be accurate; however, it is comparatively slow. So, YOLO was built to boost the speed and also to obtain super real-time performance. Since it is completely based on convolutional neural networks, it isolates a particular image into regions and envisions the confined-edge box and probabilities of every region. In YOLO, bounding box and class predictions are done concurrently, and this objective makes YOLO different from other conventional systems. The probability

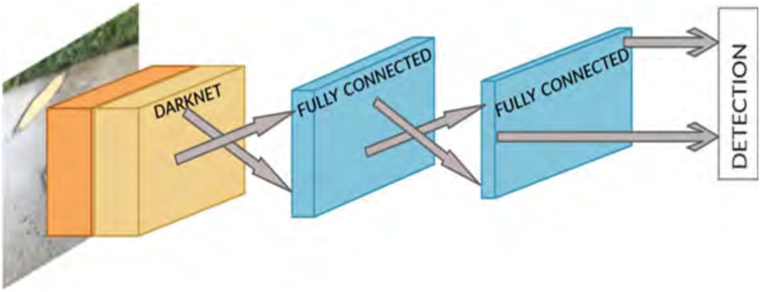


Fig. 6 YOLO architecture [10]

of an object in every bounding box is known as confidence score, which can be calculated as:

$$C = P(\text{object}) \times \text{IOU}^{\text{truth}}_{\text{predictions}}. \tag{2}$$

The term IOU refers to intersection over union. The value of IOU being close to 1 indicates that predicted bounding box is near to ground truth.

YOLOv3 tiny is a simplified version of YOLOv3, with less number of convolutional layers. Hence, the running speed is increased remarkably with a minuscule reduction in accuracy. YOLOv3 tiny uses a pooling layer for feature extraction with Darknet-53 architecture. However, its convolutional layer structure still uses the same structure as YOLOv3 but with reduction in the dimensions of the convolutional layer [18]. Similarly, YOLOv4 tiny is a simplified version of YOLOv4, an object detector that is typically pre-trained on ImageNet classification. The feature extraction in YOLOv4 tiny is done by encapsulating a few convolutional and max-pooling layers from YOLOv4 in a single layer. Based on different parameters, CSPDarknet53 is chosen as the network backbone of YOLOv4 architecture.

5 Experiment

The experiment was set up in collaboratory, which is a free cloud service by Google research, based on Jupyter notebooks with free-of-charge access to a robust GPU Nvidia K80s, T4s, P4s, and P100s [19]. The experiment is executed using different modern CNN algorithms such as FasterR-CNN, SSD, YOLOv3 tiny, and YOLOv4 tiny. Each architecture has its own trade-off between latency and accuracy. In order to get a nominal compromise between mean average precision and execution time, different architecture models are chosen and trained. TensorFlow frameworks [20] are used for training models. For FasterR-CNN, ResNet50, a pre-trained weight of 50 convolutional layers and for SSD, MobileNetv2, pre-trained weight of 53 convolutional layers are imported. For YOLOv3 tiny and YOLOv4 tiny, the reduced configuration of Darknet-53 and CSPDarknet53, respectively, is the pre-trained weight for convolutional layers. The number of classes is changed to 1 as we have only one class, i.e. pothole. Due to the low graphic capability of the collab, the batch size is fixed as 6. After amending these changes in the configuration, training is established. For every 200 iterations, the weight file is updated and metric values, i.e. average loss and mAP values are plotted using a TensorBoard. Once the loss graph stabilizes, training can be stopped by interrupting the runtime.

6 Result

Loss is a penalty for an incorrect prediction. It is calculated to enhance the algorithm, i.e. finding if the model is overfit, goodfit or underfit. Here in Fig. 7, the plot of loss drops to a stable plot indicating good fit. The exponentially declining graph of average loss while training SSD with MobileNetv2, FasterR-CNN with ResNet-50, and YOLOv4 tiny is shown in Fig. 7a–c, respectively. The amount of distortion in the graph is primarily based on the batch size used. Here, the batch sizes used are 6, 16 and 64 for SSD, FR-CNN and YOLO. Considering the capability of GPU, batch sizes are fixed accordingly. The model is trained until the graph stabilizes to a constant value. Further, the lastly updated weight file is selected for evaluation.

Generally, loss is used to compute the model's performance. Similarly to calibrate the model, mean average precision (mAP) is used to work out the model's prediction accuracy which can be computed by:

$$\text{Average Precision} = \sum_{j-1}^{n} \text{Precision}(j) \times \text{Recall change}(j) \tag{3}$$

Here, the mean average precision (mAP) at an intersection over union (IoU) threshold of 50 is examined. Intersection over union is a way to evaluate the area of the bounding box overlapping the ground truth box. The more the IoU, the more the possibility that the object is inside the predicted box. The results, after training the 4000 image dataset on different architectures, are reported in Table 2. The FasterR-CNN with ResNet50 model is trained for 90,000 steps which attains a mean average precision (mAP)@50IoU value of 74% as shown in Fig. 8a and the SSD with MobileNetv2 model is trained for 60,000 steps which attains a mean average precision (mAP)@50IoU value of 76% as shown in Fig. 8b. Similarly, YOLOv3 tiny and YOLOv4 tiny are trained using Darknet and mAP@50IoU of the models are 68% and 76.4%, respectively.

The mAP@50IoU values for different algorithms, FasterR-CNN, SSD, YOLOv3 tiny, and YOLOv4 tiny are represented in Table 2. FasterR-CNN with ResNet50 theoretically shows good accuracy, but the accuracy is reduced with 50 convolutional

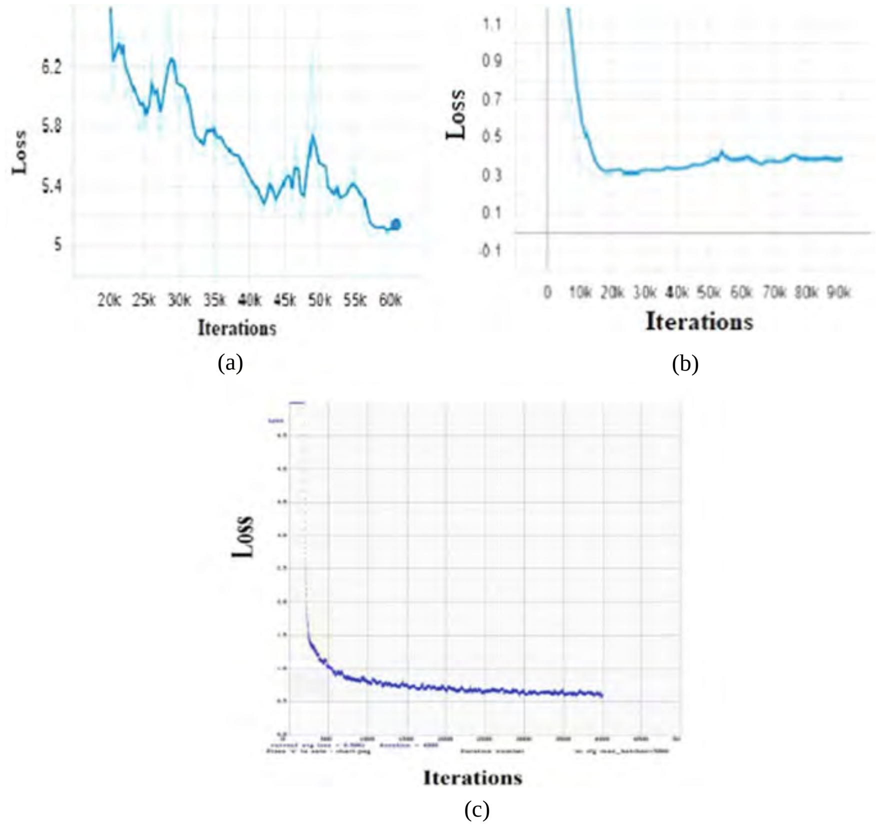
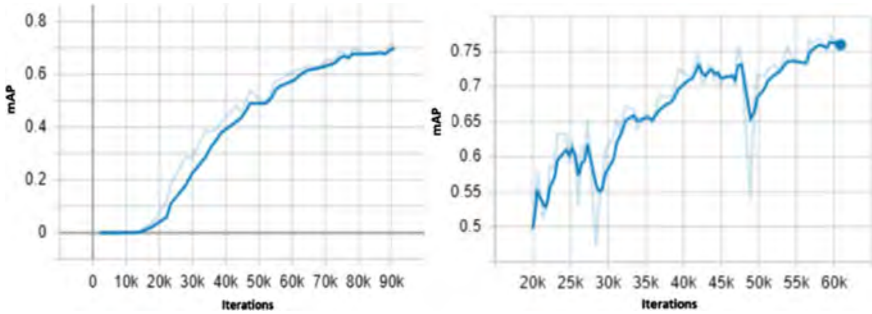


Fig. 7 Number of iterations versus average loss **a** SSD **b** FasterR-CNN **c** YOLOv4 tiny

Table 2 mAP comparison of different algorithms

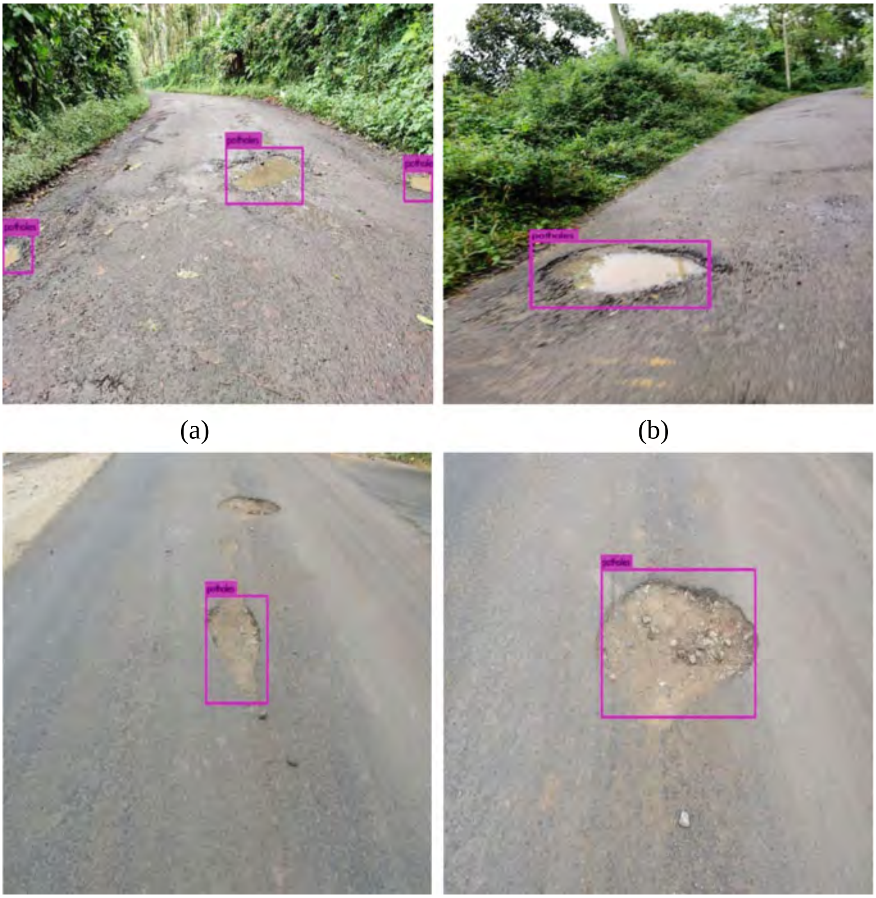
Model	mAP@50 (%)
FasterR-CNN + ResNet50	74
SSD + MobileNetv2	76
YOLOv3 tiny	68
YOLOv4 tiny	76.4

layers. YOLOv3 tiny is a fast object detector yet its accuracy is low compared to other detectors like SSD and FR-CNN. Nevertheless, SSD with MobileNetv2 provides a better speed-accuracy trade-off compared to other object detectors as they use different aspect ratios at a reasonable speed. Besides, YOLOv4 tiny is a fast detector with almost the same mAP value of SSD. Hence, Fig. 8 and Table 2 manifest that SSD and YOLOv4 tiny perform better compared to other object detection algorithms with 76% and 76.4% mAP, respectively. Figure 9a–d shows sample detection results.



(a) (b)

Fig. 8 Number of iterations versus mAP values **a** FasterR-CNN **b** SSD



(a) (b)

(c) (d)

Fig. 9 **a–d** sample detection results

7 Conclusion

Detection of potholes using object detection algorithms will be of great use as it would prevent road accidents. For this motive, we have developed a model to detect potholes using convolutional neural networks (CNN). In this work, four different algorithms are proposed and evaluated against each other. The CNN algorithms we used are FasterR-CNN, SSD, YOLOv3 tiny, and YOLOv4 tiny. In this work, the 1500 image dataset is transformed into a 4000 image dataset using data augmentation. The configuration used for FasterR-CNN is ResNet50 and for SSD is MobileNetv2. For YOLOv3 tiny and YOLOv4 tiny, the reduced configuration of Darknet-53 and CSPDarknet53, respectively, is the pre-trained weight for convolutional layers. Each technique has its own assets and liabilities with different pathways, where SSD with MobileNetv2 and YOLOv4 tiny transcend other architectures in terms of accuracy. Further, the proposed method can be used by integrating a live front view camera with ADAS to detect potholes and alert the driver. It can be implemented in real time by using a front view camera that is attached to the bonnet of a car. The camera starts recording when the car starts moving and divides the video into individual frames which can then be fed to the CNN model that will detect the presence of a pothole and warn the driver. Since SSD and YOLO have proven to be one among the fastest object detection algorithms, with an accuracy of 76.4% and 76% respectively, potholes can be easily detected considering the varying speed of the car. But the proposed approach confronts some limitations too. Potholes may go undetected due to few reasons such as water covered potholes, dark lightning conditions and high speed of the vehicle. Also, potholes can be incorrectly predicted considering the shadow types and the variety of shapes, a pothole can acquire. Thus, in order to overcome such limitations and to predict the presence of potholes more accurately, it is vital to use multiple cameras and to append more features favourable to the suggested model. This can also be integrated with an application to detect the positions of potholes and indicate it on the Google Maps for suggesting the paths with a comparatively lesser number of potholes.

References

1. T. Kim, S.-K. Ryu, Review and analysis of pothole detection methods. J. Emerg. Trends Comput. Inform. Sci. **5**(8), 603–608 (2014)
2. Y. Bhatia et al., Convolutional neural networks based potholes detection using thermal imaging. J. King Saud Univ.-Comput. Inform. Sci. (2019)
3. F. Jiménez et al., Advanced driver assistance system for road environments to improve safety and efficiency. Transp. Res. Proc. **14**, 2245–2254 (2016)
4. A. Akagic, E. Buza, S. Omanovic, Pothole detection: an efficient vision based method using RGB color space image segmentation, in *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (IEEE, New York, 2017)
5. A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks. Adv. Neural Inform. Process. Syst. (2012)

6. R.L. Galvez et al., Object detection using convolutional neural networks, in *TENCON 2018–2018 IEEE Region 10 Conference* (IEEE, New York, 2018)
7. S. Tu et al., Passion fruit detection and counting based on multiple scale FasterR-CNN using RGB-D images. Precis. Agric. pp. 1–20 (2020)
8. K.N.R. Chebrolu, P.N. Kumar, Deep learning based pedestrian detection at all light conditions, in *2019 International Conference on Communication and Signal Processing (ICCSP)* (IEEE, New York, 2019)
9. R. Gokul et al., A comparative study between state-of-the-art object detectors for traffic light detection, in *2020 International Conference on Emerging Trends in Information Technology and Engineering* (IEEE, New York, 2020)
10. J. Redmon et al., You only look once: unified, real-time object detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016)
11. H. Haritha, S.K. Thangavel, A modified deep learning architecture for vehicle detection in traffic monitoring system. Int. J. Comput. Appl. pp. 1–10 (2019)
12. J. Dharneeshkar et al., Deep learning based detection of potholes in Indian roads using YOLO, in *2020 International Conference on Inventive Computation Technologies (ICICT)* (IEEE, New York, 2020)
13. W. Liu et al., SSD: single shot multibox detector, in *European Conference on Computer Vision* (Springer, Cham, 2016)
14. S. Silvister et al., Deep learning approach to detect potholes in real-time using smartphone, in *2019 IEEE Pune Section International Conference (PuneCon)* (IEEE, New York, 2019)
15. C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning. J. Big Data **6**(1), 60 (2019)
16. Y. Lu, L. Zhang, W. Xie YOLO-compact: an efficient YOLO network for single category real-time object detection, in *2020 Chinese Control and Decision Conference (CCDC)* (IEEE, New York, 2020)
17. R. Zhang et al., An algorithm for obstacle detection based on YOLO and light filed camera, in *2018 12th International Conference on Sensing Technology (ICST)* (IEEE, New York, 2018)
18. P. Adarsh, P. Rathi, M. Kumar, YOLO v3-tiny: object detection and recognition using one stage improved model, in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)* (IEEE, New York, 2020)
19. T. Carneiro et al., Performance analysis of google colaboratory as a tool for accelerating deep learning applications. IEEE Access **6** (2018)
20. M. Abadi et al., Tensorflow: large-scale machine learning on heterogeneous distributed systems (2016). arXiv preprint arXiv:1603.04467

An Efficient Algorithm to Identify Best Detector and Descriptor Pair for Image Classification Using Bag of Visual Words

R. Karthika and Latha Parameswaran

Abstract Object identification and classification are important application in the computer vision and image processing domain. In this work, an attempt has been made to analyze various detectors and descriptors for extracting the best features and using bag of visual words (BoVW) for object classification. Experiments have been conducted on three significant different datasets COIL 100, CALTECH 101 and WANG. Features extracted from various detectors and descriptors have been used for classifying the images based on the objects. The best detector descriptor (FAST-SIFT) and optimum cluster size have been empirically determined to achieve high classification accuracy. The results obtained are used for performing further object retrieval.

Keywords Bag of visual words · Detectors · Descriptors

1 Introduction

Object classification is an important problem in computer vision as indicated by the amount of research on it. The goal is to develop an image classification system reducing the amount of manual supervision required as well as to reduce the computational cost in learning the parameters for classification; the need is to optimize between efficiency and performance. These characteristics are crucial for enabling classifiers to function in real-world applications. Research in vision-based object classification is one that has been growing for many years with different techniques ranging from fully manual to automated.

R. Karthika (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: r_karthika@cb.amrita.edu

L. Parameswaran
Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: p_latha@cb.amrita.edu

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_21

Automatic classification of images is challenging or extremely computational, especially in the presence of occlusion, background clutter, intra-class variation, different pose and intensity conditions. Global features are not able to address these challenges. Hence, an improvement of local-invariant features was necessary. The bag of visual words (BoVW) is widely used for keypoint-based representations. In BoVW, similar keypoints are grouped to form a cluster. The number of clusters denotes the size of the vocabulary, which can vary. Therefore, an image is represented as BoVW as a set of keypoints called visual words. Image classification has many applications including security (specifically face recognition), searching images, censoring images and robotics.

2 Literature Survey

Many researchers have contributed to feature detection and description on digital images. Some of the significant work has been presented here. The local features of images are identified by using boundaries, regions and points, out of which point-based feature extraction takes place in two steps—(1) keypoint detection and (2) feature descriptor generation. Harris et al. [1] identified keypoint based on two major directions specified by the Eigen values of the second-order matrix. Harris corner significantly helps in locating the corners by computing the first-order derivatives in the matrix. Lindeberg et al. [2] proposed that the image structures can be represented by a scale space representation at different levels of an image and developed a detector for blob-like features. They have built a scale space representation by smoothing high resolution images with Gaussian-based kernel at different sizes. Features from accelerated segment test (FAST) which increases the computational speed to an extent, which is required for the detection of corners, are proposed by Trajković et al. [3]. It is based on the properties exhibited by the corners, which characterizes the change in the intensity of the image. A multigrid approach is used to compute the change in arbitrary direction. The improved speed and suppression of false corner are due to the use of multigrid technique. All the features have a common property with the neurons, in the temporal cortex, which are used in the primate vision. Difference of Gaussian (DoG) is used because it accelerates the process of computation.

A different approach for the interest point's detection is proposed by Mikolajczyk et al. [4], and these interest points are not invariant to affine and scale transformations. The detector is based on the results extracted by Harris detector on interest points, the characterized scale by a local extremum for the local structures and the affine shape representation for a point in the neighborhood. Their scale detector enumerates a multi-scale representation for the interest points detected by the algorithm at which a local measure is maximum over scale. Lowe et al. [5, 6] used scale-invariant feature transform (SIFT) for extracting a unique invariant feature from the images which is used to match different views of the object. The features extracted are not invariant to scale and rotation of an image.

Ke et al. [7] improved SIFT by adding a principal component analysis (PCA). The local descriptors based on PCA are more distinctive and unique. The PCA-based SIFT makes the vector small, when compared to the vector used in SIFT. Mortensen et al. [8] presented feature descriptor that can supplement SIFT with a vector that can add curvilinear shape information from bigger region. This reduces the mismatches when two or more local descriptors are similar. It provides a better and a concrete method for the 2D non-rigid transformations, at a global scale which can be effectively matched as individuals. This establishes a vector which contains two parts: one is a SIFT descriptor and the other one is a global texture vector.

Color provides vital information in the matching task and object description. Abdel-Hakim et al. [9] enhance the SIFT descriptors so that it can be used in a color-invariant space. A descriptor called speeded up robust features (SURF) to improve the speed is proposed by Bay et al. [10]. Using the integral images and the existing descriptors reduces the complexity of the methods. Each extreme point is split into numerous 4×4 subregions, and then a Haar wavelet response is computed for all the subregions. Each and every keypoint is defined with a 64-dimensional feature vector.

Morel et al. [11] enhanced ASIFT (Affine SIFT) which simulates a set of sample views of an image obtained by varying two camera axis orientation parameters. This was done because SIFT could not perform to its capacity when it came to images containing affine changes. The ASIFT covers all the six orientation parameters, namely longitudinal and latitudinal angles. It also covers all the six parameters of the affine transform. Matas et al. [12] proposed maximally stable extremal regions (MSER) which is used as a method of blob detection in images. Correspondence is found between image elements from various images. The main contribution is the setup of new regions called extreme regions.

Leutenegger et al. [13] developed a BRISK detector which is scale as well as rotation invariant. The speed is improved by using the AGAST corner detector without compromising the detection performance. For scale invariance, BRISK performs a non-maxima suppression and interpolation by detecting keypoints. Rosten et al. [14] developed FAST-n detector has its basics related to wedge model type of a detector. The model is used to train a decision tree classifier and in turn applied to various images. The FAST-enhanced repeatability allows the option for repeatability for the detector. Patch-based image representation is divided into two types such as dense sampling and interest points. In dense sampling, patches of fixed shape and size are mounted on a regular grid.

For dense interest points, [10] cover the entire image and compute a spatial relation between features. A combination of both the local feature and the sampling schemes is proposed by Tamaki et al. [15]. DoG and sampling schemes are used to extract feature. Leonardis [16] developed kernel-based recognition method computes the geometric correspondence using pyramid matching scheme, which is introduced by Grauman et al. [17]. In this technique, the image is subdivided into various histograms and finally computing the histograms of all the local features at an increasingly fine resolution. Sivic et al. [18] matches are pre-computed based on descriptors, inverted file systems and documented rankings are used. The results that are obtained interpret that the retrieval is immediate returning a bunch of key frames ranked in the manner followed by google.

Senthilkumar et al. [19] used SURF to classify the logos in vehicles. A method for object recognition based on region of interest (ROI) and optimal bag of words was proposed by Li et al. [20]. The Gaussian mixture model is used to generate the visual codebook. Bosch et al. [21] used the concept of spatial pyramid to the development of an image signature known as pyramid histogram of oriented gradient (PHOG).The image representation is found to be effective in facial emotion recognition and facial component-based bag of words. Karthika.et al. [22] developed a face recognition system invariant to pose and orientation using Gabor wavelet features. Visual vocabularies of varying sizes with SIFT and SURF are evaluated by Schaeffer et al. [23]. The vocabularies constructed using K-means clustering and self-organizing map are compared in terms of accuracy for their own dataset. Performance of two keypoint descriptors in the context of pedestrian detector is discussed by Kottman et al. [24]. FREAK achieved a maximum accuracy using bag of words model.

Based on the literature survey, it is observed that many researchers concentrated to develop detectors and descriptors to extract features from images. This proposed work aims to develop a technique to find the best detector and descriptor combination for image classification.

3 Proposed Algorithm

The proposed algorithm to find the best detector–descriptor combination is explained here.

The notations used are as follows:

$$\text{Set of key points } K = \left\{K_1,\ K_{2,} K_3 \ldots K_n\right\}$$

$$\begin{aligned}\text{Feature Vector} Fv = &\{F_{11}, F_{12}, F_{13}, \ldots F_{1m}\}\\ &\{F_{21}, F_{22}, F_{23}, \ldots F_{2m}\}\\ &\ldots\\ &\{F_{n1}, F_{n2}, F_{n3}, \ldots F_{nm}\end{aligned}$$

Algorithm for image classification using bag of visual words:

For every detector i ∈ {BRISK, DENSE, MSER, FAST, GRID, SIFT, SURF}.

For every descriptor j ∈ {SIFT, SURF}.

1. Extract the keypoints (K) of all the images in (I) using the detector i
2. Obtain the descriptor feature vector F_v for each key point in K using the descriptor j
3. For each cluster size C = 10, 20, 40, … maximum cluster size

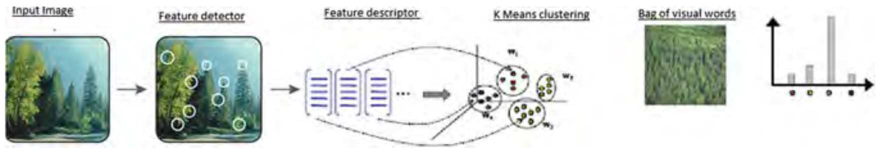
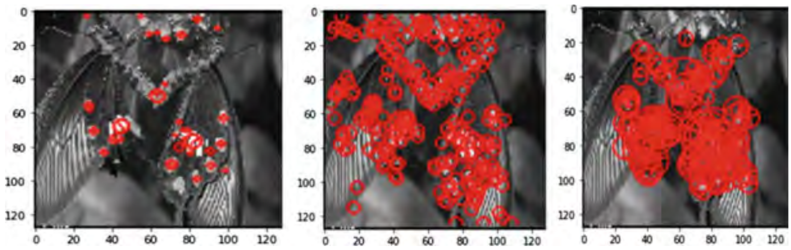


Fig. 1 Block diagram representation of BoVW algorithm



(a) (b) (c)

Fig. 2 **a** Keypoint detection using SIFT detector. **b** Keypoint detection using FAST detector. **c** Keypoint detection using BRISK detector

a. Input feature descriptor ($F_{v)}$ are clustered for the set of images (I) using the K means clustering to construct the vocabulary of visual words

$$\{V = V1, V2, V3 \ldots Vw\}$$

b. Perform a histogram plot for the occurrence of vocabulary of Words (V) using the feature descriptor (F_v)

4. Apply SVM classifier using RBF kernel.
5. Analyze results to identify best detector–descriptor pair for image classification.

Figure 1 shows different stages in the BoVW algorithm. Figure 2a–c shows the keypoint detection using different detectors.

3.1 Contribution in This Work Is Given Below

Detector and Descriptor: There have been several benchmark studies on the detectors and descriptors. In this work, we have extensively studied and extracted the image features using all possible detector and descriptor combinations for classifying the images which has shown the efficacy in classification accuracy. An efficient way to select the right combination of detectors and descriptor is addressed.

Nature of images: Most authors have tested the algorithms on set of images of a particular type or nature; whereas, in this work, we have experimented on three different datasets with varying background and large inter-class variations of object.

Cluster size: In the bag of visual words, construction identifying the best cluster size for different detectors and descriptor is addressed.

3.2 Evaluation Metrics

A confusion matrix is constructed with four essential parameters: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). These four parameters allow more detailed analysis of the classification system. We have selected two parameters such as true positive rate and accuracy of each detector and descriptor combination, for analysis and decision making.

True positive rate (TPR) is the proportion of positive cases that were correctly identified

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{1}$$

Accuracy (ACC) is the proportion of the total number of predictions that were correct.

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{2}$$

4 Experimental Results

For this experimentation, three datasets, namely COIL-100 [25], CALTECH-101 [26] and WANG [27] have been used. Testing has been done using the identified detectors and descriptors by varying the cluster size of bag of visual words.

The first dataset used is the Columbia Object Image Library COIL-100 dataset [25]. It contains a black background that has 100 objects which are placed on a motorized turntable. This table is rotated from the angles 0° to 360° at the intervals of 5°, thus taking 72 images per object. Hence, this database contains 7200 images of the 100 objects. For our experiment, we have used 57 samples per class for training and 15 samples per class for testing. All the 100 classes were tested and trained.

The second dataset used in this experimentation is CALTECH-101 [26], which contains 101 objects in a different background. Each object consists of 31–800 images. In this experiment, ten classes have been chosen randomly from CALTECH-101 for object classification. In that, 70 images were used for training and 30 images were used for testing.

The third dataset used in the experiment is the Wang dataset [27] which is a subset of 1000 images taken from the Corel stock photo database. These images have been manually selected to form ten classes of 100 images each. From this, 80 samples were used for training and 20 samples were chosen for testing.

The training has been done using support vector machine (SVM) with the RBF kernel.

4.1 Inference Based on Accuracy of Classification

The following observations have be made based on the dataset from Table 1.

The most accurate classifications for the WANG dataset, which contains 1000 images in ten classes, occur when the FAST detector is used with SIFT/SURF as the descriptor. On the other hand, the GRID detector produces the most inaccurate results of the chosen detectors. Its most accurate results are achieved with a minimum cluster size of 160. The BRISK, MSER and SIFT detectors produce slightly better results, while the accuracy of the DENSE detector is almost on par with the FAST detector for the WANG dataset.

The COIL-100 dataset, which contains 7200 images of 100 classes, is classified with higher accuracy by all of the given detector and descriptor combinations and by varying cluster sizes. The FAST detector produces the most accurate results with both SIFT and SURF detectors, with the DENSE detector once again coming in a close second. The accuracy of the detectors peaks with a cluster size of 320, showing no noticeable increase with corresponding increase in cluster size.

The CALTECH-101 dataset, containing 1000 images in ten classes, is classified best by the FAST detector with SIFT/SURF as the descriptors. The BRISK and DENSE detectors used in combination with the SIFT/SURF descriptor produce poor results as does the MSER detector when used with the SURF descriptor. The SIFT/SURF detector produces its most accurate results with a minimum cluster size of 320/160, respectively.

The following observations have be made based on the detectors from Table 1.

The accuracy is higher for the SIFT/SURF detector seen in the case of the COIL-100 dataset, with both SIFT and SURF descriptors and with a minimum cluster size of 160. For the WANG dataset, the SIFT/SURF detector produces higher accuracy in combination with the SIFT descriptor and with a maximum cluster size of 320.

The GRID detector produces better accuracy for the COIL dataset with a minimum cluster size of 160. While classifying the CALTECH-101 dataset, the GRID detector performed better in combination with the SIFT descriptor and with a maximum cluster size of 640. As with the other detectors, the DENSE detector produced better accuracy for the COIL dataset, with a minimum cluster size of 160. Classification by the DENSE detector of the WANG dataset is appreciable when used in combination

Table 1 Accuracy of classification on the three datasets

Dataset	Descriptor	BRISK		DENSE		FAST		GRID		MSER		SIFT/SURF*	
		Value	Cluster size	Value	Cluster size	Value	Cluster size	Value	Cluster size	Value	Cluster size	Value	Cluster size
WANG	SIFT	0.90	1280	0.94	640	0.94	1280	0.89	160	0.92	640	0.91	320
	SURF	0.88	640	0.92	160	0.93	320	0.86	160	0.88	160	0.90	320
COIL	SIFT	0.99	320	0.99	160	0.99	320	0.99	160	0.99	160	0.99	160
	SURF	0.99	320	0.99	160	0.99	320	0.99	160	0.99	160	0.99	160
CALTECH	SIFT	0.89	640	0.89	320	0.92	1280	0.91	640	0.91	640	0.91	320
	SURF	0.88	160	0.88	320	0.91	640	0.89	160	0.88	1280	0.91	160

*Denotes SIFT as a detector and descriptor and SURF as a detector and descriptor with the SIFT descriptor and with a maximum cluster size of 640. Despite being less accurate than the other detectors, the BRISK detector produces better results when classifying the COIL dataset with the SIFT descriptor. While classifying the CALTECH-101 and WANG datasets, the BRISK detector produces higher results in combination with the SIFT descriptor and with cluster sizes of 640 and 1280, respectively.

The FAST detector consistently shows better accuracy value across the three datasets when compared to the other detectors, classifying the COIL 100 dataset, and the highest accuracy is obtained with a minimum cluster size of 320. When classifying the WANG and CALTECH 101 datasets, the FAST detector produces its better results in both cases with the SIFT descriptor and with maximum cluster sizes of 1280. The MSER detector produces results of middling accuracy compared to the other detectors, with its own best performance being classifying the COIL dataset with a minimum cluster size of 160. It shows better results with the CALTECH 101 dataset when used with the SIFT descriptor and with a maximum cluster size of 1280.

4.2 Inference Based on TPR of Classification

The following observations have been made based on dataset from Table 2.

As far as the classification of the WANG dataset is concerned, FAST detector gives the highest TPR, with the DENSE detector showing the second highest. Using SURF as the descriptor, DENSE detector shows high TPR in cluster of size as small as 160. GRID detector performs poorly when classifying the WANG dataset regardless of the descriptor used.

The COIL dataset is classified with high TPR by all the detectors with the FAST and GRID detectors performing best and worst, respectively. SIFT descriptor produces a low TPR for this dataset when paired with the GRID detector. The SIFT/SURF detector–descriptor combination performs well with this dataset, achieving the second highest TPR after the FAST detector.

The FAST detector has the highest TPR of all the detectors when classifying the CALTECH dataset. The SIFT descriptor shows low TPR when used with the BRISK detector. The SURF descriptor performs poorly in combination with the MSER detector. When SURF is used as both detector and descriptor, it has a TPR almost as high as that of the FAST detector and achieves this at cluster size as low as 160.

The following observations are made based on detectors:

The BRISK detector shows excellent TPR in classifying the COIL dataset with SIFT as the descriptor, with cluster size of 1280. For both CALTECH and WANG, the BRISK detector achieved good TPR for each at cluster size as low as 160 when the SURF descriptor is also used.

The DENSE detector performs well in general producing its highest TPR with the COIL dataset and the SIFT descriptor. It shows high TPRs at cluster sizes as

Table 2 True positive rate for classification on the three datasets

Dataset	Descriptor	BRISK		DENSE		FAST		GRID		MSER		SIFT/SURF*	
		Value	Cluster Size	Value	Cluster Size	Value	Cluster Size	Value	Cluster Size	Value	Cluster Size	Value	Cluster Size
WANG	SIFT	0.52	1280	0.66	640	0.70	1280	0.47	160	0.6	640	0.55	320
	SURF	0.41	160	0.62	160	0.65	320	0.27	160	0.55	640	0.52	320
COIL	SIFT	0.53	1280	0.77	160	0.84	320	0.70	320	0.72	640	0.82	320
	SURF	0.40	320	0.73	160	0.84	320	0.25	320	0.67	320	0.82	640
CALTECH	SIFT	0.46	640	0.47	320	0.62	1280	0.52	640	0.53	640	0.53	320
	SURF	0.42	160	0.43	320	0.55	640	0.45	160	0.40	1280	0.53	160

low as 160 for both the COIL and CALTECH datasets. For the Wang dataset, the DENSE detector has a high TPR for cluster sizes up to 640 when used alongside the SIFT descriptor. The FAST detector had the highest TPR of all the detectors, performing its best with the COIL dataset. It hits its highest TPR for this dataset at cluster sizes starting from 320. When used along the SIFT descriptor for the WANG and CALTECH datasets, the FAST detector had high TPR with cluster sizes of up to 1280.

The GRID detector had its best and poor performances with the COIL dataset, when used alongside the SIFT and SURF descriptors, respectively. The GRID detector produces low results for the WANG dataset with the combination of SUFF descriptor. The GRID detector works better with the SIFT descriptor, showing a high TPR at a cluster size as low as 160 with the WANG dataset and as high as 640 with the CALTECH dataset.

When classifying the COIL dataset, the MSER detector produced the highest TPR when used with the SIFT descriptor, and with the SURF descriptor, it reached its highest TPR at cluster size starting from 320. With the CALTECH dataset, the detector showed high TPR with cluster size up to 1280 when used with the SURF descriptor.

When classifying the COIL dataset, the SIFT/SURF detector shows the same TPR with both SIFT and SURF descriptor. It shows its highest TPR for the WANG dataset at a cluster size of 320 irrespective of descriptor. For the CALTECH dataset, SURF detector achieved its highest TPR with the SURF descriptor and with cluster size as low as 160.

5 Discussion

In all the three datasets, SIFT descriptor performance is better than SURF descriptor. In SIFT descriptor, the features are invariant to image rotation, scaling and partially invariant to different illumination. The unique interest points are identified by using difference of Gaussian function which lead to invariance in scale and orientation. Keypoints are selected based on their stability. The cascade filtering approach reduces the cost of extracting the features.

COIL dataset performed well with all the evaluation metric parameters because it does not contain the background and all the detectors are able to classify the objects with different rotation.

In all the three datasets, FAST detector outperformed all the detectors. We can infer that the FAST detector performs superiorly when compared to other detectors. If the intensities of 16 contiguous pixels are darker or brighter beyond a threshold, then the region will be defined as a corner and these points lie on distinctive high-contrast regions. Corners are very important feature because of their ability to show two-dimensional change, and FAST detector prioritizes the detection of corners over edges by providing the advantage of a good computational efficiency.

Hence, it may be inferred that FAST could be the best detector and SIFT as the best descriptor to extract features from images. Even though the cluster size is varying, there is no much significance in the accuracy achieved. These experiments have been conducted by varying the cluster size from 10 to 2560 in steps of product of 2 (10, 20, 40, 80, 160, 320, 640, 1280, 2560). It is observed that when the cluster size is 1280, the evaluation metric parameters converge. These parameters are achieved with larger or equal cluster sizes using SIFT as a descriptor than the SURF descriptor.

Yang et al. [28] using compression network that encodes SIFT-based object histogram achieved an accuracy of 95% in COIL dataset [29]. COIL 100 with eight views exhibit an accuracy of 92.5% using nearest prime simplical complex approaches. An accuracy of about 99% has been achieved by using FAST detector and SIFT/SURF descriptor combination.

6 Conclusion

Expressing images using features is significant to many applications including object recognition and retrieval. However, many researchers have developed feature detectors and feature descriptors for various purposes, and there is not much of work to find best detector and descriptor to classify images of various types. This research work has attempted to identify the best feature descriptor–detector to represent the given image so that it is useful for image classification. We have achieved the accuracy of 99% in COIL dataset. As an extension of this work, it can be used for recognizing objects in a given image and retrieving images based on similarity. These feature descriptors and detectors can be designed as a filter in convolutional neural network to extract features as required, and hence, it will serve as the basis for deep learning.

References

1. C. Harris, M. Stephens, A combined corner and edge detector, in *Alvey Vision Conference*, vol. 15, No. 50 (1988), pp. 10–5244
2. T. Lindeberg, Feature detection with automatic scale selection. Int. J. Comput. Vision **30**(2), 79–116 (1998)
3. M. Trajković, M. Hedley, Fast corner detection. Image Vis. Comput. **16**(2), 75–87 (1998)
4. K. Mikolajczyk, C. Schmid, Scale & affine invariant interest point detectors. Int. J. Comput. Vision **60**(1), 63–86 (2004)
5. D.G. Lowe, Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision **60**(2), 91–110 (2004)
6. D.G. Lowe, Object recognition from local scale-invariant features, in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2 (IEEE, 1999), pp. 1150–1157
7. Y. Ke, R. Sukthankar, PCA-SIFT: a more distinctive representation for local image descriptors, in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2 (IEEE, 2004), pp. II–II

8. E.N. Mortensen, H. Deng, L. Shapiro, A SIFT descriptor with global context, in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1 (IEEE, 2005), pp. 184–190
9. A.E. Abdel-Hakim, A.A. Farag, CSIFT: a SIFT descriptor with color invariant characteristics, in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2 (IEEE, 2006), pp. 1978–1983
10. H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, Speeded-up robust features (SURF). Comput. Vis. Image Underst. **110**(3), 346–359 (2008)
11. J.M. Morel, G. Yu, ASIFT: a new framework for fully affine invariant image comparison. SIAM J. Imaging Sci. **2**(2), 438–469 (2009)
12. J. Matas, O. Chum, M. Urban, T. Pajdla, Robust wide baseline stereo from maximally stable extremal regions. Image Vis. Comput. **22**(10), 761–767 (2004)
13. S. Leutenegger, M. Chli, R.Y. Siegwart, BRISK: binary robust invariant scalable keypoints, in *Computer Vision (ICCV), 2011 IEEE International Conference on* (IEEE, 2011), pp. 2548–2555
14. E. Rosten, R. Porter, T. Drummond, Faster and better: a machine learning approach to corner detection. IEEE Trans. Pattern Anal. Mach. Intell. **32**(1), 105–119 (2010)
15. T. Tamaki, J. Yoshimuta, M. Kawakami, B. Raytchev, K. Kaneda, S. Yoshida, Y. Takemura, K. Onji, R. Miyaki, S. Tanaka, Computer-aided colorectal tumor classification in NBI endoscopy using local features. Med. Image Anal. **17**(1), 78–100 (2013)
16. A. Leonardis, H. Bischof, A. Pinz, *Computer Vision–ECCV 2006 9th European Conference on Computer Vision, Graz, Austria, May 7–13, 2006, Proceedings, Part IV. In Conference proceedings ECCV* (2006), p. 333
17. K. Grauman, T. Darrell, The pyramid match kernel: discriminative classification with sets of image features, in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 2 (IEEE, 2005), pp. 1458–1465
18. J. Sivic, A. Zisserman, Video Google: a text retrieval approach to object matching in videos. In: *Null* (IEEE, 2003), p. 1470
19. T. Senthilkumar, S.N. Sivanandam, Logo classification of vehicles using SURF based on low detailed feature recognition. Int. J. Comput. Appl. **3**, 5–7 (2013)
20. W. Li, P. Dong, B. Xiao, L. Zhou, Object recognition based on the region of interest and optimal bag of words model. Neurocomputing **172**, 271–280 (2016)
21. A. Bosch, A. Zisserman, X. Munoz, Representing shape with a spatial pyramid kernel, in *Proceedings of the 6th ACM International Conference on Image and Video Retrieval* (ACM, 2007), pp. 401–408
22. R. Karthika, L. Parameswaran, Study of Gabor wavelet for face recognition invariant to pose and orientation, in *Proceedings of the International Conference on Soft Computing Systems* (Springer, New Delhi, 2016), pp. 501–509
23. C. Schaeffer, *A Comparison of Keypoint Descriptors in the Context of Pedestrian Detection: FREAK Vs* (SURF vs. BRISK, Cité En p, 2013), p. 12
24. M. Kottman, Performance evaluation of visual vocabularies for object recognition. Inform. Sci. Technol. Bulletin ACM Slovakia**4**(2) (2012)
25. www.cs.columbia.edu/CAVE/software/softlib/coil-100.php
26. www.vision.caltech.edu/Image_Datasets/Caltech101/
27. https://wang.ist.psu.edu/docs/related/
28. A.Y. Yang, et al., Multiple-view object recognition in smart camera networks, in*Distributed Video Sensor Networks* (Springer, London, 2011), pp. 55–68
29. J. Zhang, Z. Xie, S.Z. Li, Nearest prime simplicial complex for object recognition (2011)arXiv: 1106.0987

GUI-Based Alzheimer's Disease Screening System Using Deep Convolutional Neural Network

Himanshu Pant, Manoj Chandra Lohani, Janmejay Pant, and Prachi Petshali

Abstract Brain is the prime and complex organ of the nervous system in all osseous and boneless organisms. Generally, it is positioned close to the sensory organs in the head. Brain is made up of trillions of networks called synapses that are used to connect and communicate with more than 100 billion cells in the human body. When the brain cells are required to be degenerated, and about to die, then Alzheimer's disease might come into the existence. Alzheimer's disease (AD) is a progressive disorder that will destroy the brain nerves. This order will profoundly lead to a continuous memory loss, which will adversely affect the mental health and daily routines. Also, Alzheimer's disease remains as one of the main causes for dementia. So, early and accurate detection and diagnosis of Alzheimer's disease remain as the most significant research area. Normally, doctors will identify and observe the AD from the visualization of the brain magnetic resonance imaging (BMRI) with naked eye. To resolve these challenges, authors have proposed and designed a graphical user interface-based AD screening system that select, detect, and predict Alzheimer's disease classes by using deep convolutional neural networks (DCNN). For this purpose, brain magnetic resonance imaging (BMRI) has been performed. These MRI images are categorized into four classes: mild demented, moderate demented, very mild demented, and non-demented Alzheimer's disease. In this paper, deep convolutional neural networks (DCNN) offer a saner solution in numerous disease controls for neurological disorder with a high precision on validation accuracy. GUI-based screening system will identify the healthy or non-demented class accurately so that anyone can easily diagnose the accurate type of Alzheimer's disease.

Keywords Alzheimer's disease · Neurological disorder · MRI · Convolutional neural network · Disease screening system

H. Pant (✉) · M. C. Lohani · J. Pant · P. Petshali
Graphic Era Hill University Bhimtal, Nainital, Uttarakhand, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_22

1 Introduction

Brain tissues damage is the main cause of Alzheimer's disease. It has a certain progressive pattern that shrinks the size of hippocampus and medullary wrapper of the brain and also expands the ventricles [1]. When the hippocampus size is reduced then spatial memory and episodic memories which are the parts of the brain is damaged. These two memories link the connections between the brain and rest of the body. Anymore neurons cannot communicate to the other cells via synapses. It is appraised that in the developing country the dispersion of this Alzheimer's disease to be around 5% after 65 years and approximate 30% after 85 years old. It is expected that around 65 crore peoples will be detected with AD by 2050 [2]. Alzheimer's is the most common cause of dementia. It is a general term for neurological brain disorder and cognitive and short-term memory loss. Due to this uncertainty, it imposes a wicked impact on the performance of routine life tasks such as writing, speaking, and reading.

Till now there is no cure of AD but patients can reduce the disease symptoms by early stage diagnosis. Brain magnetic resonance imaging (BMRI) is the one of the most appropriate tools to diagnose the AD [3]. It is a common practice for Alzheimer's disease diagnosis in clinical research. When the human brain disorder starts, then all the indicators of this disease slowly grow affectively with the passage of time. Alzheimer's disease (AD) detection and classification is a challenging task because probably brain MRI report of Alzheimer's disease person may be found normal. Prompt detection of Alzheimer's disease can help to diagnose properly and hence might be prevent brain tissue damage.

Early detection of Alzheimer's disease is taking out from the bunch of brain MRI report data along with standard healthy MRI report of older people. Magnetic resonance imaging (MRI) provides the feasibilities to study neurotic brain tissues changes that are associated with AD in vivo [4].

There are four major stages in Alzheimer's disease—non-demented, mild demented, very mild demented, and moderate demented Alzheimer's disease. Alzheimer's disease (AD) is not accurately detected until a patient reaches in moderate demented stage. But proper treatment and protect brain tissue damage in early stage can prevent the critical condition of this disease.

Earlier Alzheimer's disease is carried out by the clinical experts in elegant ways by means of powerful magnetic resonance imaging techniques which are impossible to detect by human's naked eyes. It is necessary to identify and classify the infected MRI images from the group of demented and non-demented Alzheimer's MRI dataset in early stage. But due to the lack of early detection of the disease diagnosis and screening system, persons tolerate a huge expansion loss in every year [5]. Therefore for the common person, an accurate, very fast, an automated, precise, and proficient disease screening system is required.

Numerous valuable research works have been done to diagnose and classify different stages of Alzheimer's disease [6]. Recently in image classification and computer vision, the deep learning-based convolutional neural networks (CNNs) algorithms have been applied widely [7].

Joint hippocampal segmentation and pathological score regression using brain MRI scans model were proposed on multi-task deep learning (MDL) method [8]. To extract the features from three-dimensional clinical brain MRI images for the classification, deep 3D CNNs were used [9].

Brain MRI dataset requires enormous computational resources and a huge amount of labeled data to train a deep convolutional neural network. Usually, the size of available brain MRI datasets is typically very small compared with the required datasets used in computer vision techniques for disease diagnosis. Due to this a lesser amount of the dataset, it becomes a major challenge to learn the large number of trainable parameters in deep convolutional neural network [10].

In this research, researchers had considered the four types of demented Alzheimer's disease along with the non-demented disease and proposed a graphical user interface screening system based on deep convolutional neural network for Alzheimer's disease (AD) identification. This paper also proposed a disease detection and estimation model that moderated the imprecise and erroneous manual disease diagnose. Proposed screening model improves the training and validation accuracy, viability and proficiency of the algorithm. This paper recommended a graphical user interface-based screening system which detect the various types of Alzheimer's disease from brain MRI and classify in four different Alzheimer's demented classes. Medical experts professionally analyze the brain MRI report data and then distinguish the category of Alzheimer's disease present in the MRI report.

This paper inspirations by using numerous computer vision techniques to detect and classify four different set of demented Alzheimer's disease from the brain MRI images. Three block baseline convolutional neural networks (BCNN) are applied for the same.

The primary objective of this consideration was the collection of a sufficient brain MRI (BMRI) images data from various existing open source datasets. This collected MRI dataset is further divided in four different categories and labels. The work flow describes the proposed GUI-based system Alzheimer's disease screening system architecture as shown in Fig. 1.

This research paper follows the given structure: In Sect. 1, authors described the theoretical foundations and related works for the proposed Alzheimer's disease screening model. Collected dataset descriptions and data preprocessing are provided in material and methods category in Sect. 2. Section 3 delivers the result oriented-based experimental performance on AD image dataset using deep convolutional neural network. In Sect. 4, authors proposed and create a graphical user interface-based disease screening model to predict the correct category of AD using feature extracting technique with their hidden layers in CNN model. Conclusion of proposed screening model and future scope for this project and research is scheduled in Sect. 5.

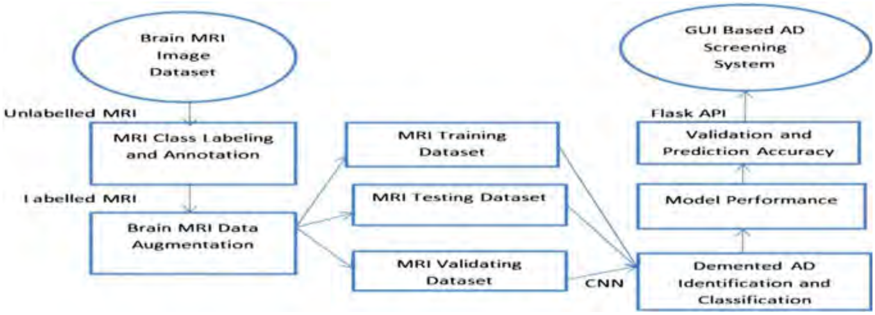


Fig. 1 Alzheimer's disease detection system design

2 Methodology

2.1 Research Design

In this proposed methodology, the neural network model depends on three prime steps. The first step is to collect the brain MRI data, preprocessing, and augmentation on that data. The second phase is to extract the features from input MRI images, and the third step is to classify and detect one of the four demented AD class. The various steps performed by the disease detection model describe in the given work flow architecture as shown in Fig. 2.

This paper established a baseline convolutional neural network approach based on VGG16 architecture for AD classification and prediction. The experimental dataset is based on the brain magnetic resonance imaging.

2.2 Dataset Descriptions

Precise data collection for the research purpose plays a stimulating and significant role in deep convolutional neural network. For that purpose, Alzheimer's disease dataset was acquired from the one of the authorized open source database, which is publicly available for the disease detection competition on the website [11]. These datasets were categorized in four different types of demented AD classes. These categorized dataset have to build in four labeled form. These four labels are the four types of demented AD class which were used to classify and predict Alzheimer's disease. The finishing dataset consists of four classes of AD having 6423 brain MRI images. The MRI images dataset distributes in training and testing phases, respectively, describe in Fig. 3. For the disease detection task, brain MRI image dataset is randomly split into 80% for training and 20% for testing phase. The proposed work used four sets of demented brain MRI images, namely mild demented AD, moderate demented AD, non-demented AD, and very mild demented AD.

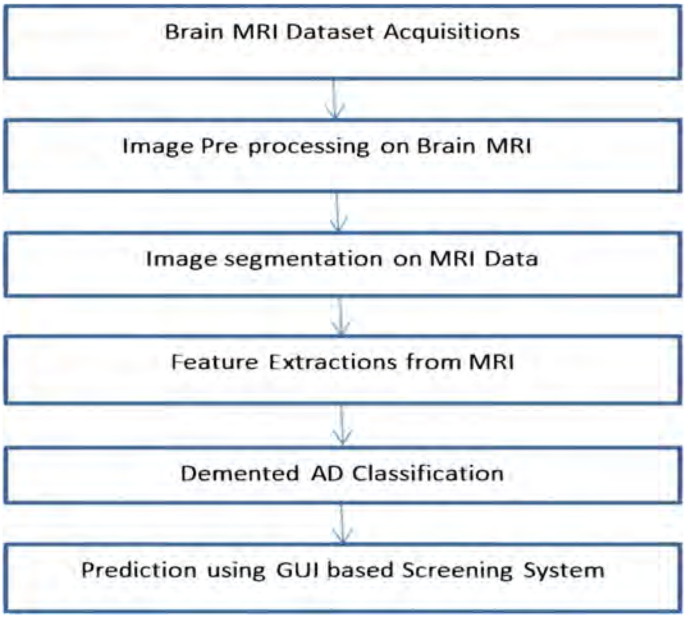
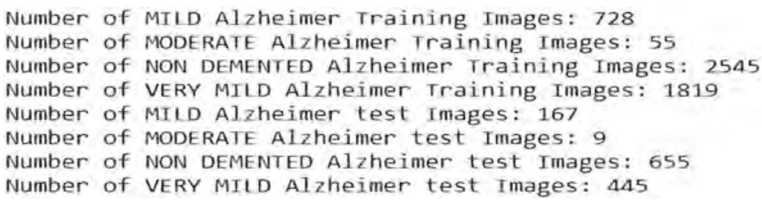


Fig. 2 Work flow architecture for AD detection



```
Number of MILD Alzheimer Training Images: 728
Number of MODERATE Alzheimer Training Images: 55
Number of NON DEMENTED Alzheimer Training Images: 2545
Number of VERY MILD Alzheimer Training Images: 1819
Number of MILD Alzheimer test Images: 167
Number of MODERATE Alzheimer test Images: 9
Number of NON DEMENTED Alzheimer test Images: 655
Number of VERY MILD Alzheimer test Images: 445
```

Fig. 3 Distribution of different AD MRI images in classes

2.3 Data Preprocessing

The collected brain MRI reports are initially in the form of un-annotated and unlabeled. Four multiple directories for four types of AD are associated with it. Authors performed the labeling and annotation process of all the images of four categories by their directory name with a word "MILD," "MOD," "NonD," and "VMD," respectively.

Initially, the four types of collected MRI are in the form of RGB color channels. All the collected MRI images have diverse measurements, sizes, and figures. It is difficult to extract the features from group of different sizes of images. So it is required to reshape all the brain MRI images in the matching size, form, and dimensions in image processing. Annotation is applied for all the brain images for the classification

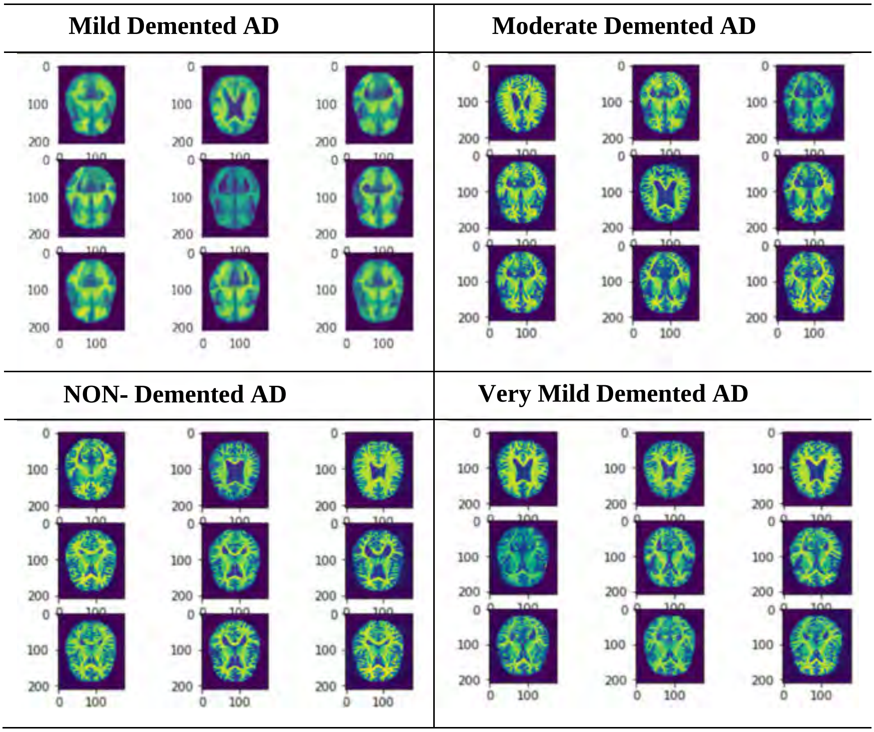


Fig. 4 Sample images for MRI demented AD

and detection. Python Keras API and tensor flow interface are used to accomplish the standard same and equal size of images in the measurement of 200 × 200 pixels photographs [12]. Image edge enhancement and noise reduction can be done by using image filtering techniques of image processing. It improves the accuracy of the model after preprocessing on the training images. Digital image filter for the noise reduction can be applied by taking the convolution of the images with the small kernal of the size n × n square matrix (where n is any odd number). Some standard fix pixel-sized sample images in all the four categories are described in Fig. 4.

Image data generator class and flow_from_directory API tools of keras are used to separate all four brain MRI class into respective directories. This API class generates a discrete train and test folders within same directory in which all four subdirectory of each class is formed [13].

The experimental performances of brain MRI for Alzheimer's disease dataset were assessed by applying three block VGG-16 model-based baseline convolution neural network model. This VGG-16-oriented model classifies and identifies four types of demented AD from the bunch of brain MRI reports data. Authors proposed a deep learning-based image classification technique that simulate brain MRI data and recognize disease spots from MRI.

In this disease detection screening system, experimental setup was accumulates by using 3 × 3-sized convolutional layers filters. These filters are used to learn the image features and different portions from an image. This filter matrix will slide from left to right and top to bottom. After this convolutional filter, a max pooling layer is applied. This max pooling layer is used to perform an operation that calculates maximum values of each feature map with respect to each patch. Single convolutional filters and single maximum pooling layer together form a single convolutional block. When we increase the number of filters in the neural network for AD, the number and size of each convolutional block are increased. These convolutional blocks can recall in the entire neural network.

The CNN kernel is used to process addition of pixels in an input image with the help of padding technique. Authors are motivated from the VGG-16 architecture and then applied the classification technique on brain MRI dataset for the four types of demented Alzheimer's disease (AD). This disease screening system accomplishes a classification task and forecast the types of AD based on brain MRI image data. After performing linear computation, every layer is triggered with a nonlinear activation function. In this proposed baseline convolutional neural network, model is activated by rectified linear unit (ReLU) function. The uniform weights from the range (−limit to +limit) are initialized by using "he_uniform" weight initializer to set the initial random weights in the proposed Keras model layers.

To detect Alzheimer's disease (AD) using baseline CNN, the total available parameters in trainable and non-trainable set were described in Fig. 5. This sequential model has total 10,334,021 parameters in this set. Stochastic gradient descent optimizer is an iterative algorithm that is used in activation function to jumps from a stochastic (random) point on a nonlinear activation function and moves its slope in down phases until the points reaches the lowermost point of that activation function [14]. Author used a standard 0.001 Conservative learning rate and 0.9 momentums to train this sequential model.

Brain MRI images are trained by using train iterator method, and dataset can be validate by applying test iterator method of keras to fit the disease screening system for classification and prediction. 81 steps are used in every 200 epoch for the training and testing the deep convolutional sequential neural model. Each training steps can

Model: "sequential_2"

Layer (type)	Output Shape	Param #
conv2d_2 (Conv2D)	(None, 200, 200, 32)	896
max_pooling2d_1 (MaxPooling2	(None, 100, 100, 32)	0
conv2d_3 (Conv2D)	(None, 100, 100, 64)	18496
max_pooling2d_2 (MaxPooling2	(None, 50, 50, 64)	0
conv2d_4 (Conv2D)	(None, 50, 50, 128)	73856
max_pooling2d_3 (MaxPooling2	(None, 25, 25, 128)	0
flatten_1 (Flatten)	(None, 80000)	0
dense_1 (Dense)	(None, 128)	10240128
dense_2 (Dense)	(None, 4)	516

Total params: 10,333,892
Trainable params: 10,333,892
Non-trainable params: 0

Fig. 5 Sequential model parameters and neural network layers be calculated with respect to one epoch used in model by dividing total number of training and total number of testing brain MRI AD images in all four categories by the described batch size [15].

3 Results

VGG-16-based baseline sequential Alzheimer's disease detection screening system can be evaluated by performing unseen testing and validating data over training MRI dataset. This performance is evaluated by measuring validation accuracy with respect to training accuracy. This validation accuracy is calculated by applying three blocks VGG-16 architecture. Three-block VGG-16 architecture is the extended version of one- and two-block architecture. This block architecture is achieved by increasing convolutional layers and pooling layers from one-block and two-block VGG models [16].

3.1 Research Findings

In this proposed architecture, the size of dataset is comparatively small to train and test the model. So it is required to augment the dataset from the existing dataset without changing the MRI image themes. This modification in the images is done by image data augmentation technique of image processing. This technique is performed by applying numerous operations on existing brain MRI dataset like cropping the images, image padding, image transformation techniques (rotation, scaling, and translation), and flipping to generate new dataset [17]. This technique is performed by shifting the existing images randomly in either X-direction or Y-direction in a plane.

After applying augmentation on the existing image dataset, the performance of this baseline convolutional classification model is represented by the mathematical arrangements of rows and columns, i.e., a matrix. This matrix is known as classification matrix or confusion matrix [18]. This classification matrix is used to visualize the performance in the form of number of rows and columns. On the basis of confusion matrix, convolutional model can also describe the classification report to achieve the highest performance in the terms of precision, recall, and F1 score [19].

$$\text{Precision} = (\text{True Positive})/(\text{True Positive} + \text{False Positive}) \tag{1}$$

$$\text{Recall} = (\text{True Positive})/(\text{True Positive} + \text{False Negative}) \tag{2}$$

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Recall})/(\text{Precision} + \text{Recall}) \tag{3}$$

Table 1 Multiclass confusion matrix and classification report

Confusion Matrix

[[23	1	83	60]
[1	0	4	4]
[81	6	310	258]
[62	2	218	163]]

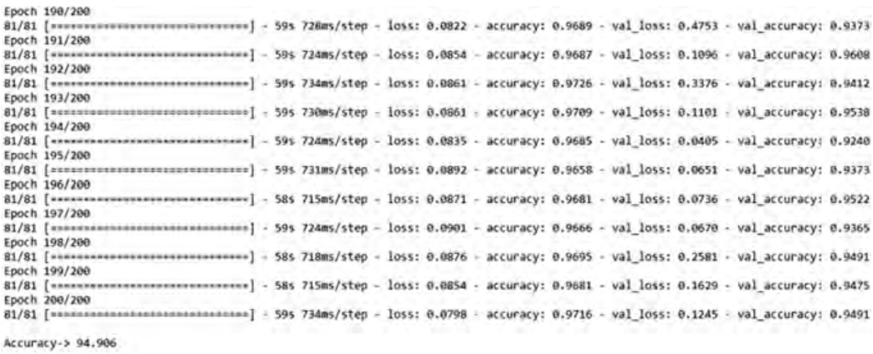
Classification Report

	precision	recall	f1-score	support
mild	0.14	0.14	0.14	167
mod	0.00	0.00	0.00	9
nond	0.50	0.47	0.49	655
vmd	0.34	0.37	0.35	445
accuracy			0.39	1276
macro avg	0.24	0.24	0.24	1276
weighted avg	0.39	0.39	0.39	1276

The multiclass confusion matrix and classification report for AD detection and classification of this proposed screening system were described in Table 1.

To achieve excellent performance of three block sequential model, hyperparameters and fine tuning is performed [20]. Model gets 97.16% training accuracy and 94.91% validation accuracy with respect to 200 epoch and 81 steps as shown in Fig. 6.

Once this sequential model gets trained by input images, there may be a chance of overfitting or underfitting. This can be controlled and visualized by exploration of the classification accuracy and cross-entropy loss with respect to each epoch in Fig. 7. This classification accuracy is much better than other traditional machine learning and convolutional models performed on Alzheimer's disease detection and classification [21].



```
Epoch 190/200
81/81 [==============================] - 59s 728ms/step - loss: 0.0822 - accuracy: 0.9689 - val_loss: 0.4753 - val_accuracy: 0.9373
Epoch 191/200
81/81 [==============================] - 59s 724ms/step - loss: 0.0854 - accuracy: 0.9687 - val_loss: 0.1096 - val_accuracy: 0.9608
Epoch 192/200
81/81 [==============================] - 59s 734ms/step - loss: 0.0861 - accuracy: 0.9726 - val_loss: 0.3376 - val_accuracy: 0.9412
Epoch 193/200
81/81 [==============================] - 59s 730ms/step - loss: 0.0861 - accuracy: 0.9709 - val_loss: 0.1101 - val_accuracy: 0.9538
Epoch 194/200
81/81 [==============================] - 59s 724ms/step - loss: 0.0835 - accuracy: 0.9685 - val_loss: 0.0405 - val_accuracy: 0.9240
Epoch 195/200
81/81 [==============================] - 59s 731ms/step - loss: 0.0892 - accuracy: 0.9658 - val_loss: 0.0651 - val_accuracy: 0.9373
Epoch 196/200
81/81 [==============================] - 58s 715ms/step - loss: 0.0871 - accuracy: 0.9681 - val_loss: 0.0736 - val_accuracy: 0.9522
Epoch 197/200
81/81 [==============================] - 59s 724ms/step - loss: 0.0901 - accuracy: 0.9666 - val_loss: 0.0670 - val_accuracy: 0.9365
Epoch 198/200
81/81 [==============================] - 58s 718ms/step - loss: 0.0876 - accuracy: 0.9695 - val_loss: 0.2581 - val_accuracy: 0.9491
Epoch 199/200
81/81 [==============================] - 58s 715ms/step - loss: 0.0854 - accuracy: 0.9681 - val_loss: 0.1629 - val_accuracy: 0.9475
Epoch 200/200
81/81 [==============================] - 59s 734ms/step - loss: 0.0798 - accuracy: 0.9716 - val_loss: 0.1245 - val_accuracy: 0.9491

Accuracy-> 94.906
```

Fig. 6 Training and validation classification accuracy

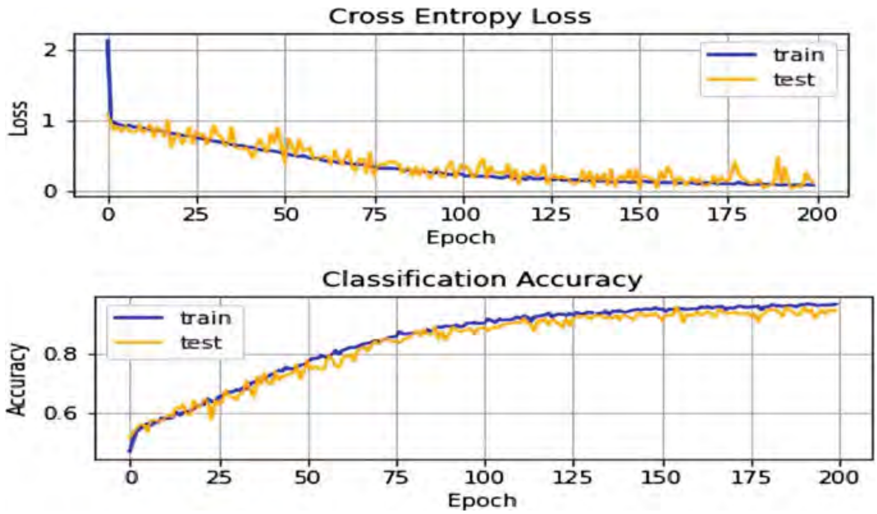


Fig. 7 Classification model performance and cross-entropy loss with respect to number of epoch

3.2 GUI-Based Alzheimer's Disease Screening System

Once the proposed convolutional model is got developed, then it required to deploy to the end-user. Different types of demented Alzheimer's disease can be detected and select accurately using a web-based graphical user interface applications. This web application for disease screening system is developed using keras model and flask API. Screening system is trained by sequential convolutional model and then makes graphical user interface for AD type classification. This graphical user interface-based screening system is categorized in three submodules. These modules are:

- Alzheimer's Disease home page
- Brain MRI Random Image selection dialog box
- Predict desired demented AD detection page.

3.2.1 Alzheimer's Disease Home Page

After development of baseline sequential model for training and validating, authors proposed a HTML web page for AD detection using flask API to interact with training and testing databases [22]. In the coding part, authors set the home page using the name index.html as shown in Fig. 8. Flask API framework has an inbuilt lightweighted web server; hence, it required minimal configuration to execute [23]. Flask server is internally managed by the own model.

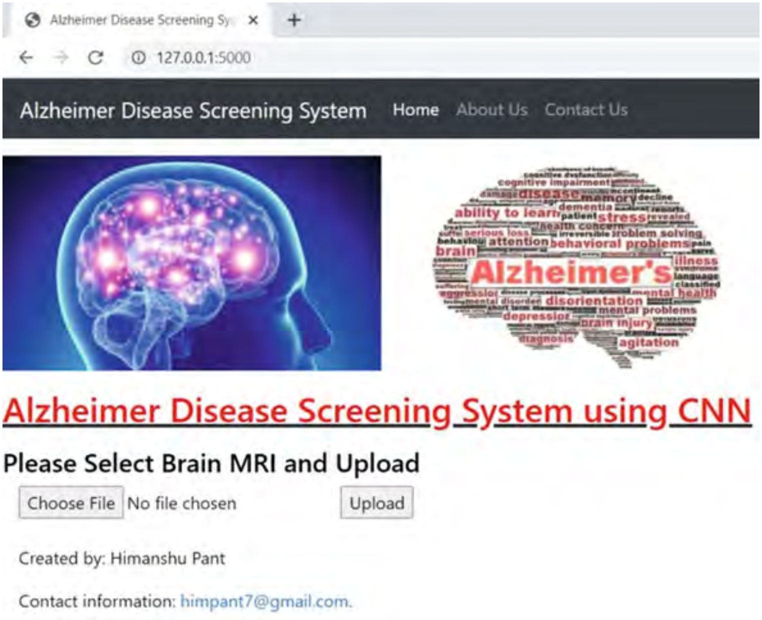


Fig. 8 Web-based AD screening system home page

3.2.2 Brain MRI Random Image Selection Dialog Box

Once the home page GUI for AD screening system is developed, end-user can select any brain MRI image randomly from four described categories or anyone can select the particular unseen testing images from our MRI demented dataset which describe in Fig. 9.

3.2.3 Predict Desired Demented AD Detection Page

After selecting a random brain MRI image from the one of the four desired demented testing class or from unknown source, disease screening model must predict the accurate demented class so doctors or patients can easily predict and classify the target disease whether it is mild demented, moderate demented, non-demented, or very mild demented Alzheimer's disease. In this paper, proposed screening system predicts approximate 98% training accurate result. The prediction of the desired image describes the types belong to respective AD as shown in Fig. 10.

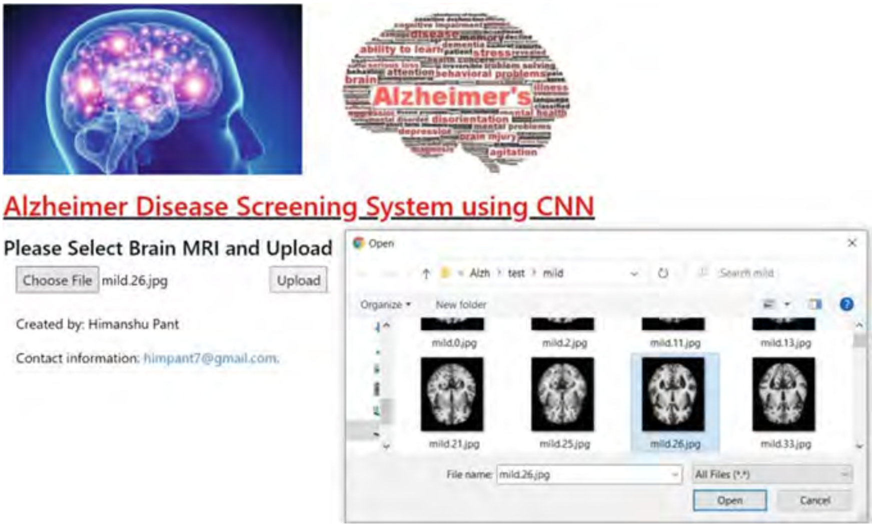


Fig. 9 Web-based AD screening system select MRI page

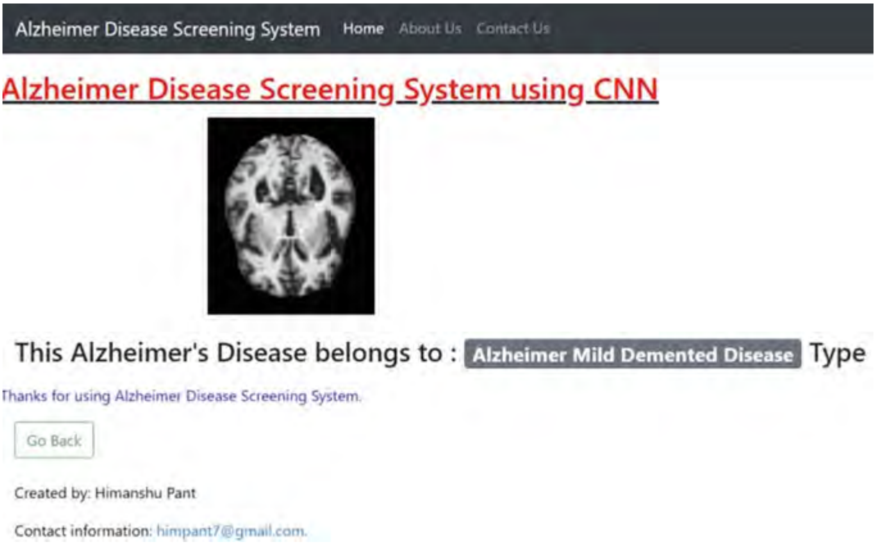


Fig. 10 Web-based AD predict class page

4 Conclusion

This research paper applied VGG-16-oriented baseline convolutional neural network to select, detect, and predict accurate Alzheimer's disease from four different demented classes. Different researchers are doing research on early detection of Alzheimer's disease to prevent the dementia with the help of different brain MRI images. Web-based graphical user interface are used to interact and select brain MRI directly so that end-user easily predict the types of demented Alzheimer's disease class. Authors described three block CNN architecture to train and fit the screening system. Apart from training and fitting the model, authors also developed graphical user interface-based AD screening system using python flask and keras API in a real world. The estimated neural network model and GUI-based disease screening detector are perfectly classified and predict demented Alzheimer's disease in early phase. Proposed model has given training classification accuracy on brain MRI images with 97.16% and validation accuracy with 94.906%. The response time for the disease detection of the proposed model is 0.07421755790710449 ms. The proposed architecture detects and classifies different types of demented AD and also performed outstanding operations with compared to other classification algorithms by means of various statistical measures.

5 Future Scopes

Validation accuracy of the disease screening system can be improved by using large-scale training MRI dataset. The accuracy of different types of AD prediction and classification can be improved by applying numerous pre-defined image classification models like VGG 16, VGG 19, AlexNet, and ImageNet. Along with these algorithms, authors will apply other machine learning and computer vision algorithms for different disease detection in future. Disease prediction and detection from the infected MRI images can be performed in other brain-related diseases. Moreover in future an android-based mobile application for the medical disease detection will be developed so that the entire layman can easily predict the disease on the human body and self-diagnose on the basis of mobile and web applications.

References

1. S. Sarraf, J. Anderson, G. Tofighi, Deepad: Alzheimer's disease classification via deep convolutional neural networks using MRİ and FMRİ, p. 070441 (2016)
2. R. Brookmeyer, E. Johnson, K. Ziegler-Graham, H.M. Arrighi, Forecasting the global burden of Alzheimer's disease. Alzheimer's Dement **3**(3), 186–191 (2007)
3. P. Vemuri, C.R. Jack Jr., Role of structural MRI in Alzheimer's disease. Alzheimers Res Ther **2**, 23 (2010)
4. M. Ewers, R.A. Sperling, W.E. Klunk, M.W. Weiner, H. Hampel, Neuroimaging markers for the prediction and early diagnosis of Alzheimer's disease dementia. Trends Neurosci. **34**(8), 430–442 (2011). https://doi.org/10.1016/j.tins.2011.05.005
5. W.W. Chen, X. Zhang, W.J. Huang, Role of physical exercise in Alzheimer's disease. Biomed Rep. **4**(4), 403–407 (2016). https://doi.org/10.3892/br.2016.607

6. L. Feng, J. Li, J. Yu, et al., Prevention of Alzheimer's disease in Chinese populations: status, challenges and directions. J. Prev. Alzheimers Dis. **5**, 90–94 (2018). https://doi.org/10.14283/jpad.2018.14
7. J. Wang, B.J. Gu, C.L. Masters, Y.J. Wang, A systemic view of Alzheimer disease—insights from amyloid-β metabolism beyond the brain [published correction appears. Nat. Rev. Neurol. **13**(10), 612–623 (2017). https://doi.org/10.1038/nrneurol.2017.111
8. J. Cao, J. Hou, J. Ping, D. Cai, Advances in developing novel therapeutic strategies for Alzheimer's disease. Mol. Neurodegener. **13**(1), 64 (2018). Published 2018 Dec 12. https://doi.org/10.1186/s13024-018-0299-8
9. E. Hosseini-Asl, M. Ghazal, A. Mahmoud, et al., Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. Front. Biosci. (Landmark Ed) **23**, 584–596 (2018). https://doi.org/10.2741/4606
10. K. Ebrahimi, M. Jourkesh, S. Sadigh-Eteghad, et al., Effects of physical activity on brain energy biomarkers in Alzheimer's diseases. Diseases **8**(2), 18 (2020). https://doi.org/10.3390/diseases8020018
11. Brain MRI dataset. Available at: https://www.kaggle.com
12. J. Islam, Y. Zhang, Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. Brain Inf. **5**, 2 (2018). https://doi.org/10.1186/s40708-018-0080-3
13. H. Pant, M.C. Lohani, A. Bhatt, J. Pant, A. Joshi, Soil quality analysis and fertility assessment to improve the prediction accuracy using machine learning approach. Int. J. Adv. Sci. Technol. **29**(3), 10032 (2020). https://sersc.org/journals/index.php/IJAST/article/view/27039
14. Stochastic Gradient Descent, https://towardsdatascience.com/stochastic-gradient-descent-clearly-explained-53d239905d31
15. T. Brosch, R. Tam, A.D.N. Initiative, et al.: Manifold learning of brain MRIS by deep learning, in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2013), pp. 633–640
16. R. Chelghoum et al., Transfer learning using convolutional neural network architectures for brain tumor classification from MRI images. Artif. Intell. Appl. Innov. **583**, 189–200 (2020). https://doi.org/10.1007/978-3-030-49161-1_17
17. S. Deepak, P.M. Ameer, Brain tumor classification using deep CNN features via transfer learning. Comput. Biol. Med. **111**, 103345 (2019). https://doi.org/10.1016/j.compbiomed.2019.103345
18. J. Islam, Y. Zhang, A novel deep learning based multi-class classification method for Alzheimer's disease detection using brain MRI data, in *International Conference on Brain Informatics* (Springer, 2017), pp. 213–222. https://doi.org/10.1007/978-3-319-70772-3_20
19. H. Pant, et al., Impact of physico-chemical properties for soils type classification of OAK using different machine learning techniques. Int. J. Comput. Appl. (0975–8887) **177**(17) (2019)
20. Z.N.K. Swati et al., Brain tumor classification for MR images using transfer learning and fine-tuning. Comput. Med. Graph. **75**, 34–46 (2019). https://doi.org/10.1016/j.compmedimag.2019.05.001
21. K. Oh, Y. Chung, K.W. Kim et al., Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning. Sci. Rep. **9**, 18150 (2019). https://doi.org/10.1038/s41598-019-54548-6
22. All the myths you need to know about Alzheimer's disease. Available at: https://solmeglas.com/wp-content/uploads/2019/07/alzheimers-disease-presenilin-protein-1.jpg
23. PDE-9 inhibitors: Potential therapeutics for the treatment of Alzheimer's disease. Available at: https://medium.com/@Innoplexus/pde-9-inhibitors-potential-therapeutics-for-the-treatment-of-alzheimers-disease-c3866c2e12b6

Performance Analysis of Different Deep Learning Architectures for COVID-19 X-Ray Classification

K. S. Varshaa, R. Karthika, and J. Aravinth

Abstract A chest radiograph is a chest projection radiograph that has been used to diagnose the disorders that affect the chest, its contents and structures in the vicinity. The chest X-ray of a pneumonia affected COVID-19 patient differs from a healthy person's chest X-ray. Differentiating between them is difficult for the untrained human eye. But deep learning networks can learn to distinguish these distinctions. This paper analyses the performance of seven different models: Xception, VGG-16, ResNet-101-V2, ResNet-50-V2, MobileNet-V2, DenseNet-121 and Inception-ResNet-V2, when differentiating between COVID-19 and normal chest X-rays. The experiment's results on the COVID-19 chest X-ray dataset inferred that the Xception model performed the best. Inception-ResNet-V2 worked fine after Xception.

Keywords COVID-19 · Pneumonia · Artificial intelligence · Chest X-ray · Deep learning architectures

1 Introduction

COVID-19 is a respiratory illness that has taken more than half a million lives since it is started spreading around the world. It is caused by (SARS-CoV-2) which is a type of Corona Virus. It is highly contagious among human beings. In adults who were hospitalized, the fatality rate of the disease ranged from 4 to 11%. It is estimated that the overall fatality rate is between 2 and 3% [1]. It is spread through droplets from an infected person to a non-infected person. The infected person does not need to be symptomatic. They can be asymptomatic as well.

K. S. Varshaa · R. Karthika (✉) · J. Aravinth
Department of Electronics and Communication Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: r_karthika@cb.amrita.edu

J. Aravinth
e-mail: j_aravinth@cb.amrita.edu

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_23

The usual symptoms of COVID-19 are fatigue, fever, myalgia, dry cough, dyspnoea and anorexia. Pneumonia can also manifest within three weeks of the infection. In fact, COVID-19 is mainly a respiratory disease. Pneumonia can be diagnosed with the help of chest X-rays [2]. However, radiologists are needed to read these X-rays. Due to the exhaustion of hospital resources in many places, radiologists are not always readily available. Artificial intelligence can be the solution to this problem. It is fast and time-efficient. Deep learning is an emerging but significant field in artifical intelligence [3]. In this paper, we compare different deep learning models capable of differentiating between COVID-19 positive and negative X-rays. The models we are going to be using are Xception, VGG-16, ResNet-101-V2, ResNet-50-V2, MobileNet-V2, DenseNet-121 and Inception-ResNet-V2.

The remainder of the paper is organized as follows. Section 2 describes related work; Sect. 3 provides the proposed models; Sect. 4 describes results and analysis; and Sect. 5 provides the conclusion.

2 Related Work

In [4], a deep neural network in combination with support vector machine (SVM) was proposed. SVM was preferred over a deep learning model because the latter requires a large dataset for training. A feature of the input image was extracted by the convolutional neural layer and was fed into the SVM. The SVM classifies the input as one of the three categories, COVID-19, pneumonia and normal. In [5], a convolutional neural network in combination with a modified AlexNet was presented. A pre-processed X-ray and CT scan dataset are fed into this network. The network classifies the images in the dataset as either COVID-19 or normal.

In [6], a new technique called COVID-ResNet is proposed. It involves fine tuning a pre-trained ResNet-50 model. This results in reduced trained time and an increase in the performance of the model. It is a three-step technique consisting of progressive resizing of input images of dimensions $229 \times 229 \times 3$, $224 \times 224 \times 3$ and $128 \times 128 \times 3$. The network is fine-tuned at each stage. COVID-ResNet technique also employs automatic learning rate selection. In [7], a completely automated tuberculosis identification system is proposed. Template matching is used to identify tuberculosis positive candidates. It is based on a two-dimensional Gaussian model. Image enhancement involving a Hessian matrix is used for feature extraction and cavity segmentation.

In [8], the use of fuzzy colour technique during pre-processing was proposed. Two sets of COVID-19 datasets were stacked together and were used for the experiment. The dataset was trained with two neural networks, SqueezeNet and MobileNetV2. Social mimic optimization method is used to process the features extracted by the models. Efficient features were extracted from the images and combined. SVM is used for the classification of the features. In [9], usage of multiple convolutional neural networks (CNNs) to evaluate the input image (chest X-ray) was proposed. The network is trained to look for any abnormalities in the input image. It classifies the X-ray as either normal or abnormal. A method to interpret the obtained results was also proposed. In [10], the performance of VGGNet and CNN for classifying the Fashion MNIST dataset was reviewed and the metrics were compared. [11] proposes and compares the performance of different deep learning networks in biometric security, specifically raw ECG data. [12] proposes a decision support system (based on image retrival) to help a physician diagonse breast cancer.

3 Proposed Models

3.1 VGG16

This model replaces the large-sized kernels of traditional convolutional networks with smaller 3×3 kernels. This increases model performance because deep nonlinear layers have the ability to learn more complex features. VGG16 has multiple convolution layers, max-pooling layers and dense layers. A sequence of convolutional layers is followed by three dense layers. Softmax is the final layer. The hidden layers have the ReLU activation function. With every pooling layer, the network width increases by a factor of 2. The maximum network width is 512 [13].

3.2 ResNet-50-V2/ResNet-101-V2

Deeper networks are harder to train. With the depth of the network increasing, model accuracy begins to saturate and degrade rapidly. But overfitting is not the cause of the deterioration in accuracy. Increasing the number of layers inside the network increases the error in training. However, deeper models produce better results in image recognition. Residual networks or ResNets are one of the solutions to the degradation problem. They present the feed-forward shortcut connections between the layers. The feed-forward connections perform identity transformation which is added to the outputs of the stacked layers.

The deep residual network is easier to optimize and has increased accuracy gains due to the significantly increased depth [14]. The 50/101/152-layer ResNet architectures have more accuracy than the 34-layer ResNet architecture. ResNet-V2 is an improved version of the ResNet-V1 architecture. ResNet-V2 performs batch normalization and ReLU activation on the input before convolution. But in ResNet-V1, batch normalization and ReLU activation is followed by convolution. ResNet-V1 has a second ReLU activation function at the end of its network but ResNet-V2 does not. ResNet-V2 improves generalization and makes training easier [15].

3.3 DenseNet-121

In DenseNet architectures, there are feed-forward connections between the layers of the network. In regular CNN's, each layer is attached to only two other layers. But in dense convolutional networks, each layer is attached to $(L + 1)/2$ layers. The activation maps of the previous layers are concatenated and given as the input to each layer.

In deep neural networks, the information given at the input can vanish before it reaches the end of the network. This is called the vanishing gradient problem. This can be reduced by using the DenseNet architecture. DenseNet strengthens the flow of information from the start to the end of the network. Each layer of the network receives the outputs of every other preceding layer. Every layer has access to the features extracted by the preceding layers. This encourages feature reuse and also decreases the number of parameters in the network.

ResNet has similar layer connections as DenseNet, but ResNet sums up the feature maps it receives from the precedent layers while DenseNet concatenates them. The width of the DenseNet layers is also narrower. The strength of the architecture lies in the feature reuse, resulting in a parameter-efficient, easily trainable model. Due to the efficient use of the parameters, DenseNet is less likely to overfit [16].

3.4 MobileNet-V2

The fundamental block of this architecture is a residual bottleneck depth-separable convolution. Traditional convolutional layers are more computationally expensive than effective depthwise convolution. MobileNetV2 performs 3×3 depthwise separable convolution. This decreases the computational cost with only a small decrease in accuracy. The initial layers are dense layers with 32 filters. The dense layers are followed by residual bottleneck layers. ReLU6 is the activation function for the network. Along with decreasing the computational cost, MobileNet architecture also decreases the memory needed for the operation [17].

3.5 Inception-ResNet-V2

Residual connections in Inception networks have been found to increase the performance of the model slightly. Residual connections also decrease the training time. Inception-ResNet-V2 is therefore a combination of Inception and ResNet architectures. In Inception-ResNet-V2, the filter concatenation stage is replaced with residual connections. A 1×1 convolutional layer follows each inception block. This compensates for the dimensionality reduction produced by the inception block. The inputs to the regular layers undergo batch normalization. But the input to the summation layers does not. Inception architectures with residual connections have a better recall and a smaller training period [18].

3.6 Xception

Xception is the expanded version of the Inception architecture. In Inception networks, convolutions are factorized into multiple blocks. Xception is similar to depthwise separable convolution but there are some notable differences. Depthwise separable convolution first performs channel-wise spatial convolution and then performs 1 × 1 convolution, whereas Xception performs 1 × 1 convolution followed by spatial convolution. After the operation, there is a ReLU nonlinearity in Xception but this is not the case in the depthwise separable convolution. Xception architecture contains 36 convolutional layers. These convolutional layers extract features from the input. Xception architecture contains 14 modules. Each of these modules contains convolutional layers. With the exception of the corner modules, all the other modules have linear residual connections [19].

4 Results and Discussion

4.1 Dataset

This dataset was created by university researchers (Qatar University and the University of Dhaka) with the help of medical doctors and other collaborators from Pakistan and Malaysia. This database contains Normal, COVID-19, and Viral Pneumonia affected chest X-rays. There are 219 COVID-19 positive images, 1341 normal images and 1345 viral pneumonia images. For our experiment, we have used 219 COVID-19 and 600 normal chest X-ray images (Fig. 1).

4.2 Metrics

We have used four metrics, namely accuracy, precision, F1-score and recall.

Accuracy: It is the ratio of correct predictions out of all the predictions. In highly unbalanced datasets, the accuracy score is misleading. In our dataset, we have 219 COVID-19 images and 600 normal images. For roughly every COVID-19 image, there are four normal images, making it an unbalanced data set. Hence, we use other metrics.

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \tag{1}$$

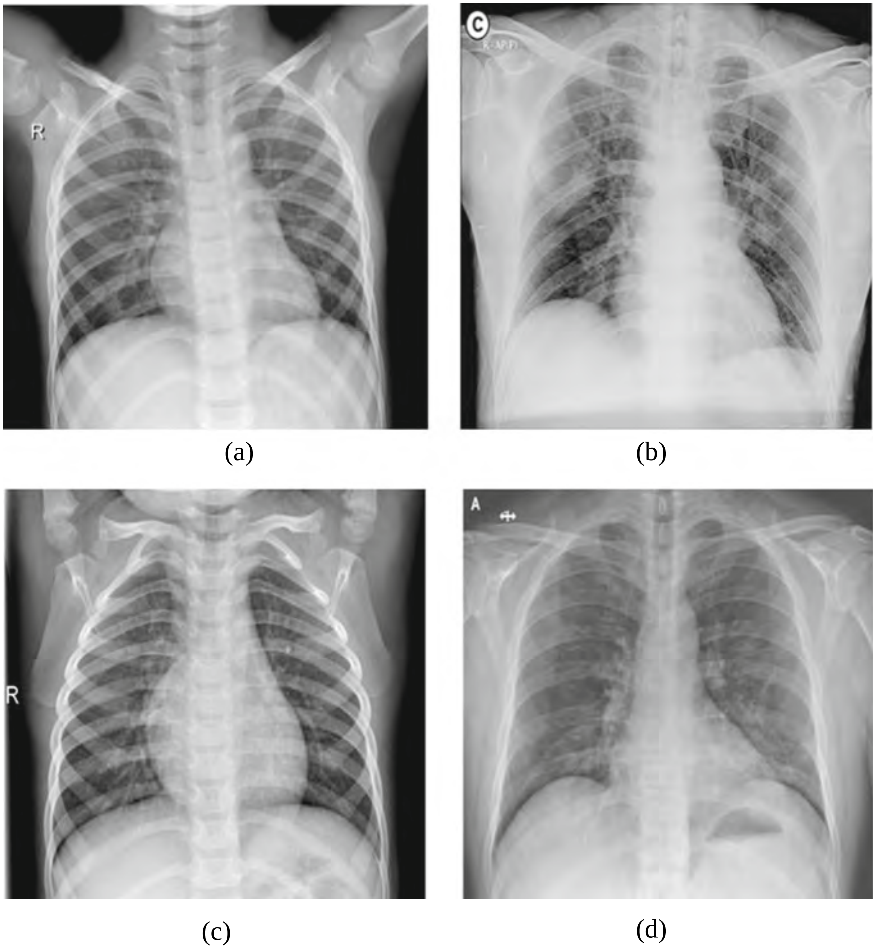


Fig. 1 **a** and **c** are normal chest X-rays. **b** and **d** are COVID-19 chest X-rays

(or)

$$\text{Accuracy} = (\text{No of correct predictions})/(\text{Total number of predictions})$$

Precision: It is the ratio of True Positives (TP) to the sum of True Positives (TP) and False Positives (FP). Precision represents the percentage of correct positive predictions out of all positive predictions.

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{2}$$

It essentially calculates the accuracy of the minority class. But it does not take into consideration the number of False Negatives (FN). Therefore, we require another metric.

Recall: It is the ratio of True Positives (TP) to the sum of True Positives (TP) and False Negatives (FN). Recall represents the percentage of correct positive predictions out of all possible positive predictions.

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{3}$$

A model can have high recall while having low precision and vice versa. As such, looking at recall and precision as isolated metrics fail to give us the complete picture. To rectify this, we use the F1-score metric.

F1-score: F1-score encapsulates the properties of both precision and recall metrics. It combines precision and recall using the formula given below.

$$\text{F1-score} = (2 * \text{Precision} * \text{Recall})/(\text{Precision} + \text{Recall}) \tag{4}$$

4.3 Experimental Analysis

Given below are the results of our experiment with the seven models Xception, VGG-16, ResNet-101-V2, ResNet-50-V2, MobileNet-V2, DenseNet-121 and Inception-ResNet-V2 when trained and tested with COVID-19 chest X-ray dataset (Figs. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12).

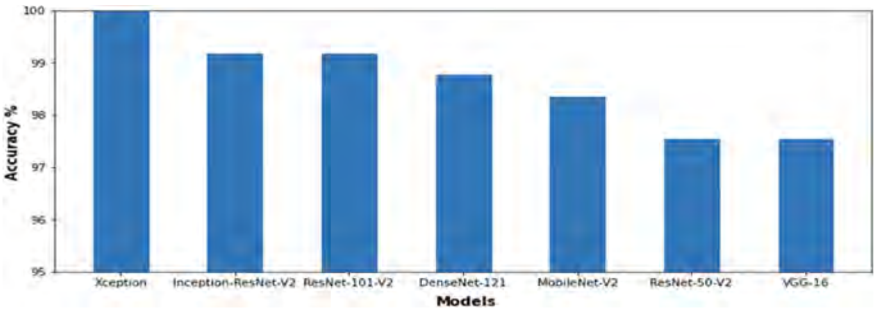


Fig. 2 Accuracy comparison for different models trained using COVID-19 dataset

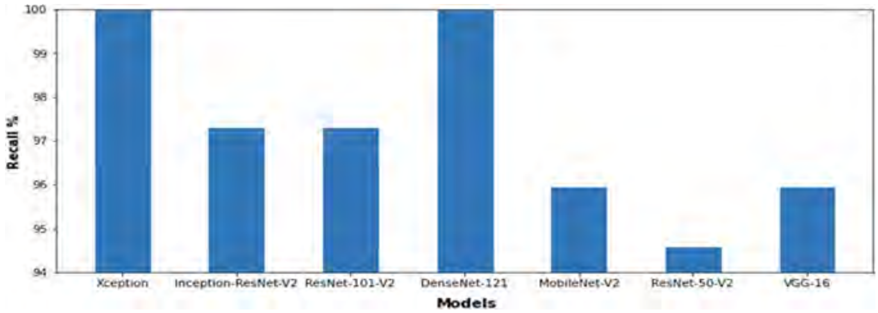


Fig. 3 Recall comparison for different models trained using COVID-19 dataset

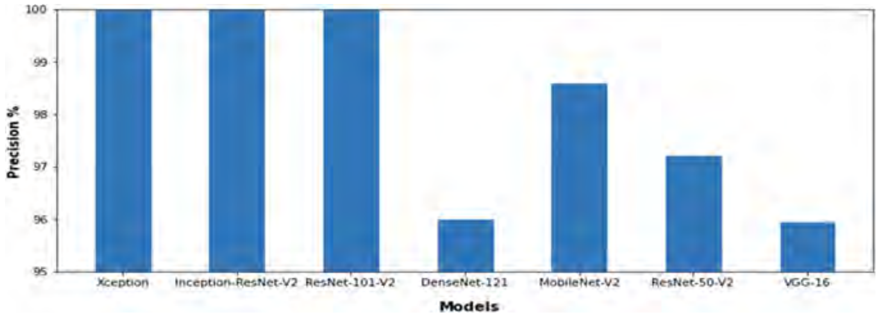


Fig. 4 Precision comparison for different models trained using COVID-19 dataset

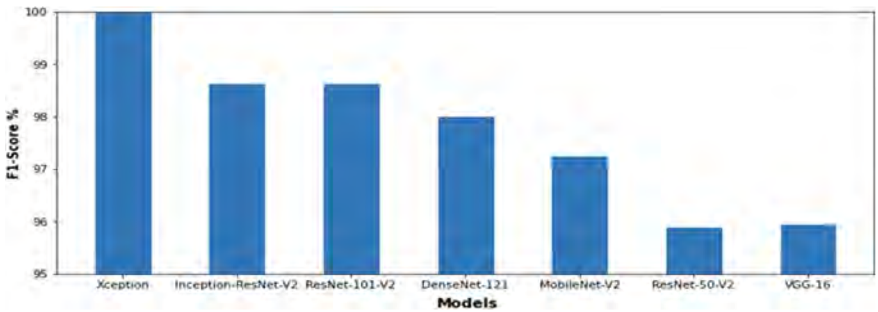


Fig. 5 F1-score comparison for different models trained using COVID-19 dataset

4.4 Inference

The size of the training set and test set is 573 and 246, respectively. The training set contains 145 COVID-19 and 428 normal images. The test set contains 74 COVID-19 and 172 normal images. The batch size is 16. The learning rate is 0.001, and it is

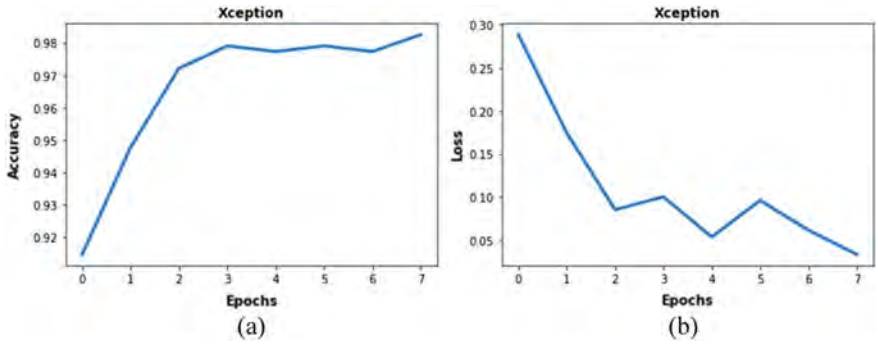


Fig. 6 **a** Accuracy versus epochs. **b** Loss versus epochs for Xception architecture

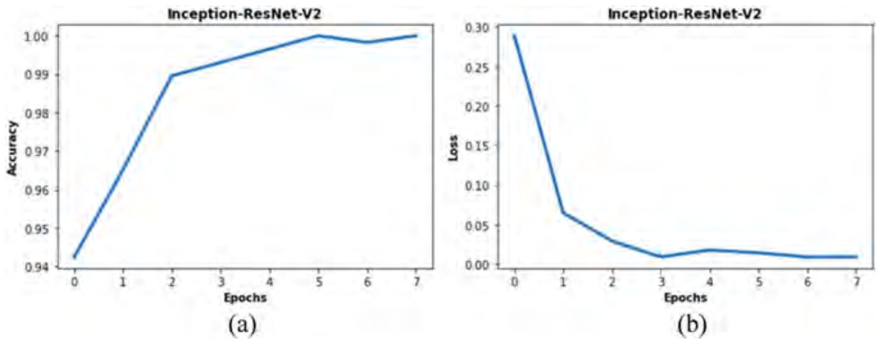


Fig. 7 **a** Accuracy versus epochs. **b** Loss versus epochs for Inception-ResNet-V2 architecture

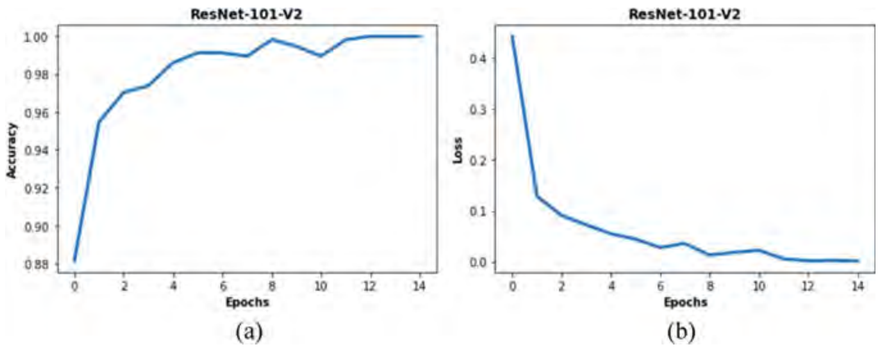


Fig. 8 **a** Accuracy versus epochs. **b** Loss versus epochs for ResNet-101-V2 architecture

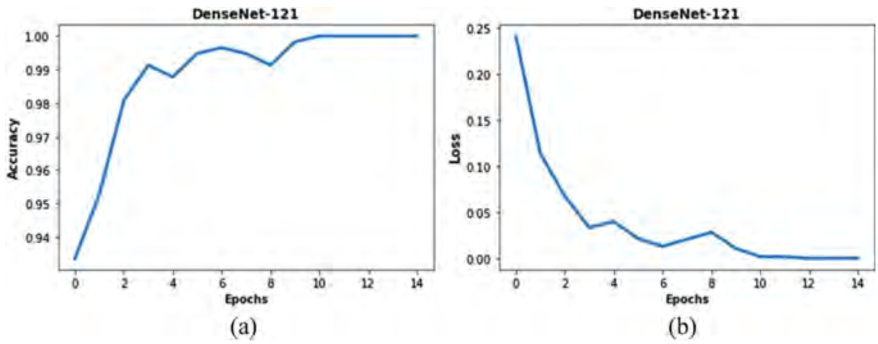


Fig. 9 **a** Accuracy versus epochs. **b** Loss versus epochs for DenseNet-121 architecture

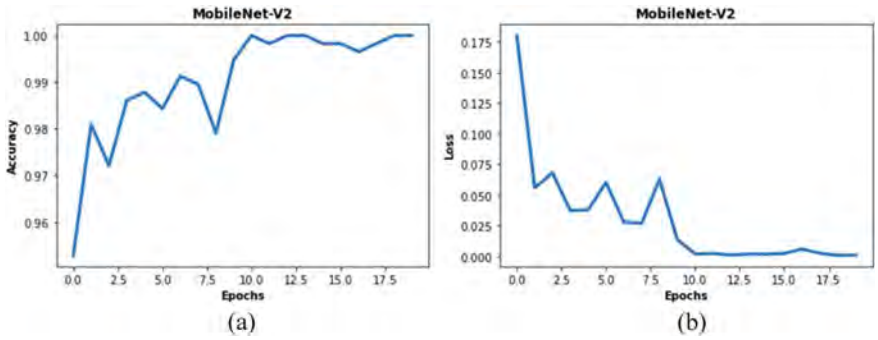


Fig. 10 **a** Accuracy versus epochs. **b** Loss versus epochs for MobileNet-V2 architecture

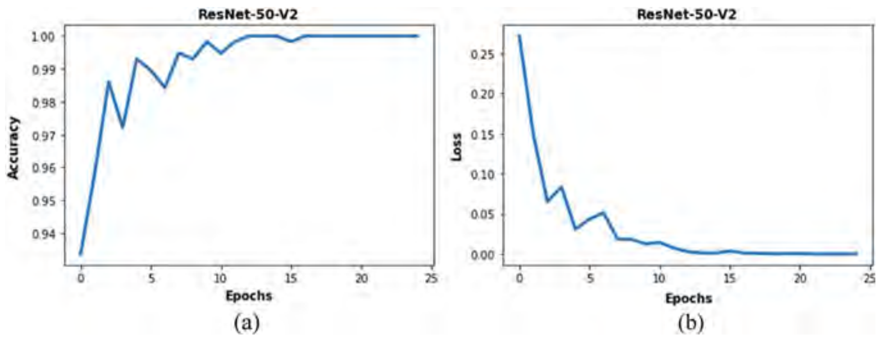


Fig. 11 **a** Accuracy versus epochs. **b** Loss versus epochs for ResNet-50-V2 architecture

constant for the first 8 epochs. After that, it reduces by 10% for every epoch. The size of the input images is 224×224.

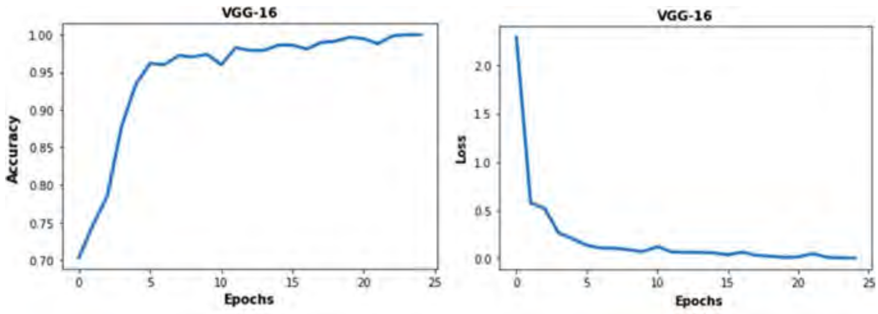


Fig. 12 **a** Accuracy versus epochs. **b** Loss versus epochs for VGG16 architecture

Looking at the results obtained above, it can be inferred that Xception and Inception-ResNet-V2 architectures have the highest accuracy scores. The architectures with slightly lower accuracy scores are DenseNet-121 and ResNet-101-V2. ResNet-50-V2 and VGG16 had the lowest accuracy scores among all the trained models.

Considering the recall metric, Xception and DenseNet-121 performed the best. A model with a high recall percentage generates fewer false negatives. With respect to the precision metric, Xception, ResNet-101-V2 and Inception-ResNet-V2 have the highest scores. A model with a high precision percentage generates fewer false positives.

Since it is a highly unbalanced dataset, F1-score is the most important performance metric. Considering F1-score, Xception, ResNet-101-V2 and Inception-ResNet-V2 have the highest score. They are closely followed by DenseNet-121. ResNet-50-V2 and VGG-16 had by far the lowest scores.

Xception was able to achieve its high accuracy and F1-score with only 8 epochs. Inception-ResNet-V2 also achieved comparable scores with only 8 epochs. The other models required 15 or more epochs to achieve the accuracy and F1-scores noted above.

VGG16 exhibited the worst performance out of the seven architectures. It is to be noted that VGG16 has the least number of layers (19 layers) among all the models. It also does not have residual connections. ResNet-50-V2 exhibited slightly better performance than VGG16. It also has a smaller number of layers (190 layers) as compared to the other models.

It is to be noted that Inception-ResNet-V2 has 780 layers while the best performing model, Xception has only 132 layers.

5 Conclusion

COVID-19 is a global pandemic that is straining hospital resources in many countries. Automating the reading of chest radiographs is a small step in alleviating the strain. Given adequate training, computers can achieve a high accuracy in reading the chest radiographs. A performance study of a few state-of-the-art deep learning models is presented in this paper. From our experimentation, we observed that Xception model achieves the highest accuracy and F1-score despite the small size of the dataset. It had a 0.81% accuracy improvement over Inception-ResNet-V2 and ResNet-101-V2, 1.22% over DenseNet-121, 1.66% over MobileNet-V2 and 2.44% over both ResNet-50-V2 and VGG-16. It had a 1.37% F1-score improvement over Inception-ResNet-V2 and ResNet-101-V2, 1.99% over DenseNet-121, 2.74% over MobileNet-V2, 4.1% over ResNet-50-V2 and 4.05% over VGG-16. It was also trained in the least time, with the least number of epochs. It would be the best model for this work. In the future, with improvement in the quality and quantity of the datasets for various respiratory diseases including but not limited to COVID-19, we can automate the process of reading radiography results to a large extent.

References

1. T. Singhal, A review of coronavirus disease-2019 (COVID-19). Indian J. Pediatr. **87**, 281–286 (2020)
2. World Health Organization, Pneumonia Vaccine Trial Investigators' Group & World Health Organization. Standardization of interpretation of chest radiographs for the diagnosis of pneumonia in children/World Health Organization Pneumonia Vaccine Trial Investigators' Group. World Health Organization (2001)
3. R. Ramachandran, D.C. Rajeev, S.G. Krishnan, P. Subathra, Deep learning—an overview. Int. J. Appl. Eng. Res. **10**, 25433–25448 (2015)
4. B. Sethy, P.K. Ratha, S.K. Biswas, Detection of Coronavirus Disease (COVID-19) Based on Deep Features and Support Vector Machine. Preprints (2020)
5. H.S. Maghdid, A.T. Asaad, K.Z. Ghafoor, A.S. Sadiq, M.K. Khan, *Diagnosing COVID-19 Pneumonia from X-Ray and CT Images Using Deep Learning and Transfer Learning Algorithms* (2020)
6. M. Farooq, A. Hafeez, *COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs* (2020)
7. T. Xu, I. Cheng, M. Mandal, Automated cavity detection of infectious pulmonary tuberculosis in chest radiographs, in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2011), pp. 5178–5181
8. M. Togacar, B. Ergen, Z. Comert, Covid-19 detection using deep learning models to exploit social mimic optimization and structured chest x-ray images using fuzzy color and stacking approaches. Comput. Biol. Med. 121 (2020)
9. P.N. Kieu, H.S. Tran, T.H. Le, T. Le, T.T. Nguyen, Applying multi-CNNs model for detecting abnormal problem on chest x-ray images, in *10th International Conference on Knowledge and Systems Engineering (KSE)* (2018), pp. 300–305
10. B. Saiharsha, B. Diwakar, R. Karthika, M. Ganesan, Evaluating performance of deep learning architectures for image classification, in *5th International Conference on Communication and Electronics Systems (ICCES)* (2020), pp. 917–922

11. Y. Muhammed, J. Aravinth, CNN based off-the-person ECG biometrics, in *International Conference on Wireless Communications, Signal Processing and Networking. WiSPNET* (2019), pp. 217–221
12. K.K. Vijayan, Retrieval driven classification for mammographic masses, in *Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing ICCSP 2019* (2019), pp. 725–729
13. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition (2014)
14. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 770–778
15. K. He, X. Zhang, S. Ren, J. Sun, Identity mappings in deep residual networks, in *European Conference on Computer Vision* (2016), pp. 630–645
16. G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 2261–2269
17. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.C. Chen: MobileNetV2: inverted residuals and linear bottlenecks, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2018), pp. 4510–4520
18. C. Szegedy, S. Ioffe, V. Vanhoucke, Inception-v4 inception-resnet and the impact of residual connections on learning. Thirty-First AAAI Conf. Artif. Intell. **131**, 262–263 (2016)
19. F. Chollet, Xception: deep learning with depthwise separable convolutions, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 1251–1258

Random Grid-Based Visual Cryptography for Grayscale and Colour Images on a Many-Core System

M. Raviraja Holla and Alwyn R. Pais

Abstract The traditional visual cryptography (VC) is a technique used to encrypt the secret images in several shares with minimum computation so that there will be no processing for decryption. A conventional random grid-based VC is a different cryptosystem that generates an encoded grid one pixel at a time based on the pixel in the original image by using a single master grid. These encrypted grids are of the same size as the original image, unlike traditional VC. No other matrices are required to generate these grids. Despite its simplicity and flexibility, it is inefficient for the real-time applications due to its reduced efficiency in the large image sizes. Thus, it is necessary to exploit the currently prevalent many-core computing power to elevate this cryptosystem for achieving a better efficiency. This paper proposes a novel (2, 2) random grid-based VC that exploits a many-core system's computational capability applied to grayscale and colour images. For more efficiency, this approach uses Compute Unified Device Architecture (CUDA) constant memory. This approach finds significance from the perspective of efficiency demand and the rapid growth in many-core computing. Experimental results proved that, the proposed method on a many-core system outperforms the normal random grid-based VC with an improved speedup.

Keywords Many-core · Image · Random grid · Cryptography · Shares · Speedup · Cuda

---

M. Raviraja Holla (✉) · A. R. Pais
Information Security Research Lab. Department of Computer Science and Engg., National Institute of Technology Karnataka, Surathkal 575025, India
e-mail: raviraj.holla@manipal.edu

A. R. Pais
e-mail: alwyn@nitk.ac.in

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_25

1 Introduction

In the days when the cryptography started, there was no attention for the future utility. There was a belief that more security should rely on more computation. Now the information is not just letters. Picture, sound and video are all accepted as information. The availability of such information elsewhere should be made safely available in real time. The image is considered as a precious medium-sized information. Conventional cryptography of extensive computation may not be more effective in achieving the efficiency. Therefore, image cryptography has been started with an objective of little processing during the encryption for medium-sized information. Image security models have new characteristics and capabilities [8].

The visual cryptography that Naor and Shamir [16] engendered is continuously appealing to researchers. This field is now with a large corpus of literature in the research [4]. The variants of this scheme cost additional storage overhead for the basis matrices and the larger shares. The dealers also transmit shadows. Hence, they consume additional traffic bandwidth. Moreover, this scheme requires cumbersome effort in designing appropriate basis matrices [2]. Besides, the image encryption proposed in [9] attracted many of the researchers in the last two decades. The inventions based on this method are generally called random grid (RG)-based techniques. The investigators came to the view that the RG models remedy the drawbacks of [16]. The first proposed refinement in [17] is an extension of the technique proposed in [9] to the colour images. But innovations in image encryption, while effective, do not place much emphasis on efficiency even with multicore central processing units (CPUs). These CPU-based models do not fit for real-time applications. The sequential model even leads to resource under-utilization [18]. The computing technology has evolved from multicore to many-core processing units. These many-core processing units are throughput optimized devices in contrast with the latency optimized multicore processing units. Although the processing elements in a many-core device are simple, their massive quantity is conducive to data-parallel tasks. Besides, each such processing unit has the multitask capability. The general-purpose graphics processing unit (GPGPU) or merely the graphics processing unit (GPU) is a data-parallel many-core system optimized for throughput. This paper reformulates the RG image encryption proposed in [17] to leverage the computing power of GPGPU. Such reformulations are novel approaches in research [7, 19]. Security brings value to the information—the faster the safety, the higher the value of it.

2 Background and Motivation

2.1 Efficiency Considerations in VC

The unique developments and features in the VC models have driven their efficiency demand. The computational complexity of the VC is of two kinds. More complex operation VC systems are one category. So the investigators in [1, 3, 5, 10, 20] used mechanisms to simplify the complexity. The calculation is more because there is more data in another category. The VC methods that allow sharing of more than one secret image require more computation on enormous data [8]. In traditional VCs, the pixel expansion takes exponential growth rapidly if the number of participants increases [12]. The premise of the image encryption is that there should be less computation in encryption and no computation in decryption. But some systems also include computation in decryption [22]. In such a situation, efficiency cannot be underestimated. The VC schemes require efficient design and implementation in cloud-based applications [15]. A generalized general access structure (GGAS) VC approach in [24] focuses on efficiency while recovering the secrecy, particularly in real-time applications. The scheme proposed in [21] reduces the pixel expansion problem and also provides flexible sharing solution. In achieving these objectives, the random grid, XOR operation for the colour pixels and GGAS are combined. Increasing visual quality in the cryptosystem [2, 6, 23] also highlights the need for efficiency. Similarly, the need for rapid communication [11] also indirectly puts demand for efficiency.

2.2 Trends in Many-Core Systems and CUDA as a Platform

The many-core systems are available in fused and discrete architectures [14]. The multicore and many-systems integrated into a single chip are the fused design. Peripheral component interconnect express (PCIe) connecting these two systems integrated into separate chips is the discrete architecture [25]. Table 1 shows future trend in multicore and many-core computing. Therefore, the cryptosystem needs to be upgraded for better use of new technology. The system needs to be changed to recognize parallel work. In this paper, there is the novelty of increasing the efficiency of the RG cryptosystem with the optimum resource utilization using a discrete GPGPU architecture.

Table 1 Future trends in multicore and many-core systems [14]

Hardware attributes	Multicore system	Many-core system
1. Number of transistors	Crosses 10 billion	8 billion
2. Number of cores	Crosses 60	3072
3. Size of LLC	96 MB	2048 KB
4. 3D integrated circuit	Exists	Exists
5. Interconnectivity bandwidth	5–12 times bandwidth compared to PCIe-Gen3 by NVLink	

3 Random Grid-Based VC for Grayscale and Colour Images

This section explains the primary random grid for the grayscale and colour images. Three basic grayscale random grid models presented in [9] generate an encrypted or encoded random grid E_2 pixels using a master random grid E_1 based on the pixel values in the binary converted plain image I. Equations (1), (2) and (3) represent these three models. The [p, q] in each equation indicates the pixel position. The E_1 is preset to have a 50% probability of zeros and ones. E_1 and E_2 when overlaid on each other reveal the plain image. Otherwise, one cannot discern any secrecy from grids.

$$E_2[p,q] = \begin{cases} E_1[p,q] & I[p,q] = 0 \\ \overline{E_1[p,q]} & \text{otherwise} \end{cases} \tag{1}$$

$$E_2[p,q] = \begin{cases} E_1[p,q] & I[p,q] = 0 \\ \text{random}(0 \ or \ 1) & \text{otherwise} \end{cases} \tag{2}$$

$$E_2[p,q] = \begin{cases} \text{random}(0 \ or \ 1) & I[p,q] = 0 \\ E_1[p,q] & \text{otherwise} \end{cases} \tag{3}$$

The research in [17] is an extension of RG model in [9] for colour images. There are three primary independent colours in the subtractive colour model: they are cyan (C), magenta (M) and yellow (Y). Other colours can be considered the linear combination of these basic colours. Alternatively, a pixel in the given image decomposed to its equivalent three monochromatic coloured-grey levels. These three coloured-grey level images are converted to coloured halftone images. Halftone equivalent of the grayscale saves memory with quality [13]. Three submaster random grids are generated. Three subencoded random grids are generated using a technique proposed in [9]. Then the corresponding colour components of subrandom grids are mixed to obtain the two coloured master and encoded random grids. When these two random grids are superimposed, the colour secret image gets revealed.

4 Performance of Sequential Random Grid-Based VC for Grayscale and Colour Images

The performance aspects of sequential (2, 2) random grid-based VC is studied using a system with dual-core Intel Core i3-2370M Processor and 4 GB RAM. The model is tested with the grayscales and the colour images of different sizes. Table 2 lists the execution times for the grayscale images of four different resolutions.

Table 2 Execution time for the grayscale image

Image size	Execution time (s)		
	T1	T2	Total
256 × 256	0.893	0.327	1.220
512 × 512	3.260	1.436	4.696
720 × 576	4.812	2.386	7.198
1024 × 1024	11.300	5.163	16.463

Note T1—time for halftoning, T2—time for generating share

Table 3 Execution time for the colour image

Image size	Execution time (s)			
	T0	T1	T2	Total
256 × 256	0.561	2.5897	0.936	4.0867
512 × 512	2.251	9.454	3.825	15.53
720 × 576	3.252	13.9548	6.014	23.3308
1024 × 1024	8.304	32.77	15.36	56.437

Note T0—time for decomposing, T1—time for halftoning, and T2—time for generating share.

T1 is the execution time for generating a halftone image in seconds. The execution time for creating two shares labelled T2. Last column $Total$ is the total execution time. The data transfer between RAM and secondary storage is not considered in obtaining these times. Hence, they are only the processing times when the images available in RAM. Similarly, Table 3 shows the execution times for the colour images at various stages while processing. An additional column labelled T0 is the time to decompose the colour image into its CMY image components. The run time to obtain corresponding three halftone images is labelled T1. The execution time to generate six shares is T2. The Total is the sum of these time components.

Figure 1 depicts the execution times for generating the halftone images for grayscale and colour images. As evident from the figure, the execution time in producing the halftone colour images increases drastically with the image size. The processing time to generate shares for grayscale and colour images is shown in Fig. 2. Again, the time for creating shares for colour images reaches prohibitive with the increase in image size. In summary, the total execution time for the VC of colour images is computationally expensive as evident from Fig. 3.

The total time includes halftone and shares generating time typical to both grayscale and colour images. These two components consume more execution time. For the grayscale images, these two components take 70% and 30% of the overall execution time. In colour images, they are 61% and 25%, respectively. These percentages remain constant independent of image sizes. The time to decompose the colour image into its CMY components is an additional time component only for the colour images. Moreover, this time component occupies only 14% of the total execution time, irrespective of the image size. This paper optimizes the process of creating halftone and share images using a many-core system.

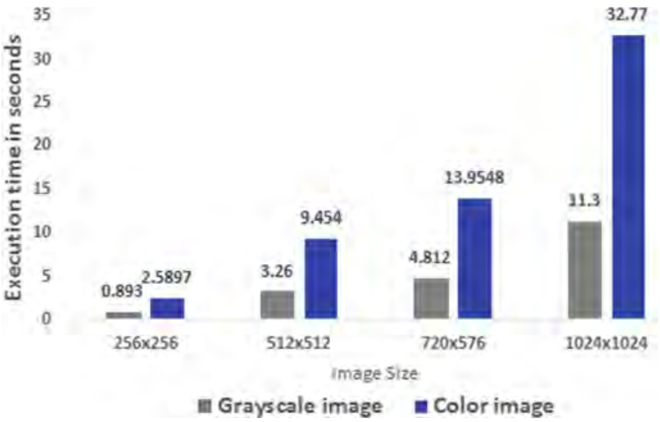


Fig. 1 Execution times for generating halftone images for grayscale and colour images in the sequential algorithm

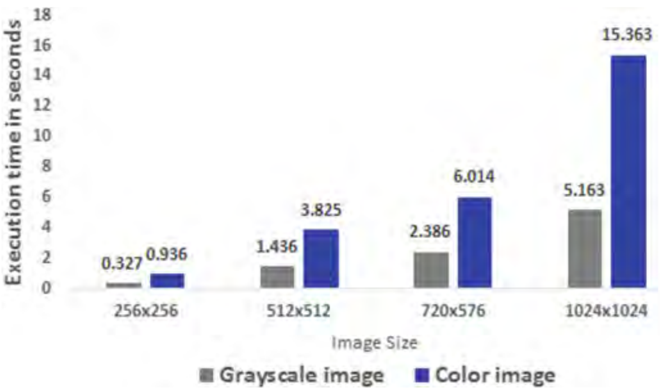


Fig. 2 Execution times for generating shares for grayscale and colour images in the sequential algorithm

5 The Proposed Random Grid-Based (2, 2) Algorithm for Grayscale and Colour Images on a Many-Core System

Figure 4 shows the block diagram of the proposed many-core random grid model. The input to the model is a colour or grayscale image. The halftone image transforms it into the corresponding halftone image. The encryption block generated encoded grid using a master grid based on the pixel values of the halftone image. As the master grid is preset to contain 50% randomness of 0s and 1s, the produced encoded grid using it also introduces randomness. Decryption requires the stacking of the master and the encoded grids to reveal the secret image. The halftoning and the encryption blocks utilize a many-core system to exploit the data-parallel tasks resulting in an

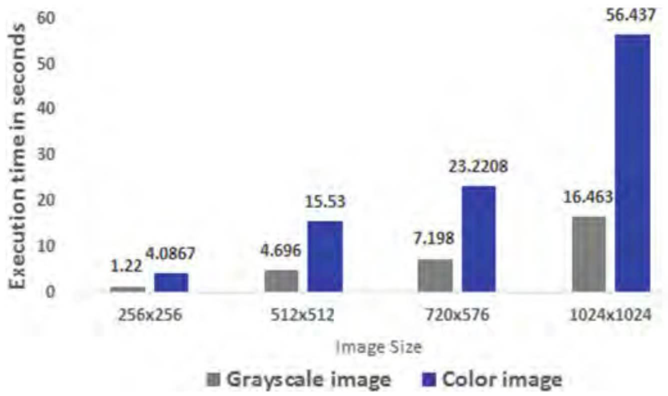


Fig. 3 Total execution times for grayscale and colour images in the sequential algorithm

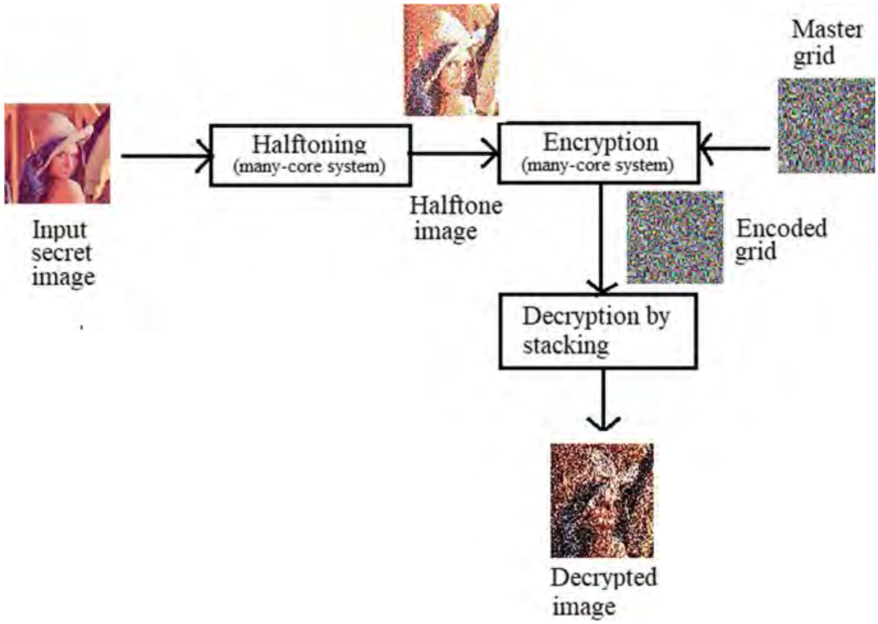


Fig. 4 Block diagram of the proposed system

improved speed. The generated encoded grid resulting from encryption is a share and hence shares generating process also refers to the encryption. Algorithm 1 on a host launches Algorithms 2 and 3 to perform halftoning and share generation on a many-core system. A host module is for the multicore system to execute. This module acts as a staging area where the necessary allocations, data transfers between multicore and many-core memories, thread creations and kernel invocation done. The CUDA API is used in the host module to perform such tasks. A device module in C language intended to execute on a many-core system. It is also called a kernel. It also uses CUDA primitives to obtain global unique thread indices. The global device memory is accessed using these indices. The number of pixels to be processed is high. So, the kernel is launched with multiple blocks and threads.

Algorithm 1 is the main host module. It relaunches the same kernels shown by Algorithms 2 and 3 for the grayscale and the colour images. These invocations are asynchronous and independent. The CUDA constant memory initialized in step 2 of Algorithm 1 with the constants that are reused by all threads. This memory facilitates efficient accessing by appropriate caching and broadcasting. The grayscale image gets processed from steps 3 to 16. Step 9 launches the kernel for generating the halftone image. It creates the number of threads equal to the height of the image. The rationale is to process each row of the image with a thread. The step 13 launches share generating kernel with the number of threads equal to the size of the image. The timer records execution times in these two stages separately. The colour image is handled from step 18 to step 33. The colour image is decomposed into C, M, Y components in step 18. Each component is converted to halftone image by reusing the same kernel in steps 22, 23 and 24. The steps 28, 29 and 30 re-invoke the kernel used for generating shares. The allocated device memory is released in step 34.

Algorithm 2 generates a halftone image. The loop in step 2, provided for a thread to iterate through each row of the image. It uses an error diffusion technique to generate halftone images. The process of generating shares given in Algorithm 3. Each thread processes a pixel by device memory read and writes. Step 1 in Algorithms 2 and 3 generates global thread indices.

6 Experimental Results

The proposed scheme implemented in CUDA with OpenCV, executed on a multi-core PARAM Shavak supercomputer with the Intel(R) Xeon(R)-E5-2670 CPU. This supercomputer contains a many-core Nvidia Tesla K40c GPGPU. Table 4 shows the hardware configuration. Many grayscale and colour images with various resolutions taken as test samples. Figure 5a shows an input grayscale image. Figure 5b shows the corresponding halftone image. Figure 5c, d are the random master grid and the

Table 4 PARAM Shavak supercomputer environment

Attributes	Multicore	Many-core
	Intel(R) Xeon(R)-E5-2670	Tesla K40c
Processing elements	2 CPUs each having 12 cores	2880 cores
Memory	8 TB	12 GB
Cache	30,720 KB	L1-64 KB, L2-1.5 MB
Clock speed	2.30 GHz	745 MHz

Algorithm 1: Allocate memory, perform data transfer, create threads, and launch kernel to generate a halftone image **(Host module)**

Input: The secret grayscale image $I1$ and colour image $I2$ of size $(h \times w)$.
Output: Reconstructed grayscale image $R1$ and colour image $R2$ of size $(h \times w)$

Declare device constant memory using $\_\_constant\_\_float\ r[4]$
Copy 7/16, 5/16, 3/16, 1/16 to the constant memory $r[4]$ using $cudaMemcpyToSymbol()$.
Use $cudaMalloc()$ to allocate device memory for $I1$ and $I2$.
Use $cudaMalloc()$ to allocate device memory for the halftone images HG, HC, HM, HY.
Use $cudaMalloc()$ to allocate device memory for the share images $SG1$,$SG2$, $SC1$,$SC2$, $SM1$, $SM2$, $SY1$,$SY2$.
Initialize $SG1$, $SC1$, $SM1$, and $SY1$ with 50% probability of binary values.
Set $timer = 0$
Use $cudaMemcpy()$ to transfer $I1$ to device memory d_I1.
Launch $Halftone\_kernel(d\_I1, HG, width, height)$ to generate halftone image with the number of threads equal to the $height$.
Use $cudaMemcpy()$ to transfer HG to CPU memory.
Store $timer$ value.
Set $timer = 0$.
Launch $Shares\_kernel(HG, SG1, SG2, size)$ with the number of threads equal to the $size$.
Use $cudaMemcpy()$ to transfer $SG1$ and $SG1$ to CPU memory.
Store $timer$ value.
Stack $SG1$, $SG2$ to output reconstructed grayscale image $R1$.
Set $timer = 0$.
Decompose $I2$ into $I2c$, $I2m$, and $I2y$.
Store $timer$ value.
Set $timer = 0$.
Use $cudaMemcpy()$ to transfer $I2c$, $I2m$, and $I2y$ to device memory d_I2c, d_I2m, and d_I2y respectively.
Launch $Halftone\_kernel(d\_I2c, HC, width, height)$ to generate halftone image with the number of threads equal to the $height$.
Launch $Halftone\_kernel(d\_I2m, HM, width, height)$ to generate halftone image with the number of threads equal to the $height$.
Launch $Halftone\_kernel(d\_I2y, HY, width, height)$ to generate halftone image with the number of threads equal to the $height$.
Use $cudaMemcpy()$ to transfer HC, HM, and HY to CPU memory.
Store $timer$ value.
Set $timer = 0$.
Launch $Shares\_kernel(HC, SC1, SC2, size)$ with the number of threads equal to the $size$.
Launch $Shares\_kernel(HM, SM1, SM2, size)$ with the number of threads equal to the $size$.
Launch $Shares\_kernel(HY, SY1, SY2, size)$ with the number of threads equal to the $size$.
Use $cudaMemcpy()$ to transfer $SC1$, $SC2$, $SM1$, $SM2$, $SY1$,and $SY2$ to CPU memory.
Store $timer$ value.
Stack $SC1$, $SC2$, $SM1$, $SM2$, $SY1$, $SY2$ to output reconstructed image $R2$
Use $cudaFree()$ function to free all allocated device memory.
End

Algorithm 2: $Halftone\_kernel(I, H, width, height)$: Generate halftone image in device memory **(Device module)**

Input: The secret grayscale image I, Device allocated halftone image H, $height$, $width$.
Output: Computed halftone image H.

$tid = blockIdx.x * blockDim.x + threadIdx.x$
For $j \in [0 \ldots, width - 1]$ do:
  if $(I[tid \times width + j] + err[tid \times width + j] < 128)$
   if $(I[tid \times width + j] + err[tid \times width + j] < 128)$
     $H[tid \times width + j] = 0$
   $else$
     $H[tid \times width + j] = 255$
  $diff = arr[tid \times width + j] + err[tid \times width + j] - crr[id \times width + j]$
  if $(j + 1 < width)$
   $err[tid \times w + j + 1] = err[tid \times w + j + 1] + diff \times r[0]$
  if $(tid < height - 1)$
   $err[tid \times width + j + 1] = err[tid \times width + j + 1] + diff \times r[1]$
  if $(tid < height - 1 \ and \ j - 1 \geq 0)$
   $err[tid \times width + j] = err[tid \times width + j] + diff \times r[2]$
  if $(tid < height - 1 \ and \ j + 1 < width)$
   $err[tid \times width + j + 2] = err[tid \times width + j + 2] + diff \times r[3]$
Return H

Algorithm 3: $Shares\_kernel(H, G1, G2, size)$: Generating shares for the halftone colour images **(Device module)**

Input: Halftone images H, Mastster grid $G1$, Allocated encoded grid $G2$, and $size$.
Output: Encoded grid $G2$.

$tid = blockIdx.x * blockDim.x + threadIdx.x$
while $(tid < size)$
  if $(H[tid] = 0)$
   $G2[tid] = G1[tid]$
  else
   $G2[tid] = \overline{G1[tid]}$
$tid = tid + blockDim.x \times gridDim.x$
Return $G2$ device global memory.

encoded random grid, respectively. They do not reveal the secrecy. Upon stacking these two grids, the resultant secret image is in Fig. 5e. Similarly, Fig. 6a shows a secret colour image. The derived C, M and Y decomposed images displayed in Fig. 6b, c and d, respectively. Figure 6e–g are their corresponding halftone images. Figure 7a, b, d, e, g, h, respectively, show the generated two shares corresponding to each halftone image. Figure 7c, f, i are the images upon stacking their respective shares. The overlaying of all shares reveal the secret image shown in Fig. 7j.

The achieved speedups with the proposed scheme is listed in Tables 5, 6, and 7. In all these tables, the execution time labels begin with the letter 'T'. They are in seconds. In Table 5, T1 is the execution time to generate the halftone image in the conventional scheme whereas T2 in the proposed scheme for the grayscale and

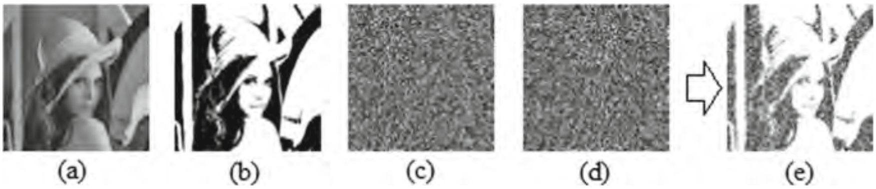


Fig. 5 **a** Sample grayscale test image Lenna (256 × 256). **b** Halftone image (256 × 256). **c** Master share (256 × 256). **d** Encoded share (256 × 256). **e** Reconstructed image (256 × 256)

Fig. 6 **a** Sample colour test image Lenna (256 × 256). **b** C component (256 × 256). **c** M component (256 × 256). **d** Y component (256 × 256). **e** Halftone of C component. **f** Halftone of M component. **g** Halftone of Y component

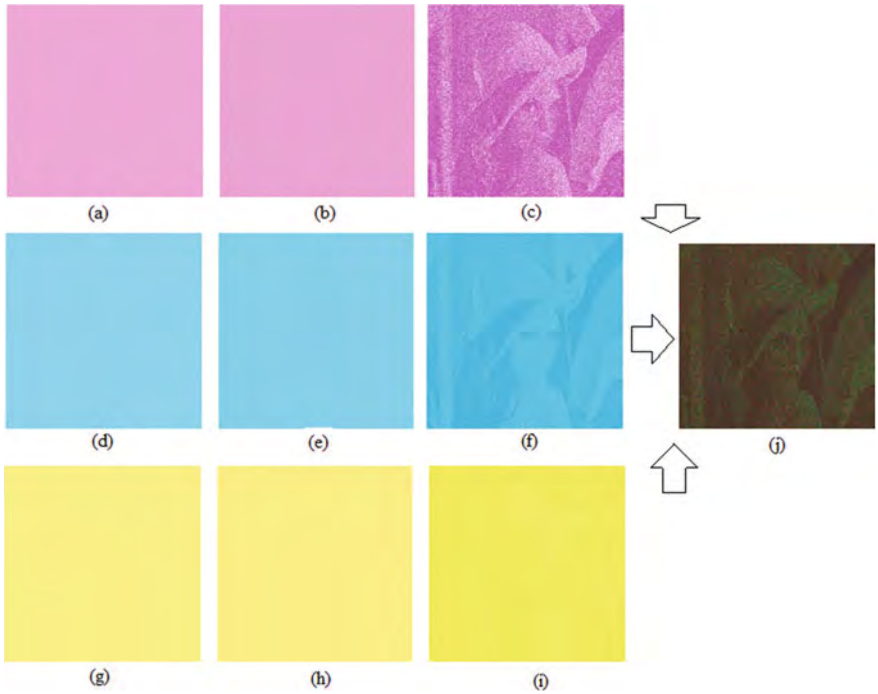


Fig. 7 **a** Master grid for C halftone. **b** Encoded grid for C halftone. **c** Stacking of C grids. **d** Master grid for M halftone. **e** Encoded grid for M halftone. **f** Stacking of M grids. **g** Master grid for Y halftone. **h** Encoded grid for Y halftone. **i** Stacking of Y grids. **j** Recovered image. *Note* Size of all images is (256 × 256)

Table 5 Speedup of the proposed system during halftoning

Image size	Grayscale image			Colour image		
	T1	T2	Speedup	T1	T2	Speedup
256 × 256	0.893	0.002984	299	2.5897	0.003753	690
512 × 512	3.26	0.003992	817	9.454	0.005621	1682
720 × 576	4.812	0.005021	958	13.9548	0.006953	2007
1024 × 1024	11.3	0.005798	1949	32.77	0.008975	3651

Note T1—time for halftoning (s), T2—time for generating share (s), and Speedup is unitless

the colour images. The speedup is the ratio of these two execution times. Figure 8 depicts the increase in these speedups as the image size grows. Table 6 lists the execution times for generating shares for the grayscale and the colour images in a sequential-RG and the proposed RG methods. The speedups recorded accordingly. Figure 9 shows the speedup in generating shares for the grayscale and colour images. The total speedup considering both the execution times for generating halftone and colour images in sequential and the proposed techniques shown in Table 7. The

Table 6 Speedup of the proposed system in generating share

Image size	Grayscale image			Colour image		
	T1	T2	Speedup	T1	T2	Speedup
256 × 256	0.327	0.00381	86	0.936	0.004932	190
512 × 512	1.436	0.00497	289	3.825	0.006031	634
720 × 576	2.386	0.00575	415	6.014	0.007641	787
1024 × 1024	5.163	0.00763	677	15.363	0.008932	1720

Note T1—time for halftoning (s), T2—time for generating share (s), and Speedup is unitless

Table 7 Total speedup of the proposed system

Image size	Grayscale image			Colour image		
	T1	T2	Speedup	T1	T2	Speedup
256 × 256	1.22	0.006794	180	3.5257	0.008685	406
512 × 512	4.696	0.008962	524	13.279	0.011652	1140
720 × 576	7.198	0.010771	668	19.9688	0.014594	1368
1024 × 1024	16.463	0.013428	1226	48.133	0.017907	2688

Note T1—time for halftoning (s), T2—time for generating share (s), and Speedup is unitless

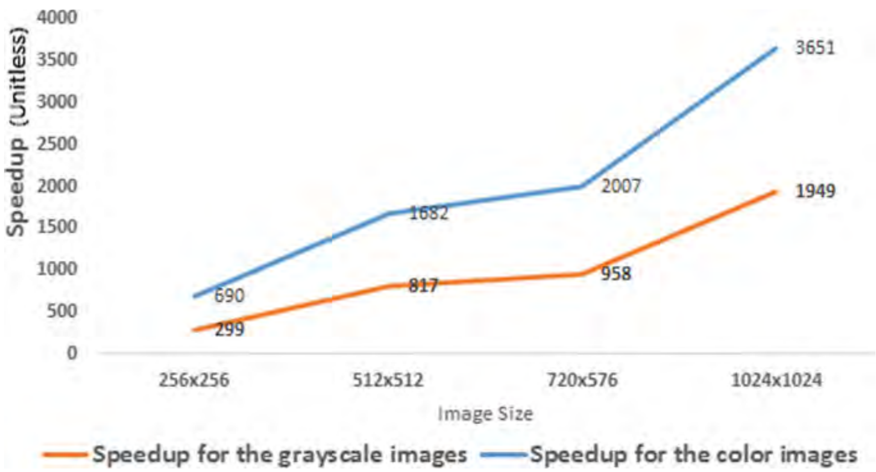


Fig. 8 Speedup in generating halftone images for grayscale and colour images

total speedup is displayed in Fig. 10. The execution time for generating the halftone images is considerably greater than that of generating the shares in a multicore random grid method. The number of memory accesses and computations is higher in obtaining the halftone images than the shares. Accordingly, the speedup achieved is also more in generating the halftone images than in producing the shares. The

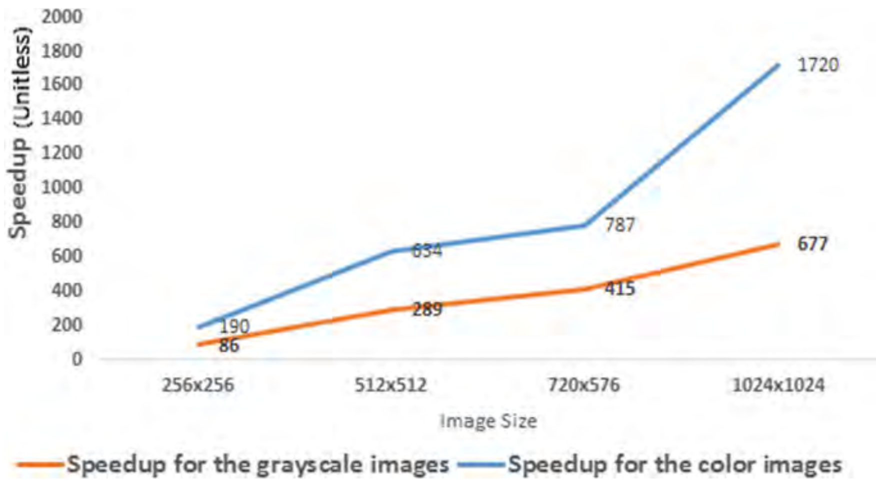


Fig. 9 Speedup in generating share images for grayscale and colour images

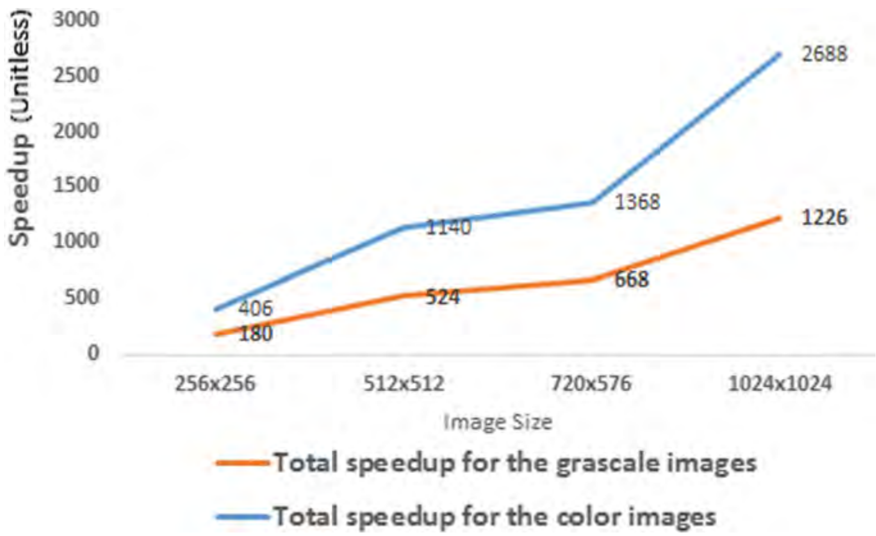


Fig. 10 Total speedup in generating halftone and share images for grayscale and colour images

experimental results reveal that the speedups achieved increase considerably as the image size grows. The performance of the proposed many-core random grid method is 406–2688 times better than the random grid in a multicore system.

7 Conclusion

This paper brings the efficiency considerations to the conventional VC to qualify it to the real-time applications. Less computation is very characteristic of any VC. However, due to the massive data to be processed, sequential VCs cannot be used in real-time applications. This work is significant in that it optimizes two stages of conventional random grid-based (2, 2) VC. These stages are generating halftone images and encrypted images using a many-core system. The use of constant memory in the kernel offers efficiency with the caching and broadcasting benefits. The reformulation of a solution to fit into a many-core system is a novel approach in the era of evolving many-core systems. The experimental results proved the efficiency of the proposed scheme over the traditional (2, 2) random grid method. The improved speedup is significant with the increased image resolutions. The proposed method outperforms 3651× and 1720× in generating halftone and share images, respectively, for a colour image size 1024 × 1024. For this image, the total performance gain is 2688× more over the conventional method.

The future work focuses on applying maturing CUDA catalysts to the proposed scheme and analysing the performance implications therein. It is possible to extend this work to the existing traditional VCs by considering other efficiency possibilities.

References

1. C.C. Chen, W.J. Wu, J.L. Chen, Highly efficient and secure multi-secret image sharing scheme. Multimed. Tools Appl. **75**(12), 7113–7128 (2016)
2. T.H. Chen, K.H. Tsao, Threshold visual secret sharing by random grids. J. Syst. Softw. **84**(7), 1197–1208 (2011)
3. T.H. Chen, C.S. Wu, Efficient multi-secret image sharing based on Boolean operations. Signal Process. **91**(1), 90–97 (2011)
4. P. D'Arco, R. De Prisco, Visual cryptography, in *International Conference for Information Technology and Communications* (Springer, 2016), pp. 20–39
5. K.M. Faraoun, Design of a new efficient and secure multi-secret images sharing scheme. Multimed. Tools Appl. **76**(5), 6247–6261 (2017)
6. T. Guo, F. Liu, C. Wu, Threshold visual secret sharing by random grids with improved contrast. J. Syst. Softw. **86**(8), 2094–2109 (2013)
7. R. Holla, N.C. Mhala, A.R. Pais, Gpgpu-based randomized visual secret sharing (grvss) for grayscale and colour images. Int. J. Comput. Appl. 1–9 (2020)
8. S. Kabirirad, Z. Eslami, Improvement of (n, n)-multi-secret image sharing schemes based on Boolean operations. J. Inform. Sec. Appl. **47**, 16–27 (2019)
9. O. Kafri, E. Keren, Encryption of pictures and shapes by random grids. Opt. Lett. **12**(6), 377–379 (1987)
10. S. Kukreja, G. Kasana, A secure reversible data hiding scheme for digital images using random grid visual secret sharing, in *2019 Amity International Conference on Artificial Intelligence (AICAI)* (IEEE, 2019), pp. 864–869
11. K.S. Lin, C.H. Lin, T.H. Chen, Distortionless visual multi-secret sharing based on random grid. Inf. Sci. **288**, 330–346 (2014)
12. F. Liu, C. Wu, L. Qian et al., Improving the visual quality of size invariant visual cryptography scheme. J. Visual Commun. Image Rep. **23**(2), 331–342 (2012)

13. D.C. Lou, H.H. Chen, H.C. Wu, C.S. Tsai, A novel authenticatable color visual secret sharing scheme using non-expanded meaningful shares. Displays **32**(3), 118–134 (2011)
14. S. Mittal, J.S. Vetter, A survey of CPU-GPU heterogeneous computing techniques. ACM Comput. Surv. (CSUR) **47**(4), 69 (2015)
15. M. Mohanty, W.T. Ooi, P.K. Atrey, Secret sharing approach for securing cloud-based pre-classification volume ray-casting. Multimed. Tools Appl. **75**(11), 6207–6235 (2016)
16. M. Naor, A. Shamir, Visual cryptography, in *Workshop on the Theory and Application of of Cryptographic Techniques* (Springer, 1994), pp. 1–12
17. S.J. Shyu, Image encryption by multiple random grids. Pattern Recogn. **42**(7), 1582–1596 (2009)
18. D. Suma, et al., Pipelined parallel rotational visual cryptography (pprvc), in *2019 International Conference on Communication and Signal Processing (ICCSP)* (IEEE, 2019), pp. 0109–0113
19. H. Wang, H. Peng, Y. Chang, D. Liang, A survey of GPU-based acceleration techniques in MRI reconstructions. Quant. Imaging Med. Surg. **8**(2), 196 (2018)
20. J. Wang, X. Chen, Y. Shi, Unconstraint optimal selection of side information for histogram shifting based reversible data hiding. IEEE Access (2019)
21. X. Wu, Z.R. Lai, Random grid based color visual cryptography scheme for black and white secret images with general access structures. Signal Process. Image Commun. **75**, 100–110 (2019)
22. X. Wu, W. Sun, Random grid-based visual secret sharing with abilities of or and xor decryptions. J. Visual Commun. Image Represent. **24**(1), 48–62 (2013)
23. B. Yan, Y. Xiang, G. Hua, Improving the visual quality of size-invariant visual cryptography for grayscale images: an analysis-by-synthesis (abs) approach. IEEE Trans. Image Process. **28**(2), 896–911 (2019)
24. X. Yan, Y. Lu, Generalized general access structure in secret image sharing. J. Visual Commun. Image Represent. **58**, 89–101 (2019)
25. Y. Yang, P. Xiang, M. Mantor, H. Zhou, CPU-assisted GPGPU on fused CPU-GPU architectures, in *IEEE International Symposium on High-Performance Comp Architecture* (IEEE, 2012), pp. 1–12

A Generic Framework for Change Detection on Surface Water Bodies Using Landsat Time Series Data

T. V. Bijeesh and K. N. Narasimhamurthy

Abstract Water is one of the important natural resources that requires immediate attention from a sustainability perspective. Depletion of surface water bodies due to various reasons has been remaining as a major concern for all growing cities across the globe. Change detection on the water bodies over the years can help the concerned authorities to implement strategies and solutions that can conserve our water bodies for future generation. This paper presents an image processing-based water body change detection method using Landsat multispectral images over the past 25 years. A hybrid level set-based segmentation algorithm is used for delineating the water bodies from the multispectral images. The surface area of the water bodies is then computed for the delineated water bodies, and a machine learning model is used for forecasting the future change from the past data. This paper also explores the possibility of building a dataset for training a deep learning-based image-to-image regression network that can forecast the shape and surface area of the water bodies.

Keywords Water body change detection · Level set-based segmentation · Image-to-image regression · Deep learning

1 Introduction

Water has always been the backbone of any civilization since time immemorial. Rapid urbanization has resulted in abrupt depletion of water bodies across all major cities in the world. Surface water bodies are the worst victims of urban developments and failure to focus immediate attention to the issue can have long-lasting effects on the future of our planet. Surface water bodies not only act as a direct source of water for the civilization, but also recharge the ground water. Disappearing and depleting water bodies hence lead to depletion in groundwater. Another important role the water bodies play is flood water mitigation and as our lakes and rivers are encroached in

T. V. Bijeesh (✉) · K. N. Narasimhamurthy
School of Engineering and Technology, CHRIST (Deemed to be University), Bangalore, India
e-mail: bijeesh.tv@christuniversity.in

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_26

the name of development, heavy floods are witnessed even with the moderate rain. Bangalore, a fast growing city in India, has witnessed this depletion of water bodies in an alarming rate. A recent study conducted by Down to Earth, a publication of the Center for Science and Environment, Bangalore's number of water bodies, has reduced by 79% in the last four decades. The main reasons for this alarming reduction are unplanned urbanization and population growth. South Africa's Cape Town had run out of drinking water and according to a BBC report, Bangalore is the next city in line. These facts prove that water body change detection studies are extremely important to analyze the surface water depletion and to subsequently device strategies to slow down this depletion thereby conserving them for the future generations.

Researchers have been working extensively on change detection studies using remote sensing images for the last few decades. Water body detection and delineation from the satellite images is the first step for change detection studies. Water body detection has been performed using techniques like spectral water indices, machine learning, spectral unmixing, and active contours. A comprehensive review on various methods of detecting and delineating water bodies from satellite images can be found in [1]. Water detection and delineation is the first step in performing change detection and temporal analysis on water bodies. Once water body delineation method is developed successfully, it is applied on multitemporal data to study the temporal changes happened to the water body over the years. Duan et al. presented a quadratic exponential smoothing-based lake area analysis using time series Landsat images. They forecasted future lake area based on the past area values without considering any external parameters [2]. The most frequently used data for change detection studies is the Landsat time series data. Researchers have already established the usefulness of Landsat time series data for performing change detection studies on water bodies [3]. A water mapping product, Water Observations from Space (WofS), was developed by Muller et al. that provides a continentally consistent model for studying and analyzing surface water bodies across the Australian subcontinent both spatially and temporally. Landsat data for the past 25 years were taken into account for developing this product. Decision tree and logistic regression were used for mapping water bodies using Landsat multispectral data [4]. An unsupervised change detection method was proposed by Acharya et al. based on spectral indices NDVI, NDWI, and MNDWI to detect the change of lakes in Pokhara city of Nepal using Landsat data of 25 years duration from 1988 to 2013. After preprocessing, a model was created in the ArcGIS for calculating the positive and negative change in the surface water of the lakes. For smaller lakes, the method was not as effective but for larger lakes the method proved to be very effective [5].

This paper proposes a generic change detection framework using level set theory-based water body delineation and deep learning-based image regression for forecasting the shape and surface area of water bodies from Landsat time series data for past 25 years.

Table 1 Landsat 7 and landsat 8 bands and details

	Landsat 5		Landsat 8	
Bands	Wavelength (μm)	Resolution (m)	Wavelength (μm)	Resolution (m)
Band 1	Blue (0.45–0.52)	30	Ultra Blue (0.435–0.451)	30
Band 2	Green (0.52–0.60)	30	Blue(0.452–0.512)	30
Band 3	Red (0.63–0.69)	30	Green(0.533–0.590)	30
Band 4	NIR (0.77–0.90)	30	Red(0.636–0.673)	30
Band 5	SWIR1 (1.55–1.75)	30	NIR(0.851–0.879)	30
Band 6	TIR (10.40–12.50)	30	SWIR1(1.566–1.651)	30
Band 7	SWIR2 (2.09–2.35)	30	SWIR2(2.107–2.294)	30
Band 8	Not applicable		PAN(0.503–0.676)	15
Band 9	Not applicable		Cirrus(1.363–1.384)	30
Band 10	Not applicable		TIRS1(10.60–11.19)	100 * (30)
Band 11	Not applicable		TIRS2(11.50–12.51)	100 * (30)

2 Materials and Methods

2.1 Dataset and Study Area

The proposed framework utilizes the freely available Landsat time series images for the past 25 years. The Landsat program is a series of Earth satellite missions operated collaborately NASA and United States Geological Survey. Researchers have been relying on Landsat images since 1972 to remotely study various natural resources and environmental phenomenon. In this work, time series data from Landsat 5 and Landsat 8 have been used to perform change detection on water bodies. The details of various EM bands at which images are acquired in sensors aboard Landsat 5 and Landsat 8 are presented in Table 1 [6].

Bands 1 and 9 are new addition in Landsat 8, and are, respectively, useful in costal/aerosol studies and cirrus cloud detection. Bands 10 and 11 are thermal bands and are used with greater precision for obtaining surface temperature information. Landsat 8 images are ideal for water-related studies because it consists of NIR and SWIR bands along with the visible bands that are utilized by most of the water

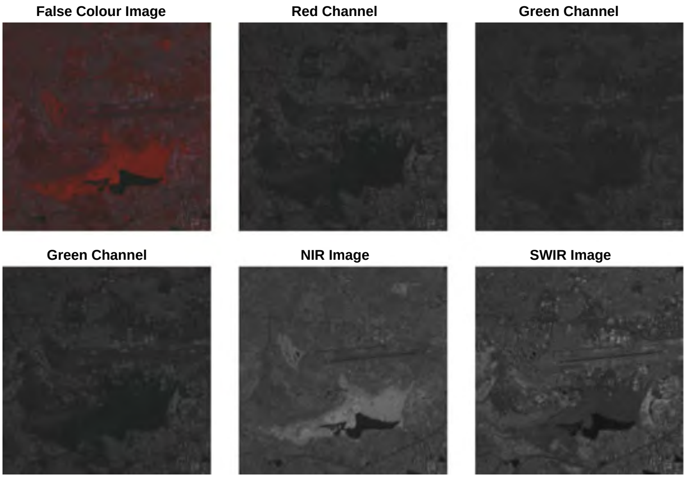


Fig. 1 Sample landsat image with important channels

detection algorithms. It also has a cirrus band for effectively detecting and removing cloud pixels from the image. Landsat 5 and Landsat 8 images are used for the study and are downloaded from the Earth Resource and Observation Science (EROS) Center-USGS Web site. Landsat 7 was initially considered but was not included because of the scan line error induced by the Scan Line Corrector (SLC) failure aboard the satellite. EROS runs the Landsat program together with NASA and maintains an immense archive of Earth's land surface satellite imagery. Implementation of all algorithms in the study is performed on a Linux computer with dual core Intel processor using MATLAB software. The study area chosen is Bellandur Lake of Bangalore, India, with latitude and longitude 12.9354°N, 77.6679°E and Varthur lake, Bangalore (12.9389°N, 77.7412°E). Landsat multispectral images from the year 1987 to the year 2019 are used to develop the change detection algorithm. The images used are of the same month of every year so that the changes in water body due to seasonal changes do not affect the analysis. Data for a few years are missing in the study as cloud-free image of the area under study was unavailable in the USGS portal. Figure 1 presents sample Landsat images in false color and red, green, blue, NIR and SWIR Channels. A false color image is an image created by using channels other than red, green, and blue and the image will look different from the corresponding RGB image. False color images are normally used to enhance or suppress some of the features in the image scene.

2.2 Preprocessing

It was established by researchers that near infrared (NIR) and shortwave infrared (SWIR) bands are the most suitable for water detection applications. This is due to the fact that water absorbs electromagnetic (EM) signals that come beyond NIR range in the EM spectrum [7]. The pixel values have to be normalized before applying level set algorithm because the satellite images are captured under varying lighting conditions and thus has variations in pixel intensity value of the same ground point based on these conditions. The proposed framework uses histogram equalization and normalized difference water index (NDWI) as a preprocessing step to normalize the pixel values so that the inconsistencies due to varying lighting conditions can be nullified. MNDWI is a spectral index that utilizes the green and SWIR bands of the Landsat image. The formula to compute NDWI is given in Eq. 1.

$$\text{MNDWI} = \frac{\text{Green} - \text{SWIR}}{\text{Green} + \text{SWIR}} \tag{1}$$

2.3 Level Set-Based Water Body Delineation

A level set is an implicit depiction of a curve. In level set segmentation, our level set is first assigned to the image arbitrarily, and then gradually evolve the level set according to an external force. Typically the force which controls the evolution of the level set is the curvature of the level set. It has been started by arbitrarily setting the contour C0 (called the initial contour) and evolve it according to a force function. An image property is used as the force function that drives the evolution of the level set which is normally the gradient of the image [9, 10].

If $\phi(x, y, t)$ is the level set function where t is the time parameter introduced to model the temporal behavior of the level set function, the evolution of the level set is governed by Eq. 2.

$$\frac{\partial \phi}{\partial t} = F\,|\nabla \phi| \tag{2}$$

The driving force F which is the curvature of the level set curve acts in a direction normal to that of the level set and is the gradient of the level set function.

To perform image segmentation using level set evolution method, a new term called edge stopping function is added to the level set evolution PDE. So the new PDE is given as Eq. 3.

$$\frac{\partial \phi}{\partial t} = g\,(x, y)\,F\,|\nabla \phi| \tag{3}$$

Here the function $g(x, y) = \left(1 + |\nabla f|^2\right)$, where f is the image, is called the edge stopping function. The value of this function becomes nearly equal to zero at the object boundaries because of the fact that the image gradient is relatively higher at the boundaries. Therefore, becomes zero at the boundaries and thus the level set evolution stops.

Based on active contour techniques, two models are obtained for boundary detection: edge-based model and region-based model. Edge-based methods usually use a measure of the changes across an edge, such as the gradient or other partial derivatives. Such methods utilize image gradient to construct an edge stopping function (ESF) to stop the contour evolution on the object boundaries. But edge-based method is not considered the best as the edges may not be sharp due to fading of ink or degradation which can prevent the gradient value at the edges from being a high value. Another disadvantage of these models is that they are very sensitive to the location of the initial contour. Edge-based models are said to have local segmentation property as they can segment the objects only if the initial contour is placed surrounding the object.

Region-based model does not work on the discontinuity in the image, rather it partitions the image into "objects" and "background" based on pixels intensity similarity. Region-based models utilize the statistical information inside and outside the contour to control the evolution, which are less sensitive to noise and have better performance for images with weak edges or without edges. This method is not very sensitive to the location of the initial contour and can detect interior and exterior boundaries at the same time. Therefore, region-based models are said to posses global segmentation property [11, 12]. The change detection framework proposed in this work employs a hybrid level set which is a combination of both edge-based and region-based techniques and hence has combined advantages of both. The final formulation of the PDE is given in Eq. 4. The terms μ and λ are weight terms and can be fine-tuned and customized based on the smoothness and texture properties of the image.

$$\frac{\partial \phi}{\partial t} = \mu \delta (\phi)\, div \left(\frac{\nabla \phi}{|\nabla \phi|} \right) + \lambda (u_0 - c_1)^2 + \lambda (u_0 - c_2)^2 + 1/1 + |\nabla \phi|^p;\ p > 1 \tag{4}$$

2.4 Multitemporal Analysis for Change Detection

The water body delineation algorithm discussed in the previous section was then applied on multitemporal Landsat data starting from 1985 to 2020 to obtain the changes on the lake. The surface area of the lake was computed from the delineated image for all these years to provide a quantitative measure of temporal changes happened to the lake. Simple machine learning-based time series forecasting can be used to provide a futuristic measure in the expected change in surface area. This paper also proposes to experiment the usefulness of image-to-image regression networks to forecast not only the change in surface area but also the shape of the lake in the

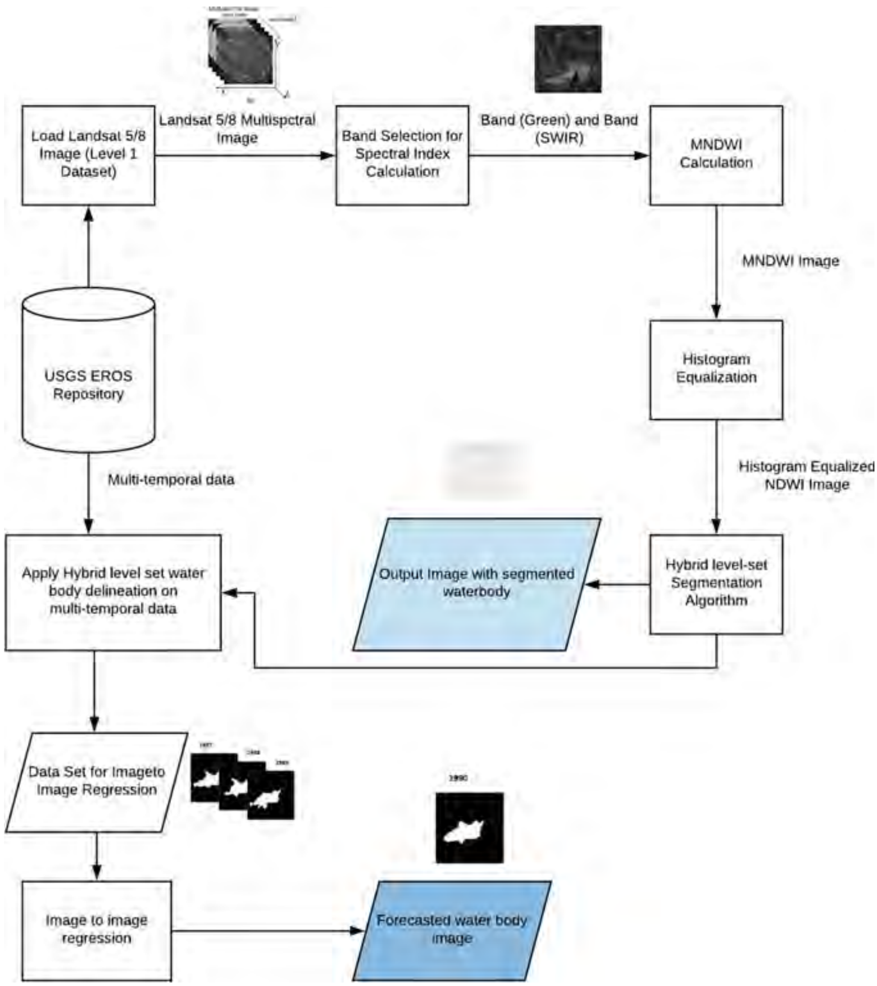
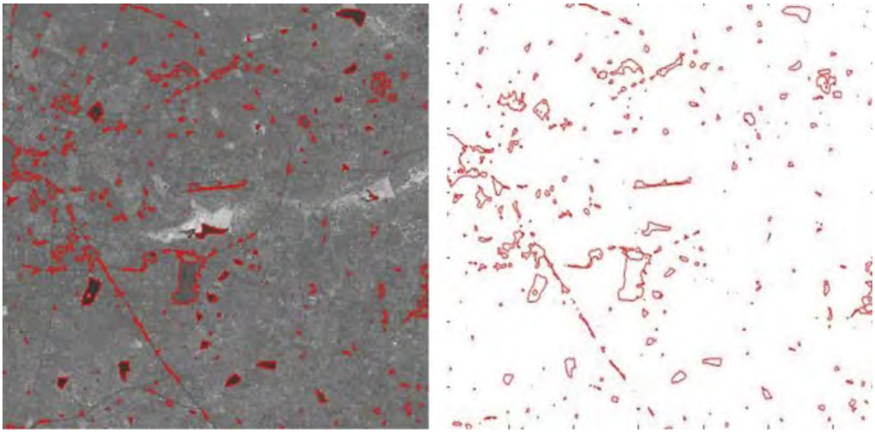


Fig. 2 Proposed generic change detection framework for surface water bodies

near future. Image-to-image regression has been widely used for image processing applications like denoising, age prediction, image generation, etc. [13–15], but is yet to be experimented for change detection using multitemporal multispectral images. The proposed framework for water body change detection is presented in Fig. 2.



(a) (b)

Fig. 3 Segmentation result for Bangalore area **a** contours juxtaposed on the SWIR image, **b** only the contour lines

3 Results and Discussions

A level set-based water body detection method is proposed and is applied on multitemporal Landsat data for change detection studies on lakes. Figure 3 shows the delineation output after applying the proposed delineation method on lake Bellandur, Bangalore. After some preprocessing on the obtained output and removing the unwanted contours, the binary output of the delineated water body is obtained as presented in Fig. 4.

The water body delineation algorithm is then applied on Landsat multispectral images of last 25 years to visualize the monitor the changes in surface area of the water bodies. The delineation output using the proposed algorithm for lake Bellandur and lake Varthur in Bangalore from the year 1987 to 2019 is presented in Fig. 5 and Fig. 6, respectively . Images for certain years may be missing in the output as cloud-free images of the study area for the year was not available in the USGS EROS portal. Bangalore is an Indian city that is growing drastically and it is evident from the results that the surface area of the lakes under study has come down steeply from 1987 to 2019 and will continue to deplete if appropriate measures are not taken by the authorities concerned.

This paper also proposes to apply the algorithm to multitemporal data for multiple lakes in Bangalore to prepare a dataset that can be used to train a deep learning image-to-image regression model that can forecast the future shape and the surface area of a given water body. The proposed image-to-image regression model takes an image cube as input where each channel corresponds to the delineated water body image for a given year and the output is a single channel image which is the forecasted

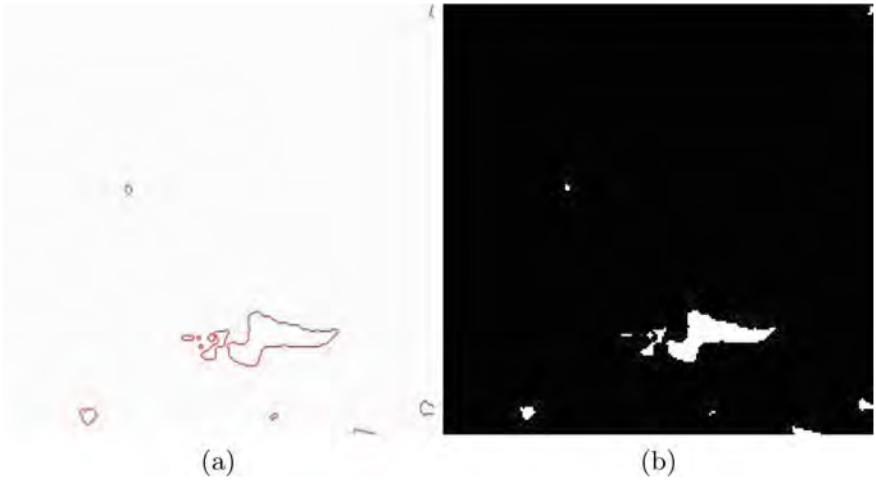


Fig. 4 Segmentation result for lake Bellandur, Bangalore **a** contour over the extracted water body. **b** Binarized output image where white pixels correspond to water pixels and black pixels correspond to non-water pixels

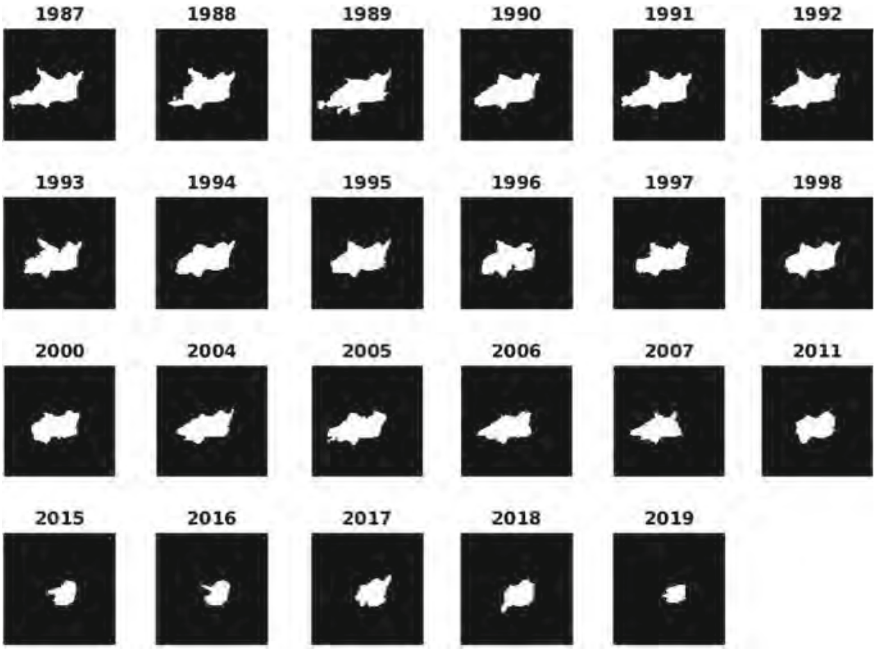


Fig. 5 Delineated output for lake Bellandur from the year 1989 to the year 2019

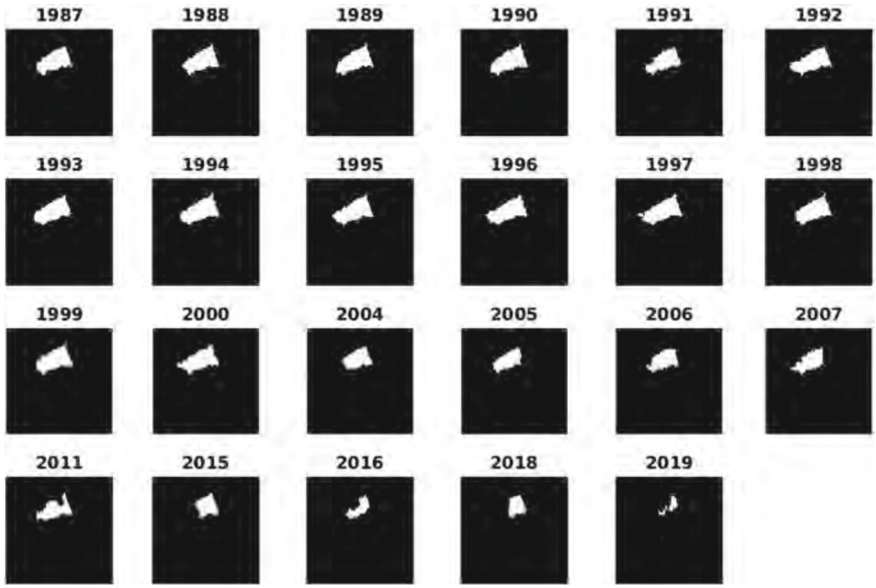


Fig. 6 Delineated output for lake Varthur from the year 1989 to the year 2019

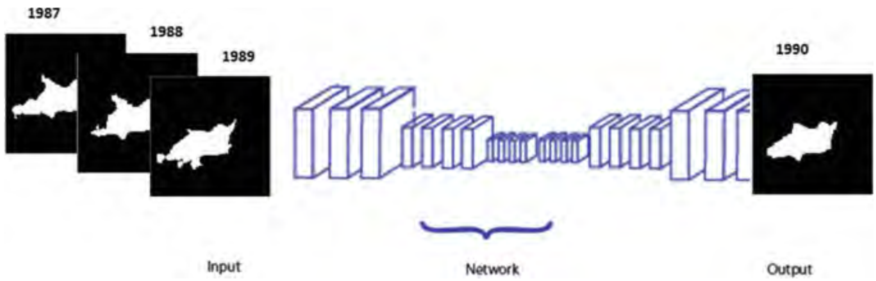


Fig. 7 Proposed image-to-image regression network that can be trained using past images to forecast future image of the water body

output. An abstract representation of the proposed image-to-image regression model is presented in Fig. 7.

Even though the framework proposed is generic in nature and can be applied on any water body, the major focus will be given to the lake in Bangalore in order to maintain the parameters like rainfall, snowfall constant. This research is still in the initial stage and once the dataset is prepared and the forecast model is built, it can help us convince the policymakers on the importance of taking important steps in conserving our fast depleting water bodies.

4 Conclusion and Future Scope

This paper presents a generic level set theory-based framework for surface water change detection studies using Landsat time series data the past 25 years. A level set-based segmentation was applied on multitemporal Landsat multispectral images of water bodies in Bangalore to visualize the changes happened to the water bodies in terms of surface area. The results clearly shows how fast the lakes under study have depleted in surface area thus emphasizing the need for taking measures to slow it down. This work has also proposed the possibility of building a deep learning image-to-image regression network that can be trained using the dataset prepared using the level set-based water delineation algorithm. The model can also be made robust by taking into consideration parameters like rainfall and temperature if the data can be collected for the last 25 years. The main challenge in building the model is the lack of available dataset and this work also proposes to prepare the dataset and the build the deep learning model.

Water is an essential natural resource that needs to be used wisely and conserved for the future generations. Many of the urban cities in the world are facing acute shortage in surface water and ground water. This research is an initial stage of developing a generic framework for fully automated monitoring and forecasting the changes happening to the surface water bodies.

References

1. T.V. Bijeesh, K.N. Narasimhamurthy, Surface water detection and delineation using remote sensing images: a review of methods and algorithms. Sustain. Water Res. Manage. **64**, 1–23 (2020)
2. Gonghao Duan, Ruiqing Niu, Lake area analysis using exponential smoothing model and long time-series landsat images in Wuhan, China. Sustainability **101**, 149 (2018)
3. W. Pervez et al., Landsat-8 operational land imager change detection analysis. Int. Arch. Photogram. Remote Sens. Spatial Inf. Sci. **42**, 607 (2017)
4. N. Mueller, et al.: Water observations from space: mapping surface water from 25 years of Landsat imagery across Australia. Remote Sens. Environ. **174**, 341–352 (2016)
5. T.D. Acharya, et al., Change detection of lakes in Pokhara, Nepal using landsat data. Multidiscipl. Digital Publ. Inst. Proc. **1**(2) (2016)
6. J.A. Barsi, et al., The spectral response of the landsat-8 operational land imager. Remote Sens. **6**, 10232–10251 (2014)
7. T. Lillesand, R.W. Kiefer, J. Chipman, *Remote Sensing and Image Interpretation* (Wiley, 2015)
8. H. Xu, Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int. J. Remote Sens. **2714**, 3025–3033 (2006)
9. M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models. Int. J. Comput. Vision **1**, 321–331 (1988)
10. S. Osher, J.A. Sethian, Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. J. Comput. Phys. **79**1, 12–49 (1988)
11. K. Zhang, et al., Active contours with selective local or global segmentation: a new formulation and level set method. Image Vision Comput. **28**(4), 668–676 (2010)
12. Tony F. Chan, Luminita A. Vese, Active contours without edges. IEEE Trans. Image Process. **10**(2), 266–277 (2001)

13. S. Pathan, Y. Hong, Predictive image regression for longitudinal studies with missing data (2018). arXiv:1808.07553
14. David Eigen, *Predicting Images using Convolutional Networks: Visual Scene Understanding with Pixel Maps* (New York University, Diss, 2015)
15. V. Santhanam, V.I. Morariu, L.S. Davis, Generalized deep image to image regression, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017)

A Machine Learning Approach to Detect Image Blurring

Himani Kohli, Parth Sagar, Atul Kumar Srivastava, Anuj Rani, and Manoj Kumar

Abstract The advent of smart mobile phones and cameras has unprecedently increased the number of photographs captured by people. Despite the hype, image blur remains as one of the distortions or quality degradation parameters, which can be raised by various parameters during the image capturing to processing phase. In the proposed paper, a machine learning-based detection technique is suggested to detect the image blur region and classify them into two categories. Image pre-processing has been applied to improve the textual information. Further, a pre-trained convolutional neural network model is used to classify the image blur. A fully automatic approach for blur detection as well classification is suggested in this paper. Three different image datasets are used to check the performance of the proposed method. Various parameters are estimated to demonstrate the results.

Keywords Blur detection · Convolution neural network (CNN) · Classification · Image blur · Laplacian enhancement

1 Introduction

Image blur arises from different natural photos due to camera shake, motion blur, defocus blur, artificial blur for highlighting important features as per the requirements. Thus, it is not required because it sometimes affects the important regions

H. Kohli (✉) · P. Sagar
Department of Computer Science, Amity University, Noida, India

A. K. Srivastava (✉)
Amity University, Tashkent, Uzbekistan
e-mail: aksrivastava1@amity.edu

A. Rani
Department of Computer Science, G L Bajaj Institute of Technology and Management, Greater Noida, India

M. Kumar
School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_27

and makes it less sharp, while in other cases, it is good to be blurred. Thus, people do it purposely to make photographs better by popping out the main content in the image and blurring the unimportant region. Many further estimated procedures have been there and tested, but with limited efforts need an enhancement in their algorithm. Understanding, blur desirability is not trivial but a challenging task to further distinguish the blur image and no-blur image. We can consider the image quality and give a scale of good, bad and ok blur type. The desirability depends on the image whether it needs to be blurred or needs to remove blur. Further, understanding the image in terms of high-level and low-level blur estimation, we proposed a novel approach training large-scale blur dataset consisting of more than 9500 images which is further categorized in natural, artificial and undistorted images. The dataset comprises of collection of different datasets which is trained and give result in terms of net accuracy.

Firstly, given an image, we pre-process the image using Laplacian operator and CLAHE. After pre-processing, our dataset is trained and evaluated using a sequential CNN model with max pooling and Swish which is end-to-end trainable network. The output is shown by giving accuracy of each training epoch and further end result is given by a net percent of accuracy. Finally, we compared the proposed technique performance with other techniques. To best of our knowledge, we designed this approach to detect the blur and enable to categorize the blur amount using blur estimation.

The proposed paper is categorized into five sections. Section 2 discusses relevant literature review followed by methodology in Sect. 3. Section 4 depicts demonstrated results and finally the conclusion of the paper is drawn in Sect. 5.

2 Related Work

In recent years, image blur detection using neural networks has proved their superiority among the conventional areas of research. Most previous work has paid attention on finding the blur in an image, as summing many users have the knowledge of blur category (good, ok, bad) or desirability criteria [1]. Different features have been used to estimate the blur amount, such as local filters [2], representations [3] and finding similarity among different neighbour elements [4]. It proposed SmartBlur dataset [5] consisting of more than 10,000 natural photos. It proposed an approach that is ABC-FuseNet to fuse low-level blur estimation and ResNet-50 is used to extract semantic feature. Evaluation has been compared with baseline methods giving as, 0.822, 0.616, 0.607, 0.785. Thus, composing a problem in two steps: generating blur responses and understanding by distilling high-level semantics failing accuracy in low-level semantics. Its limitation consists of methods that took more storage memory.

Whereas, in previous related research classification is done between two classes (blur and not blur), defining α-feature based on pixel's properties. Dataset used was DIAQ [6] and SmartDoc [7] thus consists of a drawback with image level blur desirability [8]. Other authors proposed blur quality measure for six collections, that is, Gblur, JPEG 2000 LIVE, TID2008, TID2013, IVC and CISQ by using histogram modelling [9]. In the research, it used SVM classifier and implemented SVM-RFE theory [10], includes statistic, texture [11], spectrum, low power spectrum features, for blur classification. It consists of dataset named Berkeley segmentation and Pascal VOC. This research handles complex problems but has a limitation of failing in contributing results for big sample data [12]. It proposed an approach that includes problem formulation and single-scale [13] and multi-scale deep blur detection. It consists of benchmark dataset proposed by [2, 14]. This research consists of a drawback of not considering the other features and only deals with multi-scale blur detection.

Author utilized deep learning algorithms, notably convolutional neural network to classify them either as blur or clear images. Classification procedures are done by extracting features of clear images in order to detect blur. This research has a limitation that it was failed in complex problems such as big sample dataset. In this research, comparison is drawn among three different image capture strategies for formulating the problem of motion de-blurring [15]. Experimental results are stimulated using high speed video camera. This research is sufficient for investigating the performance of a complex system but failed to investigate for one problem. Another research proposed a no-reference image blur assessment model that performs partial blur detection in the frequency domain [16]. Support vector machine (SVM) classifier is applied to the above features, serving as the image blur quality evaluator. It is using a medium size image dataset consisting of more than 2400 digital photographs. Hence, the author does not count this search for larger dataset casts as a limitation. In the recent approaches of year 2019, blur detection is labelled as an important feature of a digital image. A deep learning-based approach for detecting the blur in an image comprises of 250,000 identity images. They have done comparative analysis of our approach to statistical feature extractor, i.e. BRISQUE which was trained on SVM. Its limitations include the running time for a method applied on dataset [17]. In the research of 2020, the author proposed experimental results demonstrating the effectiveness of the proposed scheme and are compared to deterministic methods using the confusion matrix. It states an alternative method for Laplacian, and it has a limitation that it does not have a good result of classification of CNN [18, 19].

3 Methodology

In this section, the proposed methodology is discussed. Firstly, the pre-processing techniques to enhance image before classification is performed. Contrast limited adaptive histogram equalization (CLAHE) [20], Laplacian filter [21], and CNN model are designed, and dataset is trained and evaluated in this section. Figure 1 shows the proposed approach framework. It also shows the steps to encounter the blur image from the dataset. It formulates the steps to be followed in order to reach the

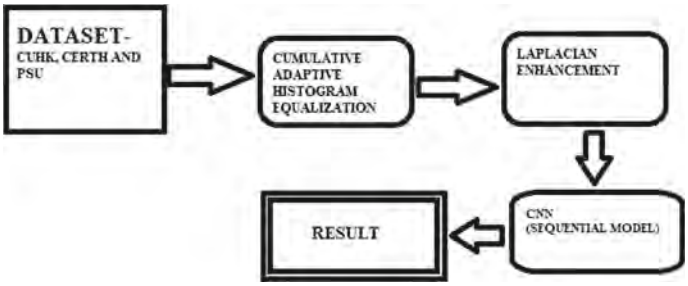


Fig. 1 Proposed framework to detect image blur

end result. In this figure, dataset is processed under CLAHE and Laplacian. Further, a model is designed to provide the detection of blur in images.

3.1 Image Pre-processing

Image pre-processing is the technique used to extract useful information from image in order to process results further.

Image pre-processing consists of the following steps:

a. Import the image which is input to the system
b. Further, analysing and manipulation of the input image
c. Output is the enhanced image after pre-processing techniques.

3.1.1 Proposed Contrast Limited Adaptive Histogram Equalization (CLAHE)

In the study, CLAHE has been used in order to overcome the noise amplification problem. In CLAHE, the contrast limiting procedure has to be applied to neighbouring pixels from which the transformation function is derived. Under the technique CLAHE, the input image is divided into sub-images, tiles and blocks. The CLAHE [20] consists of two parameters which are used to enhance image quality. The two parameters are block size (BS) and clip limit (CL). If CL is increased, the image gets brighter making histogram flatter. The dynamic range becomes larger when the BS is bigger. The CLAHE method consists of the following rules to enhance the original image which further goes to other pre-processing techniques:

a. The original intensity image is divided into non-overlapping contextual regions. The total number of sub-images is preferred to be 8×8, hereby, taking $P \times Q$.
b. According to the image level, calculating histogram equalizers for each region.
c. Using CL, further calculating the contrast histogram for each contextual region.

i. Taking CL value as:

$$Q_{avg}(Q_r\text{X} * Q_r\text{Y})/Q_{gray} \tag{1}$$

ii. where Q_{avg} is the average number of pixel, Q_{gray} is the number of grey levels and $Q_r\text{X}$ and $Q_r\text{Y}$ are the numbers of pixels in the X dimension and Y dimension of the contextual region.

d. The actual CL is expressed as:

$$\text{QCL} = Q_{\text{clip}} \times Q_{\text{avg}} \tag{2}$$

where QCL is the actual CL, Q_{clip} is the normalized CL under the range [0, 1]. If the number of corresponding pixels is greater than QCL, the pixels will be clipped.

e. The following are the rules for histogram equalizer:

$$\text{If } H_{\text{region}}(i) > \text{QCL then}$$

$$H_{\text{region\_clip}}(i) = QCL \tag{3}$$

$$\text{Else if } (H_{\text{region}}(i) + Q_{\text{avggray}}) > \text{QCL then} \tag{4}$$

$$H_{\text{region\_clip}}(i) = \text{QCL} \tag{5}$$

$$\text{Else } H_{region\_clip}(i) = H_{region} + \text{QCL} \tag{6}$$

where H_{region}(i) and $H_{\text{region\_clip}}$ (i) are original histogram and clipped histogram of each region.

f. Redistribute the remaining pixels, where, Q_{remain} is the remaining number of clipped pixels.

$$Q_{\text{gray}}/Q_{\text{remain}} \tag{7}$$

g. Enhancing intensity values in each region by Rayleigh transform.

h. Linear contrast stretch, given by,

$$X_{\text{new}} = \frac{X_{\text{input}-X_{\text{min}}}}{X_{\text{max}} - X_{\text{min}}} * 255 \tag{8}$$

When a bi or tri-modal of an image is distributed in a histogram, it stretches a certain value of histogram for increased enhancement in selected areas. Where, X_{min}

and $X_{\max}$ denote the minimum and maximum value of the transfer function, and X_{input} is the input value of transfer function.

3.1.2 Laplacian Operator

In order to detect edges in the image, we used Laplacian operator on the platform of Python. It is more efficient than Sobel and Kirsch as they are first derivative operator and Laplacian is second derivative operator. We worked with Laplacian operator because it gave more and more sharpen effect on images, therefore, we chose second derivative over first derivative. Laplacian operator is scalar differential operator for a scalar function $f(x, y)$, where x and y are spatial coordinates in a plane. Laplacian is used in digital image pre-processing to detect the edges to show the area of rapid intensity change.

$$\Delta f = \Delta^2 f = \Delta \cdot \Delta f$$

Laplacian takes out edges in the following classification:

a. Inward edges
b. Outward edge.

Laplacian consists of two types of operator that is:

1. Positive Laplacian operator

It consists of mask where corner elements are zero and centre elements are negative.

2. Negative Laplacian operator.
   It consists of mask where corner element is zero and centre element as positive and remaining should be -1.

3.2 Proposed Convolutional Neural Network (CNN)

In this research, we introduce the problem of automatically understanding image blur in terms of analysing image quality. The proposed CNN model is shown in Fig. 2.

Here, in Fig. 2, the proposed framework is consisting of steps involved in the convolutional neural network followed by input image and corresponding layers with max pooling and softmax at the end of the layer [22]. Further, Swish is used as an activation function proposed in corresponding layer of the network. This network is implemented using Python 3.7 with imported files of Keras and Sklearn Library. The model used is sequential model, where batch of images is further used as an input to the model for every training iteration. It is implemented with number of training epochs as 15. Our model consists of adaptive learning rate optimizer (ADAM), and the model has an activation function applied to the layers of network named as Swish,

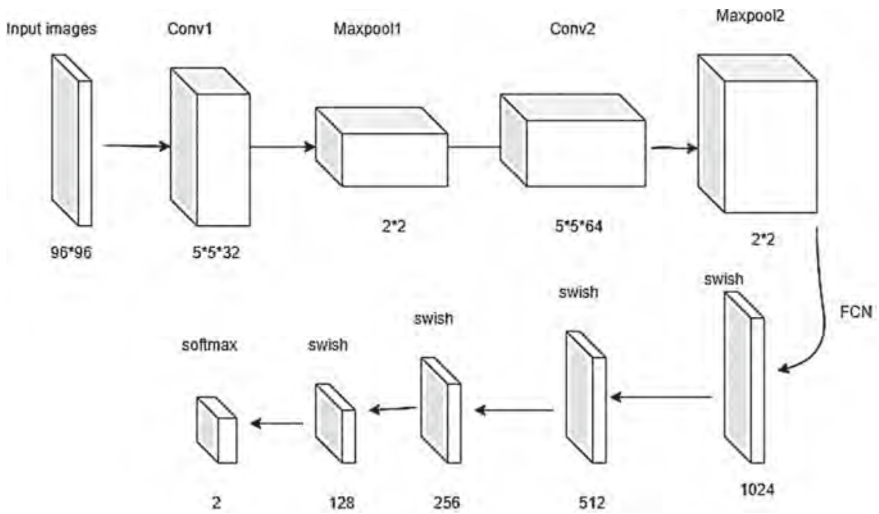


Fig. 2 Architecture of CNN model

an extension of sigmoid function. Swish is a better and more efficient function than ReLU as by replacing ReLU the model accuracy has been improved by 10%.

4 Experimental Results

Three data has been applied in our study for better results. The first one is blur image dataset granted an access from Portland State University [23], second is CUHK from The Chinese University of Hong Kong [24] and the third one is CERTH image blur dataset [25]. Our dataset consists of large-scale data with images having blur desirability with no annotations. It consists of 9500 images including testing and training dataset, which has been classified under three different categories as natural, artificial and undistorted blurred image dataset. The images in the dataset are challenging blur dataset are used for cross-dataset evaluation and further trained and tested in cross-dataset test.

The code runs effectively and reached the desired goal. The code gives the accuracy of blur detection based on image level blur detection with good image quality. Then, we measured the performance of the pre-processing to detect the blur images with threshold value,

$$\mathrm{V} = 450 \text{ for training set} \tag{9}$$

$$\mathrm{V} = 420 \text{ for testing dataset} \tag{10}$$

If the variance computed is less then threshold, then it is appended in the blur list, otherwise, it is appended in not blur list. It consists of blur label, if blur is equal to the value 1, then it is blur, otherwise, it is not blur. The following result is the evidence of the effectiveness of the program proposed, and the program run on the bucket of images to further provide the better result. Figure 3 shows pre-processing results.

In Fig. 3, it consists of results of pre-processing techniques, where an input in Fig. 3a is processed under two pre-processing techniques that is CLAHE (Fig. 3b) and Laplacian method (Fig. 3c). Here, (a), (b), (c) simplifies input, CLAHE and Laplacian methods, respectively.

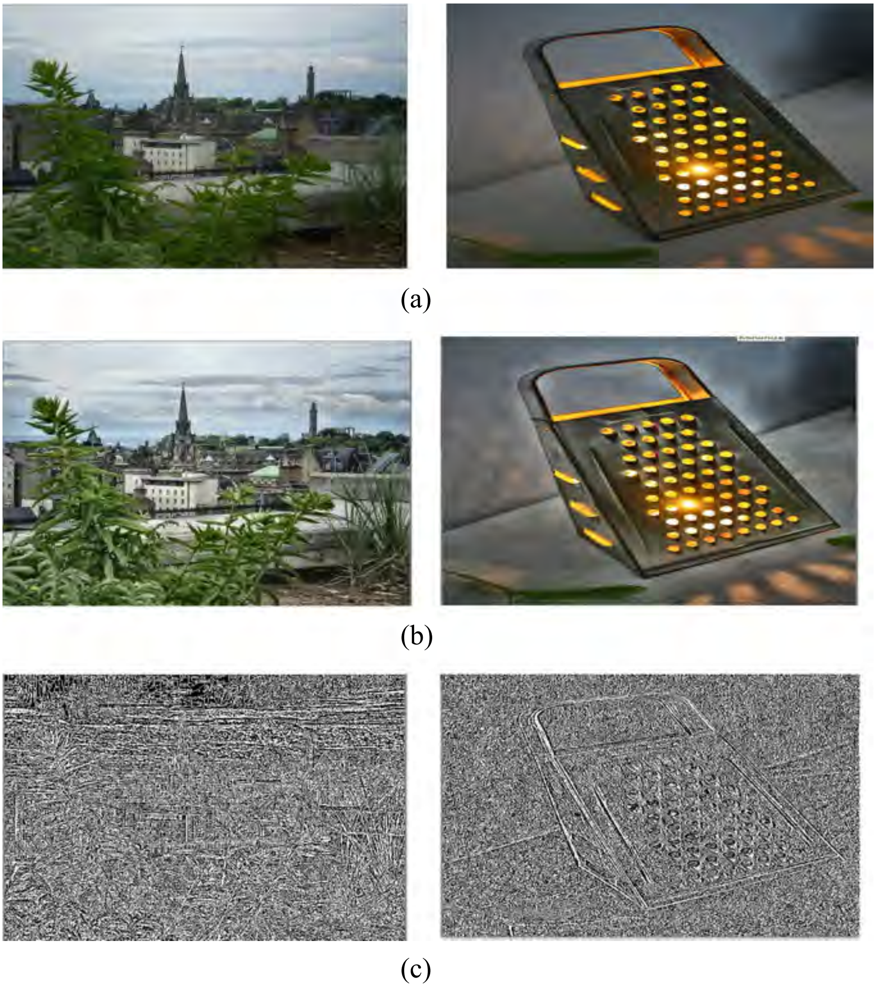


Fig. 3 Enlarged illustration of two challenging scenarios with pre-processing techniques **a** inputs. **b** CLAHE. **c** Laplacian operator

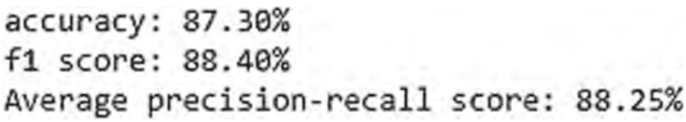


Fig. 4 Accuracy, F1and average precision-recall score for the model

Table 1 Image blur detection comparisons

Technique name (network + pre-processing + dataset)	Accuracy (%)
CNN + LAPLACIAN + CERTH	44.86
CNN + LAPLACIAN + CUHK	58.31
CNN + LAPLACIAN + CLAHE + CERTH + CUHK	65.88
Proposed CNN + LAPLACIAN + CLAHE + CERTH + CUHK	79.54

The accuracy is given as an end result for whole dataset; thus, it is the fraction of samples predicted correctly. For the following dataset, we computed its F1 score, which is the harmonic mean of precision and recall which is obtained from scikit-learn [26]. Figure 4 shows the estimated accuracy, F1 score and average precision recall for proposed blur detection technique on used datasets.

These are the techniques we implemented as a growth of our results from 44.86 to 79.54%. Where in first approach network used is CNN, pre-processing is Laplacian and dataset is CERTH. In second approach, network used is CNN, pre-processing is Laplacian and dataset is CUHK which is having high resolution images. In third approach, network used is CNN, pre-processing is Laplacian and dataset is created by including the images contained in CUHK, CERTH and CLAHE. In the last proposed sequential model of CNN, we enhanced the model by applying layers and max pooling which enabled an accuracy of 79.54%. The blur detection accuracy compares to other works is shown in Table 1. This depicts that the proposed method accuracy is more for classifying image blurring. Some of the detected image blur outputs can be seen from Table 2. Where it consists of two columns, input image and blurred image.

5 Conclusion

This work has been successfully implemented to detect the blur images. A machine learning-based approach is suggested by using CNN. The image blur region is detected and classified into two categories. After processing the image, the pre-trained neural network model is used to classify the images. Results are shown using accuracy parameters for each epoch on given datasets containing 8000 photos with elaborate human annotations. The results are tested on various datasets and show that the technique accuracy is satisfactory in terms of blur image detection. As an

Table 2 Image blur detection outputs

Input image	Outputs
	
	
	
	

evaluation parameters, accuracy, F1, average precision-recall score for the model are estimated. In the future, we can explore more techniques and giving more methodologies on removal of blur, where blurred images will be sharpened and smoothed to make it more comprehensive for human use.

References

1. S. Golestaneh, L. Karam, Spatially-varying blur detection based on multiscale fused and sorted transform coefficients of gradient magnitudes, in *IEEE Conference on Computer Vision and Pattern Recognition* (2017)
2. J. Shi, L. Xu, J. Jia, Discriminative blur detection features. in *IEEE Conference on Computer Vision and Pattern Recognition* (2014)
3. J. Shi, L. Xu, J. Jia, Just noticeable defocus blur detection and estimation, in *IEEE Conference on Computer Vision and Pattern Recognition* (2015)
4. C. Tang, C. Hou, Z. Song, Defocus map estimation from a single image via spectrum contrast. Opt. Lett. **38**(10) (2013)
5. S. Zhang, X. Shen, Z. Lin, R. Mech, J.P. Costeira, J.M. Moura, Learning to understand image blur, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018)

6. J. Kumar, P. Ye, D. Doermann, A dataset for quality assessment of camera captured documents images, in *The 2013 Camera-Baes Document Analysis and Recognition, in CBDAR'13* (Washington DS, USA, 2013)
7. J. Burie, J. Chazalon, M. Coustaty, S. Eskenazi, M. Luqman, M. Mehri, N. Nayef, J. Ogier, S. Prum, M. Rusinol, ICDAR 2015 competition on smartphone document capture and ocr(smartdoc), in *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on* (2015)
8. V.C. Kieu, F. Cloppet, N. Vincent, Adaptive fuzzy model for blur estimation on document images, in *Pattern Recognition Letters* (2017)
9. F. Kerouh, D. Ziou, A. Serir, Histogram modelling-based no reference blur quality measure, in *Signal Processing: Image Communication* (2018)
10. M. Atas, Y. Yardimci, A. Temizel, A new approach to aflatoxin detection in chilli pepper by machine vision. Comput. Electron Agric, (2012)
11. Y. Maret, F. Dufaux, T. Ebrahimi, Adaptive image replica detection based on support vector classifier, in *Signal Process, Image Communication* (2006)
12. R. Wang, W. Li, R. Li, L. Zhang, Automatic blur type classification via ensemble SVM, in *Signal Processing: Image Communication* (2018)
13. A. Krizhevsky, I. Sutskever, G. Hinton, Imagenet classification with deep convolutional neural network, in *Conference on Neural Information Processing Systems* (2012)
14. R. Huanga, W. Fenga, M. Fana, L. Wan, J. Sun, Multiscale blur detection by learning discriminative deep features, in *Neurocomputing* (2018)
15. J. Ma, X. Fan, S. X. Yang, X. Zhang, X. Zhu, Contrast limited adaptive histogram equalization based fusion for underwater image enhancement, pp. 1–27 (2017)
16. A. Agrawal, R. Raskar, Optimal single image capture for motion deblurring, in *Computer Vision and Pattern Recognition, 2009. CVPR 2009* (2009)
17. M. Kumar, M. Alshehri, R. AlGhamdi, P. Sharma, V. Deep, A DE-ANN inspired skin cancer detection approach using fuzzy c-means clustering. Mob. Netw. Appl. **25**, 1319–1329 (2020). https://doi.org/10.1007/s11036-020-01,550-2
18. Uniqtech, Understand the Softmax Function in Minutes, 31 Jan 2018. Available: https://medium.com/data-science-bootcamp/understand-the-softmax-function-in-minutes-f3a59641e86d
19. A. Long, Understanding data science classification metrics in Scikit-learn in Python, 6 Aug 2018. Available: https://towardsdatascience.com/understanding-data-science-classification-metrics-in-scikit-learn-in-python-3bc336865019
20. E. Mavridaki, V. Mezaris, No-reference blur assessment in natural images using Fourier transform and spatial pyramids, in *2014 IEEE International* (2014)
21. K. Khajuria, K. Mehrotra, M.K. Gupta, *Blur Detection in Identity Images Using Convolutional Neural Network* (IEEE, Shimla, India, 2019).
22. T. Szandała, Convolutional neural network for blur images detection as an alternative for Laplacian method. *Researchgate* (2020)
23. A. Aggarwal, M. Kumar, Image surface texture analysis and classification using deep learning. Multimed. Tools Appl. (2020). https://doi.org/10.1007/s11042-020-09520-2
24. C. Chen, Q. Yan, M. Li, J. Tong, Classification of blurred flowers using convolutional, in *Association for Computing Machinery* (China, 2019)
25. Techopedia, Convolutional Neural Network (CNN), 5 September 2018. Available: https://www.techopedia.com/definition/32731/convolutional-neural-network-cnn
26. E. Mavridaki, V. Mezaris, No-reference blur assessment in natural images using Fourier, in *Image Processing (ICIP), 2014 IEEE International* (2014)

Object Detection for Autonomous Vehicles Using Deep Learning Algorithm

E. J. Sai Pavan, P. Ramya, B. Valarmathi, T. Chellatamilan, and K. Santhi

Abstract Self-driving cars is recently gaining an increasing interest from the people across the globe. Over 33,000 Americans are killed in car accidents every year and lots of those accidents are often avoided by implementing the autonomous vehicle detection. Different ways are developed to manage and detect the road obstacles with the help of the techniques like machine learning and artificial intelligence. To resolve the issues associated with the existing vehicle detection like vehicle type recognition, low detection accuracy, and slow speed, many algorithms like the fast and faster region-based convolutional neural networks (RCNNs) are implemented but those were not supportive in real time because of the speed at which they compute and its two-step architecture with the faster RCNN, which is the enhanced version of RCNNs that runs at a speed of 7 frames per second. As it is observed that the CNN family has two steps (object detection and classification), which can reduce the response time in real time with good accuracy and high image resolution. So, the vehicle detection model like YOLOv2 and YOLOv3 is taken into consideration in this paper as they are very useful in real-time detection with a comparatively higher frame rate. As YOLO family of algorithms will mostly use the single step detection and classification. YOLO has an FPS rate of 45 which is pretty good in the real-time scenarios. We had an average of 90.4 using the taken algorithm for each image in this paper with a lower resolution image alone.

Keywords Vehicle detection · RCNN · YOLOv2 · YOLOv3 · Bounding boxes and anchor boxes · Image processing

E. J. Sai Pavan · P. Ramya · B. Valarmathi (✉)
Department of Software and Systems Engineering, School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India

T. Chellatamilan
Department of Computer Applications, School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India

K. Santhi
Department of Analytics, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_28

1 Introduction

Object detection is considered as the most important computer-based technical issue that deals with classifying and detecting the image of certain classes. Object detection before the introduction of deep learning was a level by level process, with the help of edge detection and feature extraction by using the existed implementations like HOG, which provides the edge direction as an image outline in the detection. But these images are then compared against the present templates usually at multi-scale to find and localize items present in the image.

Then, the deep learning with CNN comes into existence. This involves the two-step object detection algorithms that firstly identifies the regions which might potentially contain an object and then to independently classify each bounding box by using CNNs and region of interest (ROI) pooling layer that got added into the in fast RCNNs for the selection of the object location in the image followed by the detection and the use of anchor boxes in faster RCNNs. For the real-time object detection and classification, many one-step object detection architectures like YOLO family of algorithms, SSD have been proposed, as it uses the default boxes to calculate the scores. All these will help to detect and classify the object in a single stretch. This research work implements the YOLOv2 and YOLOv3 algorithms and provides the results.

1.1 Background

You only look once, or YOLO, is one of the fastest object detection algorithms right now. While this not the most accurate object detection algorithm, but it is considered as a very good option, when you need real-time detection, without losing time and accuracy. YOLOv2 has added additional features to the YOLO algorithm for improved performance and accuracy. Furthermore, YOLOv3 is the newest with more accuracy than the other two algorithms.

2 Literature Survey

This research work has been proposed to use 3DLiDAR sensors instead of cameras as they were more accurate and dependable for boundary detection, which is the most important for the autonomous vehicles. The whole process was proposed in four steps in the paper and was carried out for analysing different kinds of roads and conditions with different kinds of obstacles under the real-time environment and the accuracy of the outcomes of object detections with or without the obstacles have been proved to be really high [1].

In this, there were mainly three contributions to obtain the local position of the object from multiple fusion of cameras, mapping the multiple objects position form the cameras fusion and they proposed a method for curve and steep round using the information of the steering wheels angels from the autonomous vehicles and finding the blind spots in those areas. This will help in finding out the exact vehicle following coordinates for the autonomous vehicles even in the steep slopes [2].

A camera was fixed inside the car and the video was recorded, the obstacle present in the frame was taken as input for detection of the objects. The various steps for detection of objects were segmentation, where the area of interest was obtained from the frame and objects were highlighted. After this obstacle detection on different kinds of roads was performed, the results also show that by segmentation it can also identify almost all the shadows of the objects, unless they were difficult to distinguish as shadow on the roads when they were dark. Also, this algorithm was applied on X-rays, on skin colours detection, etc. [3].

In this paper, the proposed method of object detections by using the consecutive frames of a video from the cameras fixed in the vehicles. There were totally three steps in this method, first the edges of the objects were detected and then the association technique to compute edge points of consecutive images. So that we can get to know how far they were from the vehicle which was autonomously driven. And finally, symmetry was applied to find the region of interests and also the positions, coordinates of the objects were detected. Also, Adabor filter can also be used to differentiate between the vehicles and pedestrians which was not possible in symmetry technique. The results obtained were pretty good [4].

When the cameras were used for detection, it was only 2D so thus they introduced LiDAR provide scope for 3D detection, but the outcome of this process was really poor. So, in this paper, they proposed a method for the fusion of both camera and LiDAR. At the end of paper, it can be known that the accuracy of the model was increased that of compared with the CNN. The object detections which were not evident(or clear) in the normal detection methods were detected accurately by multimodal fusions. Though there were some disadvantages to this model, it does not weigh over the advantages [5].

The article proposed the on-road detection using SSD, but it has disadvantage of missing out of the small objects. In this paper, they were adjusting some of the properties responsible for those drawbacks and comparing the performance with other models. The problem can be overcome by data compensation and augmentation which would help to identify objects which would be in relatively small size. The fine-tuned SSD showed more accuracy and good performance than the basic SSD and YOLO [6].

The fusion of the sensors was taken, for taking input data, both cameras and LiDAR were arranged within the vehicle. The three-dimensional projections from LiDAR were onto four screens giving a different view of the data for better understanding. Also, HMVPN was proposed which was a convolution network as the architecture for the data. There were two convolution neural networks one for RGB and other for LiDAR projections and finally both would be integrated. The IOU was high than most of the faster RCNN [7].

This article has thrown light on the various hardware components, cameras or sensors, or other filters which were most affective and accurate for the autonomous vehicles. Also, the generic algorithm involved in the function of this was also given in different steps. Some of hardware and software components methods for were sensors, computers and OpenCV, CUDA, android, respectively. Also, the datasets for training and testing the vehicles were also mentioned which would be quiet helpful in the process [8].

In this paper, the considered detection methods are Sobel method and NGP method, and the results were obtained for them and compared to know the best of them. They have used CALTECH dataset and the whole test detection was done in phases hypothesis generation (HG) and hypothesis verification (HV). Generally, visible camera was used for detection of the object, but during the nights, infrared cameras which were present capture the obstacles on the way. Additional algorithms like SVM and HOG were also used in both methods for edge detecting [9].

For getting the better speed of matching image, the maximally stable external regions method was proposed in this article. To improve speed, some of the unwanted processes involved for region matching was removed like feature point detection. The MSER was combined with the vision-IMU-based obstacle detection and it decreased the space and time complexity in a road environment perception. Also, from results, it was noted that the method has faster processing speed with comparison with others mentioned in the article [10].

The object detection is carried in two stages, first, the image will be divided into many columns, each column was taken as “Stixel” and a net was constructed from them using neural network. In the next step, the adjustment of the results was done by using the neighbouring Stixels. In this, they used only the monocular camera but the result by using this method was better than the stereo cameras or sensors. This method was cost effective also [11].

The CAN Bus has the collection of the data from core control systems of the vehicles. In this paper, they have proposed a neural network with triplet loss network for CAN Bus anomaly detection. They have compared the detection rate with other two algorithms SVM and softmax in the same architecture of the network. In triplet loss function, three random samples were taken from the dataset. The performance was better in the triplet method than the others [12].

They have used both normal camera and thermal cameras for better results. Dataset was created by them, by recording the road drives in the real time, and tests were conducted on them for pedestrian detection. Also, there was tri-focal tensor, with the cameras from three different angles which would help in combining different features from any sources and also to get the performance validated. In this paper, they used HOG and CCF-based methods for detection results have shown and the performance was better of with the CCF than HOG [13].

The detection of the on-road obstacles was done by implementing the faster RCNN algorithm which was a deep learning technique in the GPU framework. In this article, they proposed for the detection of only the objects on the road, the rest on the footpaths and surrounding the roads were masked using the filters. This reduced the area of the detection and thus improved the speed to some extent. For the detected objected, only if the IOU was greater or equal to 0.5, then only the object was recognized inside the bounded rectangular box or else it was rejected. The dataset used for training was not of the Indian roads some of the vehicles (obstacles) which were mostly found in our country like autos could not be recognized. It was proposed in the future studies to improve that feature and make it fully adaptable to detect the Indian roads also [14].

In this paper, there were totally three kinds of obstacles that can be recognized pedestrians, cars and motorcycles. It carries out in two stages of detection for more accuracy. The camera for the car was fixed on the frontal bumper. They have carried out experiments on two different conditions under day light and night time. And finally calculated the average performance from those both. The steps for the obstacle detection were segmentation of the foreground, object detection and estimation of distance from image. The detection rate during day time was obviously more than the night time but the both the detection rates were more than the traditional HOG-based method [15].

For the object detection using the cameras, it was done in two ways using monocular vision-based method or multi-camera system based both of the methods have their disadvantages so another method stereo-vision based was proposed in the paper is better of all. First, the intersection of road was extracted to get all the possible position of the obstacles to be present in region from disparity image and then keeping track of those obstacles by using of the Kalman filter which helped to estimate their next position. The outcome of the results was good [16].

They proposed a hybrid net from SVM and CNN networking algorithms and tested on the datasets Caltech-101 and Caltech-pedestrian. They also took different hybrid neural net algorithms and tested on these datasets. Experimental results of both the datasets gave more recognition rate of the obstacles and less missing rate of the pedestrians. The LM-CNN-SVM methods improved the detection of objects in unsuitable conditions like eliminating the noise like shadows, darkness during the nights, etc. [17].

Along with the detection of vehicles and pedestrians, the autonomous vehicles have to know the location and distance from the stop line which were most essential to follow traffic rules on the roads. But it becomes almost impossible to have a sight of the stop signs once the vehicle reached near the board as the camera cannot capture them. So a method for detection was proposed in this paper. A computer vision algorithm was proposed for this to calculate the stop distance. The accuracy was really high and also the FP rate was also less. But this method proposed can be improvised by taking the input images into RGB instead of converting them grey scale reading as most of the stop sign will be in red colour. This property can be used to detect the signs where it was not clear and blurred most effectively [18].

The presented model YOLOv2 doesn't use the K-means algorithm but instead they modified it with the feature box. Then, the next step was the choice of the anchor boxes and then after to improve the performance of object detection algorithm they applied hard negative mining. The result of this was compared with the traditional YOLOv2 and faster RCNN and the proposed method was faster than both of them with moderate acceptance rate [19].

The main aim of this article was to get to know the uncertainties of the 3D LiDAR autonomous vehicles, and if it cannot detect the environment accurately, then it should alert the user about it and request them to drive by themselves in order to reduce any collisions. There were two uncertainties addressed one was epistemic, that occurs whenever there was any obstacle in the real environment which was not a part of training set. It should actively learn about the obstacle for the future references. The other was aleatoric uncertainty, the earlier can be due to lack of accuracy of detection but this one was influenced by distance of detection and occlusion. By feed-forwarding the data continuously, these uncertainties can be reduced, and the performance can also be improved [20]. The author proposed an improved procedure for handling the voluminous information. Spatiotemporal and video prediction were built using transfer learning techniques for forecasting metropolitan traffic was investigated in this paper [21].

As all the existing algorithms have their pros and cons, but the YOLO family has this one particular advantage on top of all, that is, the FPS. The average FPS for YOLO as mentioned till today is way greater that the other algorithms as it is a single stage detection algorithm. So, the YOLO algorithms can be used in the real-time detections and its accuracy is also not poor to use.

3 Dataset Description and Sample Data

The dataset COCO is used in the YOLO implementation by installing the YOLO weights file. Basically, YOLOv2 weights file is directly downloaded from the Google and preplaced in the dark flow which is provided by the YOLO developers.

COCO stands for common objects in context. So as the name suggests, the images in COCO dataset are taken out from everyday activities thus by attaching the description to the objects present in the scenes. Let us say we have to detect some object in an image, by looking at the photograph we can tell if the object really is present or not. However, it will be challenging and depends on environment where the photograph was shot without having any noise and bypassing of image.

COCO is resourcefulness to collect daily images, and the images that we see in our daily life and provides related information. Those images can contain multiple objects inside same image but each image should be labelled as a multiple object and separated. COCO dataset gives us the context and separation of the objects in the images. COCO dataset can be used by anybody to build an effective object detection model.

The coco.names files contain the list of object names the algorithm can detect and this list can be increased in the further algorithms. It is shown in Fig. 1.

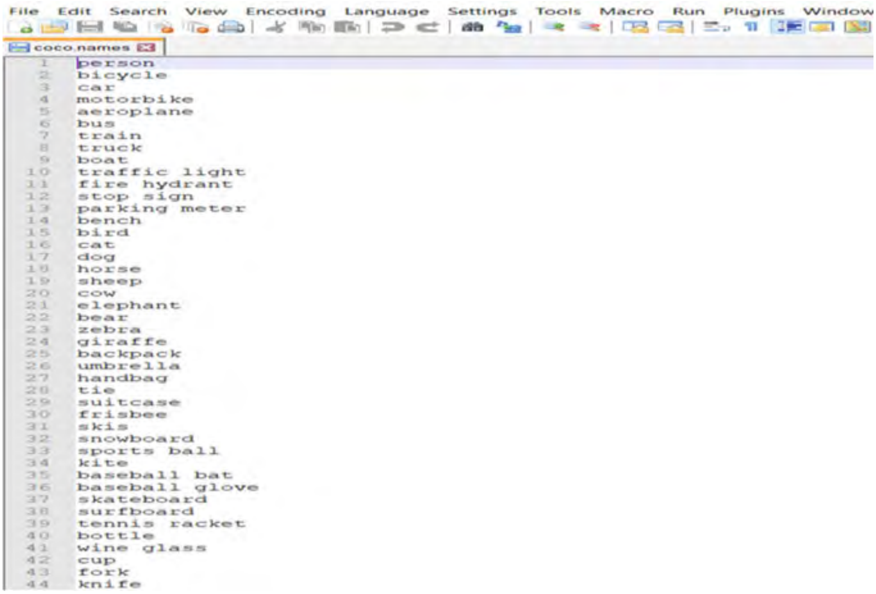


Fig. 1 Coco dataset file

4 Proposed Algorithm with Flow Chart

You only look once, or YOLO, is among the best object detection algorithms. Though it is not very highly accurate, but it is a very good to use this algorithm in real-time detection, as it takes very less time to detect without loss of too much accuracy.

The flowchart of YOLO mechanism describes the flow of and the step by step action carried out in the YOLO algorithms. As we can see that the centre point for the flow is the threshold value which asks the bounding box for an object in the image and it is shown in Fig. 2.

YOLOv2, one of the best algorithm at business, is more accurate and faster than the older versions. This is because YOLOv2 introduces the concepts like batch normalization and anchor boxes. Batch normalization (BN) normalizes the output of hidden layers which makes learning faster. Anchor boxes are assumed bounding boxes for image detection. 416 × 416 images are considered for training the detection and 13 × 13 feature map is the output. The size of the anchor boxes is pre-defined without having any prior knowledge. YOLOv2 internally takes help of k-means clustering algorithm as it provides us with good IOU scores.

YOLOv2s architecture is still behind in few of the metrices which are primary in many algorithms. No remaining blocks, no hop connections and mainly no up sampling. YOLOv3 provides all of these. We should an image of 416 × 416 and the output feature map would be again of size 13 × 13. One detection is done here by

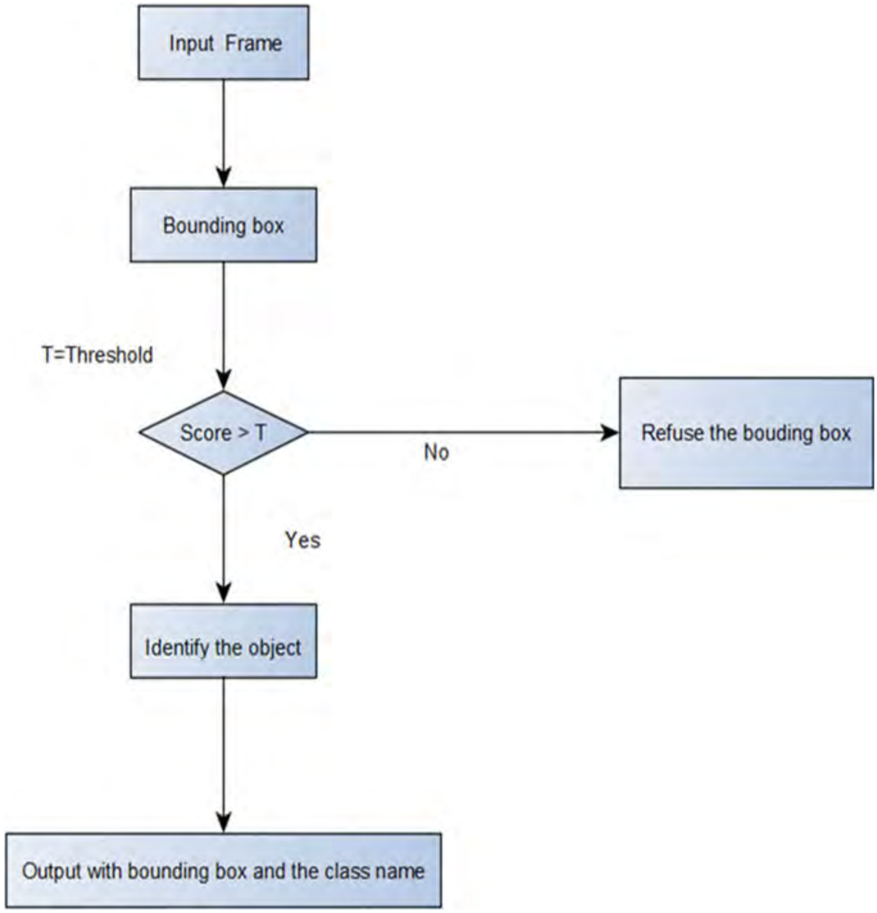


Fig. 2 Proposed algorithm flowchart

the help the 1 × 1 detection kernel, providing us with a result detection feature map of 13 × 13 × 255 (Fig. 3).

We can look at the architectural diagram for the YOLOv3 algorithm with 106 layers of convolution and the output with various sizes. The slowness of the YOLOv3 compared to YOLOv2 is because of the number of internal convolutional layers but a detailed increase in the accuracy and it is shown in Fig. 4.

The SSD, fast RCNN, faster RCNN and YOLOv3 algorithms are giving 22, 5, 17 and 75 FPS, respectively. The YOLOv3 is superior when compared to other existing algorithms like SSD, fast RCNN and faster RCNN. It is shown in Table 1.

The YOLOv3 Algorithm is provided with 106 layer of full convolution as an internal architecture. So, the reason for the low speed of YOLOv3 when compared to YOLOv2 is the huge number of fully connected layers in it. This is how the architecture of YOLOv3 now looks like. The 13 × 13 layer is used for finding larger

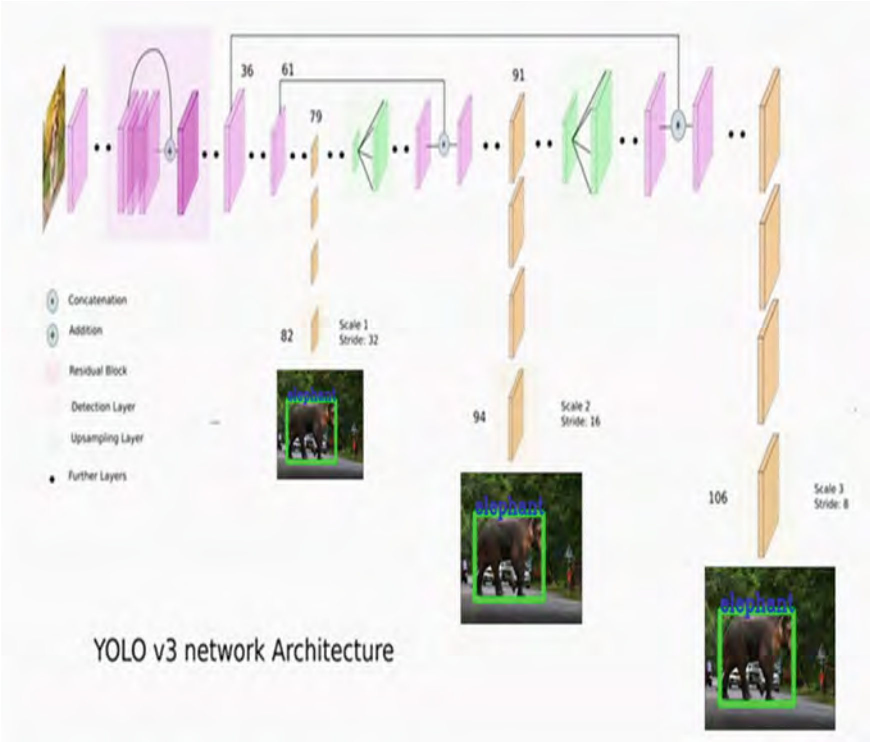


Fig. 3 YOLOv3 network architecture

Fig. 4 Experimental output

Table 1 Average FPS for an algorithm

Algorithm	Average FPS
YOLOv3	75 FPS
Faster RCNN	17 FPS
Fast RCNN	5 FPS
SSD	22 FPS

objects and 52 × 52 layer is responsible for detecting the smaller objects, finally with the 26 × 26 layer helping to detect medium objects. For an image of same size, YOLOv3 provides us with many bounding boxes than its older version YOLOv2.

5 Experimental Results

As you can see the output of an image where the elephant is absolutely detected with a proper bounding box and no overlap and the experimental output is shown in Fig. 4.

The FPS of the image depends on the GPU we use as the GPU component plays a major role in the frame rates image detection and the truck with 0.82 accuracy. It is shown in Fig. 5.

The accuracy on the car is 0.92 is seen at an average FPS 0.4. It is shown in Fig. 6.

In Fig. 7, this image, the accuracy on the person is 0.96. The above screenshots are some of the objects of the video which gave as input to the algorithm for the object detection and classification with bounding boxes around the objects along with the

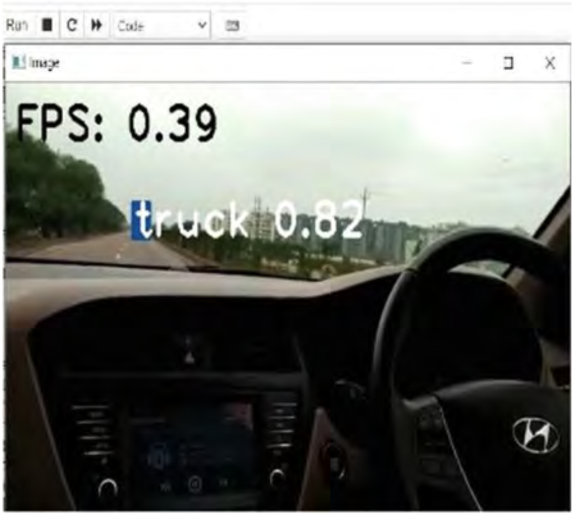


Fig. 5 Experimental result of the frame rates image detection and the truck with 0.82 accuracy

Fig. 6 Experimental result of the frame rates image detection and the car with 0.92 accuracy

Fig. 7 Experimental result of the frame rates image detection and the person with 0.96 accuracy

object classification name and accuracy. The accuracy is high and good for a real time with an average of 0.95 accuracy.

Since in major cases, the detection is done on the video files. So, every video file is cut down into frames and sent into the algorithm as an image on which the YOLO works to detect the objects in the frame. Time package is imported into the model and the start time of the video is noted and the video runs till the file run length. And, OpenCV is used to deal with the image processing. Cap.read and the BolbFromImage are used to capture frames and re-centre the size of the frame into the respective resolution. The FPS depends on the hardware mostly since the experiment is done on AMD 560X graphics card the frame rate to render and detect at the same time will be very low compared to the present hardware graphics cards like Nvidia thousand series, Amd Ryzen 5th series, etc.

6 Result and Discussion

The project implementation is based on the object detection using two algorithms YOLOv2 and YOLOv3 algorithms. Both the algorithms have some pre-trained models which are imported as weights during implementation. And threshold value would be set for both for classification of the objects. As we all know that the accuracy and threshold are always between zero and one. So, threshold is loaded into the model at runtime in this case.

When you feed any video to the YOLOv2, the runtime is comparatively less as that of YOLOv3 because it runs internally two convolutions, whereas YOLOv2 runs only once. But the accuracy of YOLOv3 is much more than that of YOLOv2. Two video files and two images were taken into consideration regarding the validation purpose. As we saw in the experimental results, the average accuracy of the objects in the image in YOLOv3 was 0.95 depending upon the threshold value and regarding the YOLOv2 the average accuracy was 0.85 and the time depends on the FPS rate. And the average FPS is 40 by the YOLO algorithms. The TensorFlow GPU versions only suits the NVidia graphic cards and allow to render the video at 30fps (frames per second). But since the experimental results are carried out on the AMD Radeon 560X, the frame rate is the only drawback with less than minimal (0.333 fps on average).

7 Conclusion and Future Work

After the implementation of both algorithms, it can clearly concluded that the quality of the videos used will affect the detection rate and accuracy of objects. And also the CPU version gives us less frames per seconds (FPS) as compared to the GPU implementation version of TensorFlow. So not only the algorithm but also the other factors like hardware, video quality can impact the accuracy of the output. However, in our implementation, YOLOv3 is most accurate but slower than the YOLOv2.

The future work will be to improve the object detection accuracy by using hybrid net and many more to work at the best as there is a risk of life or property in these conditions.

References

1. P. Sun, X. Zhao, Z. Xu, R. Wang, H. Min, A 3D LiDAR data-based dedicated road boundary detection algorithm for autonomous vehicles. IEEE Access **7**, 29623–29638 (2019)
2. J. Kim, Y. Koo, S. Kim, MOD. Multi-camera based local position estimation for moving objects detection, in *IEEE International Conference on Big Data and Smart Computing (BigComp)* (2018), pp. 642–643

3. L.A. Morales Rosales, I. Algredo, C.A. Hernandez , H.R. Rangel, M. Lobato, On-road obstacle detection video system for traffic accident prevention. J. Intell. Fuzzy Syst. **35**(1), 533–547 (2018)
4. K. Zebbara, M. El Ansari, A. Mazoul, H. Oudani, A fast road obstacle detection using association and symmetry recognition, in *2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS)* (2019), pp. 1–5
5. M. Person, M. Jensen, A.O. Smith, H. Gutierrez, Multimodal fusion object detection system for autonomous vehicles.J. Dyn. Syst. Meas. Control **141**(7) (2019)
6. H. Kim, Y. Lee, B. Yim, E. Park, H. Kim, On-road object detection using deep neural network, in*2016 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)* (2016), pp. 1–4
7. J. Zhao, X.N. Zhang, H. Gao, J. Yin, M. Zhou, C. Tan, Object detection based on hierarchical multi-view proposal network for autonomous driving, in*2018 international joint conference on neural networks (IJCNN)* (2018), pp. 1–6
8. S. Kato, E. Takeuchi, Y. Ishiguro, Y. Ninomiya, K. Takeda, T. Hamada, An open approach to autonomous vehicles. IEEE Micro **35**(6), 60–68 (2015)
9. J. Kim, S. Hong, J. Baek, E. Kim, H. Lee, Autonomous vehicle detection system using visible and infrared camera, in*2012 12th International Conference on Control, Automation and Systems* (2012), pp. 630–634
10. X. Yi, G. Song, T. Derong, G. Dong, S. Liang, W. Yuqiong, Fast road obstacle detection method based on maximally stable extremal regions. Int. J. Adv. Robot. Syst. **15**(1) (2018)
11. D. Levi, N. Garnett, E. Fetaya, I. Herzlyia, StixelNet: a deep convolutional network for obstacle detection and road segmentation. BMVC **1**(2) (2015)
12. A. Zhou, Z. Li, Y. Shen, Anomaly detection of CAN bus messages using a deep neural network for autonomous vehicles. Appl. Sci. **9**(15), 3174 (2019)
13. Z. Chen, X. Huang, Pedestrian detection for autonomous vehicle using multi-spectral cameras. IEEE Trans. Intel. Veh. **4**(2), 211–219 (2019)
14. G. Prabhakar, B. Kailath, S. Natarajan, R. Kumar, Obstacle detection and classification using deep learning for tracking in high-speed autonomous driving, in*2017 IEEE Region 10 Symposium TENSYMP* (2017), pp. 1–6
15. Y.W. Hsu, K.Q. Zhong, J.W. Perng, T.K. Yin, C.Y. Chen, Developing an on-road obstacle detection system using monovision, in *2018 International Conference on Image and Vision Computing New Zealand (IVCNZ)* (2018), pp. 1–9
16. Z. Khalid, M. Abdenbi, Stereo vision-based road obstacles detection, in *2013 8th International Conference on Intelligent Systems. Theories and Applications (SITA)* (2013), pp. 1–6
17. M. Masmoudi, H. Ghazzai, M. Frikha, Y. Massoud, Object detection learning techniques for autonomous vehicle applications, in *2019 IEEE International Conference of Vehicular Electronics and Safety (ICVES)* (2019), pp. 1–5
18. A. Arunmozhi, S. Gotadki, J. Park, U. Gosavi, Stop Sign and stop line detection and distance calculation for autonomous vehicle control, in *2018 IEEE International Conference on Electro/Information Technology (EIT)* (2018), pp. 0356–0361
19. Z. Yang, J. Li, H. Li, Real-time pedestrian and vehicle detection for autonomous driving, in*2018 IEEE Intelligent Vehicles Symposium (IV)* (2018), pp. 179–184
20. D. Feng, L. Rosenbaum, K. Dietmayer, Towards safe autonomous driving: capture uncertainty in the deep neural network for lidar 3d vehicle detection, in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)* (2018), pp. 3266–3273
21. T. Senthil Kumar, Video based traffic forecasting using convolution neural network model and transfer learning techniques. J. Innov. Image Process. 128–134 (2020)

CNN Approach for Dementia Detection Using Convolutional SLBT Feature Extraction Method

A. V. Ambili, A. V. Senthil Kumar, and Ibrahiem M. M. El Emary

Abstract Early detection of dementia remains as the best solution to prevent the development of various illness. It will help in providing an appropriate treatment to the affected individuals. So the early detection of dementia with higher accuracy is always essential for its treatment. MRI is the common biomarker for diagnosing the dementia. In this analysis, MRI image will be given as an input. The input image is pre-processed by using some morphological operations. This study introduces a convolutional shape local binary texture (CSLBT) for performing feature extraction to extract the key features from the magnetic resonance image. Linear discriminant analysis (LDA) is also used for performing feature reduction. Convolutional neural network can also be employed in the classification process.

Keywords Convolutional neural network · Dementia · Deep learning · Magnetic resonance image · Morphological operation · Shape local binary texture

1 Introduction

In the recent years, the prevalence of infectious diseases and chronic conditions has raised due to the demographic shifts in developed nations. According to the World Health Organization (WHO), approximately 47 million people are living with dementia worldwide, and this number is expected to increase to 82 million in 2030 and 150 million by 2050 [1]. Magnetic resonance imaging (MRI) is an inevitable factor in the medical field due to its high spatial resolution, soft tissue contrast, and non-invasive properties. MR imaging bestows valuable insights for prognosis and preparing brain tumor treatment [2]. To engender accurate brain images, MRI and computer tomography (CT) take advantage of computer technology. For analyzing

A. V. Ambili (✉) · A. V. Senthil Kumar
PG Research and Computer Application, Hindustan College of Arts and Science, Coimbatore, India

I. M. M. El Emary
Ph.D. Information Science Department, King Abdulaziz University Jeddah, Jeddah, Saudi Arabia
e-mail: omary57@kau.edu.sa

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_29

the structure of brain activity, MRI or CT scans are used. When compared to the CT scan, MRI scanning is more effective. There is no radiation in it [3]. In comparison with all the other imaging methods, [4] MRI is useful for the use of dementia diagnosis and identification because of its high contrast in soft tissue and its high spatial resolution, and because it produces no harmful radiation and is non-invasive [5–7]. Dementia is a brain disorder of memory, language, problem-solving, and other cognitive skills deteriorating. It incorporates the provision to do everyday tasks for a person. Different ways are available to detect dementia including brief examination, evaluation, challenging the patient's cognitive capacity, and so on [8]. The result of the analysis would also depend on the evaluator's knowledge and expertise. Also, there is a possibility of not getting a drawing test, depending on the person's characters and skills. To detect dementia quickly, such a dementia search can be done without usual constraints [9].

Deep learning (DL) is a machine learning system that has attracted tremendous interest from investigators, smashing criteria in fields like dementia tracking. The potential for efficient representation of original data by coherent nonlinear transformations increases the sophistication and complexity in DL, which varies from traditional machine learning methods [10]. Therefore, the early diagnosis of Alzheimer's disease (AD) can be naturally modeled to be a multiclass classification problem.

This research work suggests a CNN classifier for dementia classification. MRI input image is first pre-processed to reduce the noise present in the image. The pre-processed image is subjected to the morphological operation, which helps extract the brain image boundary areas. Features are extracted using GLCM, statistical, and convolutional SLBT methods. The convolutional SLBT is a combination of convolutional LBP and SLBT [11]. The extracted features are fed to CNN for the classification of dementia.

The paper is depicted in the following manner; Sect. 2 gives a literature survey of existing works. Section 3 explains the architecture of the proposed work. Results and discussions are covered in Sect. 4, and the work is concluded in Sect. 5.

2 Literature Survey

This segment discusses the research works on numerous current dementia diagnosis strategies. These research papers are taken and reviewed according to the recently research works on dementia detection techniques. Texture and shape analysis was extensively studied in recent years, and studies have discovered advanced methodologies for the classification of dementia.

Kawanishi et al. [9] had discussed an anomaly detection method for dementia detection. This system focuses drawing test on a tablet to detect dementia. MRI is an essential tool for brain disease detection [2, 3, 5]. Zhang et al. [12] developed a landmark-based feature extraction method using longitudinal structural MRI. It improves the classification efficiency and eliminates the time-consuming steps. Pachauri et al. discuss the topology-based core algorithm for cortical thickness measures [13]. It performs well on many statistical inference tasks without selecting features, reducing the dimension, or tuning parameters. In [14], a global grading biomarker is used to predict mild cognitive impairment (MCI) to Alzheimer's disease conversion based on registration accuracy, age correction, feature selection, and training data selection. This somehow provides a more accurate forecast of the conversion from MCI to AD. Zhou et al. [15] show that when combined with MRI measures, MMSE enhances the prediction performance. MMSE is the most significant statistical variable. It uses the SVM classifier for classification based on twofold cross-validation [16] and proposes a multimodal behavioral variant FTD (bvFTD) classification model to differentiate the pre-symptomatic FTD mutation carriers from healthy controls. Classification performance is calculated using receiver operating characteristic curves. It has a higher performance in classification. In shape and texture feature extraction method, it is not sufficient for extracting sharpen the image. So a combined feature extraction method is used to enhance the classification task.

2.1 Comparitive Analysis

Paper title	Methodology	Limitations
Kawanishi et al. [9]	Unsupervised anomaly detection method	Limitation of test contents
Ehab et al. [5]	A second decision for the surgeons and radiologists is proposed	Limited data base
Vieira et al. [10]	Neuroimaging correlates of psychiatric and neurological disorders	Statistical inferences are drawn from multiple independent comparisons based on the assumption that different brain regions act independently
Studholme et al. [20]	Template-based approach	Larger set of data are not considered for predicting the disease conditions
Ahmed et al. [1]	Neuroimaging and machine learning for dementia diagnosis	limited prodromal stage of dementia with Lewy bodies

3 Proposed Methodology

This study's sole objective is to design and develop a dementia detection framework based on MRI images.

Dataset

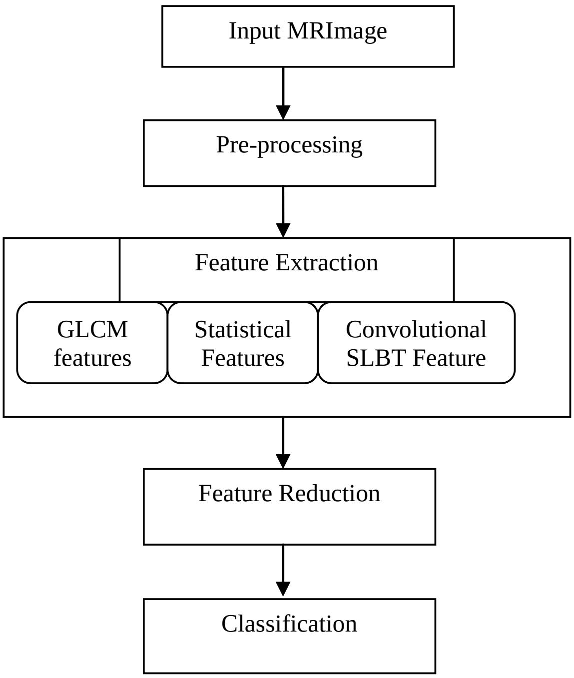
The datasets are collected UCI repository from https://www.nitrc.org/frs/?group_id=48 and https://www.smir.ch/#featured.

Here, images are taken as the base of brain image, with the complete structure of cerebrum and medulla oblongata, focusing as an input structure.

The overall procedure of the proposed approach involves pre-processing, feature extraction, feature reduction, and classification. Initially, the input image will be subjected to pre-processing for removing the noise in the image. The feature extraction will then be performed on the pre-processed image using the proposed convolutional shape local binary texture feature, gray level co-occurrence matrix (GLCM) feature, and the statistical quality like mean, variance, standard deviation, kurtosis, and the entropy. This paper proposes a new convolutional SLBT for feature extraction. The convolutional SLBPT feature will be designed by combining SLBT (Shape Local Binary Texture) [11] and convolutional LBP in which the LBP will be modified based on convolution.

At last, for the diagnosis of dementia, the classification will be carried out based on CNN. The proposed approach will be in MATLAB and the dataset that will be employed UCI repository. The efficiency of the proposed method will be assessed using measures like sensitivity, specificity, and accuracy. Figure 1 shows the flow diagram of the proposed system (Fig. 2).

Fig. 1 Flow diagram of proposed dementia detection using CNN



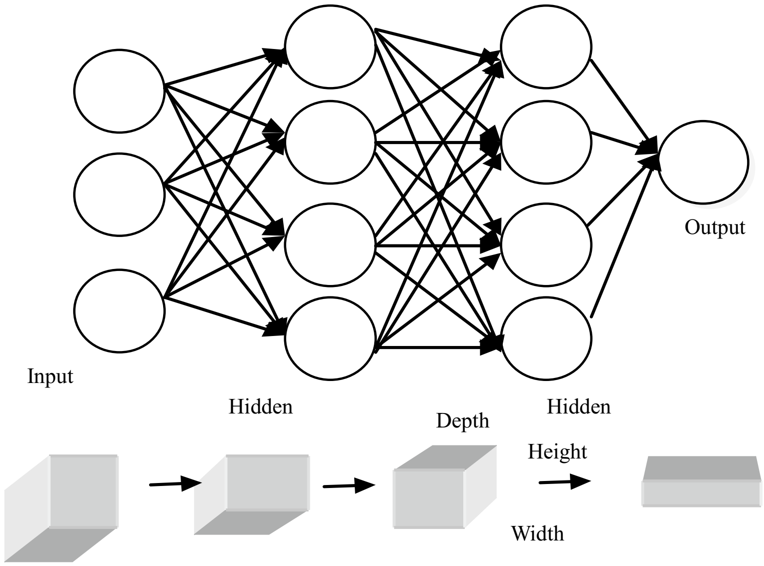


Fig. 2 Convolutional neural network architecture

3.1 Pre-processing

Pre-processing has a vital role to play in the segmentation of MRI brain image. The main goal of the pre-processing stage is the image quality and the reduction of noise. Pre-processing steps improves the quality of the MR image. The pre-processing step eliminates the unwanted noise and aversive parts of the image.

3.2 Morphological Operation

Imperfections are quite common in images. These imperfections in the images can be eradicated through morphological operations—the material modifications made on the pixels concerning the value of adjacent pixels. Dilation, erosion, opening, and closing are four elementary or fundamental morphological operations. Only dilatation and erosion are carried out in work mentioned. In erosion operation, every pixel that touches background pixels is transformed into background pixels, and the objects are converted into smaller pixels by dividing a single object into multiple objects. As against the erosion in dilation, many objects are merged into a single object [18].

3.3 Feature Extraction

The proposed work uses an MRI image for the process of identifying dementia. The pertinent and significant features from the image rendering the data provided as input into a less dimensional space are extracted, which is the main feature extraction objective.

3.4 GLCM

GLCM is the most popular statistical textural feature extraction method. The feature is derived from the gray-level covariance matrix. GLCM is a textural feature extracting strategy. Textural characteristics show a connection between the pixel values [19]. It is a second-order statistical method. The GLCM image has the same number of columns and rows as the gray forms. This gives an idea about how the co-occurring values are distributed. GLCM stands for image distance and orientation. The GLCM calculates the grayscale pixel value x that appeared in the pixel intensity y image either horizontally, vertically, or diagonally. GLCM provides gray-level data pairs. It comprises different varieties of gray values. It gives the disparity in the picture by statistical analysis compared to the original image.

$$\text{Contrast} = \sum_{x,y} (x - y)^2 p_{c,\o}(x, y) \tag{1}$$

$p_{c,\o}$ is the co-occurrence matrix and x and y are pixel intensities.

Correlation is the similarity degree of the pixel over the entire image with its neighborhood.

$$\text{correlation} = \sum_{x,y} \frac{(x - \mu_{i)}(y - \mu_{j)p_{c,\o}}(x, y)}{\sigma_i \sigma_j} \tag{2}$$

$\mu_{i \text{ and } j}$ are the means of two correlated images σ_i and σ_j are the standard deviations

$$\text{Energy} = \sum_{x,y} (p_{c,\o}(x, y))^2 \tag{3}$$

3.5 Statistical Features

Statistical features are commonly used for the statistical analysis of the image,

Mean: mean is the average value of pixels.

$$M = \frac{1}{AB} \sum_{x,y} p(x, y) \tag{4}$$

Let $p(x, y)$ is the grayscale at the point (x, y) of the picture dimension A × B.

Variance: Variance measures the pixel value distribution. The variance gives a calculation of the difference between each pixel and the average score. It is given as the average of the square of the difference between a mean and individual pixel.

$$\text{variance} = \frac{1}{AB} \sum\nolimits_{x \cdot y} p(x \cdot y) - M)^2 \tag{5}$$

Standard deviation (σ) defines the spread of the gray level around the mean. It is the square root of variance.

$$\sigma = \frac{1}{AB} \sum\nolimits_{x,y} (p(x, y) - M \tag{6}$$

Kurtosis is the measure of tailedness of an image

$$K_{tos} = \frac{1}{AB} \frac{\sum(p(x, y) - M^4}{\sigma^4} \tag{7}$$

Entropy: Entropy is the measure of dissimilarity in the MRI image. Entropy is the statistical measure of randomness that will characterize the texture of an image.

$$\text{Entropy} = \sum\nolimits_{x,y} p_{d,\emptyset}(x, y) \log_2 [p_{d,\emptyset}(x, y)] \tag{8}$$

3.6 Convolutional SLBT (Convolutional Shape Local Binary Texture)

The proposed feature extraction method combines shape local binary texture method and convolutional LBP in which the convolution operation is applied on LBP. The local binary pattern (LBP) is an elementary descriptor embraced in several applications. LBP searches for each center pixel of an image and its localized neighborhood pixel values (X) confined within the window size ascertained by Y. LBP can be evaluated as $\text{LBP}_{X,Y} = \sum_{i=0}^{X-1} T(P_c - P_n)^{2i}$, where P_n and P_c are the intensity values of neighborhood pixels and center pixels. The thresholding process is represented by $T(\cdot)$. $X = 8$ and $Y = 1$ are the values applied when high-resolution images are taken into consideration. The development of binary pattern for a window of 3 × 3 utilizing the weighted sum of the eight convolutional filters by convolutional LBP can be examined as an instance. Then the LBP using convolution is expressed as,

$$C = \sum_{i=1}^{X} \sigma(b_i * V).w_i, \text{i} = \{1, 2, \ldots, \text{X}\}, \tag{9}$$

where * represents convolution and b_i is the sparse convolutional filter with two nonzero values $\{+1, -1\}$, which convolves with the vectorized input image V. σ denotes Heaviside step function. $w = [2_{X\text{-}1}, 2_{X\text{-}2}, \ldots, 20]T$ constitutes predetermined weight values. X is the pixel count in the neighborhood (e.g., $X = 8$ with 3×3).

3.7 SLBT

SLBT is proficient in the extraction of both shape-based and texture-based attributes. A combination of the shape and texture aspects is achieved by reckoning the weights corresponding to each shape feature [11]. The texture, local, and global shapes are obtained by Eq. (10)

$$S = V_i^T \left(b_{ST} - \overline{b}_{ST}\right) \tag{10}$$

where S is the shape texture parameter, V_i^T is the eigenvectors, and $\overline{b}_{ST}$ is the mean vector $\mathbf{b_{ST}}$ is calculated using [11]. Combination of these two-parameter vectors, we get convolutional SLBT parameter vector which is formulated as,

$$\text{CSLBT} = C + S \tag{11}$$

3.8 Linear Discriminant Analysis (LDA)

LDA is used for pattern recognition. LDA is the traditional method of reducing dimensionality. It helps depict the disparity between groups of data. This transform-based approach improves the proportion between-class variance to the within-class variance throughout every dataset [17].

3.9 Classification Approach Using Convolutional Neural Network

Convolutional layers derive local characteristics from the MR image data. The extracted features are passed through convolutional neural network. CNN is a neural network's most common classification method developed with neurons that have learned with weights and biases. Each neuron receives some input and performs dot and precedes it in a nonlinear way. The entire network also shows a distinct discernible score function from input image pixels on one end to class scores on the other, and the fully connected layer has a loss function on it.

As a part of discerning the local structures, convolution layers heed the neurons. Weight allocation is carried out between the nodes to activate the local traits across the entire input channel. Kernel refers to the shared weights set, which can be utilized predominantly in the convolutional layer to discover the feature maps. CNN encompasses discrete pairs of convolution and pooling layer accompanied with fully connected layers and softmax layer as the terminal layer to foster the outcome generation.

The backpropagation (BP) algorithm is employed in the CNN training phase. The algorithm examines optimum weight value by utilizing the gradient descent approach and subsequently reduces the computational intricacy. Normalized values are move to the neural network input layer irrespective of the existence of the pre-processing phase. Every neuron's aggregate total potential in the hidden layer is computed using these input values and initialization of random weights and accordingly. The output is determined by exploring the capability of the activation function. The successive layers of these steps are repeated. Gradients are availed to rationalize the weights and biases in due course of training by propagating the error signal backward. Learning and momentum measurements are used to increase the learning speed and circumventing the local minima during the convergence. The forgoing the feedforward and feedback estimations are continual operations until the terminating parameter is acquired, which generally refers to the root mean square (RMS) an error value.

4 Result and Discussion

MRI images are considered for the detection of dementia. Data are taken from the UCI repository. Here, images are taken as the base of brain image, with the complete structure of cerebrum and medulla oblongata, which focus as an input structure, such that many fields are diverse within the image, where the images are taken from the repository UCI data, where many images are formulated differently, here in our scenario images, which are focused with a frontal view such that different correlation aspect is highlighted. Those are key facts to measure the image intensity in different scopes. 150 subjects were chosen for classification. The subjects were divided into training and testing data set. 100 subjects were used for training and 50 for testing. It is processed through the algorithm to have regular and remedy required for the patient who required corrective action with accuracy and efficiency. The proposed work uses MATLAB for performance evaluation.

The input MRI is shown in Fig. 3a. The given input is pre-processed. Figure 3b shows the morphological image used for the separation of the demented and normal brain regions. After feature extraction and reduction, classification is performed on

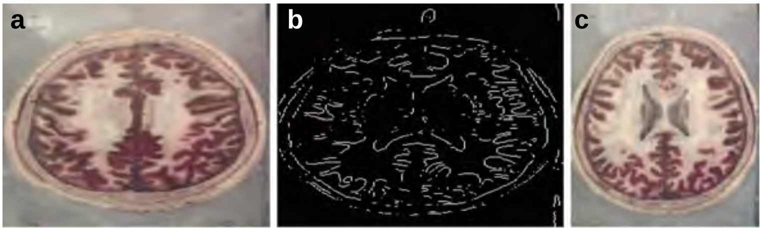


Fig. 3 **a** Input image; **b** Morphological image; **c** Abnormal MRI image

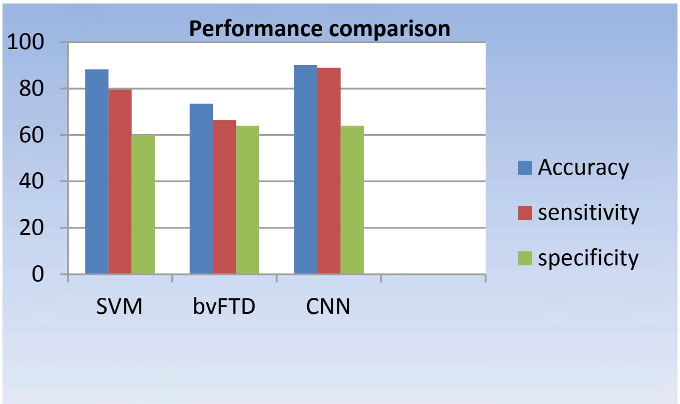


Fig. 4 Graphical representation of accuracy, sensitivity, and specificity

the enhanced image for accurate dementia detection. Figure 3c shows the abnormal brain image. Also, a judgment on the observed outcome may be either true or false. The adjudication would become one of four possible groups. The noise is observed in the dataset like homogeneous noise, texture noise, magnitude noise, non-uniformity of construction of noise.

4.1 Performance Analysis

Performance metrics are essential in the discrimination of deep learning models. In this paper, three metrics, such as accuracy, sensitivity, and specificity metrics, are used for performance evaluation. These measures are estimated based on true positive (T_p), true negative (T_n), false positive (F_p), and false negative (F_n). The evaluation methods are compared to existing methods [12, 16] (Table 1).

$$\text{Accuracy} = \frac{T_p + T_n}{T_n + T_n + F_p + F_n} \tag{12}$$

Table 1 Classification evaluation for various techniques

Techniques	Accuracy (%)	Sensitivity (%)	Specificity (%)
SVM [12]	88.3	79.61	59.90
BvFTD [16]	73.5	66.4	64
CNN	90.1	89	*64*

$$\text{Sensitivity} = \frac{T_p}{T_p + T_n} \tag{13}$$

$$\text{Specificity} = \frac{T_n}{T_n + F_p} \tag{14}$$

5 Conclusion

We establish a novel approach for feature extraction during this mission, namely convolutional shape local binary texture (convolutional SLBT) along with GLCM and statistical analysis, which improves the diagnostic accuracy. This method decreases redundancy and provides key features to clinical scores. Besides, compared to the existing methods, our approach gains good accuracy in dementia detection. Convolutional SLBT will provide smooth, intensified sharpen and enhanced shape and texture image. The proposed method achieved 90.1% accuracy and 89% sensitivity.

References

1. M.R. Ahmed, Y. Zhang, Z. Feng, B. Lo, O. Inan, H. Liao, Neuroimaging and machine learning for dementia diagnosis: recent advancements and future prospects. IEEE Rev. Biomed. Eng. 1–1. https://doi.org/10.1109/rbme.2018.2886237
2. S. Bauer, R. Wiest, L.-P. Nolte, M. Reyes, A survey of MRI-based medical image analysis for brain tumor studies. Phys. Med. Biol. **58**(13), R97–R129 (2013)
3. T.S. Armstrong, Z. Cohen, J. Weinberg, M.R. Gilbert, Imag. Tech. Neuro-Oncol. **20**(4), 231–239 (2004)
4. S. Manoharan, Performance analysis of clustering based image segmentation techniques. J. Innov. Image Process. (JIIP) **2**(01), 14–24 (2020)
5. E.F. Badran, E.G. Mahmoud, N. Hamdy, An algorithm for detecting brain tumors in MRI images, in *Proceedings of International Conference on Computer Engineering and Systems (ICCES)* (2010), pp. 368–373
6. V. Anitha, S. Murugavalli, Brain tumour classification using two-tier classifier with adaptive segmentation technique. IET Comput. Vis. **10**(1), 9–17 (2016)
7. J. Naik, S. Patel, Tumor detection and classification using decision tree in brain MRI. Int. J. Comput. Sci. Netw. Secur. (IJCSNS) **14**(6), 87 (2014)
8. T. Vijaykumar, Classification of brain cancer type using machine learning. J. Artif. Intell. **1**(2), 105–113 (2019)

9. K. Kawanishi, H. Kawanaka, H. Takase, S. Tsuruoka, A study on dementia detection method with stroke data using anomaly detection, in *Proceedings of 6th International Conference on Informatics, Electronics and Vision and 7th International Symposium in Computational Medical and Health Technology (ICIEV-ISCMHT)* (2017), pp. 1–4
10. S. Vieira, W.H.L. Pinaya, A. Mechelli, Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications. Neurosci. Biobehav. Rev. **74**, 58–75 (2017)
11. N.S. Lakshmiprabha, S. Majumder, Face recognition system invariant to plastic surgery, in *Proceedings of 12th International Conference on Intelligent Systems Design and Applications (ISDA)* (IEEE, 2012), pp. 258–263
12. Jun Zhang, Mingxia Liu, Le An, Yaozong Gao, and DinggangShen, "Alzheimer's Disease Diagnosis using landmark-based Features from Longitudinal Structural MR Images", IEEE Journal of biomedical and health informatics, vol.21, no.6, pp.1607–1616, 2017.
13. D. Pachauri, C. Hinrichs, M.K. Chung, S.C. Johnson, V. Singh, Topology-based kernels with application to inference problems in Alzheimer's disease. IEEE Trans. Med. Imag. **30**(10) (2011)
14. T. Tong, Q. Gao, R. Guerrero, C. Ledig, L. Chen, A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease. IEEE Trans. Biomed. Eng. **64**(1), 155–165 (2016)
15. Qi. Zhou, M. Goryawala, M. Cabrerizo, J. Wang, W. Barker, D. Loewenstein, R. Duara, M. Adjouadi, An optimal decisional space for the classification of Alzheimer's disease and mild cognitive impairment. IEEE Trans. Biomed. Eng. **61**(8), 2245–2253 (2014)
16. R.A. Feis, M.J.R.J. Bouts, J.L. Panman, L.C. Jiskoot, E.G.P. Dopper, T.M. Schouten, F. Vos, J. Grond, J.C. van Swieten, S.A.R.B. Rombouts, Single-subject classification of pre-symptomatic frontotemporal dementia mutation carriers using multimodal MRI. NeuroImage Clin. (2019)
17. J.-S. Wang, W.-C. Chiang, Y.-L. Hsu, T.C. Yang, ECG arrhythmia classification using a probabilistic neural network with a feature reduction method. Neurocomputing **116**, 38–45 (2013)
18. S. Goswami, L.K.P. Bhaiya, A hybrid neuro-fuzzy approach for brain abnormality detection using GLCM based feature extraction, in *2013 International Conference on Emerging Trends in Communication, Control, Signal Processing and Computing Applications (C2SPCA)* (Bangalore, 2013), pp. 1–7. https://doi.org/10.1109/C2SPCA.2013.6749454
19. P. John,Brain tumor classification using wavelet and texture-based neural network. Int. J. Sci. Eng. Res. **3**(10), 1–7 (2012)
20. C. Studholme, V. Cardenas, E. Song, F. Ezekiel, A. Maudsley, M. Weiner, Accurate template-based correction of brain MRI intensity distortion with application to dementia and aging. IEEE Trans. Med. Imag. **23**(1) (2004)

Classification of Ultrasound Thyroid Nodule Images by Computer-Aided Diagnosis: A Technical Review

Siddhant Baldota and C. Malathy

Abstract The thyroid is an indispensable gland of the human endocrine system that secretes hormones which have a significant effect on the metabolic rate and protein synthesis. The thyroid is susceptible to a variety of disorders. One among them is the formation of a thyroid nodule, an extraneous mass formed at the thyroid gland requiring medical attention and diagnosis. About 5–10 nodules out of 100 are malignant (cancerous). When a formation of nodules is discernible to the doctors, they call for a diagnostic blood test, often perfunctory, and do not differentiate malignant and benign tumours. This is where ultrasonography comes across as a better option. Automation of ultrasonography diagnosis results in a decrease in reporting time as well as provides a provisional diagnosis before the doctors' expert opinion. Thus, deep learning methods were suggested and produced better results. Initially, region of interest (ROI) and feature extraction were done before applying machine learning models like support vector machines and multilayer perceptrons. Nevertheless, the feature selection required in machine learning methods was a long-drawn process, often involving the elements of trial and error. Deep convolutional neural networks along with histogram of gradients (HOG)-aided feature extraction which was used along classifiers have yielded high specificity and sensitivity values along with the accuracy. In this paper, we have studied and compared the efficacy of the application of various deep convolutional networks for the diagnosis of malignancy in thyroid nodules.

Keywords Computer-aided diagnosis (CAD) · Deep learning · Machine learning · Malignancy · Thyroid nodules · Ultrasound (US) · Ultrasonography(USG)

S. Baldota · C. Malathy (✉)
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur 603203, India
e-mail: malathyc@srmist.edu.in

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_30

1 Introduction

The application of artificial intelligence (AI) in healthcare has been a topic of discussion, experimentation, and application for quite some time now. The rise of AI in healthcare has coincided with population growth, leading to a larger number of people affected by various kinds of diseases. Over the last ten years, the people of India, especially the youth, who have been reportedly affected by thyroid cancer, have been more than any other decade. One of the possible reasons for this is that a larger number of people have become aware and get themselves tested by a radiologist. However, due to the increased amount of testing, the time required for reporting the results for each individual has increased. To counter this problem of latency between testing and results, the automation of provisional reports via AI has come to the fore. Particularly, machine learning models have been used over the years to give a prediction of whether a thyroid tumour is malignant or not. However, machine learning models require optimal feature selection, large datasets to train themselves on, time and resources, expert interpretation, and expensive feature engineering. All of this adds to the latency of diagnosis, which defeats its purpose. To overcome this problem, supervised deep learning has been used. Supervised deep learning is a branch of machine learning dealing with deep artificial neural networks, weights, biases, loss functions, gradient optimizers, activation units, and more. The process of feature selection and engineering is automated by using deep learning.

Along with the testing, the system for reporting too has transformed. Previously, a binary system was used to classify between malignant and benign thyroid nodules. However, this system had some disadvantages. First, the misclassification rate was high as there were a number of false positive and false negatives per class—as only two classes were present: benign and malignant. Changes came about when Horvath et al. [1] presented a reporting system called Thyroid Image Reporting and Data System which consisted of a spectrum of malignancy illustrated in Fig. 2. This spectrum sets the paradigm for the classification of thyroid nodules and their diagnosis. As shown in Fig. 1, the ultrasound report of the thyroid nodule is analysed by the medical expert as well as the computation algorithm. After analysis of the report, each entity gives its opinion. If these opinions match, then a final decision is made and the nodule is deemed as either strictly malignant or strictly benign. In the case of a difference in opinion between the computer and the medical expert, a team of medical expert analyses the nodule, providing a collective perspective. In case of TIRADS, the final decision can be of six different categories as shown in Fig. 2. This paper reviews the computer-aided diagnosis (CAD) done on both the binary classification of thyroid nodules into malignant and benign and explains the need of using TIRADS.

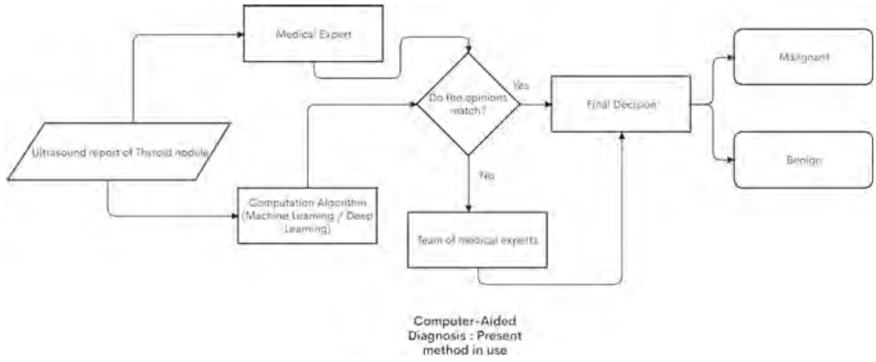


Fig. 1 Computer-aided diagnosis system currently in use

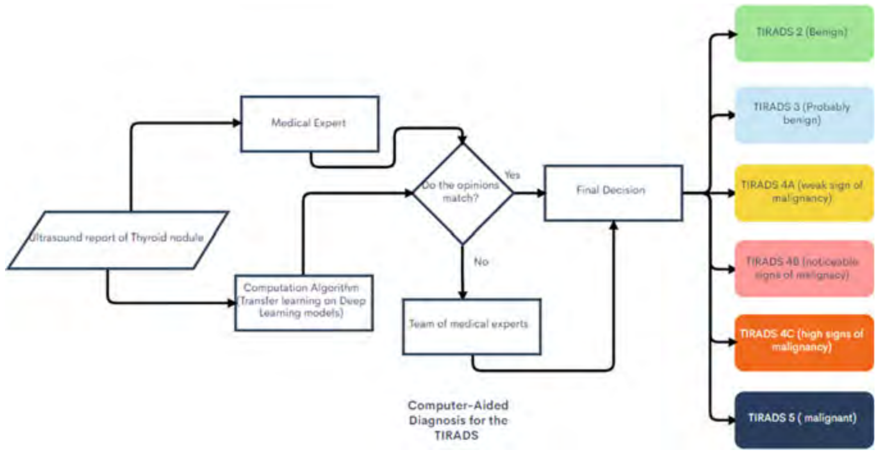


Fig. 2 Computer-aided diagnosis system using TIRADS

2 Literature Survey

We began our survey of the existing research and review work in the diagnostic approach by doctors for thyroid nodules. Hena et al. addressed the problem of malfunction of the thyroid and its diagnosis by discussing various evaluation methods for diagnosis [2], particularly in India. They discussed that the American Association of Clinical Endocrinologists (AACE) had set up a systematic step-wise approach for diagnosis. The AACE suggested ultrasonography (USG) when they were showing signs of being malignant. Moreover, USG diagnosis proves to be a brilliant supporting factor for aspiration biopsy which has a false positive rate upwards of 7% and a false negative rate ranging between 1 and 7%. Similarly, Geanina et al. evaluated the different diagnostic studies including serum markers, thyroid USG, elastography, fine needle aspiration (FNA) biopsy, cytologic diagnosis, and indeterminate cytology [3]. Out of these, FNA along with USG was one of the linchpins of thyroid diagnosis, giving a justification of lower false negatives when FNA is used with USG to support it in comparison to the methods involving palpations.

To provide a second opinion to medical professionals, automation of the USG reports is followed. Automation of thyroid nodule diagnosis has been largely due to the application of machine learning and deep learning. Machine learning techniques like support vector machine (SVM) [4], introduced by Vapnik, were used to detect cancerous diseases [5]. Training algorithms from real-world examples and problems were taken to train the support vector machine. The particle swarm optimization [6] by Kennedy et al. was used to classify using iterations. This method was analogous to the incipient movement of birds flying in a flock, looking for food. Principles of quantum physics for particles were applied to a modification of this method called quantum-behaved particle swarm optimization [7]. Apart from these complex training procedures, the process of feature selection [8] in SVM was involved. Mei-Ling et al. [9] worked on expanding the function of SVM from giving good results binary classification to multiple classes, using a recursive feature elimination (RFE) approach. Using a zoo dataset consisting of 16 features, a class label over 100 test samples, and a dermatology dataset having 33 features, one class label, and over 360 test samples and 16 features, they performed reverse order sorting on the features based on descriptive power. After sorting, new sets of features were selected by RFE, and hyperparameters were tuned. Finally, the feature sets were fed to the SVM to obtain over 95% classification accuracy. The process of SVM-RFE is depicted in Fig. 3. This long-drawn process of feature selection, however accurate, is not optimal as the time latency is high as well as sorting time complexity. Therefore, SVMs are not an ideal method to automate the process of thyroid diagnosis.

Zhi Fei et al. [10] proposed the categorization of medical images of two standard datasets, ISIC2017, and HIS2828, for computer-aided diagnosis (CAD) using high-level feature extraction using a deep convolutional neural network (DCNN), fusing these high-level features extracted from convolutional filters and rectified linear units (ReLu) [11] combined with traditional features, like the histogram of colours, moments of colour, shapes, structural and textural features, subsequently using a multilayer perceptron (MLP) classifier as shown in Fig. 4. The SVM using

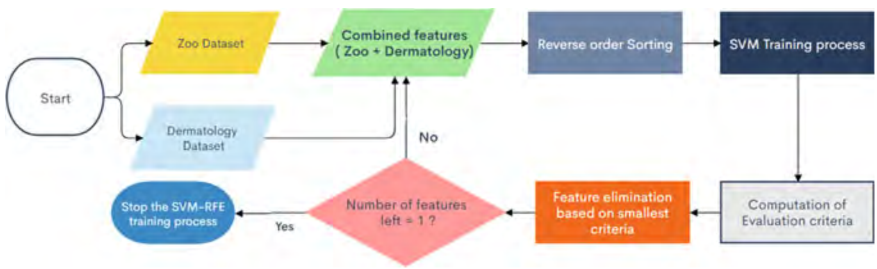


Fig. 3 Support vector machine—recursive feature elimination

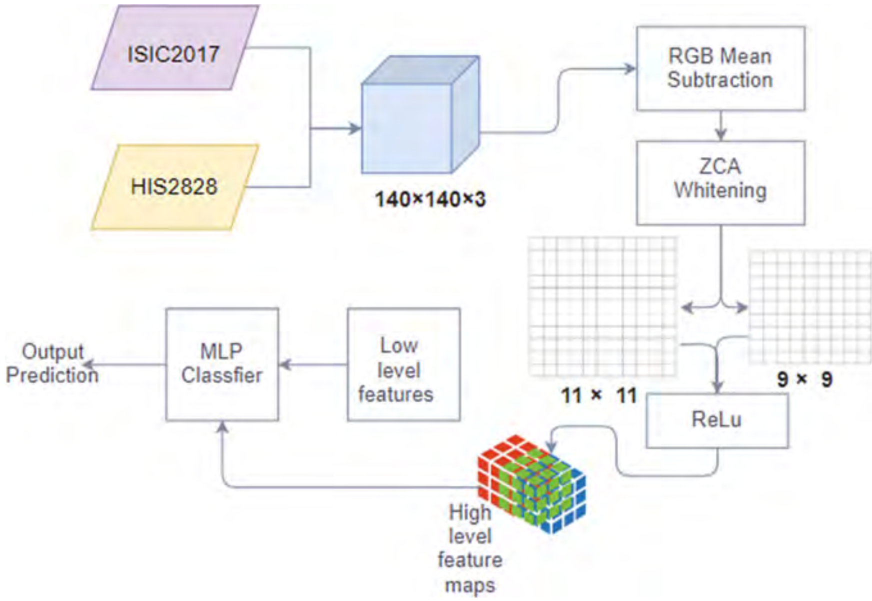


Fig. 4 Methodology used by Zhi Fei et al. which uses a combination of high- and low-level features

traditional features was outperformed by the model used by them. They obtained an accuracy of 90.1% on ISIC2017 and 90.2% on the HIS2828. However, this process relied on some traditional feature extraction, which dampens the purpose of automation.

Before we found out about deep learning methods to perform thyroid nodule diagnosis, there were a few problems that needed to be addressed. A common problem identified was class imbalance in medical datasets. The healthy cases or benign cases occur far more in quantity in comparison to malignant cases since a disease generally affects only a small part of the global population. The problem with the class imbalance in medical data leads to the model is highly biased, which makes accuracy a redundant metric.

Sara et al. were cognizant of the class imbalance problem and developed a cost-sensitive method [12] inspired by the least mean square (LMS) technique [13]. The LMS method reduces the average quadratic error to its minimum by determining the solution to a set of linear equations. Similarly, they solve the problem of class imbalance by penalizing errors of various samples using distinct weights, based on the cost-sensitive protraction of LMS. This method was an improvement from the traditional upsampling or downsampling of the deficient or superabundant class, which can change in the organization of datasets [14]. An exhaustive review of the sampling of classes is found at [15].

One more issue medical datasets face is the number of images. The quantity of the data in the form of images for thyroid nodules available through public platforms is significantly low in volume in comparison to, say, for instance, plant data or face recognition datasets. Since deep learning is a data-hungry process [16], the performance of deep learning models, even the ones used in the medical scenario, becomes better with an increase in data. Data augmentation, promulgated by Martin et al. in [17], provides a solution to this problem. A wide range of augmentation tricks exists—some of which include random cropping, adding and reducing noise, jittering, rotation, zooming, shearing, hue saturation value (HSV) filtering used by Hyuon-Koo et al. in [18], Gaussian filtering, horizontal and vertical flipping, resizing, rescaling, and introduction lighting and contrast in images. Zeshan et al. have worked in [19] to determine which augmentation strategies for datasets to be fed into deep learning models benefit medical image classifiers the most by making them more differential. Flips—both horizontal and vertical, the introduction of Gaussian noise, jittering, scaling, raising to a power, Gaussian blurring, rotations, and shearing were done on each image of a balanced dataset consisting of 1650 abnormal and 1651 normal cases with an 80–20 train validation split from the Digital Database of Screening Mammography [20]. After this, every image was then cropped to 500×500 to hold its resolution when it was fed to VGG-19 [21] in the size of 224×224. Effective plans of action were found out to be using flips, which yielded a validation accuracy of 84%, and gaussian filters, which yielded 88% on the validation set. Conversely, the addition of Gaussian noise worsened the results, rendering accuracy of 66% on the validation data.

Nima et al. [22] worked another pertaining question related to deep learning for medical images. The issue was whether to train deep learning models from scratch with random initialization of weights or to perform transfer learning pretrained on ImageNet [23] weights. They took up four different datasets from three unique medical applications related to computed tomography, cardiology, and gastroenterology. After training and inference, a conclusion appeared that in three datasets out of the four layer-wise transfer works better. Moreover, the worst performance of transfer learning was equal to the training from scratch.

Dhaya et al. [24] worked on detection of Covid-19 from chest radiographs by using transfer learning on deep learning architectures like modifications of Inception (GoogleNet) architectures like InceptionResNetV2, InceptionV3, and a residual neural network like ResNet50. Data was taken from various sources from GitHub. The dataset included 50 chest X-ray images affected by Covid-19 as well as 50 normal chest X-ray images. ResNet50 gave the best results with an accuracy of nearly 98%. The work done in [24] highlights the importance of transfer learning on pretrained architectures.

Fatemeh et al. performed image augmentation for automated thyroid nodule detection from USG images to be fed into their proposed Mask R-CNN model in [25]. It was observed that shearing, warping, and addition of noise to the images did not fit the application whereas naive strategies like flipping, rotations, scaling seemed to work well. Even if the use case of [25] was different, they too worked on ultrasound (US) thyroid nodules images belonging to the same distribution. From [19, 25], we infer that simple augmentation techniques work well with thyroid nodule US data. Sai et al. used deep learning methods for CAD of the thyroid US [26]. They used open-source TIRADS data consisting of 298 images and a local dataset that was labelled by a medical professional. For the TIRADS data, they grouped the images in malignant and benign by labelling levels 2 and 3 as benign and 4 and above as malignant. They compared three methods: (a) training a convoluted neural network consisting of three convolutional layers with 3×3 filters (as shown in Fig. 5) from scratch (b) transfer learning on VGG-16 [21] and Inception-v3 [27] by adding a bottleneck and an SVM as shown in Figs. 6 and 7 to classify. The features of the bottleneck were obtained in the previously trained CNN. (c) They used Inception-v3 and VGG-16 as their baseline models to perform fine-tuning on, using softmax to classify as shown in Figs. 8 and 9. VGG-16 with SVM resulted in giving the best results with an accuracy of 99%, a sensitivity of 1, and a specificity of 0.82.

Fig. 5 Architecture of convolutional neural network which was trained from scratch using random weights

Fig. 6 InceptionV3 or VGG-16 bottleneck features passed through convolutional neural network

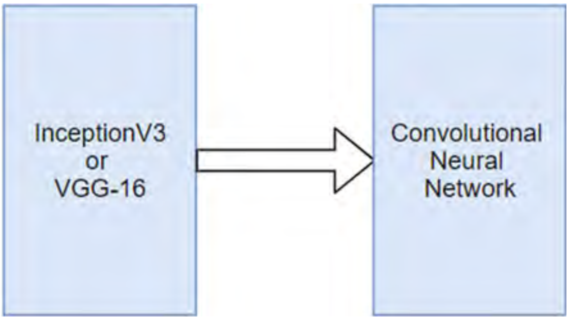
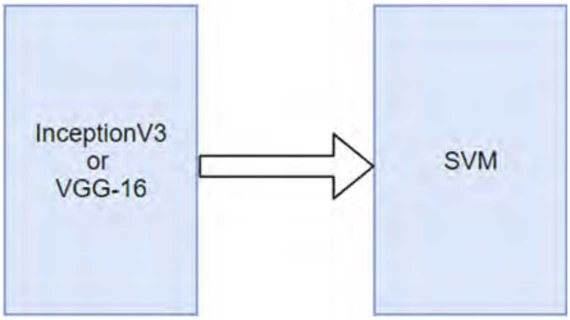


Fig. 7 InceptionV3 or VGG-16 bottleneck features passed through SVM classifier



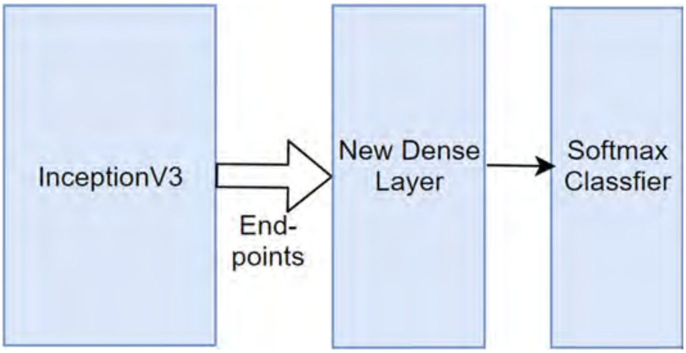


Fig. 8 Transfer learning on inceptionV3

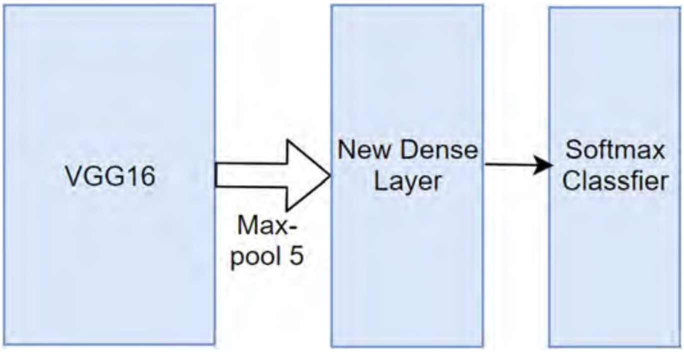


Fig. 9 Transfer learning on VGG-16

Tianjiao et al. [28] classified malignant and benign thyroid nodule US images by using a combination of deep learning feature maps and histogram of gradients (HOG) feature extraction. They performed transfer learning on the pretrained VGG-19 [21] network and obtained the feature maps from the third pooling layers and first two fully connected layers. These high-level feature maps were fused with scale-invariant feature transform (SIFT) local binary patterns (LBP) and HOG, by directly connecting the features from VGG19, LBP, SIFT and then performed feature selection based on differential sorting of the nature of samples. Another technique used for fusion was a voting strategy on each of the feature extraction methods by finding out the mode of the classification results. The feature maps of the third pooling layer of VGG-19, when fused using the feature selection strategy, with the extracted features—HOG, LBP, and SIFT—yielded the highest accuracy, sensitivity, specificity, and area under the receiver-operating curve (AUC) of magnitude 93.1%, 0.908, 0.945, and 0.977, respectively.

Olfa et al. [29] used residual convolutional neural networks (ResNet) [30] for CAD of 3771 US images, having a benign–malignant ratio of 1316 images to 2455 images.

They performed transfer learning on the ResNet50 architecture. They claimed to obtain better accuracy than VGG-19. Meng et al. [31] improved upon this by using region of interest (ROI) extraction by performing automate the segmentation process on thyroid ultrasound images using U-Nets [32]. Data augmentation was done on the ROI using variational autoencoders [33]. The augmented data along with the original ROI was fed to a pretrained Resnet50 model. An accuracy of 87.4%, 0.92 sensitivity, and 0.868 specificity were obtained by the ResNet50 when ROI extraction and autoencoder augmentation was used. Without the augmentation process, transfer learning on ResNet50 yielded 85.1% accuracy, 0.923 sensitivity, and 0.824 as specificity.

Avola et al. used a knowledge-driven approach in [34], wherein the opinions of experts on ultrasound images are concatenated and passed through dense layers. The weights from the dense layers are combined with the weights obtained from the feature maps extracted performing transfer learning on densely connected convolutional neural network (DenseNet) [35] pretrained on ImageNet [23] weights. After the fusion of weights, they are passed through three more dense layers, after which they traverse through a softmax activation unit which gives the resultant output. Ensemble learning was done where DenseNet was replaced by distinctive DCNN architectures. Moreover, while performing transfer learning, a particular percentage of layers were unfrozen from the baseline model. DenseNet, when its layers were frozen up to 25%, resulted in the accuracy over the cross-validation set.

Young Jun et al. reviewed the application of deep learning to US images to thyroid nodules in [36]. They identified that the problems faced while using deep learning in US thyroid nodule images. First, the image selection should be done by highly skilled clinicians. Next, the resolution of the US images should be of high quality and has a standard resolution throughout the distribution. Moreover, the number of images should be adequate as deep learning requires a lot of data [16]. Also, the labelling should be done accurately to identify the diseases properly. A poorly labelled dataset for classification is one of the worst things that could happen to the model, even if it is quite robust. The most impending thing though is the classification of images not just into malignant or benign but into the indeterminate TIRADS proposed in [1].

The difference between images of TIRADS [1] is best illustrated by Aishwarya et al. [37]. Some of the parameters used while TIRADS scoring is done based on morphological and physical features like the dimensions of the lesion, the structure of the lesion, the response to ultrasound waves by the nodule, the internal component of the nodule, the formation of calcium on the nodule tissue, perinodular halo and vascularity of colour Doppler. Several criteria were set at one-point increment to the final score. The final scoring decided the level of suspicion of malignancy. The authors of [37] concluded that TIRADS was reliable, accessible, and convenient. Table 1 highlights the current methods used on ultrasound data and their shortcomings, by summarizing the survey of literature.

Table 1 Summary of literature survey highlighting the shortcomings of the existing work

Title	Dataset used	Problem statement	Methodologies used	Conclusion	Limitation
"SVM-RFE-based feature selection and Taguchi parameters optimization for multiclass SVM classifier" [9]	Zoology and dermatology images	Expanding the function of SVM for achieving better results from binary classification to multiple classes	SVM with recursive feature elimination	Over 95% classification accuracy using SVM-RFE	Feature selection is tedious and not optimal as it leads to high time complexity or sorting as there is latency in results
"Medical image classification based on deep features extracted by deep model and static feature fusion with multilayer perceptron" [10]	ISIC2017 and HIS2828	Categorization of medical images	High-level features extracted from convolutional filters combined with traditional features, classified using a multilayer perceptron (MLP) classifier	They obtained an accuracy of 90.1% on ISIC2017 and 90.2% on the HIS2828	However, this process relied on some traditional feature extraction, which dampens the purpose of automation
"Exploring image classification of thyroid ultrasound images using deep learning" [28]	Open-source TIRADS data	Grouped the images in malignant and benign by labelling levels 2 and 3 as benign and 4 and above as malignant. Classification of thyroid ultrasound images into malignant and benign	Used a CNN trained from scratch. Use VGG16 and InceptionV3 bottleneck features with the CNN and SVM. Fine-tuning on VGG-16, InceptionV3	VGG-16 with SVM resulted in giving the best results with an accuracy of 99%, a sensitivity of 1, and a specificity of 0.82	Even if work was done on TIRADS data, binary classification of strictly malignant or strictly benign was performed

(continued)

Table 1 (continued)

Title	Dataset used	Problem statement	Methodologies used	Conclusion	Limitation
"Classification of thyroid nodules in ultrasound images using deep model-based transfer learning and hybrid features" [30]	Thyroid ultrasound data	Classified malignant and benign thyroid nodule ultrasound images	Used a combination of deep learning feature maps obtained from transfer learning on the pretrained VGG-19 model and HOG feature extraction. They performed	Obtained an accuracy, sensitivity, specificity and AUROC of 93.1%, 0.908, 0.945, and 0.977, respectively	Binary classification of strictly malignant or strictly benign was performed
"Thyroid nodules classification and diagnosis in ultrasound images using fine-tuning deep convolutional neural network" [31]	3771 ultrasound images, having benign:malignant ratio 1316:2455	Classified malignant and benign thyroid nodule ultrasound images	Transfer learning with residual convoluted neural networks (ResNet). Used the ResNet-50 architecture	Better accuracy than VGG-19	Binary classification of strictly malignant or strictly benign was performed
"Knowledge-driven learning via experts consult for thyroid nodule classification" [26]	Open-source TIRADS data	Classified three categories of TIRADS	Ensemble learning was done where DenseNet was replaced by distinctive DCNN architectures. Moreover, while performing transfer learning, a particular percentage of layers were unfrozen from the baseline model	DenseNet, when its layers were frozen up to 25% resulted in the accuracy over the cross-validation set	Classification done for only three categories of TIRADS

3 Discussion

The survey of existing literature helps us to get acquainted with and identify the current practices, algorithms, and methodologies used in this computer-aided diagnosis of ultrasound image segmentation. In regard to medical dataset, we address the issue of data and class imbalance, lower quantity of images, and the decision to use transfer learning or training deep learning models from scratch. From the work analysed, we conclude that the class imbalance problem in medical data is best solved by using a cost-sensitive approach. Simple augmentation methodologies, like flipping and the application of Gaussian filters, tend to work better than complex augmentation strategies, on ultrasound thyroid images. Transfer learning on pretrained imagenet weights was found out to work better than training from scratch, more often than not.

Machine learning algorithms like support vector machines and multilayer perceptrons yield high accuracies using iterative and recursive training methods but have faced a lot of latency issues in feature selection and hyperparameter tuning. Hybrid methodologies do tend to work much better but they still involve feature selection. The process of feature selection can be automated by using deep learning methods. Training with ResNet50, DenseNet, VGG-16, VGG-19 or Inception-v3, all pretrained on imagenet weights are efficient models for performing transfer learning. Transfer learning with VGG-16 gives brilliant results when paired with a bottleneck, whose features are selected from a feature map obtained by training a naive convoluted neural network from scratch and an SVM classifier. Using baseline models like ResNet50 fed with ROI extracted data from U-Nets and augmented data from variational encoders works much better than without augmentation. Nevertheless, all of this was computed on binary classification of diseases for thyroid nodules which assigns them a value strictly malignant or strictly benign. Since malignancy follows more of a spectrum than a binary discrimination, the need for Thyroid Image Reporting and Data System (TIRADS) has become more significant than ever. Some work has been carried out regarding automation of diagnosis on this system but has not been definitive.

4 Dataset

The dataset intended to be used in future work is taken from the Thyroid Digital Image Database, an open-source database of ultrasound images of thyroid nodules. Presently, this database includes a collection of beta tested ultrasound images pertaining to TIRADS, made by expert radiologists. This data is collected from of 389 patients and consists of more than 420 annotated ultrasound thyroid images. Figure 10 shows a small batch of the images from the dataset. We intend to parse the annotations and the TIRADS category to perform ROI based image classification.

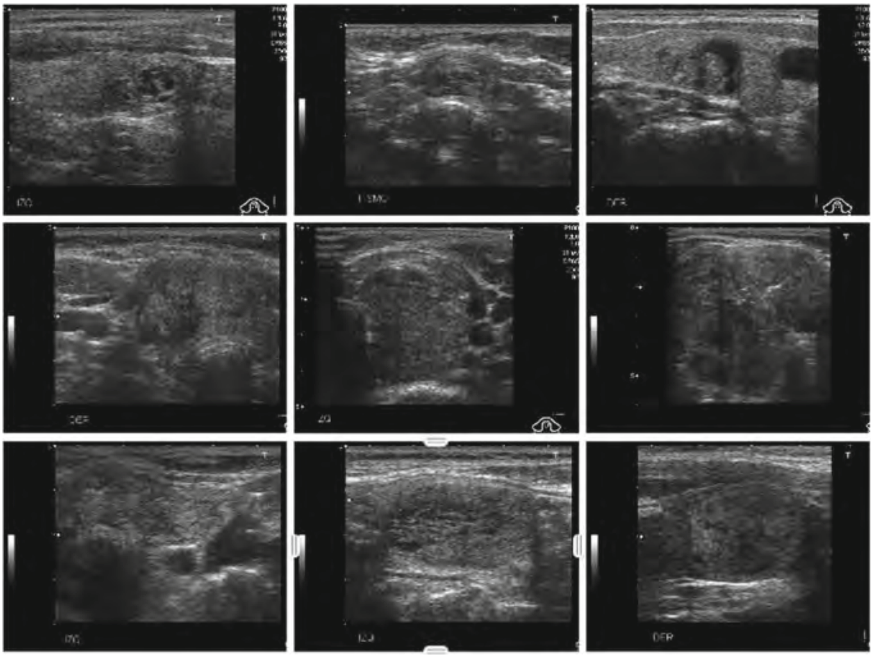


Fig. 10 Batch of thyroid ultrasound images consisting of TIRADS

5 Conclusion

The review of existing literature highlights the importance of deep learning in CAD for ultrasound thyroid images. Deep learning models like VGG-16, VGG-19, ResNet50, DenseNet, and Inception-v3 have automated the process of feature selection, resulting in the reduction of reporting time and providing a quicker and accurate second opinion to medical professionals. From the literature survey, it is observed that transfer learning may work better than other existing training models discussed. Moreover, we recognize the need for implementation of TIRADS diagnosis using deep learning.

The current work largely deals with binary classification of thyroid nodules into malignant and benign. However, one needs to understand that computer-aided diagnosis is not meant to replace medical professionals completely but to provide supporting evidence to their diagnosis. The problem with binary classification is that it deterministically gives the result for the presence or absence of cancer in the thyroid nodule. Such deterministic behaviour results in a greater number of false positive and false negative reports. Even a single false positive or false negative report, especially in terms of binary classification of the malignancy or benignity of a thyroid nodule, can prove to be fatal. This discrepancy has developed the motivation for using an indeterminate spectrum like TIRADS.

Our future work will be to perform transfer learning on existing architectures on TIRADS data by following the general thumb rules of balancing classes and using simple augmentation techniques like flipping, rotations, scaling, and Gaussian filtering. Our future work panders to our goal which is to design an accurate and robust diagnostic system for TIRADS using transfer learning on state of the art architectures.

References

1. E. Horvath, S. Majlis, R. Rossi, C. Franco, J.P. Niedmann, A. Castro, M. Dominguez, An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. J. Clin. Endocrinol. Metabol. **94**(5), 1748–1751 (2009). https://doi.org/10.1210/jc.2008-1724
2. H.A. Ansari, S.M. Vasentwala, N. Saeed, K. Akhtar, S. Rehman, Diagnostic approach to thyroid nodules. Annals Int. Med. Dental Res. (2016). https://doi.org/10.21276/aimdr.2016.2.6.PT1
3. G. Popoveniuc, J. Jonklaas, Thyroid nodules. Med. Clin. North Am. **96**(2), 329–349 (2012). https://doi.org/10.1016/j.mcna.2012.02.002.PMID:22443979;PMCID:PMC3575959
4. V.N. Vapnik, *The Nature of Statistical Learning Theory* (Springer, New York, 1995). https://doi.org/10.1007/978-1-4757-3264-1
5. N.H. Sweilam, A.A. Tharwat, N.K. Abdel, Support vector machine for diagnosis cancer disease: a comparative study. Egypt. Inform. J. **11**(2), 81–92 (2010). https://doi.org/10.1016/j.eij.2010.10.005
6. J. Kennedy, R. Eberhart, Particle swarm optimization, in *Proceedings of ICNN '95—International Conference on Neural Networks*, vol. 4 (Perth, WA, Australia, 1995), pp. 1942–194848. https://doi.org/10.1109/ICNN.1995.488968
7. X. Fu, W. Liu, B. Zhang, H. Deng, Quantum behaved particle swarm optimization with neighborhood search for numerical optimization. Math. Probl. Eng. Article ID 469723, 10 pp. (2013). https://doi.org/10.1155/2013/469723
8. M.H. Nguyen, F.D.L. Torre, Optimal feature selection for support vector machines. Pattern Recogn. **43**(3) (2010). https://doi.org/10.1016/j.patcog.2009.09.003
9. M.-L. Huang, Y.-H. Hung, W.M. Lee, B.-R. Jiang, SVM-RFE based feature selection and Taguchi parameters optimization for multiclass SVM classifier. Sci. World J. Article ID 795624. https://doi.org/10.1155/2014/795624
10. Z. Lai, H.F. Deng, Medical image classification based on deep features extracted by deep model and static feature fusion with multilayer perceptron. Comput. Intell. Neurosci. Article ID 2061516. https://doi.org/10.1155/2018/2061516
11. V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines, in *Proceedings of the 27th International Conference on Machine Learning* (Haifa, 2010), pp. 807–814
12. S. Belarouci, M. Chikh, Medical imbalanced data classification. Adv. Sci. Technol. Eng. Syst. J. **2**, 116–124. https://doi.org/10.25046/aj020316
13. D.-Z. Feng, Z. Bao, L.-C. Jiao, Total least mean squares algorithm. IEEE Trans. Signal Process. **46**(8), 2122–2130 (1998). https://doi.org/10.1109/78.705421
14. X.-Y. Liu, J.X. Wu, Z.-H. Zhou, Exploratory under-sampling for class-imbalance learning, in *Proceedings of the 6th International Conference on Data Mining (ICDM '06)* (IEEE, Hong Kong, 2006), pp. 965–969
15. H. He, E.A. Garcia, Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. **21**(9), 1263–1284 (2009)
16. C. Sun, A. Shrivastava, S. Singh, H. Mulam, Revisiting unreasonable effectiveness of data in deep learning era, in *IEEE International Conference on Computer Vision (ICCV)* (2017), pp. 843–852. https://doi.org/10.1109/ICCV.2017.97

17. M.A. Tanner, H.W. Wing, The calculation of posterior distributions by data augmentation. J. Am. Stat. Assoc. **82**(398), 528–540 (1987). https://doi.org/10.2307/2289457
18. H. Kim, J. Park, H.-Y. Jung, An efficient color space for deep-learning based traffic light recognition. J. Adv. Transp. (2018), pp. 1–12. https://doi.org/10.1155/2018/2365414
19. Z. Hussain, F. Gimenez, D. Yi, D. Rubin, Differential data augmentation techniques for medical imaging classification tasks, in *AMIA ... Annual Symposium Proceedings. AMIA Symposium* (2017), pp. 979–984
20. R. Lee, F. Gimenez, A. Hoogi, K. Miyake, M. Gorovoy, D. Rubin, A curated mammography data set for use in computer-aided detection and diagnosis research. Sci. Data **4**, 170177 (2017). https://doi.org/10.1038/sdata.2017.177
21. S. Karen, Z. Andrew, *Very Deep Convolutional Networks for Large-Scale Image Recognition.* arXiv:1409.1556v6
22. N. Tajbakhsh, J. Shin, S. Gurudu, R.T. Hurst, M.B. Gotway, J. Liang, Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Med. Imaging **35**(5), 1299–1312 (2016). https://doi.org/10.1109/TMI.2016.2535302
23. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge. Int. J. Comput. Vision (2014). https://doi.org/10.1007/s11263-015-0816-y
24. R. Dhaya, Deep net model for detection of Covid-19 using radiographs based on ROC analysis. J. Innov. Image Process. (JIIP) **2**(03), 135–140 (2020)
25. F. Abdolali, J. Kapur, J.L. Jaremko, M. Noga, A.R. Hareendranathan, K. Punithakumar, Automated thyroid nodule detection from ultrasound imaging using deep convolutional neural networks. Comput. Biol. Med. **122**, 103871 (2020). https://doi.org/10.1016/j.compbiomed.2020.103871
26. K.V.S. Sundar, K. Rajamani, S. Sai, Exploring image classification of thyroid ultrasound images using deep learning, in *Published in Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)* (2019). https://doi.org/10.1007/978-3-030-00665-5_151
27. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV, 2016), pp. 2818–2826. https://doi.org/10.1109/CVPR.2016.308
28. T. Liu, S. Xie, J. Yu, L. Nia, W. Sun, Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features, in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (New Orleans, LA, 2017), pp. 919–923. https://doi.org/10.1109/ICASSP.2017.7952290
29. O. Mouse, Khachnaoui, R. Guetari, N. Khlifa, Thyroid nodules classification and diagnosis in ultrasound images using fine-tuning deep convolutional neural network. Int. J. Imag. Syst. Technol. **30**(1) (2019). https://doi.org/10.1002/ima.22363
30. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV, 2016), pp. 770–778. https://doi.org/10.1109/CVPR.2016.90
31. M. Zhou, R. Wang, P. Fu, Y. Bai, L. Cui, Automatic malignant thyroid nodule recognition in ultrasound images based on deep learning, in *E3S Web Conference* (2020). https://doi.org/10.1051/e3sconf/202018503021
32. O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in N. Navab, J. Hornegger, W. Wells, A. Frangi (eds.), *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science*, vol 9351 (Springer, Cham, 2015). https://doi.org/10.1007/978-3-319-24574-4_28
33. D.P. Kingma, M. Welling, *Auto-Encoding Variational Bayes.* CoRR, abs/1312.6114
34. D. Avola, L. Cinque, A. Fagioli, S. Filetti, G. Grani, E. Rodol, Knowledge-driven learning via experts consult for thyroid nodule classification, 28 May 2020. arXiv:2005.14117v1 [eess. IV]
35. G. Huang, Z. Liu, L. Maaten, K. Weinberger, *Densely Connected Convolutional Networks* (Cornell University paper, 2018). arXiv:1608.06993v5

36. Y. Chai, J. Song, M. Shear, Artificial Intelligence for thyroid nodule ultrasound image analysis. Annals Thyroid **5** (2020). https://aot.amegroups.com/article/view/5429
37. K.C. Aishwarya, S. Gannamaneni, G. Gowda, A. Abhishiek, TIRADS classification of thyroid nodules: a pictorial review. Int. J. Med. Res. **4**(2), 35–40 (2019). www.medicinesjournal.com

A Transfer Learning Approach Using Densely Connected Convolutional Network for Maize Leaf Diseases Classification

Siddhant Baldota, Rubal Sharma, Nimisha Khaitan, and E. Poovammal

Abstract Cereal crops are always considered as one of the most important sources of proteins and carbohydrates to human beings. Cereal grains are consumed daily in any one form or another. Over the last decade, the production and demand of maize have shot up manifolds. As maize remains a very crucial crop for food, this research work looks at one of the profound reasons due to which maize crops might be destroyed. Generally, plant diseases will lead to crop failure most of the time. This research has been carried out on the PlantVillage dataset by using the densely connected convolutional neural networks. The PlantVillage dataset is a benchmark dataset that consists of three classes of diseases and one healthy class. Comparing to earlier proven methods on the same dataset, transfer learning on DenseNet121 has yielded better results by leveraging an accuracy of 98.45% on the test set accompanied with less storage space and low training time. The robustness of the model was proved, when it has yielded a validation set accuracy of about 94.79% and a test set accuracy of 91.49% despite being retrained with a jitter of 0.03 and lighting of 0.15, which will contribute to the noise added to the original dataset. This ability of DenseNet121 to perform well on noisy data increases the probability of the application of the proposed method in countries, where the operational costs are limited.

Keywords DenseNet · Image classification · Maize plant disease · Transfer learning

1 Introduction

Presently, the world has become more vulnerable to food security. A few factors influencing food security worldwide are the rise of environmental problems, natural calamities, crop failures, and so on. In the present scenario, crop failure is the paradigm that requires significant research attention in recent times. Failure or spoiling of crops occurs due to pests, plant diseases, excessive use of chemicals, etc. If

S. Baldota · R. Sharma · N. Khaitan · E. Poovammal (✉)
Department of Computer Science and Engineering, SRM Institute of Science and Technology, 603203 Kattankulathur, Tamil Nadu, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_31

the plant diseases are identified and detected in the early stages, it can be curbed, and the crops can be saved. The growing technology promises better and efficient agriculture by incorporating emerging technologies like artificial intelligence and machine learning. Nowadays, neural networks are also finding their application in the detection of plant diseases. Deep learning models are applied to classify the images of infected plants. Most of these models use pretrained weights that are trained for large scale competitions such as ImageNet and COCO. They are fine-tuned to be fed with an input image, which is categorized within the defined classes as either healthy or one of the diseases.

Cereals are specifically cultivated for their edible component that is grain. Cereals consist of a very rich source of minerals, carbohydrates, vitamins, and proteins. In developing countries, rice, maize, or millet will remain as a part of their daily meal. This research is narrowed down to maize crops. It is an important crop as both its silage and crop residue have heavy usage. Maize can be used to extract both oil and starch. It is used to make porridges and beer. Moreover, people eat it after roasting. Therefore, it is important to look at the maize diseases and a way to detect them.

This work intends to use densely connected convolutional networks (DenseNet) for the image classification on maize leaves to identify the nature and type of the disease. Our dataset contains both healthy and disease images. So, the model does both the jobs at once of detecting the disease and classifying it. The aim is to test the models to check the optimal model which could convey the best classification parameters. ResNets have a heavy architecture (as they can go up to 1000 layers) taking up more space and time to execute. On the other hand, DenseNets have a significantly smaller architecture. DenseNet121 trained on ImageNet taking only up to 30 MB of space. DenseNet shows promising aspects that would be tested on the dataset, split in training, validation, and hold out test sets. To check the robustness of the model, some noise will be introduced to the images till a certain level to check whether the model can sustain in low-resolution images. Since the crop plays a major role as the daily food item in developing countries, there is a pressing need to check the indefatigable nature of the model. Jitter will be introduced in images to check the accuracy in such cases where low-quality and low-resolution images are expected from the farmers of developing nations. Throughout, this research strives to find the best model which fits the requirements for detection and classification of maize plant diseases.

2 Literature Review

The study was initiated with some existing research work setting up the course for disease detection in plants. In general, various image processing, machine learning, and deep learning architectures were used to delineate plant diseases as given in Table 1. The leaf images were taken as they showed the symptoms of the disease clearly. In-depth knowledge was gathered about the plant diseases by Santhosh et al. in [1]. They have researched various diseases like rust which is common in maize

Table 1 Comparison of related works

Title	Objective	Methods	Conclusion
Analysis of artificial intelligence-based image classification techniques [6]	Fruit image classification	ML algorithms like SVM, RF, DA, and KNN	KNN gave the highest accuracy. Tuning K in KNN for test set show dependency on the training procedure
Visual tea leaf disease recognition using a convolutional neural network model [7]	Tea leaf disease classification	BOVW with ML classifiers vs DCNN from scratch	DCNN was better than ML models
Using deep learning for image-based plant disease detection [8]	Plant disease classification on PlantVillage dataset	Transfer learning on DCNN models like AlexNet and GoogleNet	Accuracy using transfer learning was much higher than training from scratch
Maize leaf disease identification based on feature enhancement and DMS-robust AlexNet [3]	Maize disease classification	DMS-robust AlexNet	Higher accuracy, but a lot of hyperparameter tuning needed to be done
Mellowness detection of dragon fruit using deep learning strategy [12]	Mellowness identification in dragon fruit images	VGGNet vs ResNet152	The performance of ResNet was better than VGGNet. ResNet152 is a large model
A large chest radiograph dataset with uncertainty labels and expert comparison [14]	Chest radiograph classification	ResNet152, SE-ResNeXt101, Inception-v4, and DenseNet121	DenseNet121 gave higher accuracy

crops. This work elucidated the plant diseases which could be seen infecting the leaves such as kole roga, yellow leaf disease, leaf curl, and so on. Such diseases could be detected by analyzing the images of their leaves. Before approaching the stated problem, which deals with the crop called maize, it has been attempted to learn the facts about its historical perspective and challenges which one might encounter in the future. Sai et al. have conducted research specifically on maize plants to study in-depth about maize crops [2]. Their work stated the current facts about the production of maize throughout India and growth during recent decades. Moreover, their research explained the challenges which the crop faces and the use of new techniques to enhance production.

Machine learning algorithms like Naïve Bayes, decision tree, and random forest to classify the maize leaf images were used by Kshyanaprava et al. [3]. These algorithms are easy to implement. They used a supervised learning approach. The dataset consisted of 3823 images with four classes including a healthy class. They subjected the images to preprocessing where images were converted to grayscale. Image segmentation and feature extraction were performed to extract the necessary features. All the algorithms were giving above 75% accuracy but random forest yielded the best accuracy of 79.23%. Jitesh et al. carried out a detailed study of various varieties of rice infected by diseases [4]. After acquiring the dataset from a village in Gujarat, the images were subject to preprocessing where RGB was converted into HSV. K-means was used for clustering the images. They used color features for feature extraction. Support vector machines use a supervised learning approach and classify the data based on the class labels. Priyanka et al. classified the cereal crops like maize, rice, and wheat [5]. After examining the images of the crops, image preprocessing and segmentation was applied. They applied six classifiers to get the desired accuracy, but when all crops were combined, Naïve Bayes stood out and gave the accuracy of 90.97%.

Shakya in [6] applied various machine learning (ML) models, namely k-nearest neighbor (KNN), support vector machine (SVM), random forest (RF), and discriminant analysis (DA) on a fruit dataset consisting of 300 images of fruits taken from Kaggle. The dataset had an 80:20 train–test split. Feature extraction on the data was performed by applying various preprocessing methods such as resizing the images into a*b picture elements, converting these resized images to grayscale, and edge detection of these grayscale images using Gaussian filter convolutions and canny edge function. The features extracted from these methods were combined with the color, textural, and spatial features. Ensemble learning was carried out on the machine learning models which acted as classifiers for the data. KNN achieved the best metrics which included the highest accuracy, specificity, and sensitivity of 93.103, 80, and 94.339%, respectively, among the classifiers. However, the use of machine learning models resulted in the need for manual feature selection. Moreover, the estimation of the K values in the KNN was done on basis of this feature extraction carried out during the training process. This indicated a high dependency of the proposed model on the training procedure. However, when used in real-time with different data distributions of fruit images, this model may result in low accuracy. Therefore, this justifies the need for using deep learning for image classification to automate the process of feature selection and reduction in the number of moving parts in the architecture. Jing et al. [7] proved this by proposing LeafNet, a deep convolutional network model to automate the process of feature extraction on a tea leaves diseases dataset consisting of high resolution 3810 images in six categories which were augmented using different transformation processes to result in a total of 7905 images which were distributed in a train: cross-validation: test split of 80:10:10. Comparison of classification was done with a bag of visual words (BOVW) model which was fed with DSIFT features and using SVM and multilayer perceptron (MLP) classifiers. Results hugely turned in favor of LeafNet which obtained an accuracy of 90.16% in comparison with SVM's 60.62% and MLP's 70.77%. Granted that the results obtained by LeafNet were high in comparison with machine learning algorithms, and the arduous process of feature extraction was eliminated. However, transfer learning on deep learning architectures pretrained on ImageNet seemed to give better results than training from scratch.

Sharada et al. applied deep learning models and algorithms on the PlantVillage dataset [8]. They applied AlexNet and GoogleNet with transfer learning. They gradually kept decreasing the learning rate. Their model achieved the best accuracy of

98.21%. Similarly, Sumita et al. applied deep convolutional neural network (DCNN) on the dataset which consisted of various categories of the corn plant. They used a pretrained DCNN model [9]. They developed the model using convolutional layers, max-pooling layers, activation, and dropout layers. The model's optimized learning rate was 0.0004. The designed model gave an accuracy of 98.40%, and using mobile phone images, 88.6% was achieved. Mingjie et al. deployed DMS-robust AlexNet for the classification of seven categories of maize leaves which includes the healthy category too [10]. The data augmentation process like rotation, flipping, etc., was applied. Their method eliminates the need for selecting specific features. Their model was able to achieve more than 98%. However, a lot of hyperparameter tuning was needed to be done for the DMS-robust version of AlexNet. Choosing these hyperparameters was a long-drawn process, and it introduced unnecessary computation overheads.

Krishnaswamy et al. [11] analyzed data from ten different diseases from four distinct crops such as eggplant, beans, okra, and lime from a field in Tamil Nadu, India. They performed transfer learning on VGG16, VGG19, GoogleNet, ResNet101, and DenseNet201. The dataset was augmented by adding transformations like rotations, flipping, and translation. For the validation data, GoogleNet gave the best accuracy of 97.30%. In real-time testing, VGG16 yielded the highest accuracy of 90.00%. One of the reasons for this is due to the variance in the distribution since there were four different crops. ResNets and DenseNets tend to perform better on a single crop having different types of diseases. This is justified in [12] where Vijaykumar et al. proposed a residual convolutional neural network ResNet152 for the identification of mellowness in dragon fruit image dataset, labeled by experts. The dataset had an 80:20 train–test split. The data was trained on the ResNet152 model for a total of 500 epochs. It was observed that the train and test losses did not increase, resulting in higher accuracies in comparison with VGG16 and VGG19. It was claimed that ResNet152 gave very good results, yielding an area under the ROC curve of 1.0. The drawback of using ResNet152 was that it was a large model requiring large storage capacity. Moreover, the depth of the model was 152 layers which meant that the number of parameters of the models was large, resulting in higher computation costs.

Gao et al. [13] proposed the convolutional network which has direct connections between any two-layers having the same size as the feature map. With growing parameters, its accuracy improves. They named it dense convolutional network (DenseNet). Tao T et al. introduced a novel approach by using transfer learning on top of DenseNet [2]. They called it sequential fine-tuning. This model achieves better accuracy than the traditional one. After applying this model, they got high accuracy on their dataset classification. Irvin et al. [14] proved the efficacy of using DenseNet121 over conventional CNNs for transfer learning. The dataset used consisted of 2, 24, and 316 chest radiology images classified for the presence, absence, and uncertainty of 14 types of diseases. Deep learning models such as ResNet152, SE-ResNeXt101, Inception-v4, and DenseNet121 were used as experimental transfer learning baseline architectures. DenseNet121 yielded the highest accuracy in these tests and therefore was trained with a learning rate of 0.0001, a momentum of 0.9, a decay of 0.99, a batch size of

16, for three epochs using an Adam optimizer. The details of training and the results of Chexpert are outside the scope of this paper since the need has been simplified to prove that ResNet121 is optimal for performing transfer learning on image classification and not actually perform the chest radiography classification. The survey of the existing literature leads us to conclude that the DenseNet121 should be used as the baseline model for transfer learning on maize leaf images.

3 Dataset

The dataset used in this paper is a part of a dataset obtained from Kaggle, which consisted of greater than 50,000 curated images from experts. Speaking of numbers, this dataset consists of a total of 3858 images comprising 3090 training images and 384 validation and test set images each. The class balancing is applied by using a cost-based approach to account for the disparity in the number of lead images per class. There are four categories of images including the healthy category. The first disease is known as gray leaf spot which is a fungal disease caused due to warm and humid weather. Initially, spots tan in color appears on the leaves which eventually turn gray when exposed to rain. The second disease is also a fungal disease called common rust that usually occurs during June. Chlorotic flecks are one of the early visible symptoms signifying common rust. Like gray leaf spot, northern leaf blight also occurs during damp and humid environments. The suitable temperature for northern leaf blight is 18–27 °C, and it is one of the easiest diseases to identify because it forms distinct cigar-shaped lesions on the leaves.

4 Methods

a. Preprocessing

The dataset used consists of 3090 training images, 384 validation set images, and 384 test set images for four categories of maize leaves including a healthy class. Train and validation loaders are set up, each having a batch size of 50. Various transformation methods are implemented on the training images to test the robustness of the proposed model and augment the training data. The first method is to try with normal transformations which include horizontal and vertical flipping of images, with a zooming up of the image up to a maximum factor of 1.1 along with a maximum distortion factor of 0.2. A random rotation up to a maximum of 10 degrees is performed on the images, both clockwise and counter-clockwise. Each image was resized to a size of 256 × 256. These transformations were part of the standard transformation path to test the general validation and test accuracy, error rate, and validation loss of the model used on these images. To verify the robustness, a noisy transformation path consisting of vertical flipping, a maximum rotation range of 15 degrees on either side, a lighting effect of the magnitude of 0.15 on a scale of 1 is included. While the distortion factor remained 0.2, jitter is introduced of magnitude 0.03 to the images. Jitter changes the pixels of an image with random replacement with pixels in the vicinity. The degree of the vicinity used to replace the pixels is defined by the magnitude of the jitter. Introducing jitter to the transformation process is a noticeable change from the previous transformation path. Jitter introduces a corruption in the image signal ratio, which makes the images noisy, thereby replicating the distribution of data which is likely to be available in places where image resolution for training is low as shown in Fig. 1.

b. **Model**

The deep convoluted neural network architectures resolve the issue of crops suffering from vanishing gradients. The gradients have to traverse such a long path that they tend to disappear before reaching the layers near the input layer, thus making the training process an almost impossible tasks after a point of time. Residual networks or ResNets were introduced to solve this problem by creating an identity path, thus adding the identity factor to the weights. However, as it turned out, ResNet models tended to be very deep too, going past 1000 hidden layers at times. To solve this problem, DenseNets were introduced. They provide a simplified solution, by connecting each layer directly to the other. This inflated the gradient flow to such an extent that maximum information was passed, reducing the need for very wide or very wide models. The feature maps extracted from other models were rendered superfluous because the model required lesser parameters than even ResNet. Rather,

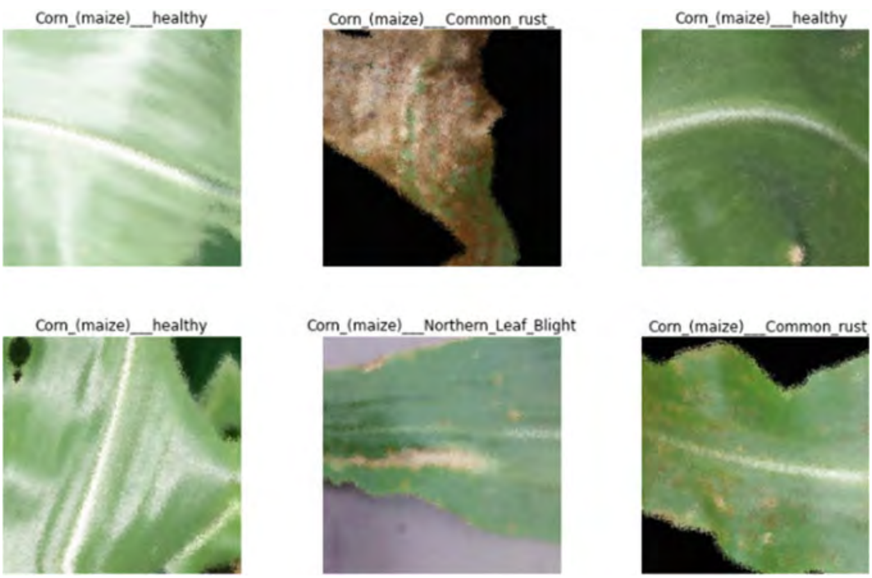


Fig. 1 A batch of maize leaf images from the noisy transformation path

DenseNet relies more on the reusability of weights and features. Another significant difference between ResNets and DenseNets was that ResNets were significantly wider than DenseNets, wherein each layer can access the penalized weights from the loss function. DenseNet applies a composite of operations. Instead of summing up the feature maps, it concatenates them. DenseNets are partitioned into dense blocks. The sizes of the maps remain constant in the block. The changes occur in the number of filters applied between the blocks. DenseNet121 model is applied for our test case, where 121 indicate the depth of the DenseNet model. Transfer learning is performed on DenseNet121 pretrained from ImageNet weights. A striking advantage of using DenseNet121 is its compact size of 30 MB as compared to DenseNet161 whose size exceeds 100 MB. Torchvision models offered by PyTorch are used for transfer learning on DenseNet121.

c. **Training**

The training was carried out on Google Colaboratory that uses an Nvidia K80 GPU and offers up to 12 GB of GPU memory with a memory clock of 0.82 GHz, a performance of 4.1 TFLOPS, an available RAM of 12 GB, and disk space of 358 GB. At first, transfer learning is carried out on DenseNet using the data from the standard transformation path (without jitter). One of the first steps of training is to find an optimum learning rate that minimizes the validation loss. To do this, the model is trained with a changing learning rate and calculated the loss and learning rate curve, where the learning rate was the independent parameter and loss was the dependent parameter. The learning rate was set as the minimum numerical gradient (2.75 e^{-04} or 0.000275) as shown in Fig. 2.

The training was carried out for a total of 25 epochs. It was in two parts of 5 epochs and 20 epochs in that order. The average training time per epoch was 38 s. Once the error rate and the validation loss had become stagnant, training was stopped. Note that the epoch rate is very low because the loss had stopped decreasing and the accuracy had stopped increasing even after training for 5 epochs. Moreover, the number of steps per epoch was high. So, it has been concluded that the epoch rate was adequate. The training procedure for DenseNet121 on the data subjected through the second transformation path was a bit different in terms of the number of epochs and hyperparameter optimization. The purpose of the second transformation was to check whether the DenseNet121 model works as well when fed on high noise and lower quality data. From Fig. 3, it is noticed that the optimal learning rate required which was computed from the minimum numerical gradient had changed a little bit from 2.75 to 2.29 e^{-04}. Then, the model was trained for 5 epochs. Apart from the first epoch which took 94 s, the rest four were done in half the time each.

The model was trained for 25 more epochs to reduce the validation loss obtained after the five epoch training process. However, this was not a standard 25 epochs training process. The model was trained for 5 epochs in an iterative process five times. The catch was that the optimal learning rate was calculated and recalculated every five epochs. Table 2 shows the different learning rate, by computing the minimum

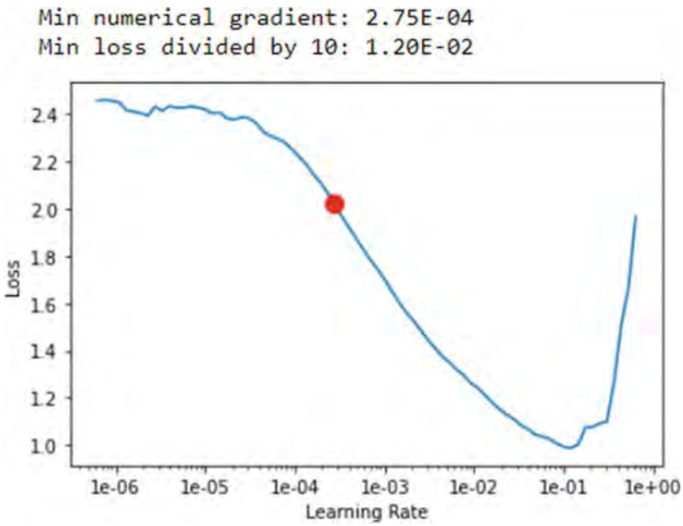


Fig. 2 Finding the near-perfect learning rate for DenseNet121 trained on the standard transformation path

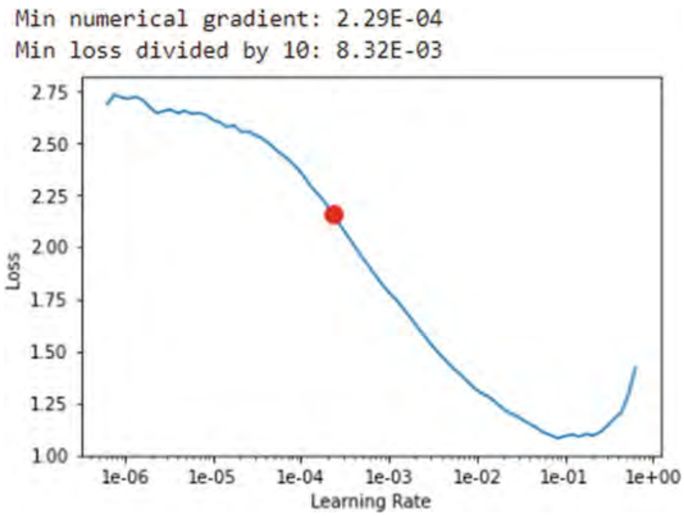


Fig. 3 Finding the near-perfect learning rate for DenseNet121 trained on the noisy transformation path

Table 2 Optimum learning rate for epochs

Range of epochs	0–4	5–9	10–14	15–19	20–24	25–29
Learning rate $*10^{-4}$	2.9	0.015	0.029	0.0063	0.0063	0.0075

numerical gradient for every fifth epoch. The average training time per epoch was maintained at 47 s.

5 Results

Over the 25 epochs of training, there are a very few exploding gradients and almost no vanishing gradients. Figure 4 plots the training and validation loss of the DenseNet121 model over 25 epochs of training on the standard transformation data.

The validation accuracy remains within 90–95% for the first three epochs. After epoch 10, it does not become lower than 96%. The validation accuracy versus epoch curve is plotted in Fig. 5. On training the model on the standard transformation data (non-jitter data), a test time augmentation (TTA) accuracy on the validation set of 97.66% was obtained as given in Table 3. A recall of 0.9765, a precision of 0.9772, and an F1 score of 0.9767 were obtained on the validation set images of non-jitter data. The accuracy on the test set which also had 384 images (non-jitter) was 98.4536%. In other words, only 6 examples out of 384 were misclassified. The precision, recall, and F1 score for these models were 0.984811, 0.98453, and 0.98460, respectively, as given in Table 3. The confusion matrix for the test set was plotted as shown in Fig. 6.

For the noisy data, DenseNet121 is retrained, with the same approach of transfer learning. The model is trained for 30 epochs tuning the learning rate every five epochs as given in Table 2. Test time augmentation (TTA) accuracy on the validation set of 94.79% is given in Table 3. A recall of 0.94791, a precision of 0.94868, and an F1 score of 0.948260 were obtained on the validation set images for DenseNet121

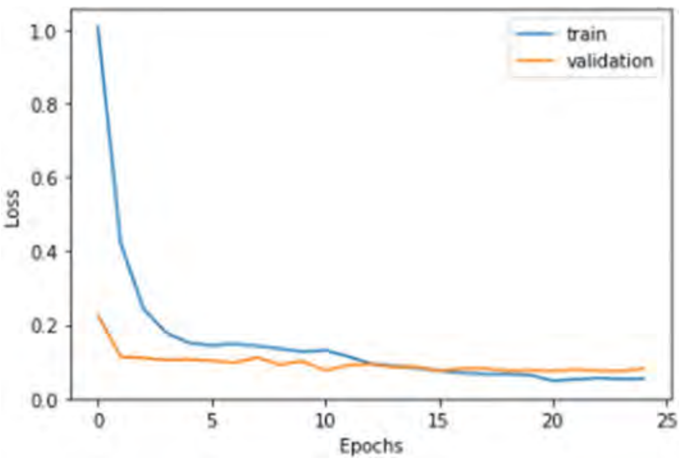


Fig. 4 DenseNet121: training on standard transformation data

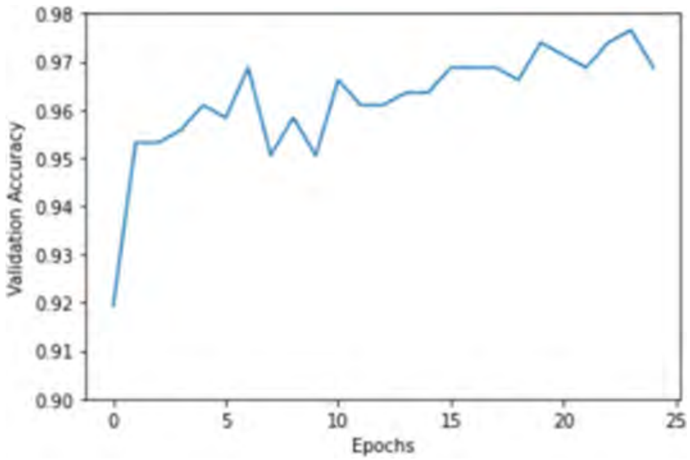


Fig. 5 Validation accuracy versus epoch curve for DenseNet on standard transformation data

Table 3 Evaluation of DenseNet

Evaluation metrics	Standard transformation	Noisy transformation
TTA accuracy on the validation set	97.66%	94.79%
F-1 score on the validation set	0.9767	0.9482
Accuracy on the test set	98.45%	91.49%
F-1 score on the test set	0.9846	0.9141

trained on jitter data. The metrics plotted 98.9 in Fig. 7 show that the validation accuracy oscillates between 92 and 93% for the larger part of the 30 epochs. The accuracy on the test set which also had 384 images was 91.49%. In other words, 33 samples out of 384 were misclassified. The precision, recall, and F1 score for these models were 0.913470, 0.914948, and 0.9141031, respectively, as given in Table 3. The confusion matrix for the test set was plotted as shown in Fig. 8.

6 Conclusion

The DenseNet121 is implemented successfully to infer the model that yielded good results. DenseNet121 was chosen as baseline model because of its low storage requirements and densely connected architecture, which reuses all feature maps instead of generating new ones. Initially, preprocessing was applied to the dataset,

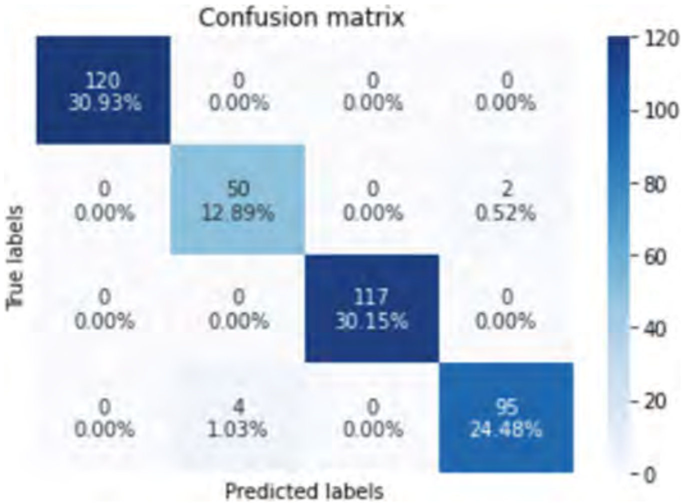


Fig. 6 Confusion matrix plot for DenseNet121 on standard transformation data

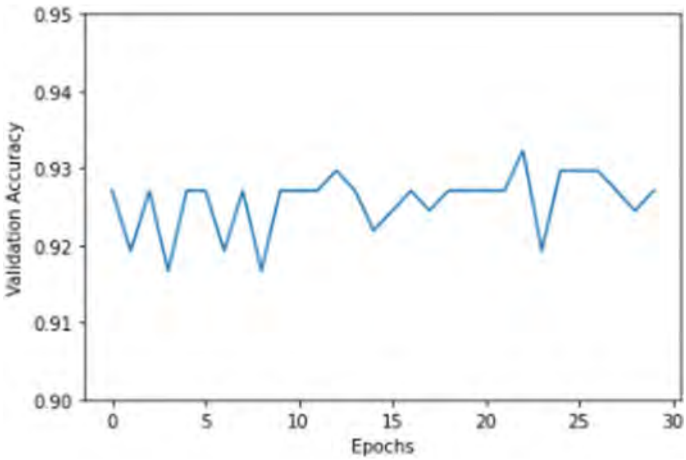


Fig. 7 Validation accuracy versus epoch curve for DenseNet121 on noisy data

and later, it was fed to the model for training. DenseNet121 gave nearly perfect accuracy for all four classes. The learning rate was optimized for various ranges of epochs. An accuracy of 98.45% was obtained on the test set. To check the robust nature of the model, jitter (noise) was added to some extent on the dataset images. The proposed model yielded an accuracy of more than 91%. This result proved that even with noise and with low-resolution images, the model classified quite well. It was concluded that the proposed model was able to achieve near human-level performance. In the future, research may be furthered by using a much larger dataset with

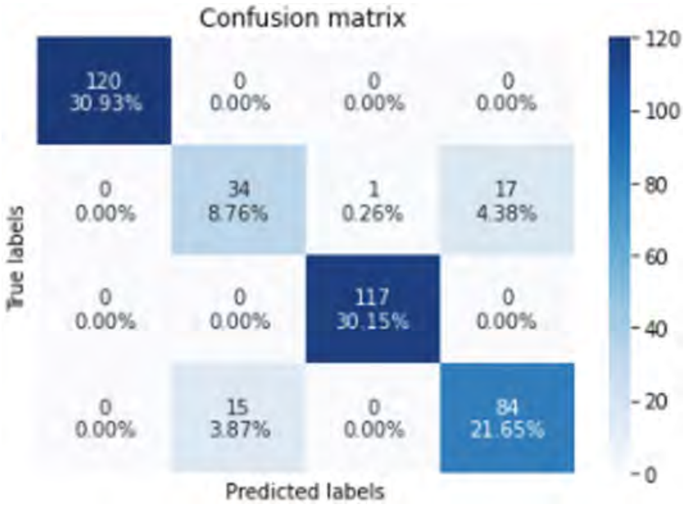


Fig. 8 Confusion matrix plot for DenseNet121 on noisy data

more categories and test more models on the datasets to find the optimal solution to the problem which exists in the developing nations.

References

1. S. Kumar, B. Raghavendra, Disease detection of various plant leaf using image processing techniques. in *International Conference on Advanced Computing and Communication Systems*(2019). https://doi.org/10.1109/ICCUBEA.2015.153
2. T. Tan, Z. Li, H. Liu, F. Zanjani, Q. Ouyang, Y. Tang, Z. Hu, Q. Li, Optimize transfer learning for lung diseases in bronchoscopy using new concept: sequential Fine-Tuning. IEEE J. Transitional Eng. Health Med. (2018).https://doi.org/10.1109/JTEHM.2018.2865787
3. K. Panigrahi, H. Das, A. Sahoo, S. Moharana, Maize Leaf detection and classification using machine learning algorithms. Advances in intelligent systems and computing, vol. 1119 (2020).https://doi.org/10.1007/978-981-15-2414-1_66
4. J. Shah, H. Prajapati, V. Dabhi, A survey on detection and classification of rice plant diseases. in *IEEE International Conference* (2016). https://doi.org/10.1109/ICCTAC.2016.7567333
5. P. Thakur, P. Aggarwal, M. Juneja, Contagious disease detection in cereals crops and classification as solid or undesirable: an application of pattern recognition, image processing and machine learning algorithms. Int. J. Eng. Technol. (2018). https://doi.org/10.14419/ijet.v7i1.2.9043
6. S. Shakya, Analysis of artificial intelligence based image classification techniques. J. Innov. Image Process. **2**, 44–54 (2020). https://doi.org/10.36548/jiip.2020.1.005
7. J. Chen, L. Qi, I. Gao, Visual tea leaf disease recognition using a convolutional neural network model. Symmetry **11**, 343 (2019). https://doi.org/10.3390/sym11030343
8. S. Mohanty, D. Hughes, M. Salathe, Using deep learning for image-based plant disease detection. Front. Plant Sci. (2016). https://doi.org/10.3389/flps.2016.01419

9. S. Mishra, R. Sachan, D. Rajpal, Deep convolutional neural network based detection system for real-time corn plant disease recognition. in *International conference on Computational Intelligence and Data Science* (2019). https://doi.org/10.1016/j.procs.2020.03.236
10. M. Lv, G. Zhou, M. He, A. Chen, W. Zhang, Y. Hu, Maize leaf disease identification based on feature enhancement and DMS-robust AlexNet. IEEE Access (2020).https://doi.org/10.1109/ACCESS.2020.2982443
11. R. Krishnaswamy, R.P. Aravind, Automated disease classification in (Selected) agricultural crops using transfer learning. Automatika. **61**, 260–272 (2020). https://doi.org/10.1080/00051144.2020.1728911
12. T. Vijaykumar, R. Vinothkanna, Mellowness detection of dragon fruit using deep learning strategy. J. Innov. Image Process. **2**, 35–43 (2020). https://doi.org/10.36548/jiip.2020.1.004
13. G. Huang, Z. Liu, L. Maaten, K. Weinberger, Densely Connected Convolutional Networks. Cornell University paper (2018). arXiv:1608.06993v5
14. J. Irvin et al., CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison. Proc. AAAI Conf. Artif. Intell. **33**, 590–597 (2019). https://doi.org/10.1609/aaai.v33i01.3301590

Predicting Embryo Viability to Improve the Success Rate of Implantation in IVF Procedure: An AI-Based Prospective Cohort Study

Dhruvilsinh Jhala, Sumantra Ghosh, Aaditya Pathak, and Deepti Barhate

Abstract In general, infertility affects one in the seven couples across the globe. Therefore, an innovative and beneficial procedure is used to fertilize an egg outside the human body with the help of in vitro fertilization (IVF) procedure. IVF is considered as the most common procedure, as it accounts for 99% of the infertility procedures. From being the most widely used procedure, its success rate for women under 35 is 39.6%, and above 40 is 11.5% depending on the factors like age, previous pregnancy, previous miscarriages, BMI and lifestyle. However, human embryos are complex by nature, and some aspects of their development are still remaining as a mystery for biologists. Embryologists will subjectively evaluate an embryo and its efficiency by making their observations manually during the embryo division process. Since these embryos are dividing rapidly, the manual evaluations are more prone to error. This paper gives a brief explanation and insights into the topic of evaluation and the success rate prediction by using artificial intelligence techniques.

Keywords In vitro fertilization (IVF) · Embryo · Blastocyst · Artificial intelligence · Machine learning · Deep learning · Infertility

1 Introduction

Infertility is a reproductive system disease that has been described as failures in clinical pregnancy after 12 months or more of regular unprotected sexual intercourse according to the World Health Organization (WHO). Infertility has been remaining as the fifth most severe disability in the world for women under 60. The Department of Economic and Social Affairs of the United Nations reports that the global average fertility rate today is just less than 2.5 children per woman. The average rate of fertility has halved in the past 50 years, and over the course of social modernisation, the number of children per wife has also decreased dramatically. In the pre-modern period, fertility rates were normal between 4.5 and 7 children per woman [1]. This

D. Jhala (✉) · S. Ghosh · A. Pathak · D. Barhate
Narsee Monjee Institute of Management Studies, Mumbai, Maharashtra, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_32

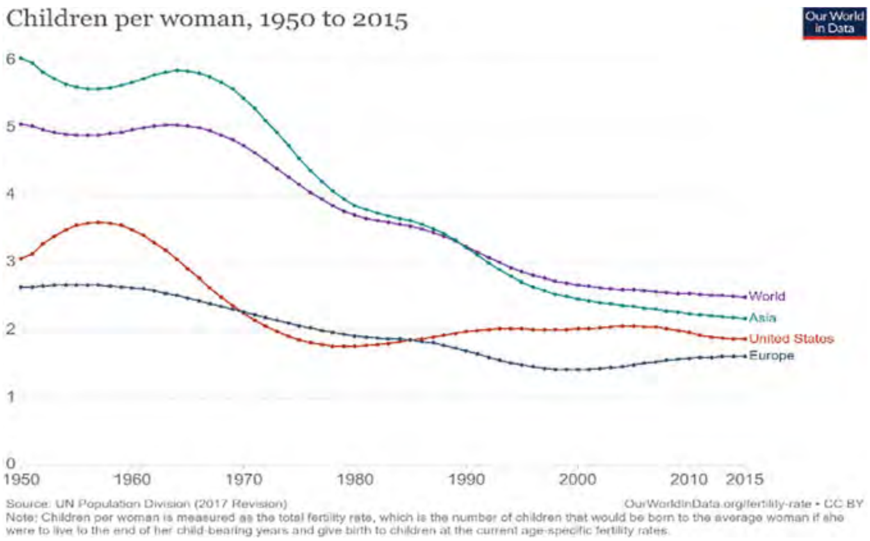


Fig. 1 (1950–2015) infertility stats by WHO [2]

is why in vitro fertilization is widely used today to overcome this challenge. In vitro fertilization (IVF) is a complex set of procedures that are used to help with pregnancy or avoid genetic defects and to assist with a child's conception (Fig. 1).

In vitro fertilization techniques have improved significantly since the first baby was born in 1978. But the research related to test-tube babies started much before. In the 1800s, Dr Sims at the Women's Hospital, New York, conducted a new intrauterine insemination as soon as the idea of reproduction became apparent. The first donor insemination was then done in 1884 by Dr William Pancoast in Philadelphia. This was a beginning to one of the first techniques related to IVF. In the 1900s, first clinic was open in Massachusetts, and after that, in a short span of time, IVF has progressed more than any other medical area [3].

In the early period of the IVF, multiple embryos per cycle were the principal strategy for increasing IVF success. Over the years, the transfer of many embryos has increased the success rates of IVF at the risk of increased multiple pregnancies. Increased numbers and complications of multiple births lead society to apply policies limiting the number of embryos being transferred. The key realistic strategy for maximizing the effects of the IVF was therefore to pick embryos with the highest transfer potential [4].

To identify the best embryo for the process, some methods are used to automate this selection saving both time and money. The methods used for this analysis are the usage of machine learning algorithms on the day 5 image of embryos after implantation of sperm [5]; another is to extract the features from the image to find the timing of pronuclear breakdown (PNB) [6]. In recent years, neural networks are used to improve the chances of pregnancy by separating the good embryos that have more possibility of cousin pregnancy from the bad ones [7]. Another emerging trend is to upload the embryo image directly into an online interface, and the online system grades the embryo image into good, ok and bad. This is an automated system which takes into consideration 24 major embryo characteristics while grading. This grading is done passing the image into three different Artificial Neural Network (ANN) [8]. The grading done by the software and professional are ought to have an agreement of 22%, which is very low. The online grading software can be made more efficient by [9] scaling the image as the size of image taken from smartphone is smaller than the one manually analysed, and using this method, the agreement between system and professional jumps to 85.7% while the online system is able to successfully segment 77.8% of the overall embryo images.

During the entire time from egg retrieval to transferring selected embryos to the woman's uterus for various examinations and monitoring procedures, the specimen is transferred to or from the IVF incubator. The embryo is in the incubator for about 5 days before it is transferred to the womb. The incubation stage is therefore crucial to the success of the IVF [10]. Chung et al. [11] shows the use of the digital microfluid (DMF) device. In this study, the DMF method has been demonstrated to be biologically consistent and is based on the scattered droplet type for the use of in vitro mouse gametes and embryo cultures. The IVF fertilization rate in DMF was found to be 34.8%, with around 25% inseminated embryos grown on DMF chips growing up into an eight-cell process. Tzeng et al. [12] helps us understand incubation devices, oocytes-zonal removal, viscosity and overall transfer channel structure.

2 Stages in IVF Procedure

See Fig. 2.

Step 1: Control Ovarian Hyperstimulation (COH)

In this step, the gonadotropin hormone generation is controlled to prevent premature ovulation. Once suppressed fully, ovulation is achieved under a controlled environment by monitoring the gonadotropin injections. Ultrasound and hormone evaluation is done until the eggs obtain the optimum size.

Fig. 2 Stages in IVF procedure

Step 2: Egg Retrieval

Eggs are retrieved from patients using surgery. Ovarian follicles are sucked using a transvaginal ultrasound-guided instrument. The embryologist scans the follicular fluids to find all viable eggs. In recent years, assistant robots are used to retrieve the eggs more safely from the patients. An auto egg detection algorithm is also used in this auto asset to find the location of the egg more accurately [13].

Step 3: Fertilization and Embryo Culture

The parameters of sperm are checked in this stage. If the parameters are normal, then 50,000–100,000 motile sperm are transferred. This process is called standard insemination. If the sperm parameters are abnormal, then a technique called intracytoplasmic sperm injection (ICSI) is used for sperm transplants in which the transfer of sperm is done by a specialist under a high-powered microscope. The spermatozoa are directly inserted into egg cytoplasm. Nowadays, research is being conducted to automate the ICSI process using logical regression and neural networks. The research takes the patient characteristics into consideration while performing regression or creating neural networks. The ICSI outcome predicted by logical regression is 75% and neural network is 79.5% [14].

Step 4: Embryo Quality

There are many criteria for classifying the quality of embryos. The embryos are examined by the embryologist early morning on the day of the transfer. The embryos are transferred usually at either on day 3 (cleavage stage) or day 5 (blastocyst stage).

5 Day Blastocyst Process

- Blastocyst initial stage–Normally, fertilized embryos show two pronuclei (PNs) in their centre.
- Blastocyst stage–Most of the embryos will now have divided and will ideally be 2–4 cells. Around 98% of embryos which have shown normal fertilization on day 1 will divide and continue to grow.
- Complete blastocyst–Embryos made another division from the previous day and to ideally have 6–8 cells.
- Prolonged blastocyst–The embryos have more than 8–10 cells and have started to compact, and this is the morula stage (Cells merging together).
- Blastocyst hatched–On day 5, a proportion of the embryos becomes blastocysts (Fig. 3).

Step 5: Embryo Transfer

Embryo transfer takes place either on day 3 or day 5 depending on the previous step. This is an easy step that does not require the patient to use anaesthesia. In this stage, embryos are inserted into a soft catheter and are put under ultrasonic supervision in the uterus of the patient via the cervix.

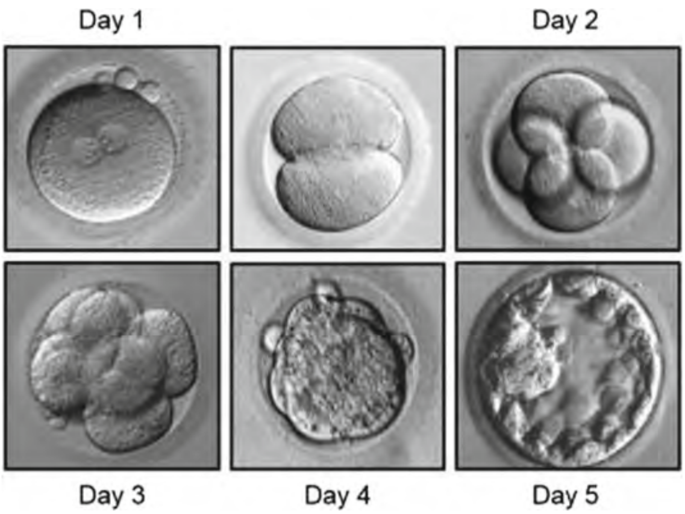


Fig. 3 Blastocyst transition [15]

Step 6: Blood Test

You will have a blood test to assess the level of hormone human chorionic gonadotropin (hCG) approximately 10–15 days after the embryo transmission. hCG in your bloodstream usually means a positive pregnancy test.

3 Embryo Morphology and Grading System

It is been more than 40 years' time since IVF started and since then all fields of assisted reproductive technology had very dynamic developments. There have been rapid improvements in the embryology methods, especially for pre-implantation embryo assessments. After the eggs have been infused with the sperm, they are in process to become embryos. During this time, they are closely watched. The selection of the best embryos starts when the eggs and the sperm have matured for about 2 or 3 days. The embryologist will inspect the embryos carefully and select the healthiest ones. Therefore, the embryo selection procedure is enhanced using a variety of different approaches.

Precise tests of embryos in several days following IVF make it possible to pick the most effective embryos for transfer. This increases the IVF procedure's success rates. The insemination of the best transfer embryo also decreases the frequency of multifetal pregnancies transferred. There are certain features on which the evaluation depends on following:

- Early cleavage.
- Polar body structure and placement.
- Blastomeres dimensions, orientation and fragmentation.
- Appearance of zona pellucida and cytoplasm.
- Pronuclear morphology.
- Number of blastomeres in particular days of culture.
- Compaction and expansion of blastomeres.
- Number of nuclei in each blastomere.

3.1 Cleaved Embryos Scoring

This embryo evaluation is made approximately 1 day after insemination when the presence of first cleavage, blastomere consistency and the degree of fragmentation is investigated. Healthy embryos consist of two symmetrical blastomeres with little or insignificant fragmentation and should have at least four cells on day 2 and at least eight cells on day 3 of culture (Fig. 4).

If there are many blastomeres and little or small fragmentation in the good quality embryo, the good blastomere number on day two is 4–6 and on day three 8–12. For example, 4A1 on the day 1 and 8A1 on day 3 are the best-performing embryos [16].

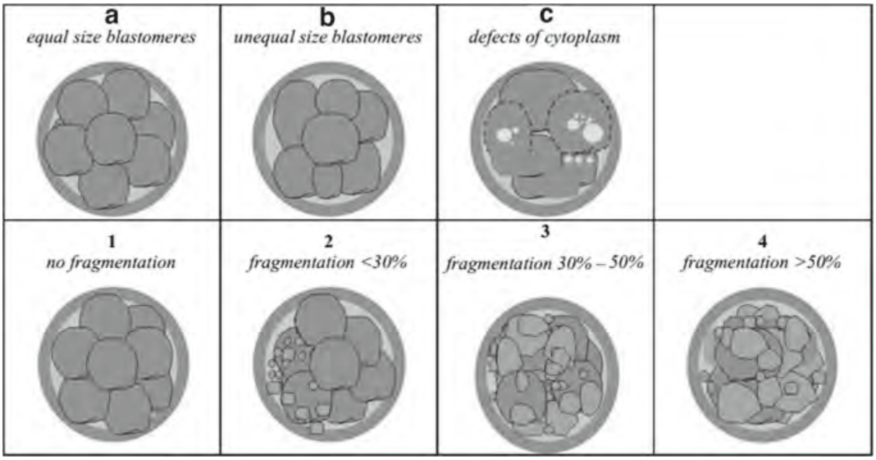


Fig. 4 Cleaving embryo scoring [16]

3.2 Blastocyst Based Scoring

Before being moved back into the uterus, blastocyst represents the ultimate stages of clinical embryonic culture and is thus the final step of morphological evaluation. The blastocyst is characterized by cavitation and blastocoel formation that correlate with the cell distinguishment between internal cell mass (ICM) and trophectoderm (TE). Any static morphological assessment is poor because blastocyst which is highly complex, and the assessment can be dramatically altered even in a short interval of time [17]. Despite the complexity of these and other grading systems, no convention exists for assigning numeric scores or for what represents a high or a low grade [18].

4 Classification of Embryos Using AI

See Table 1.

4.1 Single-Day Embryo Structure-Based Classification

There are certain ways of using artificial intelligence (AI)-based models to predict human embryo viability. One important method is to use the static images captured from an optical light microscope which works on cell distinguishment between internal cell mass (ICM) and trophectoderm (TE). VerMilyea et al. [19] performed image analysis on single static images of day 5 blastocysts using the life whisperer model based on deep learning and got an accuracy of 64.3%. The mentioned endpoint is pregnancy as measured by foetal heartbeat, and this does not mean the possibility of a living pregnancy and is limited to day 5 itself. A more accurate solution was given by [4, 33]. They used deep convolutional neural networks for segmentation of blastocyst and implantation prediction accuracy. The architecture used was compact contextualize calibrate (C3) and GoogleNet Inception v3 with mean accuracies of 70.9% and 89.01%, respectively, with only a single blastocyst image. In their research, Moradi et al. proposed a slow-fusion strategy to learn cross-modality features, and Bormann et al. replaced the final classification layer by retraining it with a dataset of 1282 embryo images captured at 115 h post-insemination.

Chen et al. used approximately 1.7 lakh images from day 5 and day 6 to create a CNN model with ResNet50 architecture. The output was classification of inner cell mass, blastocyst, and trophectoderm quality using a three-point grading system. For all three grading groups, the results showed an average predictive accuracy of 75.36%, i.e. 96.24% for blastocyst growth, 84.42% for Trophectoderm and 91.07% for ICM [42]. Similar type of research was done by [32], and it was based on a set of Levenberg–Marquardt's neural networks and uses binary patterns present in the

Table 1 Image classification embryo datasets

Reference	Data type	Data samples	AI method used	Accuracy
VerMilyea et al. [19]	Day 5	8886 images	Life whisperer	64.3%
Arsal et al. [20]	Time lapse	38 embryos	conditional random field (CRF), bag of visual words (BoVW)	96.79%
Sammali [21]	Ultrasound strain imaging	16 patients	SVM, KNN and GMM	93.8%
Sujata et al. [22]	Time lapse	535 images	Convolutional neural network	87.5%
Moradi et al. [4]	Day 5	578 blastocyst images	Deep convolutional neural networks (DCNNs), dense progressive sub-pixel upsampling (DPSU)	70 0.9%
Patil et al. [23]	Day 3	–	Hessian-based ridge detector and Hough circle transform	–
Kheradmand [24]	Day 5	8460 images augmented from 235 images	Fully Convolutional Networks (FCNs)	95.6%
Amarjot Singh [25]	Day 1 and 2	40 embryo images	Hoffman modulation contrast (HMC), Hessian-based edge detector	80%
Zhang et al. [26]	Patient details	11,190 patients	Clustering and SVM in each cluster	70%
Khayrul Bashar et al. [5]	Timelapse	10 embryos	Supervised classifier using normalized cross-correlation	
Mölder et al. [6]	Time lapse	20 embryos	Hough detection	95%
Habibie et al. [27]	Time lapse	68 images	Particle swarm optimization (PSO)-based Hough transform	6.24% error

(continued)

Table 1 (continued)

Reference	Data type	Data samples	AI method used	Accuracy
Uyar et al. [28]	Day 5	7745 embryo images	3 Bayesian networks	63.5% TPR & 33.8% FNR
Kragh et al. [29]	Time lapse	6957 images	CNN and RNN	65.5% and 72.7%
Khosravi et al. [30]	Time lapse	50,000 images	Google's inception and DNN	98%
Uyar et al. [31]	Day 2–3	3898 images	Naive Bayes, KNN, decision tree, SVM, MLP and RBF	80.4%
Manna et al. [32]	Day 2–3	269 images	Neural Network and binary patterns	AUC = 0.8
Bormann et al. [33]	Day 5	182 images	CNN	89.01%
Chen et al. (2019)	Day 5–6	1.7 lakh images	CNN	75.36%
Wang et al. [34]	Time lapse	389 videos	Three-level classification	87.92%
Chung et al. [11]	Day 3	–	Digital microfluidic (DMF) system	25%
Storr et al. [7]	Time lapse and day 5	380 images	General estimated regression models, multivariable regression model	8 parameters, AUC 0.748
Kotze et al. [35]	Day 3 image, patient details	770 image	GES-score plus sHLA-G expression	~ 60.0%
Chiang et al. [36]	day 3, patient details	47 patients	Ultrasound, basal uterine perfusion, colour	Increase in uterine (best factor)
Durairaj et al. [31]	Patient details	250 patients	ANN	73%
Uyar et al. [37]	Patient details	2453 records	KNN, SVM, decision tree (DT), naive Bayes, MLP and RBF	NB(0.739) and RBF(0.712)
Azmoudeh et al. [38]	Patient details	160 patients	Anti-mullerian hormone (AMH)	–

(continued)

Table 1 (continued)

Reference	Data type	Data samples	AI method used	Accuracy
Qiu et al. [39]	Patient details	7188 records	SVM, logistic regression, XGBoost and random forest	0.70 ± 0.003
Fábio et al. [9]	Mobile-captured image evaluation	18 embryo images, mobile captures images	Blast3Q, image segmentation	87.5%
Bertogna et al. [8]	Mobile-captured image evaluation	18 images, mobile captures images	Blast3Q, Image segmentation	22% common result
Gowramma et al. [40]	Patient details	1,21,744 records	Data analysis	–
Durairaj et al. [41]	Patient details	27 test factors	Clustering, ANN	90%

images. Area under the curve (AUC) for this study was found to be approximately 0.8 which is good enough even after using a small dataset of 269 embryos.

The selection of a model for evaluation of data is an important part of the process, but the pre-processing and preparation of data for the selected model is as important as the model. Uchida et al. [43] discusses the selection and optimization of CNN and image filter so as to improve the accuracy. They perform experiments using four different models: no filter, all filter, random filter and joint optimization. The CNN model used consists of LGG-Net. The results of the experiment convey that redundant image pre-processing can lead to adverse results, but selection of appropriate can improve the output. The accuracy obtained is 78% through joint optimization which is 9% more than using no filter. Similarly, instead of using different filters, [37] using different models to compare embryos to get the highest chance of pregnancy. The six different algorithms used were naive Bayes, K-nearest neighbour, decision tree, support vector machine, multi-layer perceptron and radial basis function network. The best result was given by naive Bayes with accuracy of 80.4%. Uyar et al. [44] used only SVM for classification upon 546 embryos and got an accuracy of 82.7%

Kotze et al. followed a different approach in his study by using GES-score. He compared selecting embryos examining day 3 images of IVF V/S GES-score plus sHLA-G expression and the effect of both the methods on implantation, miscarriage and pregnancy. The research found that GES-score plus sHLA-G expression gave less chance of miscarriages but implantation rates did not differ much [35].

4.2 Time Series Embryo Structure Based Classification

Another approach to evaluate the embryo quality is to observe the embryo as it develops through day 1 to day 5 and then use this observation to predict if embryo implantation will result in successful pregnancy. This approach gives more favourable results as embryo day 5 images have constraints in image capturing like timing and light intensity. Thus, a time-lapse image gives better results. Patil et al. had similar thoughts when they proposed a CNN model that took 350 images scaled to 256 × 256 pixels [22]. The accuracy of 87.5% is observed in initial stages. Similarly, RNN and CNN models were independently trained using 6957 time-lapse data with a single ICM and TE annotation each, with precision levels of 65.5% and 72.7%, respectively, [29]. A slightly varied approach was by Liu et al. where they propose a multi-tasking deep learning with dynamic programming model for auto classification embryo in time-lapse video [26]. They discuss CNN (ResNet) with one to one, one to many and many to one approval of model. The results received recommend one to many as many to many is computation heavy with accuracy of 86%. [45] discusses TLI and its use in embryo assessment. It also discusses a predictive algorithm for TLI, and it gives accuracy as specificity of 84.7%, sensitivity of 38.0%, PPV 54.7% and NPV 73.7%.

Another different research direction which was discussed in [30] was a model named STORK and was created based on Google's inception model which is basically composed of deep neural networks. Approximately, 50,000 time-lapse images of embryos were passed through this DNN to find high-quality embryos. STORK can predict embryo development cycle with an accuracy of more than 98%. They further improved their system by including a decision tree into STORK which helps the system to predict likelihood of pregnancy based on age and embryo quality.

In the paper [34], the three-level classification is discussed on time-lapse video of embryo development to classify the embryo. Tracking of an individual embryo cell is avoided in this research, and the classifier can easily adapt to image modularities and the problem of mitosis detection. The accuracy obtained by using multi-level classifiers was 87.92% by using 389 embryo development videos. Similarly, Uyar et al. correctly predicted the development cycle of embryos using three different Bayesian networks [31]. The first network layer had day 5 embryo image scores derived from day 3 data of embryo, the second layer contained the characteristics of the patient along with day 5 score and the third layer combined all the data collected from day 1, 2 and 3 into day 5 score. Tenfold cross-validation was used to remove the bias from 7745 embryo images along with patient records including embryo morphological characteristics. The resulting experimental data produced 63.5% true positive rate (TPR) and 33.8% false negative rate (FNR) while predicting blastocyst development.

Storr et al. used 380 images of day 5 blastocysts to evaluate whether time-lapse parameters contributed any major difference while predicting the quality of embryo. Embryo morphology evaluation was done on day 2, day 3 for symmetry fragmentation percentage and blastomeres numbers. Embryos were graded on day 5 on the basis of blastocoel cavity expansion and the number and unity of both the trophectoderm and cells inner cell mass (ICM). Eight important factors were identified using single and time variate regression models which are time durations for: division into 5-cell and then division into 8-cell, division into a 6-cell embryo and cleavage into an 8-cell embryo, full compaction, initial blastulation signs and complete blastocyst [7]. Nogueira et al. [46] used 394 time-lapse images of the human embryo taken at 110hpi, rated for inner cell mass (ICM), trophectoderm and expansion, with overall accuracy of 78.8, 78.3 and 81.5%, respectively.

4.3 Classification Based on Cell Count (Ellipse Detection)

As discussed earlier in the embryo observation process, owing to the fact of the dividing natural state of embryos, it is vulnerable to error and inaccuracy resulting from human misinterpretation. The cell number is the most significant predictor of future growth, as it can directly demonstrate the ability of an embryo to progress cell cycles. Studies show that the developmental ability of embryos with lower or higher cell numbers was substantially reduced. Day 3 human embryos with good potential for growth can typically mature into the 7–8 cell stage. Thus, by using circle or ellipse detection techniques, the number of cells can be counted (Fig. 5).

Arsa et al. used 25 patients and 38 pieces of embryo images to give an effective method to solve the prediction problem with the help of SVM, bag of visual words (BoVW) and conditional random field (CRF). They achieved an average accuracy of 80% [20]. Another study reports strategies for enhancing embryonic identification. Hough circle transformation is used to detect the presence and orientation of embryos within the acquired boundaries [23]. This method is highly accurate to get the number of cells in the blastomere. A different approach was proposed, based on the modifying of Hough transform using particle swarm optimization (PSO), to approximate the embryo as the circle. The increasing PSO particles are a period in parameter space

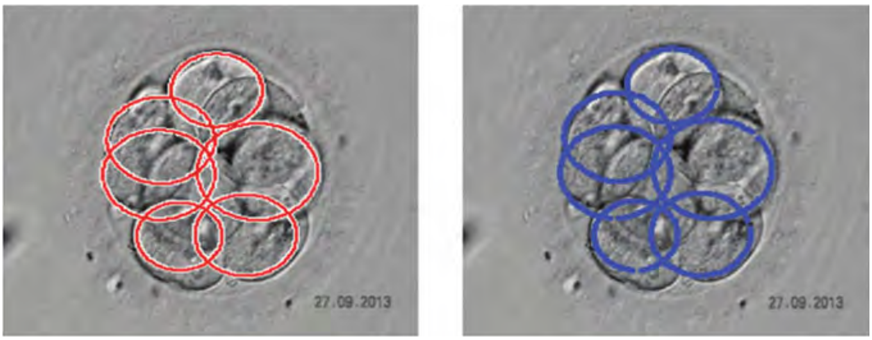


Fig. 5 Ellipse detection and verification [23]

and are used primarily to minimize Hough transform's computational complexity [27].

Embryo image analysis is a critical step in IVF, but because of the sensitivity of the embryo, the image taken is in an unstable environment, amounting to the problem of high noise in the image. To solve this issue, the extraction of features is used by various researchers to extract essential components. One method is to extract only two features, i.e. nuclear numbers and nuclear volumes in a sequential manner [5]. Automated methods were used in this method to obtain nuclear numbers while segmentation was used by the researchers to obtain nuclear volume. Constant increase of these two features during the embryo development cycle was taken as indication of a healthy embryo. The extraction of images without high noise is a difficult process so another different approach to analyse images is to study the embryo images in high levels of noise. This was done by researchers by extracting the images in circular format using Hough transformation [6]. Although this process was done under noisy conditions, the accuracy of locating embryo position was 92.9%. Hough transformation was used in this method to detect the timing of pronuclear breakdown (PNB) with 83.0% accuracy.

The shape of cell formation in the embryo is directly related to its quality and vitality. Jonaitis et al. [47] presents an idea for the detection of the embryos as well as the number of cells together by using time-lapse microscopy. Singh et al. [25] mainly proposes an algorithm for segmentation of up to four blastomeres in a single human HMC embryo image. The algorithm performs under conditions such as fragmentation, cellular and side-lit illumination of cells. Experimental findings show that the algorithm proposed is able to classify and model blastomeres in HMC images with an average accuracy of 80%.

Sammali et al. provide an overview of the uterine motion analysis and the characteristics with the help of ultrasonic imaging. They used KNN, SVM and GMM, where KNN had the best accuracy of 93.8% in providing the optimal frequency-related characteristic feature set for embryo implantation [21]. They used 4-min TVUS scans that were taken at 5.6-MHz central frequency and 30 frames/s. Similarly, Fernandez et al. discussed several algorithms for improvement of f assisted reproductive technology (ART) such as SVM, ANN, Bayesian networks and convolution neural networks (CNN). They found CNN architecture was more suitable than an MLP for classifying images. But CNN type of AI has a higher computational cost than MLP. The results showed that naïve Bayes gave the best result, and the SVM was the fourth-best area under the curve (AUC) [48].

Inner cell mass (ICM) is an important factor in determining viability of an embryo. Kheradmand et al. discussed the use of fully connected deep convolution neural network (DCNN) trained on 8460 images augmented from 235 images. They proposed a two-stage pipeline pre-processing step, a data augmentation technique. The method proposed gave a high accuracy of 95.6% [24].

4.4 Classification Based on Patient Characteristics

There is no reliable process to predict the success rate of an IVF procedure, and the success rate varies from patient to patient. The effectiveness of the specific treatment is influenced by a variety of factors, such as male and women, and different IVF test results. In assessing the effectiveness of treatment, even the pair's psychological factors play an important role, while multiple treatment cycles increase the cost and health of the patients as well as the level of stress. Hence, data like endometriosis, tubal factors and follicles in the ovaries and the physiological factors such as stress level factors can be used for reference. A research by Gowramma et al. discusses many intrinsic and extrinsic factors that affect success rate of IVF. The intrinsic factors analysed were the age of the patient, BMI, visibility of embryo, quality of sperm, genetic predisposition, hormonal balance and endometriosis. The extrinsic factors which were current medical technology, treatment method used, experience of the professionals, stress and process time over the data collected from 1,21,744 women [40].

Before starting to model over the key features, pre-processing is an essential part. Hence, an approach to cluster the data and use Johnson's reducer to find influential factors which are fed as an input to artificial neural networks is effective [41]. This process gave 90% similarity by comparing experimental results to actual results over a dataset containing 27 different test factors of 250 patients. Among these, various factors Hassan et al. worked to narrow down the key factors. The number of highly influential attributes was reduced to 19 for MLP, 16 on RF, 17 for SVM, 12 on C4.5 and eight in CART [49] by the algorithm for selection of features. Overall, age, fertility factor indices, antral follicle counts, NbreM2, sperm collection process, chamotte, in vitro fertilization rate, follicles on day 14 and embryo transfer day were the most influent defined characteristics.

One of the earlier studies was done by Durairaj et al. [50] by implementing ANN model over a dataset with 27 attributes, consisting of features like BMI, previous pregnancy, miscarrage, sperm vitality, etc. This work shows 73% of accuracy in the results. Here, the amount of data plays an important role, as it is implemented on a small dataset with 250 samples, and the accuracy is also low as embryo is not analysed and based purely on theoretical data. Later on, among 11,190 patients, [51] and conducted related research with additional features such as AFC, AMH, FSH and five pathogenic factors. The technique used was first to divide patients into different groups and then create a model for each group of SVM to achieve best overall efficiency. Similar research was performed by Qiu et al. [39], where most accurate model was found to be XGBoost which created 0.73 under the ROC curve, and it also gave an accuracy score of 0.70 ± 0.003 when cross-validated with a dataset of 7188 patients. This research helped in reduction of cost and time of pregnancy.

A research by Chiang et al. found that an increase in uterine perfusion also increased the chances of pregnancy for women having age 40 or more [36]. The results have been calculated by taking into account various factors such as day 3 FSH, patient age, number of antral follicles, artery PI in the basal uterine, day 3 estradiol (E2), endometrial thickness on the day that hCG was given, level 2 of the artery and total number of gonadal therapy ampoules. Later on, Azmoudeh et al. found a significant relationship between IVF success and anti-mullerian hormone (AMH) [38]. Logistic regression analysis showed that only AMH > 0.6 was an independent predictor of IVF success.

One other way is to use patient characteristics as well as image classification on the embryo to get a more advanced prediction. Uyar et al. [31] implemented six different classifiers on the dataset consisting of 2453 images out of which 89% are positive and 11% negative. The dataset had 18 features that characterize the embryo like women age, infertility category, infertility factor, sperm quality, early cleavage morphology, number of cells, etc. The study found that naive Bayes (0.739 ± 0.036) and radial basis function (0.712 ± 0.036) gave the best ROC curves. The only shortcoming of the research was data imbalance which had an effect on final results.

5 Conclusion

The proposed research work has discussed about four different approaches (single day embryo evaluation, time-lapse embryo evaluation, ellipse detection and patient characteristics) to predict the IVF embryo viability. Single day embryo evaluation and time-lapse embryo evaluation both are based on image processing, but more precise selection process can be carried out by using time-lapse imaging as morphological evaluation may vary after distinct time. For ellipse detection technique, Hough transform is commonly used, and the models are able to achieve great accuracies as the number of cells is directly related to embryo efficiency. Last technique discussed was using patient characteristics, which can be used to predict pregnancy outcomes. This method does not take in account the mistakes or any faults in IVF procedure itself. According to analysis, CNN and SVM were found to give the most accurate solutions. Hence, by using these AI algorithms, the accuracy can be improved as high as 89.01% for single day [33], 96.79% for time-lapse imaging [20] and 73% with patient characteristics [31].

To our best knowledge, this is the first study on discussing different AI techniques to predict the embryo viability. The predictions can help the embryologist make optimal decisions based on results obtained from these methods. This will reduce the patient's cost, time to pregnancy and improve her quality of life.

References

1. IVF Success Rates Increase Using PGD, https://www.fertility-docs.com/programs-and-services/pgd-screening/pgd-increases-ivf-success-rates.php
2. Fertility Rate https://ourworldindata.org/fertility-rate
3. J.J. Wade, V. MacLachlan, G. Kovacs, The success rate of IVF has significantly improved over the last decade. ANZJOG **55**(5), 473–476 (2015)

4. R.M. Rad, P. Saeedi, J. Au, J. Havelock, Predicting human embryos' implantation outcome from a single blastocyst image, May 19 ©2019, (IEEE, 2019)
5. M.K. Bashar, H. Yoshida, K. Yamagata, Embryo quality analysis from four dimensional microscopy images: a preliminary study. in *2014 IEEE Conference on Biomedical Engineering and Sciences*, 8–10 December 2014, (Miri, Sarawak, Malaysia, 2014)
6. A. M¨older, S. Czanner, N. Costen, G. Hartshorne, Automatic detection of embryo location in medical imaging using trigonometric rotation for noise reduction. in *2014 22nd International Conference on Pattern Recognition* (2014)
7. A. Storr, C.A. Venetis, S. Cooke, D. Susetio, S. Kilani, W. Ledger, Morphokinetic parameters using time-lapse technology and day 5 embryo quality: a prospective cohort study. Accepted: 1 July 2015 / Published online: 15 July 2015, (Springer, New York, 2015)
8. V.B. Guilherme, M. Pronunciate, P. Helena dos Santos, D. de Souza Ciniciato, M.B. Takahashi, J.C. Rocha, M.F.G. Nogueira, Distinct sources of a bovine blastocyst digital image do not produce the same classification by a previously trained software using artificial neural network. in *International Conference on Computer-Human Interaction Research and Applications CHIRA 2017: Computer-Human Interaction Research and Applications* (2017), pp. 139–153
9. M.F.G. Nogueira, V.B. Guilherme, M. Pronunciate, P.H. dos Santos, D.L. Bezerra da Silva, J.C. Rocha, Artificial intelligence-based grading quality of bovine blastocyst digital images: direct capture with juxtaposed lenses of smartphone camera and stereo microscope ocular lens. Sensors (2018)
10. K.A.S. Pramuditha, H.P. Hapuarachchi, N.N. Nanayakkara, P.R. Senanayaka, A.C. De Silva , Drawbacks of current IVF incubators and novel minimal embryo stress incubator design. in *2015 IEEE 10th International Conference on Industrial and Information Systems*, ICIIS 2015, Dec. 18–20, (Sri Lanka, 2015)
11. L.-Y. Chung, H.-H. Shen, Y.-H. Chung, C.-C. Chen, C.-H. Hsu, H.-Y. Huang, D.-J. Yao, In vitro dynamic fertilization by using EWOD device. MEMS 2015, 18–22 January (Estoril, Portugal, 2015)
12. Y.-C. Tzeng, Y.-J. Chen, C. Chuan, L.-C. Pan, F.-G. Tseng, Microfluidic devices for aiding in-vitro fertilization technique. in *Proceedings of the 12th IEEE International Conference on Nano/Micro Engineered and Molecular Systems* April 9–12, 2017 (Los Angeles, USA, 2017)
13. J. Lu, Y. Hu,A potential assistant robot for IVF egg retrieval. in *IEEE SoutheastCon 2004. Proceedings*. (Greensboro, North Carolina, USA, 2004), pp. 32–37
14. Z. Abbas, C. Fakih, A. Saad, M. Ayache, Vaginal power doppler parameters as new predictors of intra-cytoplasmic sperm injection outcome. in *2018 International Arab Conference on Information Technology (ACIT)*, (Werdanye, Lebanon, 2018), pp. 1–7
15. IVF Multi-step process, https://fertility.wustl.edu/treatments-services/in-vitro-fertilization/
16. N. Nasiri, P. Eftekhari-Yazdi, An overview of the available methods for morphological scoring of pre-implantation embryos in in vitro fertilization. Cell J. **16**(4), 392–405 (2015)
17. J. Kort, B. Behr, *Traditional Embryo Morphology Evaluation: From the Zygote to the Blastocyst Stage* (In Vitro Fertilization, Springer, Cham, 2019).
18. C. Racowsky, M. Vernon, J. Mayer, G.D. Ball, B. Behr, K.O. Pomeroy, D. Wininger, W. Gibbons, J. Conaghan, J.E. Stern, Standardization of grading embryo morphology. J. Assist. Reprod. Genet. **27**(8), 437–439 (2010)
19. M. VerMilyea, J.M.M. Hall, S.M. Diakiw, A. Johnston,T. Nguyen, D. Perugini, A. Miller, A. Picou, A.P. Murphy, M. Perugini, Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF. Submitted on October 13, 2019; resubmitted on December 23, 2019; Editorial decision on January 16, (2020)
20. D.M.S. Arsa, Aprinaldi, l. Kusuma, A. Bowolaksono, P. Mursanto, B. Wiweko, W. Jatmiko, Prediction the number of blastomere in time-lapse embryo using conditional random field (CRF) method based on bag of visual words (BoVW), ICACSIS (2016)
21. F. Sammali, C. Blank, T.H.G.F. Bakkes, Y. Huang, C. Rabotti, B.C. Schoot, M. Mischi, Prediction of embryo implantation by machine learning based on ultrasound strain imaging. in *2019 IEEE International Ultrasonics Symposium (IUS) Glasgow*, Scotland, October 6–9, (019)

22. S.N. Patil, U.V. Wali, M.K. Swamy, Non-Member IEEE, Selection of single potential embryo to improve the success rate of implantation in IVF procedure using machine learning techniques. in *International Conference on Communication and Signal Processing*, April 4–6 (2019)
23. S.N. Patil, U.V. Wali, M.K. Swamy, Application of vessel enhancement filtering for automated classification of human In-Vitro fertilized (IVF) images. in *2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques* (ICEECCOT) (2016)
24. S. Kheradmand, A. Singh, P. Saeedi, J. Au, J. Havelock, Inner cell mass segmentation in human HMC embryo images using fully convolutional network, Aug 2017-IEEE-ICIP (2017)
25. A. Singh, J. Buonassisi, P. Saeedi, J. Havelock, Automatic blastomere detection in day 1 to day 2 human embryo images using partitioned graphs and ellipsoids. in *ICIP* (2014)
26. Z. Liu, B. Huang, Y. Cui, Y. Xu, B. Zhang, L. Zhu, Y. Wang, L. Jin, D. Wu, Multi-task deep learning with dynamic programming for embryo early development stage classification from time-lapse videos. Received August 3, 2019, Accepted August 22, 2019, Date of Publication August 27 (2019)
27. I. Habibie, A. Bowolaksono, R. Rahmatullah, M.N. Kurniawan, M.I. Tawakal, I.P. Satwika, P. Mursanto, W. Jatmiko, A. Nurhadiyatna, B. Wiweko, A. Wibowo, Automatic detection of embryo using Particle Swarm Optimization based Hough Transform. in *2013 International Symposium on Micro-NanoMechatronics and Human Science*, MHS (2013)
28. A. Uyar, A. Bener, H.N. Ciray, M. Bahceci, Bayesian networks for predicting IVF blastocyst development. in *2010 International Conference on Pattern Recognition* (2010)
29. M.F. Kragh, J. Rimestad, J. Berntsen, H. Karstoft , Automatic grading of human blastocysts from time-lapse imaging. Comput. Biol. Med. **115**, 103494 (2019)
30. P. Khosravi, E. Kazemi, Q. Zhan, J.E. Malmsten, M. Toschi, P. Zisimopoulos, A. Sigaras, S. Lavery, L.A.D. Cooper, C. Hickman, M. Meseguer, Z. Rosenwaks, O. Elemento, N. Zaninovic, I. Hajirasouliha, Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. npj Received: 2 November 2018 Accepted: 1 March (2019)
31. A. Uyar, A. Bener, H.N. Ciray, M. Bahceci, ROC based evaluation and comparison of classifiers for IVF implantation prediction. (Ed.): eHealth 2009, LNICST 27, (2010), pp. 108–111
32. C. Manna, L. Nanni, A. Lumini, S. Pappalardo, Artificial intelligence techniques for embryo and oocyte classification. A 2012, Reproductive Healthcare Ltd. Published by Elsevier Ltd (2012)
33. C.L. Bormann, Thirumalaraju, Kanakasabapathy, R. Gupta, R. Pooniwala, I. Souter, J.Y. Hsu, S.T. Rice, P. Bhowmick, H. Shafiee, Artificial intelligence enabled system for embryo classification and selection based on image analysis. Fertility Sterility **111**(4), e21 (2019) Supplement
34. Y. Wang, F. Moussavi, P. Lorenzen, Automated embryo stage classification in time-lapse microscopy video of early human embryo development, medical image computing and computer-assisted intervention–MICCAI 2013. in *MICCAI 2013.* Lecture Notes in Computer Science, vol. 8150. (Springer, Berlin, Heidelberg, 2013)
35. D.J. Kotze, P. Hansen, L. Keskintepe, E. Snowden, G. Sher, T. Kruger, Embryo selection criteria based on morphology VERSUS the expression of a biochemical marker (sHLA-G) and a graduated embryo score: prediction of pregnancy outcome. Accepted: 23 February 2010 / Published online: 1 April 2010. (Springer Science+Business Media, LLC, 2010)
36. C.-H. Chiang, T.-T. Hsieh, M.-Y. Chang, C.-S. Shiau, H.-C. Hou, J.-J. Hsu, Y.-K. Soong, Prediction of pregnancy rate of in vitro fertilization and embryo transfer in women aged 40 and over with basal uterine artery pulsatility index. J. Assis. Reprod. Genet. **17**(8) (2000)
37. A. Uyar, A. Bener, H.N. Ciray, Predictive modeling of implantation outcome in an in vitro fertilization setting: an application of machine learning methods. Med Decis Making Published Online 19 May (2014)
38. A. Azmoudeh, I.D. Zahra Shahraki, F.-S. Hoseini, F. Akbari-Asbagh, D.-T. Fatemeh, F. Mortazavi, In Vitro Fertilization success and associated factors: a prospective cohort study. Int. J. Women's Health Reprod. Sci. **6**(3), 350–355 (2018)

39. J. Qiu, P. Li, M. Dong, X. Xin, J. Tan, Personalized prediction of live birth prior to the first in vitro fertilization treatment: a machine learning method. Qiu et al. J. Transl. Med. (2019)
40. G.S. Gowramma, S. Nayak, N. Cholli, Intrinsic and extrinsic factors predicting the cumulative outcome of IVF / ICSI treatment. Int. J. Innov. Technol. Explor. Eng. (IJITEE) **9**(2S) (2019)
41. M. Durairaj, R. Nandhakumar, An integrated methodology of artificial neural network and rough set theory for analyzing ivf data. in *2014 International Conference on Intelligent Computing Applications*, (Coimbatore, 2014), pp. 126–129. https://doi.org/10.1109/ICICA.2014.35
42. T.-J. Chen, W.-L. Zheng, C.-H. Liu, I. Huang, H.-H. Lai, M. Liu, Using deep learning with large dataset of microscope images to develop an automated embryo grading system. in *2019 by the Asia Pacific Initiative on Reproduction (ASPIRE) and World Scientific Publishing Co.*, Received 31 January 2019; Accepted 25 February 2019; Published 29 March (2019)
43. K. Uchida, S. Saito, P.D. Pamungkasari, Y. Kawai, I.F. Hanoum, F. H. Juwono, S. Shirakawa, Joint optimization of convolutional neural network and image preprocessing selection for embryo grade prediction in in vitro fertilization. in *ISVC* 2019, (2019), pp. 14–24
44. A. Uyar, H.N. Ciray, A. Bener, M. Bahceci, 3P: personalized pregnancy prediction in IVF treatment process. eHealth 2008, LNICST 1, (2009), pp. 58–65
45. A. Ahlström, A. Campbell, H.J. Ingerslev, K. Kirkegaard, Prediction of embryo viability by morphokinetic evaluation to facilitate single transfer. Springer International Publishing Switzerland (2015)
46. M.F.G. Nogueira, N. Zaninovic, M. Meseguer, C. Hickman, S. Lavery, J.C. Rocha, Using artificial intelligence (AI) and time-lapse to improve human blastocyst morphology evaluation. in *Conference: ESHRE 2018* October 2018, (Barcelona, Spain, 2018)
47. D. Jonaitis, V. Raudonis, A. Lipnickas, Application of computer vision methods in automatic analysis of embryo development. in *IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications* 24–26 September 2015, (Warsaw, Poland, 2015)
48. E.I. Fernandez, A.S. Ferreira, M.H.M. Cecílio, D.S. Chéles, R.C. Milanezi de Souza, M.F.G. Nogueira, J.C. Rocha, Artificial intelligence in the IVF laboratory: overview through the application of different types of algorithms for the classification of reproductive data. Received: 8 March 2020 /Accepted: 3 July 2020 Springer Science+Business Media, LLC, part of Springer Nature (2020)
49. M.R. Hassan, S. Al-Insaif1, M.I. Hossain, J. Kamruzzaman, A machine learning approach for prediction of pregnancy outcome following IVF treatment. Received: 3 July 2017 / Accepted: 24 August 2018 The Natural Computing Applications Forum (2018)
50. M. Durairaj, P. Thamilselvan, Applications of artificial neural network for IVF data analysis and prediction. J. Eng. Comput. Appl. Sci. (JEC&AS) **2**(9) (2013)
51. B. Zhang, Y. Cui, M. Wang, J. Li, L. Jin, D. Wu , In Vitro fertilization (IVF) cumulative pregnancy rate prediction from basic patient characteristics, (IEEE, 2019)
52. C.-T. Lee, H.-Y. Tseng, Y.-T. Jiang, C.-Y. Haung, M.-S. Lee, W. Hsu, Detection of multiple embryo growth factors by bead-based digital microfluidic chip in embryo culture medium. in *Proceedings of the 13th Annual IEEE International Conference on Nano/Micro Engineered and Molecular Systems*, April 22–26, (Singapore, 2018)
53. J. Herna´ndez-Gonza´lez, I. aki Inza, L. Crisol-Ortı´z, M.A. Guembe, M.J. In˜arra, J.A. Lozano , Fitting the data from embryo implantation prediction: learning from label proportions. Statistical Methods in Medical Research The Author(s) (2016)
54. T. Bączkowski, R. Kurzawa, W. Głąbowski, Methods of embryo scoring in in vitro fertilization. Reprod. Biol. **4**(1), 5–22 (2004)
55. The Birth and History of IVF, https://rmanetwork.com/blog/birth-history-ivf/
56. Understanding Embryo Grading, https://www.utahfertility.com/understanding-embryo-grading/

Breast Cancer Detection and Classification Using Improved FLICM Segmentation and Modified SCA Based LLWNN Model

Satyasis Mishra, T. Gopi Krishna, Harish Kalla, V. Ellappan, Dereje Tekilu Aseffa, and Tadesse Hailu Ayane

Abstract Breast cancer death rates are higher due to the low accessibility of early detection technologies. From the medical point of view, mammography diagnostic technology increases are essential in the detection process. This research work proposes a segmentation for each image by using improved Fuzzy Local Information C-Means (FLICM) algorithms and classification by using the novel local linear wavelet neural network (LLWNN-SCA) model. Further, the weights of the LLWNN model is optimized by using the modified Sine Cosine Algorithm (SCA) to improve the performance of the LLWNN algorithm. By applying an improved FLICM algorithm, the segmented images have undergone the process of feature extraction. The statistical features are extracted from the segmented images and fed as input to the SCA based LLWNN model. The improved FLICM segmentation achieves an accuracy of about 99.25%. Classifiers such as Pattern Recognition Neural Network (PRNN), Feed Forward Neural Network (FFWNN), and Generalized Regression Neural Network (GRNN) are also utilized for classification, and comparison results are presented with the proposed SCA-LLWNN model.

Keywords Breast cancer · Mammography · Feed-forward neural network · Sine cosine algorithm · Feed-forward neural network · Generalized regression neural network

1 Introduction

Breast Cancer is a leading reason for death across the globe including the emerging countries. In 2019, nearly more than 150,000 breast cancer survivors are living with metastatic disease, where three-fourths of whom were originally diagnosed with

S. Mishra (✉) · H. Kalla · V. Ellappan · D. T. Aseffa · T. H. Ayane
Signal and Image Processing SIG, Department of ECE, SoEEC, Adama Science and Technology University, Adama, Ethiopia

T. Gopi Krishna
Department of CSE, SoEEC, Adama Science and Technology University, Adama, Ethiopia

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*, Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_33

stage I-III according to a recent study by American cancer society epidemiologists [1]. Mammography has a false-negative rate of 10–30%, false positive of 10%. Over 90% of breast cancer can be detected using mammography. It assists interpretation classifier directly to the Region Of İnterest (ROI) image data. The feed-forward and the back-propagation neural network has been used with gradient descent learning rule with momentum. Classification methods perform innate programming to classify the masses into benign and malignant class [2]. The features extracted contain the edge, sharpness measures, shape factors, and statistical texture features. Identification of breast tissues using simple image statistics such as kurtosis, correlation, and entropy plays an important role. SVM [3] was then used to segment mammogram images and identifying the true masses from the normal parenchyma and classified the masses as benign or malignant in mammograms.

The fast development of deep learning improves the performance of segmentation and classification to identify medical imaging problems. This paper has developed a novel segmentation and classification model to detect breast cancer on mammograms using an "end-to-end" training method that efficiently influences the training datasets with the cancer status of the entire image. Li et al. [4] used Digital Database for Screening Mammography (CBIS-DDSM) database and the Deep learning model and achieved sensitivity: 86.1%, specificity: 80.1%. There are different feature selection methods available as "Principal Component Analysis (PCA)" [5], "Linear Discriminant Analysis (LDA)" [6], "filtering techniques such as chi-square test" [7–9], are used to avoid over fitting.

Riddli et al. [10] utilized the "Faster R-CNN" model for the detection of mammographic scratches and classify them into benign and malignant. Singh et al. [11] proposed a "conditional generative adversarial network (cGAN)" for segmentation from an ROI. Agarwal et al. [12] proposed deep CNNtrained on "DDSM dataset" which shown noteworthy results in identifying the semantic characteristics in the clinical decision. Gao et al. [13] proposed a "shallow-deep CNN (SD-CNN)" and "contrast-enhanced DMs (CEDM)" for detection and classification. A "multiview CNN" [14–16] is proposed to detect breast masses. Jung et al. [17] proposed a "single-stage masses detection model" using the "RetinaNet" model for mammogram images. An adaptive local linear optimized radial basis functional neural network model has been proposed [18] for financial time series prediction. Satyasis Mishra et al. [19] proposed "LLRBFNN(Local linear radial basis function neural network)" with SCA-PSO for brain tumor "detection and classification". The mentioned classifiers involve complex network models and convolutional calculations and take more "computational time". Motivated by this, to reduce the complexity in the models, a simple LLWNN-SCA model is proposed for classification of breast cancers from the images. Several studies have been conducted for FCM based segmentation such as EnFCM, [19–21], FCMS1 [20], FCMS2 [19], FLICM [22], etc. in the literature to rule out the malignancy from breast images and unfortunately couldn't acquire the desired results. To have better results of detection, and improved FLICM segmentation algorithm is proposed. Further, the features are extracted from the FLICM segmented images. The extracted features are fed to the SCA-LLWNN model to have better classification results.

The remaining part of the article is organized with different sections as follows: Sect. 2 presents material and methods which include the workflow diagram, and Improved FLICM Algorithm. Section 3 presents the generalized regression neural networks, FNN, PRNN, SCA, and LLWNN model. Section 4 presents the results of segmentation and classification. Section 5 depicts the discussion of the research and Sect. 6 draws the conclusion of the research work.

2 Materials and Methods

2.1 Research Flow Block Diagram of the Proposed Method

Different methods are proposed to detect and classify breast cancers. In addition to existing methods, this article proposes a novel local linear wavelet neural network model with SCA optimization of weights to improve the performance of the LLWNN model. The results of accuracy are compared with that of previous models. The proposed research flow is shown in Fig. 1.

At the first step(1) the breast cancer image dataset has been undergone for preprocessing to enhance the images (ii) in the second step the enhanced images are utilized for segmentation by utilizing the proposed improved FLICM algorithm to detect the tumor from the mammogram breast cancer images (iii) Further in the third step, the

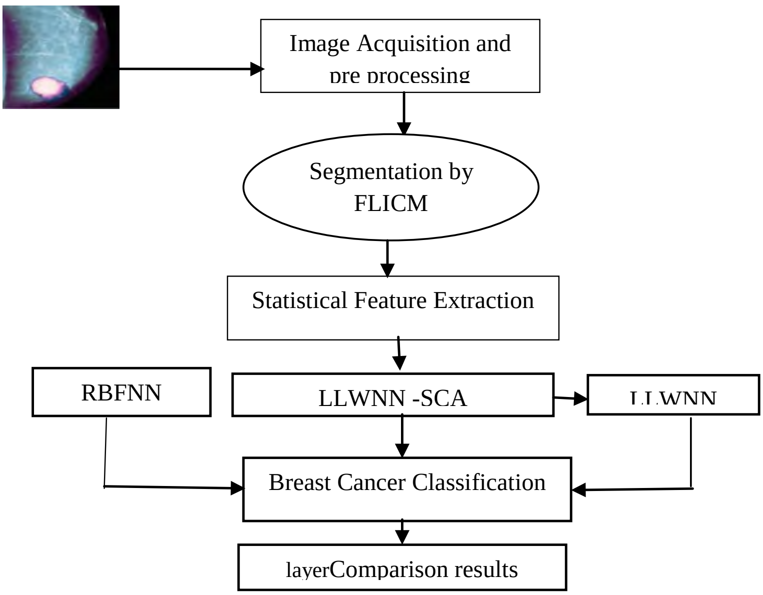


Fig.1 Workflow diagram of the proposed method

segmented images are utilized for statistical feature extraction (iv) At the 4th step the statistical features are fed as input to the proposed SCA based LLWNN model for classification breast cancer and comparison results are presented with the conventional RBFNN [23] and LLWNN models. The RBFNN(Radial Basis function Neural network) has been considered for comparison of classification as it has also proved good results for breast cancer classification, but the computational time taken by the algorithm is more in comparison to the proposed SCA based LLWNN model. At the same time, the LLWNN model also takes more time to converge which is shown in the Table 3.

2.2 *Improved FLICM Algorithm*

In order to reduce the computation in FCM algorithms [19, 20], the neighborhood term of FCM_S with variant is proposed. the objective function can be written as follows:

$$J_m = \sum_{i=1}^{c} \sum_{k \in N_k} u_{ik}^m \|x_k - v_i\|^2 + \alpha \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^m \sum_{k \in N_k} \|\overline{x_k} - v_i\|^2 \tag{1}$$

where $\overline{x_k}$ is a means of neighbouring pixels lying within *a* window around x_k.

The EnFCM [21] algorithm proposed to speed up the segmentation process where as a linearly-weighted sum image ξ and its local neighbor average image is presented as

$$\xi_k = \frac{1}{\alpha}\left(x_k + \frac{\alpha}{N_R} \sum_{j \in N_k} x_j\right) \tag{2}$$

where ξ_k denote the gray value of the *k*th pixel of the image ξ,

The FLICM [24] introduces a fuzzy factor to improve the performance of the segmentation better than the EnFCM but fails to preserve the robustness of the image quality. To improve the performance further the fuzzy factor of the FLICM has been modified as

$$G_{kv} = \sum_{\substack{r \in N_v \\ v \neq r}} \frac{1}{\exp(\mathrm{d}_{vr} + 1)} (1 - u_{kr})^m \|x_r - v_k\|^2 \tag{3}$$

With the modification of the fuzzy factor with the exponential function to the spatial Euclidean distance d_{vr}, the updated cost function with the modification factor is given by

$$J_s = \sum_{t=1}^{c} \sum_{l=1}^{q} \gamma_l u_{il}^m (\xi_l - v_t)^2 + \exp(G_{kv}) \tag{4}$$

With the new cost function, the accuracy has been calculated and compared to the other FCM based segmentation algorithms.

3 LLWNN Model with SCA Weight Optimization

In Fig. 2, the proposed SCA-LLWNN model the weights of the LLWNN [25–27] model with the SCA algorithm is presented and the pseudo-code for the algorithm for optimization also been shown.

With the data inputs $x_1, x_2, \ldots x_n$ which are treated as features and $Z_1, Z_2,$ ——-Z_N are the wavelet activation function in the hidden units and is demarcated by a wavelet Kernel as

$$Z_n(x) = |p_i|^{\frac{1}{2}} \Psi\left(\frac{x - q_i}{p_i}\right) \tag{5}$$

where the parameters *p, q* is the scaling and translation parameters, respectively

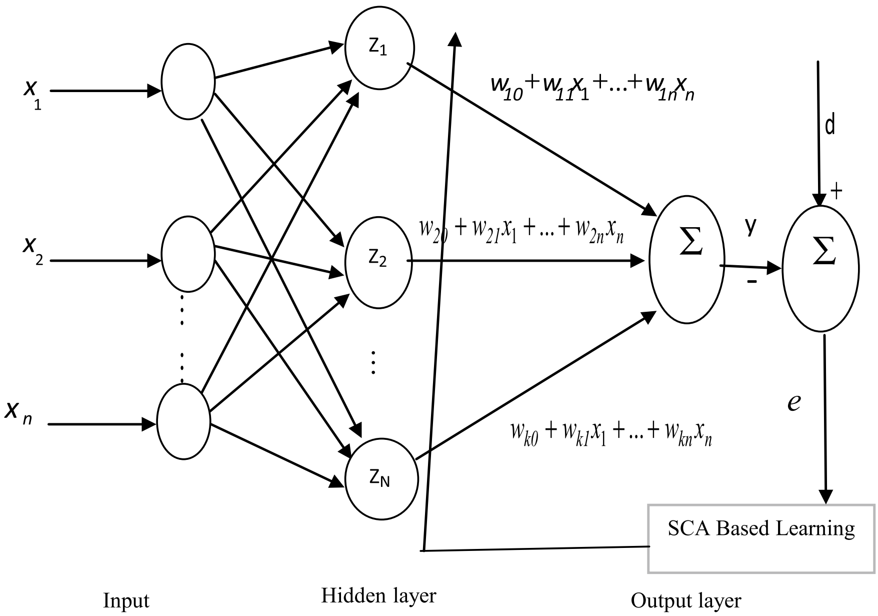


Fig. 2 SCA weight optimization based LLWNN model

$$y_n = \sum_{i=1}^{N} (w_{i0} + w_{i1}x_1 + \ldots w_{iN}x_N) Z_n(x) \tag{6}$$

The objective function is to minimize the error and the MSE is given by

$$\mathrm{MSE(e)} = \frac{1}{N} \sum_{n=1}^{N} (\mathrm{d}_n - y_n)^2 \tag{7}$$

where "d" is the desired vector.

3.1 Modified Sine Cosine Algorithm

According to the sine cosine algorithm [19], the position equation is updated as

$$X_i^{n+1} = \begin{cases} X_i^n + \alpha_1 \times \sin(\alpha_2) \times \left|\alpha_3 p^{g\,\mathrm{best}} - X_i^n\right|, & \alpha_4 < 0.5 \\ X_i^n + \alpha_1 \times \cos(\alpha_2) \times \left|\alpha_3 p^{g\,\mathrm{best}} - X_i^n\right|, & \alpha_4 \geq 0.5 \end{cases} \tag{8}$$

where $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ are the random variables and α_1 is given by

$$\alpha_1 = a\left(1 - \frac{n}{K}\right) \tag{9}$$

where n is the current iteration, K is the maximum number of iterations.

The current position X_i^n and X_i^{n+1} updated position has been mentioned in Eq. (12). The parameter α_1 governs the next place regions, α_2 evaluates the direction of movement from $x_i(n)$. The parameter α_3 controls the current movement, and the parameter α_4 equally changes among the sine and cosine functions. To have faster speed of convergence of the parameter α_1 is improved as

$$\alpha_m = \exp\left(\frac{1}{1 + (\alpha_1)}\right) \tag{10}$$

And the corresponding updated position equation is given as

$$X_{ij}^{n+1} = \begin{cases} X_{ij}^n + \alpha_m \times \sin(\alpha_2) \times \left|\alpha_3 p^{g\,\mathrm{best}} - X_i^n\right|, & \alpha_4 < 0.5 \\ X_{ij}^n + \alpha_{m1} \times \cos(\alpha_2) \times \left|\alpha_3 p^{g\,\mathrm{best}} - X_i^n\right|, & \alpha_4 \geq 0.5 \end{cases} \tag{11}$$

Further, the weights are mentioned as $W = [w_{i0} + w_{i1}x_1 + \ldots w_{iN}x_N]$ and the weights are mapped and updated using

$$W_{ij}^{n+1} = \begin{cases} W_{ij}^{n} + \alpha_m \times \sin(\alpha_2) \times \left|\alpha_3 p^{g\,\text{best}} - W_i^n\right|, & \alpha_4 < 0.5 \\ W_{ij}^{n} + \alpha_{m1} \times \cos(\alpha_2) \times \left|\alpha_3 p^{g\,\text{best}} - W_i^n\right|, & \alpha_4 \geq 0.5 \end{cases} \tag{12}$$

3.2 Data Collection

The breast cancer databases are collected from the "Cohort of Screen-Aged Women (CSAW)" which is a population-based data of all women "40 to 74 years of age" in the "Stockholm region, Sweden". "CSAW" [28] included 499,807 women invited for screening examinations. A total of 500 images are considered for training and testing. The features such as mean, standard deviation, variance, energy, entropy, minimum, and maximum are obtained from the segmented images. So there is $500 \times 7 = 3500$ number of data re-utilized for the purpose of classification. The simulations are accomplished with MATLAB2019a with an 8 GB RAM,2.35 GHz system.

Pseudo-code: SCA for weight optimization of LLWNN Model
1. Initialize the parameters of SCA parameters $\alpha_1, \alpha_2, \alpha_3, \alpha_m, K$
2. Initialize the position equation and map with the weights
3. Initialize the weights as W_{ij}^n
4. % starting of loop for $l = 1{:}K$ update(W_{ij}^n) as $W_{ij}^{n+1} = \begin{cases} W_{ij}^{n} + \alpha_m \times \sin(\alpha_2) \times \left\|\alpha_3 p^{g\,\text{best}} - W_i^n\right\|, & \alpha_4 < 0.5 \\ W_{ij}^{n} + \alpha_{m1} \times \cos(\alpha_2) \times \left\|\alpha_3 p^{g\,\text{best}} - W_i^n\right\|, & \alpha_4 \geq 0.5 \end{cases}$ End update W_{ij}^{n+1} to obtain minimum weight values end of for loop
6. Stopping criteria: Continue till optimization gets minimum error values
7. If not converges, repeat until nearly zero error is satisfied

4 Results

4.1 Segmentation Results

4.2 Classification Results

The X-axis of the above graph represents the simulation result iteration (epochs) and Y-axis determines Mean Squared Error (MSE). A set of dataset observations were randomly split into test error (as showed in red), validation set (showed in blue), and training set (showed in cyan). This graph showed test error at the beginning is quite higher when epoch increase error linearly decreased.

5 Discussions

Figures 3, 4, 5, shows the segmentation results. The segmentation accuracy is measured with rician noise[19]. Improved FLICM also reduces the noise from the breast cancer images and provides better performance in segmentation. The segmentation accuracy has been presented in the Table 1 with rician noise. The classifier LLWNN-SCA is focused with the help of the training method and validation method. For comparison “Feed-Forward Neural Network (FFWNN)”, Fit Function Neural

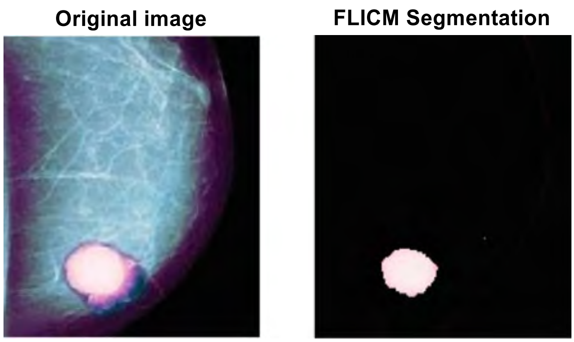


Fig. 3 Segmentation of the breast cancer images

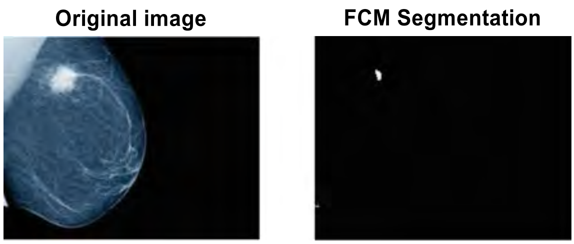


Fig. 4 Segmentation of the breast cancer images

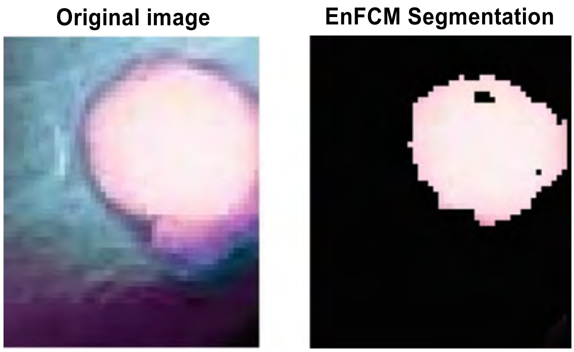


Fig. 5 Segmentation of the breast cancer images

Table 1 Segmentation accuracy

Algorithm	Noise level	
	Rician noise ($\sigma_n = 10$)	Rician noise ($\sigma_n = 20$)
FCM S1	90.21	89.83
FCM S2	95.43	92.37
En FCM	98.21	96.52
FLICM	98.81	97.75
IFLICM	**99.25**	98.83

The bold indicates higher values of accuracy in comparison to other mentioned algorithms

Network (FFNN), and pattern regression neural network (PRNN). The output simulation results of the same data used for the Generalized Regression Neural Networks (GRNN), PNN, and RBFN are considered. The output simulation result of comparison performance by hold-validation, cross-validation for LLWNN-SCA is shown in Fig. 6. Among training set methods LLWNN-SCA training performs best accuracy among the others as observed in Table 2. As the model LLWNN-SCA is a novel method, and the functions of the network are not available, MSE is taken through mathematical calculation. In this research work, performances of different parameters have been considered. All parameters in detection schemes are used to measure the enhancement of mammogram image by removing noise and minimize the variation of neighboring pixels and increase contrast image. MSE is the key parameter to measure the performance of the error in classification. Network error (net error) is also one of the parameters which measure the error at each mentioned neural network. Bias is an important factor that has a great impact on the radial basis function. Another property of radial basis function is the disparity of output configuration and distance between input and weights. In order to manifest the statistics of the simulation in terms of different parameters for all scheme are provided in Table 2. The test errors and accuracy are calculated by using the machine learning models such as PRNN, GRNN, RBFNN, PNN, and FFWNN that are simulated with the dataset newly created for this research work. The sensitivity and specificity performance

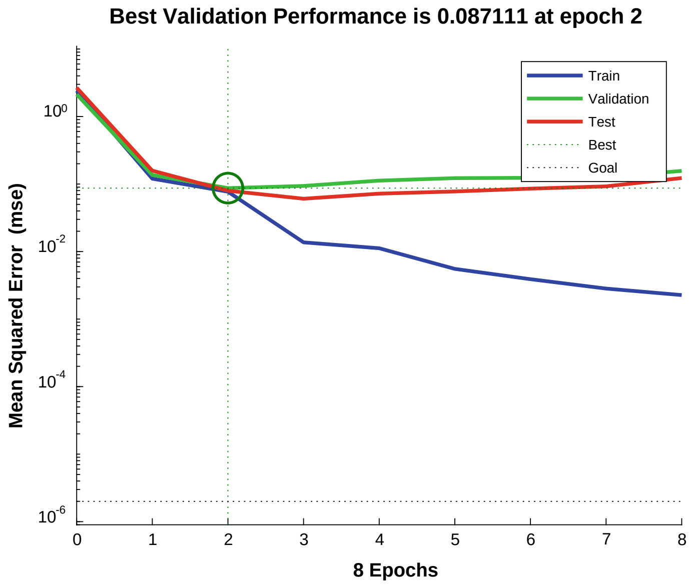


Fig. 6 Comparison of performance for hold-validation of LLWNN-SCA

evaluations are presented in Table 3. The sensitivity is defined as True Positive (TP) divided by the True Positive (TP) + False Negative(FN).Similarly, the specificity is defined as true negative (TN) divided by true negative (TN) + False positive (FP). The sensitivity and specificity are calculated by considering the classification measure. The computational time taken by the proposed SCA + LLWNN took 18.4521 secs which is lower than the RBFNN, and LLLWNN method.

6 Conclusion

This research work proposes a novel Improved FLICM segmentation and hybrid SCA-LLWNN classification machine learning model. The breast cancer images are segmented by the improved FLICM method and the segmented images are utilized for feature extraction. The statistical features are extracted and fed as input to the proposed SCA-LLWNN model. The FFNN, FFWNN, GRNN also has been taken into consideration and the classification results are presented in Table 3. The SCA-LLWNN has shown better accuracy compared to all other mentioned methods. PRNN also showed good performance among training neural networks due to reduction of the false negative and false positive predicted class as well as improve true positive and true negative predicted class of breast cancer dataset. As observed from Table2 that, when the number of hidden layers decreases, the error also decreases, and the accuracy increases. The key existing parameters such as MSE, test error, accuracy in

Table 2 Performance validation for classification

Neural network scheme	Parameters (%)	Hidden layer (N)	
		$N = 20$	$N = 40$
GRNN	MSE	0.427	0.427
	Net error	0.683	0.683
	Accuracy	96.667	98.333
PNN	MSE	0.428	0.428
	Net error	15.409	15.409
	Accuracy	98.333	96.667
RBFNN	MSE	0.030	0.030
	Net error	0.0876	0.876
	Accuracy	95	98.333
FFWNN	MSE	3.70	1.86
	Test error	0.0381	0.032
	Accuracy	97.145	98.095
PRNN	MSE	1.83	1.9
	Test error	0.046	0.0485
	Accuracy	97.145	98.968
LLWNN	MSE	0.99	0.99
	Test error	0.029	0.051
	Accuracy	98.095	99.048
LLWNN-SCA	MSE	0.130	0.130
	Test error	0.023	0.048
	Accuracy	**99.020**	**99.307**

The bold indicates higher values of accuracy in comparison to other mentioned algorithms

Table 3 Performance evaluation

Model	No. of data	Computational Time in sec	Sensitivity in %	Specificity in %	Accuracy in %
RBFNN	3500	39. 5234	97.25	88.59	98.333
LLWNN	3500	26.6754	98.43	95.67	99.048
LLWNN + SCA	3500	18.4521	98.89	99.15	**99.307**

The bold indicates higher values of accuracy in comparison to other mentioned algorithms

mammogram image classification has been utilized to present the performance of the models. With these experimental results, it can be concluded that the performance of detection and classification achieved by improved FLICM segmentation and SCA-LLWNN for mammogram image is superior to the other FCM based segmentation methods and classification models.

References

1. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2019–2020.pdf in American cancer society 2020
2. S. Mojarad, S. Dlay, W. Woo, G. Sherbet,Breast cancer prediction and cross validation using multilayer perceptron neural networks. in *Proceedings 7th Communication Systems Networks and Digital Signal Processing*, Newcastle, 21st–23rd July (IEEE, 2010), pp. 760–674
3. Y. Ireaneus Anna Rejani, S. Thamarai Selvi Noorul, Early detection of breast cancer using SVM classifier technique. Int. J. Comput. Sci. Eng. **1**(3), 127–130 (2009)
4. L. Shen, L.R. Margolies, J.H. Rothstein, Deep learning to improve breast cancer detection on screening mammography. Sci Rep **9**, 12495 (2019)
5. H.P. Chan, D. Wei, M.A. Helvie, B. Sahiner, D.D. Adler, M.M. Goodsitt, Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space. Phys. Med. Biol. **40**(5), 857–876 (1995)
6. X. Jin, A. Xu, R. Bie, P. Guo, Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles. in *Proceedings of the 2006 International Conference on Data Mining for Biomedical Applications* 2006 Presented at: BioDM'06; April 9, (Singapore, 2006), pp. 106–115
7. G.I. Salama, M.B. Abdelhalim, M.A. Zeid, Breast cancer diagnosis on three different datasets using multi-classifiers. Int. J. Comput. Sci. Inf. Technol. **1**(1), 36–43 (2012)
8. E. Saghapour, S. Kermani, M. Sehhati, A novel feature ranking method for prediction of cancer stages using proteomics data. PLoS One **12**(9), e0184203 (2017)
9. M.M. Eltoukhy, S.J. Gardezi, I. Faye, A method to reduce curvelet coefficients for mammogram classification. in *Proceedings of the Region 10 Symposium.* 2014 Presented at: IEEE'14; April 14–16, 2014, (Kaulalumpur, Malaysia, 2014), pp. 663–666
10. D. Ribli, A. Horváth, Z. Unger, P. Pollner, I. Csabai, Detecting and classifying lesions in mammograms with deep learning. Sci. Rep. **8**(1), 4165 (2018)
11. V.K. Singh, S. Romani, H.A. Rashwan, F. Akram, N. Pandey, M. Sarke, conditional generative adversarial and convolutional networks for x-ray breast mass segmentation and shape classification. in *Proceedings of the Medical Image Computing and Computer Assisted Intervention.* Presented at: MICCAI'18; September 16–20. (Granada, Spain, 2018), pp. 833–840
12. V. Agarwal, C. Carson, Stanford University. Using deep convolutional neural networks to predict semantic features of lesions in mammograms (2015). https://cs231n.stanford.edu/reports/2015/pdfs/vibhua_final_report.pdf
13. F. Gao, T. Wu, J. Li, B. Zheng, L. Ruan, D. Shang, SD-CNN: a shallow-deep CNN for improved breast cancer diagnosis. Comput. Med. Imaging Graph **70**, 53–62 (2018)
14. Y.B. Hagos, A.G. Mérida, J. Teuwen, Improving breast cancer detection using symmetry information with deep learning. in *Proceedings of the Image Analysis for Moving Organ, Breast, and Thoracic Images.* RAMBO'18; September 16, (Granada, Spain, 2018), pp. 90–97
15. J. Teuwen, S. van de Leemput, A. Gubern-Mérida, A. Rodriguez-Ruiz, R. Mann, B. Bejnordi, Soft tissue lesion detection in mammography using deep neural networks for object detection. In: *Proceedings of the 1st Conference on Medical Imaging with Deep Learning.* 2018 Presented at MIDL'18; July 4–6. (Amsterdam, The Netherlands, 2018) pp. 1–9
16. R. Dhaya, Deep net model for detection of Covid-19 using radiographs based on ROC analysis. . J. Innov. Image Process. (JIIP) **2**(03), 135–140 (2020)
17. H. Jung, B. Kim, I. Lee, M. Yo, J. Lee, S. Ham, Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network. PLoS ONE **13**(9), e0203355 (2018)
18. S.N. Mishra, A. Patra, S. Das, M.R. Senapati, An adaptive local linear optimized radial basis functional neural network model for financial time series prediction. Neural Comput. Appl. **28**(1), 101–110 (2017)

19. S. Mishra, P. Sahu, M.R. Senapati, MASCA–PSO based LLRBFNN model and improved fast and robust FCM algorithm for detection and classification of brain tumor from MR image. Evol. Intel. **12**, 647–663 (2019)
20. S. Chen, D. Zhang, Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure. IEEE Trans. Syst. Man Cybern B Cybern **34**(4), 1907–1916 (2004)
21. L. Szilagyi, Z. Benyo, S.M. Szilagyii, H.S. Adam, MR brain image segmentation using an enhanced fuzzy c-means algorithm. In: *Proceeding of the 25th annual international conference of the IEEE EMBS*, (2003), pp. 17–21
22. W. Cai, S. Chen, D. Zhang, Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. Pattern Recognit. **40**(3), 825–838 (2007)
23. K.T. Vijay, Classification of brain cancer type using machine learning. J. Artif. Intell. **1**(2), 105–113 (2019)
24. S. Krinidis, V. Chatzis, A robust fuzzy local inform ation cmeans clustering algorithm. IEEE Trans. Image Process **19**(5), 1328–1337 (2010)
25. M.R. Senapati, P.K. Dash, Intelligent systems based on local linear wavelet neural network and recursive least square approach for breast cancer classification. Artif. Intell. Rev. **39**(2), 151–163 (2013) Springer, ISSN 0269–2821
26. W.S. Tamil Selvi, J. Dheeba, N. Albert Singh, Computer aided detection of breast cancer on mammograms: a swarm intelligence optimized wavelet neural network approach. J. Biomed. Info. **49**(2014), 45–52 (2014) Elsevier Inc
27. V. Chakkarwar, M.S. Salve, Classification of mammographic images using gabor wavelet and discrete wavelet transform. Int. J. Adv. Res. ECE 573–578 (2013) ISSN
28. K. Dembrower, P. Lindholm, F. Strand, A multi-million mammography image dataset and population-based screening cohort for the training and evaluation of deep neural networks—the cohort of screen-aged women (CSAW). J Digit Imaging **33**, 408–413 (2020)

Detection of Diabetic Retinopathy Using Deep Convolutional Neural Networks

R. Raja Kumar, R. Pandian, T. Prem Jacob, A. Pravin, and P. Indumathi

Abstract Diabetic retinopathy (DR) is the most prevalent disease among diabetic patients. DR affects retinal blood vessels and causes total loss of vision if it is not treated earlier. Around 50% of the diabetic population is affected by DR and thus necessitates the need for DR detection. Manual processes available consume more time, and hence a convolutional neural network is employed to detect DR at an earlier stage. The proposed method has three convolutional layers and a fully connected layer. This method gives higher accuracy (94.44%) with reduced hardware requirement than conventional approaches to detect and classify DR into five stages, namely no DR, mild, moderate, severe, and proliferative DR.

Keywords Diabetic retinopathy · Deep learning · Convolutional neural network · Convolutional layer · Fully connected layer · Radial basis function

1 Introduction

Diabetes is a disease that turns up when adequate insulin is not secreted by the pancreas or the cell does not respond to the produced insulin. The increase in blood sugar level causes damage in the retinal blood vessels called diabetic retinopathy. This disease at a severe stage leads to a total loss of vision. There are two important stages

R. Raja Kumar (✉)
Mathematics Department, Sathyabama Institute of Science and Technology, Sholinganallur, 600119 Chennai, India

R. Pandian
Department of Electronics and Instrumentation Engineering, Sathyabama Institute of Science and Technology, Sholinganallur, 600119 Chennai, India

T. Prem Jacob · A. Pravin
Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Sholinganallur, 600119 Chennai, India

P. Indumathi
Department of Electronics Engineering, MIT Campus, Anna University, 600044 Chennai, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_34

Fig. 1 **a** An eye affected by diabetic retinopathy (Medical News Today, August 2017)

of diabetic retinopathy, namely non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR). In NPDR, the growth of new blood vessels which affect the already present blood vessels is less, whereas, in the PDR, the new abnormal blood vessels grow over the retina causing severe loss of vision (Fig. 1).

As per the World Health Organization (WHO), about 70% of persons with diabetic retinopathy live in developing countries. Sometimes the diabetic retinopathy symptoms do not show up at an earlier stage so the ophthalmologist cannot treat the disease within the stipulated time as the screening process takes more time. Diabetic retinopathy is categorized into five stages, namely stage 1 denotes No DR—The person is normal and does not possess any symptom of DR, and stage 2 denotes Mild—a feature called microaneurysm is formed that are small balloon-like swellings in the retinal blood vessels. This prevents light from entering into the retina from leaking fluid into the retinal area. Stage 3 denotes Moderate—the blood circulation in the retinal blood vessels gets damaged due to the formation of exudates, and stage 4 denotes Severe—new abnormal blood vessels emerge in the retina affecting necessary blood vessels. Stage 5 denotes Proliferative DR—the growth of new blood vessels increases drastically causing the detachment of the retina from the basal blood vessels. This leads to permanent loss of vision.

Deep learning is a subgroup of machine learning with enhanced performance. It does not require separate feature extraction, unlike machine learning. It goes deep into the network for analysis based on the extracted features in each layer. In deep learning, convolutional neural network (CNN) is an approach that is mostly applied to perceptible images. It has various convolutional layers to detect features in an image and fully connected layers to aggregate the results for class separation. Our proposed model consists of three convolutional layers and a fully connected layer. The convolutional layer gives a feature map of the extracted features like shape, color, texture, and size, and there is a layer called softmax after the fully connected layer to normalize the output into probabilities. This convolutional neural network requires nominal preprocessing. This is achieved with the help of the data augmentation process. The comparative results with the data augmentation process and without the data augmentation process are discussed in the following sections.

2 Related Work

Shanthi et al. [1] used an algorithm by modifying basic AlexNet architecture to classify DR. This approach contains only four classes. Frazao et al. [2] used holistic texture and local retinal features to diagnose DR. This method produces low accuracy. Jebaseeli et al. [3] proposed a new algorithm named tandem pulse coupled neural network (TPCNN) algorithm for feature extraction and support vector machine (SVM) for classification. This method provides low sensitivity and also needs a separate algorithm for feature extraction. Devaraj et al. [4] have surveyed extract features like microaneurysms and exudates. Wan et al. [5] have done comparative work with different classifiers for analyzing classifier performance. Uma et al. [6] proposed a morphological-based approach for extraction of features like blood vessels, optic disk, finally the exudates. This method is applicable for a small set of images, and it requires an additional step called feature extraction. Omar et al. [7] used morphological operations-based approach to classify DR. The accuracy obtained with this method is not too good. Pratt et al. [8] used 10 convolutional layers and 3 fully connected layers to detect and classify DR. the accuracy and specificity obtained with this method are too low. Omar et al. [9] used a multiscale local binary pattern (LBP) texture model for feature extraction and radial basis function (RBF) for classification in their proposed work. This approach fails to classify DR into five stages as it concentrates only on exudate extraction. To detect shapes, higher-order spectra (HOS) were used in [10] for its nonlinear features. The SVM classifier was used to classify DR into five classes. Though very involved, the achievable performance measures were only a sensitivity of 82% and a specificity of 88%. Different stages of DR are identified in [11]. The neo-normal pandemic is to be detected which was discussed [12]. The authors have made use of deep neural networks for the detection of Covid-19. Interestingly in [13], retrieval of complex images was done using cognitive classification. The risk factors of DR are investigated in [14], after applying statistical techniques.

The proposed work involves the classification of DR and various methods for investigation. Convolutional neural networks are employed to solve the problem. The reason to select CNN is that it is involved with the extraction of features (from image) and later split the data for training and testing purposes.

3 Proposed Framework

The fundus camera is used to obtain the fundus images of the eye (i.e., taking a photograph of the rear of the eye) and databases like the Kaggle database were used to store those fundus images. The concept of a convolutional neural network is used to illustrate the framework of the five different classes of DR.

The proposed framework is shown in Fig. 2b origins with the data collection. The input retinal fundus image from the Kaggle database is used in this work. The input retinal image is of size around 1 MB. It takes more time to process. As an initial stage, the input image is preprocessed wherein resizing and cropping is done to reduce the image size and also to remove unwanted portions from an image. The convolutional neural network plays a major role in extracting features from an image and classifying them based on the obtained features. The data are split for training as well as testing. The convolutional layer does the process of feature extraction by convolving the input image with the filter or kernel used. In each convolutional layer output, a feature map is obtained. The entire feature points grabbed from the

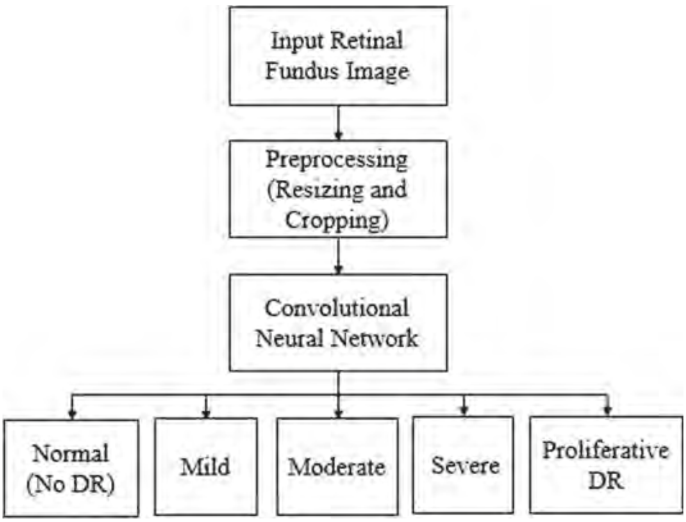


Fig. 2 **b** Proposed framework

convolutional layer do not contain useful information, and therefore, the max-pooling is done to reduce the dimension. Finally, the fully connected layer connects all the neurons, such that the obtained features are converted into classes. The training and testing data are compared to obtain the performance metrics like accuracy, sensitivity, and specificity.

3.1 Preprocessing

The images acquired from the Kaggle database are of various sizes. This makes the task convolutional neural network complex. Therefore, the images from the dataset are resized and cropped to a size of 227*227*3 to remove unwanted portions and to reduce memory consumption [5]. Some of the authors extracted the green channel from the image as it gives high sensitivity and moderate saturation than red and blue channels. The proposed work does not necessitate the process of green channel extraction as the data augmentation process and collects fine details from the image by scaling, translation, and rotation.

3.2 Convolutional Neural Network

Convolutional neural network (CNN) is mainly applied to visual images as it requires less preprocessing time. A simple convolutional neural network architecture is given in Fig. 3. It consists of an input layer, output layer, and hidden layers. The convolutional layer, pooling layer, and fully connected layer form the hidden layer. Each convolutional layer has some kernel/filters which help in the extraction of features from the image. The filter operates on the original image to produce a feature map.

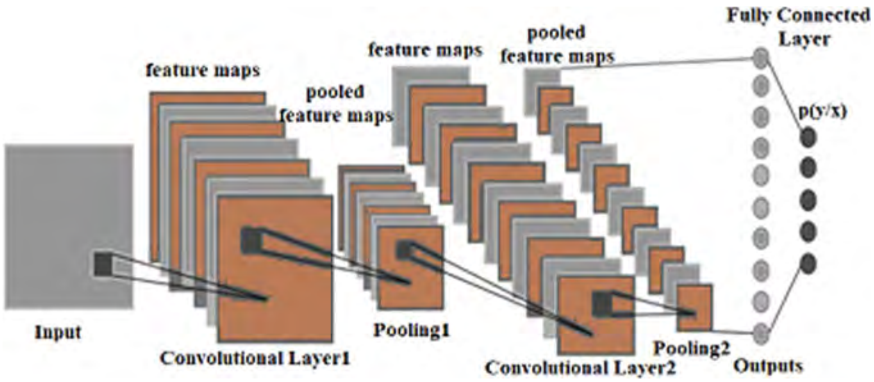


Fig. 3 Simple convolutional neural network architecture

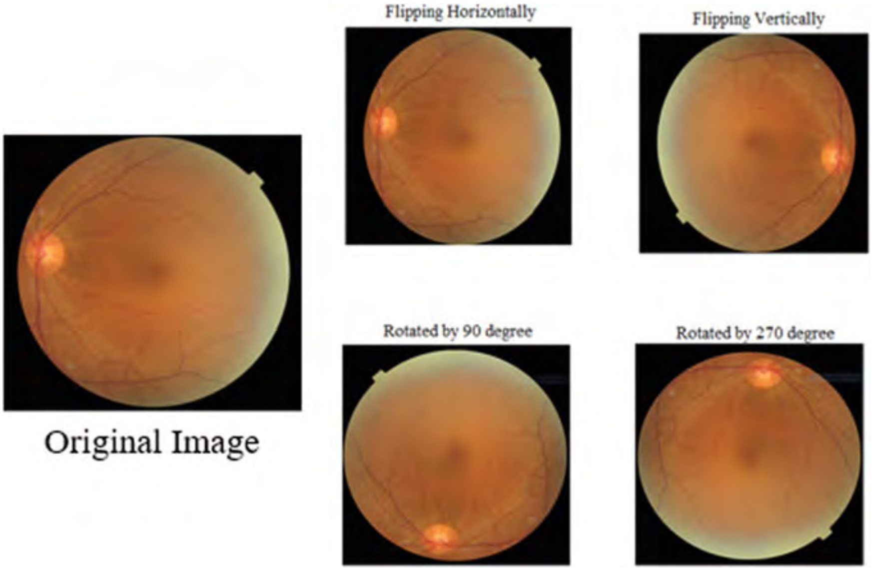


Fig. 4 Data augmented image

The obtained feature map contains feature points. All the feature points in the feature map are not necessary and hence max-pooling, which is the process of sub-sampling is done to reduce the dimensions of an image. Thus, the processing time is reduced. The fully connected layer neurons are fully connected to the previous layer neurons for class separation.

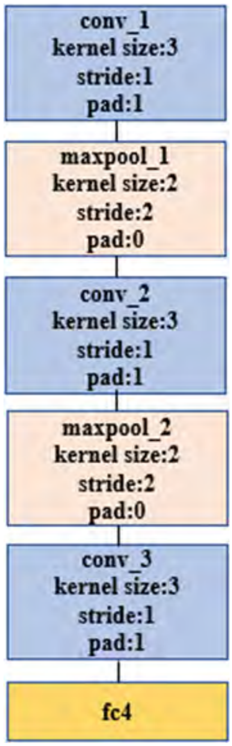
3.3 Data Augmentation

The data augmentation is the process wherein the data are flipped horizontally, vertically, tilted, scaled, rotated, and translated to collect fine details of an image. In simple convolutional neural networks, the training performance is much higher than the testing performance. Hence, methods such as data augmentation, normalization, and regularization are adopted. The data augmentation process is portrayed in Fig. 4

3.4 Proposed CNN

The proposed CNN architecture is comprised of three convolutional layers such a max pool layer, a fully connected layer, and a softmax layer. In a brief sense, the layers are as follows: conv_1 consists of eight 3*3 filters, maxpool_1 of size 2*2, conv_2

Fig. 5 Proposed CNN architecture



consists of sixteen 3*3 filters, maxpool_2 of size 2*2, conv_3 consists of thirty-two filters of size 3*3, fully connected layer fc4, and finally ended with softmax layer. The softmax layer gives the class scores by normalizing the output of the fully connected layer into probabilities. The proposed architecture is given in Fig. 5.

The activation function used in this proposed work is the rectified linear unit (ReLU). It decides which neurons to be activated such that the desired classification result is achieved. This function eliminates the drawbacks in the previous functions like gradient exploding or gradient vanishing and nonzero centered problem. The ReLU activation function with a feature point 'y' is given by

$$f(y) = \max(0, y) \tag{1}$$

The softmax function $\sigma(y)$ is given by

$$\sigma(y) = \frac{e^{y_j}}{\sum_{k=1}^{K} e^{y_k}} \text{ for } j = 1, \ldots K \text{ and } y = (y_1, \ldots, y_K) \in R^K \tag{2}$$

where y_j is the element of the input vector y.

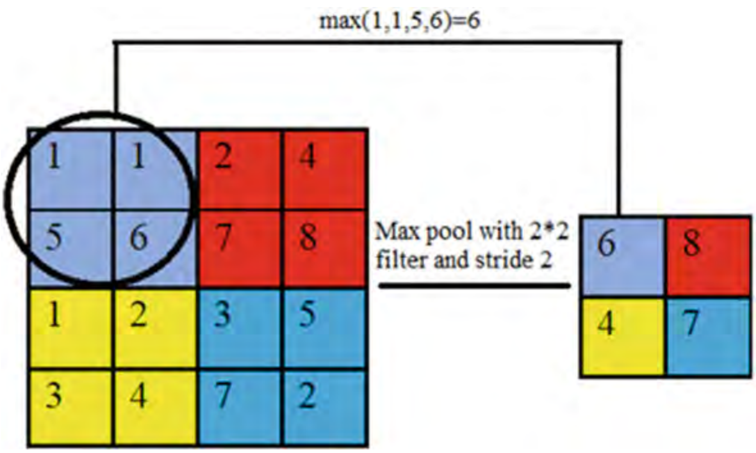


Fig. 6 Max-pooling operation

The max-pooling layer performs max-pooling which is a process of downsampling the feature map obtained and taking the maximum values. Therefore, the dimensionality along the width and height is greatly reduced with this process. High contrast features are obtained by this process. The operation of max-pooling is given in Fig. 6.

Each layer has a dimension that can be calculated by the following formula. The calculation of layer size plays a crucial role in the design of a convolutional neural network. The output layer size (O) is given by

$$O = \frac{I - K + 2P}{S} + 1 \tag{3}$$

where 'I' represents the input layer width or height, 'K' represents the filter/kernel size, and 'P' represents the padding length, 'S' represents the stride length provides the number of shifts of the filter takes over the input image.

4 Experimental Results

4.1 *Software Used*

The proposed framework uses a simulation platform of Matlab R2018b. An interactive mode is available in MATLAB 2018b enables the user to create the individual neural network using a deep neural network designer application. MATLAB performs matrix operations; i.e., it converts the input image into the matrix and then into a vector for processing. This is highly used for image processing applications.

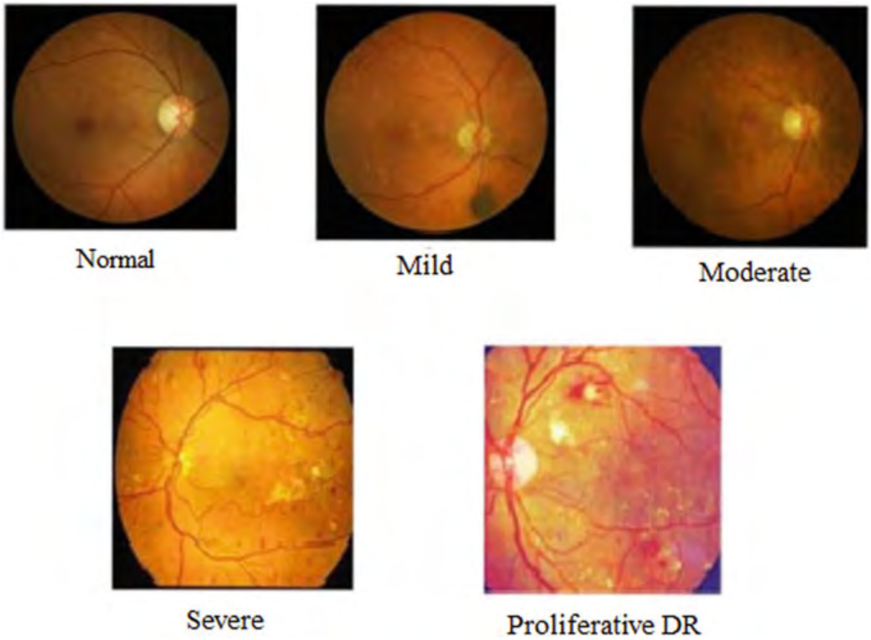


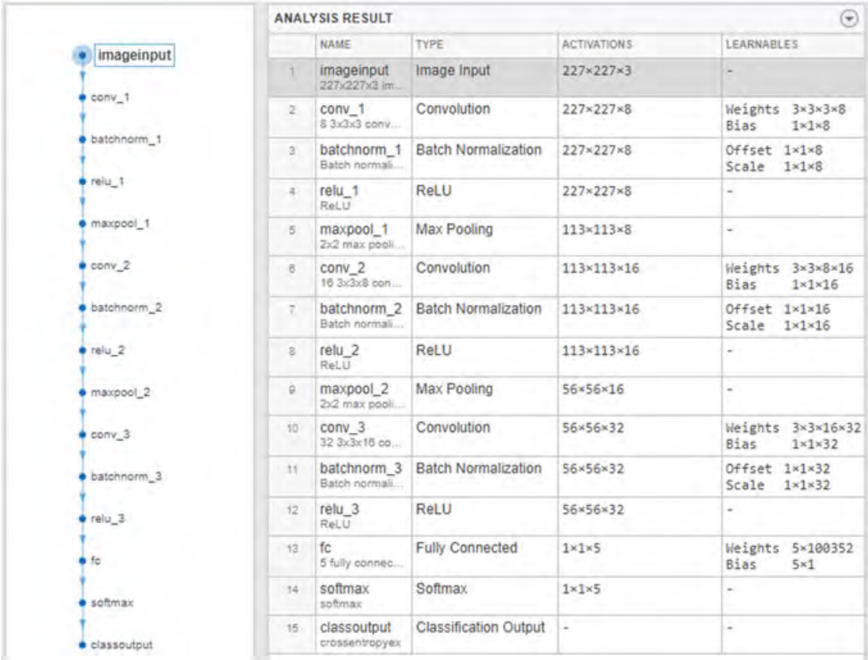
Fig. 7 Sample classification dataset

4.2 Dataset

The dataset used in the framework is obtained from the Kaggle database [1]. The training and the testing images in the dataset consist of five stages, namely no DR, mild, moderate, severe, and proliferative DR. The images in the dataset consist of both the inverted and the non-inverted images. If a square or circular notch is present in the image, then the image is not inverted and vice-versa. Here, a zip application is recommended to extract zip files in the Kaggle database. 300 images from the Kaggle database are used for training as well as testing. This requires a high-end general processing unit (GPU). The sample classification dataset for DR stages is given in Fig. 7.

4.3 Simulation Results

The proposed CNN architecture is comprised of three convolutional layers and a fully connected layer [10–15]. Out of 300 images, 70% (210) is used for training, and 30% (90) is used for testing. The Kaggle dataset is used in the proposed work. The image is resized to a size of 227*227*3. This is done to reduce memory consumption and



	NAME	TYPE	ACTIVATIONS	LEARNABLES
1	imageinput 227x227x3 im...	Image Input	227×227×3	-
2	conv_1 8 3x3x3 conv...	Convolution	227×227×8	Weights 3×3×3×8 Bias 1×1×8
3	batchnorm_1 Batch normali...	Batch Normalization	227×227×8	Offset 1×1×8 Scale 1×1×8
4	relu_1 ReLU	ReLU	227×227×8	-
5	maxpool_1 2x2 max pooli...	Max Pooling	113×113×8	-
6	conv_2 16 3x3x8 con...	Convolution	113×113×16	Weights 3×3×8×16 Bias 1×1×16
7	batchnorm_2 Batch normali...	Batch Normalization	113×113×16	Offset 1×1×16 Scale 1×1×16
8	relu_2 ReLU	ReLU	113×113×16	-
9	maxpool_2 2x2 max pooli...	Max Pooling	56×56×16	-
10	conv_3 32 3x3x16 co...	Convolution	56×56×32	Weights 3×3×16×32 Bias 1×1×32
11	batchnorm_3 Batch normali...	Batch Normalization	56×56×32	Offset 1×1×32 Scale 1×1×32
12	relu_3 ReLU	ReLU	56×56×32	-
13	fc 5 fully connec...	Fully Connected	1×1×5	Weights 5×100352 Bias 5×1
14	softmax softmax	Softmax	1×1×5	-
15	classoutput crossentropyex	Classification Output	-	-

Fig. 8 Proposed CNN network analysis

processing time. The architecture of the proposed work and the layer description is mentioned in the network analysis figure which is given in Fig. 8.

The training images used for the proposed framework is shown in Fig. 9.

The model accuracy and loss graph without data augmentation are shown below. As the images from the dataset are not flipped, rotated, scaled, or translated, only less detail from the image is captured. If the feature obtained from the image is less, the accuracy, thus resulted in 66.67%. Due to this, another problem occurs called over fitting where the performance of the training dataset is too high, and the testing dataset is too low. This can be well seen in the model accuracy and loss graph shown in Fig. 10.

The data augmentation process overcomes the problem of overfitting, and hence, the performance metrics of the proposed framework are considerably enhanced. Backpropagation is the algorithm used in the convolutional neural network for loss calculation and minimization. The accuracy obtained with this process is 94.44%. This is clearly shown in Fig. 11.

The classified result is shown in Fig. 12. The softmax layer produces class scores. Based on the class scores, the input image is classified into one of the following stages, namely no DR, mild, moderate, severe, and proliferative DR.

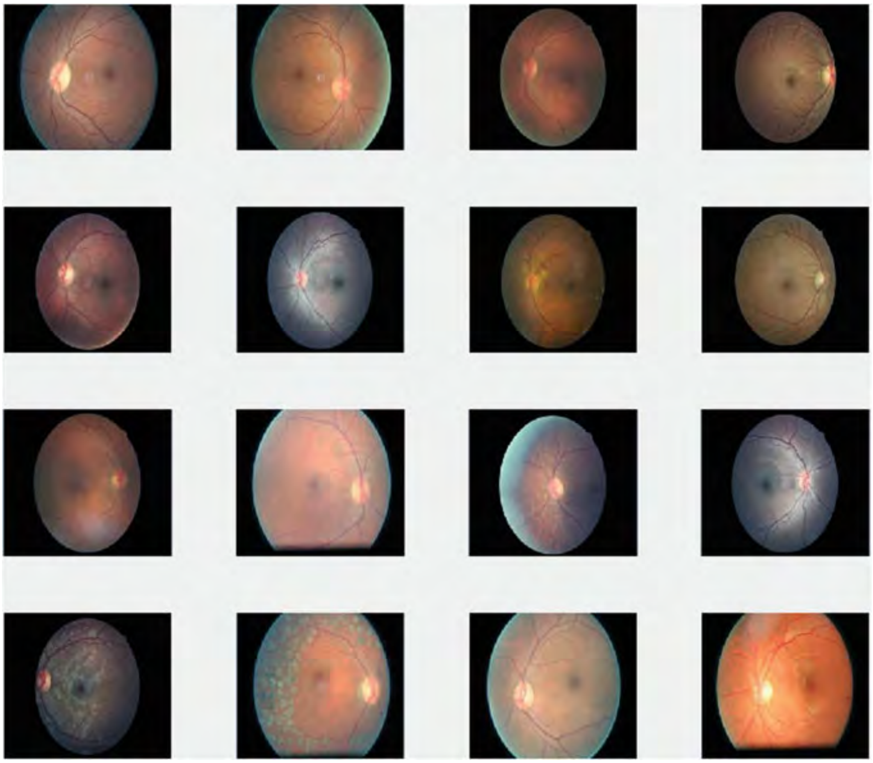


Fig. 9 Training images

The performance metrics like sensitivity, specificity, accuracy, f-score, and precision are calculated with the help of the obtained confusion matrix as shown in Fig. 13.

The performance metrics are given by the following formula and are shown in Table 1.

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \tag{4}$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{5}$$

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{6}$$

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN} + \text{FP})} \tag{7}$$

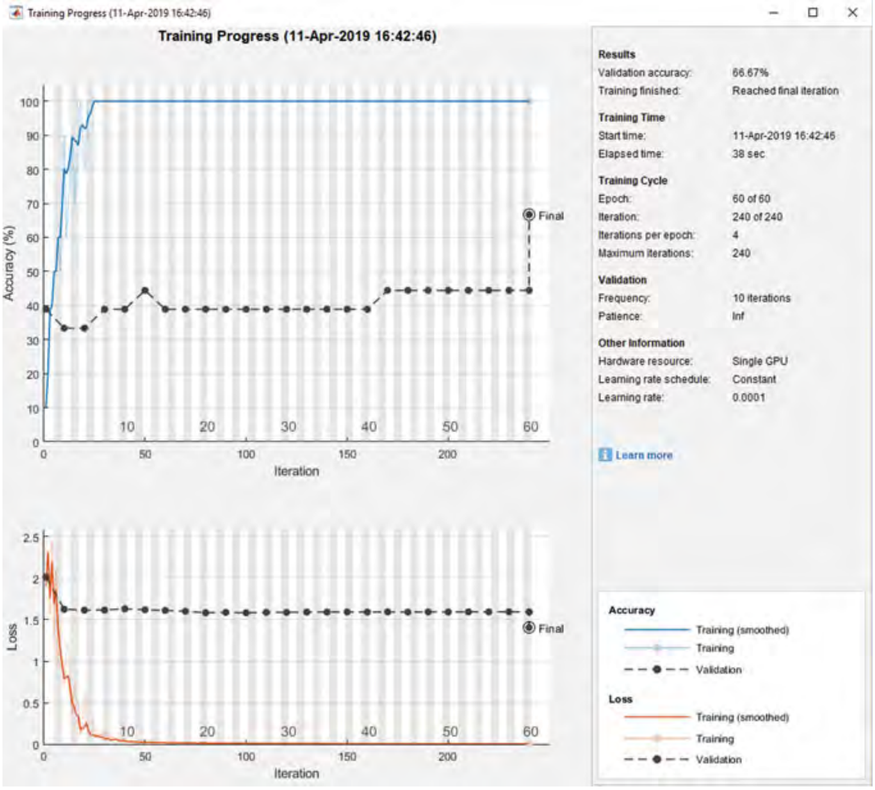


Fig. 10 Model accuracy and loss without data augmentation

$$f - \text{score} = \frac{2 * \text{TP}}{(2{*}\text{TP} + \text{FP} + \text{FN})} \tag{8}$$

The parameters such as accuracy, sensitivity, specificity, precision, and f-score achieved as a whole are 94.44, 93.96, 98.54, 93.96, and 93.96%, respectively. If suppose the accuracy was 100%, there will not be any indication whatsoever about the various stages of DR. As an example, the 100% accuracy concerning the absence of DR being accurately diagnosed. Once that happens, there is no chance of any other stage of DR affecting the patient.

5 Conclusion

In this proposed work, a deep convolutional neural network is designed to detect and classify diabetic retinopathy into five stages, as it is a deadly disease among diabetic patients. The results obtained from our proposed framework are quite high

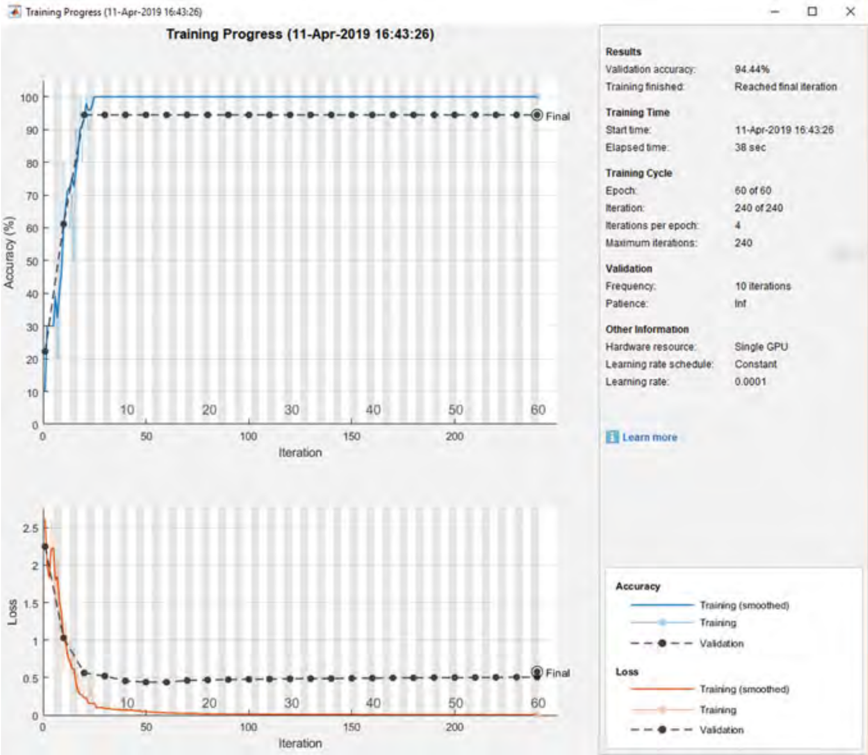


Fig. 11 Model accuracy and loss with data augmentation

when compared to the performance metrics obtained. Therefore, convolutional neural network-based diabetic retinopathy classification is more reliable and accurate. This consumes less time for classification to occur. Hence, the disease can be treated earlier and reduces the risk of vision loss. In the future, the images with higher resolution can be used for gaining high accuracy and other performance metrics.

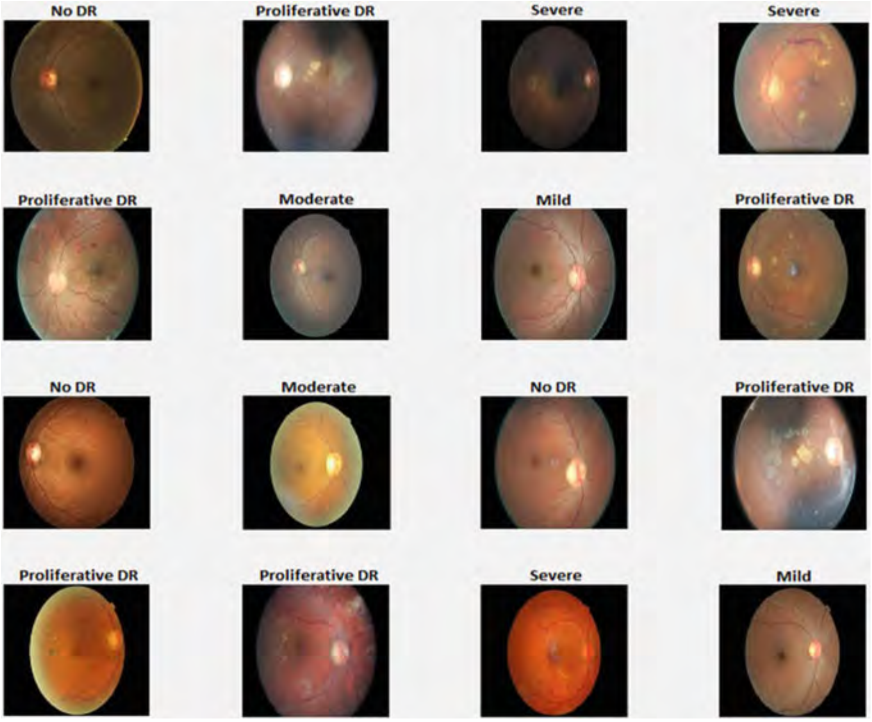


Fig. 12 Classified images

Table 1 Performance metrics of proposed CNN

Performance metric per classes	Mild (%)	Moderate (%)	No DR (%)	Proliferative DR (%)	Severe DR (%)
Precision	95.23	93.75	96.15	90.9	93.75
Sensitivity	95.23	93.75	96.15	90.9	93.75
Specificity	98.48	98.59	98.36	98.68	98.59
f-score	95.23	93.75	96.15	90.9	93.75
Accuracy*	–	–	100	–	–

* implies if suppose the accuracy was 100%, there will not be any indication whatsoever about the various stages of DR

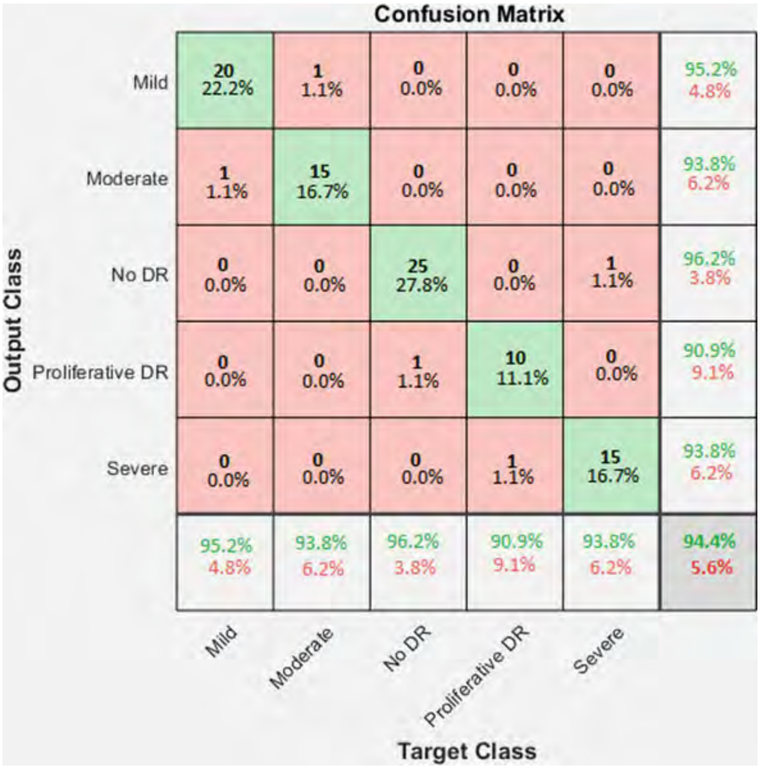


Fig. 13 Confusion matrix

References

1. T. Shanthi, R.S. Sabeenian, Modified alexnet architecture for classification of diabetic retinopathy images. Int. J. Comput. Electri. Eng. **76**, 56–64 (2019)
2. L.B. Frazao, N. Theera-Umpon, S. Auephanwiriyakul, Diagnosis of diabetic retinopathy based on holistic texture and local retinal features. Int. J. Info. Sci. **475**, 44-66 (2019)
3. T. Jemima Jebaseeli, C. Anand Deva Durai, J. Dinesh Peter, Segmentation of retinal blood vessels from ophthalmologic diabetic retinopathy images. Int. J. Comput. Electri. Eng. **73**, 245–258 (2019)
4. D. Devaraj, R. Suma, S.C. Prasanna Kumar, A survey on segmentation of exudates and microaneurysms for early detection of diabetic retinopathy. ILAFM2016 **5**(4), 10845–10850 (2018)
5. S. Wan, Y. Liang, Y. Zhang, Deep convolutional neural networks for diabetic retinopathy detection by image classification. Int. J. Comput. Electri. Eng. **72**, 274–282 (2018)
6. P. Uma, P. Indumathi, Remote examination of exudates-impact of macular edema. Healthcare Technol. Lett. IET Publishers **5**(4), 118–123 (2018)
7. Z.A. Omar, M. Hanafi, S. Mashohor, N.F.M. Mahfudz, Automatic diabetic retinopathy detection and classification system. in *7th IEEE Conference on System Engineering and Technology*, (2017), pp. 162–166
8. H. Pratt, F. Coenen, D.M. Broadbent, S.P. Harding, Y. Zheng, Convolutional neural networks for diabetic retinopathy. in *International Conference on Medical Imaging Understanding and Analysis*, vol. 90 (2016), pp. 200–205

9. M. Omar, F. Khelifi, M.A. Tahir, Detection and classification of retinal fundus images exudates using region based multiscale LBP texture approach. in *International Conference on Control, Decision and Information Technologies* (2016), pp. 227–232
10. U. Rajendra Acharya, C.K. Chua, E.Y. Ng, W. Yu, C. Chee, Application of higher order spectra for the identification of diabetic retinopathy stages. J. Med. Syst. **32**(6), 481–488 (2008)
11. J. Nayak, P.S. Bhat, R. Acharya, C. Lim, M. Kagathi, automated identification of diabetic retinopathy stages using digital fundus images. J. Med. Syst. **32**(2), 107–115 (2008)
12. R. Dhaya, Deep net model for detection of Covid-19 using radiographs based on ROC analysis. J. Innov. Image Process. (JIIP) **2**(03), 135–140 (2020)
13. T. Vijayakumar, R. Vinothkanna,Retrieval of complex images using visual sa ‘ liency guided cognitive classification. J. Innov. Image Process. (JIIP) **2**(02), 102–109 (2020)
14. V.C. Lima, G.C. Cavalieri, M.C. Lima, N.O. Nazario, G.C. Lima, Risk factors for diabetic retinopathy: a case– control study. Int. J. Retina Vitreous **2**, 21 (2016)
15. M. Akter, M.H. Khan, Morphology-based exudates detection from color fundus images in diabetic retinopathy. in *International Conference on Electrical Engineering and Information and Communication Technology* (2014), pp. 1–4

Hybrid Level Fusion Schemes for Multimodal Biometric Authentication System Based on Matcher Performance

S. Amritha Varshini and J. Aravinth

Abstract The performance of the multimodal system was improved by integrating both the physiological and behavioral characteristics of an individual. Usually, the fusion is carried out either in the score or feature level. Multimodal biometric system against the unimodal system is explained below. This system was considered in order to overcome several demerits that were found in the former system. The overall recognition rate of the biometric system was improved by implementing the multimodal systems. ECG, face, and fingerprint were integrated in the new level of fusion named hybrid fusion scheme. In the hybrid fusion scheme, the scores from the feature level fusion were fused along with the best unimodal system (ECG) by using score level fusion techniques. Feature vectors were obtained by processing the signal as well as the images obtained from the databases FVC2002/2004, Face94, and PhysioNet (MIT-BIH Arrythmia) after the process of feature extraction. Matching scores and individual accuracy were computed separately on each biometric trait. Since the matchers on these three biometric traits gave different values, matcher performance-based fusion technique was proposed on the specified traits. The two-level fusion scheme (score and feature) was carried out separately, for analyzing their performances with the hybrid scheme. The normalization of the scores was done by using overlap extrema-based min–max (OVEBAMM) technique. Further, the proposed technique was compared with Z-Score, tanh, and min–max by considering the same traits. The performance analysis of these traits with both unimodal and multimodal systems was done, and they were plotted using receiver operating characteristic (ROC) curve. The proposed hybrid fusion scheme has leveraged the best TPR, FPR, and EER rates as 0.99, 0.1, and 0.5, respectively, by using the normalization techniques with the weighted techniques like confidence-based weighting (CBW) method and mean extrema-based confidence weighting (MEBCW) method.

Keywords Multimodal biometric system · Matcher performance · Hybrid level fusion · Weighted methods

S. Amritha Varshini (✉) · J. Aravinth
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*, Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_35

1 Introduction

Biometric recognition systems are the technology that involves both physiological and behavioral traits of a person. These traits are given as the input, and the assurance of the person's identity is confirmed by the system. These systems are used in almost all the fields. Hence, it is mandatory for securing the data. The three major stages of biometrics are as follows authentication or identification, verification, and authorization. Identification is performed with the process of matching. It confirms the identity of a person by comparing the images in the entire database with the scanned image. If the compared data matches with the scanned image for a percentage of about 85%, then the verification is successful and it authenticates for accessing the data. Biometrics is highly secured when compared to the conventional methods. The basic components of the biometric systems are as follows input data, processing unit (identification, verification, and authentication), and output data. Input data are acquired from the sensors. It can be either fingerprint or any other behavioral parameters (face, speech, iris, etc.). The processing unit includes various operations like image enhancement, normalization, feature extraction, and comparison of acquired data from the database. The general working of the system takes the biometric sample from an individual. The features are extracted, and it is then compared with the data that is already stored in the database. If the input sample matches with the data that are present in the storage, then it allows to access the resources. Multimodal biometric systems are designed using more than one trait like both physiological and behavioral. The inputs are acquired from multiple sensors, and they are processed.

ECG biometric is advantageous when compared to other biometrical characteristics that are commonly involved. In this paper, face and fingerprints are interfaced along with ECG. Though these two are used commonly in recognition systems, it has several demerits and it is prone to many vulnerable activities. The major issues found in face recognition systems are 1. It is sensitive to artificial disguise 2. It has troubles with image quality and size 3. It is subjected to many false identifications 4. The influences of camera angle are more 5. Two-dimensional face recognition is commonly practiced in all the fields so it is considered to be insecure and it can be easily spoofed 6. People can be identified from the distance; it leads to privacy issues mostly. In the case of fingerprint, it can be recreated using latex. To overcome these disadvantages, fusion of ECG biometric along with face and fingerprint is suggested. When ECG is used as the biometric, many illegal activities can be easily avoided. Because it is completely unique to an individual, measuring of ECG from an individual is also not a tedious process, where it can be detected by using single-electrode system. Just like acquiring the verification data from an individual, ECG system also captures the data by placing the electrode on the surface of skin. ECG signal is considered to be heterogenous, because in the ECG there is a slight difference in their amplitude and time interval for every individual. The morphology of heartbeats also differs from person to person depending upon their body structure, living environment, and lifestyle. The only threat to ECG biometric is the presence of any cardiovascular diseases in the heart. Even then it works efficiently if the electrical transverse regions like sino-atrial node (SA node) and atrioventricular node (AV node) are not affected. The ECG biometric can be designed by measuring arrythmia or normal sinus rhythm.

The level of fusion initiates the process of extraction from the raw data that are collected from the sensor level. In score level fusion, the matching scores are obtained from different matchers and it is easier to fuse the scores. But additionally, it requires various normalization or weighted techniques for enhancing the overall recognition rate. Similarly, in feature level fusion the performances of the matching techniques are not considered. In order to rectify these issues, a new level of fusion called hybrid fusion scheme is suggested in this the final scores from feature level fusion and the scores of best unimodal systems is combined by implementing score level fusion techniques. The weighted techniques are incorporated in the hybrid scheme, and hence, better accuracy rate for the proposed multimodal system is achieved using this method.

The paper is arranged as follows, Sect. 2 deals with the literature survey that were executed previously related to the work. Section 3 gives the various methods that were used to process the data followed by the experimental results in Sects. 4 and 5 which gives the conclusion of the paper.

2 Related Works

The authors "Aboshosha, Ashraf et al." in their paper [1] "Score Level Fusion for Fingerprint, Iris and Face Biometrics" described the score level fusion using min–max normalization. The feature extraction of face and fingerprint was explained in detail. Face and fingerprint images were preprocessed. Minutiae algorithm was used in fingerprint for extracting the points. Whereas in face images local binary patterns (LBP) were implemented. Face images were divided into small cells, and the important features were extracted. The scores between 0 and 1 were obtained using min–max normalization technique. The technique has less accuracy, and it took more computational time. The authors "Bashar et al. and the authors "Abhishek et al." in their paper [2–4] "ECG-based biometric authentication using multiscale descriptors: ECG-based biometric authentication" explained about ECG authentication system wherein the techniques like multiscale features extraction from ECG signals were implemented. In the preprocessing stage, nonlinear filters were used. Further, the filtered signal was divided into multiple segments. Mean classifier method was adopted by computing simple minimum distance. After determining the matching and non-matching factors, the accuracy rate of ECG biometric system was found out. This paper [5] by the authors "Mingxing He et al." explains about the performance evaluation in the multimodal biometric system. Wherein the fusion of various traits like fingerprint, finger vein, and face were execution. The performances were validated for sum rule-based score level and support vector machine (SVM)-based classifiers. The new normalization scheme derived from min–max normalization technique was proposed in this paper. This normalization is exclusively for reducing high-score effect. Then the final results were compared between sum rule-based fusion and SVM-based fusion. The authors "Kabir et al." described about the two-level fusion for multimodal biometric system in their paper [6–8]. The new techniques, namely matcher performance-based (MPb) fusion scheme and overlap extrema variation-based anchored min–max (OEVBAMM) normalization scheme, were proposed in the paper. The fusion was carried out in both score and feature levels, and the decision was made based on their performances. The major drawback of this paper is that the performances were indicated with respect to ROC and DET curves. The values of ROC are lesser than GAR, and the DET values were greater than FAR values. The fusion scheme was not described in dual or trilevel system. In the paper [9, 10] the authors gave the overview of multimodal biometrics, that includes the limitations of unimodal systems as well. He also explained in detail regarding the fusion strategies with various transformation or normalization techniques that can be incorporated. The papers [11, 12] from the authors "Rattani et al. and Aravinth et al." explained about the techniques used to implement feature level fusion using the different biometric sources, namely face and fingerprint [13]. It gave a vital information about various techniques that were involved in the process of features compatibility and features reduction. The authors "Singh et al." gave different techniques for fusing ECG with unobtrusive biometrics like face and fingerprint in their paper [12, 14]. This paper explains about the procedures involved in processing of ECG signals acquired from the online datasets. The first and foremost method involves ECG delineation. QRS complex delineator employs fiducial of QRS complex, i.e., QRS onset and QRS offset P wave delineation is calculated from QRS onset (beginning of heartbeat). Similarly, T wave is computed as follows QRSoffset + 80 ms to QRSoffset + 470 ms. The authors "Vishi et al." explained about the various evaluation approaches in their paper [15]. They used score normalization and fusion techniques for the two modalities fingerprint and finger vein. The individual scores from the two sources were combined, and they were normalized using the traditional techniques, namely min–max, Z-score, and TanH. The best improvement rates were achieved up to 99.98%. The authors "Aravinth et al." explained about the implementation of multimodal biometric systems for remote authentication purpose using multi-classifier-based score level fusion in their paper [12, 15].

3 Methodology

See Fig. 1.

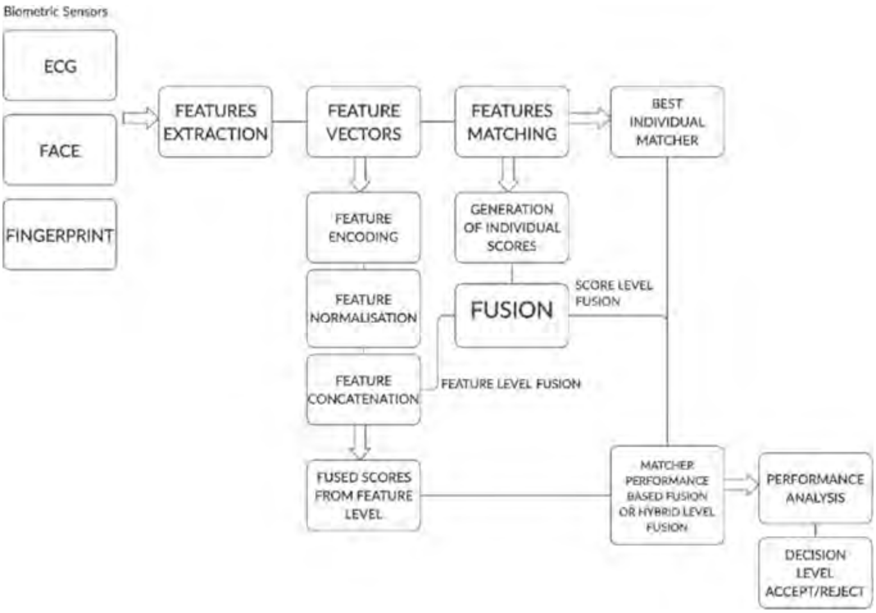


Fig. 1 Block diagram of hybrid fusion-based multimodal biometric system

3.1 ECG Signal for Individual Authentication

The acquired ECG signal undergoes several processes for assuring the quality of the signal to provide best recognition system. 1. Preprocessing of the signal is done in order to remove several artifacts like baseline wandering and noises. The noises were extracted from the signal using median filters. 2. Segmentation is the technique that is used for detecting the beats that are present in the signal. It is applied mainly to detect the *R*—peaks from the ECG signal. 3. Features are extracted from the signal for subsequent analysis of ECG. The features are described based on three different categories, namely (i) Morphology of RR peaks (ii) Interval, and (iii) Normalization of detected RR peaks. The morphology involves higher-order statistics (HOS) and QRS complex delineation. HOS gives the normal distribution of the signal in both the extreme ends with a center point. QRS complex delineation is done using very old algorithm called Pan–Tompkins algorithm. This repository consists of two different classes, namely QRS off and QRS on. QRS off is used for extracting the features from the signal. Euclidean distance is computed among the two different R peaks, and the remaining four points can be extracted from it by assigning the maximum and minimum intervals. The second category includes the intervals that are present in the signal. They are computed with respect to the R peaks, and they are as follows PQ, QR, RS, and ST and the last one is the normalized RR that provides the average value for the detected intervals. 4. Classification of the features is done using the machine learning algorithm called KNN. The individual scores and accuracy rate for the ECG biometric system are computed, and their receiver operating characteristic (ROC) for true positive and false positive values are plotted.

3.2 QRS Delineation

QRS delineation is determined by using Pan–Tompkins algorithm. The raw signal is used as an input. According to the algorithm, the signal is processed in several stages, namely filtering and thresholding. The filtering is done using low-pass and high-pass filters, and this process is very much useful in removing the noises and other distortions present in the signal. The filtered signal is then differentiated to find the signal segments. Later then squaring and integration are done in order to make the signal more distinct. Hence, the integrated signal is further used for the peak detection purposes. The QRS complex is identified by adjusting the thresholds. The process of thresholding and filtering enables to detect the sensitivity and relatively false positive QRS complex. Two values, namely QRS on and QRS off, are assigned to detect the different waves in the signal.

3.3 Fingerprint Recognition

The set of two databases was used for fingerprint authentication system, namely FVC2002 and FVC2004. Several image processing techniques were adopted for extracting the minutiae points. Image processing techniques, like histogram equalization and binarization, were carried out for removing the dots and for smoothening the edges. Fingerprint recognition includes various steps like preprocessing, feature extraction, post-processing, and matching. Preprocessing involves two different stages, namely histogram equalization and binarization, and these two techniques were executed to enhance the images. Binarization is used to convert the grayscale images into binary images. It is also used to indicate the number of ridges and furrows present in an image. Image segmentation is done in order to locate the region of interest (ROI) for extracting the minutiae points. Thinning and skeletonization were performed in the image for obtaining the minutiae points. The minutiae points were extracted by considering the two phenomena, namely bifurcation and termination. After extracting the minutiae points, the false points were removed in the postprocessing stage and the final stage is the fingerprint matching which was done using hamming distance and the matching scores were generated using the Flann-based matcher.

3.4 Facial Recognition

The face samples were collected from the database Face94. People having different facial expressions were taken. Among many images in the database, the top 150 images were taken and processed for finding its individual recognition accuracy. The set of features like eyes, nose, and lips was extracted from the image using haar cascade classifiers. The data transformation was done using principal component analysis (PCA). The eigenfaces were projected using the orthonormal basis, and the obtained eigenvectors from the analysis were compared in accordance with the Euclidean distance. The normalized images were used for the purpose of training using PCA algorithm. The face recognition with local binary patterns (LBP) was also performed in order improve the accuracy rate. The regions of face were first divided into smaller parts, and the histogram was computed. Later then, they were acquired in the form of vectors, which was extracted and concatenated. The regions were labeled based on the pixels of an image. 3*3 neighborhood pixel for each value is considered. The results were computed in both binary and decimal numbers. The matching scores computed using hamming and Flann-based matcher. The individual scores and accuracy were determined for face recognition, and the ROC is obtained against true positive and false positive values.

3.5 Score Level Fusion Technique

Score level fusion is one of the common techniques that is used in case of the multimodal biometric system. It involves the matching scores. Because it is sufficient to determine the true positive and false positive values for the biometric system, these scores were acquired after the process of feature extraction and features matching. Initially, the features matching was executed separately for all the traits and the scores of unimodal systems are recorded. Further, the fusion is carried out by applying certain algorithms like sum and product rule and then the feature matching was carried out once again for the fused system. Since the matchers gave different values, the new system called "Matcher performance-based fusion (Mpb)" is performed, wherein the three biometric sources produced various set of values at the matching stage after the successful completion of feature extraction. The feature values from ECG, fingerprint, and face are denoted as $x(k)$, $y(k)$, and $z(k)$, they were allotted these annotations with respect to the positions (x, y). The features encoding was carried out, and they were encoded as $X_{xy}(k)$, $Y_{xy}(k)$, and $Z_{xy}(k)$. The hamming distance was computed for all these features after encoding them. The matching of the features was for the set of feature values $x(k)$, $y(k)$, and $z(k)$ to find its true positive (TP) and false positive (FP) values. The threshold was automatically fixed with respect to maximum TP values and minimum FP values. The EER, FPR, and TPR values were found out for the fused modality system using the below formula, where the values of TPR was higher when compared to FPR and EER.

$$\text{Threshold (Th)} = \max(\text{TP}) - \min(\text{FP})/K \tag{1}$$

where K represents the empirical parameter. The parameter K is used to control the variable threshold values.

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{3}$$

$$\text{EER} = \text{TPR} + \text{FPR}/2 \tag{4}$$

3.6 Feature Level Fusion Technique

In the case of feature level fusion, the extraction of features from the different modalities was carried out individually. Then the features were made compatible in order to perform feature concatenation as the features from three different biometric sources were varying in their dimensions. By performing features reduction and concatenation, the dimensions of the entire feature pointsets were reduced. The feature reduction was done separately for the traits, to make their dimensions equal. This process was executed prior to features pointsets fusion. The various dimensionality reduction techniques were implemented "neighborhood minutiae elimination" and were used in the case of fingerprint. "Points belonging to specific regions" were performed in face, and for ECG, "Neighborhood features elimination" were done. The reduced features were recorded in each category, and the fusion is done using the reduced feature pointsets. The fused features of the query images were matched with the features that were already stored in the database using the technique "point pattern matching" by Euclidean distance. This fusion process is none other than concatenating the features from various sources, where these features were said to have good differentiation compared to individual values. The TP and FP values were computed for the reduced pointsets in both the stage, i.e., before and after fusing. The TPR, FPR, and EER values were recorded.

3.7 Hybrid Level Fusion Technique

By considering the several disadvantages in the above-mentioned fusion scheme, hybrid level fusion is suggested in case of multimodal biometric systems for improving its overall recognition rate. In score level fusion, the values are improved and the TPR is higher, but still the mismatch in the FPR and EER values along with

TPR is prevailing. This is because, score level fusion requires additional transformation technique or weighted technique to give appropriate results. Similarly, in feature level fusion reduction in features is one of the major factors that reduce the TPR and increases the other two. Feature fusion can be considered for best authentication system in case if it includes the rate of features matching. Since it does not have the relevant information, the single level fusion scheme alone cannot be made reliable in the multimodal biometric system. The new level of fusion scheme can be used to overcome the limitations. Hybrid level fusion scheme includes the fusion of final scores from the feature level fusion and the score of best unimodal system. These two scores were combined by means of score level fusion. The results of hybrid level fusion were recorded. It was found that the TPR was higher and the mismatch among FPR and EER values was corrected. The performance of HBF scheme is compared with other two-level of fusion. The main aim of weighted hybrid level fusion was achieved by implementing confidence-based weighting (CBW) method and mean extrema-based confidence weighting (MEBCW) method. It was observed that the values were reduced comparatively in FPR and EER. The overall recognition rate was achieved using WHBF. The performances were compared with score level fusion under two cases, namely with and without involving weights.

3.8 Weighted Hybrid Level Fusion

The improved recognition rate is degraded in the case of commonly used fusion techniques, due to dissimilar EER or TPR values. In order to provide better performance of the multimodal system, "Weighted rule-based fusion" is used. The two different weighted techniques, namely "Confidence-based weighting technique (CBW)" and "Mean Extreme-based confidence weighting technique (MEBCW),"are implemented. These techniques are applied on the fused system using the generated final scores.

Confidence-based weighting technique (CBW)—In this technique, the highest weights were set to the system, resulted in best matching. The non-overlapping or false positive scores were considered and discarded. So, that the remaining true positive values gave better recognition rate.

Mean extreme-based confidence weighting technique (MEBCW)—In the case of MEBCW method, it considers both overlapping and non-overlapping between the true positive and false positive values.

Hence, the weighted techniques are applied in both score and hybrid level fusion, and their results are compared using the below formulae. To compute the confidence weights between the scores of feature level fusion (S_1) and the scores of best unimodal systems (S_2), two parameters were defined namely "a" and "b" and it can be represented as follows,

$$a = [(\max(\mathrm{FP}) - \mathrm{mean}\ (\mathrm{FP})) + (\mathrm{mean}(\mathrm{TP}) - \min(\mathrm{TP}))] \tag{5}$$

$$b = [(\max(\mathrm{TP}) - \mathrm{mean}(\mathrm{TP})) + (\mathrm{mean}(\mathrm{FP}) - \min(\mathrm{FP}))] \tag{6}$$

3.9 Overlap Extrema-Based Min–Max (OVEBAMM) Normalization Technique

The suggested techniques for the ECG, fingerprint, and face are overlap extrema variation-based min–max normalization technique (OEVBAMM). From this technique, the normalized scores can be determined after executing the process of fusion. The normalization is carried out using the final fused scores. In this technique, only the overlapping scores in TP values are taken into account. As the first step, overlap extrema variation is calculated. Let us assume the anchor value "A" and the equation is as follows,

$$A = \mathrm{Max}(\mathrm{FPR}) - \mathrm{Min}(\mathrm{TPR})/\mathrm{Std}(\mathrm{FPR}) - \mathrm{Std}(\mathrm{TPR}) \tag{7}$$

Equation (7) shows the difference between the maximum value of FAR and the minimum value of TPR, and in the denominator, standard deviation is determined for the values and the anchor value is generated. With these values, it is very much easier to find the normalized scores for the fusion technique under two cases.

(i) When $(S1) \leq A$

$$S = S1 - \frac{\mathrm{Min}(\mathrm{FPR},\ \mathrm{TPR})}{2(A) - \mathrm{Min}(\mathrm{FPR}, \mathrm{TPR})} \tag{8}$$

(ii) When $(S1) > A$

$$S = 0.5 + (S1) - \frac{A}{\mathrm{Max}(\mathrm{TPR}, \mathrm{FPR}) - A} \tag{9}$$

where $S1$ represents the scores from the fused system. In Eq. (9), 0.5 represents the threshold value that has been assigned. Hence, Eqs. (8) and (9) are used for generating the normalized scores (S) by employing OVEBAMM technique.

4 Results and Discussions

4.1 Dataset Preparation

The datasets used in the experiment are PhysioNet, Face 94, FVC2002/2004. In face & fingerprint, 80 users were selected. The features were extracted, and the samples were trained and tested. In the sample data, five of the images were training data and the remaining three were the testing data. All these databases are independent from each other because there is no common database contains "fingerprint and face" biometrics of the same person. Hence, the labeling of these databases was done and the custom dictionary is created. ECG recordings were extracted from MIT-BIH Arrhythmia Database and European ST-T database from Physionnet.com. ECG recordings were taken from healthy persons (Men Age Groups 20–45, Female Age Groups 20–50). The sampled frequency of the signal was about 128 Hz. Lead I ECG recordings were used in the experiments. Both raw and filtered signal were used, and they were represented in "mV." The maximum duration of the signal taken was about 1 min. The first 40 s was considered to be the training data, and the remaining 20 s was the testing data. Each recording was sampled at 256 samples bits/seconds.

4.2 Feature Extraction

Figure 2 shows the QRS delineation that was executed using Pan–Tompkins algorithm. QRS complex is shown. It also detects *P*, *Q*, *R*, *S*, and *T* waves. Figure 3 shows the detected *R* peaks that were derived after applying the algorithm in the input signal.

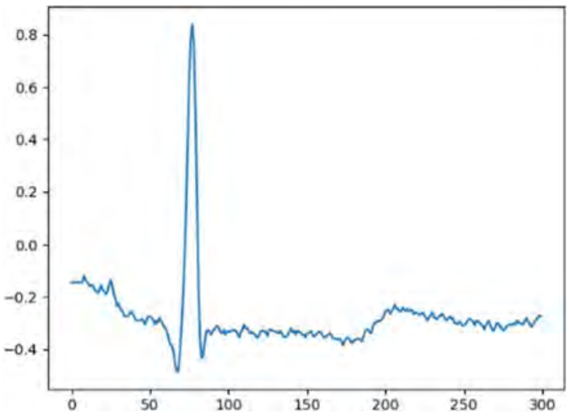


Fig. 2 Extraction of QRS complex

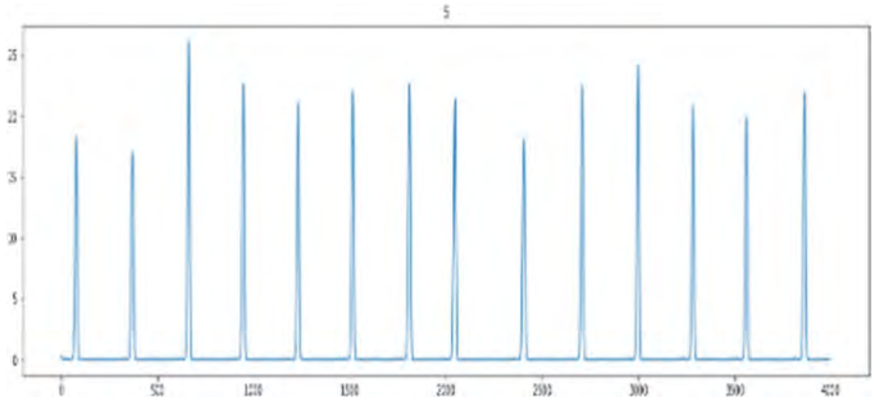


Fig. 3 Detection of *R* peaks

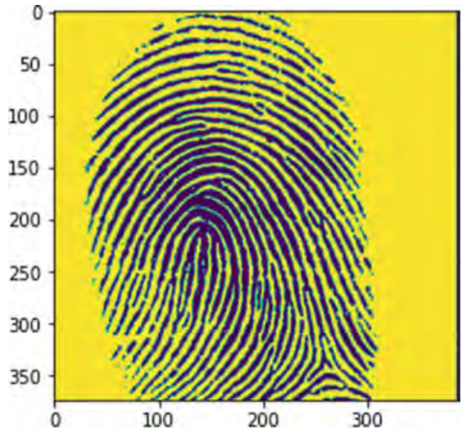


Fig. 4 Original image

Figure 4 shows the original image, and Fig. 5 is the minutiae points extracted image. The points were marked in red. The major points considered are termination and bifurcation.

Figure 6 shows the set of features like eyes, nose, and lips were detected using haar cascade classifiers, and they were extracted for the training data and these were the features used for finding the matching scores.

4.3 Performance Analysis

Score Level Fusion

See Fig. 7.

Fig. 5 Minutiae points extraction

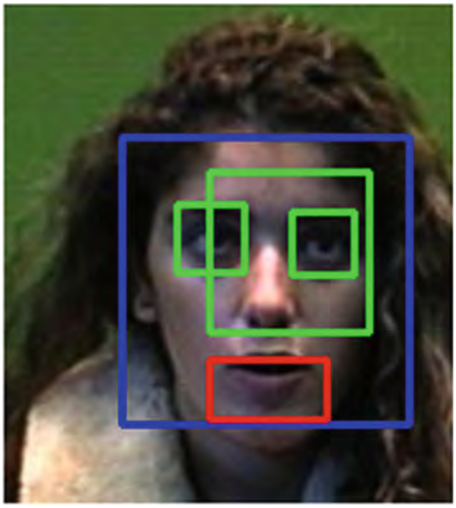


Fig. 6 Face–feature extraction

The performance analysis of the system in single trait as well as the combination of two traits was analyzed. The traits were fused by implementing "Matcher performance-based fusion scheme.' Because of this method, better performance was achieved. It was evaluated using ROC curve analysis. The highest rate of 0.99 was obtained in this designed system. The FPR and EER were 0.34 and 0.66. False positive rates were reduced gradually, and the desired rates were acquired (Table 1).

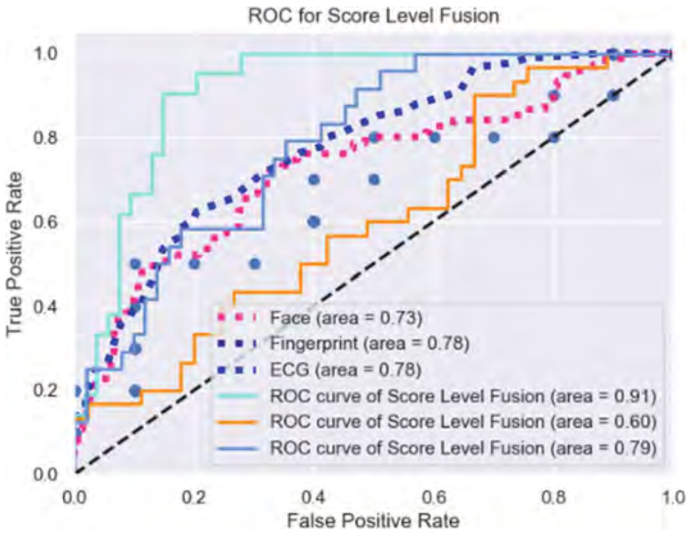


Fig. 7 Performance analysis of face, ECG, and fingerprint employing matcher performance-based fusion scheme–score level fusion

Table 1 Performance evaluation of score level fusion techniques

Recognition system	FPR	TPR	EER
Fingerprint	0.64	0.95	0.79
Face	0.67	0.94	0.80
ECG	0.63	0.96	0.79
ECG + Fingerprint	0.42	0.98	0.7
ECG + Face	0.50	0.97	0.73
Face + Fingerprint	0.57	0.95	0.76
ECG + Face + Fingerprint	0.34	0.99	0.66

Feature Level Fusion

See Fig. 8.

Due to the issues in dimensions, the feature reduction techniques were used in all the three modalities individually. The fusion techniques were carried out with the set of reduced features. After analysing the feature fusion in the combination of two traits, the fusion of ECG, face, and fingerprint is performed. The performance analysis of feature level fusion was analyzed using ROC curve. It was found that the FPR and EER values were reduced. The TPR values were also less when compared with the previous fusion state. It may be because of feature elimination process. The values were TPR 0.97, FPR 0.28, and EER 0.62 (Table 2).

Hybrid Level Fusion

See Fig. 9 and Table 3.

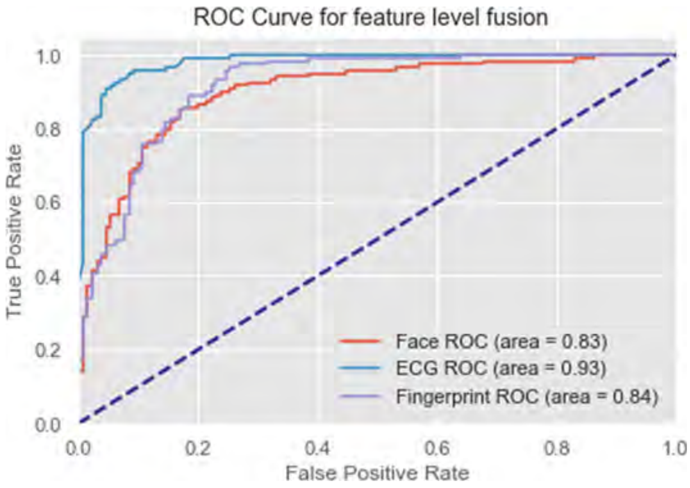


Fig. 8 Performance analysis of face, ECG, and fingerprint employing matcher performance-based fusion scheme–feature level fusion

Table 2 Performance evaluation of feature level fusion techniques

Recognition system	FPR	TPR	EER
Fingerprint	0.54	0.95	0.74
Face	0.46	0.94	0.7
ECG	0.29	0.97	0.63
Fingerprint + Face	0.62	0.88	0.75
Fingerprint + ECG	0.46	0.91	0.68
ECG + Face	0.41	0.93	0.67
ECG + Face + Fingerprint	0.28	0.97	0.62

The scores of feature level and the score of best unimodal system were fused using score level techniques. As observed ECG was the best system that gave very high rates. Hence by fusing, the scores were TPR 0.99, FPR 0.13, and EER 0.56. It gave the results without any dissimilarities. It was observed that both the FPR and EER rates were reduced, and the better TPR rate was obtained.

5 Conclusion

The fusion technique named "Hybrid Fusion Scheme" was proposed for the biometric sources' ECG, face, and fingerprint. This technique was executed by the final scores of feature level fusion for the same specified traits. The scores were FPR 0.28, TPR 0.97, and EER 0.62. These scores were fused along with the scores of best unimodal systems by implementing score level fusion techniques. By completed analysis, ECG

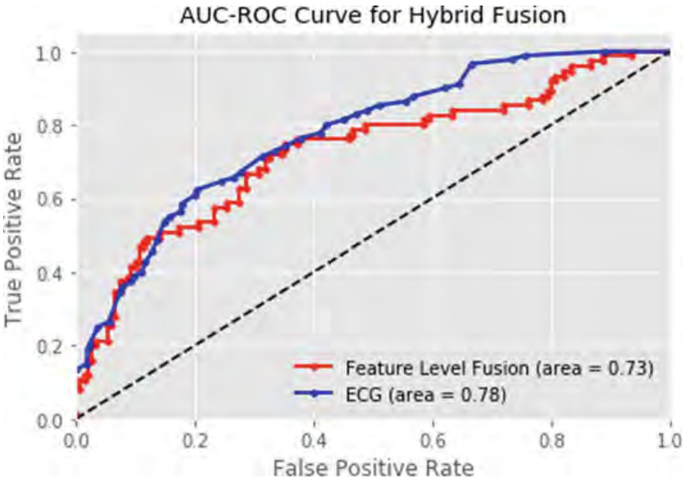


Fig. 9 Performance analysis of matcher performance-based fusion scheme–hybrid level fusion

Table 3 Performance evaluation of hybrid level fusion techniques

Recognition system	FPR	TPR	EER
(ECG + Feature Level Fusion)	0.13	0.99	0.56

was found be the best single modality system among the three and the scores were FPR 0.29, TPR 0.97, and EER 0.63. After combining it was found that the scores were reduced ultimately in the case of FPR and EER, and they were FPR 0.13, TPR 0.99, and EER 0.56. Our aim providing reduced FPR and EER for the authentication system was achieved. In order to stabilize the system and to establish the accuracy rates of the designed system, weighted techniques like confidence-based weighting technique (CBW) and mean extreme-based confidence weighting technique (MEBCW) have been incorporated with the proposed hybrid fusion scheme. Hence, it was observed that system produced very low EER and FPR in the case of weighted hybrid fusion scheme. The scores were FPR 0.1, TPR 0.99, and EER 0.5. The dissimilarities among the EER and TPR values were commonly found in the case of other fusion schemes, but it was not found in hybrid level of fusion with weighting techniques. Hence, it is an added advantage of the proposed method. Thus, it was proved that if the TPR is high, then the system is considered as the best authentication system.

References

1. A. Aboshosha, K.A. El Dahshan, E.A. Karam, E.A. Ebeid, Score level fusion for fingerprint, iris and face biometrics. Int. J. Comput. Appl. **111**(4), 47–55 (2015)
2. M. Be, A.M. Abhishek, T.J, V. K R, L.M. Patnaik, Multimodal biometric authentication using ECG and fingerprint. Int. J. Comput. Appl. **111**(13), 33–39 (2015)

3. M. He et al., Performance evaluation of score level fusion in multimodal biometric systems. Pattern Recognit. **43**(5), 1789–1800 (2010)
4. W. Kabir, M.O. Ahmad, M.N.S. Swamy, A multi-biometric system based on feature and score level fusions. IEEE Access **7**, 59437–59450 (2019)
5. S.A. Israel, W.T. Scruggs, W.J. Worek, J.M. Irvine, Fusing face and ECG for personal identification. in *Proceedings–Application Imaging Pattern Recognition Working.* vol. 2003. (2004) , pp. 226–231
6. W. Kabir, S. Member, M.O. Ahmad, M.N.S. Swamy, Weighted hybrid fusion for multimodal biometric recognition system. (2018), pp. 3–6
7. A. Rattani, D.R. Kisku, M. Bicego, M. Tistarelli, Feature level fusion of face and fingerprint biometrics. (2007), pp. 0–5
8. Y.N. Singh, S.K. Singh, P. Gupta, Fusion of electrocardiogram with unobtrusive biometrics: an efficient individual authentication system. Pattern Recognit. Lett. **33**(14), 1932–1941 (2012)
9. S. Udhary, R. Nath, A multimodal biometric recognition system based on fusion of palmprint, fingerprint and face. in *2009 International Conference Advanced Recent Technology Communications Computing* (2009), pp. 596–600
10. K. Vishi, V. Mavroeidis, An evaluation of score level fusion approaches for fingerprint and finger-vein biometrics. no. Nisk (2017)
11. J. Aravinth, S. Valarmathy, Multi classifier-based score level fusion of multi-modal biometric recognition and its application to remote biometrics authentication. Imaging Sci. J. **64**(1) (2016)
12. S. Prasad, J. Aravinth, Serial multimodal framework for enhancing user convenience using dimensionality reduction technique. in *Proceedings of IEEE International Conference on Circuit, Power and Computing Technologies*, ICCPCT (2016), art. no. 7530162. https://doi.org/10.1109/ICCPCT.2016.7530162
13. P. Sharma, K. Singh, Multimodal biometric system fusion using fingerprint and face with fuzzy logic. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **7**(5), 482–489 (2017)
14. S. Veni, S. Thushara, Multimodal approach to emotion recognition for enhancing human machine interaction—a survey. Int. J. Adv. Sci. Eng. Inf. Technol. **7**(4), 1428–1433 (2017)
15. B. Saiharsha, A. Abel Lesle, B. Diwakar, R. Karthika, M. Ganesan,Evaluating performance of deep learning architectures for image classification. *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, (2020), pp. 917–922. https://doi.org/10.1109/ICCES48766.2020.9137884

Evolutionary Computation of Facial Composites for Suspect Identification in Forensic Sciences

Vijay A. Kanade

Abstract Suspect identification is an integral and significant part of forensic sciences. The detection of face is an important issue when the face of a suspect is described by a witness. The article aims to simplify the process of creating facial composites and thereby generate the images of facial likeness by utilizing an advanced genetic algorithm. The disclosed procedure is comfortable for a forensic technician and natural for a witness at the same time. The proposed genetic algorithm evolves from the original algorithm by employing a feedback-loop mechanism, wherein the most optimal solution is obtained for the generated facial composite.

Keywords Facial composite · Genetic algorithm (GA) · Suspect identification · Interactive evolution strategy

1 Introduction

Evolutionary algorithms (EAs) are nature-inspired problem-solving methods that help in the optimization of the associated solutions. In the current research, the genetic algorithms are used for designing a computer-based automatic face generation platform.

As it is well known in forensic sciences, suspect identification forms the most significant part of any forensic investigations. Such investigations generally depend on the witness's description of the suspect. Currently, there are various techniques and algorithms of facial composite generation and face recognition that assists in visualizing a suspect's face [1]. Today, 'facial recognition technology' helps law enforcement officials and police departments identify an individual of interest by scanning the available photos or video footages (Fig. 1).

In the above process, the 'extracting step' is of great importance where the facial recognition is performed using genetic algorithms to track the facial data points and develop a face structure that matches the data residing in the database [2]. Hence,

V. A. Kanade (✉)
Researcher, Pune, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_36

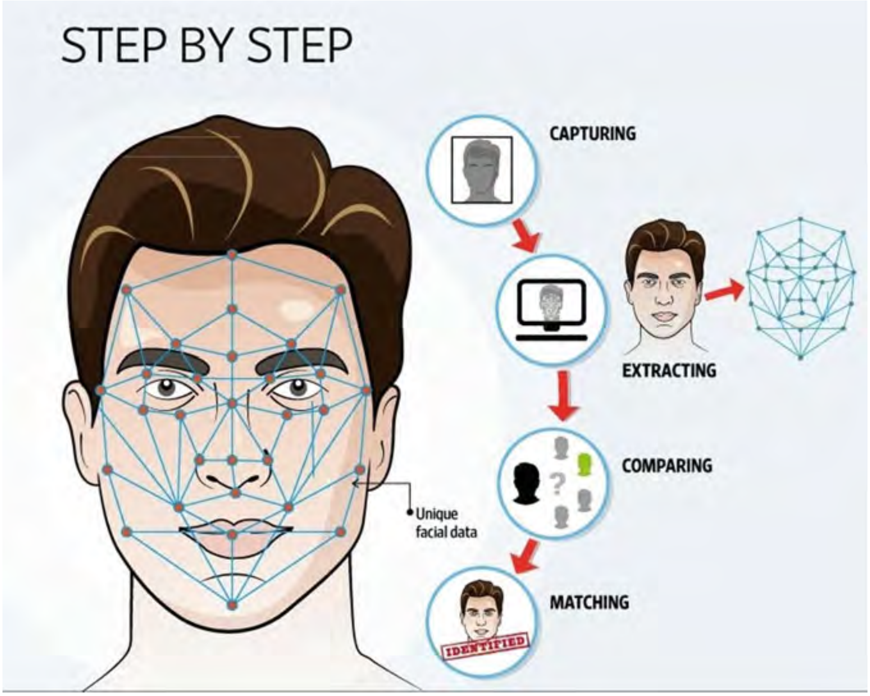


Fig. 1 Facial recognition technology [7]

face recognition technology essentially requires stored data to perform the matching step of facial recognition.

Further, with today's technological advancements, even the criminals have advanced their modus operand, wherein the suspected criminals rely on using masks that can trick the face recognition algorithms embedded in smart cameras, or change their walking posture such that it eludes person identification. In such cases, the role of the witness becomes crucial. The proposed technique is inspired by the face generation tool EvoFIT; however, the proposed evolutionary method fine-tunes the results generated by previously known tools [1]. This optimization is achieved by evolving the results in a manner that simplifies the work of the involved technician or personnel and makes it easier for the witness to command the developed platform. Besides, our approach is independent of any stored data as opposed to facial recognition technology.

2 Genetic Algorithm for Facial Composite Generation

Genetic algorithms fall under the roof of evolutionary algorithms [3, 4]. A population is defined by a chromosome set where each chromosome symbolizes a potential solution of the algorithm. Chromosomes have numeric values that are identified as "genes." The genetic algorithm begins with a random generation of the initial population. Further, the chromosomes undergo an iterative evolutionary process, which involves crossover, mutation, and replacement. On completion of each iteration, the initial population is replaced by the fittest population, and the feedback loop is taken into consideration which further optimizes the results in its second iteration. And the process of feedback-based optimization continues until the most appropriate facial composite structure is generated by the developed tool.

The evolutionary process designed in this research consists of the following primary operators applied on a population size 'N', along with chromosome length denoted by 'n':

1. Selection:
   In the selection step, 'N' chromosome pairs are randomly selected from the parent population.
2. Crossover:
   For each selected pair, two cross-points are selected randomly in the range [0, n]. Every gene identified between these points is switched with probability 'pc'.
3. Mutation:
   The swap mutation is used to obtain the gene values of each offspring chromosome which are swapped and changed with the probability 'pm'.
4. Replacement:
   In the final step, the previous population is replaced by the offspring in case the fitness value of the offspring is optimal than the parent population or vice versa. [5, 6].

2.1 *Evolving Genetic Algorithm: Implementation*

The proposed genetic algorithm (GA) differs from traditional GAs in fitness evaluation and feedback-loop-based image optimization, wherein the users get involved in the algorithm. Such an arrangement helps the witness to visualize the suspect's face and refine it without causing any burden on the technician involved as the algorithm operates on the evolutionary process.

Here, as the witness tries to recollect the suspect's face, the technician plots facial data points on the developed tool. Upon plotting the data points, the GA runs its course, and an optimal result is generated for the plotted data points. The optimal result is based on the fitness function, which is defined as a measure of the minimal distance between the plotted data points. The fitness represents the similarity of individual faces concerning the face structure described by the witness. As the results are generated, the primary face structure of the suspect is displayed to the witness for feedback. Based on the face generated, the witness further describes the facial structure of the suspect, and the additional data points are further plotted by the technician. This leads to refinement and fine-tunning of the facial composite generation process. Thus, an appropriate face is thereby generated by applying evolving computation to the traditionally known GAs.

Pseudocode:

Random initial population generation
Repeat

  Fitness evaluation of individuals
  Reproduction

    Pair selection
    Crossover: Recombine pairs
    Mutation: Swap mutation
    Replace the initial population
    Feedback loop

Run until an optimal solution is produced.

2.2 Experiments and Preliminary Results

The platform is developed in Visual Studio 2017 by using C# language.

In the first step, facial data points are plotted on the developed UI based on the description of the suspect given by the witness. On plotting the data points, GA is executed to produce an optimal result, wherein the basic facial structure of the suspect is generated from the plotted data points (Fig. 2).

Further, in the next step, the generated facial structure is shown to the witness, and feedback is taken *i* to fine-tune the facial generation process. The witness may then provide additional input by describing the suspect further. The technician may then further plot the data point based on the witness's description. The feedback from the witness further refines and molds the generated facial structure (Fig. 3).

The above process continues until 'm' iterations based on the feedback input provided by the witness. Thus, the final generated face may potentially be that of the suspect as described by the witness (Fig. 4).

Similar to the above case, the results are obtained using the evolved GA for five different facial composites separately, to determine the efficiency of the developed evolutionary computation in comparison to the standard GA. The obtained results for each facial composite are given in the table below:

Initially, the facial data points (i.e., 12 in first case) are plotted for facial composite generation by the technician on the developed UI based on the description of the suspect provided by the witness. In the next step, the conventional GA is executed

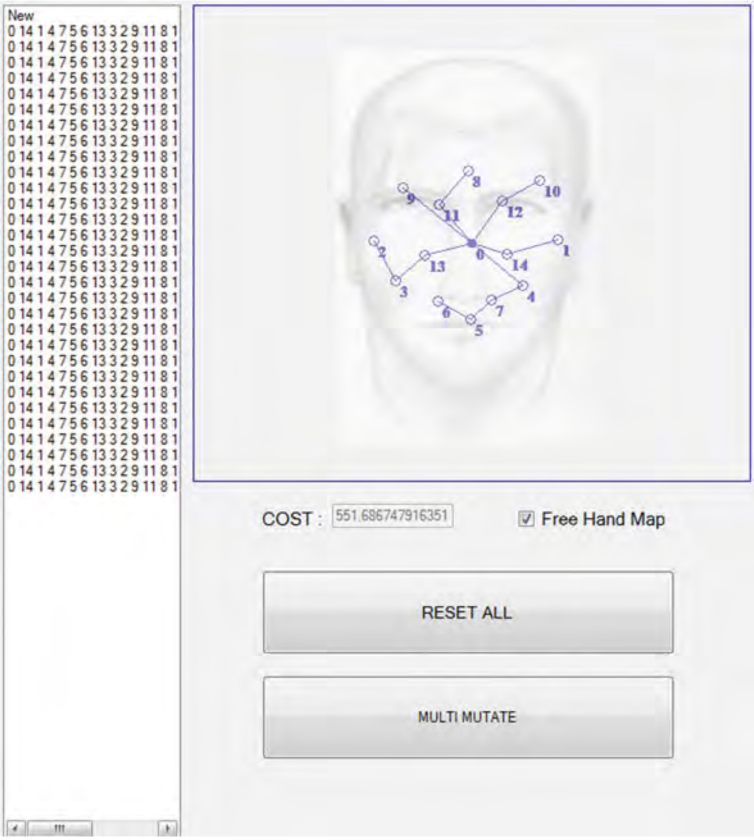


Fig. 2 Snapshot I: Primary facial structure generation based on '15 facial data points'

based on the fitness function, thereby yielding optimal fitness results (820.315 units) for corresponding plotted facial data points (12). In the next step, the optimized result produced during first iteration of GA undergoes evolution (selection, crossover, and mutation) for 'm' iterations to further yield the most optimized fitness result (735.896 units) for the respective facial data points (12). On obtaining the final optimized result, the facial composite is finally generated and displayed to the witness for further feedback. Thereby, the process continues until the final facial composite is accurately generated. Thus, the above table shows that the first facial composite where 12 facial data points were plotted, yielded the optimal fitness result of 735.896 units on employing the designed evolved GA, whereas simple GA yielded 820.315 units for the same facial composite having 12 facial data points. It has been identified that the fitness result of evolved GA generates a more accurate facial composite than when simple GA is used.

Furthermore, for the next four facial composites generated via evolved GA process, 15, 9, 20, and 18 facial data points were potted, and the optimized fitness

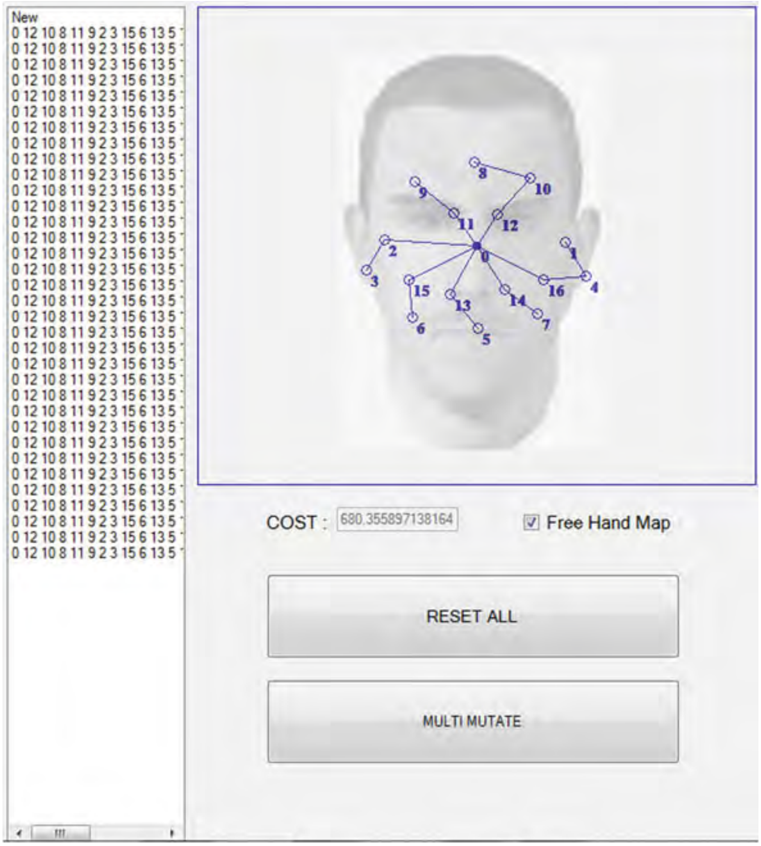


Fig. 3 Snapshot II: Facial optimization based on feedback loop with additional data (17 facial data points)

results obtained were 1053.774 units, 644.082 units, 1480.207 units, and 1144.917 units, respectively. Table 1 therefore validates the efficiency of the developed evolutionary GA as compared to the simple GA, for the plotted facial data points (corresponding to the generated facial composites), in the context of their fitness evaluation (i.e., results). The evolved GA also helps in generating better and accurate facial composites than when traditional GA is used.

3 Conclusion

In this research, a face generation tool is designed to optimize the face generation based on the feedback loop. The tool uses GA for face generation and further incorporates the inputs from the witness to evolve the results further. Hence, the developed

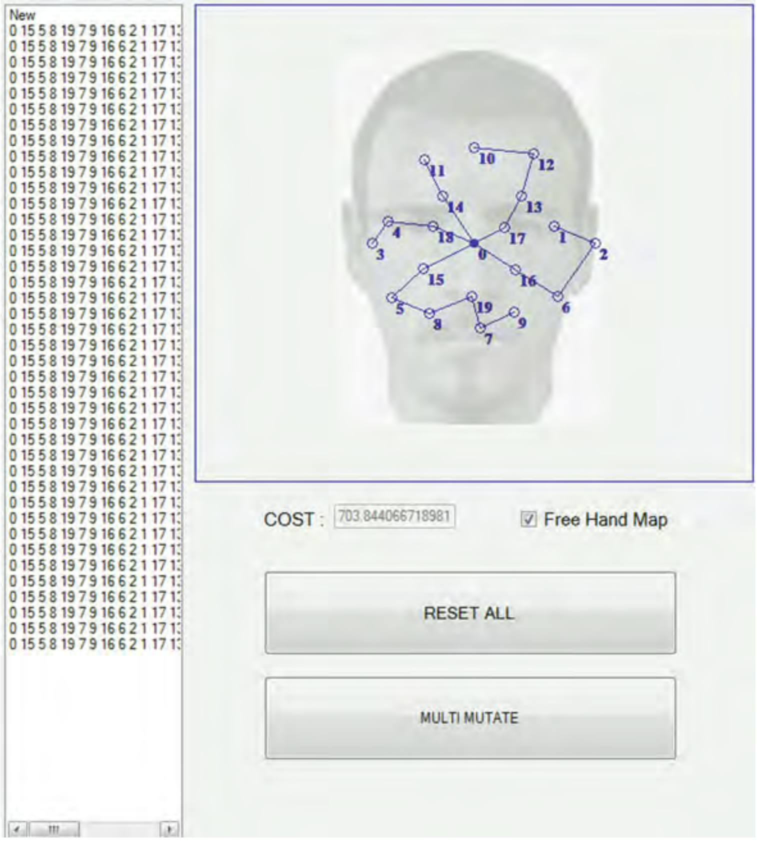


Fig. 4 Snapshot III: Optimized face after '*m*' iterations (20 facial data points)

Table 1 Evolved GA results

Sr. No.	Facial data points for facial composite generation	Fitness evaluation	
		G.A (1st iteration) (units)	E.A ('*m*' iterations) [Optimal result] (units)
1	12	820.315	735.896
2	15	1099.583	1053.774
3	9	717.845	644.082
4	20	1537.701	1480.207
5	18	1252.584	1144.917

platform explores the interactive evolutionary algorithm for optimal face generation of the suspect. The tool is useful in criminal investigations that heavily rely on the descriptions provided by the witnesses.

4 Future Work

The preliminary results of the conducted experiments are encouraging; however, in the future, we intend to widen the data set and validate the face generation process. This includes considering facial properties such as size, shape, hairstyle, beard, and aging factor

Acknowledgements I would like to extend my sincere gratitude to Dr. A. S. Kanade for his relentless support during my research work.

Conflict of Interest: The authors declare that they have no conflict of interest.

References

1. T. Akbal, G.N. Demir, E. Kanlikilicer, M.C. Kus, F.H. Ulu, Interactive nature-inspired heuristics for automatic facial composite generation. in Genetic and Evolutionary Computation Conference Undergraduate Student Workshop (2006)
2. Q. Qing, E.C.C. Tsang, Face recognition using genetic algorithm, 05 December (2014)
3. A.E. Eiben, J.E. Smith, in *Introduction to Evolutionary Computing*, (Springer, 2003)
4. A.E. Eiben, M. Schoenauer, Evolutionary computing. Info. Process. Lett. **82**(1), 1–6 (2002)
5. B. Zahradnikova, S. Duchovicova, P. Schreiber, Generating facial composites from principal components. in *MATEC Web of Conferences, EDP Sciences* (2016)
6. G.N. Demir, Interactive genetic algorithms for facial composite generation. in *GECCO'06*, July 8–12, (Seattle, WA, USA (ACM), 2004)
7. The Future of Face Recognition (2017)

Automatic Recognition of Helmetless Bike Rider License Plate Using Deep Learning

K. V. L. Keerthi, V. Krishna Teja, P. N. R. L Chandra Sekhar, and T. N. Shankar

Abstract Traffic incidents have risen exponentially in the recent years, and most of the people have suffered a severe or dead injury due to the practice of motorbike helmet-free driving. Even though the government has announced it as a punishable crime for not wearing a helmet, it has been violated at many places. Especially, when the traffic is heavy, the traffic constable cannot apprehend all the violators at once. This research work has proposed a model that can automatically identify the bike rider without a helmet and retrieves the bike owner's information by recognizing the license plate. The TensorFlow object detection API is used to identify objects from video frames. The proposed model is trained by using faster R-CNN to recognize bike riders without a helmet. Then, the license plate is recognized using the Tesseract OCR engine and extract owner's information. The experimental results of the model show superior performance with state of the art.

Keywords Image processing · Helmet detection · Faster R-CNN · Optical character recognition · License plate recognition

1 Introduction

Road injuries are a significant cause of concern nationwide. Lakhs of road accident cases are reported each year. India has about 1% of the world's vehicle population and is responsible for around 6% of global road accidents. Around 70% of injuries include [1, 2] young people. It is a fact that about four people die every hour from road accidents because they do not wear a helmet when riding motorbikes. Despite its severity, governments have made it a legal crime not to wear a helmet while riding a motorcycle and have introduced various methods to capture the violators [3].

K. V. L. Keerthi · V. Krishna Teja · P. N. R. L. Chandra Sekhar (✉)
Gandhi Institute of Technology and Management, Visakhapatnam, Andra Pradesh, India
e-mail: cpnrl@gitam.edu

T. N. Shankar
Koneru Lakshmaiah Educational Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_37

Nevertheless, the current techniques are not successful because they require considerable human intervention. In identifying violators, the captured images gave to the model through video monitoring or smartphone photographs taken manually at traffic signals [4, 5]. Therefore, automation is important for accurate and efficient identification of violators and identifying their license plate number to get their information. It then significantly decreases human involvement [6, 7].

However, identifying a bike rider without a helmet and number plate brings many challenges. (i) It requires very high-quality video and proper alignment of camera angle for detecting number plate accurately, (ii) capture the vehicle in motion and any weather condition (iii) requires a large amount of information processed in a short time, (iv) the region of interest objects that occur with other objects and classification is more complicated.

In keeping with these difficulties, a framework for identifying bike riders without a helmet and recognizing their license plate is proposed to obtain the vehicle's owner information.

The rest of the paper organizes as follows: Sect. 2 describes existing work done to address this problem; Sect. 3 describes the proposed methodology, Sect. 4 gives the experimental setup and results, and the last section summarizes the paper.

2 Existing Work

Over the years, several techniques have been proposed to address this problem, and each technique uses a different approach to obtain better results. In video surveillance and character recognition, automatic identification of bike riders number plate falls on the category of anomaly detection. Active surveillance systems usually consist of modeling, identification, monitoring, and classification of moveable objects [8].

In [9], Silva et al. proposed a methodology for a computational vision of helmet detection on public highways. They used the AGMM algorithm to segment bikes and the HOG features to classify the helmet and achieved 91.37% accuracy. However, the technique is computationally expansive due to Hough transform to locate the bike riders head. In [10], Chiverton uses the helmet's geometrical structure and the difference of illumination at various portions to detect the helmet. They use circle arc detection and Hough transform. This technique's drawback is often computationally expansive due to the helmet's full-frame position and sometimes classify similarly shaped items. To minimize computational complexity in the classifier [11], Rattapoom et al. proposed to detect helmet by K-nearest neighbor (KNN). They extracted circularity, average strength, and average hues from each helmet quadrant and fed to classifier KNN. The device displays measured wearing of the helmet at a rate of 74%.

To improve the classification accuracy in [4, 12], a hybrid method is proposed for determining bike rider by extracting HoG, SIFT, and LBP features and classify the region around the bike rider's head to determine the helmet. SVM classifier is used to determine whether or not the bike rider wears a helmet. With a processing time of 11.58 ms per frame, they achieved 93.80% detection accuracy.

Deep learning algorithms such as convolution neural networks (CNN) have recently taken helmet detection and are considered comparably better than conventional methods. Faster R-CNN developed by Microsoft accelerates bikes' detection and recognition rate compared to other deep learning networks such as CNN, fast R-CNN, and YOLO.

In [13], the authors propose two convolutional neural networks (CNNs) to classify whether the bike rider wears the helmet or not. One CNN is used to classify the frame artifacts into motorcyclists and non-motorcyclists, and another CNN is used to classify for the helmet.

In [14–16], the authors classified the helmet, then recognized license plate for those without the helmet and passed to the OCR to identify the registered vehicle number with optical character recognition. This approach obtained more than 91% precision to locate the helmet and achieved 97% and 94% accuracy, respectively, for identifying alphabets and numbers.

In summary, the methods discussed above have two drawbacks: They are either computationally expansive or passive, which implies that they detect only the helmet but do not identify the number plate. The proposed approach overcomes the drawbacks mentioned above and recognizes the number plate of helmetless bike riders automatically.

3 Proposed Method

This section presents the proposed method of recognizing a helmetless bike rider license plate, which performs in five stages. In the first stage, the input frame is segmented and detects bike objects. Then, in the second stage, using the head of the bike rider, it detects whether he is wearing a helmet or not. In the third stage, the bike rider's object without a helmet is fed to the model to detect the license plate. The license plate recognizes in the 4th stage and extracts the bike owner's information in the final stage. The block diagram in Fig. 1 describes the proposed method where the input is a still video frame and passes through the object detection model to detect the bike, head, helmet, and number plate.

3.1 Bike Detection

To detect the bike in the input frame, Faster R-CNN model is adopted to detect objects and labeled as bikes from various objects. Faster R-CNN extracts features from the Region of Interest (RoI)—bike in this case—and uses a single model for extracting features, classifying, and returning the bounding boxes.

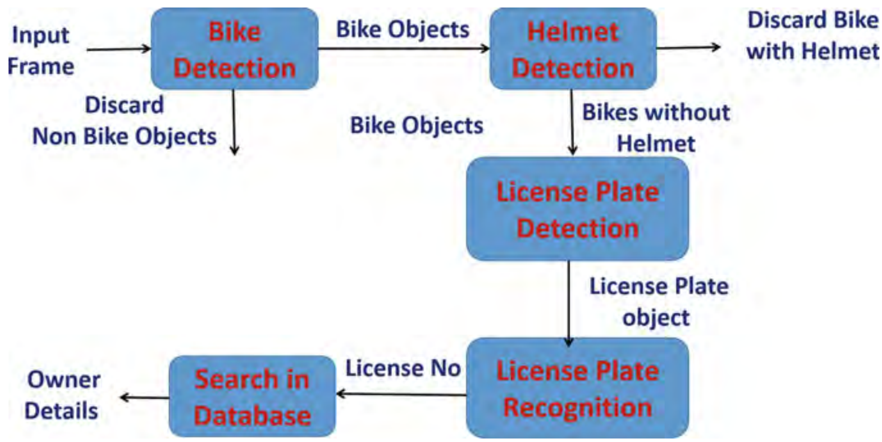


Fig. 1 The proposed method

3.2 Helmet Detection

After the bike is detected, the objected is passed to Faster R-CNN to classify further whether there is a rider. VGG-16 and Faster R-CNN are used as a pre-trained model to distinguish persons, horses, and chairs to identify the rider's head. The upper part of the object is considered as a region of interest while training the Faster R-CNN. Gabor filter is used over the occlusions, confirming the stability and orientation in representing facial segments by that even in different environmental conditions the head can also detect. The frame is then passed to the classifier to detect whether the head is with a helmet. The input frame without a helmet passes further to extract the number plate.

3.3 Detection and Recognition of License Plate

Those frames classified as head without a helmet are further processed to detect the number plate using the Faster R-CNN model. Upon detecting the license plate, using the Tesseract optical character recognition model recognizes those characters in the license plate. The license number is then searched in the database to find the owner's details to further action.

4 Experimental Setup and Results

The proposed method performs multiple tasks to detect helmetless bike riders and recognizes their license plate. The images are trained using the Faster R-CNN-Inception-V2 model, which uses mean average precision (MAP) to get the highest accuracy with TensorFlow object detection API. For better performance, the TensorFlow CPU Configuration is used. After training the model, the bike rider's license plate who wears the helmet is extracted and recognized using the Tesseract OCR model, which takes an image and returns the string detected from the image. The resultant string is the license plate number and is used to extract the registered bike owner's information.

The model was implemented on the Windows operating system with Core i3-7100U processor 2.40 GHz and used Python 3.6.1, TensorFlow 1.15.0, and Tesseract 5.0.0.

4.1 Helmet and License Plate Detection

For training, a dataset of 500 still images is taken manually from video surveillance where 250 images are bike rider with helmet and the rest are bike rider without a helmet. Initially, using OpenCV, the video frames are converted into frames using video capture with fps =0.5 s and the collected raw images are preprocessed, such as data formatting, data cleansing, and converted into feature transformation. The resultant optimal data are fed into the model. After data preprocessing, the images are labeled with the help of the labeling tool. These labeled images are used for training the model. The images are divided into training and testing in the ratio of 80:20. Figure 2 shows the sample dataset for helmet detection and recognition of the license plate.

Fig. 2 Sample inputs frames extracted from video surveillance

Four regions of interest are identified, and four labels are created : bike, helmet, head, and license plates. The labeled image data are stored in the XML format to convert them into the CSV file, further storing all the information about the labeled objects' coordinates. After this, TF records are created to serve as input data for the object detector training. Then, a label map is created as a training configuration file before training the image on the model.

Faster R-CNN-Inception V3 model has been used for training in 2072 steps and in the last step exported the inference graph to run the model. After trained the model, tested the output for the test images given initially during the test directory.

Figure 3 shows the experimental results of the Faster R-CNN model in detecting with helmet and without a helmet on the head of a bike rider. In (a), the head is recognized with 93% accuracy from the back angle and the bike's license plate with 99%. In (b), head is recognized with 99% accuracy from the front angle as facial matching will help the model, and the license plate recognizes with 99% accuracy. Both a & b are the frames without a helmet. In (c), the helmet is recognized with 74% accuracy from the back angle and a license plate with 91% accuracy. The downfall in the accuracy is due to the long gap between the object and the camera. In (d), the helmet is recognized with 98% accuracy from the front angle and a license plate with 99% accuracy. Accuracy, in this case, increased compared to case 3 because of the camera's front angle and nearness by which it can quickly identify as a helmet.

Table 1 and Fig. 4 present the results of the experiments for the classification with bike and helmet from the backside and front side, with helmet and without helmet using the proposed Faster R-CNN method with the existing CNN method as its performance is highest among all other methods presented [13]. Faster R-CNN gives an accuracy of 99%, which is superior to the existing CNN method on the created dataset. The results show that Faster R-CNN gives better accuracy when the front side of video frames are used. There is a slight oscillation of performance is observed when the frames are captured from the backside. As the distance from the camera is increasing, the accuracy rate is decreasing. Whereas, the Gabor-Wavelet filters helped the model to gain better accuracy from the front side. In summary, for still images, the success rate of the model is 90% and for real time it is in the range of 70–80%. From the results it is evident that the model is best suitable for less traffic.

The TensorBoard tool is used to evaluate the model. The TensorBoard evaluation tool provides metrics such as precision, recall, loss. The threshold score is computed based on the precision and recall by which the model removes the unnecessary boxes. In this model, the default threshold score is set as 0.5. Figure 5a represents the precision graph to measure how much accuracy the model predicts objects correctly and is usually in the range of 0–1. The proposed model obtains a precision value of 0.55, and it also observes over time, as precision increases on the increase in the number of steps. Figure 5b shows a visualization graph of recall, which measures how much accuracy the model will find all positives, and it is also between 0 and 1. The proposed model obtained a recall of 0.6. Figure 5c shows a visualization of

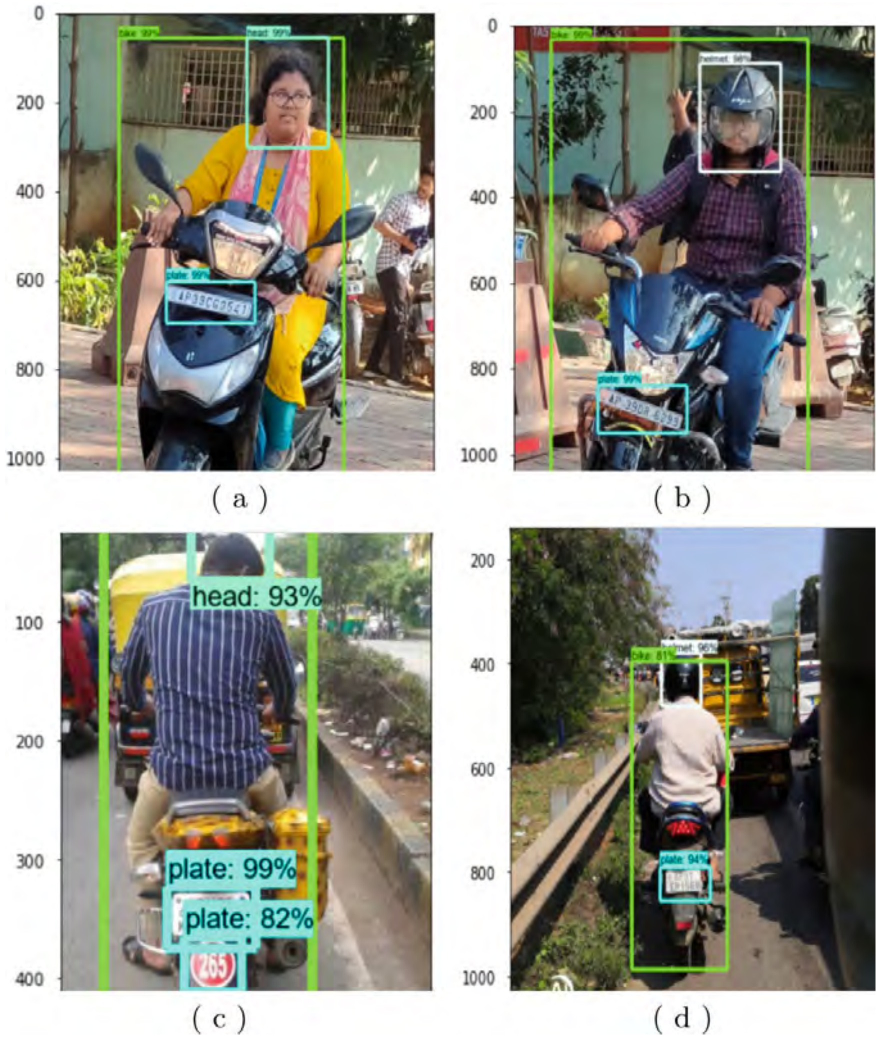


Fig. 3 Detection of bike, head, helmet, and number plate both front and rear

Table 1 Accuracy comparison

Method	Bike from back	Bike from front	Helmet from back	Helment from front
CNN	72	98.88	72	87.11
Proposed method	69	99	74	99

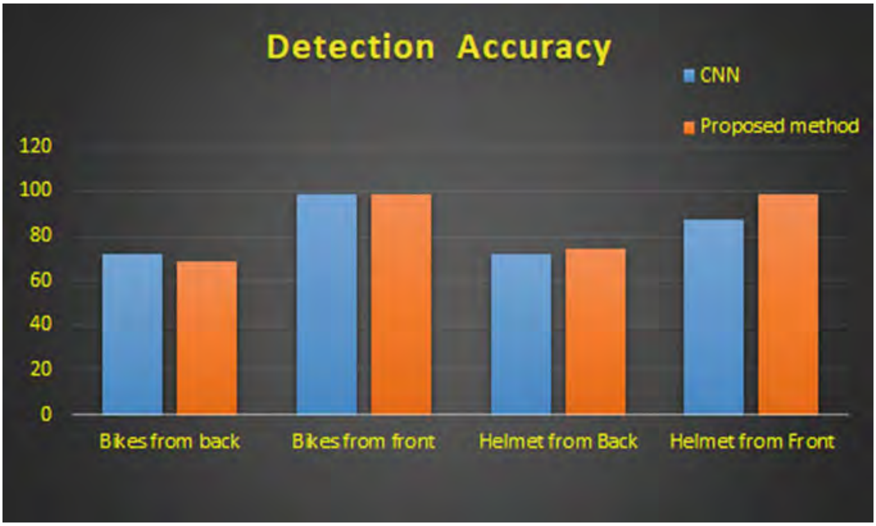


Fig. 4 Detection accuracy graph

loss, which is a measure of the difference between a model's predicted output and the ground truth label. The total loss obtained in the proposed model is 0.35. From Table 2, it is evident that the proposed method using Faster R-CNN performs much superior to the existing CNN method.

4.2 Detection of License Plate

For character recognition, the Tesseract OCR engine is used. The Tesseract takes the input of cropped image of license plate in bike rider without helmet case. Tesseract needs a high-quality image as an input, and it cannot recognize the text from a rotated image. So, it is necessary to rotate the cropped license plate image before passing as input. In this model, the images are rotated manually for better accuracy.

In Fig. 6a, all the alphabets and numbers are recognized correctly. In Fig. 6b, all the string literals except '3' is recognized correctly. Here, 3 is identified as 'S'. In Fig. 6c, all the string literals except 'C' are recognized correctly. Here, 'C' is identified as a special character. The Tesseract engine identifies the characters based on image constraints like lighting, structure, and clarity. Overall the model achieves 88% accuracy on the test dataset. The recognized characters are used to extract the bike owner's information to take further action by the concerned.

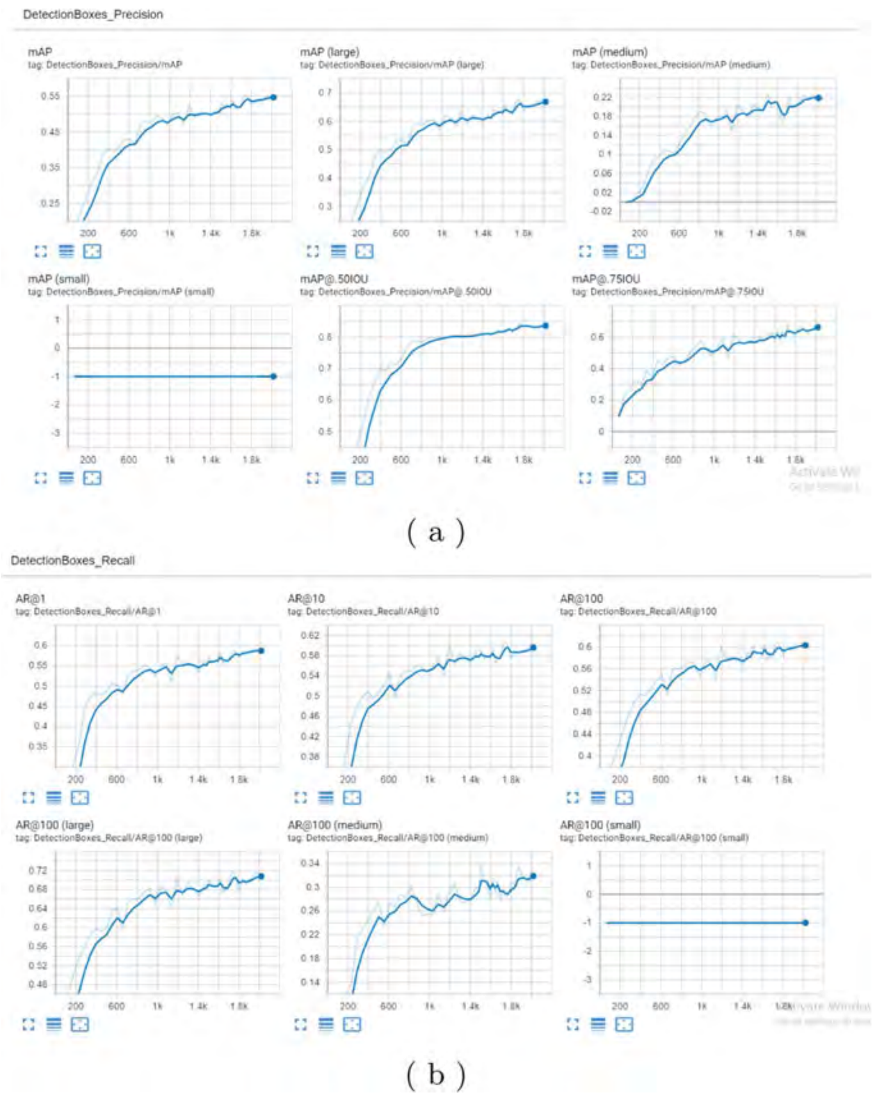
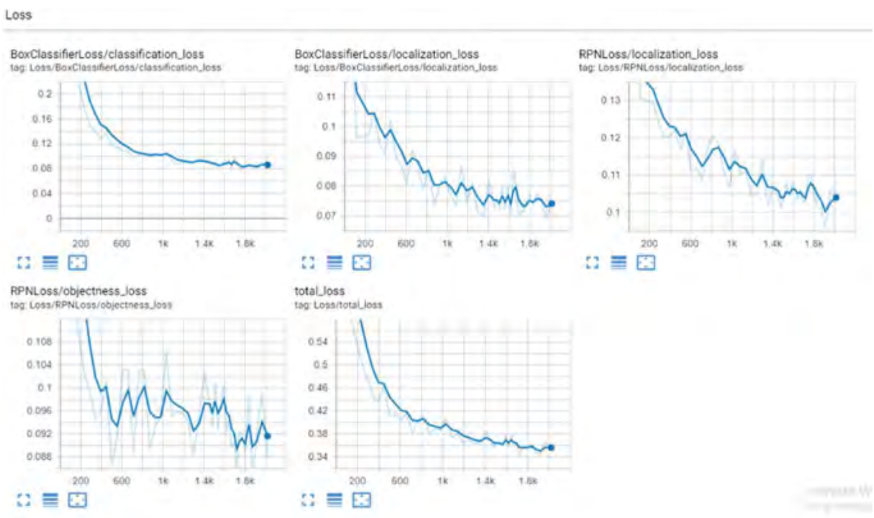


Fig. 5 Evaluation metrics of the proposed method: **a** precision, **b** recall, **c** loss



(c)

Fig. 5 (continued)

Table 2 Comparison of evaluation metrics

Method	Precision	Recall	Loss
CNN	0.51	0.65	0.42
Proposed method	0.55	0.6	0.35

Fig. 6 Results of Tesseract OCR Engine

5 Conclusion

This paper proposes a model to automatically detect bike riders who do not wear a helmet and recognize their license plate to facilitate traffic authorities to take further action. Faster R-CNN is used with TensorFlow object detection API to detect bike rider without a helmet and then detect its license plate. Using the Tesseract optical character recognition tool, the characters are extracted on the license plate and extracted the owner's details from the database. The experimental results reveal the best performance of the model with the existing CNN model. This model can be extended for night time surveillance and trained with more real-time data for detecting in dynamic scenes and is considered future work.

References

1. Y.S. Malik, Road accidents in India 2017, in *Transport Research Wing* (Ministry of Road Transport & Highways)
2. P. Mishra, P. Mishra, Vital stats-road accident in India, in *PRS Legislative Research* (Institute for Policy Research Studies, New Delhi)
3. R. Thakur, M. Manoria, RBFNN approach for recognizing Indian license plate. Int. J. Comput. Sci. Network (IJCSN) **1**(5) (2012). http://www.ijcsn.org. ISSN 2277-5420
4. K. Dahiya, D. Singh, C.K. Mohan, Automatic detection of bikeriders without helmet using surveillance videos in real-time, in *International Joint Conference Neural Networks (IJCNN)*, Vancouver, Canada, July 24–29, 2016, pp. 3046–3051
5. M.H. Dashtban, Z. Dashtban, H. Bevrani, A novel approach for vehicle license plate localization and recognition. Int. J. Comput. Appl. (0975 - 8887) **26**(11) (2011)
6. S. Manoharan, An improved safety algorithm for artificial intelligence enabled processors in self driving cars. J. Artif. Intell. **1**(02), 95–104 (2019)
7. M.H.J.D. Koresh, J. Deva, Computer vision based traffic sign sensing for smart transport. J. Innov. Image Process. (JIIP) **1**(01), 11–19 (2019)
8. W. Hu, T. Tan, L. Wang, S. Maybank, A survey on visual surveillance of object motion and behaviors. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **34**(3), 334–352 (2004)
9. R. Silva, K. Aires, T. Santos, K. Abdala, R. Veras, A. Soares, Automatic detection of motorcyclists without helmet, in *Proceedings of Latin American Computing Conference (CLEI)*, Puerto Azul, Venezuela, October 4-6, 2013, pp. 1–7
10. J. Chiverton, Helmet presence classification with motorcycle detection and tracking. Intell. Transp. Syst. (IET) **6**(3), 259–269 (2012)
11. R. Waranusast, N. Bundon, V. Timtong, C. Tangnoi, Machine vision techniques for motorcycle safety helmet detection, in *28th International Conference on Image and Vision Computing*, New Zealand (IVCNZ, 2013), pp 35–40
12. S.A. Ghonge, J.B. Sanghavi, Smart surveillance system for automatic detection of license plate number of motorcyclists without helmet. Int. J. Comput. Sci. Eng. **2**(1) (2018)
13. C. Vishnu, D. Singh, C.K. Mohan, S. Babu, Detection of motorcyclists without helmet in videos using convolutional neural network, in *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK (2017), pp. 3036–3041. https://doi.org/10.1109/IJCNN.2017.7966233
14. M.J. Prajwal, K.B. Tejas, V. Varshad, M.M. Murgod, R. Shashidhar, Detection of non-helmet riders and extraction of license plate number using Yolo v2 and OCR method. Int. J. Innov. Technol. Exploring Eng. (IJITEE) **9**(2) (2019). ISSN: 2278-3075
15. A. Mukhtar, T.B. Tang, Vision based motorcycle detection using HOG features, in *IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* (IEEE, 2015)
16. L. Allamki, M. Panchakshari, A. Sateesha, K.S. Pratheek, Helmet detection using machine learning and automatic License Plate Recognition, in *International Research Journal of Engineering and Technology (IRJET)*, vol. 06(12), December 2019

Prediction of Heart Disease with Different Attributes Combination by Data Mining Algorithms

Ritu Aggrawal and Saurabh Pal

Abstract Heart disease is considered the most dangerous and fatal infection in the human body. This globally fatal disease cannot be identified easily by a general practitioner, and it requires an analyst or expert to detect it. In the field of medical science, machine learning plays important role in disease prediction to identify the heart infection features. In this perspective, this research work proposes a new technique to predict heart disease by using various classifier algorithms such as random forest, gradient boosting, support vector machine, and K- nearest neighbor algorithms. For this purpose, the classification accuracy and the obtained results of each predictor have been compared. In each analysis, machine learning classifier algorithms: random forest, gradient boosting, support vector machine, and K-nearest neighbor algorithms are used and finally defect, gradient boosting, which has calculated high accuracy with low error values and high correlation value when compared to other used algorithms.

Keywords Feature selection methods: extra tree · Random forest · Gradient boosting · Support vector machine · K-nearest neighbor algorithms · Contingency coefficient · Adjusted contingency coefficient · Correlation coefficient and phi coefficient · Root mean square error

1 Introduction

This research has used heart disease prediction-based analysis by using machine learning techniques. In recent days, machine learning plays an important role in various areas like e-commerce, business, diagnosis of disease, development of expert systems, etc. The main work of machine learning is to discover and identify the hidden pattern in a complex huge dataset. Heart disease is one of the most popular reasons for death all over the world. This disease does not cover particular sex but both males and females globally suffer from these problems. Every year the number of

R. Aggrawal · S. Pal (✉)
Department of Computer Applications, VBS Purvanchal University, Jaunpur, India
e-mail: drsaurabhpal@yahoo.co.in

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_38

heart infected patients increased very rapidly, and due to this reason, various facts of disease being collected. These facts provide help in identify a heart infection. Data mining assists the medical healthcare system. The main objective of machine learning is to support the disease expert in diagnosis but does not replace the expert.

1.1 K-Nearest Neighbors (KNN)

K-NN algorithm covers feature similarity and matching of the data point in the training set.

(a) Initially, load training dataset and implement algorithms.
(b) Next, select the values of K as integer values which are the nearest data point.
(c) Next, evaluate the distance between the row of training data and test data.
(d) Next, arrange the value of distance in ascending order.
(e) Next, select peak k rows of the sorted array to test.
(f) End.

1.2 Random Forest

Random forest is used for both classification and regression. This algorithm creates decision tree on the data sample for better prediction.

(a) First, select the random sample of the dataset.
(b) Next, it generates a decision tree for every sample and predicts the results for every decision tree.
(c) Next, the voting method supports selecting predicted results.
(d) At last, the final prediction result is selected by the voting method.

1.3 Gradient Boosting

Gradient boosting is a technique to solve regression and classification problems and generate prediction model for weak tree models.

(a) Initially, select training dataset and fit decision tree on it.
(b) Next, provide the result of actual and predicted result.
(c) Next, fit the target variable as error residual by a new model.
(d) Next, previous predictions combine predicted residuals.
(e) Next, adjust the model on residuals that is still left.
(f) Next, control overfitting and observed accuracy.

1.4 Support Vector Machine

Support vector machine fined an optimal boundary between possible outputs.

(a) Initially, select the training dataset.
(b) Next, train all training samples by SVM.
(c) Next, search the best decision for categories each feature.
(d) Next, ordering the features of the top percentage.
(e) Next, the remaining sample to retrain by SVM.
(f) Next, select a pair of features to the last sample in training.
(g) End.

2 Related Work

Amin et. al [1] considered a cardiovascular disease and predicted accuracy by machine learning algorithms. The author used naïve Bayes and logistic regression for better prediction. Finally, they used hybrid techniques to calculate high accuracy. The hybrid model is organized by naïve Bayes and logistic regression algorithms.

Verma et al. [2] discussed skin disease by various classifier algorithms in data mining as PAC, LDA, RNC, MNB, NB, and ETC. They generate an ensemble models by bagging and boosting algorithms. The authors measured high classification accuracy by gradient boosting algorithms.

Gokulnath and Shantharajah [3] selected heart disease attributes by feature selection method in machine learning. They generate a model by support vector machine and calculate high classification accuracy.

Wu et al. [4] considered coronary artery disease by various classification algorithms as deep learning and filtered method. The authors used a monocardiograph of deep learning and calculated high accuracy, sensitivity, and specificity.

Alaa et al. [5] analyzed the risk in cardiovascular disease by various classifier algorithms. They imputed data by pipeline method and used 473 variables of cardiovascular disease. The authors calculated (0.77) values of receiver operating characteristics.

Haq et al. [6] generated a model using lasso features selection algorithms in disease attributes, selected more important features, and then predicted 85% receiver operating characteristic by K-NN, ANN, DT, and NB.

Vijayashree and Sultana [7] discussed heart disease problems by various machine learning algorithms as particle swarm optimization metaheuristic algorithms and support vector machine. The authors calculated high accuracy by (PSO + SVM) on selected features.

Vivekanandan and Iyengar [8] identified heart disease problems by various machine learning algorithms. They used a differential evolution algorithm with optimal features selection and calculated high accuracy by fuzzy AHP and feed-forward neural network.

Khateeb and Usman [9] considered heart disease problems by the K-NN technique. They collect similar factors in heart disease by machine learning algorithms and select risk factor. The authors calculated high classification accuracy by K-nearest neighbor algorithm.

Ramotra et al. [10] considered heart disease problems using a decision tree, naïve Bayes, and support vector machine classifiers. The authors used Weka and SPSS tool for heart disease prediction and calculated high accuracy by support vector machine by SPSS tool. They selected 24 important features from 46 related features. They calculated high accuracy for five different classes in machine learning algorithms.

Narayan and Sathiyamoorthy [11] considered heart disease features and identified heart infection by machine learning algorithms. They develop a hybrid model as x^2 – DNN that calculated high classification accuracy for DNN and ANN.

Gonsalves et al. [12] predicted the information the facts of a coronary heart infection and identify correlated attributes by machine learning algorithms as naïve Bayes, support vector machine, and decision tree. Finally, the authors calculated high accuracy by naïve Bayes.

Manogaran et al. [13] developed a hybrid model by multiple kernel learning with adaptive neuro-fuzzy inference for heart infection. He modify in cuckoo search algorithm and calculated high sensitivity with specificity.

Jayaraman and Sultana [14] identified heart infection seriousness by machine learning. The authors consider correlate features by cuckoo search algorithms and used neural networks for better prediction. They calculated high accuracy with low error rate with time by a neural network.

3 Methodology

This section discussed the detailed study of the heart disease dataset and applied various machine learning techniques.

3.1 Data Description

The dataset shown in Table 1 has been taken from the UCI repository. The various heart disease attributes and their domain values are given in table [15, 16].

3.2 Histogram

A histogram is a graphical display of data points by various bars of different heights. It appears as a chart, but histogram groups have a collection of various ranges. The height of each bar describes various fall with different range [17, 18] (Fig. 1).

Table 1 Representation of heart disease features description

S. No.	Code	Feature domain values
1	Age	Age in years
2	Sex	Sex (1 = male; 0 = female)
3	Cp	Chest pain type: 1 = typical angina; 2 = atypical angina; 3 = non-angina pain; 4 = asymptomatic
4	Trestbps	Resting blood pressure (mg)
5	Chol	Serum cholesterol (mg/dl)
6	Fbs	Fasting blood sugar > 120 mg/dl: 1 = yes; 0 = no
7	Restecg	Resting electrocardiographic results: 0 = normal; 1 = ST-T wave abnormal; 2 = left ventricular hypertrophy
8	Thalach	Maximum heart rate achieved [71, 202]
9	Exang	Exercise induced angina: 1 = yes; 0 = no
10	Oldpeak	ST depression induced by exercise relative to rest: 0 and 6.2
11	Slope	The slope of the peak exercise ST segment: 1 = upsloping; 2 = flat; 3 = downsloping
12	Ca	Number of major vessels colored by fluoroscopy (values 0–3)
13	Thal	Exercise thallium scintigraphy: 3 = normal; 6 = fixed defect; 7 = reversible defect
14	num	0 = no presence; 1 = presence

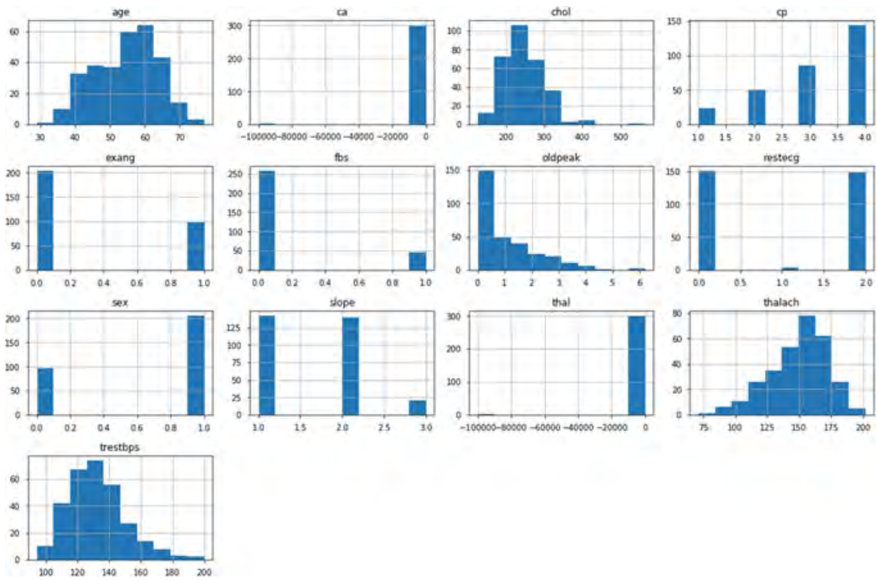


Fig. 1 Representation of heart disease attributes by various range of bar

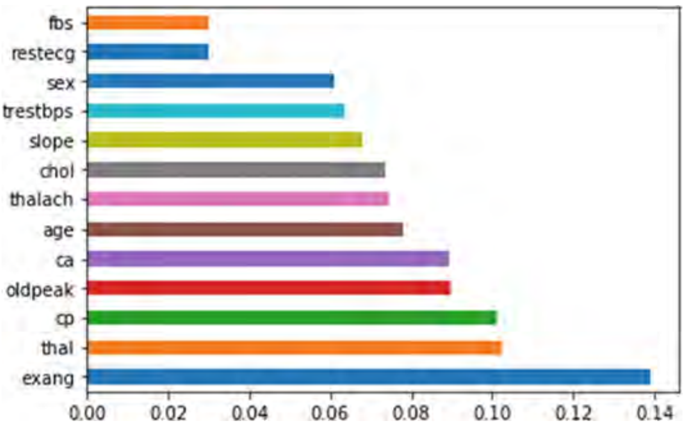


Fig. 2 Representation of important features of heart disease by extra tree

3.3 Extra Tree Feature Selection

Extremely randomized trees are an ensemble technique. This technique generates an original training sample and aggregates the results of multiple de-correlated decision trees collected in a forest. It is differing from random forest because of the construction of the decision trees in the forest [19–21] (Fig. 2).

[0.07784633 0.06093372 0.10118647 0.06361897 0.07383465 0.02986432
0.03003928 0.07437752 0.1390865 0.08973886 0.06789341 0.08924798
0.102332]

3.4 Matrix Evaluation

Different performance measures are evaluated using different matrices. These matrices are discussed in this section.

$$\text{Classification Accuracy (CA)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

where

TP, TN, FP, and FN are true positive, true negative, false positive, and false negative values in the dataset, respectively.

$$\text{Root Mean Square Value(RMSE)} = \sqrt{\frac{\sum_{l}^{m}\left(\text{Predicted}_j - \text{Actual}_j\right)^2}{m}} \tag{2}$$

where

$M =$ No. of rows in table.

The contingency coefficient decided the dependency and independency between dataset variables. This coefficient is defined as the Pearson coefficient and based on chi-square statistic as:

$$\text{Contingency Coefficient } (C) = \sqrt{\frac{l^2}{N + l^2}} \tag{3}$$

where

l^2 is chi-square.

N total cases.

The C is adjusted so it reaches a maximum of 1. when there is complete association in a table of any number of rows and columns by dividing C by $C_{\max}$. Therefore, $C_{\max}$ is calculated as:

$$\mathrm{C}_{\max} = \sqrt{\frac{(m-1)}{m}} \tag{4}$$

$$\text{Adjusted Contingency Coefficient } (C^*) = \frac{C}{C\max}\sqrt{\frac{m \cdot l^2}{(m-1)\left(n + l^2\right)}} \tag{5}$$

where

$m =$ Number of rows or number of columns [22, 23].

3.5 Proposed Model

This research work has proposed new techniques, which are shown in Fig. 3 to search the best algorithms in different machine learning techniques. Initially, important features are selected by the extra tree and then applied various machine learning techniques: random forest, gradient boosting, support vector machine and K- nearest neighbor algorithms with different features group as like 3, 7, 11, and 14 and check their correlation strength as contingency coefficient, adjusted contingency coefficient, correlation coefficient, and phi coefficient with classification accuracy and root mean square error values.

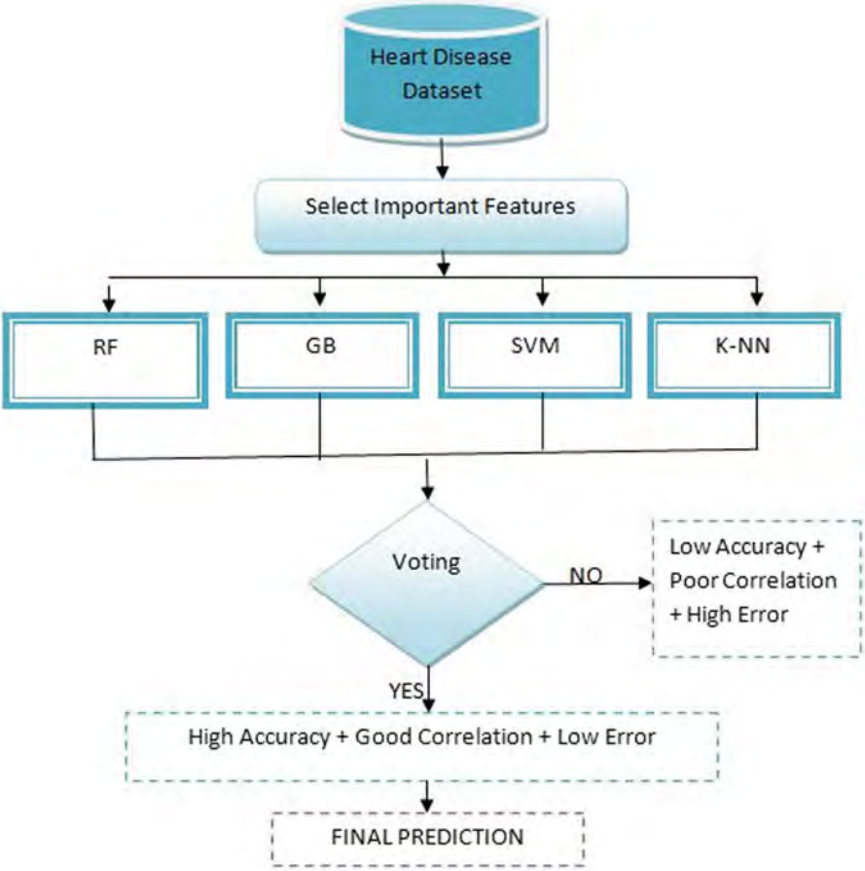


Fig. 3 Representation of proposed model for heart disease prediction

4 Results

The machine learning efficiency is monitored by matrix values like TP, FP, TN, and FN values. In disease dataset, features valuable problems move between binary and decimal values with target variable 0 (zero) present and 1 (one) absent of heart infection in human.

Pearson correlation calculates and measures the strength with the direction of a linear relationship between two variables in the dataset. The correlation matrix decide range between −1 (strong negative relationship), +1 (strong positive relationship), and 0 (weak or no linear relationship) [24] (Fig. 4).

The confusion matrix is shown in Table 2. The results are obtained by classification on binary and decimal values with tenfold cross-validation. The experimental matrix

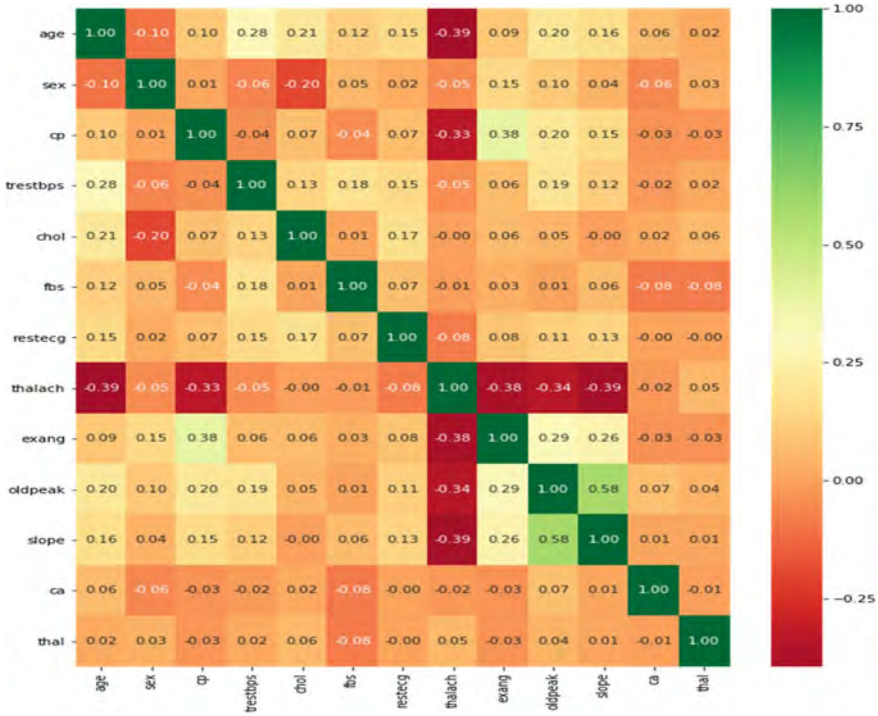


Fig. 4 Representation of features correlation matrix for heart disease

Table 2 Representation of confusion matrix for heart disease attributes

Algorithm	RF	SVM	K-NN	GB
Attributes (3)	a, b classified as 134 30 | $a = 0$ 36 103 | $b = 1$	a, b classified as 137 27 | $a = 0$ 40 99 | $b = 1$	a, b classified as 136 28 | $a = 0$ 44 95 | $b = 1$	aa, b classified as 144 20 | $a = 0$ 31 108 | $b = 1$
Attributes (7)	a, b classified as 129 35 | $a = 0$ 30 109 | $b = 1$	a, b classified as 127 37 | $a = 0$ 37 102 | $b = 1$	a, b classified as 133 31 | $a = 0$ 50 89 | $b = 1$	a, b classified as 138 26 | $a = 0$ 28 111 | $b = 1$
Attributes (11)	a, b classified as 127 37 | $a = 0$ 37 102 | $b = 1$	a, b classified as 133 31 | $a = 0$ 50 89 | $b = 1$	a, b classified as 132 32 | $a = 0$ 51 88 | $b = 1$	a, b classified as 138 26 | $a = 0$ 37 102 | $b = 1$
Attributes (14)	a, b classified as 128 36 | $a = 0$ 29 110 | $b = 1$	a, b classified as 135 29 | $a = 0$ 40 99 | $b = 1$	a, b classified as 123 41 | $a = 0$ 46 93 | $b = 1$	a, b classified as 142 22 | $a = 0$ 42 97 | $b = 1$

Table 3 Representation of computational table for heart disease three attributes

Statistical analysis	A3 (RF)	A3(SVM)	A3(K-NN)	A3(GB)
Classification accuracy	78.21%	77.88%	76.23%	83.16%
RMSE	0.39	0.47	0.41	0.36
Contingency coefficient	0.54	0.489	0.484	0.551
Adjusted contingency coefficient	0.763	0.691	0.685	0.78
Correlation coefficient	0.846	0.773	0.77	0.866
Phi coefficient	0.641	0.56	0.554	0.661

Table 4 Representation of computational table for heart disease seven attributes

Statistical analysis	A7(RF)	A7(SVM)	A7(K-NN)	A7(GB)
Classification accuracy	78.54%	75.57%	73.26%	82.17%
RMSE	0.41	0.45	0.41	0.36
Contingency coefficient	0.495	0.453	0.418	0.551
Adjusted contingency coefficient	0.701	0.641	0.591	0.78
Correlation coefficient	0.782	0.717	0.671	0.866
Phi coefficient	0.57	0.508	0.46	0.661

represents the highest and lowest values of true positive, true negative, false positive, and false negative [25–27].

In Table 3, only three features (age, sex, and num) are used and check their correlation strength as contingency coefficient, adjusted contingency coefficient, correlation coefficient, and phi coefficient with classification accuracy and root mean square error values by machine learning classifier algorithms: Random forest, gradient boosting, support vector machine, and K-nearest neighbor algorithms. With the results, three attributes of gradient boosting have calculated a high accuracy of about 83.16% with 0.36 error value and high correlation value of 0.86 compare to other used algorithms.

In Table 4, only six features (age, sex, cp, trestbps, chol, fbs, and num) are used and checked their correlation strength with classification accuracy and root mean square error values by machine learning classifier algorithms. With the results, seven attributes gradient boosting have calculated a high accuracy of about 82.17% with 0.36 error values and a high correlation value of 0.86 compared to other used algorithms.

In Table 5, 11 features (age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, and num) are used. With the results observed for 11 attributes, gradient boosting has calculated a high accuracy of about 79.2% with a 0.39 error value and a high correlation value of 0.79 compared to other used algorithms.

In Table 6, 13 features (age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, and thal) have used. It has been observed for 13 attributes, and gradient boosting has calculated a high accuracy of about 78.87% with a 0.38 error value and a high correlation value of 0.79 compare to other used algorithms.

Table 5 Representation of computational table for heart disease 11 attributes

Statistical analysis	A11(RF)	A11(SVM)	A11(K-NN)	A11(GB)
Classification accuracy	75.57%	73.59%	72.6%	79.2%
RMSE	0.4	0.43	0.47	0.39
Contingency coefficient	0.453	0.429	0.407	0.502
Adjusted contingency coefficient	0.641	0.606	0.576	0.71
Correlation coefficient	0.717	0.682	0.655	0.795
Phi coefficient	0.508	0.474	0.446	0.58

Table 6 Representation of computational table for heart disease 14 attributes

Statistical analysis	A14(RF)	A14(SVM)	A14(K-NN)	A14(GB)
Classification accuracy	78.54%	77.22%	71.28%	78.87%
RMSE	0.39	0.42	0.43	0.38
Contingency coefficient	0.495	0.475	0.388	0.499
Adjusted contingency coefficient	0.7	0.672	0.548	0.705
Correlation coefficient	0.781	0.755	0.616	0.798
Phi coefficient	0.57	0.54	0.42	0.575

5 Discussion

From Tables 3, 4, 5, and 6, it is observed that the classification accuracy is continuously decreased, root mean square error values increases, and correlation values decreases. It is clear that the gradient boosting algorithm better performs in each experiment compare to other used algorithms: random forest, support vector machine, and K-nearest neighbor algorithms.

The main objective of the model is to predict heart disease on different feature combinations and compare how accurate is the model output with different feature correlations. Figures 5 and 6 represent correlation strength as contingency coefficient, adjusted contingency coefficient, correlation coefficient, and phi coefficient

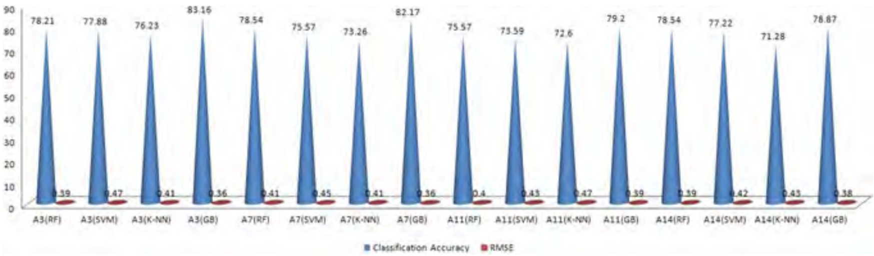


Fig. 5 Representation of accuracy and RMSE by different algorithms in heart disease 14 attributes

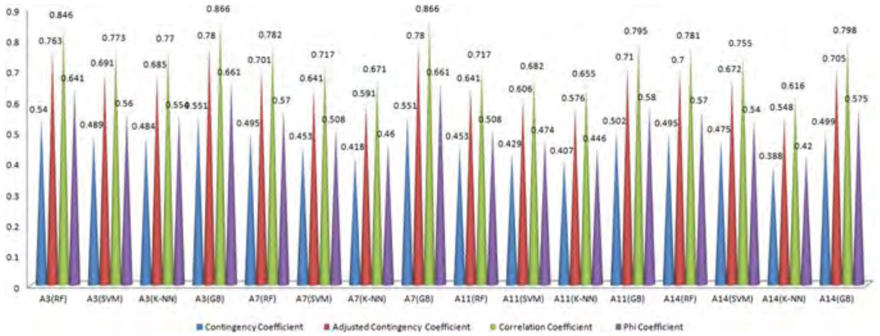


Fig. 6 Representation correlation strength as contingency coefficient, adjusted contingency coefficient, correlation coefficient, and phi coefficient by different classifier algorithms

by machine learning different classifier algorithms. The correlation values of each algorithm are detected to continue to be different in each experiment.

6 Conclusion

The dataset has been taken from the UCI repository for various heart disease problems. Total 14 attributes and 303 instances are taken in this experiment and test machine learning algorithms efficiency on different feature combinations such as classification accuracy, root mean square values and check their correlation strength as contingency coefficient, adjusted contingency coefficient, correlation coefficient, and phi coefficient. In each experiment, machine learning classifier algorithms: random forest, gradient boosting, support vector machine, and K-nearest neighbor algorithms are compared. As a result, it is observed in each experiment that the gradient boosting algorithm evaluated the highest accuracy with low error values and higher correlation value as compared to other used algorithms. In the future, the neuro-fuzzy, fuzzy genetic, and neuro-genetic combined model can be used and tested on various different dataset combinations for achieving a better prediction.

References

1. M.S. Amin, Y.K. Chiam, K.D. Varathan, Identification of significant features and data mining techniques in predicting heart disease. Telematics Inform. **36**, 82–93 (2019)
2. A.K. Verma, S. Pal, S. Kumar, Prediction of skin disease using ensemble data mining techniques and feature selection method—a comparative study. Appl. Biochem. Biotechnol. **190**(2), 341–359 (2020)
3. C.B. Gokulnath, S.P. Shantharajah, An optimized feature selection based on genetic approach and support vector machine for heart disease. Cluster Comput. **22**(6), 14777–14787 (2019)

4. J.M.T. Wu, M.H. Tsai, Y.Z. Huang, S.H. Islam, M.M. Hassan, A. Alelaiwi, G. Fortino, Applying an ensemble convolutional neural network with Savitzky-Golay filter to construct a phonocardiogram prediction model. Appl. Soft Comput. **78**, 29–40 (2019)
5. A.M. Alaa, T. Bolton, E. Di Angelantonio, J.H. Rudd, M. van Der Schaar, Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423, 604 UK Biobank participants. PloS one **14**(5), e0213653 (2019)
6. A.U. Haq, J.P. Li, M.H. Memon, S. Nazir, R. Sun, A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile Info. Syst. **2018**, 1–21 (2018)
7. J. Vijayashree, H.P. Sultana, A machine learning framework for feature selection in heart disease classification using improved particle swarm optimization with support vector machine classifier. Programm. Comput. Softw. **44**(6), 388–397 (2018)
8. T. Vivekanandan, N.C.S.N. Iyengar, Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease. Comput. Biol. Med. **90**, 125–136 (2017)
9. N. Khateeb, M. Usman, Efficient heart disease prediction system using K-nearest neighbor classification technique. in *Proceedings of the International Conference on Big Data and Internet of Thing*, (2017), pp. 21–26
10. A.K. Ramotra, A. Mahajan, R. Kumar, V. Mansotra, Comparative analysis of data mining classification techniques for prediction of heart disease using the weka and SPSS modeler tools. in *Smart Trends in Computing and Communications* (Springer, Singapore, 2020), pp. 89–96
11. S. Narayan, E. Sathiyamoorthy, A novel recommender system based on FFT with machine learning for predicting and identifying heart diseases. Neural Comput. Appl. **31**(1), 93–102 (2019)
12. A.H. Gonsalves, F. Thabtah, R.M.A. Mohammad, G. Singh, Prediction of coronary heart disease using machine learning: an experimental analysis. in *Proceedings of the 2019 3rd International Conference on Deep Learning Technologies*, 51–56 (2019)
13. G. Manogaran, R. Varatharajan, M.K. Priyan, Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system. Multimedia Tools Appl. **77**(4), 4379–4399 (2018)
14. V. Jayaraman, H.P. Sultana, Artificial gravitational cuckoo search algorithm along with particle bee optimized associative memory neural network for feature selection in heart disease classification. J. Ambient Intell. Humanized Comput. 1–10 (2019)
15. M. Tanveer, A. Sharma, P.N. Suganthan, Least squares KNN-based weighted multiclass twin SVM. Neurocomputing (2020). https://doi.org/10.1016/j.neucom.2020.02.132
16. D.C. Yadav, S. Pal, Prediction of heart disease using feature selection and random forest ensemble method. Int J Pharmaceutical Res. **12**(4), 56–66 (2020)
17. H. Lu, S.P. Karimireddy, N. Ponomareva, V. Mirrokni, Accelerating gradient boosting machines. in *International Conference on Artificial Intelligence and Statistics* (2020) pp. 516–526
18. B. Richhariya, M. Tanveer, A reduced universum twin support vector machine for class imbalance learning. Pattern Recogn. **102**, 107150 (2020)
19. Yuan, B. H., Liu, G. H.,: Image retrieval based on gradient-structures histogram. Neural Computing and Applications, 1–11 (2020)
20. M. Alizamir, S. Kim, O. Kisi, M. Zounemat-Kermani, Deep echo state network: a novel machine learning approach to model dew point temperature using meteorological variables. Hydrol. Sci. J. **65**(7), 1173–1190 (2020)
21. D.C. Yadav, S. Pal, Prediction of thyroid disease using decision tree ensemble method. Human-Intell. Syst. Integra. 1–7 (2020)
22. M. Baak, R. Koopman, H. Snoek, S. Klous, A new correlation coefficient between categorical, ordinal and interval variables with Pearson characteristics. Comput. Stat. Data Anal. **152**, 107043 (2020)
23. D.C. Yadav, S. Pal, To generate an ensemble model for women thyroid prediction using data mining techniques. Asian Pac. J. Cancer Prev. **20**(4), 1275 (2019)

24. M.A. Hasan, M.U. Khan, D. Mishra, A computationally efficient method for hybrid EEG-fNIRS BCI based on the pearson correlation. Biomed. Res. Int. **2020**, 1–13 (2020)
25. R. Aggrawal, S. Pal, Sequential feature selection and machine learning algorithm-based patient's death events prediction and diagnosis in heart disease. SN Comput. Sci. **1**, 344 (2020)
26. A.K. Verma, S. Pal, S. Kumar, Prediction of different classes of skin disease using machine learning techniques. in *Smart Innovations in Communication and Computational Sciences.* Advances in Intelligent Systems and Computing, vol 1168. (Springer, Singapore, 2021)
27. V. Chaurasia, S. Pal, Machine learning algorithms using binary classification and multi model ensemble techniques for skin diseases prediction. Int. J. Biomed. Eng. Technol. **34**(1), 57–74 (2020)

A Novel Video Retrieval Method Based on Object Detection Using Deep Learning

Anuja Pinge and Manisha Naik Gaonkar

Abstract The recent research in computer vision is focused on videos. Image and video data has been increased drastically in the last decade. This has motivated researchers to come up with different methods for image and video understanding and other applications like action recognition from videos, video retrieval, video understanding, and video summarization. This article proposes a novel and efficient method for video retrieval by using the deep learning approach. Instead of considering low-level features, videos could be best represented in terms of the high-level features for achieving efficient video retrieval. The idea of the proposed research work is novel, where objects present in the query video are used as features and are used to match against all other videos in the database. Here, object detection is based on YOLOv3, which is the current state-of-the-art method for object detection from videos. This method is tested against YouTube action dataset. It was found that the proposed method has obtained comparable results as the other state-of-the art video retrieval methods.

Keywords Video retrieval · Deep learning · Yolov3 · Object detection

1 Introduction

Advances in technology in the field of data capturing, storage, and communication techniques have resulted in the evolution of a huge amount of image and video data. The availability and ease of to access the camera and social networks have resulted in the drastic increase in visual media content shared among people. With this ever-increasing growth of digital videos, there is a need to develop efficient algorithms to get useful information from this video data. As a lot of work has been carried out upon video data, this decade has witnessed many significant research outcomes on digital data, e.g., videos and images. Taking into consideration video data, the focus is mainly on extracting semantic content from the videos. The various purposes of

A. Pinge (✉) · M. N. Gaonkar
Computer Science and Engineering Department, Goa College of Engineering, Farmagudi, Ponda, Bhausaheb Bandodkar Education Complex, Ponda, Goa 403401, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_39

video understanding are video summarization, video captioning, video classification, video retrieval, action recognition from videos, etc.

Computer vision is a domain that aims to develop algorithms that can make computers understand the contents present in digital data such as images and videos. In the last decade, computer vision gave state-of-the-art results with machine learning techniques. But over the last few years, deep learning models have outperformed the earlier machine learning methods in several fields; among them, computer vision is the significant one. Deep learning has multiple processing layers from which it learns the features at different levels of abstraction level. The factor that contributed to the huge boost in the use of deep nets was mainly because of the publicly available large datasets labeled datasets and also with the power of parallel computing with GPU. The use of GPUs has reduced the training time drastically, thus resulting in the acceleration of deep learning models.

Convolutional neural networks (CNNs) have been found to be effective for understanding content in the images. CNN's yield state-of-the-art results on image segmentation, recognition, detection, retrieval, etc. The main reason for the suitability of these networks is because of the capability to scale up to tens of millions of parameters, and also it handles a very large dataset for the learning process. With these features, CNNs are capable to learn prominent features from images. The results obtained by the use of CNN's have encouraged to use of video data. To obtain these results, CNNs require an extensively long time for training by optimizing millions of parameters to obtain the final optimized trained model. These networks can learn invariant representation from the videos by back-propagating the information, via stacked convolution and pooling layers.

Increase video retrieval is a process of effective search through the dataset and retrieves the most relevant videos to query presented by the user from the dataset [1]. Videos contain complex and variety of patterns including low-level features in-between the frames and also high-level features across several frames. This makes the video retrieval more challenging. The main focus of this project is on developing an algorithm for fast and efficient video retrieval. After the great success of object detection in images, the idea is now extended to videos. And this same idea is used in this method proposed. Videos can be described by the objects present in it, and similar videos have the same kinds of objects. This forms the base for the proposal of a fast and efficient method for video retrieval.

The layout of this article is organized as follows. Section 2 describes the related work done in video retrieval. Section 3 represents the proposed method of research work. Section 4 demonstrates the implementation details, and Sect. 5 depicts the performance analysis of the proposed method. Section 6 concludes the research work. Section 7 explains the discussions, and finally, Sect. 8 outlines the future scope of the research work.

2 Related Work

The main aim of video retrieval is to retrieve relevant videos from the database similar to the query video provided by the user. The entire video retrieval process is complex as it is difficult to understand what exactly the user is looking out for. This requires the analysis of the given query video and makes the task of video retrieval challenging. The analysis is mainly based on semantic concepts.

The user always has high-level features in mind while searching for videos. But there were no good concept detectors for the high-level features. Design of the concept detectors is a lengthy process. It involves data pre-processing, extraction of low-level features, and machine learning techniques for classification [2]. Surveillance cameras capture a huge amount of videos around the daily. Hence, searching a video sequentially amongst such huge data is time-consuming and costly. As a result, video content analysis has emerged as a field in the computer vision domain.

In 2002, Dimitrova et al. [3] described the entire process of content-based video retrieval. It has four processes. These processes are feature extraction, structure analysis, abstraction, and indexing the video. Feature extraction is a crucial step in video retrieval. The entire indexing scheme is dependent upon attributes chosen for feature representation. It is not easy to map very easily extractable features such as color, texture, edges, shape, etc. into semantic concepts objects, people, and scenes. Considering audio domain features, e.g., energy, pitch, or bandwidth can be used for classification. Although in the video, the visual content is focused. Hence, the major features that could be present are text or captions present in the video. The second step in video retrieval is structural analysis. In this step, attempt is made to extract structure information from within the frames of the video. This represents temporal information. The third step is video abstraction. Video abstraction involves a process of creating a representation of video from the structure analyzed in the previous step. This abstraction is similar to extracting keywords from the text. And the final step is video indexing. This means tagging the videos from the database.

Traditional methods focus on extracting low-level features from videos, and these features are then used to find the match with others. The main aim of researchers is to develop algorithms to automatically parse the text, video, and audio. Earlier search solutions were based on text matching which failed for the huge number of videos that have little relevance with metadata or with no metadata at all as the videos captured on mobile phones, wearable devices, or surveillance cameras do not have any metadata attached. To overcome this problem, content-based video semantic retrieval methods were developed. These methods depend on semantics rather than on textual metadata or low-level features, e.g., color, edge, etc. These semantic features include people, objects, actions, activities, scenes, text, or voice involved in the videos.

In 2009, Karpenko [4] extended the tiny image data mining techniques to tiny videos. The dataset of 50,000 videos was collected from YouTube. Content-based copy detection experimented on this dataset. Based on the similarity matrix, the result for text-only video retrieval was improved. The tiny image descriptor was color and affine invariant. This same descriptor was used for the tiny videos.

This decade witnessed e-lecturing. This increased the lecture data available on the internet. Hence, there was a need for an efficient lecture video retrieval algorithm. In 2014, Yang and Meinel [5] proposed a method to retrieve the lecture videos widely present on the World Wide Web. It proposed automated indexing of video and also video search through the large video lecture database. Firstly, automatic video segmentation was done along with key-frame detection for visual content identification. Then, it used optical character recognition (OCR) technique on the key-frames to extract metadata and automatic speech recognition (ASR) on the audio present in the video to convert speech to text. It focused on two main parts visual content (text) and audio tracks (speech). For visual content, a slight difference between the frames was detected, and a unique slide frame is extracted considering it as a video segment. Then, video OCR analysis is done for extracting the textual data present on those frames. In a speech to text task, CMU Sphnix is used as ASR. The proposed might not work well when videos of different genres are embedded in the slides. To improve this, support vector machine (SVM) classifier was used. Also in order to make comparison, histogram of gradients (HOG) was applied.

Another approach for video retrieval was proposed by Gitte et al. [6] in 2014. Here, the system for video mining from the multimedia warehouse has two steps: first is building the multimedia warehouse, and the second is retrieving the video from that warehouse. To retrieve the video from the warehouse, the presented query is first processed, and then, similarity matching is performed. Similarity matrix used here is Euclidean distance. Then, the final similar videos are listed. The two steps are called off-line processing and online processing. The videos are uploaded, and feature extraction is performed on each video. Key frame is chosen from the available frames, and the video is indexed. These indexes are stored along with key-frames and videos. In online processing, user presents a query video. This video is analyzed, and features are extracted. These features are then matched against the features of the videos stored in the warehouse. Euclidean distance is used as a similarity measure for matching query video and candidate video from the warehouse.

In 2018, Iqbal et al. [7] proposed content-based video retrieval method for unconstrained video. The content-based retrieval segment the video and detect the objects of interest. To detect objects, the foreground is separated from the background, and then, localization is done on the frames extracted to obtain features. For video retrieval, the user submits a frame or short video. The image or the extracted frames are then converted to grayscale. Filters are then used to remove noise. This image is then segmented into four co-ordinates. Each of the segmented frames is then converted into eight orientations. For each orientation, various feature extraction methods and classification/clustering algorithms are used to retrieve videos.

Object detection is one of the most challenging problems in the computer vision domain. It focuses on object localization and objects classification in the given input. The deep neural network has shown excellent results in object detection compared to other machine learning approaches.

Year	Paper title	Summary
2002	Applications of video content analysis and retrieval [8]	The proposed the entire process of video retrieval into four major steps. Feature extraction, structure analysis, abstraction, and indexing the video
2009	50,000 tiny videos: A large dataset for non-parametric tent-based Retrieval and Recognition [9]	It used a large dataset of 50,000 tiny videos collected from YouTube. Here, these tine videos are for classification instead of tiny images
2014	Content-based lecture video retrieval using speech and video text information	This paper proposed automated video indexing and video search for a large video database. Video segmentation is performed for key-frame detection. OCR is used to extract data textual data from these key-frames. Automatic speech recognition is used to extract information from audio tracks
2014	Content-based video retrieval system	This approach has two major steps: Building a multimedia warehouse and retrieving the videos from the warehouse. Video retrieving has sub-steps: video segmentation, key-frame detection, feature extraction
2018	Content-based video retrieval using convolutional neural network	Object detection is used for video analysis. The foreground is separated from the background, and localization is performed on each key-frame extracted. Eight-oriented frames are used for feature extraction

Objects are considered as features from videos in the process of video retrieval. The best-known algorithm for object detection is You Look Only Once (YOLO) [10]. Earlier methods of object detection are used classifiers for the task of classification. Instead, this technique treats object detection as a regression problem to bounding boxes and their associated class probabilities. An entire neural network can predict bounding boxes and class probabilities in a single evaluation. It works exactly as the name suggests You Look Only Once [11] at the image. This approach of predicting bounding boxes and class probabilities at the same time is called the unified approach. This algorithm treats the image globally and hence capable of extracting conceptual information. It is also capable capturing a general representation of the objects. Hence, it gives good results when applied to new domains. Bounding boxes are predicted based upon the entire image. It divides the image into an $S \times S$ grid. If a grid cell contains the center of an object, then that grid cell is responsible for detecting that particular object. Each of the grid cells predicts certain bounding boxes and confidence scores for the boxes. YOLOv3 [12] can perform multi-scale detections. The entire YOLO method is divided into two parts: feature extraction and detector. For a new image, first, the features are extracted at multiple (three) scales. These features are then passed onto three different detectors to get bounding boxes and probabilities. YOLOv3 uses Darknet-53 for feature extraction.

3 Method Proposed

Step 1: **Train the YOLOv3 model using Transfer Learning**

Transfer learning is a machine learning technique in which a model developed earlier for some task is reused for some other specific task as a starting point [13]. Transfer learning can be applied in two major ways. One is developing a model, and the other is using a pre-trained model approach. In developing a model approach, first, the source task is selected. Model is developed for that particular task. This is then reused for other tasks by fine-tuning. Another approach is directly using pre-trained weights instead of developing a model. Here, best fit pre-trained model is selected, and this is used as a starting point. The model is fine-tuned according to the task. Here, the second approach is used (Fig. 1).

A. **Dataset Preparation**
YouTube action dataset is used. It has 11 action categories. To train the model to identify the objects present in these videos, a dataset was created by taking the images from the videos of this dataset. For example, diving video has a diving board, water, and a person present in it or a golf game has a golf stick, person, or maybe a golf ball. The images of these objects to were taken. Approximately, 150 images of each class comprise the entire new dataset making total images for training as 1500 images for training and 150 images for validation. Finding the images from the videos was challenging because these videos had view-point changes, and the background was cluttered in many videos, illumination variations even with respect to object scale.

B. **Training the Model**

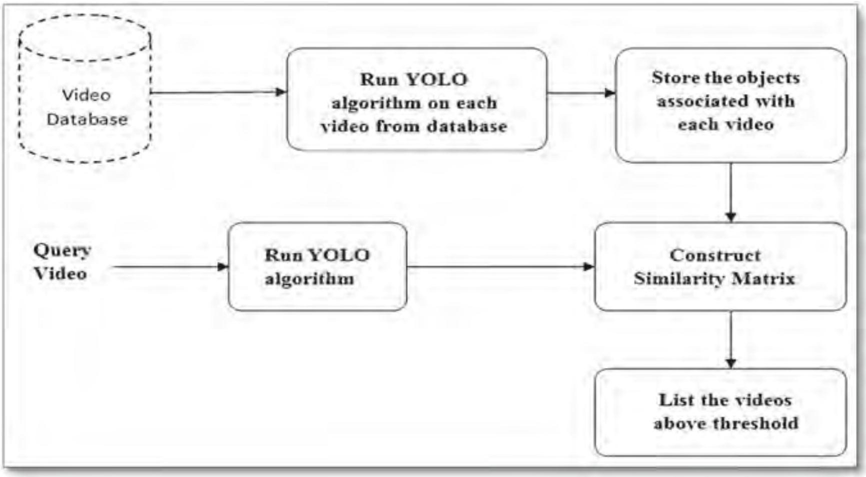


Fig. 1 Overview of the proposed video retrieval method

YOLOv3 pre-trained model was trained for 100 epochs using Google Colab. This model was able to identify the objects from action videos of the database correctly. It took 19 h to train the model.

Step 2: **Use the trained model to find objects from the videos in the database**

For each of the videos in the database, the prominent objects were detected using the above-trained model. These objects were saved in a data structure. These are later required for similarity matching between the given query video and a particular video from the database. These objects now represent the features from each video, and hence, the actual videos are not visited again. And hence becomes an abstraction of the videos.

Step 3: **Use the trained model on the query video to retrieve similar videos**

Once all the videos have the objects detected, the next step is to identify the objects present in the query video. Hence, query video is passed to the YOLO algorithm to detect objects, and the objects from the query videos are retrieved.

Step 4: **Construct a similarity matrix**

Once objects from the query are video is extracted, the next step is to compute the similarity measure between query video and each video present in the database. To compute similarity measure, binary vectors are created depending upon the presence or absence of a particular object. In order to create two binary vectors, first unique objects present in query video and the objects present in a particular video are taken into consideration. The methods used for similarity measure are the Jaccard coefficient and simple matching coefficient.

Given two vectors the following values are computed depending upon the pair of values from each of the vectors. Consider the two vectors *a* and *b*.

M_{01} = the number of attributes where a was 0 and b was 1.
M_{10} = the number of attributes where a was 1 and b was 0.
M_{00} = the number of attributes where a was 0 and b was 0.
M_{11} = the number of attributes where a was 1 and b was 1.
The two measures used to compute are.

A. Jaccard Similarity Measure

The Jaccard similarity measure also known as the Jaccard similarity *coefficient* compares members for two binary sets to see which members are shared and which are distinct. It is a measure of similarity for the two sets of data, with a range from 0 to 100%. The higher the percentage, the more similar the two videos.

$$J(A, B) = \frac{\text{Number of } 1 - \text{to} - 1 \text{ matches}}{\text{Number of all matches and mismatches (except both zero)}} \quad (1)$$

$$J(A, B) = \frac{M_{11}}{M_{01} + M_{10} + M_{11}} \quad (2)$$

B. **Simple Matching Coefficient**

It is used for matching and comparing the similarity between two vectors.

$$\mathrm{SMC} = \frac{\text{Number of matches}}{\text{Number of attributes}} = \frac{M_{00} + M_{11}}{M_{00} + M_{11} + M_{10} + M_{01}} \tag{3}$$

C. **Dice Similarity Coefficient**

It is used to measure the similarity between two given samples.

$$\text{Dice Coefficient} = \frac{2.|a \cap b|}{|a| + |b|} \tag{4}$$

Example of the similarity measures with respect to the proposed method.

Video 1 = [Basketboard, Person]		
Video 2 = [Diving Board, Water, Person]		
Total unique objects = [Basketboard, Person, Diving Board, Water]		
Video 1 = [1, 1, 0, 0]		
Video 2 = [0, 1, 1, 1]		
Jaccard coefficient	Dice coefficient	Simple matching coefficient
J(A,B) = 1/4 = 0.25	DSC = 2/8 = 0.25	SMC = 1/4 = 0.25

In most of the cases, the length of the unique vector obtained from the objects of two videos is limited to 4 to 5 according to the method proposed. Hence, any of the above similarity measures will yield almost the same result. The threshold determined here is 0.75.

When the three measures considered Jaccard coefficient, Dice similarity coefficient, and simple matching coefficient, it was observed that the Jaccard and Dice coefficient and also simple matching coefficient work in a similar manner when used for this proposed method. Simple matching coefficient works the same as Jaccard as this method does not provide an increase to M_{00} state. As the objects are considered only if present in either of the two videos. Similarly, the Dice coefficient also works the same as the Jaccard coefficient. Hence, the Jaccard coefficient is used further.

Step 5: **Retrieve the most similar videos**

Once the similarity measure for each video from the database is calculated, the next step is to retrieve the videos with similarity measure above the threshold. The threshold is decided to depend upon the fact that few objects may be in common in the video even if the videos are dissimilar. Here, it is considered as 0.75.

4 Implementation Details

4.1 Hardware

Processor requirements: A CPU with i3 or higher.

Memory requirements: A minimum memory of 500 GB is required to process a video and store the output.

4.2 Software

Languages used: Python.

Software required.

Desktop: Anaconda enabled with Jupyter Notebook for Python.

Cloud: Google Colab for GPU support.

4.3 Dataset

It has 11 categories of action [14]. This includes basketball shooting, cycling, horseback riding, soccer juggling, swinging, tennis, trampoline jumping, walking with dog, volleyball spiking, diving, and golf. This dataset is very challenging due to large variations in camera motion, object appearance and pose, object scale, view point, cluttered background, illumination conditions, etc.

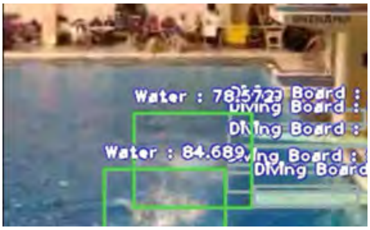
4.4 Output Obtained

See Table 1.

5 Performance Evaluation

The accuracy of this method is basically dependent upon accurate object detection. And since YOLOv3 is till now the best algorithm for object detection, it yields perfect objects as trained using the newly created dataset. The performance of the system proposed is evaluated on precision and recall metrics. Precision–recall is a metric for measuring the outcome of the prediction. Precision is a measure of relevant information retrieved. The recall is a measure of actually relevant data retrieved. In information retrieval, the precision score for an exact match is 1.0. This means every

Table 1 Output obtained by proposed approach

Cycling	Diving
	
Tennis	**Basket board**
	

retrieved document was relevant. When the recall score is 1.0, it means every relevant data was retrieved (Fig. 2; Table 2, 3 and 4).

$$\text{Precision} = \frac{\text{No. of videos reteirved relavant to query video}}{\text{Total Number of videos retrieved}} \tag{5}$$

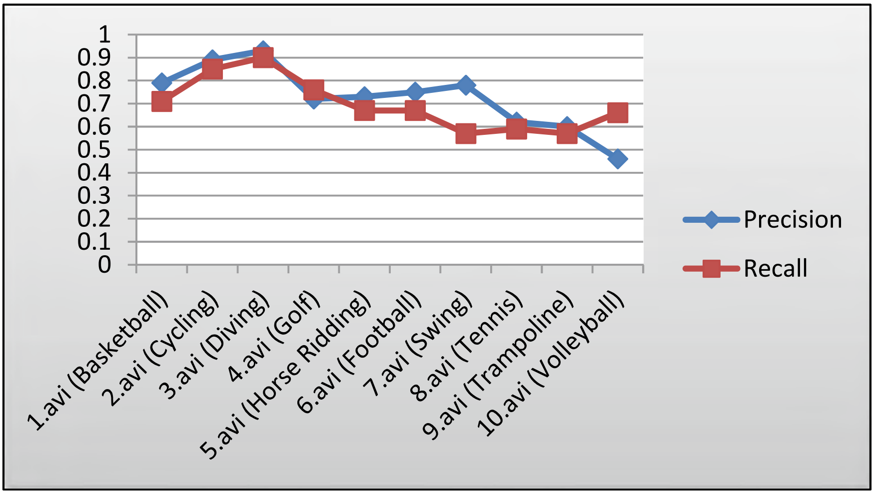


Fig. 2 Precision–recall plot for YouTube action dataset

Table 2 Precision of the proposed method in detail

Sr. No	Query video	Relevant videos	Total retrieved	Precision
1	1. avi (Basketball)	100	127	0.79
2	2. avi (Cycling)	124	140	0.89
3	3. avi (Diving)	140	150	0.93
4	4. avi (Golf)	108	150	0.72
5	5. avi (Horse ridding)	132	189	0.73
6	6. avi (Football)	105	140	0.75
7	7. avi (Swing)	117	150	0.78
8	8. avi (Tennis)	98	158	0.62
9	9. avi (Trampoline)	68	112	0.60
10	10. avi (Volleyball)	77	166	0.46

Table 3 Recall of the proposed method in detail

Sr. No	Query video	Relevant videos	Total relevant videos present in database	Recall
1	1. avi (Basketball)	100	141	0.71
2	2. avi (Cycling)	124	145	0.85
3	3. avi (Diving)	140	156	0.90
4	4. avi (Golf)	108	142	0.76
5	5. avi (Horse ridding)	132	198	0.67
6	6. avi (Football)	105	156	0.67
7	7. avi (Swing)	117	207	0.57
8	8. avi (Tennis)	98	167	0.59
9	9. avi (Trampoline)	68	119	0.57
10	10. avi (Volleyball)	77	116	0.66

Table 4 Precision and recall for YouTube action dataset

Sr. No	Query video	Precision	Recall
1	1. avi (Basketball)	0.79	0.71
2	2. avi (Cycling)	0.89	0.85
3	3. avi (Diving)	0.93	0.90
4	4. avi (Golf)	0.72	0.76
5	5. avi (Horse ridding)	0.73	0.67
6	6. avi (Football)	0.75	0.67
7	7. avi (Swing)	0.78	0.57
8	8. avi (Tennis)	0.62	0.59
9	9. avi (Trampoline)	0.60	0.57
10	10. avi (Volleyball)	0.46	0.66

$$\text{Recall} = \frac{\text{No. of the videos retrieved videos relavant to query video}}{\text{Total number of videos relvant to query video in database}} \tag{6}$$

6 Conclusion

A fast and efficient video retrieval method is proposed in this work. This method is basically dependent on object detection for video retrieval. Since the objects are the inseparable part of videos, the main idea here is to treat these objects as features to retrieve the video. Hence, the state-of-the-art algorithm for object detection is used to extract features "objects" as features. The algorithm used here YOLOv3. Transfer learning is performed using pre-trained weights of YOLO on the YouTube action dataset.

For transfer learning, a dataset of images representing objects of interest was created. The YOLOv3 model was trained on this dataset for 100 epochs. This model gave satisfactory results despite the challenges involved in videos from the database. Similarity measures tried were Jaccard coefficient, Dice coefficient, and simple matching coefficient. It was observed that all three work in the same manner for this proposed method. Hence, Jaccard is used later.

It can be said that this method is efficient because only features (objects) from the videos are extracted only once and stored, unlike other video retrieval methods. This proposed method is fast as to retrieve the video, only task involved is to obtain the features from the given query video and retrieve the similar videos based on the similarity matrix.

7 Discussion

Objects are the prominent part of the videos that have motivated us to use the objects as features. The existing methods are using different approaches. Using the objects from videos is a novel approach and is expected to give better results. The entire proposed approach is dependent upon objects. Object detection should be accurate and efficient. The algorithm used here for object detection is YOLO which is the state-of-the-art algorithm for object detection. This yields accurate results after training the model on the objects of interest.

8 Future Work

This algorithm can be extended to detect and identify the more complex objects. This can be done by training the YOLOv3 model on more number of clear images. Also, the same concepts of object detection can be used for action recognition. The objects and context information can be utilized to detect the action undergone the videos.

References

1. A. Podlesnaya, S. Podlesnyy, Deep learning based semantic video indexing and retrieval. in *Proceedings of SAI intelligent systems conference*. (Springer, Cham, 2016), pp. 359–372
2. Y. Gu,, C. Ma, J. Yang, Supervised recurrent hashing for large scale video retrieval. in *Proceedings of the 24th ACM international conference on Multimedia* (2016), pp. 272–276
3. N. Dimitrova, H.J. Zhang, B. Shahraray, I. Sezan, T. Huang, A. Zakhor, Applications of video-content analysis and retrieval. IEEE Multimedia **9**(3), 42–55 (2002)
4. A. Karpenko, 50,000 Tiny Videos: A Large Dataset for Non-parametric Content-based Retrieval and Recognition (Doctoral dissertation) (2009)
5. H. Yang, C. Meinel, Content based lecture video retrieval using speech and video text information. IEEE Trans. Learn. Technol. **7**(2), 142–154 (2014)
6. M. Gitte, H. Bawaskar, S. Sethi, A. Shinde, Content based video retrieval system. Int. J. Res. Eng. Technol. **3**(06) (2014)
7. S. Iqbal, A.N. Qureshi, A.M. Lodhi, Content based video retrieval using convolutional neural network. in *Proceedings of SAI Intelligent Systems Conference*. (Springer, Cham, 2018), pp. 170–186
8. HOLLYWOOD2: Human Actions and Scenes Dataset. Retrieved from https://www.di.ens.fr/~laptev/actions/hollywood2/
9. M. Jain, J.C. Van Gemert, C.G.M. Snoek,What do 15,000 object categories tell us about classifying and localizing actions?. in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015)
10. J. Redmon, et al., You only look once: unified, real-time object detection. in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016)
11. J. Redmon, A. Farhadi,YOLO9000: better, faster, stronger. in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017)
12. J. Redmon, A. Farhadi,Yolov3: an incremental improvement (2018). *arXiv preprint* arXiv:1804.02767
13. G. Kordopatis-Zilos, S. Papadopoulos, I. Patras, I. Kompatsiaris, FIVR: Fine-grained incident video retrieval. IEEE Trans. Multimedia **21**(10), 2638–2652 (2019)
14. UCF Sports Action Data Set, Retrieved from https://www.crcv.ucf.edu/data/UCF_Sports_Action.php
15. Recognition of human actions Retrieved from https://www.nada.kth.se/cvap/actions/

Exploring a Filter and Wrapper Feature Selection Techniques in Machine Learning

V. Karunakaran, V. Rajasekar, and S. Iwin Thanakumar Joseph

Abstract Nowadays, huge amounts of data are generated by many fields such as health care, astronomy, social media, sensors, and so on. When working with such data, there is a need for the removal of irrelevant, redundant, or unrelated data. Among various preprocessing techniques, dimensionality reduction is one such technique used to clean data. It helps the classifiers by reducing training time and improving the classification accuracies. In this work, the most widely used feature selection techniques were analyzed in machine learning for improving the classification as well as prediction accuracies.

Keywords Feature selection · Machine learning algorithms · Dimensionality reduction

1 Introduction

Attribute selection (AS) plays a very vital role in preprocessing tasks in machine learning or data mining techniques. It is the process of finding the optimal features from a given dataset, i.e.., in the original dataset, those features which contribute more information are retained in the training dataset and the remaining features are eliminated. Attribute selection methods are broadly classified into two:

i. Filter method
ii. Wrapper method.

In the filter method, the selection criterion is purely based on the filter function, i.e., the features are selected based on the filter function and not based on the classifiers. In the wrapper method, the selection criterion is based on classification accuracy, i.e.,

V. Karunakaran (✉) · S. I. T. Joseph
Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India

V. Rajasekar
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani campus, Chennai, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_40

selecting the best features is purely based on classification accuracy obtained from the classifiers. In the wrapper method, a classifier is selected based on the problem domain. The full article is organized in the following manner. Section 2 explains the process of feature/attribute selection by filter method, and Sect. 3 illustrates the process of feature/attribute selection by wrapper method. Finally, Sect. 4 concludes the research work.

2 Attribute Selection by Filter Method

In this method, the selection criterion is purely based on the filter function. In this section, the most widely used techniques for attribute selection is described using the filter method. The general functionality of the filter method is given in Fig. 1.

In feature selection techniques, information gain (IG) is a prominent method and it is also called as a Kullback–Leibler (KL) divergence. Given probability functions of *A* and *B*, KL is a non-symmetric measure of divergence between *A* and *B*:

$$D(X||Y) = \sum_j X(j) \log X(j)/Y(j) \tag{1}$$

In reality, when given probability models *A* and *B*, KL divergence is the expected logarithmic difference between *A* and *B*. It is zero if and only if *A* and *B* are equal.

Information gain can be defined as mutual information. Information gain IG (*K*) can be defined as the reduction in entropy archived by learning a variable *K*:

$$\mathrm{IG}(K) = M(L) - \sum (L_i/L) M(L_i) \tag{2}$$

where

H (*L*) stands for the entropy of the dataset.

H (*Lm*) stands for the entropy of the *m*th subset generated by partitioning *L* based on attribute *K*.

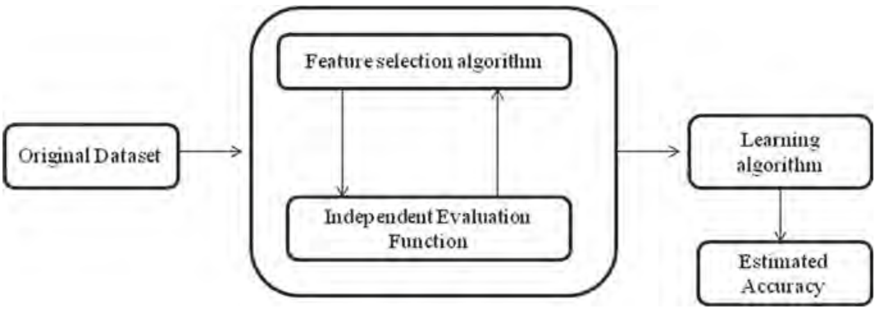


Fig. 1 Attribute selection using filter approach

Here, the attributes are ranked based on the values obtained from the information gain—for attributes having high information gain, those features will be ranked higher than others because those features have substantial power in classifying the data. It evaluates each feature one by one and eliminates those features which contribute less information or unwanted features from the training set. This method is simple and cost-effectiveness when compared to other filter methods. This method is fast, even in such cases where the number of attributes is larger than the number of instances [1].

The gain ratio is also one of the well-known methods for feature selection, proposed by Mitchell in the year 1997. The ID3 algorithm is the best subset of features identified by the information gain method. C4.5 algorithm is a successor of the ID3 algorithm. In the C4.5 algorithm, the best subset of features was identified by using a gain ratio method (Salzburg 1994). In the information gain method, the split is determined based on the feature having the highest information gain—measured in bits. This yields better results but will favor splitting the features and adding to the dataset—which already have too many features. The limitation of the information gain method can be overcome by the gain ratio method. The gain ratio method will incorporate the value of split and will determine what proportion of IG is valuable for those splits [1, 2].

Mutual information is a well-known method for attribute selection, proposed by Battiti in the year 1994. In this article, the best subset is identified by using mutual information. The effectiveness of the mutual information feature selection method is evaluated using a neural network classifier and compared with principal component analysis (PCA) and random feature selection method (RFSM). In terms of classification accuracy, the mutual information feature selection method provides a better classification accuracy compared with random feature selection method (RFSM) and principal component analysis (PCA) [3].

The conditional mutual information concept was proposed by Cheng et al. in the year 2011 for selecting the best subset of attributes. This method not only identifies the best subset of attributes and also analyzes the synergy and redundancy. Mutual information feature selection is used for many applications, but this method does not have features synergy and analyzing redundancy. The conditional mutual information method considers the redundancy and synergy between the attributes and identifies the favorable attributes. The CMIFS method reduces the probability of choosing the best features as a redundant feature. The experimental result shows that the conditional mutual information feature selection method provides a better classification accuracy than the mutual information feature selection method [4].

Kira and Rendell proposed a RELIEF algorithm, a simple and well-known method for feature weight estimation. If the features are tightly dependent on each other, then it is difficult to identify which feature is contributing more information or quality of the feature. The relief algorithms identify the quality of attributes based on the values that are distinguishable between records/data and the near points.

The pseudocode of the basic RELIEF algorithm is given as follows:

```
Set all weights M [attributes] =0
for i =1 to j do
Begin
Randomly select an instance A;
Find the nearest hit L and nearest miss O;
For F= 1 to all_ attributes do
M [attributes] = M [attributes] – diff (attributes, H) / j + diff (attributes, M) / j;
end;
Where,
A-stands for instances in the dataset
L- stands for a Nearest hit of instances
O- stands for Nearest miss of instances
```

In the RELIEF algorithm, a diff() function is used for finding the difference between the values of features K for 2 instances. If the diff() function returns ‘1′ then both instances have different attribute values. If the diff() function returns ‘0′ then both instances have the same value. The RELIEF algorithm identifies the relevant attributes and irrelevant attributes using diff() function [5].

3 Feature Selection Using Wrapper Method

Feature selection was carried out by the following methods, namely sequential forward selection (SFS) and sequential backward elimination (SBE) method. Both methods belong to the iterative method. In the sequential forward selection method, the training starts with an empty set ‘A’ gradually by adding the features into the training set one by one, using some evaluation function. In the end, the training set ‘A’ contains those features which contribute more relevant information during classification. The working model of sequential backward elimination is a reverse of sequential forward selection. The sequential backward elimination method starts with an entire training set ‘A’, gradually removing the attributes one by one using some evaluation function. In the end, training set ‘A’ contains those features which contribute more relevant information during the classification task. Sequential forward selection as well as sequential backward elimination methods should give an effectively reduced dataset and better classification accuracy. Both methods tend to be fascinated by local minima.

The caveats of sequential forward selection as well as sequential backward elimination methods were overcome using randomized algorithms. All randomized algorithms tend to include randomness into their search procedure to escape from the local minima. One of the well-known randomized algorithms is a genetic algorithm. Mohamad (2004) used a genetic algorithm to identify the best subset of attributes from both high dimensional datasets and low dimensional datasets. The performance of attribute selection using a genetic algorithm was evaluated by using support vector machine. The experimental result shows that the selected subsets of the feature were good and obtained better classification accuracy for training data for both small dimensional and large dimensional datasets [6].

Zhang and Sun tried a feature selection by using Tabu search for the high dimensional dataset. The proposed method is compared with three other methods, namely sequential forward selection (SFS), sequential backward selection, and genetic algorithm (GA). The experiment result shows that the reduced subset of features produces better classification results in a minimum amount of time. This method is capable of gaining local minima and global optimization [7].

Huang et al. proposed a hybrid genetic algorithm for feature selection. In this article, the author worked on both the wrapper method and filter method. Here, the work is divided into phases. The first phase is the outer stage optimization, the best subset of features was identified by using a wrapper method, and this method achieves global optimization. The second phase is an inner stage optimization, the best subset of features was identified by using filter method and this method achieves local optimization. Both the methods will assist each other to provide a better global predictive accuracy and high local search efficiency. Experimental results show that the proposed hybrid genetic algorithm provides excellent classification accuracy compared to the recursive feature elimination method and Battiti's greedy feature selection method (MIFS) [8].

Ahmad tried both feature extraction and feature selection using principal component analysis and particle swarm optimization, respectively. This proposed method is compared against existing methods, namely principal component analysis (PCA) and principal component analysis with genetic algorithm (PCA-GA). The classification is carried out by a modular neural network classifier. The obtained experimental result depicts that the combining principal component analysis (PCA) and particle swarm optimization (PSO) provides a better detection rate and lesser false alarm rate compared to principal component analysis (PCA) and Principal Component Analysis + Genetic Algorithm methods (PCA + GA) [9].

Suganthi and Karunakaran tried a combination of instance selection and feature extraction using cuttlefish optimization algorithm and principal component analysis, respectively. The experiment was performed using four large datasets under the 4 criteria's such as.

i. With original dataset,
ii. Feature extraction using PCA,
iii. Instance selection using the cuttlefish optimization algorithm,
iv. Combination of instance selection (IS) and feature extraction (FE) using the cuttlefish optimization algorithm (COA) and principal component analysis (PCA).

The experimental results show that the combination of instance selection plus feature extraction using cuttlefish optimization algorithm and principal component analysis will take only a small amount of training time when compared to the other criteria's [10].

Karunakaran et al. proposed cuttlefish optimization algorithm (COA) through Tabu search (TS) for both feature selection and instance selection. The experiment was analyzed with four large datasets obtained from the UCI repository. The method cuttlefish optimization algorithm through Tabu search was better in terms of detection rate (DR), accuracy, false positive rate (FPR)/type one error, and training time (TT) of classifiers. The experiment was carried out with SVM and KNN classifier [11].

Karunakaran et al. tried the bees algorithm to find the optimal subset of features from the entire dataset. Dataset used in this article is a weather dataset obtained from the UCI repository. The experimental result shows that the proposed method is better in terms of accuracy, type 1 error/false positive rate and detection rate compared to the entire dataset [12]. The below-mentioned methods tried with different metaheuristic for solving feature selection problems and achieved better classification results [13–15] (Table 1).

Table 1 Analyzed various filter and wrapper feature selection methods

S. No	Authors and year	Description	Inference
1	Kullback and Leibler [16] Cover and Thomas [17]	Information gain (IG) is used for ranking features. According to the information gain value, the features were arranged in descending order. (First feature have a higher rank and the last feature have the lowest rank)	1. Simple and cost-effective 2. It also provides good performance if features are larger instances 3. It will perform poor if the attributes have a large number of distinct values
2	Mitchell [1] Salzberg [2]	Gain ratio (GR) is used for identifying the best features among the entire features	1. Simple and cost-effective 2. It will overcome the limitation of the information gain method. For example, if two features are having the same information gain. Then, the gain ratio will select which attribute has less number of categories/distinct values 3. If the attributes are more dependent on each other, then the performance leads to poor values
3	Kira and Rendell [5]	RELIEF algorithm will estimate the quality of features based on how well their values are unique between the instances and near points	1. If the features are strongly dependent it will produce a good result 2. The basic RELIEF algorithm supports only two-class problems

(continued)

Table 1 (continued)

S. No	Authors and year	Description	Inference
4	R.Battiti [3]	Sequential forward selection (SFS) algorithm begins with an empty feature set 'S' and gradually adds the features in the set by using some evaluation function. Finally, the feature set S contains only those features were contributing more information during classification Sequential backward selection (SBS) begins with all features and repeatedly removes a feature; the feature which contributes very limited information	1. Implementation is simple 2. Both the method provides an effectively reduced dataset and they give better classification accuracy 3. Both methods achieve only local optima
5	Mohamad [6]	Genetic algorithm (GA) is used to search out and identify the potential informative features and support vector machine (SVM) is used for classification	1. This method will overcome the limitation of tendency to become trapped in local optima 2. The experiment result shows the selected features are good to get better classification accuracy
6	Zhang and Sun [7]	Tabu search (TS) method used for solving high dimensionality problem by selecting the best features from others in the dataset	1. The method is compared with genetic algorithm (GA), sequential forward selection (SFS), and sequential backward selection method (SBS) 2. Tabu search (TS) method provides quality optimal subset and less computational time compared to existing systems

(continued)

Table 1 (continued)

S. No	Authors and year	Description	Inference
7	Huang et al. [8]	1. Both wrapper and filter methods are applied for finding the best subset of features 2. It consists of two stages i. Selecting the best subset of features by using the wrapper method (global search) ii. Selecting the best subset of features using the filter method (local search)	1. It performs both global search and local search 2. The method is compared with the feature elimination method and greedy feature selection (GFS) method 3. The hybrid genetic algorithm (HGA) provides better classification accuracy than existing systems
8	Ahmad [9]	1. Principal component analysis (PCA) is used for feature extraction 2. Particle swarm optimization (PSO) is used for feature selection 3. The performance is evaluated by modular neural network classifier	The combination of PCA and particle swarm optimization (PSO) provides a better detection rate, false positive rate, and classification accuracy than PCA and PCA + Genetic algorithm
9	Suganthi and Karunakaran [10]	1. Instance selection is carried out by using the cuttlefish optimization algorithm (COA) 2. Feature extraction using PCA 3. Combination of instance selection and feature extraction	The combination of the cuttlefish optimization algorithm (COA) and principal component analysis (PCA) provides a small amount of training time than other criteria
10	Karunakaran et al. [11]	Both feature selection (FS) and instance selection (IS) was carried out by using the cuttlefish optimization algorithm (COA) through Tabu search (TS)	This method provides better in terms of accuracy (ACC), false positive rate (FPR), and detection rate (DR) Computational time (CT) is less
11	Karunakaran et al. [12]	1. Weather dataset used 2. Optimal subset features identified by using the bees algorithm (BA)	This method provides better in terms of accuracy, false positive rate, and detection rate

4 Conclusion

This article delivers a survey on attribute selection methods. The basic concepts of attribute selection methods were analyzed and discussed. The various attribute selection methods were broadly categorized into filter and wrapper methods, and an overview of the filter as well as wrapper methods was discussed. Many approaches were proposed for attribute selection. Most of the approaches still suffer from the problem of stagnation, i.e., the searching process will be stopped in local optima, instead of reaching into global optima and the methods that belongs to global optimization techniques does not contribute good accuracy and takes a huge amount of time to train the classifiers. Further, efficient searching techniques are required for identifying the best subset of attributes.

References

1. T.M. Mitchell, in *Machine Learning.* (WCB/McGraw-Hill, Boston, Massachusetts, 1997)
2. S.L. Salzburg, C4.5: Programs for machine learning. Morgan Kaufmann publishers, inc., Machine Learn. **l6**(3), 235–240 (1994)
3. R. Battiti, Using mutual information for selecting features in supervised neural net learning. IEEE Trans. Neural Netw. **5**(4), 537–550 (1994)
4. G. Cheng, Z. Qin, C. Feng, Y. Wang, F. Li, Conditional mutual information-based feature selection analyzing for synergy and redundancy. Etri Joixnal **33**(2), 210–218 (2011)
5. K. Kira, L.A. Rendell, The feature selection problem: Traditional methods and a new algorithm. Aaai **2**, 129–134 (1992)
6. M.S. Mohamed, Feature selection method using genetic algorithm for the classification of small and high dimension data. in *Proceeding International Symposium. Information Communication Technology* (2004) pp. 13–16
7. H. Zhang, G. Sun, Feature selection using tabu search method'. Pattern Recogn. **35**(3), 701–711 (2002)
8. I. Huang, Y. Cai, X. Xu, A hybrid genetic algorithm for feature selection wrapper based on mutual information. Pattern Recogn. Lett. **28**(15), 1825–1844 (2007)
9. I. Ahmad, Feature selection using particle swarm optimization in intrusion detection. Int. J. Distrib. Sens. Netw. **11**(10), 806954 (2015)
10. M. Suganthi, V. Karunakaran, Instance selection and feature extraction using cuttlefish optimization algorithm and principal component analysis using decision tree. Cluster Comput. **22**(1), 89–101 (2019)
11. V. Karunakaran, M. Suganthi, V. Rajasekar, Feature selection and instance selection using cuttlefish optimisation algorithm through tabu search. Int. J. Enterprise Netw. Manage. **11**(1), 32–64 (2020)
12. V. Karunakaran, S.I. Joseph, R. Teja, M. Suganthi, V. Rajasekar, A wrapper based feature selection approach using bees algorithm for extreme rainfall prediction via weather pattern recognition through svm classifier. Int. J. Civil Eng. Technol. (IJCIET) **10**(1) (2019)
13. O. Gokalp, E. Tasci, A. Ugur, A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification. Expert Syst. Appl. **146**, 113176 (2020)
14. S. Mahendru, S. Agarwal, Feature selection using Metaheuristic algorithms on medical datasets. in *Harmony Search and Nature Inspired Optimization Algorithms* (Springer, Singapore, 2019), pp. 923–937

15. A.A. Lyubchenko, J.A. Pacheco, S. Casado, L. Nuñez, An effective metaheuristic for bi-objective feature selection in two-class classification problem. J. Phys.: Conf. Series **1210**(1), 012086 (2019) IOP Publishing
16. S. Kullback, R.A. Leibler, On information and sufficiency. Ann. Math. Stat. **22**(1), pp. 79–86. (1951)
17. T.M. Cover, J.A. Thomas, 'Elements of information theory', 2nd edition. Hoboken, Wiley-Inter science (2006)

Recent Trends in Epileptic Seizure Detection Using EEG Signal: A Review

Vinod J. Thomas and D. Anto Sahaya Dhas

Abstract Epilepsy is a neurological brain disorder. Seizures are the key characteristics of epilepsy. Seizures occur in a community of neurons in certain parts of the cerebral cortex during which rapid bursts occur in uncontrolled electrical activity. More than 2% of the people in the world suffer from this brain disorder. Any abnormality in brain functionality can be easily identified by the use of an electroencephalogram (EEG). As the EEG recordings are usually a few hours long, the visual analysis of the EEG signal is a time consuming and laborious job. With the advancement in signal processing and digital methods, several systems can automatically detect seizures from EEG signal. This article reviews the state-of-the-art methods and concepts which give an orientation for future research work in the field of seizure identification. Here, various methods for the detection of epileptic seizures that mainly differ in feature extraction techniques were discussed. Different feature extraction techniques that involve extracting features from the frequency domain, time domain, frequency and time domain combinations, wavelet transform, and powerful deep learning methods are discussed with a comparison of the methods. When considering the EEG signal's nonstationary existence and different artifacts affecting the signal, feature extraction based on deep learning approaches is found to be robust and appropriate for the identification of seizures.

Keywords Seizure detection · Discrete wavelet transform · Epilepsy · Encephalogram · Artificial neural network · Deep learning

V. J. Thomas (✉) · D. Anto Sahaya Dhas
Vimal Jyothi Engineering College Chemperi, APJ Abdul Kalam Technological University, Thiruvananthapuram, Kerala, India
e-mail: vinodkurisinkal@gmail.com

D. Anto Sahaya Dhas
e-mail: dr.anto@vjec.ac.in

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*, Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_41

1 Introduction

World Health Organization (WHO) estimates about 50 million people are affected by epilepsy. At a given period, the average proportion of the general population with active epilepsy is between 4 and 10 per 1000 individuals. An estimated five million individuals are diagnosed with epilepsy each year worldwide. It is estimated that 49 per 100,000 people with epilepsy are diagnosed each year in high-income countries. This number may be as high as 139 per 100,000, in low- and middle-income countries [1]. A person with epilepsy suffers from seizures characterized mainly by recurrent, unpredictable, and uncontrolled electrical surges which really affect the normal life of an individual. Some seizures are very hard to observe because they just lead to a minor confusion of the mind, while others seldom cause loss of consciousness leading to injury or fatality. Epileptic seizures are generally categorized as generalized seizures and partial seizures depending on the place at which the seizure begins. Partial seizures are also known as focal epilepsy. Some partial seizures that require a surgical operation in which a small portion of the cortex is removed or may be cured. The EEG is the primary examination for epilepsy diagnosis and the compilation of information on the type and location of seizures. Figure 1 illustrates the different forms of EEG signals that are used in the detection of epileptic seizures with the corresponding spectrum. They were taken from the Bonn University database. Set-A displays the EEG signals taken with eyes open from healthy subjects and Set-B suggests the same with eyes closed. Set-C and Set-D are interictal type, and epileptic signals. The epileptic signals reported during the seizure are indicated by Set-E.

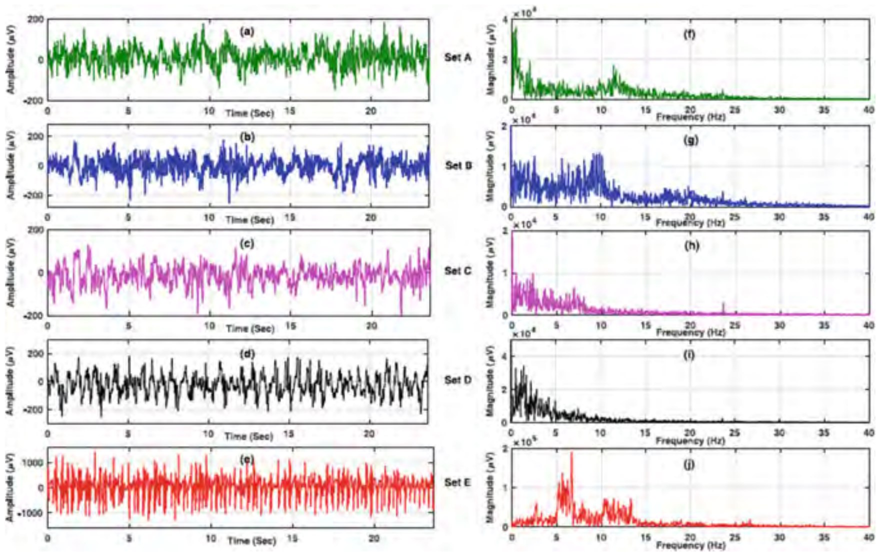


Fig. 1 Different types of EEG signals and their spectra. (*Source* Reference [43])

Several feature extraction techniques have been developed for the automatic detection of epileptic seizures. Most techniques use hand-designed features taken from the frequency domain, time domain, time and frequency domain combinations. Tzallas et al. [2], Correa et al. [3], Chan et al. [4], Aarabi et al. [5], Srinivasan et al. [6], Meier et al. [7, 8], Minasyan et al. [9], and Abibullaev et al. [10] have suggested that many of the other researches use features derived from the wavelet domain. Nowadays, there are more efficient systems that use deep learning/machine learning approaches for feature extraction. Some other works use a different technique. For example, the authors in [11–13] use empirical mode decomposition. The authors in [14, 15] use a rational function, and the authors in [16] use a statistical-based approach. But due to the adaptability and self-learning capabilities, machine learning approaches are found to be robust and appropriate in many situations. Suppose, a dataset that includes positive and negative cases for a particular two-class problem have true positives (TP) and true negatives (TN). The algorithm which is selected may predict true positive cases as negative cases and vice versa; therefore, a false positive (FP) and false negative (FN) cases based on these values. They are indicated by FP and FN, respectively. Following are the main performance metrics used to evaluate the usefulness of different approaches.

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \times 100$$

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN} + \text{FP})} \times 100$$

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{FN} + \text{TP} + \text{FP} + \text{TN})} \times 100$$

This article is structured as follows. Section 2 gives an overview of the seizure detection system. Section 3 provides an idea of wavelet transformation and the domain of time–frequency. Section 4 addresses different seizure detection methods in which characteristics are obtained from the time domain, frequency domain, and a mixture of both. Section 5 illustrates the latest techniques which employ deep learning approaches for feature extraction. Each section outlines the central idea of the papers along with the methods adopted for feature extraction and classification. The results obtained for each work and the input data used for the corresponding system also are discussed. Section 6 includes a conclusion and future research direction in this area.

2 Overview of the Detection Systems

A general block diagram representation of all seizure detection methods is represented in Fig. 2. The brain's electrical activity is often polluted by different noise and artifacts. It can impact the precision of detection. Therefore, to remove all sources

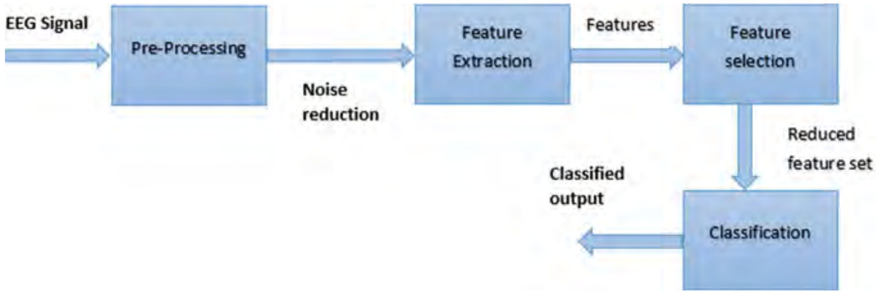


Fig. 2 General block diagram of the seizure detection system

of noise from the EEG data, proper preprocessing has to be performed. EOG, ECG, and EMG are the key sources of artifacts influencing the EEG signal. There are many techniques in the literature for signal conditioning. Klados et al. [17] use regression techniques for the removal of artifacts. Safieddine et al. [18] use a wavelet-based decomposition approach for the removal of artifacts. Casarotto et al. [19] used principal component analysis (PCA) to eliminate ocular artifacts and showed that PCA is more effective computationally than the methods of linear regression. The artifacts are removed by the independent component analysis (ICA) which are used by many authors [20, 21]. In addition to this, different filtering approaches are often used to eliminate artifacts, including adaptive filtering [22], wiener filtering [23], and sparse methods of decomposition [24]. The pre-processed signal is applied to the feature extractor block and various features are derived. Then the dimensionality of the feature set is be reduced by feature selection algorithms. Finally, the reduced feature set is applied to the classification unit and the output signal is an indication of seizure/non-seizure activity.

3 Wavelet Transform

Fourier transform cannot provide time–frequency localization. Hence, the wavelet representation can eliminate this problem. These wavelets are wave-like functions and can have a finite duration, zero average value, and finite energy. The wavelet series uses a continuous-time signal that can be represented as

$$f(t) = \sum_{s} a_{r0,s} \Phi_{r0,s}(t) + \sum_{r=r0}^{\propto} \sum_{s} b_{r,s} \Psi_{r,s}(t) \tag{1}$$

The signal is decomposed into different levels having a different resolution. The first part in the above equation is called the approximation function and the second part is called detail functions. Equation (1) represents the basic multi-resolution framework for representing any signal using a wavelet transform.

Here $\Phi_{r0,s}(t)$ and $\Psi_{r,s}(t)$ are the basis functions used to represent the signal and are analogous to the complex exponential signal in the Fourier transform. They actually represent the dilated and translated versions of basic functions called Φ(t) and Ψ(t), respectively. Here, Φ(t) is called the scaling function and Ψ(t) is called the wavelet function. There are different combinations of scaling and wavelet functions depending on which type of wavelet is used in a particular study. In Eq. (1), $a_{r0,s}$ and $b_{r,s}$ are called wavelet transform coefficients and they are analogous to Fourier coefficients in Fourier transform.

where,

$$\Phi_{r0,s}(t) = 2^{r0/2}\Phi\left(2^{r0}t - s\right) \tag{2}$$

Equation (2) represents the multiplication of scaling function translated by a unit "s" and dilated by 2^{r0}, with the factor $2^{r0/2}$.

$$\Psi_{r,s}(t) = 2^{r/2}\Psi\left(2^{r}t - s\right) \tag{3}$$

Equation (3) represents the multiplication of wavelet function translated by a unit "s" and dilated by 2^{r}, with the factor $2^{r/2}$.

The factors $a_{r0,s}$ and $b_{r,s}$ in (1) are given by

$$a_{r0,s} = \int f(t)\Phi_{r0,s}(t)\mathrm{d}t \tag{4}$$

$$b_{r,s} = \int f(t)\Psi_{r,s}(t)\mathrm{d}t \tag{5}$$

Equations (4) and (5) represent wavelet transform coefficients and they are also known as approximation and detail coefficients. The advantage of wavelet coefficients is that each wavelet coefficient is a function of two variables r and s. Here, r and s represent the scaling and shifting of the wavelet, respectively. Therefore, the wavelet coefficients actually insist the contribution of a wavelet which is dilated and translated by some value. By taking the Fourier transform of the dilated and translated version of the basic wavelet, we can determine the frequency content. The parameter s gives information about the time instant where this wavelet occurs. Thus, the wavelet transform can provide time–frequency localization. One can choose a wavelet that is suitable for a particular application. This feature is very important when a nonstationary signal is analyzed like EEG. In any application, finding the suitable wavelet function, the number of decomposition stages, and then finding the appropriate features from these sub-bands are a challenging task.

4 Time Domain/Frequency Domain-Based Methods

There are twelve articles summarized here. The basic aim is to demonstrate various approaches, interrelationships between approaches, and different opportunities in the same area to achieve changes in seizure detection.

Hernández et al. [25] suggested a method to detect epilepsy by extracting 52 features from the time domain, frequency domain, and a combination of the two domains above. The set of temporal features that are implemented in this work are signal power, average value, first and second differences, the normalized version of first and second differences, and standard deviation. Some other temporal characteristics introduced in this work that attempt to account for the EEG signal's nonstationary existence are Hjorth features [26], nonstationary index [27], higher-order crossings [28]. Some of the most common spectral features are derived from the continuum of power density that can be obtained by measuring the signal's Fourier transform. Short-time Fourier transform is used here with the Hamming window and the last set of features derived in this research was from the time–frequency domain. To obtain the characteristics, the discrete wavelet transform used here is 10 distinct classification models. The freely accessible Bonn University dataset is used in this work. In terms of accuracy in the state of the art, the authors find that their models are same in performance and, in some cases, outperform the classifiers.

Sriram and Raghu [29] present an approach in which a total of 26 features have been derived from the time domain, frequency domain, information theory, and statistical approach for the detection of focal and non-focal epileptic seizures. They used the Bern-Barcelona database for this purpose. Wilcoxon rank-sum test was conducted and they found that five features are insignificant for classification. They also found that removing outliers from the feature set improves classification accuracy. By using Turkey's range test, the outliers were eliminated. Then, the 21 characteristics selected are fed to the optimized support vector machine classifier and obtained an accuracy of 92.15%, 89.74% precision, and 94.56% sensitivity. The result shows that the proposed method outperforms the existing methods which use the same database and employ some signal splitting techniques like empirical mode decomposition and discrete wavelet transform.

Rational transform is a depiction of the time–frequency domain based on rational functions. It is a free parameter-based approach that uses some optimization algorithms such as particle swarm optimization (PSO) [30] to pick optimal bases. Samiee et al. [31] carried out seizure detection by using a rational discrete short-time Fourier transform. Here, the authors used PSO to find the optimal location of each EEG epoch's pole, which gives the device of a compact time–frequency representation. To determine the efficiency of the proposed framework, the authors used the Bonn University dataset. With various algorithms such as Naïve Bayes, logistic regression, support vector machine, k-nearest neighbors, and multi-layer perceptron architectures (MLP), the system's efficiency is evaluated. The selected feedforward MLP serves as the optimal classifier after a sequence of experiments. They had a sensitivity of 99.9%, a precision of 99.6%, and an accuracy of 99.8% for the two-class Bonn university dataset (E-A) classification.

Wang et al. [32] suggested the identification of epileptic seizures by using extraction of multi-domain features and nonlinear analysis. A wavelet threshold denoising procedure is used to eliminate the dynamic low-frequency noise associated with the EEG signal. The signal is broken down into five frequency sub-bands using the fourth-order Daubechies wavelet, and the maximum, minimum, mean, and total variance measurements of the wavelet coefficients are derived from each subband. The characteristics derived from the time domain include mean, variance, variation coefficient, and total variation. The relative power spectral density determined from the fast Fourier transform coefficients is derived from the frequency domain. Due to the nonstationary nature of the EEG signal, features extracted via linear analysis may be sufficient. Therefore, using empirical mode decomposition and intrinsic mode functions, several additional features are extracted. By using principal component analysis and variance analysis, the dimensionality of the feature set is reduced. The dataset used for this study is Bonn university data. The set of input features are fed to a set of classifiers including linear discriminant analysis, Naïve Bayesian, k-nearest neighbor, logistic regression, and support vector machine. The proposed seizure detection method achieved an average accuracy of 99.25%.

A seizure detection method based on time–frequency analysis was proposed by Tzallas et al. [33] using short-time Fourier transform, power spectral density, and many other time–frequency distributions. Different distributions of time–frequency used in this analysis are: Margenau-Hill, Wigner-Ville, Rihaczek, pseudo Margenau–Hill, pseudo Wigner-Ville, Born-Jordan, Butterworth, Choi-Williams, generalized rectangular, reduced interference, smoothed pseudo Wigner-Ville, and Zhao-Atlas-Marks. By dividing the time by three equal-sized windows and frequency by five sub-bands, a time–frequency grid is created. By integrating the power spectral density over the time–frequency windows, the characteristics are extracted. Additonally, the total signal energy is also used as a feature. The principal component analysis is used here to decrease the size of the feature set and an artificial neural network is used as a feature set to be graded. A seizure detection approach based on time, frequency, time–frequency parameters and nonlinear analysis was proposed by Gajic et.al [34]. The five sub-bands of EEG signals that are of clinical interest [delta (0–4 Hz), theta (4–7 Hz), beta (14–30 Hz) and gamma (31–64 Hz)] are considered in this technique. Then, using time, frequency, time–frequency, and nonlinear methods, the different characteristics were extracted from these sub-bands. By using scatter matrices, the dimensionality of the function space is minimized. The reduced two-dimensional feature space is applied to a quadratic classifier to detect epileptic activity. The authors used the Bonn University dataset to detect the accuracy of the model. They reached a detection accuracy of 98.7% overall.

Observation: The frequency domain approaches are ideal to have long data records. But time information will be absent. Even though the time domain methods are fast, it does not give information about frequency contents of the signal. A combination of the features from these two domains will give a better result. But a transform that can inform us about time and frequency behavior simultaneously is preferable for efficient signal analysis. The wavelet transform can serve this purpose and the following are a set of papers that employ different possibilities of wavelet transform for feature extraction.

Li et al. [35] used wavelet-based envelope analysis and neural network ensemble to detect normal, interictal, and epileptic signals. Here, the signal is decomposed into five levels or sub-bands using discrete wavelet transform. Daubechies wavelet of fourth order is used and the sub-bands (d1-d5 and a5) are applied to a Hilbert transformer and the sub-band envelopes resulting from it are obtained. The features selected are the average value, signal energy, standard deviation, and the peak value of the envelope spectrum in every sub-band. The features are classified by using the neural network ensemble (NNE). NNE is a mixture of neural networks and each has no association with other networks. Here also Bonn university dataset is used and the accuracy obtained is 98.78%. The advantage of this method is that by combining DWT with envelope analysis, they could improve the discrimination of the features. The authors tested the performance of the algorithm with different classifiers and they found that compared to other classifiers such as back propagation algorithm, KNN, support vector machine, and linear discriminant analysis, NNE gives better accuracy.

For the identification of normal, pre-seizure and seizure EEG signals, Harpale and Bairagi[36] used pattern-adapted wavelet transform. An estimated wavelet was constructed by taking into account the seizure waveform from a single recording source. Continuous wavelet transform coefficients were calculated using pattern-adapted wavelet transform. The average power, variance coefficient, RMS value, power spectral density, etc., are different features derived using the above coefficients. These characteristics are fed to the fuzzy classifier for seizure, pre-seizure and regular EEG signs to be identified. The data collection used for this analysis was taken from the Boston Children's Hospital CHB-MIT Scalp Dataset. The advantage of the method is that here the wavelet adapted to the situation is used to improve the accuracy. Moreover, pre-seizure EEG signal also is identified so that the patient can be alerted well before the seizure occurs. They achieved an overall classification accuracy of 96.48% and an accuracy of 96.02% for pre-seizure detection which is greater than the accuracy obtained by some other methods which use the same fuzzy classifier.

For the detection of focal and non-focal EEG signals, Sharma et al. [37] used localized orthogonal wavelet filter banks with time–frequency. Epileptic signals are referred to as focal EEG signals, whereas signals obtained from the non-epileptic region of the brain are referred to as non-focal EEG signals. In this paper, the authors built a new class of filter banks that have a better localization of time–frequency than the equivalent Daubechies filter. The signal is decomposed using the above filter banks and wavelet coefficients were calculated. Then using these coefficients, different entropy-based features are extracted. The dataset used for this analysis was taken from the database of Bern-Barcelona. They could achieve a maximum accuracy of 94.25%, which is much greater than many other works using the same database to distinguish focal and non-focal EEG signals. The importance of this method is that some focal epilepsy can be cured by surgically removing the portions of the brain which contribute to epilepsy.

For the detection of seizures, Kumar and Kolekar [38] have used a wavelet-based study. The signal is decomposed into five frequency sub-bands, namely delta, theta, alpha, beta and gamma, using discrete wavelet transform. But the relevant information regarding the seizure is carried by only three bands: theta, alpha, and beta. From these three sub-bands, distinct characteristics such as energy, variance, zero-crossing rate, and fractal dimension were assessed. The above features were applied to support vector machine classifiers. When compared to the current approaches, they achieved very high sensitivity and accuracy. Features such as fractal dimension, zero-crossing rate, and SVM kernel functions such as polynomial and Gaussian radial basis functions were responsible for the improved sensitivity.

A new method using wavelet decomposition of multichannel EEG data is proposed by Kaleem et al. [39]. EEG signals from all channels are divided into segments of four seconds duration. These segments are decomposed into four frequency sub-bands like the previous work using discrete wavelet transformation [38]. Then, from each of these sub-bands, three characteristics are extracted. Feature 1 is the energy of the wavelet transform-related components of approximation and detail functions. Feature 2 is the amplitude spectrum sparsity, and feature 3 is the sum of amplitude spectrum derivative. CHB-MIT dataset is used for validation of the proposed methodology. They obtained 99.6, 99.8 and 99.6% accuracy, sensitivity, and specificity values, respectively. The obtained results are found to be outperforming some of the existing methods. Two aspects distinguish this method from many other methods. The first one is that there is no feature processing is required. The second is the feature that are extracted in a computationally efficient manner.

A system in which the input signal is decomposed into six levels was proposed by Chen et al. [40]. But the important difference from all other approaches is that the inherent down-sampling operation associated with the wavelet decomposition is absent here. This method is called non-subsampled discrete wavelet transform. Therefore, the number of coefficients at any level is the same as that of the original input signal. Then forward one-dimensional FFT is applied to the detail coefficients at levels 3, 4, 5, and 6. Because these levels contain significant information associated with the seizure. The resulting Fourier spectra are used as feature vectors and are fed to the nearest neighbor classifier to identify epileptic seizure activity. Bonn university data is used in this study. The experimental results showed that the proposed method achieved 100% accuracy for the identification of EEG seizures.

Observation: So far, various methods for epilepsy identification that uses domain-based (time, frequency, time–frequency) approaches were discussed. But these domain-based methods have some disadvantages. First, they are susceptible to detrimental variations in seizure patterns. This is because EEG data is nonstationary and its statistical characteristics can differ over time across patients and for the same patient. The presence of various artifacts that affect EEG signals like eye blinking, muscle artifacts, etc., were ignored for the data acquisition systems. These artifacts

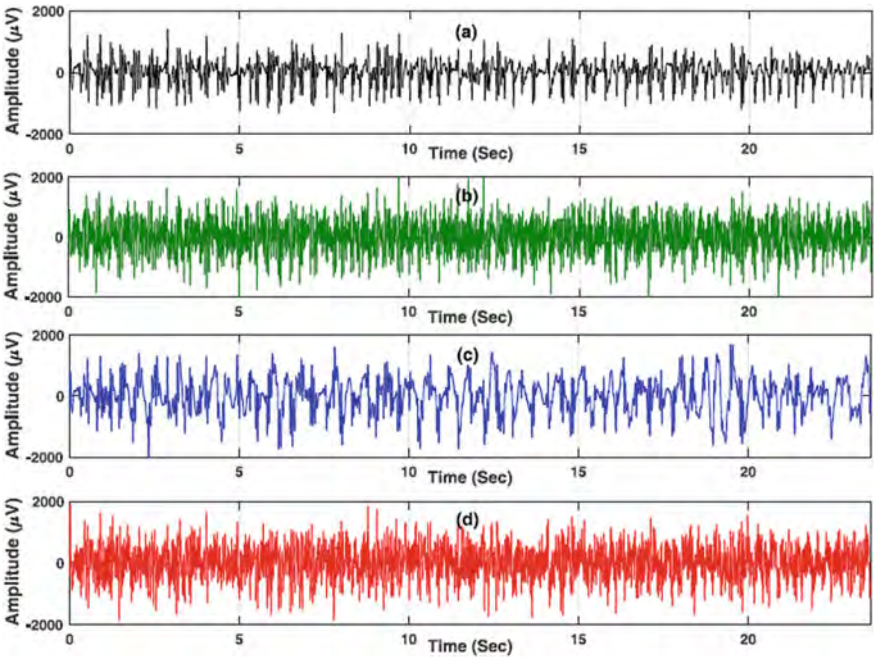


Fig. 3 EEG signals corrupted by artifacts *Source* Reference [43]

are capable of modifying the original EEG characteristics and can influence the efficiency of the detection system. The effect of artifacts on the pure EEG signal is represented in Fig. 3. In this, (a) represents the EEG signal as clean. (b)–(d) reflects ictal signals damaged by muscle artifacts, eye blinking and white noise. Finally, most of the existing seizure detection systems use small-scale datasets collected from a small number of patients and hence may not be useful in clinical applications. Therefore, to overcome these disadvantages, the automatic seizure detection systems use deep neural networks for feature extraction. Features extracted using deep learning models are more discriminative and powerful than the above methods. The following are a set of articles that use machine learning/deep learning approaches for feature extraction.

5 Deep Learning/Machine Learning Approaches

Several works in the seizure recognition literature use deep learning techniques for extracting features. Of these, the first was by Acharya et al. [41]. For seizure recognition, they used a 13-layer deep convolutional network. The first 10 layers were specifically for the extraction of features. For classification, the last three completely connected layers were used. In this research, the University Dataset of Bonn was used with a ten-fold cross-validation approach. The gain was that the separate steps for feature extraction and feature selection were not necessary. But to achieve the best results, the proposed approach needs a huge amount of training data. With an accuracy of 88.7%, this method could detect normal, preictal, and seizure groups.

Lu et al. [42] identified epileptic seizure activity using a convolutional neural network with residual blocks. The residual networks involve neglect connections that use shortcuts to traverse over some layers. The performance of the layers is avoided to eradicate the gradient problem. A fixed-size kernel slides over an input that has a fixed stride. For each kernel, this operation gives a function map. Nonlinear activation functions are used to map input and output to simulate the nonlinear behavior of nerve cells. To avoid the problem of overfitting, batch-normalization and drop-out methods are used. Then fully linked layers are used for classification with the SoftMax activation function. The Bonn dataset is used to perform a three-class classification problem which categorizes the input signal into healthy, unhealthy, and seizure types. The obtained accuracy was 99%. To classify the given signal into focal and non-focal EEG signals, they used the Bern-Barcelona dataset. In this case, the accuracy obtained was 91.8%.

Hussein et al. [43] developed a method that uses L1-penalized robust regression for feature learning. Here, rather than considering a large number of features, only a small number of very relevant features is used. The insignificant features are suppressed. This is especially good for less training data. L1 regularization essentially makes the feature vector sparse (smaller) by making most of its components zero. The remaining nonzero components are very useful. Therefore, by using this method, the most useful features associated with seizures can be identified. Here, EEG spectrum is considered and the extracted information is fed to a random forest classifier. One advantage of this method is that it performs well in ideal as well as real-life conditions. Since the used Bonn dataset did not have artifacts, models developed for the artifacts are used in this study to mimic the behavior of muscle, eye blinks, and white noise artifacts. The authors could obtain an accuracy of 100% for ideal conditions and an accuracy in the range of 90–100% for EEG data corrupted with noise, which is much better than the existing methods.

Ulla et al. [44] employed a deep learning approach focused on one-dimensional pyramidal convolutional neural network (P1CNN) ensembles. Even though the CNN model excels in the domain (time, frequency)-based methods, the primary issue is that large amounts of training data are needed. It uses a refinement technique to solve this problem such that 61% fewer parameters are involved compared to traditional CNN approaches. This method used the Bonn university dataset. To overcome the limitation associated with a lesser amount of data, the authors propose two augmentation schemes which basically generate many instances from one dataset by using a sliding window method. This is the main feature of this work. The obtained accuracy was $99.1 \pm 0.9\%$.

A seizure detection system based on signal transform and convolutional neural networks was proposed by San-Segundo et al. [45]. For feature extraction, it uses two convolutional layers and a fully linked network of three layers for feature classification. Various models used as input are the deep learning model, Fourier transforms, wavelet transform, and empirical mode decomposition. The inputs are arranged as an M × N matrix and the dimension depends on the signal transform used. The activation function rectified linear unit (ReLU) was used in all intermediate layers. The probability is reduced for the gradient vanishing problem. For two-class classification, the number of output layer has only one output and sigmoid activation function was used. For multi-class problem, the output layer contains many outputs corresponding to the number of classes and in that case SoftMax activation function was used. This work focuses on two key problems, the first of which is the detection of EEG focal and non-focal signals. The second one is the identification of normal seizures. Therefore, in this research, two sets of data were used, namely Bern Barcelona dataset and the epileptic seizure recognition dataset. The best accuracy was obtained when the Fourier transform was used to generate inputs to the deep learning model. The accuracy obtained was 99.5% for a two-class problem.

Hussein et al. [46] used a deep neural network with optimized model architecture design for seizure detection as an alteration to their previous work [43]. Initially, to learn the high-level representation of various EEG waveforms, a deep long short-term memory network (LSTM) is used. To consider the most important features associated with seizures, a fully connected layer is utilized. For classification, the extracted features are fed to a SoftMax network layer. The basic function of LSTMs is the recall of data over long periods of time. There are many advantages to this method. In the noisy and real-life circumstances, this method is efficient. It considered the presence of various artifacts that can harm an EEG signal. Since recurrent neural network (RNN) and LSTM are used, it can efficiently exploit the time dependencies in the EEG signals. This study used a dataset from Bonn University and obtained 100 percent accuracy for two-class, three-class, and five-class problems. But the pre-seizure activity was not identified by this approach.

A cost-sensitive deep learning approach for seizure detection was proposed by Chen et al. [47]. A double deep neural network was created by them. As a classifier, the first DNN is used and the second DNN is used to classify the expense of misclassification. Thus, in the unlabeled data pool, they have created a utility feature to select the most insightful samples. The authors first used a one-dimensional CNN with several filters for feature learning. The layer progresses with a fully layer having a SoftMax function in the last layer. Then, they employed a recurrent neural network (RNN) with an LSTM unit and finally an RNN with a gated recurrent unit (GRU). All three methods were implemented in the double DNN model. Bonn university data was used in this study. The results showed that the best accuracy was obtained with one-dimensional CNN (97.27%). The accuracies obtained with LSTM and GRU were 96.82% and 96.67%, respectively.

In the classification of safe, interictal, and ictal EEG signals, Zhang et al. [48] used a deep learning approach focused on the temporal convolutional neural network (TCNN). This method automatically learns features from the input data without any preprocessing. It is found that TCNN is more accurate, simpler, and clearer than recurrent neural networks such as LSTM, especially, in sequence modeling. In a temporal convolutional neural network, there exists a causal relationship between the network layers, so that there will be no missing information from the data. It mainly involves a process called causal convolution in addition to the one-dimensional fully convolutional neural network. Even though LSTM has a memory gate, it cannot remember all the data. Moreover, the model architecture of TCNN can be adjusted to any length. The proposed method used the Bonn University dataset for training and testing. They addressed fourteen different classification problems and the best accuracy of 100% was obtained with two-class problem (epileptic and non-epileptic).

Observation: Feature extraction based on deep learning approaches eliminates the need for hand-crafted features and avoids the pre-processing step associated with feature extraction and makes the whole process fully automatic. Many deep learning methods in the literature perform better in feature extraction and classification. One-dimensional convolutional neural networks and recurrent neural networks are found to perform better among all the available methods for extracting features associated with the EEG time series data (Table 1).

6 Conclusion

The objective of this proposed work is to review different methods for detecting seizures and to provide researchers in the same field with useful research guidelines. In this article, various methods for epileptic seizure detection that mainly differ in feature extraction techniques are discussed. Most of the works use temporal and spectral features or a combination of both. It is found in many cases that when the features from two domains are combined, it yields better results compared to the methods which use features from a single domain. This is because the EEG signal is nonstationary in nature and can differ between patients. The wavelet transform is a very powerful tool in EEG analysis as it can provide time–frequency localization. The selection of appropriate wavelets with suitable decomposition levels produces promising results. The EEG analysis was similarly performed by raw signal processing, extraction of features, choice of features, and classification of features. Most of the recent works use deep learning approaches for feature extraction. The deep learning approaches can remove the separate steps of feature extraction, feature selection, and can render the whole process for fully automated. Among various deep learning approaches, one-dimensional CNN and recurrent neural networks are very much appropriate for analyzing the EEG time series data. Various classifiers like support vector machine, k-nearest neighbor, artificial neural network, decision trees, random forest classifier, etc., are found to be producing very good results in seizure detection under different circumstances. But the selection of best algorithm for feature extraction and classification still need further investigations.

Table 1 Summary of various seizure detection methods

Sl. No.	Author	Year	Feature extraction method	Classifier	Performance metrics	Dataset used	Advantages/limitations
1	Hernández et al. [25]	2018	Time domain, frequency domain	Ten different classifiers	Accuracy: 94.25% (Five-class problem with SVM classifier)	Bonn University data	
2	Sriram & Raghu [29]	2017	Time domain, frequency domain, information theory, statistical based	Support vector machine (SVM)	Accuracy: 92.15% Precision: 89.74% Sensitivity: 94.56%	Bern-Barcelona database	(Sl. No. 1–6) Time domain features lack frequency information. Frequency domain features lack temporal characteristics. A combination of the two features gives better results. But major drawback is that the time frequency localization is absent
3	Samiee et al. [31]	2013	Rational discrete STFT	Naïve Bayes, logistic regression, SVM, kNN etc	Accuracy: 99.8% Precision: 99.6% Sensitivity: 99.9%	Bonn University data	
4	Wang et al. [32]	2017	Time domain, frequency domain, nonlinear analysis	Naïve Bayes, logistic regression, SVM, kNN etc	Accuracy: 99.25%	Bonn University data	
5	Tzallas et al. [33]	2009	Short-time Fourier transform, various time–frequency distributions	Artificial neural network	Accuracy: 89% (Five-class problem)	Epileptic seizure recognition dataset	
6	Gajic et al. [34]	2015	Time domain, frequency domain, nonlinear analysis	Quadratic classifier	Accuracy: 98.7%	Bonn University data	

(continued)

Table 1 (continued)

Sl. No.	Author	Year	Feature extraction method	Classifier	Performance metrics	Dataset used	Advantages/limitations
7	Miyang Li et al. [35]	2017	Wavelet transform	Neural network ensemble	Accuracy: 98.78%	Bonn University data	
8	Harpale and Bairagi[36]	2018	Pattern-adapted wavelet transform	Fuzzy classifier	Accuracy: 96.48%	CHB-MIT scalp dataset Boston	
9	Manish Sharma et al. [37]	2017	Localized orthogonal wavelet filter banks	Least squares-support vector machine	Accuracy: 94.25% Specificity: 96.56% Sensitivity: 91.95%	Bonn University data	(Sl. No. 7–12) It provides time–frequency localization and produces good results. But the signal detection in the presence of noise and artifacts makes the situation worse. The nonstationary existence of EEG signal also makes the results unpredictable. Therefore, better methods that eliminate these drawbacks are needed
10	Kumar and Kolekar [38]	2014	Discrete wavelet transform	Support vector machine	Sensitivity: 98%	Epileptic seizure recognition dataset	
11	Kaleem et al. [39]	2018	Discrete wavelet transform	SVM, Naïve Bayes, k-nearest neighbor, linear discriminant analysis, classification tree	Accuracy: 99.6% Specificity: 99.6% Sensitivity: 99.8%	CHB-MIT scalp dataset Boston	
12	Chen et al. [40]	2017	Discrete wavelet transform (Non subsampled)	k-nearest neighbor	Accuracy: 100%	Bonn University data	

(continued)

Table 1 (continued)

Sl. No.	Author	Year	Feature extraction method	Classifier	Performance metrics	Dataset used	Advantages/limitations
13	Acharya et al. [41]	2017	Deep convolutional neural network	Fully connected layer (3 layers)	Accuracy: 88.7%	Bonn University data	
14	Lu et al. [42]	2019	Convolutional neural network with residual blocks	Fully connected layers	Accuracy: 99% (Bonn dataset) Accuracy: 91.8% (Bern-Barcelona dataset)	Bonn University, Bern-Barcelona datasets	
15	Hussein et al. [43]	2018	L1 penalized robust regression	Random forest classifier	Accuracy: 100% (ideal conditions) 90–100% (real-life conditions)	Bonn University data	(Sl. No. 13–20) Feature extraction based on deep learning approaches eliminates the need of hand-crafted features and avoid the pre-processing step associated with feature extraction and makes the whole process fully automatic. One-dimensional convolutional neural networks and recurrent neural networks are found to perform better among all the available methods
16	Ulla et al. [44]	2018	Pyramidal CNN	Convolutional neural network (CNN)	Accuracy: 99.1 ± 0.9%	Bonn University data	
17	Segundo et al. [45]	2019	Convolutional layers(two)	Fully connected convolutional layers	Accuracy: 99.5% (two-class problem)	Bern Barcelona dataset and epileptic seizure recognition dataset	
18	Hussein et al. [46]	2019	Deep neural LSTM networks	SoftMax network layer	Accuracy: 100%	Bonn University data	
19	Chen et al. [47]	2018	One-dimensional CNN	Double deep neural network	Accuracy: 97.27%	Bonn University data	
20	Zhang et al. [48]	2018	Temporal convolutional neural network (TCNN)	Convolutional neural network	Accuracy: 100% (Two-class problem)	Bonn University data	

References

1. https://www.who.int/news-room/fact-sheets/detail/epilepsy
2. A.T. Tzallas, M.G. Tsipouras, D.I. Fotiadis, Automatic seizure detection basedon time-frequency analysis and artificial neural networks. Comput. Intell. Neurosci. **2007**(4), 80510 (2007)
3. A.G. I Correa, E. Laciar, H. Patiño, M. Valentinuzzi, Artifact removal from EEG signals using adaptive filters in cascade. J. Phys. Conf. Ser. **90**, 012081 (2007)
4. A.M. Chan, F.T. Sun, E.H. Boto, B.M. Wingeier, Automated seizure onset detection for accurate onset time determination in intracranial EEG. Clin. Neurophysiol. **119**, 2687–2696 (2008)
5. A. Aarabi, R. Fazel-Rezai, Y. Aghakhani, A fuzzy rule-based system for epileptic seizure detection in intracranial EEG. Clin. Neurophysiol. **120**, 1648–1657 (2009)
6. V. Srinivasan, C. Eswaran, N. Sriraam, Artificial neural network based epilepticdetection using time-domain and frequency domain features. J. Med. Syst. **29**(6), 647–660 (2005)
7. R. Meier, H. Dittrich, A. Schulze-Bonhage, A. Aertsen, Detecting epileptic seizures in long-term human EEG: a new approach to automatic online and real-time detection and classification of polymorphic seizure patterns. J. Clin. Neurophysiol. **25**, 119–131 (2008)
8. J. Mitra, J.R. Glover, P.Y. Ktonas, A.T. Kumar, A. Mukherjee, N.B. Karayiannis et al., A multistage system for the automated detection of epileptic seizures in neonatal EEG. J. Clin. Neurophysiol. **26**, 218 (2009)
9. G.R. Minasyan, J.B. Chatten, M.J. Chatten, R.N. Harner, Patient-specific early seizure detection from scalp EEG. J. Clin. Neurophysiol. **27**, 163 (2010)
10. B. Abibullaev, H.D. Seo, M.S. Kim, Epileptic spike detection using continuous wavelet transforms and artificial neural networks. Int. J.Wavelets Multiresolut. Inf. Process. **8**, 33–48 (2010)
11. A.K. Tafreshi, A.M. Nasrabadi, A.H. Omidvarnia, Epileptic seizure detection using empirical mode decomposition, in *Proceedings of the IEEE İnternational Symposium on Signal Processing and İnformation Technology*, ISSPIT, Sarajevo, December 16–19, 2008, pp. 16–19
12. L. Orosco, E. Laciar, A.G. Correa, A. Torres, J.P. Graffigna, An epileptic seizures detection algorithm based on the empirical mode decomposition of EEG, in *Proceedings of the International Conference of the IEEE EMBS*, Minneapolis, MN, USA, September, 2009, pp. 3–6
13. C. Guarnizo, E. Delgado, EEG single-channel seizure recognition using empirical mode decomposition and normalized mutual information, in *Proceedings of the IEEE İnternational Conference on Signal Processing (ICSP)*, Beijing, October 24–28, 2010, pp. 1–4
14. P.S.C. Heuberger, P.M.J. Van den Hof, B. Wahlberg, *Modelling and Identification with Rational Orthogonal Basis Functions* (Springer, London, 2005).
15. S. Fridli, L. Lócsi, F. Schipp, *Rational function systems in ECG processing the project is supported and financed by the European Social Fund* (grant agreement no. TAMOP 4.2.1./B-09/1/KMR-2010–0003)
16. A. Quintero-Rincón , M. Pereyra, C. D'Giano, M. Risk, H. Batatia, Fast statistical model-based classification of epileptic EEG signals. Biocybern. Biomed. Eng. **38**, 877–889
17. M.A. Klados, C. Papadelis, C. Braun, P.D. Bamidis, REG-ICA: a hybrid methodology combining blind source separation and regression techniques for the rejection of ocular artifacts. Biomed. Signal Process. Control **10**, 291–300 (2011)
18. D. Safieddine, A. Kachenoura, L. Albera, G. Birot, A. Karfoul, A. Pasniu, Removal of muscle artifact from EEG data: Comparison between stochastic (ICA and CCA) and deterministic (EMD and wavelet-based) approaches. EURASIP. J. Adv. Signal Process. **2012** (2012)
19. S. Casarotto, A.M. Bianchi, S. Cerutti, G.A. Chiarenza, Principal component analysis for reduction of ocular artefacts in event-related potentials of normal and dyslexic children. Clin. Neurophysiol. **115**, 609–619 (2004)
20. R. Vigário, Extraction of ocular artifacts from EEG using independent component analysis. Electroencephalogr. Clin. Neurophysiol. **103**, 395–404 (1997)
21. R. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen, E. Oja, Independent component approach to the analysis of EEG and MEG recordings. IEEE Trans. Biomed. Eng. **47**, 589–593 (2000)

22. P. He, G. Wilson, C. Russell, Removal of ocular artifacts from electro-encephalogram by adaptive filtering. Med. Biol. Eng. Comput. **42**, 407–412 (2004)
23. B. Somers, T. Francart, A. Bertrand, A generic EEG artifact removal algorithm based on the multi-channel Wiener filter. J. Neural Eng. **15** (2018)
24. D.L. Donoho, Sparse components of images and optimal atomic decompositions. Constr. Approx. **17**, 353–382 (2001)
25. D.E. Hernández, L. Trujillo, E.Z. Flores, O.M. Villanueva, O. Romo-Fewell, Detecting epilepsy in EEG signals using time, frequency and time-frequency domain features, in *Computer Science and Engineering—Theory and Applications*, *Studies in Systems, Decision and Control*, ed. by M.A. Sanchez et al. (eds.) (Springer International Publishing AG, part of Springer Nature, Berlin, 2018), p. 143
26. H. Bo, EEG analysis based on time domain properties. Electroencephalogr. Clin. Neurophysiol. **29**(3), 306–310 (1970)
27. J.M. Hausdorff, A. Lertratanakul, M.E. Cudkowicz, A.L. Peterson, D. Kaliton, A.L. Goldberger, Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. J. Appl. Physiol. **88**(6), 2045–2053 (2000)
28. P.C. Petrantonakis, L.J. Hadjileontiadis, Emotion recognition from eeg using higher order crossings. IEEE Trans. Inf. Technol. Biomed. **14**(2), 186–197 (2010)
29. N. Sriraam, S. Raghu, Classification of focal and non-focal epileptic seizures using multi-features and SVM classifier, J. Med Syst Arch. **41**(10) (2017)
30. J. Kennedy, R.C. Eberhart, Particle swarm optimization, in *Proceedings of IEEE İnternational Conference on Neural Networks*, vol IV (IEEE Service Center, Piscataway, NJ, 1995), pp. 1942–1948
31. K. Samiee , P. Kovacs, M. Gabbouj, Epileptic seizure classification of EEG time-series using rational discrete short time fourier transform. IEEE Trans. Biomed. Eng. (2013)
32. L. Wang, W. Xue, Y. Li, M. Luo, J. Huang, W. Cui, C. Huang, Automatic epileptic seizure detection in EEG Signals using multi-domain feature extraction and nonlinear analysis. Entropy (2017)
33. A.T. Tzallas, M.G. Tsipouras, D.I. Fotiadis, Epileptic seizure detection in EEGs using time–frequency analysis. IEEE Trans. Inform. Technol. Biomed. **13**(5) (2009)
34. D. Gajic, Z. Djurovic, J. Gligorijevic, S. Di Gennaro, I. Savic-Gajic, Detection of epileptiform activity in EEG signals based on time-frequency and non-linear analysis. Front. Comput. Neurosci. **9**, 38 (2015). https://doi.org/10.3389/fncom.2015.00038
35. M. Li, W. Chen, T. Zhang, Classification of epilepsy EEG signals using DWT-based envelope analysis and neural network ensemble. Biomed. Signal Process. Control **31**, 357–365 (2017) (2017)
36. V. Harpale, V. Bairagi An adaptive method for feature selection and extraction for classification of epileptic EEG signal in significant states. J. King Saud Univ. Comput. Inform. Sci. (2018)
37. M. Sharma, A. Dhere, R.B. Pachori , U. Rajendra Acharya, An automatic detection of focal EEG signals using new class of time–frequency localized orthogonal wavelet filter banks. Knowl. Based Syst. **118,** 217–227 (2017)
38. A. Kumar, M.H. Kolekar, Machine learning approach for epileptic seizure detection using wavelet analysis of EEG signals, *IEEE, 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom)*
39. M. Kaleem, A. Guergachi, S. Krishnan, Patient-specific seizure detection in long-term EEG using wavelet decomposition.Biomed. Signal Process. Control **46**, 157–165 (2018)
40. G. Chen, W. Xie , T.D. Bui , A. Krzyzak, Automatic epileptic seizure detection in EEG using nonsubsampled wavelet–fourier features. J. Med. Biol. Eng. https://doi.org/10.1007/s40846-016-0214-0
41. U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, H. Adeli, Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Comput. Biol. Med. (2017). https://doi.org/10.1016/j.compbiomed.2017.09.017
42. D. Lu, J. Triesch, Residual deep convolutional neural network for eeg signal classification in epilepsy (2019). arXiv preprint arXiv:1903.08100

43. R. Hussein, M. Elgendi , Z. Jane Wang, R.K. Ward, Robust detection of epileptic seizures based on L1-penalized robust regression of EEG signals. Expert Syst. Appl. **104**, 153–167 (2018)
44. I. Ullah, M. Hussain, E. Qazi, H. Aboalsamh, An automated system for epilepsy detection using EEG brain signals based on deep learning approach. Expert Syst. Appl. 107 (2018) 61–71
45. R. San-Segundo, M. Gil-Martín, Luis Fernando D'Haro-Enríquez, José Manuel Pardo", Classification of epileptic EEG recordings using signal transforms and convolutional neural networks". Comput. Biol. Med. **109**, 148–158 (2019)
46. R. Hussein, H. Palangi , R.K. Ward, Z. Jane Wang, Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals. Clin. Neurophysiol. **130**, 25–37 (2019)
47. X. Chen, J. Ji, T. Ji, P. Li, Cost-sensitive deep active learning for epileptic seizure detection, in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (2018), pp. 226–235
48. Zhang, J., Wu, H., Su, W., Wang, X., Yang, M., Wu, J., A new approach for classification of epilepsy eeg signals based on temporal convolutional neural networks, in *2018 11th International Symposium on Computational Intelligence and Design (ISCID)*, vol. 2 (IEEE, 2018), pp. 80–84

Measurement of Physiological Parameters Using Video Processing

M. Spandana, Pavan Arun Deshpannde, Kashinath Biradar, B. S. Surekha, and B. S. Renuka

Abstract This paper proposes a non intrusive and contactless method for estimation of physiological boundaries such as Heart Rate (HR), Inter Beat Interval (IBI) and Respiration Rate (RR), these are imperative markers of patients physiological state and essential to screen. Be that as it may, the majority of the measurements methods are association based, for example sensors are associated with the body which is regularly confused and requires individual assistance. The proposed strategy recommends a straightforward, ease and non-contact approach for estimating different physiological boundaries using a web camera continuously. Here, the breath rate and heartrate are gotten with the assistance of facial skin shading variation caused by body blood course. Signal preparing techniques such as autonomous segment examination (ICA), Fast FourierTransform (FFT), and Principal segment investigation (PCA) have been applied on the shading diverts in video accounts and the blood volume beat (BVP) is removed from the facial regions. The elements to be measured and looked at are IBI, HR, and RR. A decent understanding was accomplished between the measurements across every single physiological boundary. The proposed strategy has significant potential for progressing telemedicine and personal health care.

Keywords Principal segment investigation · Independent component analysis · Fast fourier transform · Eulerian magnification · Inter beat interval · Blood volume beat · Heart rate · Respiratory rate

1 Introduction

The human physiological signs, for instance, breath rate, beat, beat whim partner degreed blood O immersion and its checking acknowledges a work in end of flourishing conditions and irregular functions, Routinely, sensors, anodes, leads, wires and chest ties square measure utilized for seeing of cardiorespiratory development,

---

M. Spandana (✉) · P. Arun Deshpannde · K. Biradar · B. S. Surekha · B. S. Renuka
JSS Science and Technology University, Mysore, India
e-mail: renuka@sjce.ac.in

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_42

which can cause nervousness and oblige the patient at whatever point utilized for wide stretches of time. Focal points in diminishing the issues related with contact seeing frameworks has impelled investigation exploitation clear elective assessment ways for checking of physiological signs, for instance, the usage of the Photoplethysmography (PPG), Eularian video improvement and Independent parcel assessment [1].

At the regular heart cycle, meter changes inside the facial veins change the way length of the encompassing lightweight incidented with the tip objective that the resulting changes in live of reflected lightweight exhibit the plan of vehicle diovascular events. This facial video is considered as partner degree data and PPG signals are removed from the got video and independent component Analysis is utilized for extraction of Red, green and blue signals and winnowing methodology square measure utilized for isolating the sign and after sign is free and demonstrated [2].

The people rate is measurable by counting the measure of breath severy second through checking anyway as a rule the chest rises. Breath rates may increment with fever, infection, or option clinical conditions [3]. Respiratory rate ends up being a basic pointer of ailment. Eulerian Video Magnification (EVM) is partner degree stimulating video preparing approach that reveal simple breath developments inside video groupings [4]. It for the premier half contains of spatial decomposition,phase comparison, signal smoothing and prime acknowledgment

2 Literature Survey

The paper [4] they stone-broke down regarding applied mathematics info from thetraffic division shows that physiological and mental elements causes the amount the amount accidents. To conquer that they projected the procedure that contains A digital camera that's utilised to quantify the 60 min of drivers, therefore on monitor their actual exercises in real time. The 60 min is gotten by catching the colour varieties materializing attributable to blood course in facial skin. Blind supply, supported by the RGB shading amendment embrace in video following, is utilized for info extraction. Contrasted and existing business location prepare ment, the non-contact identification strategy will helpfully quantify the 60 min with acceptable preciseness, significantly throughout driving circumstances that need further pre-alert.

The this paper [5] they dicused regarding pulse variation, it's causes, effects and moniter. Heart rate fluctuation is aproportion of types between each heartbeat that demonstrates the impacts of weight on a human body. With the ascent of unfortunate dietary patterns and inactive ways in which of life over the globe. To screen the heart beat selection they develoved a model that catches video utilizing the camera, were isolated into Red, inexperienced and Blue (RGB) shading channels that, were then modified over to (HSI) shading model. Cheeks were chosen because the district vital to that strategy (BBHE) was applied. Application of Principal section investigation (PCA) on the 3 shading channels separated new head elements. Using these ways they live the heart beat to screen pulse selection.

The paper [6] they counseled regarding the importance of remote sensing element in heath care to screen pulse then on and there applications. To screen they projected a framework, that framework utilizes wearable remote multimodal fix sensors, planned utilizing off-the-shelf components. These wearable sensors utilize a low-power nine-pivot mechanical phenomenon estimation unit to live the metabolic process development moreover, a MEMs electronic equipment to record sound signals. Data handling moreover, combination calculations square measure utilised to work out the metabolic process return and also the hacking events. This framework utilizes wearable remote multimodal fix sensors, planned utilizing off-the-shelf elements. These wearable sensors utilize a low-power nine-hub mechanical phenomenon estimation unit to judge the metabolic process development additionally, a MEMs mouthpiece to record sound signs. Info getting ready what is additional, combination calculations square measure utilised to work out the metabolic process return and also the hacking functions.

The paper [7] They documented awkward of convolution ways, for instance, for estimating heart rate there square measure distinctive commonplace ways accessible, for instance, cardiogram that is expensive and discomforts. Another business widget is oximetry sensing element that wants association to fingertips, is likewise inconvenient. They documented higher approach for pulse mensuration. They saw that in the guts cycle, because of volumetrical changes within the facial veins, the manner length of the incidence close lightweight is changed. And the circumstance of was functions are|ar|area unit|square live incontestable as a results of the following changes within the measure of mirrored lightweight. They have utilised sign partition ways like freelance section analysis(ICA) to isolate inexperienced shading band from red, blue and inexperienced band. After analytic inexperienced band they figure repose Beat Interval to work out the heart beat.

The paper [8] they describe their observation on deceases relating to heart and respiration. To monitor these physiological parameters projected technique contains RGB band separation from video captured, selecting inexperienced band for pulse rate and red and blue for rate mensuration and for hard O containt within the blood they used optical sensing element placed at tip.

The paper [9] they processed concerning would like of contact less estimation of physiological parameters. Heart rate could be a basic indispensable sign for clinical diagnosing. There is developing revenue in extricating it while not contact, particularly for populaces, as an example, untimely kids and therefore the recent for whom the skin is delicate and damageable by customary sensors. Hence they planned strategy to measure pulse through video handling technique. According to their technique, they take away pulse and beat lengths from recordings by estimating invisible head movement led to by the Newtonian response to the deluge of blood at every beat. Method tracks highlights on the pinnacle and performs head half examination (PCA) to decay their directions into a bunch of phase motions. It at that time picks the phase that best relates to pulses passionate about its transient repetition spectrum. Finally, investigate the movement extended to the present phase and distinguish pinnacles of the directions, that compare to pulses.

The paper [10] As per their summary Human visual framework has restricted spatio-worldly sensitivity, but various signs that fall below this limit are often informative, example human skin tone differs marginally with blood circulation. This variety, whereas unperceivable to the unaided eye, are often abused to separate heartbeat rate by amplifying caught video. They have used Eulerian video magnification(EVM)method. Eulerian Video Magnification, takes a typical video arrangement as data, and applies spacial disintegration, trailed by transient separation to the frames. The coming concerning sign is then increased to uncover shrouded information.

The paper [11] the foremost half focuses on home eudaimonia and distant heath observant of crucial signs incorporates the high exactitude symptomatic gadgets additionally as simple ones and open for everybody. In this paper they planned basic and powerful strategy for estimating the beat rate. The technique contains for every rous' channel, pixels esteems were extra severally for every frame. The signals noninheritable alongthese lines were sifted utilizing a FIR bandpass filter, Next autonomous and head half examinations were performed. After that boundaries square measure removed

The paper [12] creators conceivably introduced precise estimation and observant of physiological boundaries, as an example, internal heat level, pulse, metabolic process patterns. Among them their elementary focus concerning rate measuring. According to planned technique, the video is gathered at a collection edge pace of thirty rate, that is sufficient to discretize the respiration movements. After recording, the pel at the degree of pit of the neck within the main fringe of the video. The respiration rate are often far from either in repetition or time domain. Obtained rate are often used to determine hospitalization, and it offers the open door for early intervention. Moreover, the rate has been discovered to be a lot of detrimental boundary among steady and shaky patients.

Checking of human physiological signs, as an example, pulse, pulse inconstancy, breath rate, and blood chemical element immersion assumes a operate finally of medical issue what is a lot of, uncommon functions, as an example, cardiac arrhythmia, arrhythmia, bradypnoea, tachypnoea, apnoea, and hypoxemia. Conventionally, checking of cardio-respiratory action is accomplished by utilizing glue sensors, terminals, drives, wires what is a lot of, chest ties which can cause distress and compel the patient whenever used for in depth stretches of your time. In addition, these sensors might cause skin damage, contamination or unfavorable responses on people with delicate skin. In this paper [13] creators used Christian Johann Doppler measuring system result and Thermal imaging strategy to quantify and screen human physiological signs. The Doppler shift is associate degree estimation procedure equipped for distinctive invisible development of the chest divider materializing thanks to mechanical action of the center and lungs later on uncovering the cardio-respiratory signal. Thermal imaging is associate degree inactive estimation procedure, which can be used to spot the created radiation from specific areas of the form within the infrared (IR) scope of the magnetism vary to free the cardio-respiratory sign.

3 Methodology

This section discusses the steps methodologies involved in Measurement of Heart Rate using Video Processing.

3.1 Methodology Involved in Measurement of Heart Rate Using Video Processing

This section contains basic building blocks which represent methods involved in the heart rate measurement and respiratory rate system. Figure 1 represents building blocks of heart rate measurement system. It shows various techniques used in different stages of the implementation.

The system consists of three main modules

1. **Data Collection**
   - **VideoCapturing**
     For video catching Open Computer vision library (Open CV) is used to consequently distinguish the directions of the face area in the primary casing of video recording.
   - **ROI Selection**
     Initially, the rectangle shape face region is chosen at the first edge of video recording. Selected face locale facilitators continued as before for entire succession of images. The ROI is a rectangle shaped part of forehead area.
   - **Extracting PPG signal**
     Separate the red, green, blue raw traces from the selected ROI and normalize them.
2. **Data Pre-processing**
   - **Signal Separation**
     The Independent Component Analysis is a computational and factual procedure utilized to separate free signals from blend of signals. Assume that we have n blends $x_1, \ldots, x_n$ of n independent segments.

$$X_j = a_{j_1} * s_1 + a_{j_2} * s_2 + \cdots + a_{j_n} * s_n \tag{1}$$

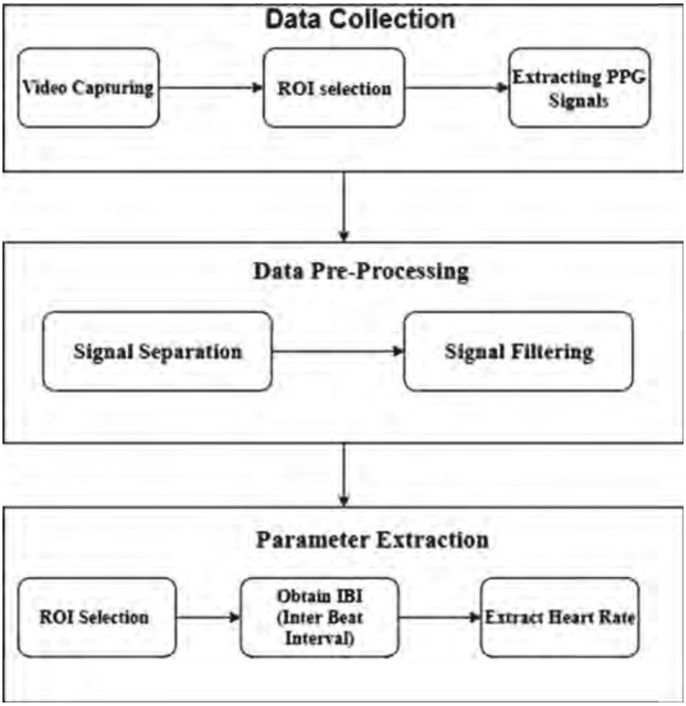


Fig. 1 Block diagram for heart rate measurement

The time list t has dropped in ICA model, since according to expect singular parts are random factors rather than an ideal time signals. Thus the watched qualities $x_j(t)$, for example the microphone signals in the mixed drink party issue, are then a sample/acknowledgment of this irregular variable. The condition can be communicated utilizing vector-matrix notation.
The equation can be expressed using vector-matrix notation.

$$X = A * S \tag{2}$$

$$X = \begin{bmatrix} X1 \\ . \\ . \\ X_n \end{bmatrix} S = \begin{bmatrix} S1 \\ . \\ . \\ S_n \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} a_{11} & \cdot\cdot & a_1 n \\ . & ... & . \\ . & ... & . \\ a_{n1} & \cdot\cdot & a_{nn} \end{bmatrix} \tag{3}$$

X random vector whose parts square measure the mixtures $x_1 \ldots x_n$
S random vector whose parts square measure the sources $s_1 \ldots s_n$
A mixing matrix with elements a_{ij}

- **Signal Filtering**
  For each casing pixels esteems are included independently for each ROI's channel. From them just green signs are thought of and apply Fast Fourier Trans-structure (FFT)filtering calculation.

3. **Parameter Extraction**

- **Obtain Inter Beat Interval (IBI)**
  The pinnacles esteems are found from green flag and compute a normal of the mean of sign and most extreme power of the signal. Then various pinnacles which is more prominent than this edge are calculated. Then pulse can get and show.

3.2 Methodology Involved in Measurement of Respiratory Rate Using Video Processing Measurement

Figure 2 represents the block diagram of respiratory rate measurement.

The Respiratory Rate System Consists of

1. **Data Acquisition**
   Two kinds of information are acquired. Both video and the breath signals are gotten during a polysomnography. RGB values are changed over into Gray scale worth and afterward choosing the ROI that is chest and mid-region area. Here buffer is created for storing the image pyramid captured during data acquisition. Management of this buffer helps to handle the overhead problem.
2. **Data Processing**
   The recorded video information is prepared in various steps. First step is, explicit movements are intensified utilizing Eulerian video magnification. The developments are separated utilizing optical stream calculation from the video information. The yield from the optical stream is then adjusted so as to acquire a sign of which the quality can be surveyed in examination with the control signals.
3. **Signal Analysis**
   By including the optical stream estimations of each edge we can get recurrence substance of the sign and it is examined by fourier change. Optical stream calculation contains a few presumptions:

   - An object pixel intensities do not change between consecutive frames.
   - Neighbouring pixels have similar motion.

   we should Consider a pixel I(x,y,t) in first frame. In the following casing it moves by separation (dx, dy) taken after dt time. Since pixels are the equivalent and force doesn't transform, we can scientifically say that,

$$I(x, y, t) = I(x + \mathrm{d}x, y + \mathrm{d}y, t + \mathrm{d}t) \tag{4}$$

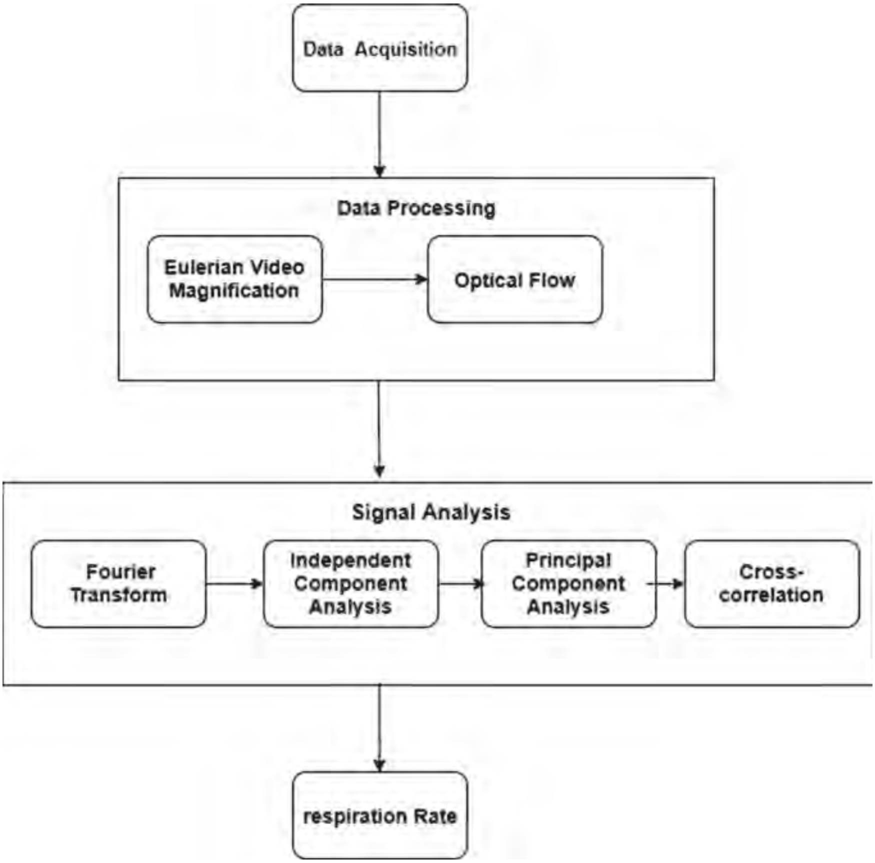


Fig. 2 Block diagram for respiratory rate measurement

Then apply taylor series for approximation of right-hand side, to get the following equation

$$f_x u + f_y v + f_t = 0 \tag{5}$$

$$f_x = \frac{\partial f}{\partial x}; \quad f_y = \frac{\partial f}{\partial y} \tag{6}$$

$$u = \frac{\mathrm{d}x}{\mathrm{d}t}; \quad y = \frac{\mathrm{d}y}{\mathrm{d}t} \tag{7}$$

Above condition is called Optical Flow equation. It is utilized to discover fx and fy, they are picture angles. What's more, *ft* is the angle along time. Be that as it may, (u, v) is unknown. Solving two obscure factors utilizing this one condition is preposterous. So there are a few techniques to take care of this issue and we utilized Lucas–Kanade strategy.

Lucas-Kanade strategy takes a 3×3 sub-part around the pixel. So all the 9 focuses have the equivalent motion. For these 9 focuses we have find (fx, fy, ft). So now our concern becomes explaining 9 conditions with two obscure factors which is over-determined with least square fit technique better arrangement can be acquired. The following is the last arrangement which is two condition two obscure issue and illuminate to get the arrangement.

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \sum_i {f_{x_i}}^2 & \sum_i f_{x_i} f_{y_i} \\ \sum_i f_{x_i} f_{y_i} & \sum_i {f_{y_i}}^2 \end{pmatrix}^{-1} \begin{pmatrix} -\sum_i f_{x_i} f_{t_i} \\ -\sum_i f_{y_i} f_{t_i} \end{pmatrix} \tag{8}$$

In the basic way the thought is to follow some given focuses and get the optical stream vectors of those focuses. Be that as it may, this will be fine Until managing little motions, it will bombs when there is enormous motion. To defeated this difficult we have utilized pyramid method. In the pyramid, little movements are expelled when cross up and huge movements turns out to be little movements. Utilizing Lucas–Kanade strategy, we can get optical stream alongside the scale. These signs can have a low sign to noise proportion, making breath recognition difficult. Independent part examination (ICA) and head segment examination (PCA) are performed to improve the sign quality.

4. **Extract Respiratory Rate**

   The last yield signal is the normal stretch inside the estimation buffer. Using **peak detection** technique resultant sign is acquired.

   Top identification calculation follows as, using picture histogram a pinnacle location signal is produced. At that point, utilizing the zero-intersections of the pinnacle recognition signal and the nearby extrema between the zero-intersections, the histogram tops are found.

   The pinnacle identification signal, rN, is produced by convolving the histogram h with the pinnacle identification part pN, i.e.,

$$r_N = p_N \circledast h \tag{9}$$

4 Results

This section deals with results part of the executed project. It includes the results obtained during every step of heart rate and respiratory rate measurement.

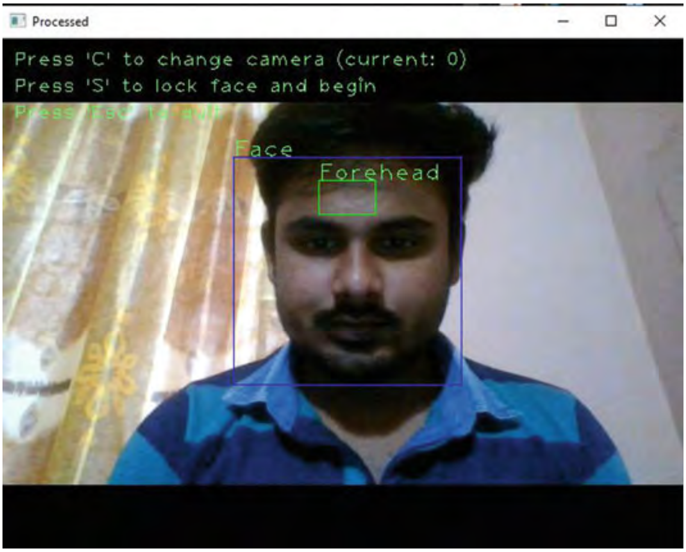


Fig. 3 Initial application UI

4.1 Heart Rate Measurement Results

Figure 3 shows initial application UI, which contains key handler menu, and two rectangular boxes to detect face and forehead.

The initial application UI contains key handler menu, which contains different key options for different functionalities. To measure heat rate, for that first should select region of interest and should lock it. To do so we need to press key "S". Figure 4 shows window selecting and locking of ROI.

The ROI window contains green channel band segregated from ROI region and also key handler menu. Key handler menu consists of option like restart, display data and exit option.

After locking region of interest next to plot or display then we need press key "D". Figure 5 shows display window which display hear rate in beat per minute(bpm).

Figure 6 shows the command prompt output which displays the tag results in each steps of the heart rate measurement application.

4.2 Respiratory Rate Measurement

The respiratory rate measurement application starts by executing python file. Figure 7 shows initial application widow contains video capture frame, raw signal display plot, and processed signal plot.

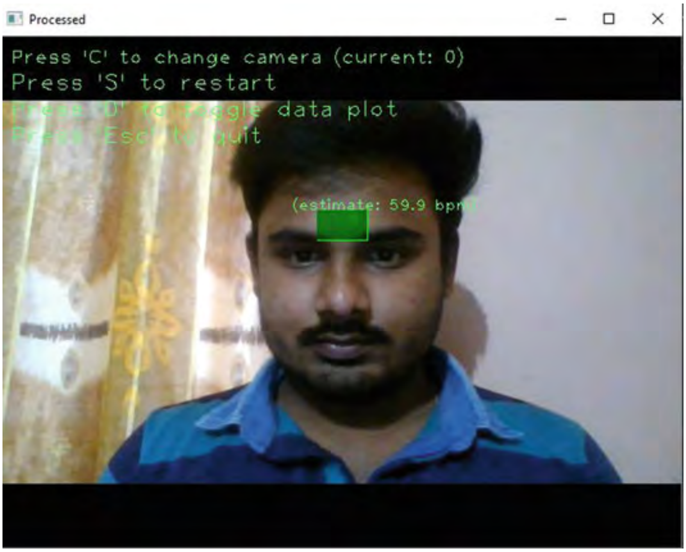


Fig. 4 Window showing locking of ROI

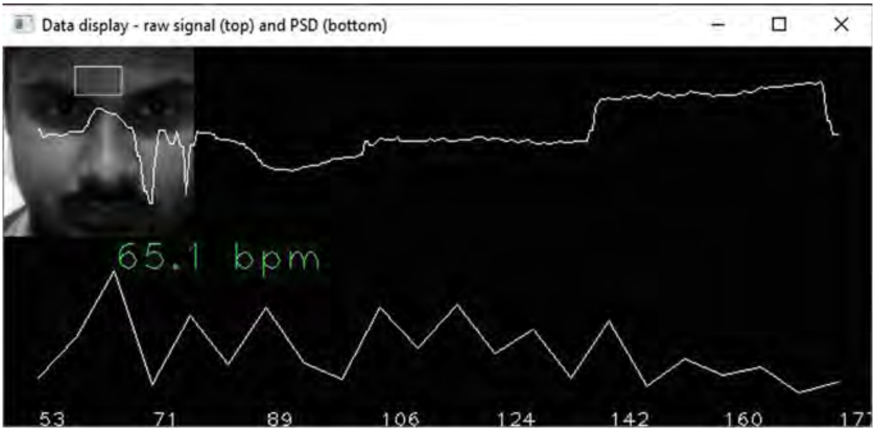


Fig. 5 Display window

After initialization next step will be calibration step to identify the ROI using image pyramid method and then measurement process will start. Figure 8 shows the measurement window, which contains the raw signal display, processed signal plot and digital real time respiratory rate display.

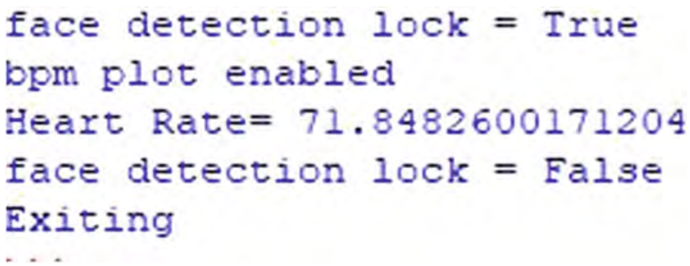


Fig. 6 Command prompt output

Fig. 7 Initial application window

5 Conclusion

Based on the results, it will be terminated that the mensuration of vital sign and rate will be performed victimization the planned non-invasive and cheap methodology. Conventionally, observance of cardio-respiratory activity is achieved by victimization adhesive sensors, electrodes, leads, wires and chest straps which can cause discomfort and constrain the patient if used for long periods of your time contactless physiological parameter. It provides easy, user friendly and harmless mensuration methodology that each Individual needs.

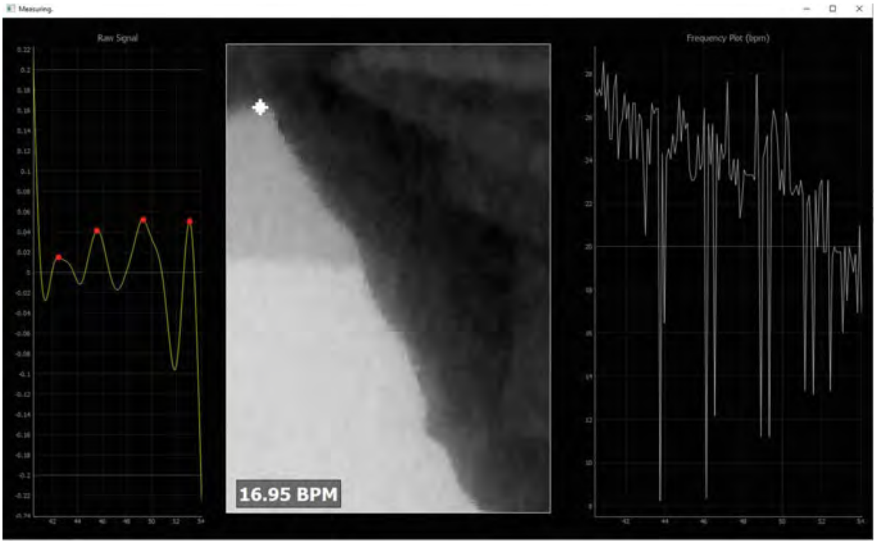


Fig. 8 Measurement window

5.1 Applications

(a) Heart rate and respiratory rate can be used to diagnose the cardiovascular and respiratory disease.
(b) The proposed method can be used for monitoring the status of the driver such as drowsiness and the mental stability of the driver.
(c) The proposed method can also be used to monitor the status of infants without any wearable devices.
(d) The proposed method is helpful in remote areas for immediate health monitoring.

5.2 Future Scope

The goals of this project were purposely kept within what was believed to be attainable within the allotted timeline. As such, many improvements can be made upon this initial design. For future works, the results of the heart rate and respiratory rate can be utilised as the basic criteria for detection of the brady cardia, trachy cardia and many other respiratory diseases. The techniques such as PPG and Eularien Magnification can also be used to measure the body temperature and oxy-haemoglobin and carboxy-haemoglobin content in the blood based on the movements of the blood in the certain region.

The obtained results can be inerfaced with the android device so that the proposed method is accessible by everyone. Further the android application can provide real-time doctor support for the patient who are using. The results of the patient will be accessible by the doctor in real-time. For this there is a need for application developed for doctors. This application will display all the measurement details of the patients. The doctor can also get alerts about serious patients who require instant medical care. Further the app can have a real-time chat with patients personal doctor.

The other pshyological parameters like body temperature and SpO2 can also be mmeasured and used for diagnosis.

References

1. S.L. Bennett, R. Goubran, F. Knoefel, The detection of breathing behavior using Eulerian-enhanced thermal video, in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (2015), pp. 7474–7477
2. Y.D. Patil, G.J. Chhajed, Heart rate measurement system using facial video processing. Int. J. Emerg. Trends in Technol. (IJETT) **4** (2017) (Special Issue)
3. S. Alam, S.P.N. Singh, U. Abeyratne, Considerations of handheld respiratory rate estimation via a stabilized video magnification approach, in *IEEE-EMBS International Conference on Biomedical and Health Informatics*, NV, USA, Las Vegas (2016), pp. 41–44
4. Z. Qi Zhang, G. Xu, W. Ming, Y. Zhou, W. Feng, Webcam based non-contact real-time monitoring for the physiological parameters of drivers, in *The 4th Annual IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems*, China, Hong Kong, June 4–7, 2016
5. G. Uppal, P. Kalra, N.R. Prakash, Heart rate measurement using facial videos. IEEE 0973-6107 **10**(8), 2343–2357 (2017)
6. T. Elfaramawy, C.L. Fall, S. Arab, M. Morissette, F. Lellouche, B. Gosselin, A wireless respiratory monitoring system using a wearable patch sensor network. IEEE Sens. J. **19**(2) (2019)
7. S.M. Kshirsagar, G.J. Chhajed, Heart rate, respiration rate and oxygen saturation level monitoring from video recording by smart phone camera. Int. J. Eng. Develop. Res. (IJEDR1402254) **2**(2) (2014)
8. G. Balakrishnan, F. Durand, J. Guttag, Detecting pulse from head motions in video, in *MIT CSAIL* (2016)
9. H. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, W. Freeman, Eulerian video magnification for revealing subtle changes in the world, in *Quanta Research Cambridge and MIT CSAIL* (2015)
10. M. Lewandowska, J. Rumiński, T. Kocejko Jędrzej Nowak, Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity, in *IEEE Proceedings of the Federated Conference on Computer Science and Information Systems* (2011)
11. C. Massaroni, D. Simões Lopes , D. Lo Presti, E. Schena, S. Silvestri, Contactless monitoring of breathing patterns and respiratory rate at the pit of the neck: a single camera approach. J. Sens. **2018**, ID 4567213 (2018)
12. A. al-naji, K. Gibson , S. Lee, J. Chahl, Monitoring of cardiorespiratory signal: principles of remote measurements and review of methods, in *IEEE Translations* (2017)
13. H. Fitriyah, A. Rachmadi, G.E. Setyawan, Automatic measurement of human body temperature on thermal image using knowledge-based criteria. J. Inform. Technol. Comput. Sci. **2**(2) (2017)

Low-Dose Imaging: Prediction of Projections in Sinogram Space

Bhagya Sunag and Shrinivas Desai

Abstract Computed tomography (CT) is one of the preferred medical diagnostic tools according to the medical survey. In CT, X-ray projections are acquired from different view angles to generate the tomographic images of the body. The current study has been evident that there is an adverse effect on health due to the excess radiation exposure. In this context, low-dose imaging is becoming as a clinical reality. Low-dose image is achieved by a sparse-view CT and usually possess complex artifacts and noise. Reconstructing the high-quality images from low dose is always remaining as a challenging task. To address this issue, the simple-averaging method is presented to estimate the missing projection data in the sinogram space. Reconstructed image quality is assessed by using the parameters such as PSNR, RMSE, and SSIM. Experimental results show that the proposed technique improves the image quality as compared to the conventional low-dose image.

Keywords Computed tomography · Low-dose · Sparse-view

1 Introduction

Medical imaging technology is rapidly growing over the past 30 years. It uses imaging modalities and processes such as CT, MRI, ultrasound, and X-ray to get more detailed information on the human body, which helps the doctor to diagnose and treat the patients effectively.

B. Sunag (✉) · S. Desai
School of Computer Science & Engineering, KLE Technological University, Hubli, Karnataka 580031, India
e-mail: sunagbhagya@kletech.ac.in

S. Desai
e-mail: sd_desai@kletech.ac.in

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*, Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_43

The usage of CT has increased rapidly all over the world in the last two decades. Image reconstruction in CT is a mathematical process where X-ray projections from different angles are taken to generate the tomographic images of the human body for proper diagnosis of disease. From sufficient projection angles, a high-quality image can be acquired. However, repeated scan or high radiation dose can damage body cells, including DNA molecules, which may lead to radiation-induced cancer. Therefore, the radiation dose is one of the critical issues in the medical field.

Two promising schemes introduced to reduce the radiation dose [1]. One is low-dose CT, which is to change the hardware condition to lessen the tube current [2]. Another is sparse-view CT, which is to reduce the number of projections which will produce the complex artifacts and noise in the reconstructed image. Therefore, to improve the image quality due to sparse view, several conventional algorithms have been proposed. They are analytical and iterative reconstruction algorithm [3].

The proposed methodology uses an analytical filter back projection (FBP) algorithm to reconstruct a sparse-view CT image called a low-dose image. The main objective of the experiment is to generate a high-quality image from low dose. Therefore, to realize this, missing projection data due to sparse-view in sinogram space are filled with simple averaging of neighboring projection values.

The paper is organized as follows: In Sect. 2, mathematical background along with its implementation details is discussed, and in Sect. 3, sparse-view reconstruction is discussed. In Sect. 4, the proposed method performance is compared and analyzed for sparse-view projection, respectively. Lastly, in Sect. 5, the conclusion and future work are given.

2 Literature Survey

This section includes a survey on state-of-the-art techniques to improve the low-dose CT image quality for medical diagnostics.

In the paper [5], authors have worked on 197 trauma patients of young age group to evaluate abdominal organ injuries and graded them to ASST scales. They compared the image quality and noise ratio of low-dose CT using ASIR-V and FBP with routine dose CT. ASIR-V with significant reduction in radiation dose performs better in assessing the multi-organ abdominal injury without harming image quality.

ASIR-V hybrid iterative algorithm [6, 7] is used to study the human lung specimen and phantom for lesion detection, image noise, resolution, and dose reduction potential. The ASIR-V with low radiation dose minimizes the image noise and improves the image quality but did not considerably influence airway quantification values, while variation in measurements such as %WA and WT slightly increased with dose reduction.

In [8], authors have considered 59 children to conduct experiment. MBIR is an iterative algorithm that has a better performance compared to ASIR in improving the image quality of contrast dose abdominal and low radiation dose CT to meet the diagnostic requirements.

In [9], 21 patients brain 3D-CTA axial and volume-rendered (VR) images were reconstructed from the 3D-CTA data using ASIR and MBIR. MBIR provides better visibility of small intracranial arteries such as AchoA and TPA without increasing the radiation dose.

In [10, 11], the authors have compared deep learning reconstruction (DLR) images with state-of-the-art techniques reconstructed images for phantom, submillisievert chest, and abdominopelvic CT in improving image quality, noise, and detectability.

In recent literature, some of the authors worked on the reconstruction of sparse-view CT images using DenseNet and deconvolution [12], GoogleNet [13], and convolution neural network [14, 15].

Most of the literatures have not explored sinogram space for addressing low-dose imaging; hence based on this research gap, our proposed methodology is designed.

3 Mathematical Background

3.1 Radon Transform

Radon and inverse radon transform mechanisms are used in the reconstruction of CT images from projections. Radon transform produces a line integral from object $f(x, y)$ as a projection shown in Fig. 1. The projection of a 2D object at an angle is one dimensional. Series of 1D projection at different angles are stacked together to form a sinogram. The radon transforms mathematical equation is given below [4].

In CT, 1D projection of an object $f(x, y)$ at an angle is given by,

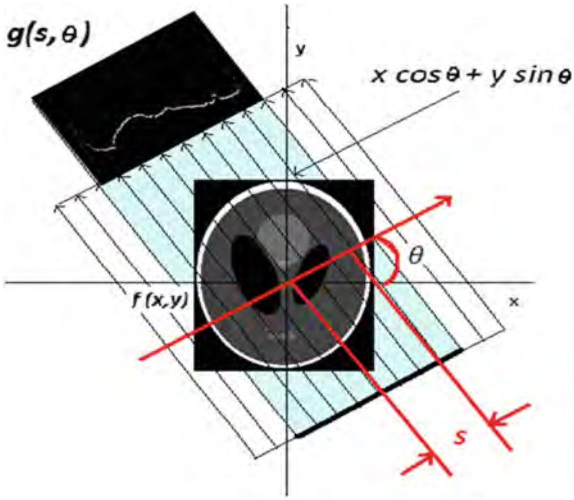


Fig. 1 Radon transform maps f on the (x, y)-domain into Rf on the (α, s)-domain

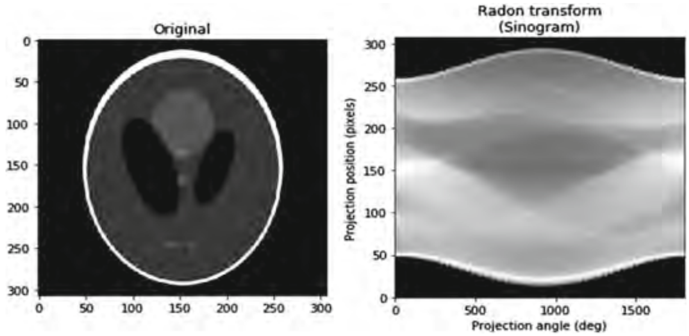


Fig. 2 Radon transform (sinogram *p*)

$$g(s,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x\sin\theta + y\cos\theta - r)\mathrm{d}x\mathrm{d}y \tag{1}$$

where θ is the angle of the line and s is the perpendicular offset of the line. The gathering of $g\ (s, \theta)$ at different angles of θ is called the radon transform of an image $f\ (x, y)$.

Figure 2 presents the sinogram p from the image f using the radon function.

$$p = \mathrm{radon}(f, \theta)$$

The inverse radon transform is used in CT to reconstruct a 2D image from a sinogram.

Reconstructed image $f'(x, y)$ is given by,

$$f'(x,y) = B\{g(r,\theta)\} = \int_0^{\pi} g(x\cos\theta + y\sin\theta, \theta)\mathrm{d}\theta \tag{2}$$

Figure 3 presents the reconstructed image f from projection data using the iradon function.

$$\mathrm{i} = \mathrm{iradon}(p, \theta)$$

3.2 *Filtered Back Projection*

Filter back projection is an analytical reconstruction algorithm, which is one of the fasted methods to perform inverse radon transform. Hence, it is widely used in CT to

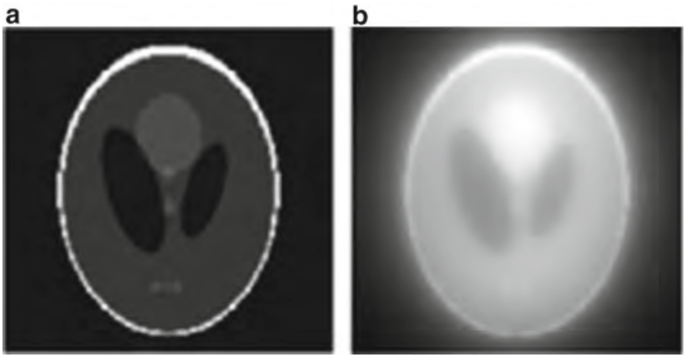


Fig. 3 **a** Original image f, **b** unfiltered reconstructed image

overcome the limitations of back projection. The only tunable parameter in FBP is a filter and is used to remove blurring (artifacts) in the reconstructed image. Using a ramp filter is the optimal way to suppress complex artifacts. So, FBP is a combination of ramp filtering and back projection.

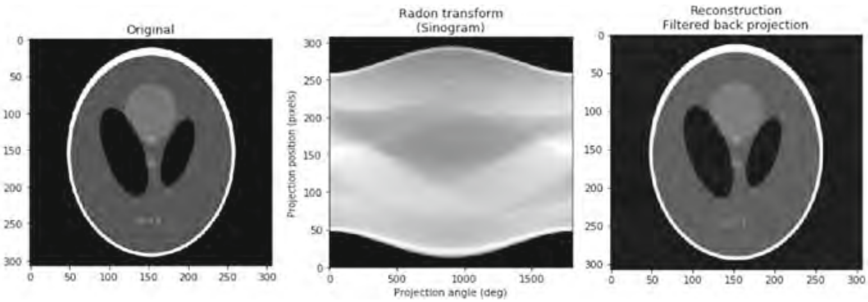
The ramp filter is given by,

$$H\left[\hat{h}\right](w) = [w]\widehat{h}(w) \tag{3}$$

Applying the projection-slice theorem changing the integration variables, it can be observed for f,

$$f = \frac{1}{4\pi} R * H[p] \tag{4}$$

This means that the original image f can be reconstructed from sinogram p as shown in Fig. 4.



(a) Original image (f) (b) Sinogram (p) (c) Reconstructed Image

Fig. 4 Image reconstruction from projections

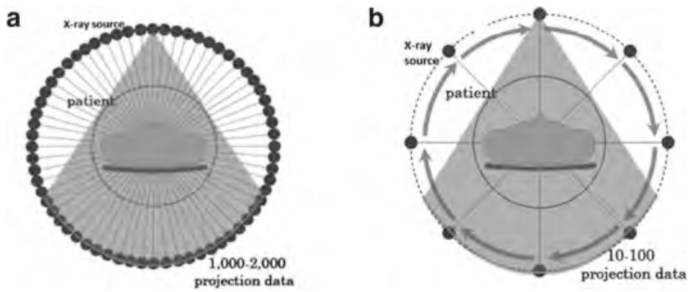


Fig. 5 **a** Conventional CT, **b** sparse-view CT

3.3 Sparse-View CT

In the conventional CT, the X-ray source positions are distributed uniformly over the angular range of 0°–360°, and 1000–2000 projection data are measured. In the sparse-view CT, projections $g(f, \theta)$ are known for only a small set of increment angles θ distributed over 0–π as shown in Fig. 5b. The number of X-ray projections reduced to less than 100. By implementing a sparse view, there is a substantial reduction in the radiation dosage, and the images generated are hence called low-dose images. They will be suffering from complex artifacts and noise but lower radiation risk on the human body.

3.4 Sparse-View CT Reconstruction with FBP

The experiment is carried out by varying the number of views. The objective is to analyze how the image quality is affected by drastically reducing the number of views. The proposed research work have considered 0–360° with an increment of 1° (360 views) as complete data and 0–360° with an increment of 2°, 4°, 6°, and 8° (i.e., 180, 90, 45, 22 views) as incomplete data. In Fig. 6, it can be observed that as the radiation dose is reduced, the quality of the reconstructed image is degraded. It suffers from artifacts and noise. These images are called low-dose images.

4 Proposed Methodology

4.1 System Model

The Shepp–Logan phantom head model is used as an original image. The original image is subjected to radon transform for 0°–360° with an increment of 2°, 4°, 6°, and 8° separately. Due to a sparse view, data acquired at sparser angles. For example,

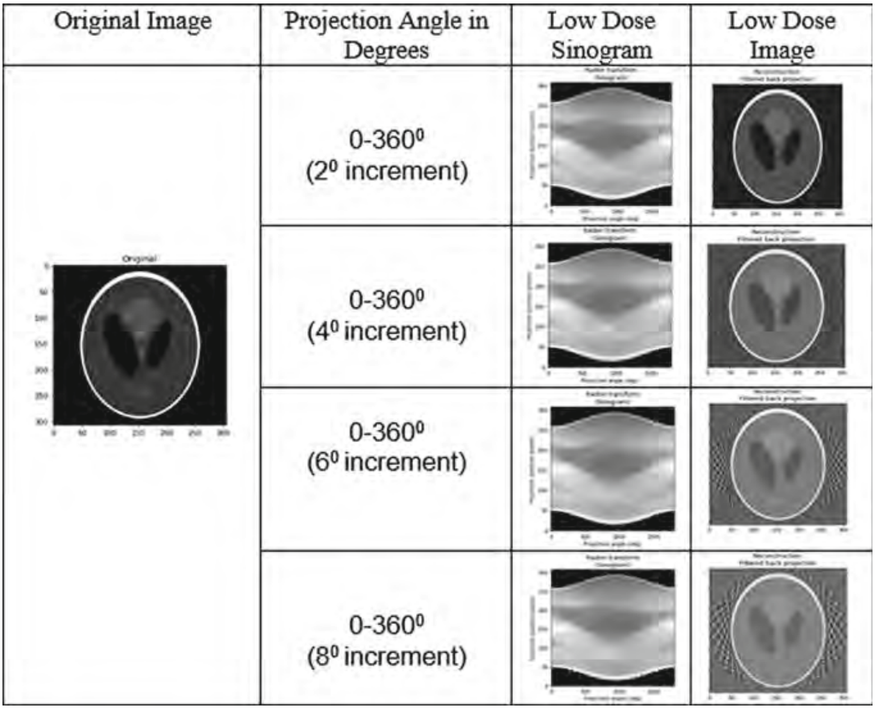


Fig. 6 Reconstruction of image from sparse projections

with 4° increment, angle projection data is acquired at 0°, 4°, 8°, 12° …, and the intermediate angle data are not acquired. An effort is made to predict projections at these intermediate angles using a simple-averaging method. The detailed process is explained in the next section. After predicting the missing intermediate projection values, the manipulated sinogram is transformed into an image domain using filtered back projection, which is inverse radon. As a result, the better quality image is reconstructed with minimal artifacts and noise as compared to the image reconstructed with a 4° increment angle without manipulation. Figure 7 represents the proposed model for image reconstruction.

4.2 Filling the Missing Projection Data in Sinogram Space

The missing projection data in sinogram space due to sparse view is filled with a NULL (zero (0)) value. In the proposed method, only the alternate degrees between the sparse angles are filled with value 0. For example, 2° is the alternate degree between sparse angle 0° and 4° filled with 0. Figure 8 presents the missing projection value in sinogram space filled with 0.

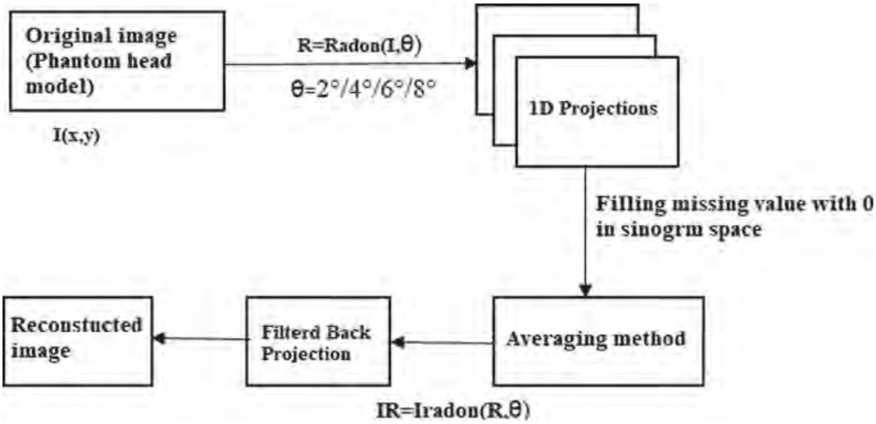


Fig. 7 Image reconstruction process model

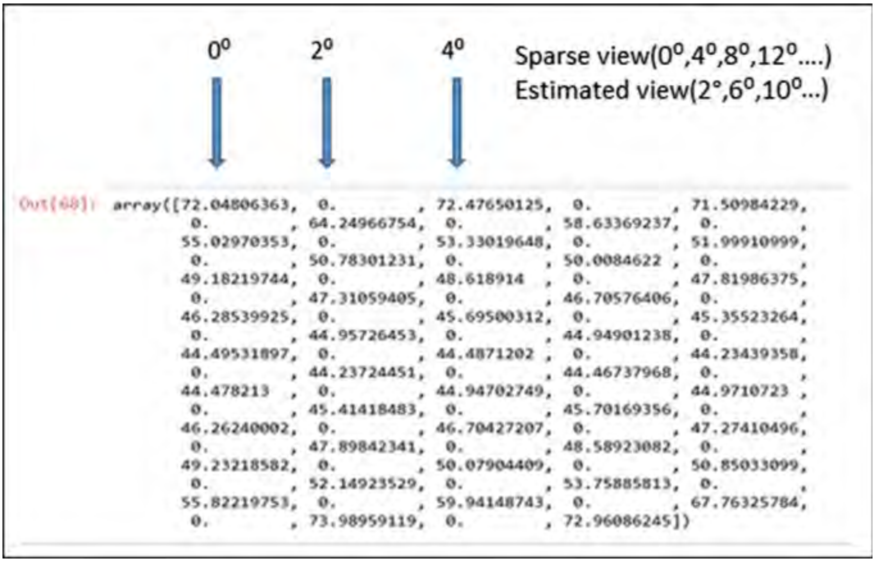


Fig. 8 Sparse projections and predicted values filled with 0 in sinogram space

The Shepp–Logan phantom head model is used for experiment. Figure 6 represents reconstructed images for a sparse view. The reconstructed images for sparse angle 4°, 6°, and 8° are called low-dose images, as the number of dose on the object are reduced. These images exhibit complex artifacts and low spatial resolution. The challenge is to improve the quality of the image by applying suitable technique.

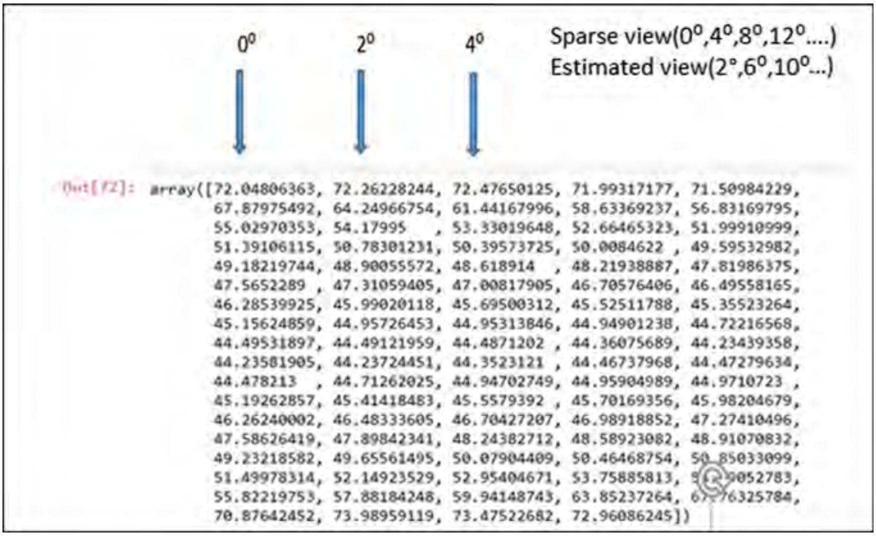


Fig. 9 Estimated projection value using simple average method

4.3 *Predicting the Missing Data*

The missing projection data due to sparse views are estimated using simple averaging of neighboring projection data in sinogram space. In Fig. 9, it can be observed that the missing projection value is estimated and filled by using the averaging method.

5 Results and Discussion

Comparison of low-dose and reconstructed image using proposed method is shown Fig. 10. It depicts that the proposed methodology yields a better quality image compared to the low-dose image for a different sparse angle.

5.1 *Result Analysis*

The performance measuring parameters such as structural similarity index matrix (SSIM), peak signal-to-noise ratio (PSNR), and root mean square error (RMSE) were calculated and compared between low-dose and image reconstructed using the proposed method. Table 1 presents the recorded values.

From Table 1, it is witnessed that the proposed technique produces better SSIM score, PSNR, and RMSE value compared to the conventional low-does method except

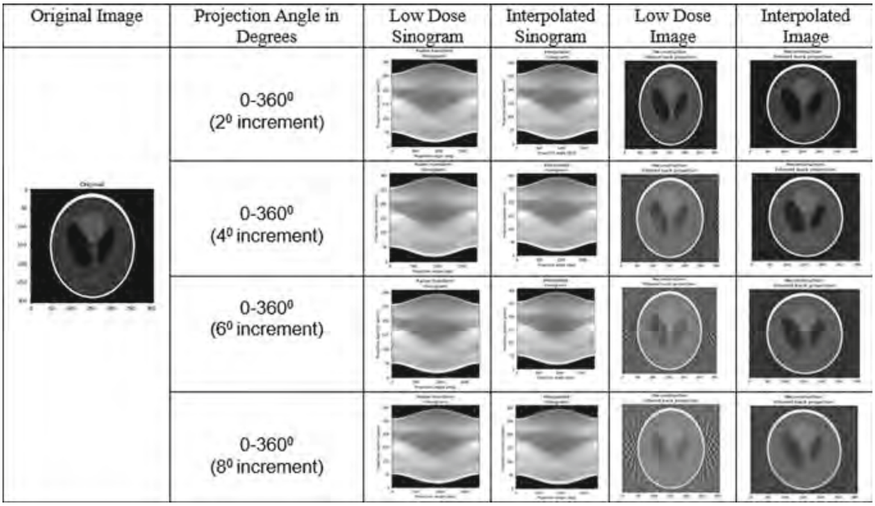


Fig. 10 Comparison of low-dose and reconstructed images with proposed method

Table 1 Performance measures

Increment angle in degrees	PSNR		SSIM		RMSE	
	Low dose	Low dose (with correction)	Low dose	Low dose (with correction)	Low dose	Low dose (with correction)
2°	79.8	78.5	0.91	0.95	0.026	0.030
4°	72.5	74.2	0.74	0.88	0.060	0.049
6°	68.0	70.5	0.67	0.83	0.101	0.075
8°	65.3	68.4	0.62	0.77	0.138	0.096

for lesser sparse angle, i.e., for 2° incremental angle (180 views). This means that the proposed method is not viable for a lesser sparse angle. SSIM up to 0.8 is clinically accepted (from literature study); hence, a sparse view with 6° is the optimal incremental angle beyond which image may not have diagnostic visibility.

6 Conclusion and Future Scope

The sparse-view CT (low dose) images usually exhibit low spatial resolution and complex artifacts. To deal with such an issue, the proposed research work has presented the simple-averaging method which predicts the intermediate missing projection data in sinogram space to yield a better quality image. After incorporating sinogram-based correction for low-dose imaging, there are few chances of missing finite tissue structure details that may affect the diagnostic decision. However, the proposed method overcomes the lost medical information due to sparse view. From the experimentation results, it is concluded that the proposed method performs better compared to the conventional low does method except for a lesser sparse angle. SSIM up to 0.8 is clinically accepted; hence, a sparse view with 6° is the optimal incremental angle beyond which image may not have diagnostic visibility.

In the future, real-time sparse-view CT images will be considered for experiment, and the performance of FBP will be compared with proposed model.

References

1. A.C. Kak, Computerized tomography with X-ray, emission, and ultrasound sources. Proc. IEEE **67**(9), 1245–1272 (1979)
2. M. K. Kalra, M. M. Maher, T. L. Toth, L. M. Hamberg, M.A. Blake, J. A. Shepard and S. Saini," Strategies for CT radiation dose optimization," Radiology, vol.230, pp.619–628,2004.
3. S.D. Desai, L. Kulkarni, A quantitative comparative study of analytical and iterative reconstruction techniques. International Journal of Image Processing (IJIP) **4**(4), 307 (2010)
4. S.D. Desai, Reconstruction of image from projections-an application to MRI & CT Scanning, in *Proceedings of International Conference ICSCI-2005* (2005)
5. Nam Kyung Lee, Low-dose CT with the adaptive statistical iterative reconstruction V technique in abdominal organ injury: comparison with routine—Dose CT with filtered back projection. Am. J. Roentgenol. **213**, 659–666 (2019)
6. L. Zhang, Airway quantification using adaptive statistical iterative reconstruction-V on wide-detector low-dose CT: a validation study on lung specimen. Jap. J. Radiol. (2019)
7. A. Euler, A third-generation adaptive statistical iterative reconstruction technique: phantom study of image noise, spatial resolution, lesion detectability, and dose reduction potential, in *AJR:210*, June 2018
8. J. Sun, *Performance Evaluation of Two Iterative Reconstruction Algorithms, MBIR and ASIR, in Low Radiation Dose and Low Contrast Dose Abdominal CT in Children* (Italian Society of Medical Radiology, 2020)
9. N. Hamaguchi, *Improved Depictions of the Anterior Choroidal Artery and Thalamoperforating Arteries on 3D CTA Images Using Model-based Iterative Reconstruction* (Association of University Radiologists, 2020).
10. T. Higaki, *Deep Learning Reconstruction at CT: Phantom Study of the Image Characteristics* (The Association of University Radiologists, 2019)
11. R. Singh, Image quality and lesion detection on deep learning reconstruction and iterative reconstruction of submillisievert chest and abdominal CT, in *AJR* (2020)
12. Z. Zhang, A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution. IEEE Trans. Med. Imaging **37**(6) (2018)
13. S. Xie, Artifact removal using improved GoogLeNet for sparse-view CT reconstruction. Sci. Rep. (2018)
14. D.H. Ye, *Deep Back Projection for Sparse-View CT Reconstruction* (IEEE, 2018)
15. H. Nakai, *Quantitative and Qualitative Evaluation of Convolutional Neural Networks with a Deeper U-Net for Sparse-View Computed Tomography Reconstruction* (The Association of University Radiologists, 2019)

Transfer Learning for Children Face Recognition Accuracy

R. Sumithra, D. S. Guru, V. N. Manjunath Aradhya, and Raghavendra Anitha

Abstract Identifying the missed and kidnapped children at the later age will be a quit challenging process. To overcome the challenge, this research work has proposed a new Children Face Recognition (CFR) application using Artificial Intelligence (AI) system. To the best of our knowledge, the existence of children face image dataset, which is created in a fruitful process has not been reported in the earlier literature. Hence, to this consequence, this research work has addressed the problem of developing the children face recognition model with a suitable dataset. A model has been proposed by using machine learning pipeline that consists of pre-processing, feature extraction, dimensionality reduction and learning model. To this end, an attempt has also been made to classify the face images of children by training a multi-classification algorithm with the ensemble techniques such as Bagging and Boosting. During the dataset creation, 40,828 longitudinal face images of 271 young children of age from 4 to 14 years are captured over the duration of 30 months. This extensive experimentation has analyzed that few projection vectors of k-NN classifier has achieved a high accuracy of about 93.05%.

Keywords Children face recognition · Bagging · Boosting · Longitudinal face image

---

R. Sumithra (✉) · D. S. Guru
Department of Studies in Computer Science, University of Mysore, Manasagangotri, Mysuru, Karnataka 570006, India

V. N. Manjunath Aradhya
Department of Computer Application, JSS Science and Technology University, Mysuru, Karnataka, India

R. Anitha
Department of ECE, Maharaja Institute of Technology, Mysuru, Karnataka, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_44

1 Introduction

Face Recognition (FR) is a challenging area, where a great deal of work has been accomplished in biometric applications for pose variation, partial face identification and illumination variation to help the process of face recognition [11]. Face images are beneficial not just for the identification of individuals but also for disclosing other characteristics such as gender identification, expression recognition, a person's emotional state detection, age estimation, ethnicity etc. A desirable biometric modality such as its uniqueness, universality, acceptability and easy collect ability is generated by the facial parameters. In the field of biometrics, face recognition is well-studied problem that is still unsolved because of the inherent problems presented by human faces [10]. An aging recognition is now attracting more attention from the face recognition community as it is an inevitable phase and the appearance of human faces changes remarkably with the progression of age [4]. Aging involves differences in the shape, the texture and size of the face with regard to the facial recognition system. Output degeneration would be triggered by these temporal changes. Therefore, the ID issued by the government, e.g. driving license, Passport, has to be updated every 5–10 years. Aging refers to improvements over a period of time in a biometric trait, which can theoretically influence a biometric system's accuracy [17].

Several efforts are being made to explain the aging impact on systems of face recognition. However, as a subject of the growth and development process of childhood, algorithm efficiency with respect to human aging has only begun to be thoroughly explored by researchers. As the face changes over time, the capacity to identify the individual becomes more difficult. Anthropological and forensic research have made a major contribution to demonstrating that children's age-related changes vary from adult facial aging [20]. A model for the extraction of facial landmarks for the face recognition method was introduced by [12]. Deb et al. [3], examined the method of face recognition as a "man-machine" method, where human experts had to manually locate certain facial landmarks on a matching image. Then, based on 20 normalized distances obtained from these facial landmarks, matching was performed automatically. By exploiting both texture and shape characteristics of face images, the morph-able model proposed by Blang and Vetter [2] improves the use of 3D models in face recognition.

Some of the most important developments in the field of face recognition in the last decade are Sparse Representation Coding (SRC) [26] and face recognition based on deep learning [22]. Most facial recognition techniques presume that faces (both geometrically and photo-metrically) can be positioned and correctly normalized where the alignment can be performed based on the locations of two eyes on the face. A face detection model considered to be a breakthrough was created by [25] because it allows faces to be identified even in the presence of background clutter in real time. On order to extract age-related characteristics, decomposition of facial attributes such as identity, expression, gender, age, race, pose etc., is necessary. Nevertheless, a few attempts has been made in the literature to identify a person through photographs of facial aging. Anthropomorphic [21], active appearance model [5], aging pattern subspace [9], and aging manifold [8] are some of the typical aging face models. In infants, facial aging primarily includes craniofacial growth and development. The variations of the cranium and face year after year from childhood to young adulthood are identified in Karen T. Taylor's book "Forensic Art and Illustration". In non-adults, the rate of face changes is maximum, especially between birth to 5 years of age [18]. The precision of identification is highly dependent on the age of the subject. Nevertheless, the bulk of the study on adults aging found in the literature faces recognition. Reference [14] concluded that face recognition technology is not yet ready to recognize very young children reliably, especially face images taken at 3 years and above.

As observed in the previous literature that the success of a face recognition system largely depends on the stability of the facial parameters in the face. Any highly sophisticated and efficient face recognition system will fail if the parameters of a face change regularly. One of the biggest issues is the vast amount of data that is required to fully understand the human face and its maturation process. To the best of our knowledge, there is no existing children face image dataset created in a very fruitful process. Hence, to this consequence, the problem of developing the children face recognition model has been addressed by creating a suitable database. During the dataset creation, 40,828 longitudinal face images of 271 young children of age from 4 to 14 years are captured, over the duration of 30 months. A model has been proposed by using machine learning pipeline consists of pre-processing, feature extraction, dimensionality reduction and learning model. To this end, an attempt has also been made to classify children face images by training a multi-classification algorithm using ensemble techniques, such as Bagging and Boosting.

The following are the contribution of this study:

- Creation of large sized longitudinal face images of a young children by maintaining the quality and quantity of our dataset.
- Study of conventional techniques such as feature engineering, feature transformation and supervised classification for the recognition of a children.
- Exploitation of traditional approach for the fusion of multi- classification using OR rule.
- Employed an Ada-boosting and Random forest classifiers for boosting and bagging techniques.

The rest of the paper is organized as follows: explanation of proposed model in Sect. 2. Section 3 reports a dataset creation and experimental results with analysis, and Sect. 4 gives a conclusion and future enhancement.

2 Proposed Model

The overview of the our proposed model and its pictorial representation is shown in Fig. 1. The model has two stages, first one with feature representation using PCA and transformed FLD, and second one with the decision level fusion of multi-classifiers called ensemble technique, which is depicted in Fig. 2.

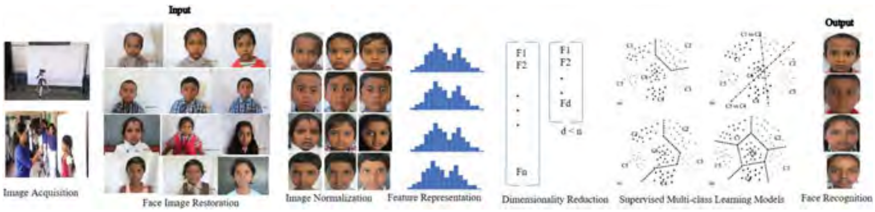


Fig. 1 The pictorial representation of our proposed model

Fig. 2 The technical description of our proposed model

2.1 *Feature Extraction and Its Representation of Children Face Images*

Pre-processing The aim of the pre-processing technique is to extract the face region from an image, using Viola-Jones face detection method by fixing the threshold and region size [25]. This research work has extracted only the face part from each image and pre-processed the images of same child at different age interval are shown in Fig. 3.

Handcrafted Feature Computation During the feature extraction, the local and handcrafted features viz., Histogram of Orientation Gradient (HOG) and Multi-scale Local Binary Pattern (M-LBP) are utilized. The adopted features have widely used and highly discriminating in face recognition. Hence, they have utilized in our study, by hyper-parameter tuning.

Fig. 3 The Illustration of pre-processing **a** captured face image; **b** pre-processed image

HOG: The features are extracted by dividing the face image window into small spatial regions. Each detection window is divided into cells of size 32 × 32 pixels and each group of 4 × 4 cells are integrated into a block in a sliding mode, so blocks overlap with each other. Each cell consists of an oriented gradient (HOG) 9 bin histogram, and each block includes a concatenated vector of all its cells.

MLBP: This operator works by thresholding a 3×3 neighborhood with the value of the center pixel, thus forming a local binary pattern, which is constructed as a binary number. Features calculated in a local 3 × 3 neighborhood cannot capture large-scale structures because the operator is not very robust against local changes in the texture. Therefore, an operator with a larger spatial support area is needed. The operator was extended to facilitate rotation invariant analysis of facial textures at multiple scales such as 3 × 3, 5 × 5, 7 × 7, and 9 × 9 (Ojala et al. 2002).

Dimensional Reduction Using Subspace Technique To reduce the feature dimension, subspace methods called Principal component Analysis (PCA) [24] and Fisher Linear Discriminant (FLD) [15] is used to preserve the most dominating projection vectors.

PCA Let W represent a linear transformation matrix mapping the function points from m-dimension to p-dimension, where $p \ll$ m, as follows:

$$Y_p = w^T X_p \tag{1}$$

is the linear transformation of the extracted features. Where $\{Wi|i = 1, 2, \ldots, m\}$ is the set of n-dimensional projection vectors corresponding to the m largest non-zero eigenvalues in Eq. (1).

FLD An example of class specific approach is the Fisher's Linear Discriminant (FLD). This method selects W in such a way that maximizes the ratio of the scatter between class and the scatter within class, Eq. (2).

$$W = \arg\max_{W} \left| \frac{W^T S_B W}{W^T S_W W} \right| \tag{2}$$

where S_B is between-class scatter matrix $S_B = \sum_{i=1}^{C} N_i\,(x_i - \mu)\,(x_i - \mu)^T$ and the S_W is within class scatter matrix be defined as $S_W = \sum_{i=1}^{C} \sum_{X_k \in X_i}^{C}\ (x_k - \mu_i)$ $(x_k - \mu_i)^T$. Where μ_i is the mean image of class X_i, and N_i is the number of samples in class X_i. Where $\{W_i|i = 1, 2, \ldots, m\}$ is the set of generalization eigen vectors of S_B and S_W corresponding to the m largest generalized eigenvalues $\{\lambda\ i|i = 1, 2, \ldots m\}$. The number of images in the learning set in general much smaller than the number of features in the image. This means that matrix W can be chosen in such a way that the projected samples' in-class scatter can be rendered exactly null. This has been achieved by using PCA to reduce the size of the feature space to N-c, and then applied the standard FLD defined by Eq. (3), to reduce the size to c-1 [1].

More formally, W_{FLD} is given by:

$$W_{\text{FLD}}^T = W_{\text{fld}}^T \; W_{\text{pca}}^T \cdots \tag{3}$$

where $W_{\text{fld}} = \arg\max_W \frac{|W^T W_{\text{pca}}^T S_B W_{pca} W|}{|W^T W_{\text{pca}}^T S_W W_{pca} W|}$ and $W_{\text{pca}} = \arg\max_W \|W^T S_T W\|$ In computing W_{pca}, only the largest c-1 projection vectors are selected.

2.2 Supervised Learning Models

Learning Models In the literature, there are many learning models used for face recognition system. Thsi study has evaluated the quality of our dataset by conducting a number of baseline learning models. Therefore, ifferent classification method viz., k-Nearest Neighbor (k-NN) [13], Support Vector Machine (SVM) [23], Navies Bayes [19] and Decision tree [6], and the fusion of multi- classifiers using majority voting technique are used as shown in Fig. 4.

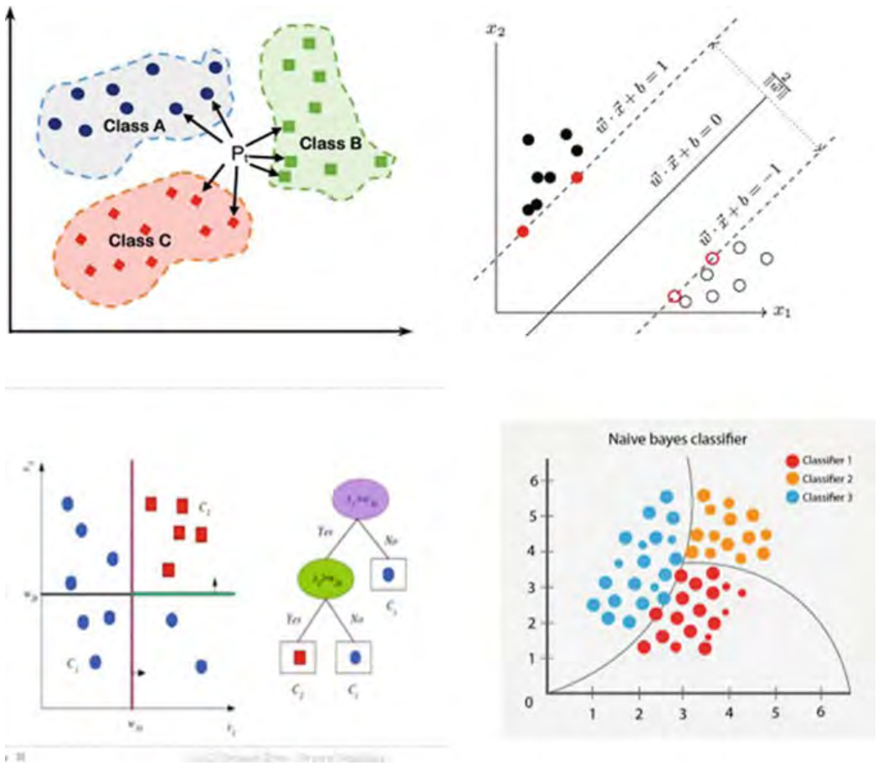


Fig. 4 Fusion of multi classifier using majority voting

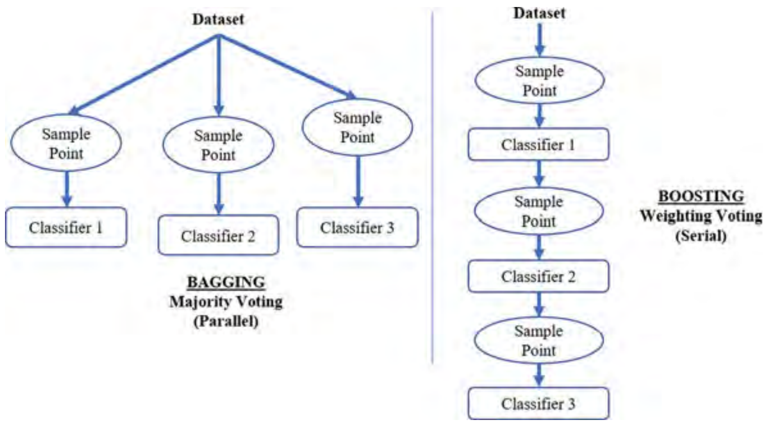


Fig. 5 Ensemble technique: bagging and boosting

Ensemble Leaning A standard decision fusion technique called ensemble has been adopted for enabling a deeper understanding of our dataset. Ada-Boosting classifier in boosting [7] and Random Forest in bagging [16] techniques are used during by transfer learning as shown in Fig. 5.

3 Experiment Results

3.1 Longitudinal Face Images of Children

This research work has created a own dataset of longitudinal face images of young children of 4–14 years old. In order to capture the face images of school children, a permission has been taken from the Director, the Board of Education (BOE), Northern region, Mysore, India. Our longitudinal face data collection has conducted in different Government schools in and around Mysore, India. To capture the children face images, parents and teachers have required to sign a consent form giving their permission to provide their child face images. In addition to the image, other metadata has also collected such as date-of-birth, gender, and date of capture, child name and father's name. The face images have captured using 16 MP rear end camera of canon. Face images have collected in the school premises with a suitable setup made by the author as shown in Fig. 6. To maintain a degree of consistency throughout the database, the same physical setup and location with semi-controlled environment have used in each session. However, the equipment have to reassemble for each session; therefore, there is a variation from session to session.

Due to the rapid growth in facial features of young children and to analyze the minute growth rate in more effectively and efficiently, a longitudinal face images of

Fig. 6 Process of our dataset creation

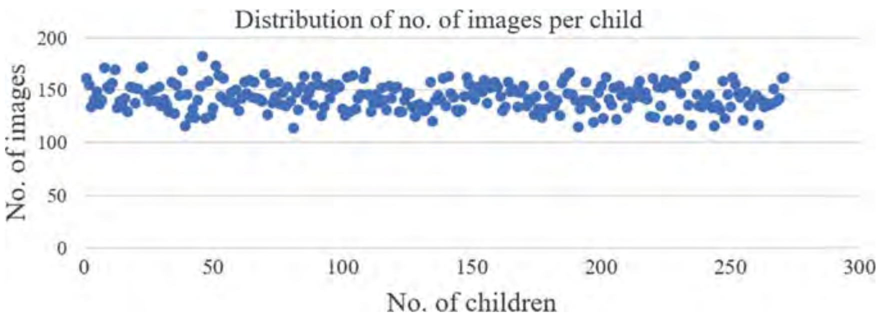


Fig. 7 Histogram of our dataset used during experiment

children of 4–14 years old are created, over the period of 30 months in 10 different sessions for every 3–4 months interval. Our dataset comprises of 271 classes each class represents a child. Hence, the total number of children encountered in our study have 271. The minimum and maximum number of images taken from each child have 114 and 182 respectively. The sum of 40,858 longitudinal face images of 271 children have been considered for experimentation, among which 137 are male and 134 are female. The histogram of number of children and images are shown in Fig. 7 and the Longitudinal face images are shown in Fig 8.

3.2 Experimental Analysis

In this section, the recognition has been computed by fixing the setup for its maximum performance through an extensive experimentation. During HOG feature extraction,

Fig. 8 Longitudinal Face image dataset

the block of windows represented by a 36-D feature vector that is normalized to unit length have been considered. For each 64 × 128 detection window is represented by 7 × 15 blocks, gives a total of 4184 features. Similarly for M-LBP, the rotation invariant analysis of facial textures are facilitated at multiple scales such as 3 × 3, 5 × 5, 7× 7, and 9 × 9. Obtained with 59 features from each scale therefore, sum of 256 features are computed. Then, the feature transformation has represented by using PCA and transformed FLD. After applying PCA and transformed FLD, features have been reduced to 25% from the total number of projection vector. Hence after feature transformation, it has been obtained with 64 (25% of 256) projection vector from the M-LBP feature and 1046 (25% of 4184) projection vector from HOG feature. By adopting this reduction techniques, it has been noticed that, there is an improvement in the rate of recognition and the computational time after feature transformation.

The classification results of accuracy, f-measure, precision and recall for the HOG features are shown in Fig. 9, similarly, for MLBP features are shown in Fig. 10. From the computed performance, it has been observed that the rate of recognition is high with the combination HOG feature and k-NN classifier. Hence, it has experimented by using k-NN classifier for cumulatively increasing the number of projection vectors, during feature selection.

The accuracy for cumulatively increased projection vectors is also increased and saturated (Curse of Dimensionality) at some projection point as shown in Table 1.

The results obtained from the fusion classifiers have also appreciated. Ensemble classifier such as Ada-boosting for boosting technique and random forest for bagging technique have been utilized in this study. Ensemble of 10 decision trees have considered with row and column replacement. For Ada-boosting, the observation weights are equal to the number of samples taken for experimentation. The number

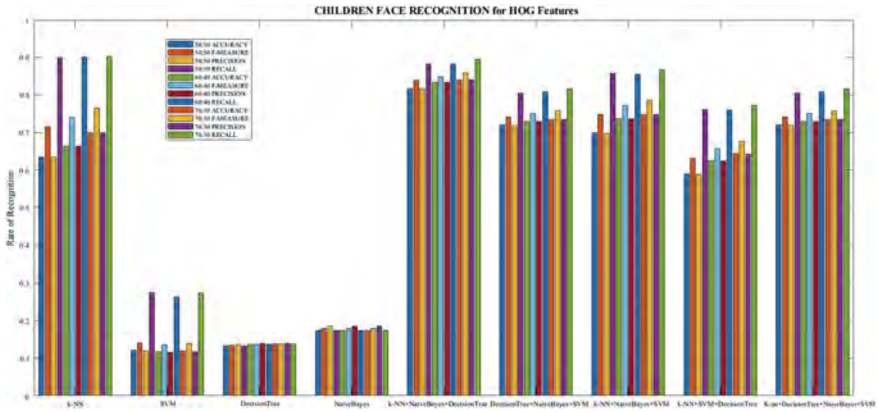


Fig. 9 Classification results for HOG feature by varying different training and testing samples

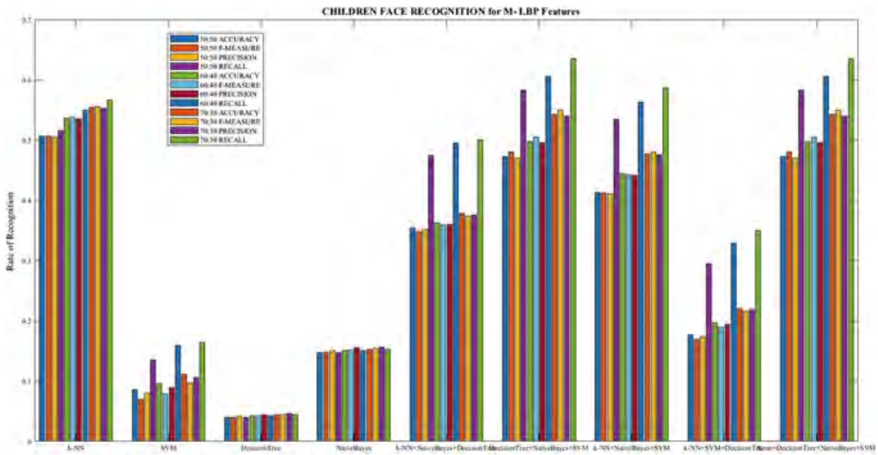


Fig. 10 Classification results for MLBP feature by varying different training and testing samples

of learning cycle has 100 for all predictor combinations. 10k-fold validation with prior probabilities has been done empirically with classification margin distribution speed of 0.1 s.

In Table 2, the combination of HOG features with Bagging classifier has the highest accuracy of 86.44%, at 70:30 training and testing split. There has been a huge variation in the rate of recognition between Bagging technique and Boosting. Hence Bagging is much superior then Boosting for our created dataset, which is depicted in Table 2.

Table 1 Accuracy for a Selected projection vectors using HOG feature for k-NN classifier

Projection	Accuracy		
	50:50	60:40	70:30
5	50.80 ± 0.10382	51.44 ±0.38295	51.60 ±0.3977
10	81.39 ± 0.07987	81.98 ± 0.20951	82.39 ± 0.24057
15	88.23 ± 0.20792	88.46 ± 0.20859	88.55 ± 0.15816
20	88.63 ± 0.12335	88.96 ± 0.16279	89.22 ± 0.20576
25	90.11 ± 0.12572	90.33 ± 0.17087	90.53 ± 0.16804
30	91.23 ± 0.18625	91.28 ± 0.16475	91.24 ± 0.18816
35	91.87 ± 0.09057	91.98 ± 0.14208	92.04 ± 0.18402
40	92.64 ± 0.09848	92.71 ± 0.13732	92.64 ± 0.17166
45	**93.05 ± 0.13014**	**93.07 ± 0.08559**	**93.13 ± 0.1488**
50	91.68 ± 0.14302	91.71 ± 0.15777	91.82 ± 0.16175
55	90.03 ± 0.14787	90.32 ± 0.21474	90.41 ± 0.21342
60	90.25 ± 0.13816	90.43 ± 0.17375	90.52 ± 0.15426
65	88.75 ± 0.16017	89.15 ± 0.16972	89.38 ± 0.28191
70	89.01 ± 0.15279	89.35 ± 0.20781	89.66 ± 0.24056

Table 2 Results for ensemble classification recognition

Classifier	Features	Split	Accuracy	F-Measure	Precision	Recall
Boosting	HOG	50-50	18.24	13.44	13.5	17.46
		60-40	17.6	13.1	14.12	16.79
		70-30	17.9	13.93	14.26	17.17
	MLBP	50-50	2.64	0.61	0.47	2.46
		60-40	2.69	0.57	0.56	2.48
		70-30	2.51	0.49	0.71	2.32
Bagging	**HOG**	50-50	71.39	77.42	71.43	90.11
		60-40	86.16	86.21	86.22	86.5
		70-30	**86.44**	**86.5**	**86.5**	**86.95**
	MLBP	50-50	36.4	35.99	36.19	38.14
		60-40	37.76	37.23	37.6	39.41
		70-30	38.8	38.62	38.65	40.91

4 Conclusion and Future Enhancement

The aim of this study is to find the rate of recognition of young children using aging face images. Due to the drastic growth pattern in young children, the recognition through aging is esoteric task. To address such problem, a simple and efficient model that consists of feature representation and transfer learning models has been built. During feature extraction, local features such as M-LBP and HOG, and feature transformation such as PCA and FLD are adopted. For transfer learning model, different baseline models with their fusion using majority voting technique has been adopted. Also an attempt has been made with ensemble techniques such as Bagging (Random Forest) and Boosting (Ada-Boosting) for higher analysis. For experimentation, our own dataset of 40,828 longitudinal face images of 271 young children of age from 4 to 14 years are created, over the duration of 30 months. It has been concluded that, the high rate of recognition is obtained with the combination of HOG features and k-NN classifier by varying the projection vector. In future, the Deep Convolution Neural Network (D-CNN) technique can be adopted to enhance the performance of the model.

References

1. P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 711–720
2. V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model. IEEE Trans. Pattern Anal. Mach. Intell. **25**(9), 1063–1074 (2003)
3. W.W. Bledsoe, in *Man-machine face recognition*. Technical report PRI 22. Panoramic Research, Inc.
4. D. Deb, N. Nain, A.K. Jain, in *Longitudinal study of child face recognition*. arXiv preprint arXiv: 1711.03990 (2017)
5. G.J. Edwards, T.F. Cootes, C.J. Taylor, *Face Recognition Using Active Appearance Models*. Image Analysis Unit, Department of Medical Biophysics, University of Manchester, Manchester M13 9PT, UK
6. M.A. Friedl, C.E. Brodley, Decision tree classification of land cover from remotely sensed data. Remote Sens. Environ. **61**(3), 399–409 (1997)
7. J.H. Friedman, Stochastic gradient boosting. Comput. Statistics Data Anal. **38**(4), 367–378 (2002)
8. Y. Fu, T.S. Huang, Human age estimation with regression on discriminative aging manifold. IEEE Trans. Multimedia **10**(4), 578–584 (2008)
9. X. Geng, Z.H. Zhou, K. Smith-Miles, Automatic age estimation based on facial aging patterns. IEEE Trans. Pattern Anal. Mach. Intell. **29**(12), 2234–2240 (2007)
10. A.K. Jain, A. Ross, K. Nandakumar, *Introduction to Biometrics* (Springer, Berlin, 2011)
11. A. K. Jain, K. Nandakumar, A. Ross, 50 years of biometric research: accomplishments, challenges, and opportunities. Pattern Recogn. Lett. (2011). https://doi.org/10.1016/j.patrec.2015.12.013
12. T. Kanade, in *Picture processing system by computer complex and recognition of human faces*. Ph.D. thesis, Kyoto University (1973)
13. J.M. Keller, M.R. Gray, J.A. Givens, A fuzzy k-nearest neighbor algorithm. IEEE Trans. Syst. Man Cybern. **4**, 580–585 (1985)
14. B.-R. Lacey, Y. Hoole, A. Jain, Automatic face recognition of newborns, infants, and toddlers: a longitudinal evaluation, in *2016 International Conference of Biometrics Special Interest Group (BIOSIG)* (IEEE, 2016)
15. S. Mika, G. Ratsch, J. Weston, B. Scholkopf, K.R. Mullers, Fisher discriminant analysis with kernels, in *Neural networks for signal processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop* (cat. no. 98th8468) (IEEE, 1999), pp. 41–48
16. M. Pal, Random forest classifier for remote sensing classification. Int. J. Remote Sens. **26**(1), 217–222 (2005)

17. K. Ricanek, S. Bhardwaj, M. Sodomsky, A review of face recognition against longitudinal child faces, in *BIOSIG2015* (2015)
18. K. Ricanek, Y. Karl, Y. Wang, Y. Chang, C. Chen, Demographic analysis of facial landmarks. U.S. Patent 9,317,740, issued April 19, 2016
19. I. Rish, An empirical study of the Naive Bayes classifier, in *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, Vol. 3(22) (2001), pp. 41–46
20. L. Rowden, Y. Hoole, A. Jain, Automatic face recognition of newborns, infants, and toddlers: a longitudinal evaluation, in *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)* (IEEE, 2016)
21. A.S. Sohail, P. Bhattacharya, Detection of facial feature points using anthropometric face Model Concordia University, 2-9525435-1 SITIS (2006), pp. 656–664
22. Y. Sun, X. Wang, Tang, Deep learning Face Representation from predicting 10,000 classes, in *Proceedings of IEEE Conference on Computer Vision and PR.PP*, pp. 1891–1898
23. J.A. Suykens, J. Vandewalle, Least squares support vector machine classifiers. Neural Process. Lett. **9**(3), 293–300 (1999)
24. M.A. Turk, A.P. Pentland, Face recognition using Eigen faces, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'91)* (IEEE, 1991)
25. P. Viola, M.J. Jones, Robust real-time face detection. Int. J. Comput. Vision **57**(2), 137–154 (2004)
26. J. Wright, A.G. Yang, S. Sastry, Y. Ma, Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. **31**, 361–369 (2009)

A Fact-Based Liver Disease Prediction by Enforcing Machine Learning Algorithms

Mylavarapu Kalyan Ram, Challapalli Sujana, Rayudu Srinivas, and G. S. N. Murthy

Abstract Health is prominent in mortal cheerfulness and financial progress across the world. Being healthy and active is also very important to the human lifestyle. Numerous major problems like liver disease, heart disease, and diabetes may decrease one's activity level. One of the major issues is liver abnormality. In India, most individuals suffer from the deficiency of the liver. Most liver diseases occur due to some human faults like consumption of heavy alcohol, consuming fast foods, usage of identical needles while injecting drugs, etc. This may lead to many complications in the future. Due to liver malfunctioning, many disorders may take place such as liver cancers, hepatic angiosarcoma, pediatric hepatoblastoma, hepatitis B, and hepatitis C, etc. Machine learning techniques have been strived on numerous liver disorder datasets with Python programming language to find out the accurate and exact analysis of disease. The ultimate goal of this investigation is to predict the best accuracy after successfully executing machine learning algorithms that have been used on various liver disease datasets.

Keywords BUPA · Classification algorithms · Feature selection · ILPD · Liver disorder · Performance measures · Prediction · Machine learning

1 Introduction

The liver is the major organ that acts a massive role in the human body. It acts as a real filter that retrieves and eradicates many toxins. The liver ensures the metabolism

M. K. Ram (✉) · R. Srinivas · G. S. N. Murthy
CSE Department, Aditya Engineering College (A), Surampalem, A.P., India
e-mail: kalyanram.mylavarapu@aec.edu.in

R. Srinivas
e-mail: srinivas.rayudu@aec.edu.in

C. Sujana
IT Department, Pragati Engineering College (A), Surampalem, A.P, India
e-mail: sujana.ch@pragati.ac.in

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_45

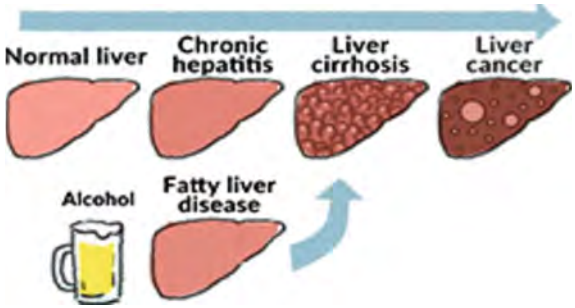


Fig. 1 Various liver disorders

of carbohydrates and fat, which accomplishes several activities like showing of bile, decomposition of RBC, etc. Bile is an essential element for the human digestive system. The toxins generated by the liver are wasted due to human faults or by increasing the content of ammonia in the human body. The pessimistic point in medical management is increasing the count of liver disorder patients. In India, patients suffering from liver disorders undergo with high morbidity and mortality rates. Liver tissue is incorporated with too many lobules and each lobe is organized with internal cells, the elemental metabolic cells of the liver. Several actions can influence and escalate liver disease risks. Figure 1 depicts the major liver disorders. Popular liver disorders are fatty liver, liver cirrhosis, hepatitis B and hepatitis C, and liver cancer. Major origin of liver diseases **is** genetic liver disorder, smoking, consumption of alcohol, consuming fast foods, obesity, and diabetes.

This article aims to develop the best model to assist medical practitioners to find the finest treatment at preliminary stages for liver disorders. Numerous medical diagnosis mechanisms can be predicted by using machine learning techniques. Various liver complications are not resolved in the prior days because of abundant tools and a lack of awareness of a peculiar disease.

The layout of the research work is described in the following manner. Section 2 describes the literature survey, and Sect. 3 represents the operational requirements. Section 4 illustrates the proposed methodology that considers different algorithms for predicting liver disorders. Section 5 exhibits the experimental results, and finally, Sect. 6 proposes the conclusion and future work.

2 Literature Survey

In the past few years, plenty of research has been performed on liver disorders prediction using machine learning techniques throughout the world. In this section, various authors contributed and essence their work by using machine learning algorithms to predict liver disorders.

Wu et al. [1] used various classification algorithms such as RF, NB, ANN, and LR to predict liver fatty disease early with accurate results. A comparison of these four algorithms has been considered based on their classification accuracy measure. To evaluate the performance, they were used ROC of all models with cross-validation. Finally, they concluded random forest as the best algorithm for the prognosis of fatty liver abnormality with the highest accuracy of 87.48 than other algorithms.

Kumar et al. [2] described a new hybrid fuzzy-ANWKNN classification approach to predict liver disease with the highest accuracy and efficient manner. The existing fuzzy-ADPTKNN was not satisfied to predict liver disease. They were implemented on different datasets like the Indian liver dataset and Madhya Pradesh liver dataset. Finally, the author concluded a new hybrid approach gave the best performance than the existing approach.

Sivakumar et al. [3] analyzed liver prediction based on some human quality factors. Chronic liver disease was predicted using some machine learning algorithms such as k-means clustering and statistical classifier decision tree approaches. Finally, all measurements have been taken then concluded C4.5 algorithm is superior to other classification calculations.

Idris et al. [4] the authors analyzed liver disease prediction as early by implementing various supervised learning algorithms such as LR, SVM, and RF. The above algorithms got different accuracies but the authors increased accuracy using new algorithms, i.e., Ada boost classifier logistic and bagging random forest algorithms which got accuracies of 74.35 and 72.64, respectively.

Nahar et al. [5] worked on liver disease prediction on various decision tree techniques like RF, J48, random tree, REPT tree, decision stump, LMT, and hoeffding. The authors used an Indian liver dataset to evaluate these techniques and obtained performances of all these algorithms. From the result analysis, the decision stump performance is excellent than other algorithms and obtained 70.76% accuracy.

Arshad et al. [6] data mining classification algorithms such as Naïve Bayes, SMO, Bayes Net, and J48 are used to predict liver disease. In this article, the authors implemented the WEKA tool on the BUPA dataset and concluded that SMO gained the highest accuracy of 97.3% where Naïve Bayes gained less accuracy of 70.72%.

Kumar et al. [7] various classification algorithms namely Naïve Bayes, C5.0, k-means, random forest, and C5.0 with adaptive boosting have been implemented for finding out the liver disease prediction. Based on vulnerability, the latter classification algorithm has been introduced and it proved to predict the disease more accurately. Among all algorithms, random fast secured the highest accuracy with 72.10% but after implemented C5.0 with adaptive boosting, it got a more accurate result of 75.19%.

Kuzhippallil et al. [8], the author described a new genetic algorithm, namely the XGBoost classifier and compared various algorithms and introduced envisaging procedures for liver disorder prediction. The author also explained anomaly detection techniques. Finally, the author concludes the new algorithm and obtained good accuracy and reduced the classification time.

Pathan et al. [9] proposed a methodology using several classification algorithms like RF, Naive Bayes, J48, Ada Boost, and bagging. The author implemented all algorithms on ILDP using WEKA Tool. Finally, the author compared all results and finalized random forest provides accurate of 100% result.

Muthuselvan et al.[10] proposed a liver disease prediction based on a computer-aided diagnosis. Through computers, all liver data analyzed and stored all images in the database. This was very useful for further prediction. The author compared all four algorithms, namely K-star, Naïve Bayes, J48, random tree, and concluded that random tree secured the highest accuracy of 74.2% with a minimum time of 0.05 secs.

Shapla Rani Ghosh and Sajjad Waheed [11] proposed liver disease prediction on various machine learning algorithms, namely logistic tree, REP tree, Naive Bayes, bagging, and K-star. The author was implemented on ILDP using WEKA Tool and evaluate the accuracy, precision, sensitivity, and specificity. Finally, the author revealed when compared to Naïve Bayes, K-star got efficient accuracy of 100%.

Thirunavukkarasu et al. [12] illustrated several classification algorithms such as KNN, SVM, and logistic regression were utilized for liver disorder prediction. Based on severity, the classification algorithm has become a solution to predict the disease in advance.

Dr. Vijayarani et al. [13] proposed two classification algorithms, namely SVM and Naïve Bayes for predicting liver disease. When compared to these two algorithms based on performance and execution time, SVM generated high accuracy.

Dhamodharan [14] proposed two classification algorithms for predicting liver disorders such as Naïve Bayes and FT tree to uncover diseases like liver cancer, cirrhosis, and hepatitis, etc. After compared these algorithms on WEKA Tool, Naïve Bayes obtained good accuracy than FT tree.

Bahramirad [15] proposed 11 classification algorithms such as K-star , Gaussian processes, logistic, linear logistic multilayer perceptron, RIPPER, rule induction, support vector machine, regression, logistic model trees, neural net, and classification and regression trees. The author used two datasets, namely the AP liver dataset and BUPA liver dataset. After observing the results, the accuracy performance of APL was better than BUPAL.

Venkata Ramana et al. [16] reported that four classification algorithms such as Naïve Bayes, SVM, backpropagation NN, and C4.5 were used to predict early liver disease problems. These algorithms were implemented on various datasets and evaluated based on the performance of accuracy.

Venkata Ramana et al. [17] reported that the common attributes in both NL and INDIA datasets are ALKPHOS, SGPT, and SGOT [22]. Two methods were applied for the analysis of these attributes. These methods are ANOVA and MANOVA. When compared these two methods, MANOVA was analyzed better than ANOVA.

Venkata Ramana et al. [18] proposed that the combination of Bayesian classification with bagging and boosting techniques obtained the best accuracy of male 97.91 and female 91.16. for liver disease prediction [19].

3 Operational Requirements

In this article, an intake of two liver patient datasets was used to estimate the performance.

1. The Indian Liver Patients Dataset (ILPD) retrieved from UCI ML database which consists of 583 records and 11 attributes.

Out of 583 records, 416 are liver disorder records and 167 are non-liver disorder patients records. This dataset consists of a chronicle of 441 male patients and 142 female patients [20]. Table 1 displays the particulars of ILPD dataset.

2. The BUPA dataset from the UC Irvine Machine Learning [20, 21] Repository California state of the USA which consists of 345 records and 7 attributes.

Table 2 displays the particulars of the BUPA dataset. The first five variables are acquired from blood tests which may arise from the consumption of too much alcohol [21].

Table 1 Indian liver data with attributes

Index	Attribute	Description	Range	Type
1	Age	Patient's age	4–90	Real numbers
2	TB	Total bilirubin	0.4–75	Real numbers
3	DB	Direct bilirubin	0.1–19.7	Real numbers
5	ALB	Albumin	10–2000	Real numbers
6	A/G ratio	Albumin and globulin ratio	10–4929	Real numbers
7	SGPT	Alamine aminotransferase	2.7–9.6	Integer
8	SGOT	Aspartate aminotransferase	0.9–5.5	Integer
9	ALP	Alkaline phosphotase	0.3–2.8	Real numbers
10	Gender	Gender of patient	F/M	Categorical
11	Selector field	Diseased or not	0/1	Binominal

Table 2 Bupa liver dataset and attributes

Index	Attribute	Description	Range	Type
1	Mcv	Mean corpuscular volume	[65, 103]	Integer
2	Alkphos	Alkaline phosphatase	[23, 138]	Integer
3	Sgpt	Alamine aminotransferase	[4, 155]	Integer
4	Sgot	Aspartate aminotransferase	[5, 82]	Integer
5	Gammagt	Gamma-glutamyl transpeptidase	[5, 297]	Real numbers
6	Drinks	No. of alcohols drunk per day	[0.0, 20.0]	Real numbers
7	Selector	Diseased or not	{1,2}	Binominal

4 Proposed Methodology

4.1 Structure for Classification of Liver Patient Dataset

The below flowchart represented the classification process of the Indian Liver Disease Dataset and BUPA dataset. This process begins with the dataset selection. In Fig. 2, the origin begins with data selection from both ILPD and BUPA datasets in CSV file format; thereby, preprocessing is done which is used to transform the raw data into a useful format. Repositories were classified based on the parameters such as liver and non-liver patients. Individually, both the datasets were imported to the Jupyter notebook, for ILPD, the training set and test set are 451 and 113 samples were considered, respectively. For BUPA dataset, 310 for training set and 27 elements for testing set. After the selection of the boundaries, the algorithms are determined for the execution using the current dataset.

The outcomes of this classification model were isolated. The output can be shown distinctly analyzed as a classifier output.

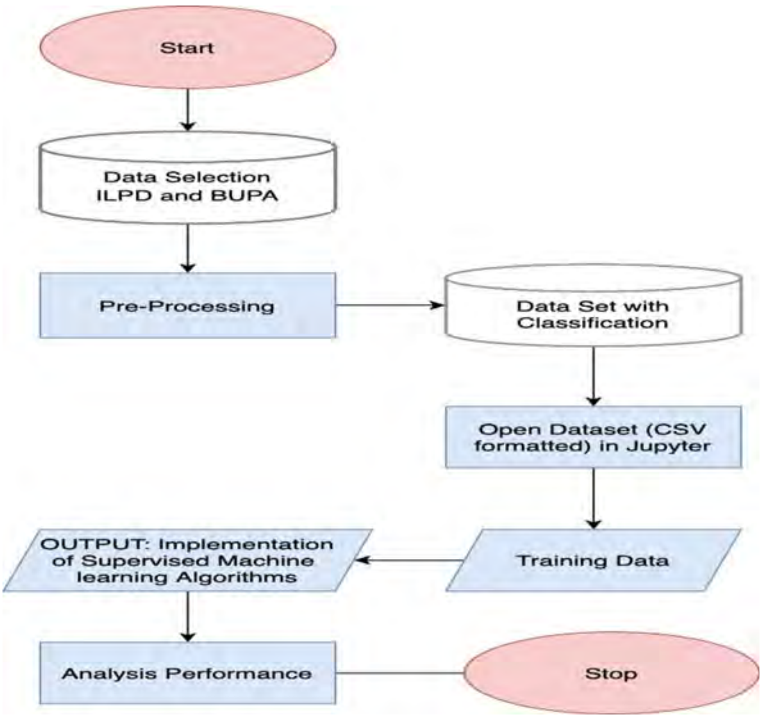


Fig. 2 Flowchart for classification and prediction of the liver patient dataset

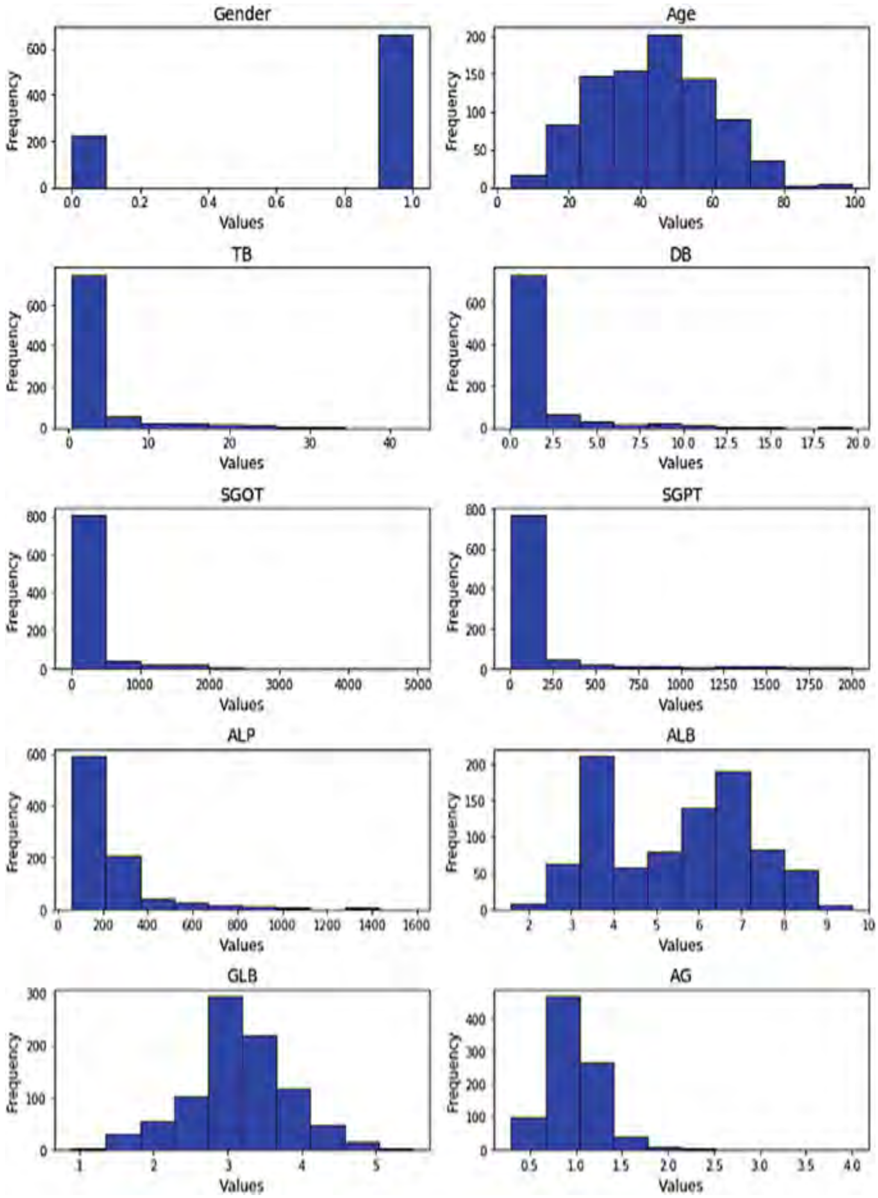


Fig. 3 Histogram—numerical features for ILPD dataset attributes (No. of records on X-axis and range of features on Y-axis)

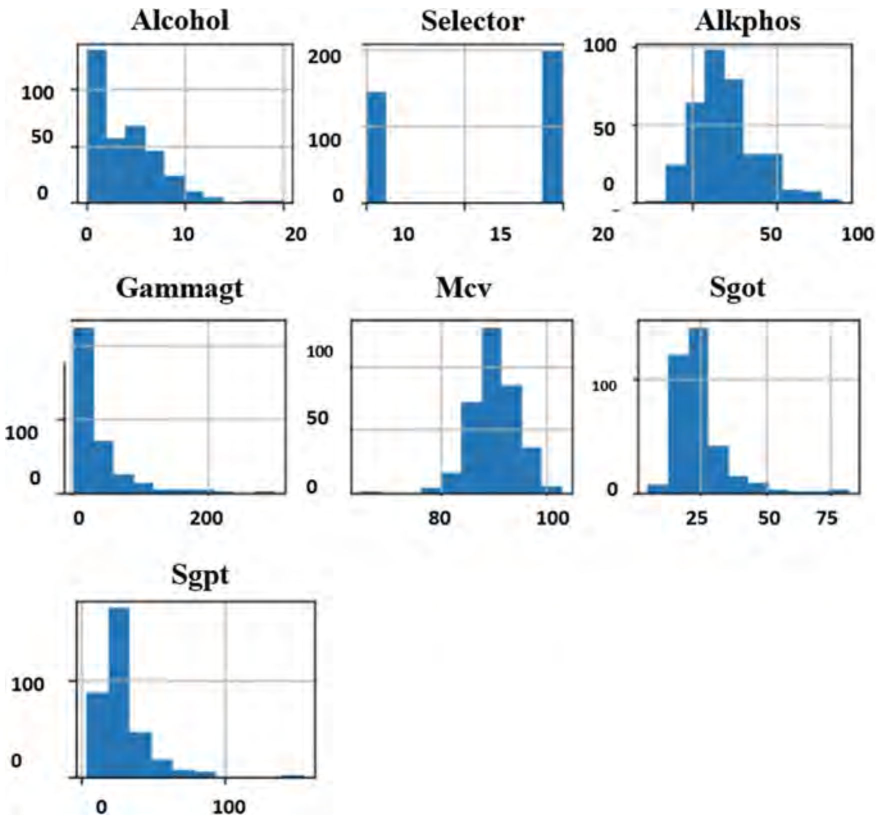


Fig. 4 Histogram—numerical features for BUPA dataset attributes (Number of records on X-axis and range of features on Y-axis)

Figures 3 and 4 illustrate the pre-processed visualized histogram for ILPD and BUPA repositories after successful completion of the data regularization and conversion. This procedure will make all different values into one group.

In the implementation phase, the proposed liver disease prediction an accurate classification algorithm has been considered and accomplished on the above two datasets. The algorithms are logistic regression, decision tree induction, Naive Bayes, SVM, RF, KNN, and gradient boosting selected for implementation on the datasets. The execution is performed by using Python Jupyter. After the successful completion of attaining all the supervised machine learning algorithms, the confusion matrix will be generated, analyzed, and discussed for all the intake algorithms.

4.2 Feature Selection

The feature selection is the finest idea in machine learning which highly impacts the performance of the classification model at concrete levels. It is an activity where one can select the features manually or in a mechanized way which provides us a prediction variable. The performance of the designed model may diminish if the data consists of irrelevant features. Feature selection methods are classified into three types such as filter-based, wrapper-based, and embedded.

4.3 Normalization

In machine learning, normalization is one of the best techniques. Using this technique, all numeric attributes values are to be changed within the range of the same scale. In the current research, normalization is applied to two datasets (ILPD and BUPA).

4.4 Classification Using Machine Learning Algorithms

Machine learning (ML) is one of the eminent domains in research that utilizes various algorithms to construct as well as analyze the models. Figure 5 describes the machine learning algorithms such as supervised learning techniques, unsupervised learning techniques, semi-supervised learning techniques, and reinforcement learning techniques [1].

In this study, the demonstration of different types of strategies applied to liver disorder prognostics is projected, several machine learning algorithms like multilayer perceptron, SVM, logistic regression, random forest, KNN, Naïve Bayes,

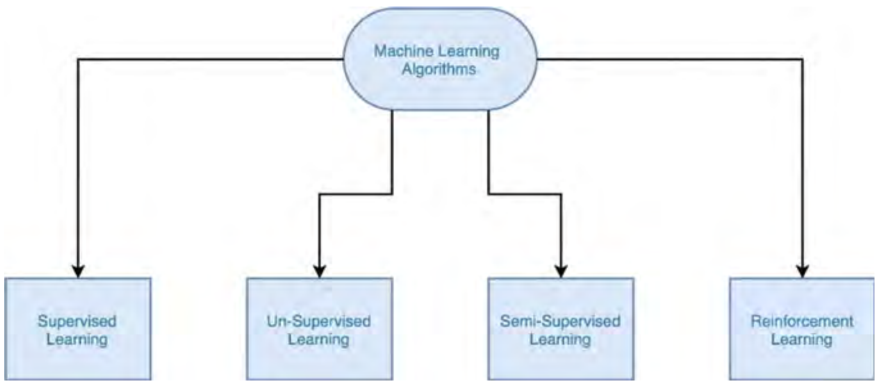


Fig. 5 Categories of machine learning

AdaBoost, XGBoost, decision tree, gradient boosting, Bayesian and bagging, etc., are implemented [2].

The main task of this research is:

1. Distinct classification algorithms are deployed for liver disease forecasting.
2. Comparing different algorithms
3. Finding the best algorithm for liver disease prediction.

In the current scenario, several classification algorithms are obtainable to forecast the disease at preceding stages and increase the life span of the sufferer [6]. For every algorithm, ten-fold cross-validation is furnished. The algorithms are discussed below.

4.4.1 Multilayer Perceptron Classifier

This classifier is originated from the Weka library and closely related to the neural network operator. There are three layers in MLP such as an input layer, an output layer, and in between hidden layers as depicted in below Fig. 7. In the research, all complex problems are easily solved by MLP. Some applications of MLP are health prediction, speech recognition, NLP, and image processing.

4.4.2 K-Nearest Neighbor (KNN) Classifier

The k-nearest neighbor is a statistical method that uses machine classification algorithms and regression lines. The outputs of this classifier are class membership and KNN regression is the value for the object. KNN applications are gene expression, get missing values, protein–protein prediction, pattern recognition, etc. KNN used to calculate the Euclidean distance to predict the class as follows:

$$\sqrt{\sum_{i=1}^{k} (x_i - y_i)^2}$$

4.4.3 Logistic Regression Classifier

Logistic regression is an algorithm associated with machine learning and is used to solve the classification difficulties like prediction, detection, or used to diagnosis of an issue. It is a predictive algorithm based on probability. Various categories of logistic regression include binary logistic regression, multinomial logistic regression, and ordinal logistic regression. Binary logistic regression is used for liver disease prognosis that concerns intended variables which are present or not. In logistic regression, one of the cost functions is known as the sigmoid function.

4.4.4 Decision Tree Classifier

In machine learning, a decision tree is one of the supervised algorithms that resolve classification and regression problems. Various categories of decision trees are like categorical variable decision tree and continuous variable decision tree. In the categorical variable decision, tree splitting is done through the Gini method whereas continuous variable decision looks after a continuous target variable.

4.4.5 Random Forest Tree Classifier

Random forest tree is used to solve classification and regression problems. Each decision tree trains on different observations. Random forests are very functional and give good accuracy in prediction.

4.4.6 Gradient Boosting Classifier

Gradient boosting is a distinct approach in machine learning algorithms that solved various classification and regression problems. This algorithm is used with CART trees of all base learners.

4.4.7 Support Vector Machine (SVM) Classifier

The SVM classifier was coined by Vapnik in 1979 [13]. In machine learning algorithms, one of the best-supervised learning algorithms is SVM. By using this algorithm, different prediction problems are solved. For the currently deployed model, SVM obtained the highest accuracy with more than 85%.SVM is based on learning theory and useful in ML. In Fig. 6 represents the performance accuracy of SVM classification algorithms.

Figure 8 represents a 2D visualization of data points that are linearly separable related to SVM.

There are three important concepts in SVM.

1. **Support Vectors**: Data points near the hyperplane.
2. **Hyperplane**: It is a decision plane which is divided between a set of classes.
3. **Margin**: The space between two lines on the nearest data points of data classes.

The SVM algorithm works as follows for a linearly separable dataset.

Pseudo Code for Support Vector Machine (SVM)

1. Initialize the dataset D as (X_1, y_1), (X_2, y_2), … (X_n, y_{n}), where X_i is training tuple, $X_i \in R_d, y_i$ is labeled class corresponding to X_i and let $y_i \in$ $(+1, -1)$

Fig. 6 System architecture

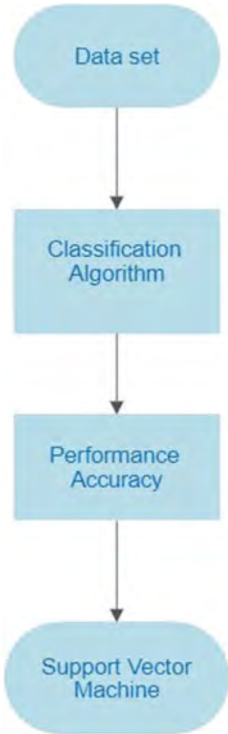
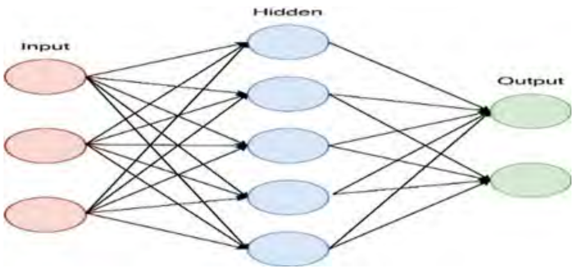


Fig. 7 Multilayer perception



2. Corresponding to each training tuple X_i let the weight Vector W_i then the optimal hyperplane is computed as

$$W \cdot X + b = 0 \tag{1}$$

where b is the bias.

3. If any point lies above the separating hyperplane, then that point follows the following inequality

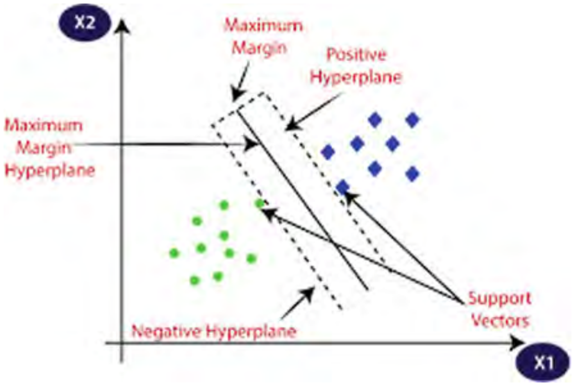


Fig. 8 Two-dimensional case where the data points are linearly separable

$$W \cdot X + b > 0 \tag{2}$$

4. If any point lies below the separating hyperplane, then that point follows inequality

$$W \cdot X + b < 0 \tag{3}$$

5. To define the sides of the margin, the weight can be adjusted by computing the following equations

$$H_1 : W \cdot X + b \geqslant 1 \quad \text{for } y_i = +1 \tag{4}$$

$$H_1 : W \cdot X + b \leqslant 1 \quad \text{for } y_i = -1 \tag{5}$$

   Equations (4) and (5) are jointly written as

$$y_i(W \cdot X + b) \geqslant 1\ \forall_i \tag{6}$$

6. The X_i that satisfies the Eq. (6) is called support vector.
7. To maximize the margin between two separating hyperplane, compute the following equation

$$\frac{\text{Max } 2}{\|W\|} \quad \text{(or)} \quad \text{Min } \|W\| \tag{7}$$

   Subject to the condition Eq. (6).
8. By solving Eq. (7) using Lagrangian formulation and Karush–Kuhn–Tucker (KKT) conditions

The equation can be rewritten in the following manner

$$f(X) = \Sigma a_i y_i X_i^T X + b \tag{8}$$

where X represents the test point, X_i represents the support vectors, y_i is the class label of support vector X_i; a_i and b are numeric parameters automatically determine by the SVM algorithm.

Naive Bayes Classifier

Naive Bayes is derived from the Bayes theorem of conditional probability. Those are independent of each other. There are three models in Naive Bayes, namely the Gaussian model, multinomial model, Bernoulli model.

AdaBoost Classifier

Adaptive Boosting algorithm is one of the best algorithms which solves both binary classification and regression. AdaBoost is used with small decision trees. Mostly, it will work for weak learners.

XGBoost Classifier

XGBoost algorithm is one of the freely available libraries that provide high performance on data.

Bayesian Classifier

Bayesian classification model is based on Bayes theorem. In this, the data class can be classified by using probability. The Bayesian logistic regression model was implemented by using Gaussian and Laplace priors.

Bagging Classifier

Bagging is a plain and robust ensemble method. Every element of the classifier is being trained by a discrete sample. Through plurality voting, the classification is done.

4.5 Performance Metrics Analysis

To compute the data analysis and prediction, numerous machine learning classification algorithms are employed on the liver datasets of both ILDP and BUPA. These are implemented using Python programming.

1. **Confusion Matrix**: is a record used for the computation of performance analysis by allowing easy visualization. The confusion matrix confesses the difference between true-positives, false-negatives, false-positives, and true-negatives.
2. **Accuracy**: Accuracy determines the sum of positive liver disorder occurrences of the defined model to the total number of occurrences. The accuracy margins define the probability of rightly classified occurrences.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

3. **Precision**: Precision defines the correctly classified disorder occurrences to the total obtained occurrences and is defined as follows

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

4. **Recall**: The fraction of related occurrences retrieved to the aggregation of all occurrences.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

5. **F-measure**: The evaluation done on the weighted mean of Precision and recall is defined as F-measure.

$$\text{F} - \text{Measure} = \frac{2 * \text{Presion} * \text{Recall}}{\text{Presion} + \text{Recall}}$$

6. **Time Complexity**: Time complexity evaluates the total amount of time to compute the predictions

5 Results

In this study, twelve classification algorithms have been applied with sampling techniques. The work is implemented using Python programming. The detailed outcomes of the performance are shown in Tables 3 and 4. According to Table 3, it can be seen that SVM on ILPD gives better results for the parameter accuracy 85.70%, precision 83.70%, recall 83.90%, F-score 85.70%. Table 4 also shows that the SVM on BUPA obtained the best result for the accuracy of 83.65%, precision 81.23%, recall 80.09%,

Table 3 Classification results from ILPD dataset

S. No.	ILPD dataset performance measures					
	Classification algorithm	Accuracy %	Precision %	Recall %	F-measure	Time complexity
1	Multilayer perceptron	72.50	51.40	70.50	72.50	0.15
2	KNN	74.20	69.10	73.50	74.20	0.11
3	Logistic regression	74.90	71.30	68.40	74.90	0.99
4	Decision tree	65.70	68.00	67.00	65.70	0.89
5	Random forest tree	74.60	72.78	74.20	74.60	0.99
6	Gradient boosting	69.40	67.00	65.00	69.40	0.15
7	Support vector machine	85.70	83.70	83.90	85.70	0.03
8	Naive Bayes	62.00	59.20	59.00	62.00	0.05
9	AdaBoost	68.70	62.87	65.10	68.70	0.36
10	XGBoost	70.30	69.30	70.00	70.30	0.56
11	Bayesian	71.40	73.20	68.20	71.40	0.54
12	Bagging	75.30	73.45	76.34	75.30	0.45

Table 4 Classification results from BUPA dataset

S.No.	BUPA dataset performance measures					
	Classification algorithm	Accuracy %	Precision %	Recall %	F-measure	Time complexity
1	Multilayer perceptron	69.54	75.23	69.45	69.54	0.13
2	KNN	73.50	65.40	72.10	73.50	0.12
3	Logistic regression	72.30	69.90	68.40	72.30	0.76
4	Decision tree	69.34	65.40	62.00	69.34	0.89
5	Random forest	74.60	71.91	73.40	74.60	0.91
6	Gradient boosting	69.40	67.00	65.00	69.40	0.12
7	Support vector machine	83.65	81.23	80.09	83.65	0.04
8	Naive Bayes	67.23	58.45	54.30	67.23	0.05
9	AdaBoost	65.67	69.12	63.23	65.67	0.36
10	XGBoost	71.34	69.12	65.30	71.34	0.51
11	Bayesian	70.34	71.30	69.23	70.34	0.52
12	Bagging	71.30	69.98	72.40	71.30	0.45

F-score 83.65%. Figures 9 and 10 show a comparative graphical representation of performance for both datasets.

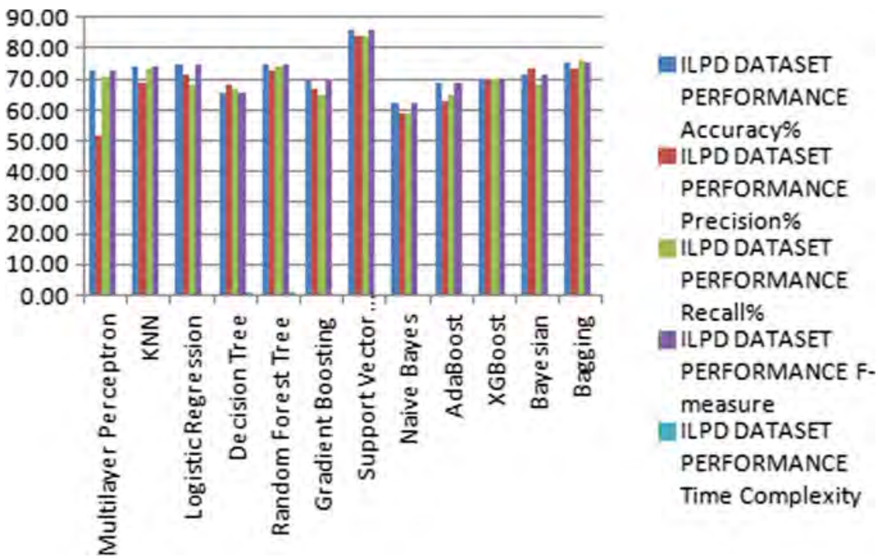


Fig. 9 Comparison of the performance of selected classification in percentages with all attributes of ILDP dataset

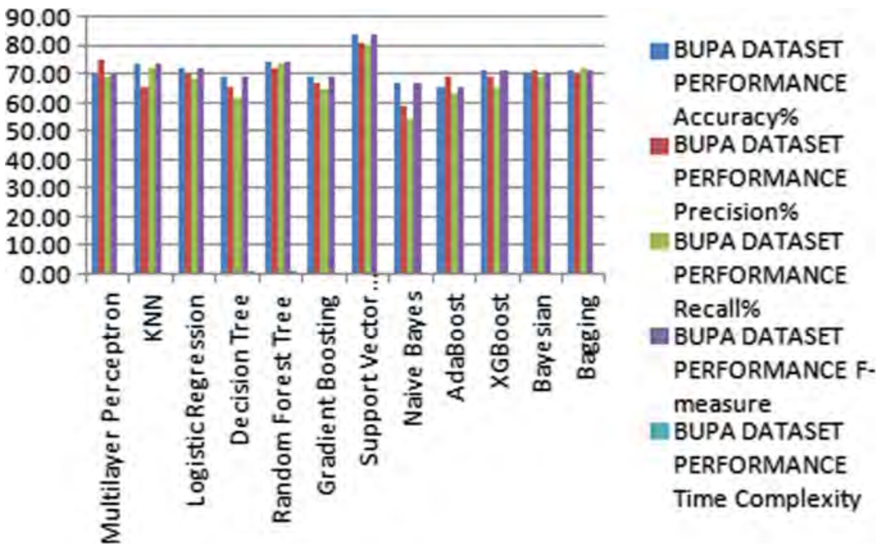


Fig. 10 Comparison of the performance of selected classification in percentages with all attributes of BUPA dataset

Table 5 Execution time analysis for liver disease dataset

Classification algorithm	ILPD execution time in ms	BUPA execution time in ms
Multilayer perceptron	0.1532	0.134
KNN	0.11	0.123
Logistic regression	0.99	0.769
Decision tree	0.897	0.897
Random forest tree	0.999	0.91
Gradient boosting	0.15	0.1239
Support vector machine	0.0391	0.0401
Naive Bayes	0.059	0.059
AdaBoost	0.365	0.365
XGBoost	0.564	0.514
Bayesian	0.543	0.523
Bagging	0.456	0.456

Table 3 presents the implementation of the supervised algorithm on the ILPD dataset and shows the accuracies in the prognosis of liver disease. Also, the SVM operates with the highest prediction accuracy than other algorithms. Figure 9 depicts the performance metrics concerning the achieved results of ILPD datasets (in terms of %).

Table 4 represents the implementation of the supervised algorithm on the BUPA dataset and shows the accuracies in the prognosis of liver disease. Out of all, high accuracy is achieved by the SVM algorithm. Figure 10 depicts the performance metrics concerning the achieved results of BUPA datasets (in terms of Percentages).

Table 5 depicts the evaluation of execution time between classification algorithms for forecasting liver disorder prognosis from both ILPD and BUPA datasets. Support vector machine executes within optimum time 0.0391 ms and 0.0401 ms and at the same time, Fig. 11 depicts the pictorial graph representation of optimum results.

6 Conclusion and Future Scope

Liver infection identification at the prior stage is crucial to combatThe frequent and terrible deaths of patients in India. The patients must be shield based on early symptoms for the development of customized therapy. In this research work, implementation on ILPD and BUPA datasets from the UCI repository has been taken as an input to predict liver disorders and applied 12 machine learning algorithms on these datasets. Out of all comparisons, SVM gives the best accurate results with 85.70% using

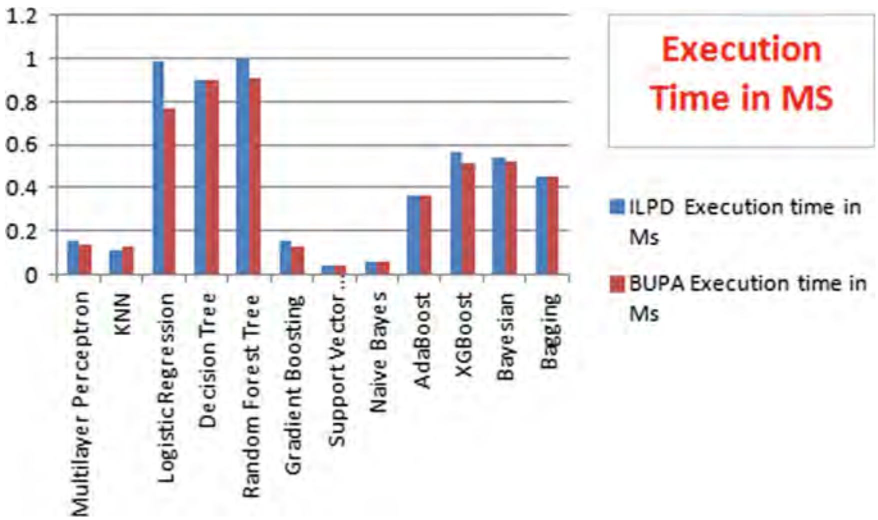


Fig. 11 Execution time analysis

ILPD, 83.65% using BUPA, and execute within optimal time 0.0391 ms, 0.0401 ms, respectively.

The future work, the outcomes of this research will support the development of a novel approach by implementing a hybrid classification model along with a smart mobile system for uninterrupted observation of the risky liver disorder patients, and it can be extended up to by appending an automatic emergency call mechanism when the impaired is in threat.

References

1. C.C. Wu, et al., Prediction of fatty liver disease using machine learning algorithms. Comput. Methods Programs Biomed. **170**, 23–29 (2019). https://doi.org/10.1016/j.cmpb.2018.12.032
2. P. Kumar, R.S. Thakur, Diagnosis of liver disorder using fuzzy adaptive and neighbor weighted K-NN method for LFT ımbalanced data, in *6th IEEE International Conference on Smart Structures and Systems (ICSSS)* (2019). https://doi.org/10.1109/ICSSS.2019.8882861
3. D. Sivakumar, M. Varchagall, L.G. Ambika, S. Usha, Chronic liver disease prediction analysis based on the impact of life quality attributes. Int. J. Recent Technol. Eng. **7**(6), 2111–2116 (2019)
4. K. Idris, S. Bhoite, Applications of machine learning for prediction of liver disease. Int. J. Comput. Appl. Technol. Res. **8**(9), 394–396 (2019). https://doi.org/10.7753/ijcatr0809.1012
5. N. Nahar, F. Ara, Liver disease prediction by using different decision tree techniques. Int. J. Data Min. Knowl. Manag. Process. **8**(2), 01–09 (2018). https://doi.org/10.5121/ijdkp.2018.8201
6. I. Arshad, C. Dutta, T. Choudhury, A. Thakral, Liver disease detection due to excessive alcoholism using data mining techniques, in Proceedings of 2018 International Conference

on Advanced Computing and Communication Engineering (ICACCE) (2018), pp. 163–168. https://doi.org/10.1109/ICACCE.2018.8441721
7. S. Kumar, S. Katyal, Effective analysis and diagnosis of liver disorder by data mining, in Proceedings of International Conference on Inventive Research in Computing Applications (ICIRCA) (2018), pp. 1047–1051. https://doi.org/10.1109/ICIRCA.2018.8596817
8. M.A. Kuzhippallil, C. Joseph, K.A, Comparative analysis of machine learning techniques for Indian Liver disease patients (2020), pp. 778–782. https://doi.org/10.1109/icaccs48705.2020.9074368
9. A. Pathan, comparative study of different classification algorithms on ILPD dataset to predict liver disorder. Int. J. Res. Appl. Sci. Eng. Technol. **6**(2), 388–394 (2018). https://doi.org/10.22214/ijraset.2018.2056
10. S. Muthuselvan, S. Rajapraksh, K. Somasundaram, K. Karthik Classification of liver patient dataset using machine learning algorithms, Int. J. Eng. Technol. **7**(3.34), 323–326 (2018) (Special Issue 34)
11. G. Ahmad, M.A. Khan, S. Abbas, A. Athar, B.S. Khan, M. S. Aslam Automated diagnosis of Hepatitis B using multilayer mamdani fuzzy ınference system. J. Healthcare Eng. **2019** (2019)
12. K. Thirunavukkarasu, A. S. Singh, M. Irfan, and A. Chowdhury Prediction of liver disease using classification Algorithms, 2018 4th Int. Conf. Comput. Commun. Autom. ICCCA 2018, vol. 6, no. 9, pp. 1–3, (2018)
13. S. Vijayarani, S. Dhayanand, Liver disease prediction using SVM and Naïve Bayes algorithms. Int. J. Sci. Eng. Technol. Res. **4**(4), 816–820 (2015)
14. S. Dhamodharan, Liver disease prediction using bayesian classfication, in *4th National Conference on Advance Computing and Application Technology* (2014), pp. 1–3
15. S. Bahramirad, A. Mustapha, M. Eshraghi, Classification of liver disease diagnosis: a comparative study, in *2nd International conference on Informatics and Applications* (ICIA) (2013), pp. 42–46
16. B. Venkata Ramana, M.S.P. Babu, N. Venkateswarlu, A critical study of selected classification algorithms for liver disease diagnosis. Int. J. Database Manag. Syst. **3**(2), 101–114 (2011)
17. B. Venkata Ramana, P.M. Surendra, P. Babu, P.N.B. Venkateswarlu, A critical comparative study of liver patients from USA and INDIA: an exploratory analysis. Int. J. Comput. Sci. Issues **9**(3), 506–516 (2012)
18. A. Pradesh, A. Pradesh, A critical evaluation of Bayesian classifier for liver diagnosis using bagging and boosting methods. Int. J. Eng. Sci. Technol. **3**(4), 3422–3426 (2011)
19. M. Banu Priya, P. Laura Juliet, P.R. Tamilselvi, Performance analysis of liver disease prediction using machine learning algorithms. Int. Res. J. Eng. Technol. **5**(1), 206–211 (2018)
20. ILPD (Indian Liver Patient Dataset) Dataset, in *UCI Repository of Machine Learning Databases.* Available in https://archive.ics.uci.edu/ml/datasets/ILPD. Last accessed December 15, 2012
21. BUPA Liver Disorders Dataset, in *UCI Repository of Machine Learning Databases.* Available in https://archive.ics.uci.edu/ml/datasets/Liver+Disorders. Last accessed December 15, 2012
22. J. Han, M. Kamber, *Data Mining Concepts and Techniques*, 2nd ed. (Morgan Kaufmann, San Francisco, 2006)

Novel Approach to Data Hiding in Binary Images Minimizing Distortion

Gyankamal J. Chhajed and Bindu R. Garg

Abstract In today's digital communication, much important information is sent as picture images in binary form. To secure this information, data is hidden in these images to preserve originality, authenticity of their visible content or to covert communication as invisible contents. The proposed idea is to hide data in such a way that the flipping pixels in the cover are minimized. The very first ciphertext is generated by encrypting original data with a key. Then, the binary image is scanned in four ways, and out of these four scannings, one that is having maximum pixel and ciphertext bit matching count is selected for hiding the data by flipping the unmatched pixel. The header with cipher embedding details and its length is embedded from the end of the image in reverse order applying odd–even feature to 3 × 3 blocks. During extraction, the exact opposite method is applied so that the hidden data can be accurately extracted.

Keywords Binary image · Data hiding · Distortion

1 Introduction

Nowadays, digital media has become a daily need. Documents, transaction details, medical images, etc. are transferred frequently as a binary picture image. Authentication of such digital content is very important. Usually, watermarks are embedded in these digital images to prove their authenticity or originality. Watermarks are either visible or invisible. If invisible watermarks are introduced, then the security of digital images increases. Achieving invisible watermarking is possible by hiding watermark

G. J. Chhajed (✉)
VPKBIET, Baramati, India
e-mail: gjchhajed@gmail.com; gyankamal.chhajed@vpkbiet.org

B. R. Garg
BVDUCOE, Pune, India
e-mail: brgarg@bvucoep.edu.in

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_46

details in cover media in such a way that it is unnoticeable to the human visual system. Similarly, for secret communication, secret data is hidden in images in such a way that it will not be noticeable to anybody else than the communicators that are sender and receiver. Such type of secret communication is called steganography. Most of the research for watermarking and steganography is on color and gray-scale images, where pixels can have 256 values for each RGB colors and 256 values, respectively. In such images, pixel values are modified in such a way that changes are not easily visible. This property of the human visual system needs to be maintained for binary images as well. In binary images, the pixels can have only two values, and hiding data without causing visual distortion is a great challenge. The key idea in a binary image to achieve this is to identify areas of the picture that would be utilized as much as possible to hide data so that less number of pixels are required to be changed. In existing work, many schemes were proposed in the spatial and frequency domain. In some techniques, either limited pixels are available for embedding purposes, or if more pixels are used to carry data, then they cause more visual distortion. The efficiency of the data hiding technique can be extended by using maximum pixels for embedding with exiting arrangements of pixels in image. In some steganography methods, embedding is done in the 3×3 [1–3] block of a cover image. In such blocks, one or two bits can be embedded using different algorithms.

2 Literature Survey

In existing work, various methods and techniques are proposed for hiding data in the binary image for watermarking and steganography. The hiding data in the eight-connected boundary of a character of binary text documents was proposed in [4]. A fixed set of pairs of five-pixel boundary patterns was identified for embedding data. Deletion and addition of the center foreground pixel in a pair are required to embed data. The scheme proposed in [5] trades in such a way that image quality is maintained. In this, any modified bit in the host image and its adjacent bits are compared to keep hiding effect which is quite invisible. It checks for watermark embedding possibility in the discrete cosine transform (DCT) domain for binary images [6]. In the method, locates "flippable" pixels and maintains relationships within the block to hide data without causing visual distortion [1–3]. Kim and de Queiroz [7] propose a method that detects visual alteration which maintains visual quality for all types of binary images. In this, a variation of the proposed scheme is also included that locates the updated region. A blind data hiding method is proposed in [8], which is preserving the connectivity of pixels in a neighborhood. The "flippability" is identified by using 3×3 moving windows centered at the pixel. A two-layer blind binary image authentication scheme was proposed in [9], in which the first layer was for the authentication and the second layer was used for locating the tampering locations. A morphological transform domain is used in [10] to hide data for authentication purposes. A reversible data hiding scheme is proposed in [11], where run-length (RL) histogram is modified to embed data. The sequence of black RL and white RL was combined to get RL couple. Such couples were used to hide data. Review and analysis of exiting watermarking and steganography techniques are presented in [12] considering image processing in spatial and transform domain. The steganalysis method was proposed in [13] which checks whether the secret messages are embedded in black and white images using the steganographic techniques. The coding tables were prepared for data hiding and extraction considering HVS. It verifies the condition that whether the content of the image is obviously changed or not and tries to achieve higher hiding capacity than other techniques. In [14], coding tables for data hiding and extraction is constructed based on HVS in suitable blocks of binary images. Under the condition that the content of the image is not obviously changed, the proposed technique achieves higher hiding capacity than the state-of-the-art homogeneous techniques. In [15] the proposed scheme, a secret position matrix is designed to improve the hiding capacity which is capable of preventing the least distortion based on combination theory.

3 Proposed Data Hiding Method

In our proposed work, the binary image is scanned in four ways as follows:

1. Plain horizontal scan.
2. Plain vertical scan.
3. Block horizontal scan.
4. Block vertical scan.

As shown in Fig. 1a, b are images of size 3 × 3 while in c and d images size is 9 × 9 considering the block of size 3 × 3, representing four scanning ways. During the scanning, the count for bits was checked by which are matched with bit pattern of ciphertext. From these four scanning, one that is having maximum count is selected for embedding the data bits. At the sender side, first, take a binary image as an input and then scan that image in four scanning methods. The ciphertext is inserted into the original binary image using one of the scanning methods. Hiding details are stored in the header, and this header and its details are also inserted in the image for exact

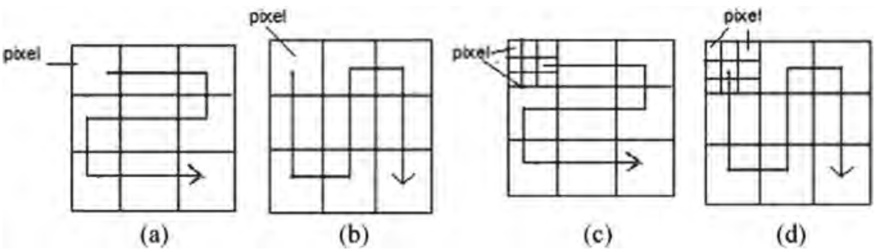


Fig. 1 Four-way scanning **a** plain horizontal scan, **b** plain vertical scan, **c** block horizontal scan, and **d** block vertical scan

extraction of hidden data. At the receiver side, cipher bits are extracted from the image through the header.

3.1 Data Hiding and Extraction Processes

Proposed work is represented in two processes, embedding, i.e., hiding data in an image and extraction, i.e., taking out hidden data. Figure 2 shows the proposed embedding process. In this, very first get the data to be hidden, and key is converted into bit streams, for example, if both data and key are text characters, then their ASCII codes or any binary representation can be considered. The data bit stream is encrypted using key bit streams applying encryption technique. In our proposed work, the XOR operation is applied to form cipher bit streams provided data using supplied key. Next, get the image which is in binary form in which data to be hidden as Y and scan it in four ways. During the scanning, map the cipher bit pattern with the pattern in the image. Note the count of pixels matching during mapping. The scanning method that gives maximum matching count is selected for embedding. The starting location of the mapped pattern in a selected scanning method is determined, and ciphertext is embedded at that location changing the unmatched pixels. The header is formed containing details of selected scan method, starting location where data is embedded, length of cipher text in cipher and actual length of cipher. This header length is identified, and the length and header are embedded from the end of the

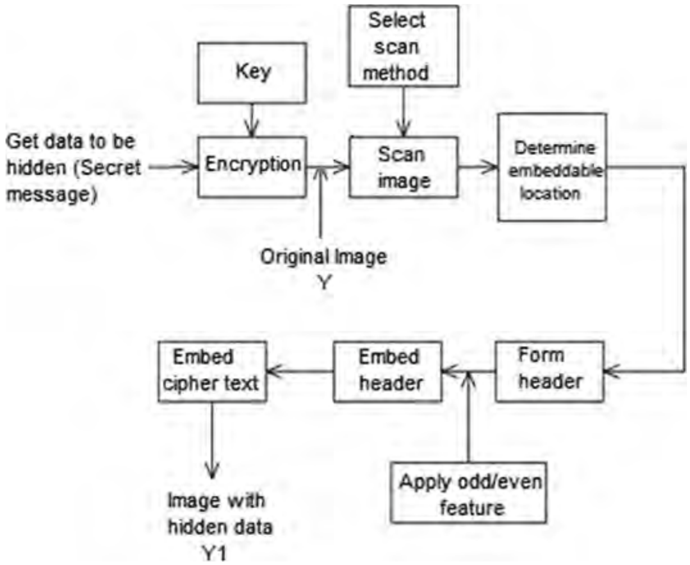


Fig. 2 Embedding process

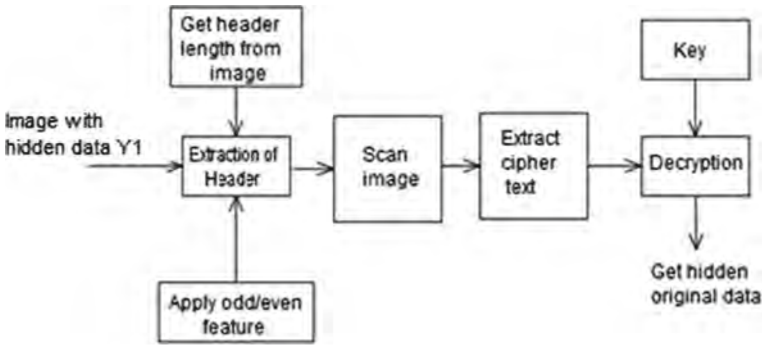


Fig. 3 Extraction of hidden data

image using the odd–even feature. The image with embedded ciphertext and header details is sent to the receiver nominated as Y1.

In the extraction process, the receiver receives image Y1 with hidden data as shown in Fig. 3 and extracts the header information using odd–even feature from the end of the image. From the extracted header, the information of scan method, starting location of cipher text, length of cipher text are extracted. The image is scanned from the starting location using the scanned method specified in the header, and ciphertext is extracted. The original data is extracted by decrypting the ciphertext using a key.

The header is formed containing the information as shown in Fig. 4.

The odd–even feature is used for embedding data [1]. To embed "1," block must contain an even number of the black pixels if not then 1 pixel is changed so that the total number of black pixels in that block will be an even number. Similarly, as shown in Fig. 5, to embed a "0," the number of black pixels in the block is enforced to an odd number.

Scan method	Start Index	Length of cipher in Header	Actual length of cipher

Fig. 4 Header format

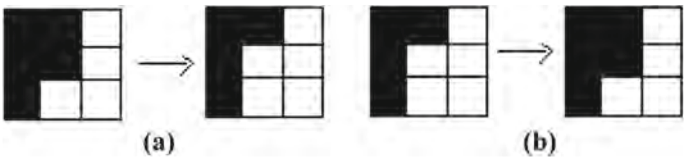


Fig. 5 Odd–even feature used to embed **a** "1" value **b** "0" value [1]

3.2 Mathematical Model

Let I be the image of size $m \times n$, where each pixel p_i is represented as a binary bit

$$I = \{p_1, p_2, p_3 \ldots p_{m\times n}\}$$

Let D represents data p bitstream of size p

$$D = \left\{d_1, d_2, d_3, \ldots d_p\right\}$$

Let K represents key bitstream of size q, and if $q < p$, then starting bits of K corresponding key characters are repeated to make it equal to p

$$K = \left\{k_1, k_2, k_3, \ldots k_q\right\}$$

Let C represents cipher bitstream of size p

$$C = \left\{c_1, c_2, c_3, \ldots c_p\right\}$$

Let S represents a set of four scanning methods

$$S = \{H, V, BH, BV\}$$

Set SMC is the scan match_count of cipher bits and image bits in all scanning methods

$$\mathrm{SMC}[1\ldots4] = \mathrm{match\_count}(\mathrm{C}, \mathrm{I}, \mathrm{S})$$

$$\mathrm{SMC} = \{\mathrm{HC}, \mathrm{VC}, \mathrm{BHC}, \mathrm{BVC}\}$$

MSC represents the maximum count from SMC

$$\mathrm{MSC} = \max(\mathrm{HC}, \mathrm{VC}, \mathrm{BHC}, \mathrm{BVC})$$

ES represents scan method S selected for embedding corresponding to MSC

$$\mathrm{ES} = \mathrm{S}(\mathrm{MSC})$$

3.3 Algorithms

The proposed system consists of main modules that handle scanning images, the formation of ciphertext and header, embedding of ciphertext, extraction of ciphertext, and decryption of the secret message. The algorithms used for embedding and extraction are mainly dependent on the scanning method selected.

(a) **Embedding algorithm**

1. Select a binary image with high resolution.
2. Create a cipher bitstreams by encrypting data to be hidden using key.
3. Scan the image in one of the following four ways:
   i. Plain horizontal scan.
   ii. Plain vertical scan.
   iii. Block horizontal scan.
   iv. Block vertical scan.
4. During the scanning, count the matching bits with the bit pattern of ciphertext.
5. Select the scanning format which is having the maximum count for matched bits.
6. Embed the ciphertext in the image.
7. Form the header as shown in Fig. 3.
8. Find out the header length and embed it from the end of the image in reverse.
9. Then embed header using the odd–even feature.

(b) **Extraction algorithm**

1. Receive an image with embedded data.
2. Extract the header first.
   i. First, take out the length of the header from the bits at the end of image.
   ii. If the numbers of black pixels in the block are even, then extract 1 else 0.
   iii. Then extract the header using length.
3. After extracting the header, get all information about the scan method and cipher such as start location and length. Using that, extract the cipher from the image.
4. Decrypt the cipher with a key. As the key is shared between sender and receiver, hidden data is extracted correctly.

Algorithm: Embedding_cipher (I,D,K)

```
1. C←D⊕K
2. for i←1 to 4
3.     for  j←1 to p  //
           SI←scan(I,i)  //scanning image with each scanning method
                          and get scan bit-stream using window shifting
           SMC[i] ←match_count(SI[j],C[j])
4. MSC←max(SMC)
5. ES←scan_method _selected(MSC)
6. for  j←1 to p  //
           SSI←scan(I,ES)  //scanning image with selected scanning
                            method and get scan bit-stream
           if(SSI[j] ≠ C[j])
              SSI[j]←C[j]       //flipping pixel for unmatched image
                                 bit  and  cipher   bit i.e. cipher bits
                                 embedded
```

Algorithm: Embedding_header (I,ES,starting_index,C)

```
1. H←append (ES,starting index, no_of_characters(C), length(C)
//H header bit stream, ES is represented in 2 bit for 4scan methods

2. for k←1 to n       //embedding n number of header bits
      B←scan(I,B)  //scanning image from end of the image in 3X3
                     Blocks
         if((black_pixel_count(B)%2)==0 & H[k]==0)
             flip(one_black_pixel_in_B)       //enforce block to have
                                                odd no. of black pixel to
                                                embed 0
         else if((black_pixel_count(B)%2)==1 & H[k]==1)
             flip(one_black_pixel_in_B)       //enforce block to have
                                                Even no. of black pixel to
                                                embed 1
```

Table 1 Experimental results

Image	Data to be hidden (secret essage)	Key	No. of cipher bits (n)	Selected scan method	No. of maximum match bits	No. of bits changed (m)	Ratio of change (n/m)
Img1.jpg 235 × 239	Attack	Alpha	32	Plain vertical scan	26	6	0.18
	Hello	abcde	30	Plain horizontal scan	25	5	0.15
Img1.jpg 150 × 140	Attack	Alpha	32	Plain horizontal scan	25	7	0.21
	Hello	abcde	30	Plain vertical scan	24	6	0.18

3.4 Experimental Setup

JAVA jdk1.8.0_101 and NetBeans 8.0.1 are used to execute the work with a set of binary images of variable sizes.

4 Results and Performance Analysis

As per Table 1 and graph in Fig. 6, result shows that very few bits changed to embed the cipher bitstream. Hence, this method introduces very limited distortion in the cover image (Fig. 7).

5 Conclusion

Most steganographic worked so far carried out on pictures has been color or gray-scale pictures, and limited work has been carried out on black and white pictures because black and white images are sensitive to changes. The four scan methods are introduced in which the image is scanned in four different ways to find out the match bits in the ciphertext. By selecting the scan method with maximum match bit count, the data is embedding in the image. From the experiments, it can be achieved that the minimum distortion in the image is selected with the assistance of the scan method with a maximum match bit count. Further, a header is formed by having the information as selected scan method, length of cipher text, starting index, and

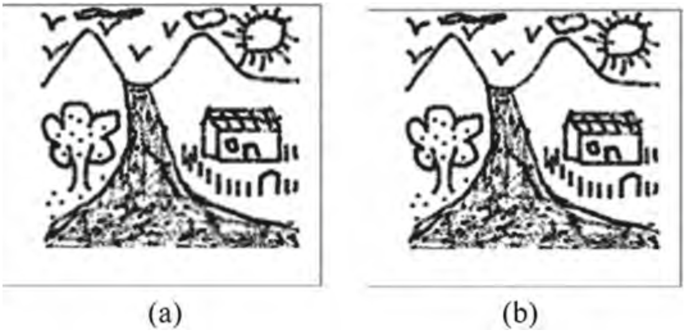


Fig. 6 Sample Img1.jpg 235 × 239 **a** original image and **b** image after embedding 32-bit data cipher bits and header details

ciphertext so that accurate information can be extracted. If the image size is small, the amount of the black and the white pixels would be less. On the contrary, if the image size is large, a large amount of pixels can be achieved. To select the proper scan method such that there should be less distortion in image and also have enough capacity for the embedding information.

6 Future Work

In our method, only 3 * 3 block scheme and considering the block size 3 * 3 which is fixed one (FB). There is another scheme to partitioning blocks interlaced block scheme (IB). For the interlaced block scheme, any two vertically or horizontally neighboring blocks share. The interlaced block scheme is used to partition the blocks of variable size, i.e., 4 × 4 IB or any other size. More pixels can be chosen to increase the probability of a block to be embeddable.

If more number of blocks is available, then it is possible to embed more data. For block horizontal scan and for block vertical scan, we can number the block, respectively, as shown in Fig. 8.

To perform the inter-block scanning as shown in Fig. 9 is to increase the match bit count, and thus, efficiency of embedding is also increased.

Scanning can be performed in any order. To obtain the maximum match bit count and to reduce the distortion in image, scanning is operated in horizontal block and vertical block.

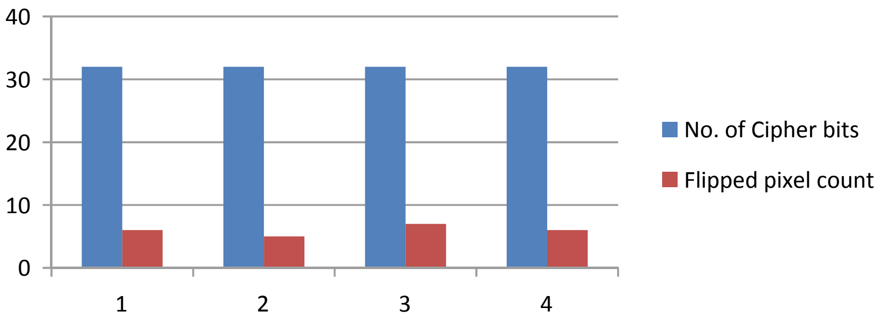


Fig. 7 Comparative graph of the number of cipher bits versus flipped pixel

1	2	3
4	5	6
7	8	9

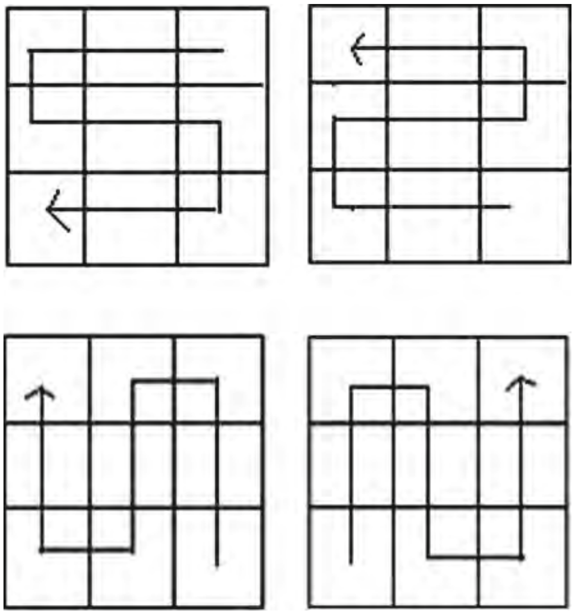
(a)

1	4	7
2	5	8
3	6	9

(b)

Fig. 8 a Block horizontal numbering and b block vertical numbering

Fig. 9 Various ways for block horizontal and block vertical scanning



References

1. M. Wu, B. Liu, Data hiding in binary images for authentication and annotation. IEEE Trans. Multimed. **6**(4), 528–538 (2004)
2. G.J. Chhajed, V.S. Inamdar, V.Z. Attar, Steganography in black and white picture ımages, in *2008 Congress on Image and Signal Processing, CISP 2008, IEEE Computer Society*, pp. 141–144. 978-0-7695-3119-9/08. https://doi.org/10.1109/CISP.2008.626
3. G.J. Chhajed, S.A. Shinde, Efficient embedding in B&W picture images, in *2nd IEEE International conference ICIME 2010* (2010), pp. 525–528. 978-1-4244-5265-1/10
4. Q. Mei, E.K. Wong, N. Memon, Data hiding in binary text document. Proc. SPIE **4314**, 369–375 (2001)
5. Y.C. Tseng, H.-K. Pan, Data hiding in 2-color images. IEEE Trans. Comput. **51**(7), 873–878 (2002)
6. H. Lu, X. Shi, Y.Q. Shi, A.C. Kot, L. Chen, Watermark embedding in DC components of DCT for binary images, in *Proceedings, IEEE Workshop on Multimedia Signal Processing* (2002), pp. 300–303
7. H.Y. Kim, R.L. de Queiroz, Alteration-locating authentication watermarking for binary images, in *Proceedings of International Workshop Digital Watermarking, 2004* (Springer, Berlin, Heidelberg, 2005), pp. 125–136
8. H. Yang, A.C. Kot, Pattern-based date hiding for binary images authentication by connectivity-preserving. IEEE Trans. Multimed. **9**(3), 475–486 (2007)
9. H. Yang, A.C. Kot, Binary image authentication with tampering localization by embedding cryptographic signature and block identifier. IEEE Signal Process. Lett. **13**(12), 741–744 (2006)
10. H. Yang, A.C. Kot, S. Rahardja, Orthogonal data embedding for binary images in morphological transform domain—a high capacity approach. IEEE Trans. Multimed. **10**(3), 339–351 (2008)
11. G. Xuan, Y.Q. Shi, P. Chai, X. Tong, J. Teng, J. Li, Reversible binary ımage data hiding by run-length histogram modification, in *International Conference on Pattern Recognition* (IEEE, 2008), pp. 1651–1653. 978-1-4244-2175
12. G.J. Chhajed, K. Deshmukh, T. Kulkarni,Review on binary image stegnography and watermarking, in *International journal in computer Science and Engineering (IJCSE)*, vol. 3, no. 11 (2011), pp. 3545–3651, ISSN: 0975-3397 (Citation: 18)
13. B. Feng, J. Wenga, W.L. BeiPeic, Steganalysis of content-adaptive binary image data hiding. J. Visual Commun. Image Rep. **46** (Elesvier, 2017), pp. 119–127
14. Wu. Nan-I, M. Hwang, S. Hwang, Taiwan Development of a data hiding scheme based on combination theory for lowering the visual noise in binary images ScienceDirect. Elsevier J. Displays **49**, 116–123 (2017)
15. W. Ding, Y. Wang, Data hiding in binary image with high payload. Arab. J. Sci. Eng. **43** (Springer, 2018), 7737–7745. https://doi.org/10.1007/s13369-018-3130-5

Deep CNN-Based Fire Alert System in Video Surveillance Networks

P. J. Sunitha and K. R. Joy

Abstract In the modern world, technology has unimaginably impacted our lives. Advancement in daily usage of gas, petrol, and electricity made a drastic increase in fire accidents. There is a need for a robust fire alert system for improving surveillance in the early stages. Existing systems provide a large number of false alarms and low accuracy. They depend solely on smoke or temperature sensors, which require proximity and human involvement for conforming to the fire. To overcome these deficiencies, a deep CNN-based fire alert system is proposed in the surveillance networks. The main segment of this system is deep CNN for fire detection, which improves the prediction accuracy in comparison with the traditional handcraft methods. WEMOS D1 mini and GSM help to monitor and control fire at the early stages. The proposed system is CCTV-oriented, which reduces the extra need for sensors and provides a large surveillance area. In this proposed system, CNN takes feeds from different cameras and all the videos were processed simultaneously. The video frame is taken and preprocessed one by one, and then given to the trained model, which predicts fire or no fire. If a fire is detected, then this information will be sent to the concerned authority by uploading the details on the web using the GSM. The experimental results show that the proposed CNN-based FACP system depicts accurate prediction in indoor and outdoor applications. Using emerging technologies, the proposed system is very robust and highly efficient.

Keywords Convolutional neural network · Deep learning · Fire detection · NodeMCU · GSM · Fire extinguisher

1 Introduction

Nowadays, surveillance is a major concern for human lives. Smart cameras improve surveillance by detecting various abnormal events such as fire, accidents, and medical

P. J. Sunitha (✉) · K. R. Joy
Department of Electronics and Communication Engineering, Sahrdaya College of Engineering and Technology, Thrissur, Kerala, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_47

emergencies at the early stages. Thus, the concerned authorities can be informed without losing valuable time for taking appropriate actions.

Fire is an anomalous event, and it can spread over a vast area and can destroy things in a matter of time. It can cause houses, trees, and many other things to burn into ashes. Wildfire often destroys large areas of forests. Most of the time, fire outbreaks unexpectedly and spreads over a large area within a very short time. Fire detection at an early stage is very important for surveillance networks and is a very interesting area for researchers. Till now, the two broad categories for fire detection include:

(1) Traditional fire alarms
(2) Vision sensor-assisted fire alarms.

Traditional fire alarm systems [1, 2] work based on sensors that require proximity for activation. They cannot predict the intensity of the fire nor the object which seized fire. These limitations open a door for optical sensors. It has a lot of advantages considering the conventional handcraft methods. Optical sensors [3–7] do not need any human intervention and have better coverage. Despite these benefits, the existing issues within this system include the intricacy of the scenes under observation, lighting issues, and shabby frames. Considering the above issues, there is a need for a better fire monitor and control system.

Fire alarm control panel (FACP) is a device used to monitor and control the fire alarm system [8]. FACP receives input from fire detectors and controls all fire notification subsystems connected to this network. Thus, it helps in the early fire rescue process and also helps in improving safety (Fig. 1).

Two types of FACP systems available are:

(1) Conventional FACP system
(2) Addressable FACP system.

Conventional systems notify the occurrence of fire by a change in electrical current, and they never give proper information about the location of the fire. Addressable systems give the exact address of the detector which identified fire. The proposed system is an addressable type, which gives the identifying number of each CCTV camera.

In fire alert systems, the preeminent one is the fire detector [9, 10]. CNN-based fire detection is the major part of this system. This article proposes a computationally efficient CNN architecture. It utilizes smaller convolutional kernels and contains no dense and fully connected layers. This gives much better accuracy and efficiency in comparison with the existing handcraft methods [11]. In this article, today's most emerging deep learning technology is recommended for fire detection. In comparison with the traditional methods, this method gives better accuracy and reduces the extra need for a large number of sensors.

Fig. 1 FACP system

2 Related Works

Traditional fire detection methods are point-based sensors, which depends on smoke particle sampling, atmosphere temperature sampling, and relative humidity sampling. These have some issues regarding applicability. The major issue is related to proximity and less coverage area. To overcome these limitations, the image-based methods were used to improve the accuracy in prediction that are more effective.

Yuan [12] proposed an unmanned aerial vehicle (UAV)-based forest fire detection and tracking method. Forest fire monitoring and detection mission are composed of three stages:

(1) Fire search
(2) Fire confirmation
(3) Fire observation.

The lab color model is used for fire segmentation, and the blob counter approach is used for object tracking.

Khan [13] proposed an automated firefighting system with smoke and temperature detection. The firefighting system incorporates smoke and temperature sensors for fire detection. If the data exceeds a certain threshold, the corresponding information will be sent to an alarm and a water spraying nozzle to extinguish the fire. The microcontroller controls these processes.

Bayoumi [14] focused on real-time fire detection and notification system based on computer vision. It identifies color and motion by image processing. A notification system helps to inform the concerned authority.

Azmil [15] presented a fire detection and monitoring system based on wireless technology, which had an Arduino controller, buzzer, smoke sensors, and image sensors. The camera converts the input image into an electronic signal and transmits it to a wireless receiver. If the received data exceeds a certain threshold, the alarm automatically turns on.

Gupta [16] proposed an automatic fire alert system. This system consists of a microcontroller, RF module, GSM, IR, and temperature sensors. Google Map identifies the location where the fire is affected. In this system, the RF transmitter sends the information to the receiver. At the receiver, the microcontroller processes the data and computes the affected area. Then, the fire brigade receives the number of extinguishers required for processing.

The existing literature transpired through mainly depends on the color features of the scene, smoke, or temperature sensors. Color feature-based fire detection systems lack efficiency due to misclassification by color like objects, varying lighting conditions, etc. Sensor-based systems have less coverage area and need proximity for activation. Regarding the precedent issues in the existing technologies, there is a need for a highly precise, effective fire alert system.

Our CNN-based FACP system overcomes these limitations of the existing systems by increased coverage area and accurate prediction. This article proposes an emerging deep CNN-based FACP system for fire detection and control (Fig. 2).

3 Overview of the System

The proposed CNN-based FACP system mainly consists of five main parts: CCTV, CNN, Wi-Fi, GSM, and fire extinguisher. CCTV cameras placed in different locations of the building will capture the images, frame by frame, and these different real-time videos are transmitted to the digital video recorder (DVR) from where the videos are taken simultaneously by CNN, which was pre-trained. CNN detects fire at an early stage and informs the details via the Wi-Fi module and GSM to the concerned authority by uploading the details on the web and sending SMS (Fig. 3).

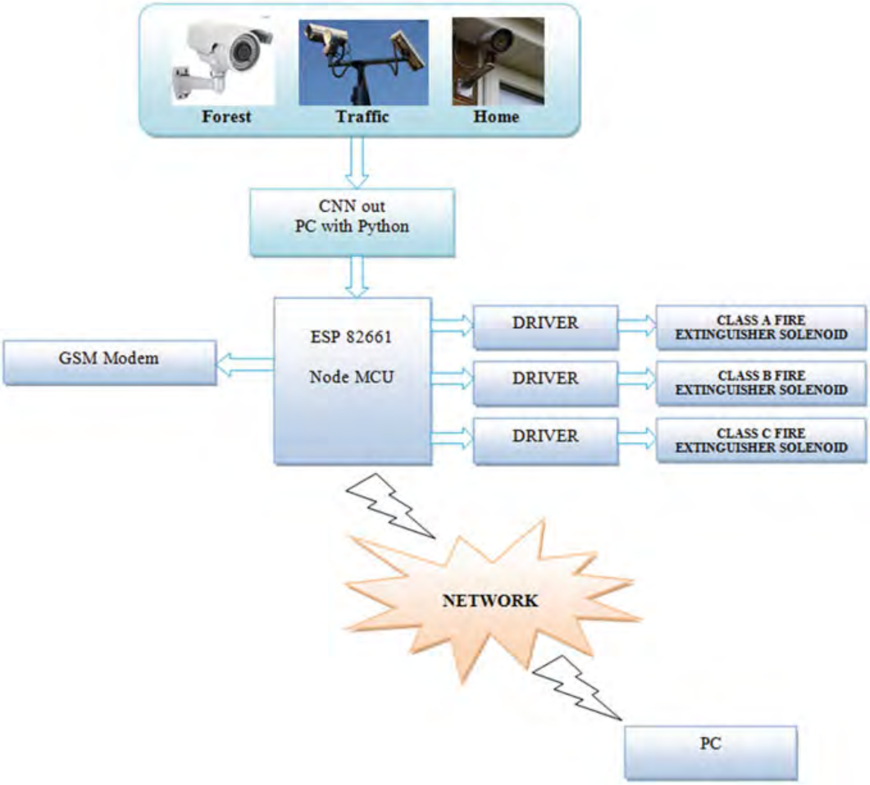


Fig. 2 Block diagram of the proposed system

4 Detailed Design

4.1 CCTV System

At present, surveillance is most important for human lives. CCTV cameras or surveillance cameras are used to monitor home, business, and other types of properties. Surveillance cameras allow us to remotely monitor a scene and playback recorded images. This will help system assessment and validation for emergency responders. CCTV system provides surveillance for extensive areas.

In this closed-circuit system, it transmits the signal through a restricted path only, unlike broadcast television. Some examples of different CCTV cameras suitable for surveillance applications are dome CCTV cameras, bullet CCTV cameras, pan tilt and zoom (PTZ) cameras, infrared/night vision CCTV cameras, and network/IP CCTV cameras (Fig. 4).

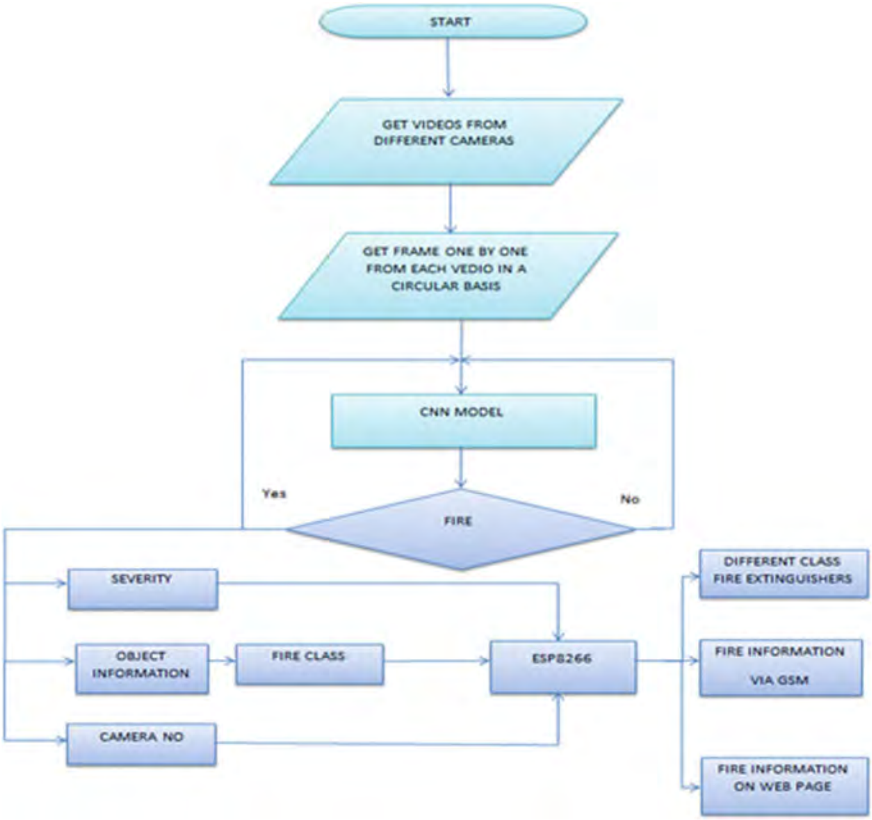


Fig. 3 Flow diagram of the proposed system

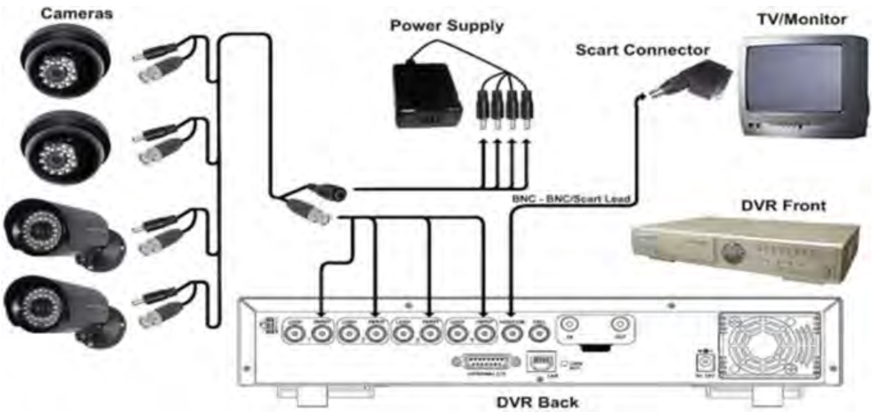


Fig. 4 CCTV system

4.2 Convolutional Neural Network

The convolutional neural network (CNN) includes a class of deep learning neural networks. They are frequently used to analyze visual imagery and are continuously working behind the scenes in image classification. Image classification inputs taken are images, and it outputs a particular class. CNN works by extracting features from images. This eliminates the necessity for manual feature extraction. The features are not trained but they are learned while the network trains on a given set of images. This makes deep learning models extremely accurate and efficient for computer vision-based tasks. CNNs learn feature detection through tens or hundreds of hidden layers of data. Each layer increases the intricacy of the learned features and provides maximum accuracy for calculations.

4.2.1 CNN Architecture

CNNs are made up of multiple layers. Generally, CNN has four layers, convolutional layer, ReLU layer, pooling layer, and fully connected layer.

The base of CNN is a convolutional layer, and this layer refers to the filtering process. CNN-based fire detection systems in real-world surveillance networks are the milestone of this proposed system. This article proposes computationally efficient CNN architecture based on [17]. It utilizes smaller convolutional kernels and contains no dense and fully connected layers. This gives much better accuracy and efficiency in comparison with the existing handcraft methods. When considering the works, this proposed system developed a custom trained CNN model.

The model is designed with four convolutional layers, three max pooling layers, and one average pooling layer. The output of this average pooling layer is given to the softmax classifier to find the probabilities of the output, fire, or no fire. In deep convolutional layer 1, it applies 64 3×3 filters with ReLU activation function, these 64 feature maps obtained from this convolutional layer are then given to the pooling layer: This will perform max pooling with a 3×3 filter and stride of 2 (which specifies that pooled regions do not overlap). Later, this reduced out is given to Convolutional Layer 2: It applies 128 1×1 filters, with ReLU activation function then again Pooling Layer: this also performs max Pooling with a 3×3 filter and stride of 2. Again to convolutional layer 3: It applies 256 3×3 filters, with a ReLU activation function. Its output is given to pooling layer: again, performs max pooling with a 3×3 filter and stride of 2. Convolutional layer 4**:** It applies 512 3×3 filters, with ReLU activation function. Once again in pooling layer: Again, it performs average pooling with a 3×3 filter and stride of 2. Then, the dropout regularization rate of 0.2 (where the probability of 0.2 that any given element will be dropped in training) was taken; then the last dense layer: Softmax classifier is used.

CNN extracts images from various cameras in this proposed system. All videos are processed at the same time. Each video frame is taken one by one and applied for preprocessing and then given to the trained model which predicts fire or no fire.

Algorithm 1: Fire Zone and Object Detection

Input: CNN model, video from cameras, AlexNet.
Step 1: Place the camera in fire-prone areas.
Step 2: Connect the camera and capture videos.
Step 3: Read frame by frame of the chosen videos
Step 4: Forward propagate these samples through CNN model.
Step 5: No action will be taken until a fire is found in each video frame.
Step 6: If a fire is detected, take feature maps.
Step 7: Calculate the mean activation map for F8, F26, and F32.
Step 8: Apply binarization.
Step 9: Segment fire area from the frame.
Step 10: Detect the object from the segmented area using AlexNet architecture.
Output: Fire detection and object in which fire is affected caught information.

If there is a fire, then that particular image is again given to the model. Feature maps are took from the first two layers. Considering feature maps 8, 26, and 32, calculate their mean activation function. Regarding the greatest and least pixel values in the mean activation map, obtain the average value. Considering the average, largest, and mean activation values, set the threshold value. Values above the threshold are assigned 1 and below the threshold as 0, thus binarizing the image. Computing the number of 1 s in the binarized image provides information regarding the severity of the fire. Subtracting the binarized image from the reshaped input image gives the ZOI. Algorithm 1 describes how the CNN fire localization and object detection were performed. The proposed system uses AlexNet for object detection.

AlexNet is a convolutional neural network which has 5 convolutional layers and 3 fully connected layers. It acquired the image classification in 2012. In this article, a pre-trained AlexNet model [18–22] is being used, which shows the output as one of the 1000 classes of the softmax layer. The AlexNet model receives the preprocessed image as 227 by 227.

4.3 Wi-Fi Module

In this segment, the WEMOS D1 Mini is used as an ESP8266-based miniature wireless 802.11 (Wi-Fi) development board. Programming in this module is easy, as this module incorporates a built-in micro-USB interface, enabling the module to be coded directly from the Arduino IDE [23, 24] (Fig. 5).

This board supports both serial and OTA programming. In our proposed system, CNN out is given serially to this module. As this project uses the ESP-8266 chip, it acts as a controller for the hardware section. The CNN out conveys the information to the concerned authority via GSM and web server. The proposed system also includes fire extinguishers, which activates corresponding to the class of fire.

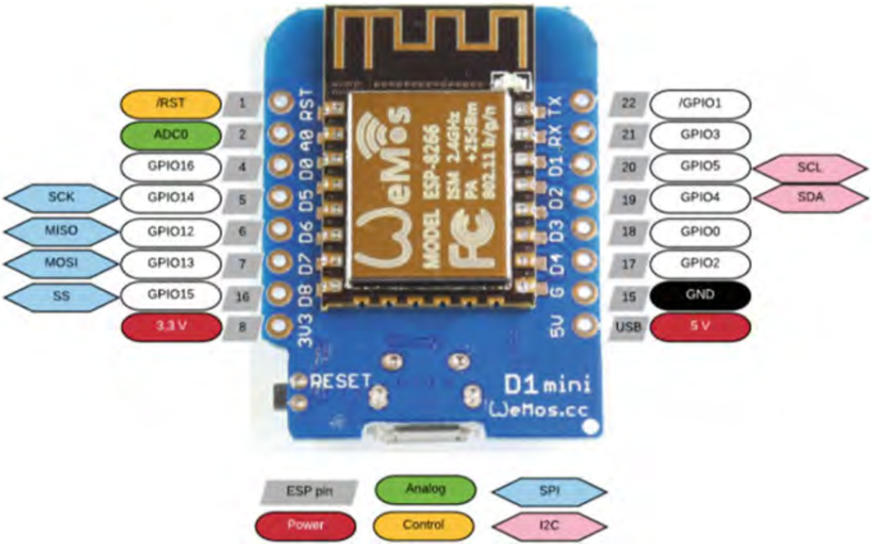


Fig. 5 Wi-Fi module

JavaScript AJAX is used to update the data without any refresh. AJAX is Asynchronous JavaScript and XML, which is applied for creating fast and dynamic web pages. AJAX uses a combination of (Fig. 6):

(1) XMLHttpRequest object
(2) JavaScript/DOM

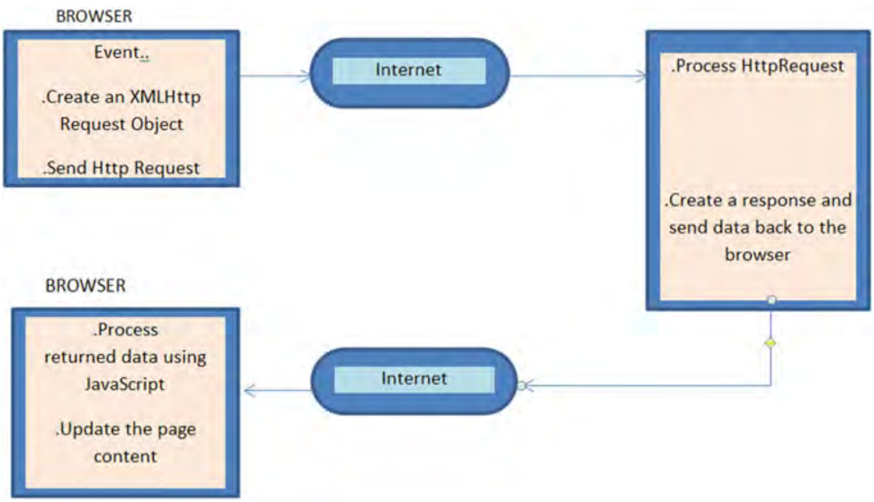


Fig. 6 Working of ESP32 AJAX

(3) CSS
(4) XML.

In this system, the Wi-Fi module has mainly three functions:

(1) To switch on the corresponding fire extinguisher (classes A, B, or C)
(2) Convey the information to the concerned authority via the GSM module
(3) To put the information on the web, so that the persons connected on the Wi-Fi network will get the fire information.

4.4 GSM Modem

In security-based systems, the SM is used in conjunction with Wi-Fi to ensure more secure connections.

SIM800L module is a minute cellular GSM/GPRS breakout board that allows GPRS communication, sending and receiving messages and calls. This module maintains a quad-band GSM/GPRS network, which is available for both GPRS and remote SMS data transmission. A tiny footprint is included to have quad-band frequency support, and the low price makes it perfect for any project that demands long-range connectivity.

The device uses a serial communication method, so it communicates with almost every controller via the UART port. After giving a voltage supply, the module boots up, searches for the cellular network, and connects automatically.

4.5 Fire Extinguisher

Fire extinguishers are designed to undertake specific types of fire. A single fire extinguisher does not work on all types of fire. To ensure true safety, the right extinguisher is confirmed to be installed in the right place. The three classes of fire extinguishers are:

(1) Class A extinguishers are used to tackle fires that contain solid materials as main components.
(2) Class B extinguishers are designed for situations involving flammable liquids like oil, gasoline, and grease.
(3) Class C extinguishers are for use on burning gases like natural gas or petroleum gas.

5 Results and Discussion

CNN-based FACP system provides better performance. Considering the existing fire alert systems, the FACU uses better technologies and more efficient methods.

Table 1 Comparative study of existing and present system

Technologies used	Previous systems					Present system	Results
	Ref. [1]	Ref. [12]	Ref. [13]	Ref. [14]	Ref. [15]		
Smoke sensor	No	Yes	Yes	Yes	Yes	No	False alarms and not suited for outdoor
CNN	Yes	No	No	No	No	Yes	Reduce false positives and more accurate prediction
Wi-Fi module	No	No	No	Yes	No	Yes	Improves rescue operation
GSM	No	No	No	No	No	Yes	Improves rescue operation
Fire extinguisher	No	No	Yes	No	No	Yes	Improves rescue operation

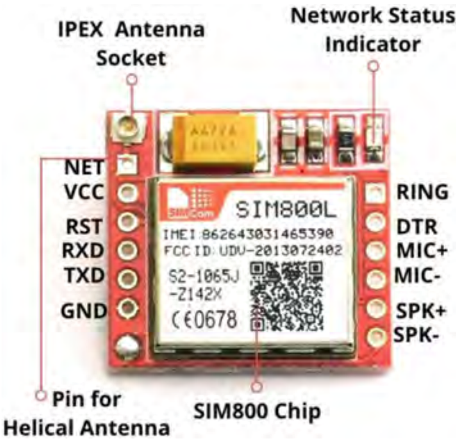
Table 1 shows a comparative study of existing systems and our CNN-based FACP system. This proposed system gives information about the object which got fire. According to that information, classify the fire, and corresponding fire extinguishers can be used.

5.1 CNN Out

Initially, CNN is trained with fire and no fire images, and for this purpose, it uses 1200 fire images and 1500 no fire images, thus a total of 2700 images in the training dataset. The dataset is taken from https://www.kaggle.com/. CNN is trained using this dataset and is validated. Here, 70% of data is used for training purposes, and the remaining 30% is used for validation. This CNN model is used for fire detection. The accuracy of prediction is checked during this section which gives an accuracy of 99.61%. Figure 7 shows CNN's prediction for an object which received fire. Figure 8 shows the CNN layers and the parameters used for the prediction of the severity of the fire and the object which is on fire. Here, the proposed model shows CNN error of 0.38%, recall: 0.99544, f1 score: 0.9969. Confusion matrix (Fig. 9):

$$\begin{bmatrix} 388 & 1 \\ 3 & 655 \end{bmatrix}$$

Fig. 7 GSM module



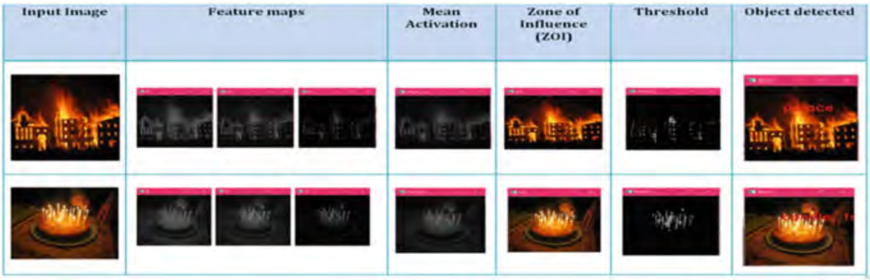


Fig. 8 Fire localization using algorithm 1

5.2 *Simultaneous Access of CCTV Out*

In this section, a different CCTV camera was focused to check the prediction accuracy of our fire detection module by considering both the indoor and outdoor situations. In our proposed system, four CCTV videos were given as input simultaneously and each video specifies various fire or no fire situations. Here, three outdoor and one indoor situation were examined and Python programming is used as a tool. It includes all videos in an array, and then the CNN model takes each video frame one by one.

It is compared to other conventional handcraft methods as shown in Fig. 8, and fire prediction and object detection from videos were processed. This CCTV-based surveillance system provides very accurate results. It is analyzed frame by frame and gives an improved prediction rate of the fire and its intensity.

```
Loaded model from disk
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 222, 222, 64)      1792
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 74, 74, 64)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 74, 74, 128)       8320
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 24, 24, 128)       0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 22, 22, 256)       295168
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 7, 7, 256)         0
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 5, 5, 512)         1180160
_________________________________________________________________
average_pooling2d_1 (Average (None, 1, 1, 512)         0
_________________________________________________________________
dropout_1 (Dropout)          (None, 1, 1, 512)         0
_________________________________________________________________
flatten_1 (Flatten)          (None, 512)               0
_________________________________________________________________
dense_1 (Dense)              (None, 224)               114912
_________________________________________________________________
dense_2 (Dense)              (None, 2)                 450
=================================================================
Total params: 1,600,802
Trainable params: 1,600,802
Non-trainable params: 0
_________________________________________________________________
fire
(224, 224, 3)
(1, 222, 222, 64)
(1, 222, 222, 64)
severe fire
candle, taper, wax light
severe fire in candle, taper, wax light
```

Fig. 9 CNN model for input candle image

5.3 Information Passing Using WEMOS Mini

In this section, the hardware section which contains NodeMCU, GSM, and fire extinguisher was checked for processing. This module receives CNN out, and when there is no fire, there is no action.

If there is a fire, then the data corresponding to the severity, class, and the camera details is uploaded on the web.

Figure 10 shows how the web server displays the data. ESP8266-based web server is more responsive and can access it with any device with a browser in our local network. Here, AJAX helps with updates in real-time data. In this project, the CNN model verifies real-time video outs, and it updates with no delay. At the same time, the concerned authority receives information via GSM, and it also activates the corresponding fire extinguishers as shown in Fig. 11. This project can de intended for a real-time application, which activates fire extinguisher according to each class of fire (Fig. 12).

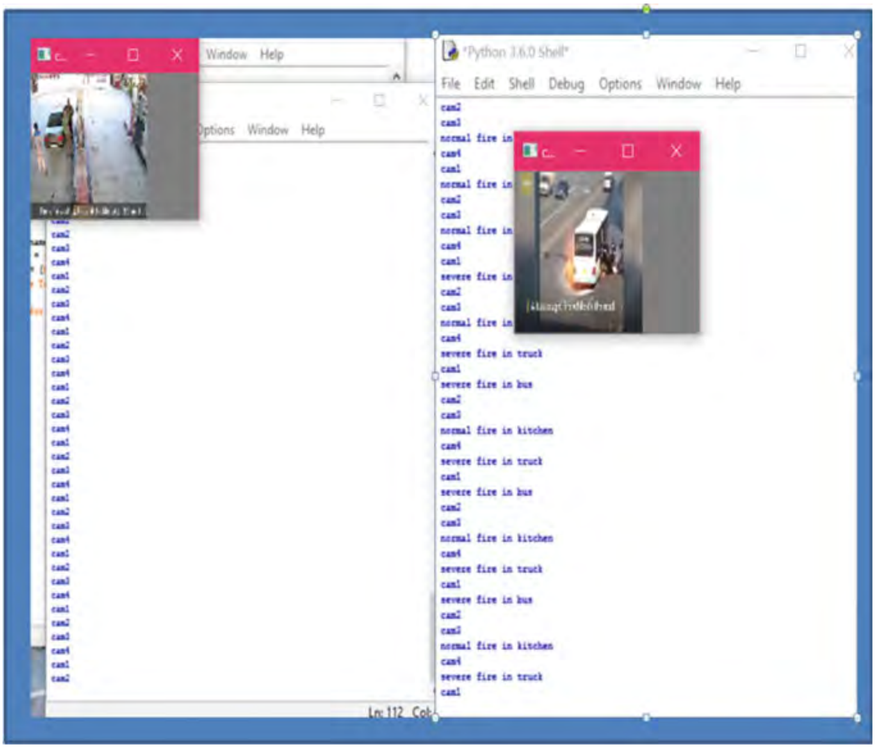


Fig. 10 Fire localization using CNN

Fig. 11 Web server data in case of fire

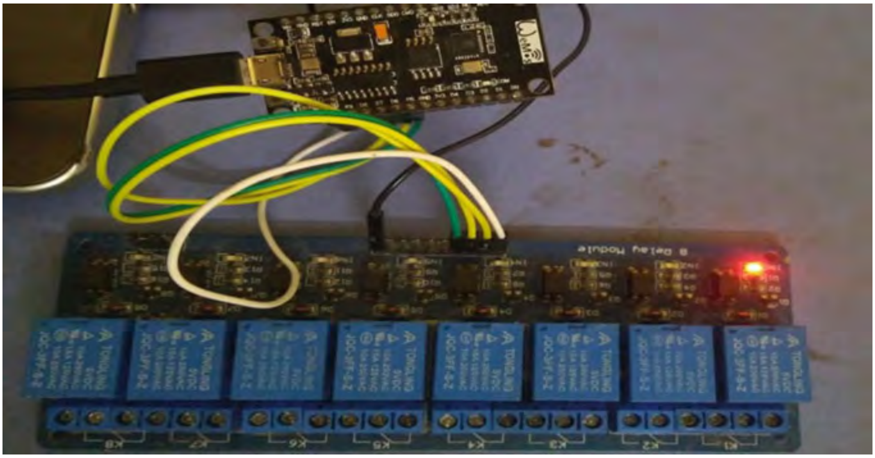


Fig. 12 Activating relays corresponding to extinguisher class

6 Discussion

According to the experimental results, the overall system performance gives very high accuracy. Conventional handcraft methods will give a large number of false positives in the case of fire detection. This CNN-based fire detection and control system is cost-effective and highly precise. Because the proposed system provides information about the object that caught fire, the fire can be classified and the appropriate action can be taken by the appropriate authority.

This system gives information on fire at the early stages, which will be very effective in surveillance networks. Future work related to this project is to include smoke detection, which may cause a fire. Using satellite images further improves fire detection rates in various regions like a forest fire or fire in remote regions. Using this technology, it is possible to protect the forest before it causes major destruction to plants and animals.

7 Conclusion

The embedded processing capabilities of smart cameras have improved the performance of CCTV surveillance systems. The concatenation of CCTV and CNN gives a great improvement in the early fire detection process. Because CNN is a new image classification technology, it reduces the number of false alarms, which was a common problem with existing sensor-based fire alert systems. As this proposed system gives an object which seized fire, from that information, the class of fire can be distinguishable, and the corresponding fire extinguisher will be switched on without any time delay. This will be very suitable for indoor applications. In the outdoor environment, the information is conveyed through GSM, and the web server also enables the fire rescue process without any delay. So considering the overall performance and the benefits, our proposed system based on CNN and CCTV is best suited for early fire detection and control, respectively. The experiments were conducted using two benchmark datasets and verify the feasibility of the proposed system for deployment in real CCTV networks. The CNN models provide reasonable accuracy for fire detection and localization, its size, and reduced rate of false alarms, and the system can be helpful to disaster management teams in controlling fire disasters on time, thus avoiding huge losses.

References

1. Z. Kun, S. Hu, J. Li, Automatic fire alarm system based on MCU. in *Proceeding of IEEE, International Conference on Electrical and Control Engineering* (2010), pp. 517–520
2. Z. Chaczko, F. Ahmad, Wıreless sensor network based system for fıre endangered areas. in *Proceedings of the Third International Conference on Information Technology and Applications (ICITA'05)*
3. P. Foggia, A. Saggese, M. Vento, Real-time fire detection for video surveillance applications using a combination of experts based on color, shape and motion. in *2015 IEEE Transactions on Circuits and Systems for Video Technology* (2015)
4. L. Ma, K. Wu, L. Zhu, Fire smoke detection in video images using Kalman filter and Gaussian mixture color model. in *IEEE 2010 International Conference on Artificial Intelligence and Computational Intelligence* (2010)
5. Thou-Ho (Chao-Ho) Chen, Ping-Hsueh Wu, Yung-Chuen Chiou, An early fire-detection method based on ımage processing. in *2004 International Conference on Image Processing (ICIP)*
6. P. Morerio, L. Marcenaro, Carlo S. Regazzoni, Early fire and smoke detection based on colour features and motion analysis. in *2012 19th IEEE İnternational Conference on İmage Processing*

7. M. Mueller, P. Karasev, I. Kolesov, Optical flow estimation for flame detection in videos. IEEE Trans. Image Process. **22**(7), (2013)
8. Fire alarm systems: devices and components, www.tedsystems.com
9. A. Khan, A. Sohail, U. Zahoora, A.S. Qureshi, A Survey of the Recent Architectures of Deep Convolutional Neural Networks. Published in Artificial Intelligence Review. https://doi.org/10.1007/s10462-020-09825-6
10. S. Frizzi, R. Kaabi, M. Bouchouicha, Jean-Marc Ginoux, Eric Moreau, Farhat Fnaiech, Convolutional neural network for video fire and smoke detection. in *IECON 2016–42nd Annual Conference of the IEEE Industrial Electronics Society*
11. T.S. Kumar, Video based traffic forecasting using convolution neural network model and transfer learning techniques. J. Innov. Image Process. (JIIP) **2**(03), 128–134 (2020)
12. C. Yuan, Z. Liu, Y. Zhang, UAV-based forest fire detection and tracking using ımage processing techniques. in *IEEE 2015 International Cnference on Unmanned Aircraft Systems (ICUAS)* (references)
13. Mohammad Jane Alam Khan, Muhammed Rifat Imam, Jashim Uddin, M.A. Rashid Sarkar, Automated fire fighting system with smoke and temperature detection. in *Proceedings of IEEE 7th International Conference on Electrical and Computer Engineering* (20–22 Dec. 2012), pp. 232–235
14. Sahar Bayoumi, Eham A. Sobky, Moneerah Almohsin, Manahel Altwaim, Monira Alkaldi, Munera, A real-time fire detection and notification system based on computer vision. in IEEE Software and Knowledge Engineering Research Group from the Malaz Research Centre (RGP-VPP157) (16–18 Dec. 2013), pp.1–4
15. Muhammad Salihin Ahmad Azmil, Norsuzila Yaacob, Khairul Nizam Tahar, Wireless fire detection monitoring system for fire and rescue application. in *Proceedings of IEEE 11th International Colloquium on Signal Processing and Its Applications* (6–8 March 2015), pp. 84–89
16. Sarita Gupta, Ajay Mudgil, Prashant Bhardwaj, Mahendra Gupta, Design and development of automatic fire alert system. in *2016 8th International Conference on Computational Intelligence and Communication Networks (CICN)*
17. Muhammad Khan, Jamil Ahmad, Zhihan Lv, Efficient deep CNN-based fire detection and localization in video surveillance applications. in *2018 IEEE Transactions on Systems, Man, And Cybernetics Systems*
18. Luis Tobìas, Aurelien Ducournau, François Rousseau, Gregoire Mercier, Ronan Fablet, Convolutional neural networks for object recognition on mobile devices: a case study. in *2016 23rd International Conference on Pattern Recognition (ICPR)* (2016)
19. Ronnel R. Atole, Karen Michelle A. Alarcon, Garry P. Dacillo, A Pre-trained deep convolutional neural network for the detection of tungro in rice plants. RR Atole, KMA Alarcon, GP Dacillo—academia.edu
20. D. Tomè, F. Monti, L. Baroffio, L. Bondi, M. Tagliasacchi, S. Tubaro, Deep convolutional neural networks for pedestrian detection
21. S. Ren, K. He, R. Girshick, X. Zhang, J. Sun, Object detection networks on convolutional feature maps. IEEE Trans. Pattern Anal. Mach. İntell. **39**(7), (2017)
22. J.E. Espinosa, S.A. Velastin, Vehicle detection using Alexnet and faster R-CNN Deep learning Models. in *International Visual İnformatics Conference IVIC* (2017)
23. Dlnya Abdulahad Aziz, Websever based smart monitoring system using ESP2866 nodemcu module. Int. J. Sci. Eng. Res. **9**(6), (2018). ISSN 2229–5518
24. A. Sungheetha, R. Sharma, Real time monitoring and fire detection using ınternet of things and cloud based drones. J. Soft Comput. Paradigm. (JSCP) **2**(03), 168–174 (2020)

Implementation of Chassis Number Recognition Model for Automatic Vehicle Identification

Khine Htoo and Myint Myint Sein

Abstract Document image recognition is required for information retrieving, searching, editing, and reporting of image text. There are several kinds of document images such as text, image, figure, signature, car license plate, chassis number, and so on. In this research, the SVD-based Document Image Recognition System is proposed for Automatic Vehicle Identification Number (VIN identification). Canny edge detection algorithm and morphological processing are applied for text area extraction. For all the segmented characters, singular value decomposition (SVD) features are calculated for the recognition process. The system is implemented by C# and the result shows that the recognition method is feasible and easy to use for users, and it can be put into practice. It can be extended to the real-time vehicle examination and identification system by using mobile device thought over the network.

Keywords VIN identification · Singular value decomposition · Chassis number · Canny edge detection

1 Introduction

For a developing country like Myanmar, robust vehicle identification and detection system are required for regular administrative tasks such as license registration, vehicle valuation assessment, traffic violation cases, to identify the theft cars, etc. [1]. In this research, an automatic vehicle identification system is developed based on a singular value decomposition approach. Especially, the type of vehicle is classified from the image of the vehicle identification number (VIN). The vehicle identification number (VIN) is a unique code, including a serial number, used by the automotive industry to identify individual motor vehicles, and VIN is sometimes called the car's chassis number. There are a lot of character recognition systems and optical character recognition systems (OCR) and chassis number recognition systems [1–12].

K. Htoo (✉) · M. M. Sein
University of Computer Studies, Yangon (UCSY), Myanmar

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_48

OCR-based chassis number recognition system using artificial neural networks is proposed by Shah et al. [2]. To reduce the characters to single-pixel thickness, normalization and morphological thinning processes are performed on the individual image. The backpropagation algorithm of artificial neural network (ANN) is used for the character recognition process. They only tested for digits of VIN number and are not implement the chassis identification. Karthick et al. [3] discuss the five phases of the problem in handwritten text recognition. Prewitt and Canny edge detection methods are used for text detection from images and Otsu's method is used for isolating the information from its background. Natei et al. [4] presented text extraction from documents image, the combination of the two powerful methods such as connected component and edge-based method were utilized. The following Table 1 illustrated the investigation of the existing character reorganization approach [1–10] and proposed approach. This comparison contains the type of input image, method for feature extraction and character recognition and implementation of chassis number

Table 1 Investigation of existing and propose approach

Ref. No.	Input image	Feature /edge extraction	Recognition	Chassis implementation	Other-application and remark
1	Five characters of the VIN	Sobel	ANN	No	Only recognize the digits
2	Text	Prewitt and Canny	Otsu	–	OCR
3	Document	Canny	Connected component	–	Text extraction
4	Handwritten text	Canny	HMM	–	Myanmar text document
5	Passport	Color segmentation	Gaussian	–	Auto data entry of passport
6	Passport		Hopfield Neural Net	–	Passport holder identification
7	Bank check	Zoning	HMM	–	Extraction of Payee's Name
8	Malayalam character	Binarization	SVD	–	Recognize the malayalam character
9	Malayalam character	KNN	SVD and Radon projection	–	Recognize the malayalam character
10	Text	Slide window	Connected component, deep learning	–	Review of scene text detection and recognition
	17 characters of VIN	Canny	SDN	Yes	Proposed approach

identification, and so on. By our proposed system, the information of a car can be obtained from chassis number.

In many existing vehicle identification systems using VIN, the 17 characters of digits and capital letters are needed to type manually for checking with the database. It will be affecting the result of vehicle identification when typing errors occur. To overcome this problem, the proposed system can implement directly to VIN document images for the vehicle identification process. To accurately identify cars in all the places where necessary, the system of the vehicle identification number is developed using singular value decomposition (SVD) features.

2 VIN and Locations

A VIN is a unique identification number for every car manufactured and it is constant throughout the lifetime of the car. That does not include O, I, and Q in 17 alphanumeric characters of VINs (to avoid confusion with numerals 0, 1, and 9). Figure 1 illustrated the locations of VIN in a car and an explanation of the VIN is shown in Fig. 2, respectively [12].

Fig. 1 Locations of VIN

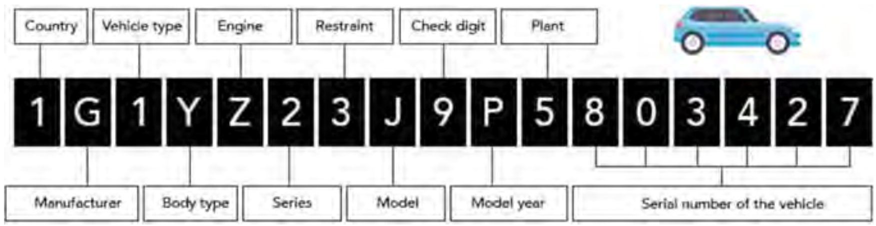


Fig. 2 Explanation of vehicle identification number (VIN) [12]

3 System Overview

The system consists of four main stages: text area detection, image segmentation, VIN/ chassis number recognition, and vehicle identification. The Canny edge detection algorithm is used to detect the text area of the input document image [4, 5]. For image segmentation, the morphological dilation operation and labeling method are used for text area segmentation and individual character segmentation followed by histogram methods. Then, the system extracts the features from the individual character of the VIN/Chassis number by using SVD [13]. The Euclidean distance measure is used for the matching process in character recognition [14]. Finally, the system will classify the recognized characters for vehicle identification.

The system is generated by C# programming language and the result shows that the recognition method is feasible and easy to use for the user. Overview flow of the SVD-based vehicle identification system is illustrated in Fig. 3.

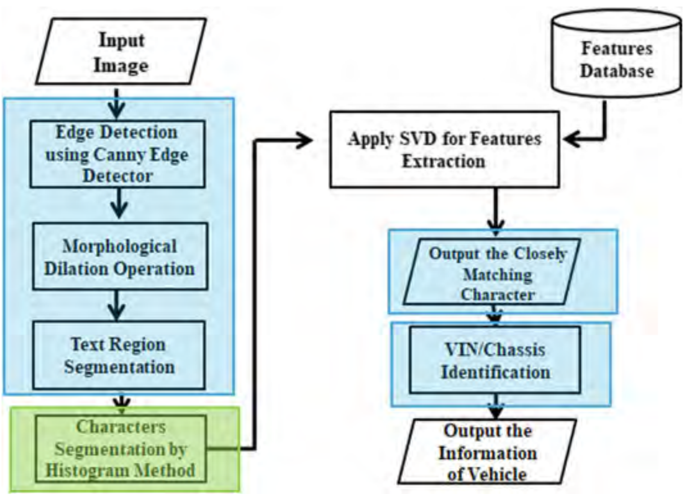


Fig. 3 Overview of the system

4 Implementation

The image of the vehicle identification number (VIN) is acquired for an input image, firstly. Then, the possible VIN areas are extracted from the input image. After removing the unnecessary part and cropping the VIN area, the characters are extracted by object pixel histogram method. For all the segmented characters, singular value decomposition (SVD) features are calculated for the recognition process. The Euclidean distance measure is used to find the most similar VIN in the recognition process.

4.1 Character Area Extraction

The first and most important part of the VIN recognition system is the extraction of VIN area from the image. The Canny edge detector is used to fine the edges of all objects in the input image as Fig. 4b. Detected objects that represent VINs are located side by side horizontally. Morphological processing tools are very useful for

(a) Acquired Image (b) Edge Image

(c) Dialeted Image. (d) Connected Image.

Fig. 4 Result of labeling connected objects feature extraction and removing the unwanted parts from the image. To detect the text area, the dilation operation is provided for obtaining the connected text region. The main purpose of dilation horizontally is to receive the VIN area as a big object. After dilation that image horizontally, all the adjacent objects are connected as shown below in Fig. 4c.

Obviously, so many candidate possible areas are detected and the most possible VIN area needs to be cropped among them. To do that first all the detected objects are labeled and consequently, the size and shape properties of detected areas are accounted to choose and crop the VIN area from the input image. Figure 4d described the connected image. The big size, the ratio of the width, and the height of the labeled area can give information about the VIN area [15–18].

4.2 Character Segmentation on VIN Area

After labeling the connected region process, the VIN area is extracted from the source VIN image. And the cropped image is binarized to distinguish background and text objects. Each character object is easily extracted by using the object count histogram illustrated in Fig. 5. As shown in the figure, the segmented character image can be saved to the database by the desired name tag.

The algorithm of the character segmentation is as follow:

1. Binarization to input VIN image
2. Created objects count (White pixels) histogram along the width of the image
3. Find the starting and ending points of each adjacent non-zero area
4. Crop each character according to the result of step (3)

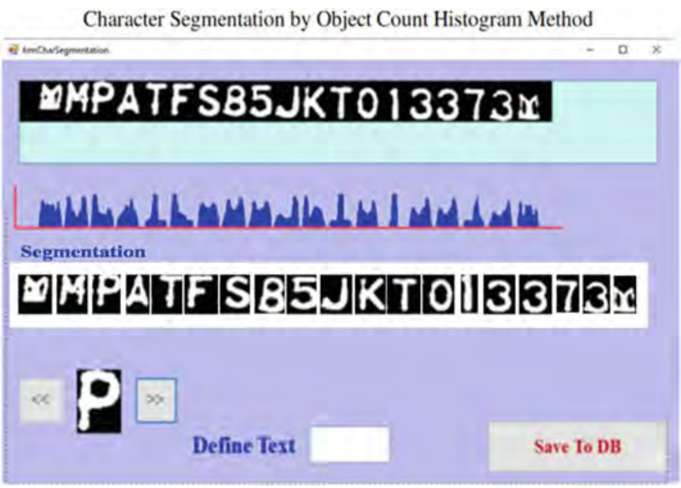


Fig. 5 Illustration of the character segmentation

5. Temporary save these cropped character images in local disk for feature calculation process.

5 Feature Extraction and Classification

The feature extraction is performed from each cropped character image. In this work, singular value decomposition (SVD) features are used for the recognition process. The special feature of SVD is that it can be performed on any real (m, n) matrix.

Given any $m \times n$ matrix A, an algorithm to find matrices U, $\sum$, and V such that

$$A = U \sum V^T \tag{1}$$

where Matrix U is an $m \times m$ orthogonal matrix, $U = [u_1, u_2, \ldots, u_r, u_{r+1}, \ldots, u_m]$. Column vectors u_i, for $i = 1, 2, 3, \ldots., m$ form an orthonormal set:

$$u_i^T u_j = \delta_{ij} = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases} \tag{2}$$

Matrix V is an $n \times n$ orthogonal matrix, $V = \left[v_1, v_{2,}, \ldots., v_{r,}, v_{r+1,}, \ldots, v_{m,}\right]$ Column vectors v_i, for $i = 1, 2, 3, \ldots., n$, form an orthonormal set:

$$v_i^T v_j = \delta_{ij} = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases} \tag{3}$$

The matrix $\sum$ is an $m \times n$ diagonal matrix with singular value (SV) on the diagonal.

The operation procedure is as follow:

1. Create a data matrix, say ‘A’
2. Apply SVD, $A = U \sum V^T$
3. Compute feature descriptor matrix, $H = \sum V^T$
4. Given a test vector ‘a’ and compute $a^T U$
5. Repeat $a^T U$ to match the dimension of $H = \sum V^T$
6. Find the minimum of $\text{repmat}\left(a^T U, \text{size}(H, 2)\right) - H$ to get the desired class.

The classification is carried out using SVD features and Euclidean distance to extract VIN number information. All the feature vectors are stored with the definition of each character. Totally, about 20,000 image features are stored for the experimental work. The two-dimensional Euclidean geometry, the Euclidean distance between two points $p = (p_1, p_2)$ and $q = (q_1, q_2)$ are defined as:

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2} \tag{4}$$

where $d(p, q)$ is the distance between point p and point q.

Distance "d_i" is selected as the most similar one among all the characters in the database, where i is the position that makes "$d_{i\ i}$" minimum value among all the feature vectors,

$i = \overline{1, N}$, $N =$ Total feature vectors in DB.

6 Experimental Results

According to the extracted VIN result, VIN information is given as the result of this system. A VIN is a unique identification number for every car manufactured and it is constant throughout the lifetime of the car. It is composed of 17 alphanumeric characters which do not include O, I, and Q. A total of 12,000 VIN number images are collected for the generated database. The camera type is not fixed for chassis image acquisition. Several chassis images are acquired for database creation. Some acquired images are shown in Fig. 6. The segmented character images are stored in the database. The images of typing characters A to Z except for O, I, Q, and digit 0 to 9 are included in the database.

The camera captures the VIN images and processes the image such as resizing, grayscale converting, and edge detection. And morphological operation dilation process is used to extract the connected text region. After skewing and small region removing process, the text region is cropped from the original input chassis image. The histogram method is applied for segmentation and matching of the segmented image with the database is performed by modified SVD. The result is shown in Fig. 7.

The testing result of a VIN is also shown in Table 2 for the required decision-making processes. An experiment is carried out to know the accuracy of the recognition of the proposed system. Results show that almost all the characters are recognized from the image. But in a very rare case, the system may give the wrong result. The resulting Table 3 describes that about just 5 training features are enough to know the accurate result in this system.

Fig. 6 Some acquired images

Fig. 7 Illustration result of the character recognition

Table 2 Result of actual VIN number information

VIN Number = MPATFS85JKT013373		
1	Brand	ISUZU
2	Vehicle type	Double cab
3	Model year	2019
4	Country of origin	Thailand

Table 3 Accuracy rate

	Number of features in DB	Accuracy (%)
1	1 feature	95.6
2	3 features	97.7
3	5 features	99.9
4	10 features	99.9
5	20 features	99.9

7 Conclusion

A VIN recognition system is developed for the need of actual requirement in Myanmar. According to the implementation and results, the system can truly crop the interested VIN area and segmented each character. This proposed model can be applied to vehicle number recognition, person identification, passport, and other security verification work. Furthermore, a car information management system can easily be added for more information about the inspected vehicle. Currently, recognition processes are only available offline. The real-time vehicle examination and identification system should be done by using the available network and Web camera in the future. Web applications and mobile applications need to be created to use everywhere in Myanmar.

References

1. Mya Mya Thinn, Myint Myint Sein, Implementation of automatic data entry system using Myanmar passport. in *The Proceedings of the Third Malaysian Software Engineering Conference, MySEC'07* (3–4 Dec 2007)
2. Parul Shah et al., OCR-based Chassis-number recognition using artificial neural networks. in *ICVES* (2009)
3. K. Karthick, K.B. Ravindrakumar, R. Francis, S. Ilankannan, Steps involved in text recognition and recent research in OCR; a study. Int. J. Recent Technol. Eng. (IJRTE) 8(1), (2019). ISSN: 2277-3878
4. K.N. Natei, J. Viradiya, S. Sasikumar, Extracting text from image document and displaying its related information. IJERA (2018)
5. San San Mon, Myint Myint Sein, Recognition of Myanmar handwriting text based on hidden Markov model. in *Proceedings of the Fifth International Conference on Computer Application (ICCA2007)* (Myanmar, 8–9 Feb 2007), pp. 329–333
6. Mya Mya Thinn, Myint Myint Sein, Automatic data entry of passport for security system. in *Proceedings of the Fifth International Conference on Computer Application* (Myanmar, 8–9 Feb 2007), pp. 407–410
7. Nang Aye Aye Htwe, San San Mon, Myint Myint Sein, Recognition on user-entered data from Myanmar bank cheque. in *Proceedings of the Sixth International Conference on Computer Application* (Yangon Myanmar, 14–15 Feb 2008), pp.114–118
8. R. Anil, Arjun Pradeep, E.M. Midhun, K. Manduca, Malayalam character recognition using singular value decomposition. Int. J. Comput. Appl. (0975–8887), **92**(12), (2014)
9. Manjusha et al., Experimental analysis on character recognition using singular value decomposition and random projection. Int. J. Eng. Technol. (IJET) 7(4), 1246–1254 (2015)
10. Han Lin et al., Review of scene text detection and recognition. Arch. Comput. Meth. Eng. **27**, 433–454 (2020)
11. Chaitanya R. Kulkarni, Ashwini B. Barbadekar, Text detection and recognition: a review. Int. Res. J. Eng. Technol. (IRJET), **04**(06), (2017)
12. Sanjyot Kajale, Vehicle Identification Number (VIN)–a survey. Int. J. Innov. Res. Comput. Commun. Eng. **6**(9), (2018)
13. Zecheng Kuang, *Singular-Value Decomposition and its Applications* (Department of Mathematics, University of California, San Diego, CA, 2012)
14. Leo Liberti, Carlile Lavor, Nelson Maculan, Antonio Mucherino, Euclidean distance geometry and applications. SIAM Rev. (2012)
15. Text reading from images. https://www.codeproject.com/Questions/660834/Text-reading-from-image
16. Image Processing in C#. https://www.c-sharpcorner.com/article/image-processing-in-C-Sharp
17. Image Processing using C#. https://www.codeproject.com/Articles/33838/Image-Processing-using-C
18. https://www.researchgate.net/publication/316629965_Novel_Method_for_Energy_Consumption_Billing_Using_Optical_Character_Recognition

Identification of Artificial Body Marks and Skin Disorder Marks Using Artificial Neural Network Approach

Dayanand G. Savakar, Danesh Telsang, and Anil Kannur

Abstract Artificial and skin disorder body marks like a tattoo and white patch are used to identify the victims, such as suspected, unidentified bodies like in mass death in a plane crash and the tsunami it is a very complex situation to identify the body. This article presents the identification of artificial and skin disorder body marks like tattoos and white patches. Active contour segmentation is used to segment the image. There are 28 features extracted from each tattoo and white patch image, where 18 are color features, 4 are texture features, and 6 are shape features. The artificial neural network is used to classify artificial and skin disorder body marks, and classification accuracies obtained 93.6%. The designed algorithm works based on the features that are being extracted. The designed and proposed algorithm within this paper incorporates such kind of techniques to identify the artificial and skin disorder body marks.

Keywords Active contour · Classification · Tattoo · Segmentation · White patch · Image processing

1 Introduction

Tattoos include letters, pictures, names of beloved drawn on some parts of the human body as a mark of love and devotion. For thousands of years, this is practiced. Sometimes, they are used for protection against caste conversion, medicine, social status symbol, or as a sign of religious or spiritual belief. Egyptologists have identified the practice of tattooing during ancient times in Egypt, Africa, Greek, Rome, Japan, and Asian and American countries. During ancient times, tattooing was done by using a long needle piercing the skin with natural color ink, usually green. This

D. G. Savakar · D. Telsang (✉)
Department of Computer Science, Rani Channamma University, Belagavi, Karnataka, India

A. Kannur
Department of Computer Science and Engineering, Nagarjuna College of Engineering and Technology, Bengaluru, Karnataka, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_49

was very painful [1]. But in modern times, tattooing has become a hobby and more sophisticated equipment is used to create beautiful pictures and designs on the human body. This simplified and less painful process has attracted the modern community to have tattoos. Usually, motorcyclists, sailors, and members of criminal gangs were having tattoos. But nowadays, tattooing has become a fashion in young males and females and has considered it as a mark of reputation. White patch/vitiligo is a natural body mark. A person develops vitiligo due to medical problems like lack of immunity, insufficient vitamins, and the death of pigmentation cells. White patches of 1–4 cms are found on the face, hands, and legs, and some persons may get on other parts of the body also. It is a disease caused due to destruction of pigment cells, but most experts believe that it is an autoimmune condition in which the body's immune system attracts and destroys cells within the body by mistake [2, 3].

The identification of different body marks is useful in the medical field. It is also very useful in the case of mass death due to floods, plane crash, tsunami, and earthquake. Tattoos (artificial body mark) and white patches (skin disorder mark) are useful to identify the victim [4–6]. The algorithm has been proposed for the identification of artificial body marks and skin disorder marks such as tattoos and white patches. In a preprocessing stage to remove the hairs from the DULL RAZOR and to segment the object from the image, active contour is used. Based on the artificial neural network (ANN), the artificial and skin disorder body marks are identified. The detailed description of the proposed method is present in subsequent Sect. 3.

2 Review of Literature

The back-propagation neural network (BPNN) is used for the artificial neural network for recognition and classification of almost the same food grain images like cumin seeds, fennel seeds, mung beans, black gram, finger millet, mustard, soybeans, and black-eyed beans. The color and texture features are selected for classification, considering color and texture 18 and 27 features are extracted, respectively, and combining color and texture 4 and 5 features are extracted [7]. Artificial Neural Network for Identification of multiple grain image samples from tray images like green gram, alasandi, wheat, metagi, red gram has been considered to be an important step in identification and classification based on the color and texture features. In this proposed work, back-propagation neural network (BPNN) is used to classify and recognize the food grain image samples, here the color and texture features are selected for classification, considering color and texture 18 and 24 features are extracted, respectively, and combining color and texture 42 features are extracted [8].

The author has discussed the identification and classification of foreign bodies in mixed food grains like Groundnut, jawar, rice, wheat, and green gram for training and identification of the unknown types of grain mixed with foreign bodies, the color, and textural features are presented to the neural network. Here the author selected the color and texture features are for classification, considering 18 colors and 27 textures features extracted and the combined color and texture (total 45) features are extracted in this work and feedforward network ANN model and backpropagation algorithm for training were used [9] for identification and classification. In this article, the author discussed a computer-aided diagnosis system for vitiligo; here, vitiligo area scoring index (VASI) was used to assessing vitiligo [10].

In this work, the author used to match tattoos in the database and to retrieve similar tattoos using the content-based image retrieval (CBIR) approach. Throughout this study, a new active contour CBIR method has been introduced for tattoo segmentation, texture analysis, Haar wavelet decomposition, and hue saturation value histograms that are used for color representation at the end using a new approach glocal (glocal–local) image feature methods for testing the dataset [11]. The presence of foreign bodies in food grains will impact on recognition and classification of the largest food grain images for the sake of sorting here the author used the color and texture features which are selected for classification, considering color and texture 18 and 27 feature are extracted respectively and combined color and texture 4 and 5 features are extracted in this work feed-forward network ANN model and back-propagation algorithm are used for training [12].

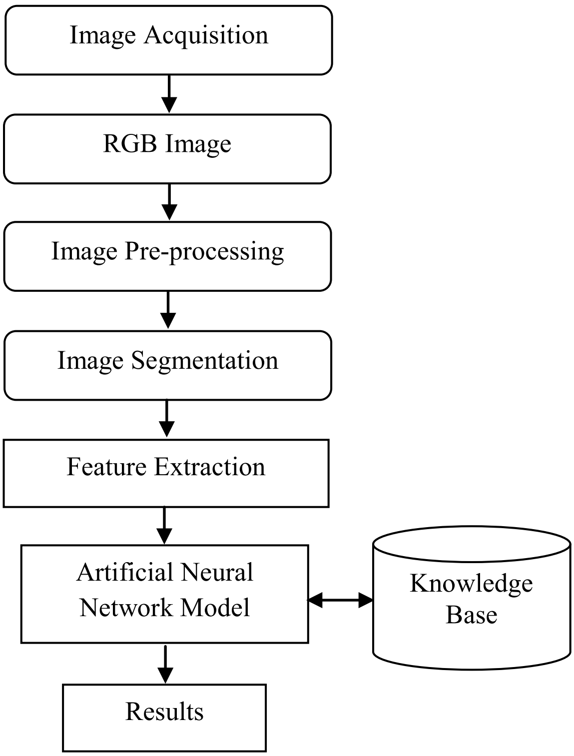
3 Problem Statement

One of the challenging jobs in the image processing is the identification of artificial and skin disorder body marks. In the modern era, the tattoo has become a fashion and everyone likes to have a tattoo on hand, leg, neck, and other parts of the body. These tattoos and white patches (skin disorder) are becoming essential tools in identifying persons who expired in the event of a plane crash and tsunami. To resolve this problem of identification of persons in mass death and culprits, the proposed method artificial neural network has been used. The proposed method steps are shown in Fig. 1.

4 Methodology

This section presents a detailed description of image acquisition, image preprocessing, segmentation, and feature extraction. Finally, an artificial neural network is used to classify the artificial and skin disorder body marks as explained in the subsequent section.

Fig. 1 Proposed system architecture



4.1 *Image Acquisition and Preprocessing*

It is a process of capturing an image from a camera, and it is the first step in the workflow sequence. The images are captured using canon 1300D digital camera model with 1.5× to 10× lens zooming capacity and keeping the object and the camera in a fixed position with maintaining clock word distance between them were in some of the images collected from an online source. In the preprocessing stage, unnecessary noise is removed from the image and resized too [300 400]. The DULL RAZOR software [13] is used to alteration and removes unnecessary hairs from images to improve the quality of the image, it is a medical imaging software, and it uses an algorithm the same as that of average filtering. The filtering technique also plays a major role in image processing to upgrade the images, sharpening, edge detection, and noise reduction. For convolution operation, spatial filtering is used as shown in Eq. (1).

$$S(x, y) = \sum_{m=-M/2}^{M/2} \sum_{n=-N/2}^{N/2} h(m, n) f(s - m, y - n), \tag{1}$$

where $h(m, n)$ is the Gaussian filtering mask of size $M \times N$.

4.2 Segmentation (Active Contour)

The active contour is a refinement of an object boundary, and the curve is formed by connecting their edged points. So, there is no clear idea about the boundary needed to connect those edge points to be formed by a curved. Later, the set of connected points moves to minimize a specified energy function. There is also another name for this process called a snake method. The active contour segmentaion flows along with the different perspectives on the image itself and then it comes down to convergence from the particular attribute of the contour itself, from this particular process it gets its name called an active contour.

The model is chosen as an active model for the segmentation process. In the first step, it should have an image that is extracted from its background layer, it is considered as energy function like gray level value, gradient and the initial step is segment a boundary by a general segmentation technique object with boundary has been found. The next step refines that boundary with the technique used to wiggle the snake; in this, it compares the pixel on each point on the boundary with energy calculated for the point in its neighborhood. The pixel on the boundary is extended to the neighborhood point that has the lowest energy and operate once in all points on the boundary after that repeat the iteration until the termination [14, 15]. The advantages of the active contour are computational efficiency and relative simplicity, and the disadvantage is decision criteria on initial points and its complexity of dependency. Active contour is used in various fields of segmentation based on the application. The tattoo and white patch images are segmented with several segmentation techniques in that active contour produced better results as compared to the other methods [16]. Hence in this work, active contour is used for segmenting an image as shown in Fig. 2.

Algorithm 1: Active contour (Region-based)-based image segmentation
Input: RGB image
Output: Color segmented image
Start:
Step 1: Read RGB image

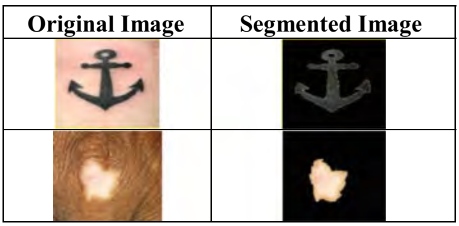


Fig. 2 Samples of the original image and segmented image using active contour segmentation technique

Step 2: Convert RGB to gray
Step 3: Create a mask
Step 4: Initialize number of iteration
Step 5: Cover the object of an image using the region and a mask value based on the number of iteration
Step 6: Set, background = 0 and foreground = 1 (covered area of an image is 1)
Step 7: Segment the covered area of an image
Step 8: Reshape the original image.

Stop

4.3 Feature Extraction

The image is segmented by active contour, where the features are extracted for the process identification of different body marks such as artificial marks like tattoos and skin disorder images like white patches which are depicted in Table 1. The color features are selected because the tattoo and white patch images are representing different colors. Texture feature is also extracted from the images based on the surface area of the image. Finally, shape features are extracted to identify the tattoo and white patch images based on shape area and shape solidity. There are a total of 28 features extracted from each image and stored in vector form as a feature vector. Table 1 represents the total number of features like color features (18), texture features (04), and shape features (06).

The detailed description of the color, texture, and shape features is discussed in the subsequent section.

4.3.1 Color Feature Extraction

In a color image, there are several components such as red, green, blue, hue, saturation, and value. Three features are selected from each component like mean, variance, and range. There are 18 color features extracted from each image based on the color of the tattoo and white patch image [17].

Algorithm 2: Color Feature Extraction
Input: Color image.
Output: 18 color features.
Start
Step 1: Color image separated from RGB components.
Step 2: Established the HIS components from RGB components using Eq. (2)-7
Step 3: From each RGB and HSI component, extracted range, variance, and mean
Stop.

Table 1 List of color, texture, shape features

Sl. No.	Color features
01	Red mean
02	Red variance
03	Red range
04	Green mean
05	Green variance
06	Green range
07	Blue mean
08	Blue variance
09	Blue range
10	Hue mean
11	Hue variance
12	Hue range
13	Saturation mean
14	Saturation variance
15	Saturation range
16	Value mean
17	Value variance
18	Value range
Texture feature	
19	Texture contrast
20	Texture correlation
21	Texture energy
22	Texture homogeneity
Shape feature	
23	Shape area
24	Shape convex area
25	Shape eccentricity
26	Shape solidity
27	Shape filled area
28	Shape equivDiameter

4.3.2 Texture Feature Extraction

Texture feature extraction refers to the appearance of the characteristic of an object given by the density, shape, size, arrangement, the proportion of its elementary parts of artificial and skin disorder body marks in a preliminary stage. The analysis of the tattoo and white patch image through the texture analysis then extracted four features from each tattoo and white patch image like contrast, correlation, energy, and homogeneity [17].

Algorithm 3: Extraction of textural feature
Input: Original image RGB
Output: 4 texture features
Start
Step 1: 24-bit input color image separated from RGB components, and obtain the gray-level co-occurrence matrices (GLCMs).
Step 2: Compute the co-occurrence matrix
Step 3: GLCM features (texture features) contrast, correlation, energy, homogeneity
Stop.

4.3.3 Extraction of Shape Feature

The tattoo and white patch features are extracted by different measures like shape area, convex area, eccentricity, solidity, filled area, and equivDiameter. The area of the object present in the image is measured by the shape area and measured area of the particular object as consider as a feature, stored in the feature vector. The other measures also extract the feature of the tattoo and white patch images based on their properties. There are six features extracted from each image [17].

Algorithm 4: Extraction of shape feature
Input: RGB image
Output: 6 shape features
Start
Step 1: RGB-to-gray conversation
Step 2: Detection of an object area from an image
Step 3: Shape features (shape area, shape convex area, shape eccentricity, shape solidity, shape filled area, shape equivDiameter)
Stop.

4.4 Artificial Neural Network

The artificial neural network is the most powerful network architecture in today's world. In this work, ANN [14, 15] is used to classify the skin disorder and artificial body marks such as white patches and tattoos based on color, texture, and shape features. There are three different layers with feed-forward architecture that are used for the formation of the neural network. The number of input features is considered as equal to the number of input layers. The outcome of the number of body marks is equal to the output layers shown in Fig. 3. From each image, 28 features are extracted in that 18 are color, 6 are shape, and 4 are texture, based on that white

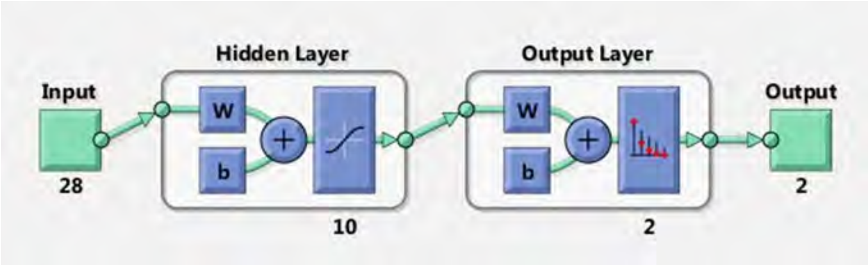


Fig. 3 Structure of ANN

patch and tattoo samples are recognized. The output layers have 2 nodes in all cases. The classification output is represented as a pattern vector of 2 bits Q (q1, q2) for two classes and is set to 1 and remaining bits to 0 s, the image sample belongs to the ith type of body images. The vectors Q1 (1 0), Q2 (0 1) represent white patches and tattoos, respectively. The number of hidden layers is calculated using the formula as shown in Eq. (2).

$$n = \frac{I + O}{2} + y^{0.5} \tag{2}$$

where

n number of nodes in the hidden layer
I number of input features
O number of outputs
y number of input patterns in the training set

4.4.1 Training and Testing

The identification of artificial and skin disorder body marks on different datasets like a tattoo and white patches is summarized as follows: Training and testing of the neural network are accomplished using body mark datasets, and these datasets are divided into two parts: First one is training and testing is the second one. Totally, 295 sample images were taken for the experiment, in that 145 images are tattoo and 150 images are a white patch. Here, 50% of the datasets are taken for training, 25% are taken for validation and 25% are taken for testing purposes. The ANN model performance for training and testing of the selected dataset is shown in Fig. 4, the blue color indicates the training, the red color indicates the testing, and the green color indicates the validation of the experiment. The graph shows the performance of the ANN model achieving the target with the number of iterations in a particular period.

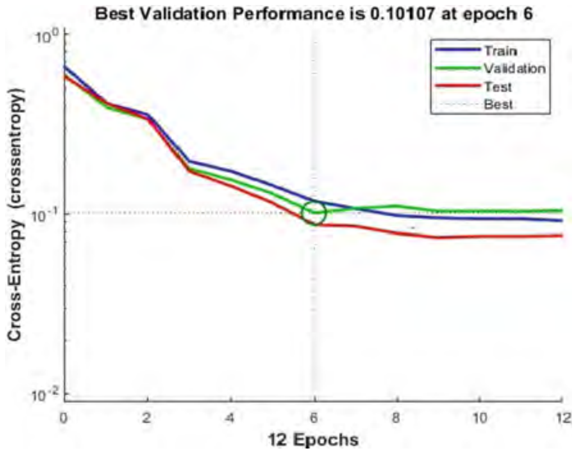


Fig. 4 Best validation performance of ANN

5 Result and Discussion

The experiment is evaluated on 295 images of which 145 contain tattoo images and 150 consist of white patch images. The total number of features extracted from each image is 28. The number of input layers is 28, the number of hidden layers is 10, and 2 output layers are set to experiment. Then, 50% of the dataset is used for training, the remaining 25% for validation, and 25% for testing. The proposed experimental results are shown in the confusion matrix, the prediction value of white patches is 91.9% true and failure case 8.1% out of 150 samples, whereas in tattoos 95.6% true and the failure case 4.4% out of 145 samples and the overall accuracy of the proposed method is 93.6% true and 6.4% failure out of 295 samples present in Fig. 5. The precision value is 0.96 and basically provides a positive predicated result by the model of the actual positive result. The precision is calculated using the formula as shown in Eq. (3).

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{3}$$

The recall value is 0.91, the total number of images retrieved is relevant images in the database called recall, and it is calculated using the formula as shown in Eq. (4).

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{4}$$

The f1 score value is 0.93, the accuracy measure is based on the weighted average of precision, and recall is calculated using the formula as shown in Eq. (4).

$$f = \frac{2 \times (\mathrm{Recall} \times \mathrm{Precision})}{(\mathrm{Recall} + \mathrm{Precision})} \tag{5}$$

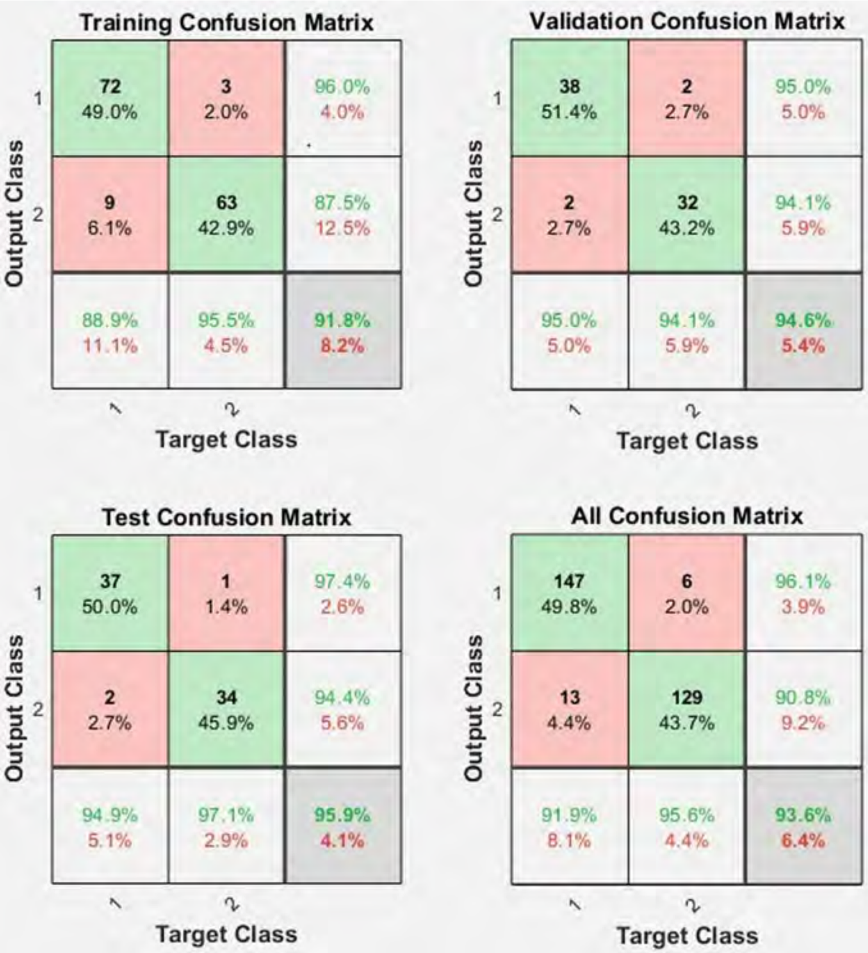


Fig. 5 Target class confusion matrix between class1 white patches and class2 tattoo

This is a commonly used graph that summarizes the performance of a classifier's overall possible thresholds. It is generated by plotting the true positive rate (*y*-axis) against the false positive rate (*x*-axis) as you vary the threshold for assigning observations to a given class. The blue color represents class one (white patch), and tan color represents class two (tattoo). The curves are presented in all ROC, and the performance of the proposed method is shown in Fig. 6.

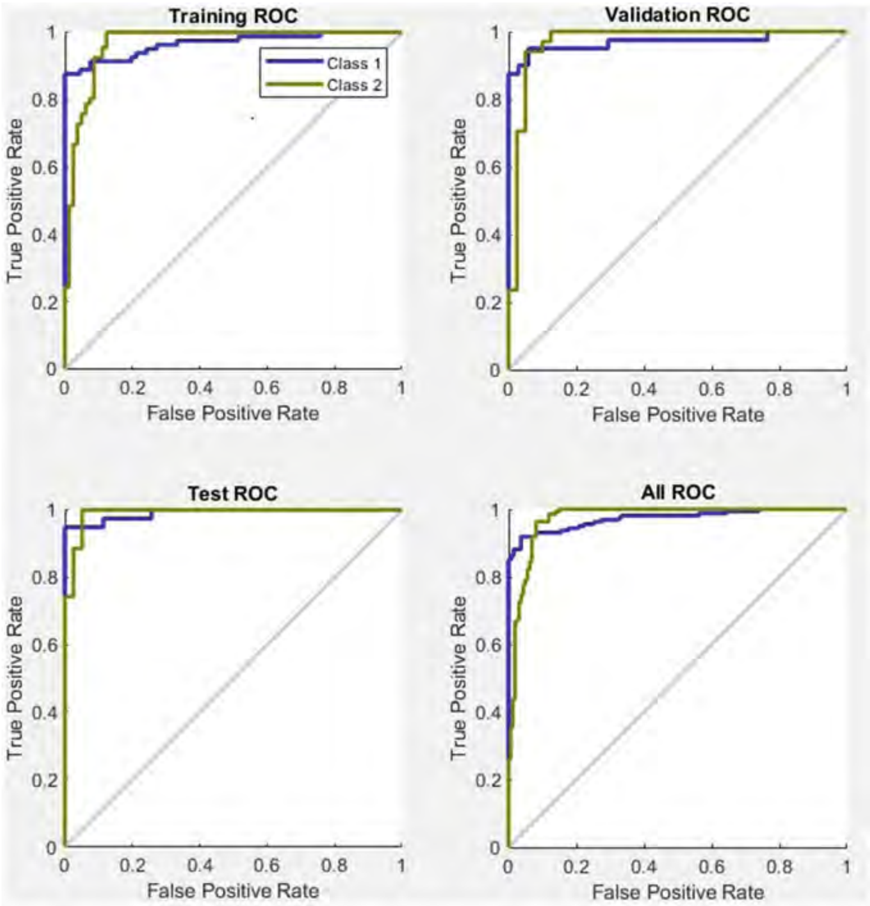


Fig. 6 Classifier accuracies for a selection of different white patches and tattoo classes

5.1 Experimental Results of the Work

The white patches and tattoo images are collected from various hospitals and Web sites, and the total number of images that are considered for the experiment is 295 images. These images were preprocessed with the following techniques like hair removal using DULL RAZOR software, and spatial filtering is used to reduce edge detection, noise reduction, and segmentation processes, and finally obtained the segmented images after the preprocessing technique. The ANN toolbox is used from MATLAB Platform, in that random is for data division and scaled conjugate gradient for training and cross-entropy for performance, and the numbers of epochs are set to 12 iterations. For identification of white patch, it is set to network 1, and for the tattoo, it is set to network 0. The result of white patch identification is about 91.9% and the failure case is about 8.1% of 150 samples. In tattoo, the result is 95.6%

Table 2 Comparison of the proposed method results with other methods

Sl. No.	Methods	Results (%)
1	Neural network with back-propagation	75
2	Neural network with back-propagation	75.6
3	Auto-associative neural network	80.8
4	Auto-associative neural network	82.6
5	**Proposed method**	**93.6**

The significance of the results is with respect to the implementation of method and the results generated from the testing dataset

and the failure case 4.4% of 145 images, and the combined images' overall accuracy result is 93.6%, and the failure case 6.4% of total 295 samples.

5.2 *Comparative Analysis*

The experiment is evaluated on skin disorder and artificial body marks, and it contains 150 and 145 images. The proposed method results are compared with other methods and sorted that the proposed method is the novel method which is not dominated by any other methods. The proposed method is compared with the neural network with back-propagation [18, 19] and auto-associative neural network [20, 21] methods. The results of the proposed method are compared based on the method accuracy shown in Table 2.

The proposed method was implemented using MATLAB 2018b and trained and tested on different datasets, after testing the proposed method has given better results as compared to the other methods present in the literature review. The manner in which a system interprets an image is completely different from humans. Computer vision uses image processing algorithms to analyze and understand visuals from a single image or a sequence of images. Considering the growing potential of computer vision, many organizations are investing in image recognition to interpret and analyze data coming primarily from visual sources for a number of uses such as medical image analysis. Image recognition is the ability of a system or software to identify objects, people, places, and actions in images. It uses machine vision technologies with artificial intelligence and trained algorithms to recognize images through a camera system. Much fuelled by the recent advancements in machine learning and an increase in the computational power of the machines, image recognition has taken the world by storm.

6 Conclusion and Future Scope

The proposed method has shown the classification and identification of body mark using ANN. The proposed method obtained 91.9% accuracy for white patches, and 95.6% for a tattoo, and the overall accuracy is about 93.6% when both images were combined. The color, texture, and shape features are used to set for white patches and tattoos, respectively. The proposed method is very useful for identification between skin disorder marks and artificial body marks like white patches and tattoos. The work carried out has relevance to real-world identification of body mark type, and it involves both image processing and pattern recognition techniques. The future work is to identify and classify the natural, artificial, and accidental body marks, which is useful in several areas like identifying the criminal and identify the bodies where mass death occurred.

References

1. A Brief History of Tattoos. https://www.designboom.com/history/tattoo_history.html
2. H. Nugroho, *Computerised Image Analysis of Vitiligo Lesion: Evaluation Using Manually Defined Lesion Areas* (Wiley, A/S Skin Research and Technology, 2011), pp. 1–6
3. H. Nugroho, Determination of skin repigmentation progression. in *Proceedings of the 29th Annual International Conference of the IEEE EMBS Cité Internationale* (Lyon, France, 23–26 Aug 2007)
4. A.K. Jain, J.E. Lee, R. Jin, Tattoo-ID: automatic tattoo image retrieval for suspect and victim identification. in *PCM* (2007), pp. 256–265
5. M. Ngan, P. Grother, Tattoo recognition technology—challenge (Tatt-C): an open tattoo database for developing tattoo recognition research (National Institute of Standards and Technology (NIST))
6. J.-E. Lee, A. Jain, R. Jin. Scars, marks, and tattoos (SMT): soft biometric for suspect and victim ıdentification. in *Biometrics Symposium* (September 2008)
7. Dayanand G Savakar Recognition and classification of similar looking food grain ımages using artificial neural networks. J. Appl. Comput. Sci. Math. **13**(6), (2012)
8. B.S. Anami, D.G. Savakar, Identification of multiple grain image samples from tray. Int. J. Food Sci. Technol. **44**, 2452–2458 (2009)
9. B.S. Anami, D.G. Savakar, Improved method for ıdentification and classification of foreign bodies mixed food grains ımage samples. ICGST-AIML J. **9**(1), 2009. ISSN: 1687-4846
10. A. Nurhudatiana, *A Computer-Aided Diagnosis System for Vitiligo Assessment: A Segmentation Algorithm* (Springer-Verlag Berlin Heidelberg, 2015), pp. 323–331
11. Scott T. Acton, Adom Rossi, Matching and retrieval of tattoo images: active contour cbir and glocal image features. in *IEEE SSIAI* (March 2008)
12. B.S. Anami, D.G. Savakar, Effect of foreign bodies on recognition and classification of bulk food grains image samples. J. Appl. Comput. Sci. **6**(3), (2009)
13. T. Lee, V. Ng, R. Gallagher, A. Coldman, D. McLean, Dullrazor a software approach to hair removal from images. Comput. Biol. Med. **27**(6), 533–543 (1997)
14. J.S. Raj, J.V. Ananthi, Recurrent neural networks and nonlinear prediction in support vector machines. J. Soft Comput. Paradigm (JSCP) **1**(01), 33–40 (2019)
15. M.H.J.D. Koresh, J. Deva, Computer vision based traffic sign sensing for smart transport. J. Innov. Image Process. (JIIP) **1**(01), 11–19 (2019)

16. D.G. Savakar, D. Telsang, A. Kannur, Comparative analysis of segmentation technique for different body marks. Int. J. Adv. Sci. Technol. (IJAST) **29**(4), 8605–8614 (2020)
17. D.G. Savakar, D. Telsang, A. Kannur, Artificial neural network for identification and classification of natural body marks. in *International Conference on Communication, Computing and Electronics Systems (ICCCES 2020) Accepted in Conference*
18. H.R. Mhaske, D.A. Phalke, Melanoma skin cancer detection and classification based on supervised and unsupervised learning. in *Circuits, Controls and Communications (CCUBE), 2013 International Conference* (IEEE, December 2013), pp. 1–5
19. M.K.A. Mahmoud, A. Al-Jumaily, M. Takruri, The automatic identfication of melanoma by wavelet and curvelet analysis: study based on neural network classfication. in *Hybrid Intelligent Systems (HIS), 2011 11th International Conference* (IEEE, December 2011), pp. 680–685
20. Ho Tak Lau, Adel Al-Jumaily, Automatically early detection of skin cancer: study based on neural network classification. in *International Conference of Soft Computing and Pattern Recognition* (IEEE, 2009), pp. 375–380
21. S. Srivastava, D. Sharma, Automatically detection of skin cancer by classification of neural network. Int. J. Eng. Tech. Res. **4**(1), 15–18 (2016)

Whale Optimization Algorithm Applied to Recognize Spammers in Facebook

R. Krithiga and E. Ilavarasan

Abstract Social networks have become an inevitable communication system in the everyday life of people. In recent years, the number of users of social networks has only been increasing exponentially due to the attractive services that it offers. Despite these benefits, it also suffers from the threat of spammers. The spamming is executed through texts, photos, videos, and other multimedia forms. The presence of spammers hinders the user's experience and causes inconvenience. The existing spammer detection algorithms fail to accurately detect the spammers. In this work, the spammer detection problem is viewed as a binary classification problem that classifies the user profiles into either spam or non-spam account. A set of IP-address-based features is adapted and classifies Facebook's user profiles. Further, the whale optimization algorithm is applied as a feature selection method to select the optimal feature subset. The experimental results indicate that the WOA outperforms the existing methods in terms of accuracy.

Keywords Social network mining · Spammer detection · Spam profile identification · Facebook · WOA

1 Introduction

Online social networks (OSN) have attained its popularity due to the cost-effective and seamless communication it offers. The OSN exists in various forms that differ in structure, purpose, and services offered. People irrespective of geographical constraints can stay connected with their acquaintances and friends' circle. The user of OSN posts contents in the form of text, photo, video, and other multimedia files depending on the OSN platform being used. Thus, the associated users enlisted as

R. Krithiga (✉) · E. Ilavarasan
Department of Computer Science and Engineering, Pondicherry Engineering College, Pondicherry, India

E. Ilavarasan
e-mail: eilavarasan@pec.edu

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_50

friends could witness the sharing of the individuals. More often, users accept friend requests from unknown sources that would seed for stalking. Unsolicited messages in the form of URL's or texts are sent to the users responding to which would bring in the consequences of spamming. The spamming is caused by user accounts called "spammers," which could cause financial loss, defamation, loss of reputation, and stealing of personal information, etc, [1]. The spamming has been continuously executed since the launch of social networks due to the prevailing attraction and popularity. Social networks pose several research problems such as link prediction, rumor detection, friend recommendation, identification of opinion leaders, user profiling, fake profile identification, fake news detection, and many more. However, in this work, the spam account is addressed by identification problem. Various methods have been suggested in the literature to deal with the problem of spamming. However, one of the serious problems faced by the research community is the changing nature of spammers [2]. The spammers have been adopting peculiar strategies to invade the detector systems and stay hidden in the network in the guise of legitimate users.

The behavior of spammers is to be recorded to comprehend the strategies employed by them. In social network mining, this behavior is defined in the form of feature values. The feature set describes the characteristics of spammers. Various categories of features have been proposed in the literature to best discriminate the spammers from legitimate users [3]. Extensive experiments have been performed to identify the spammers exclusively for a particular type of OSN. The smart spammers manipulate and pretend to be legitimate users by mimicking the feature values that are close to genuine users. Hence, there is a need to devise features that would be impossible to manipulate or influence. Also, the presence of noisy and irrelevant features would diminish the accuracy of spam classification. To address the aforementioned issues, the IP-address-based features are adopted as it is challenging to overrun and also aid in identifying spammers. Further, the whale optimization algorithm (WOA) is employed in the feature selection process to select the optimal set of features that best discriminates the spam profiles in the network. For the experiment, the Facebook network has been considered, and user profiles were extracted using the Facebook API. Performance comparison and analysis is also done as part of the experiment against the most widely used wrapper algorithms such as genetic algorithm (GA) and particle swarm optimization (PSO).

The full article is organized as follows: Sect. 2 summarizes the related work prevails in the literature. Section 3 details the wrapper algorithms such as GA, PSO, and WOA. Section 4 provides the experimental settings and methodology incorporated to execute the techniques under consideration. Section 5 concludes the research work.

2 Related Work

Since the launch of social networks, several types of research have been in execution to address various issues related to social networks. In [4], the author crawled a huge dataset that consisted of 23 million Arabic tweets and performed spam tweet detection. The profile samples were defined in terms of profile and content attributes. In [5], the author proposed a novel set of features appropriate for real-time spam filtering. The features were proposed considering the time of tweet postings, account creation date, and alike. In [6], the author devised a hybrid methodology that incorporated both content-based and graph-based features for spam detection on Twitter on a dataset that consisted of 11K users. In [7], the author considered the time correlation among the tweets, and the method outperformed the account-based features. The author performed a comprehensive survey of features proposed in the literature and evaluated them [8]. The experiment indeed revealed some of the most key features that aid in the spam detection process. In [9], the author proposed three graph-based and four tweet content-based features for spam detection on Twitter. The random forest classifier yielded an accuracy of 91% using the proposed feature set. In [10], the author grouped similar Twitter users into categories and introduced a dynamic feature selection to use context-specific feature set for every type of user group instead of utilizing the same feature set for the entire system. In [11], the author proposed a novel technique to identify the spam groups in the Facebook social network. A chrome extension called "Itus" was developed using the support vector machine classifier which would extract the features from the Facebook groups. In [12], the author considered temporal evolution factors to distinguish the behavioral patterns and the outperformed other existing techniques. Though several models have been proposed in the literature, all of them become obsolete as the spammers change strategies over time. All the models proposed in the literature by academicians, and industrialists are applicable only for a particular period of time. By analyzing the detecting pattern, the spammers invade the detector mechanism. Hence, to tackle this problem, a set of IP-address-based features are adopted to recognize spammers in the Facebook social networks as IP-addresses are difficult to manipulate and evade. The article in [13] attempted to apply WOA using content, graph-based, profile-based features and effectively detected spammers in the Twitter network. Hence, in this work, the WOA is selected to detect spammers in Facebook using the proposed set of features and evaluate its applicability.

3 Feature Selection Algorithm

In social network mining, the feature selection (FS) phase plays an important role to reduce the dimension of the data and also increase the prediction accuracy by eliminating the irrelevant features. Thus, performing feature selection not only reduces the size of the data but also increases prediction efficiency [14]. The FS is broadly classified into the filter, wrapper, and hybrid methods. The first determines the top best features by setting a threshold to calculated values by performing a statistical scoring. The second class of methods utilizes an optimization algorithm to choose the ideal set of features. Studies state that no single feature is strongly discriminative to distinguish one class from the other. Instead, when used collectively yields the

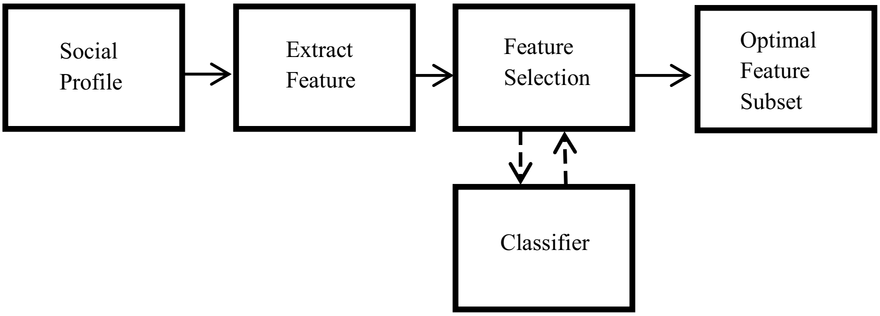


Fig. 1 Scheme of feature selection

optimal set of features that possess stronger discrimination ability. The third class combines the advantages of both classes of FS methods. Due to these reasons, the most representative optimization algorithms such as genetic algorithm (GA), particle swarm optimization (PSO), and whale optimization algorithm (WOA) attempted to analyze the performance (Fig. 1).

3.1 The Genetic Algorithm (GA)

The GA is a class of optimization algorithms that works based on the principle of natural gene selection [15]. The individual solutions are represented as chromosomes and a group of such chromosomes comprises the population. The new off-springs are produced from the current population through crossover and mutation operators. The survival of the fittest is retained in the population based on the fitness value. The entire process of gene selection and offspring production is repeated for generations until the stopping criterion is met. Finally, the optimal sequence of the chromosome is output as the optimal set of features.

3.2 The Particle Swarm Optimization (PSO)

The PSO is a stochastic optimization technique inspired by the bird flocking behavior [16]. The positions of individuals change during iteration based on two ways: (i) own flying experience (pbest) and (ii) concerning the best particle of the swarm (gbest). One of the unique aspects of PSO is the ability to recall as the individuals in the swarm hold the history of the previous best position. During iteration, the position of the particles is updated based on the velocity, and the best particle of the swarm shares the coordinates of its position to the entire swarm. Depending on the stopping criterion, the algorithm ceases and outputs the optimal sequence of features.

3.3 The Whale Optimization Algorithm (WOA)

The WOA has been developed inspired by the prey hunting behavior of whales [17]. The whales hunt for their prey following three kinds of motions, namely (a) encircling prey, (b) upward spiral movement, and (c) prey search. The first two movements intensify exploitation and the latter facilitate exploration. The WOA has been successfully applied to solve many real-world problems and attained outstanding results. The movements of the whales are mathematically modeled to be fitting for the optimization task. In the first movement, the whales surround the prey to migrate toward it and are mathematically formulated using the below equation,

$$X(t+1) = X^* - A \cdot D \tag{1}$$

where

$$D = \left|C \cdot X^* - X(t)\right| \tag{2}$$

$$A = 2a \cdot r_1 - a \tag{3}$$

$$C = 2 \cdot r_2 \tag{4}$$

$$a = 2 - t\frac{2}{t_{\mathrm{Max}}} \tag{5}$$

Here, $X(t+1)$ denotes the position of a whale during the next iteration, X^* represents the position of the best whale, $X(t)$ denotes the position of the whale in the current iteration, D is the distance between the current position of the whale and the best position of a whale, C is a coefficient, A is a random number in the range [-a, a], where a decreases from 2 to 0 during iterations as in Eq. (5), r_1 and r_2 are random numbers in the range [0,1], t is the current iteration, and t_{Max} is the maximum iterations defined in the algorithm.

In the second type of exploitation, the whales exhibit a spiral movement parallelly propelling an upward movement and are articulated as follows,

$$X(t+1) = D' \cdot ebl \cdot cos(2\pi l) + X^* \tag{6}$$

$$D' = \left|X^* - X(t)\right| \tag{7}$$

where D' is the absolute distance between the best position of a whale and its current position, b defines the shape of the spiral, and l is a random value in the range [−1, 1]. The WOA chooses between the above position updating mechanisms, based on a probability p as defined in Eq. (8).

$$X(t+1) = \begin{cases} X^* - A \cdot D & p < 0.5 \\ D' \cdot e^{bl} \cdot \cos(2\pi l) + X^* & p \geq 0.5 \end{cases} \tag{8}$$

The last movement involves a random search that achieves exploration, where a random whale is chosen, and the rest of the whales move toward the random whale. This mode of movement is mathematically accomplished using the following Eq. (9),

$$X(t+1) = X_{\text{rand}} - A \cdot D \tag{9}$$

$$D = |C \cdot X_{\text{rand}} - X(t)| \tag{10}$$

Here, X_{rand} denotes the position of a whale that is chosen randomly. The exploitation in the hyperplane is achieved by updating the positions based on the best position of a whale. The diversification is ensured through the randomization process in the algorithm (Fig. 2).

4 Experiment and Results

The algorithms were coded in Python 3.0 using Scikit package on a Windows 10 machine with Intel Core i7-3630, 2.40 GHz processor, and a RAM capacity of 8 GB. The parameter settings of GA, PSO, and WOA are detailed in Table 1. The dataset consists of 2634 non-spam profiles and 867 spam profiles. This was crawled using the Facebook developer API and transformed into a dataset. The manual labeling was done by following the URLs in the content, and accordingly, the ground truth was established. The SVM and AdaBoost are used as classification algorithms. A total of 24 features are used in the problem. Nineteen features based on profile and content in [18] are adopted, and five IP-address-based features proposed in our previous research for the Twitter social network [19] have been mapped to the Facebook OSN, and 70:30 ratios were followed for training and testing set split.

The proposed optimization algorithm-based wrapper methods are primarily used for the feature selection process. The chosen subset is evaluated using various classification algorithms, and the results are recorded. The results are compared based on the four evaluation metrics as given in Eqs. (11)–(14).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{11}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{12}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{13}$$

Fig. 2 WOA-based feature selection

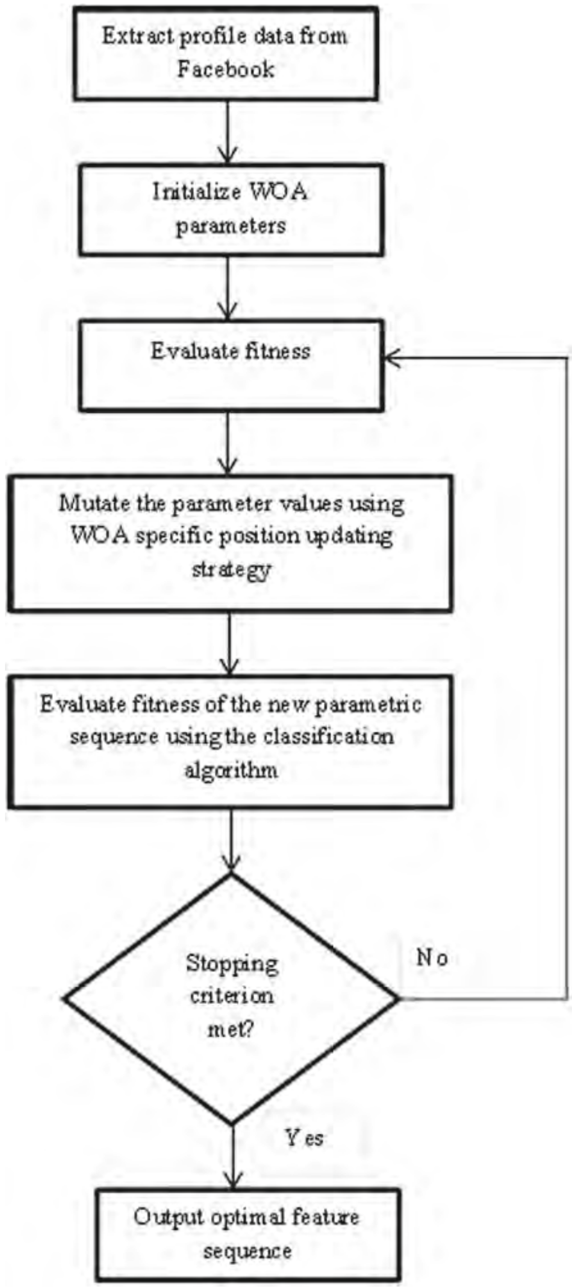


Table 1 Proposed novel set of features

IP-address-based features	
F20	No. of the distinct IP-address in a week time
F21	No. of the distinct IP-address in the last 24 h
F22	No. of posts deleted
F23	No. of IP addresses used for posts
F24	No. of IP addresses used for comments

$$F - \text{Measure} = \frac{2 \times \text{Precision} \ \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

where

TP Number of Facebook profiles correctly recognized as spam
TN Number of Facebook profiles misrecognized as non-spam
FP Number of Facebook profiles incorrectly recognized as spam
FN Number of Facebook profiles misrecognized as non-spam

The above metrics are calculated by constructing the confusion matrix.

Since the feature set size is high dimensional, a feature selection process is required to be performed to reduce the columns by choosing the optimal set of features. For this purpose, the GA, PSO, and WOA are employed by including the proposed set of novel features in the dataset. The parameter settings of these algorithms are displayed in Table 2. These feature selection algorithms in union with the classifier algorithms such as SVM and Adaboost are compared in terms of the metrics considered.

Table 2 Parameter settings

GA parameters	
Population	10
Generation	25
Mutation ratio	0.2
Crossover function	Single point
PSO parameters	
Population	8
Iteration	25
Velocity	Vmax = 2.0, Vmin = 0
Acceleration constants	C1, C2 = 3
Inertia weight	Wmax = 0.9, Wmin = 0.4
WOA parameters	
Population	8
Iteration	15

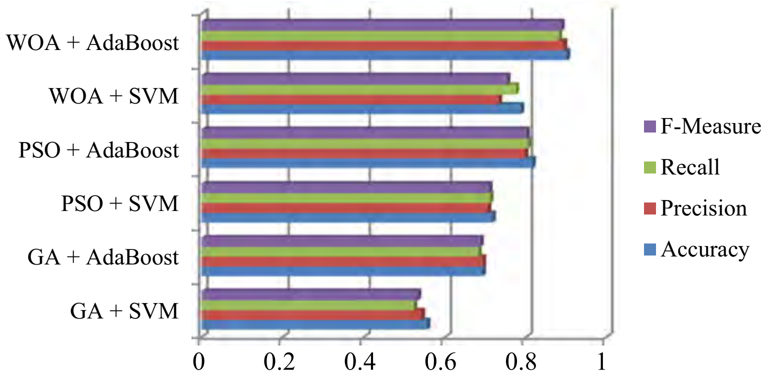


Fig. 3 Scheme of feature selection

The results of the experiment are given in Fig. 3 which clearly shows that the WOA + AdaBoost has attained the highest accuracy followed by PSO + AdaBoost, WOA + SVM, PSO + SVM, GA + AdaBoost, and GA + SVM.

5 Conclusion

In the proposed work, the whale optimization algorithm is attempted to detect spam profiles on Facebook. The IP-address based on features was adopted to characterize the spammers, and WOA was used in the feature selection phase. The performance was compared with other existing wrapper algorithms GA and PSO. The experimental results indicate that the WOA was more accurate in distinguishing the spam profiles.

References

1. C. Yang, R.C. Harkreader, G. Gu, Die free or live hard? Empirical evaluation and new design for fighting evolving Twitter spammers. In: *Recent Advances in Intrusion Detection. RAID 2011*, ed. by R. Sommer, D. Balzarotti, G. Maier. Lecture Notes in Computer Science, vol. 6961 (Springer, Berlin, Heidelberg, 2011)
2. A.A. Amleshwaram, N. Reddy, S. Yadav, G. Gu, C. Yang, CATS: characterizing automation of Twitter spammers. in *2013 Fifth International Conference on Communication Systems and Networks (COMSNETS)* (Bangalore, 2013), pp. 1–10
3. R. Krithiga, E. Ilavarasan, Machine learning techniques for spammer identification: state of the art and analysis. J. Crit. Rev. **7**(1), 446–448 (2020)
4. N. El-Mawass, S. Alaboodi, Detecting Arabic spammers and content polluters on Twitter. in *2016 Sixth International Conference on Digital Information Processing and Communications (ICDIPC)* (Beirut, 2016), pp. 53–58
5. M. Washha, A. Qaroush, F. Sedes, Impact of time on detecting spammers in Twitter. (2016)

6. M. Mateen, M.A. Iqbal, M. Aleem, M.A. Islam, A hybrid approach for spam detection for Twitter. in *2017 14th International Bhurban Conference on Applied Sciences and Technology (IBCAST)* (Islamabad, 2017), pp. 466–471
7. M. Washha, M. Mezghani, F. Sedes, Behavioural account-based features for filtering out social spammers in large-scale Twitter data collections. Ingenierie Des Systemes D'Information **22**, 65–88 (2017)
8. W. Herzallah, H. Faris, O. Adwan, Feature engineering for detecting spammers on Twitter: Modelling and analysis. J. Inf. Sci. (2017). 016555151668429
9. Z. Alom, B. Carminati, E. Ferrari, Detecting spam accounts on Twitter. in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (Barcelona, 2018), pp. 1191–1198
10. M.S. Karakaşlı, M.A. Aydin, S. Yarkan, A. Boyaci, Dynamic feature selection for spam detection in Twitter. in *International Telecommunications Conference*, ed. by A. Boyaci, A. Ekti, M. Aydin, S. Yarkan. Lecture Notes in Electrical Engineering, vol. 504 (Springer, Singapore, 2019)
11. F.-H. Hsu, M.-J. Yan, K.-W. Chang, C.-W. Ou, H.-M. Sun, Itus: behavior-based spamming group detection on Facebook. J. Comput. (Taiwan) **29**, 69–81 (2018)
12. Q. Fu, B. Feng, D. Guo, Q. Li, Combating the evolving spammers in online social networks. Comput. Secur. (2017)
13. A.M. Al-Zoubi et al., Evolving support vector machines using whale optimization algorithm for spam profiles detection on online social networks in different lingual contexts. Knowl. Based Syst. **153**, 91–104 (2018)
14. A.A. Bidgoli, H. Ebrahimpour-Komleh, S. Rahnamayan, A many-objective feature selection algorithm for multi-label classification based on computational complexity of features. in *2019 14th International Conference on Computer Science and Education (ICCSE)* (Toronto, ON, Canada, 2019), pp. 85–91
15. D. Thierens, D.E. Goldberg, Mixing in genetic algorithms. in *Proceedings of the Fifth International Conference on Genetic Algorithms*, ed. by S. Forrest. (Morgan Kaufmann, 1993).
16. J. Kennedy, R. Eberhart, Particle swarm optimization. in *Proceedings of ICNN'95—International Conference on Neural Networks*, vol. 4 (Perth, WA, Australia, 1995), pp. 1942–1948
17. N. Rana, M.S.A. Latiff, S.M. Abdulhamid et al., Whale optimization algorithm: a systematic review of contemporary applications, modifications and developments. Neural Comput. Appl. **32**, 16245–16277 (2020)
18. S. Rathore, V. Loia, J.H. Park, SpamSpotter: an efficient spammer detection framework based on intelligent decision support system on Facebook. Appl. Soft Comput. J. (2017)
19. R. Krithiga, E. Ilavarasan, A novel hybrid algorithm to classify spam profiles in Twitter. Webology **17**(1), 260–279 (2020)

FNAB-Based Prediction of Breast Cancer Category Using Evolutionary Programming Neural Ensemble

Vijaylaxmi Inamdar, S. G. Shaila, and Manoj Kumar Singh

Abstract The process of diagnosis and survival rate along with endurance period is greatly decided by the type of cancer and the current existing stage of cancer disease. The information that is accurate could provide lots of help during patient treatment management and increase the chance of survival. The biopsy process through fine needle aspiration has generally been applied to rule out the presence of malignancy lesions. Cytologists analyzed the details and tried to establish the correlation among the various observations to deliver the outcome. The limitation of expertise and possibilities of natural human error can cause a significant impact on the diagnosis process as well as patient economical and cognition conditions. In this article, computational intelligence has developed over the neural network platform to predict the possibilities of lesions' categorical belongings. The feed-forward architecture has considered developing the predictor because of its universal approximation quality. To obtain more robust outcomes in the decision instead of a single classifier, ensembles of classifiers have proposed where every individual classifier is having variability in the training data to avail the quality of knowledge diversity. The formation of weight-oriented ensemble has been done by evolutionary programming. It has been observed that the proposed method has delivered a high value of correct recognition in comparison to individual classifiers. The benefit of the proposed form of the ensemble has compared against conventional methods like majority voting and mean value decision. The strength of ensemble forming capability by evolutionary programming has compared with particle swarm optimization. The proposed ensemble benefit has further been tested over receiver operating characteristics environment to meet the practical challenges of variability in decision threshold value.

V. Inamdar · S. G. Shaila
Dayananda Sagar University, Bangalore, India
e-mail: vijaylaxmi.inamdar@gmail.com

S. G. Shaila
e-mail: shaila-cse@dsu.edu.in

M. K. Singh (✉)
Manuro Tech Research Pvt.Ltd, Bangalore, India
e-mail: mksingh@manuroresearch.com

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_51

Keywords Breast cancer · Benign · Malignant · Neural network · Ensemble · Evolutionary programming · Receiver operating characteristics

1 Introduction

The origination of breast cancer started from the breast tissue and commonly existed in the milk duct's inner lining or the lobules which provides the milk to the duct. Broadly, depends upon the origin of cancer either from ducts or from lobules, cancer is defined as ductal carcinomas or lobular carcinomas correspondingly [1]. The quantity and kind of treatment are a complex function of various characteristically parameters that appeared with breast cancer disease like size, stage, and rate of growth, age, etc. Various possible forms of treatment may occur in the breast cancer disease such as surgery, radiation therapy, immunotherapy, and drug administration which may include hormonal therapy as well as chemotherapy. In most cases, the primary screening of breast cancer happens through mammography. But mammography outcomes carried a high value of false outcomes especially at the beginning stage of breast cancer [2]. Hence, a more confirmative approach has been applied by examination of biopsy tissue samples from the breast [3]. The tissue samples were examined by cytologists to define the category of tissue in the domain of normal or cancerous by observing the various cell characteristics. The manual examination and analysis of tissue are a very tedious and challenging task and needed great expertise in delivering the right outcome. The manual approach carried the obvious limitations of a human in the decision outcome, and in general, the accuracy level appeared in the range of 62–89%. Hence, a need and considerable scope are existing to place the automated decision aids which have intelligence like a human but at the same time do not have limitation like a human. So, unlike in the past, where, fundamentally the handcrafted features bounded in the rule-based process have been applied in the CAD system to define the diagnosis decision, machine learning-based artificial intelligence can be the better model.

In this article, the research work has contributed to increasing the reliability of the computational decisions by forming an optimal ensemble of predictors using computational intelligence where the confidence of individual classifiers rewarded with the optimal weight value. This layout of the article is organized as follows: Sect. 2 describes the literature review, and Sect. 3 proposed the evolutionary programming-based ensemble of the classifier. The experimental results have been discussed and analyzed in Sect. 4. Finally, Sect. 5 draws the conclusion of the research work.

2 Literature Survey

A numbers of researchers have given attention to breast cancer for automated detection. Based on a combination of SVM, K-nearest neighbors and PNN [4] have proposed a model to classify the category of breast tumor in the benign and malignant. Single nucleotide polymorphisms from the BRCA1, BRCA2, and TP53 genes have been utilized in [5] to predict the presence of breast cancer through machine learning [6] The mammography-based approach has included in [7] to develop the machine learning-based classifier to classify the feature vectors on *craniocaudal* (CC) and/or *mediolateral oblique* (MLO) mammography image views. To predict the eventual therapeutic response of breast cancer patients after a single cycle of NAC, [8] proposed a machine learning method where artificial meta-plasticity learning algorithms have been used. The artificial meta-plasticity in MLP is applied to detect breast cancer [9]. A comparison over performances of breast cancer detection by different machine learning algorithms (SVM, C4.5, NB, and K-NN) has been presented in [10]. Prediction of breast cancer recursion has been proposed in [11, 12]. There could be a remarkable contribution to the diagnosis and treatment by identifying the genetic factor with microarray technology. The breast cancer data based on identifying the active genes are applied to cluster the data [13]. The weighted K-means support vector machine (wKM-SVM) and weighted support vector machine (wSVM) have been applied in [14] to identify breast and kidney cancer. The period of survival can help a lot in defining the diagnosis process as well as for the patient itself. Hence, the survival period prediction is important, and a rule-based approach in integration of machine learning has been proposed in [15] SVM and its ensemble-based approach has been proposed in [16] for breast cancer prediction. FFBPN-based classifier for breast cancer has been discussed in [17] to define in the category to malignant or benign. Based on machine learning, the various existing method has been reviewed in [18, 19].

3 Proposed Solution

The development of ensemble architectures has been fulfilled in the two different levels. In the first level, three individual classifiers have developed as shown in Fig. 1, where each classifier has received training with a different part of the data set carried new samples of data to achieve knowledge diversity. The individual classifier is made of multilayer perceptron feed-forward architecture contains the unipolar sigmoid function as a transfer function in their active nodes. Each individual has received the learning using the gradient descent algorithm. Finally, after completion of learning, ready to use predictor classifiers CLF1, CLF2, and CLF3 were available to be ensembled in the second stage. In the second stage, the overlapping set of training data (contains the samples from the training data set of individual classifiers) has been taken, and at a time, the same data have given as input to all the three classifiers. The generated outcomes of each classifier are a numeric value in the range of 0–1. These outcomes multiplied with corresponding weights $w1$, $w2$, and $w3$, and a summation operation has performed before passing to the Heaviside function which has a threshold to transfer the input into binary form as 0 or 1 to represent the final decision. The error in ensemble module decision corrected by minimizes

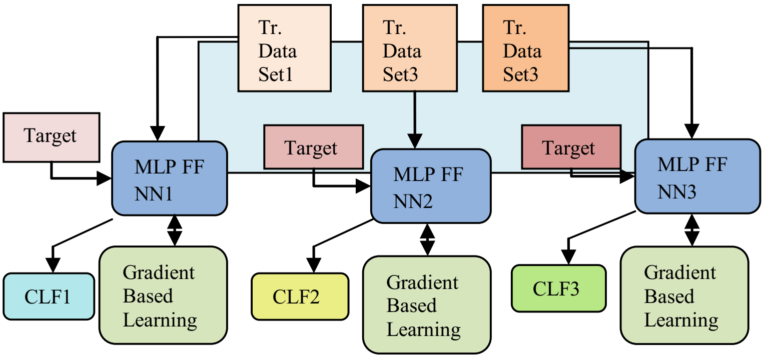


Fig. 1 Individual classifier development for the first stage

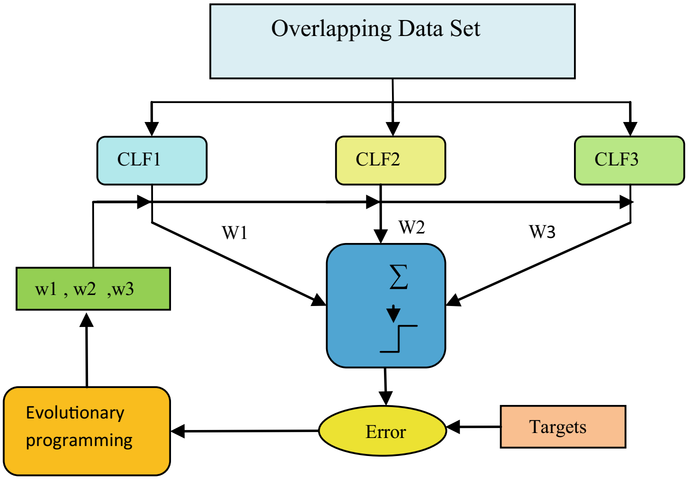


Fig. 2 EP-based development of ensemble architecture

the error value through evolutionary programming which has given the job to find the more suitable weights corresponding to each classifier. After convergence, the best possible value of minimum error has been achieved, and corresponding weights values are the final decision weights for the individual classifier. The whole process is described in Fig. 2.

3.1 *Learning Algorithm with Gradient Descent*

The change in the value of the weight at any instant follows the proportionality of the derivative value of the available nonlinear transfer function and the local error value which has been evaluated by Eq. (1) for output layer weights, and Eq. (2) has applied for hidden layer weight. The hidden layer and output layer weights upgrading has been given by Eq. (3) where the momentum term has been included to increase the rate of learning.

$$\delta_i^s = \left(d_q - x_{\text{out},i}^s\right) g\left(u_i^s\right) \tag{1}$$

$$\delta_i^s = \sum_{h=1}^{n2} \delta_h^{s+1} w_{hi}^{s+1} g\left(u_i^s\right) \tag{2}$$

$$w_{ij}^s(t+1) = w_{ij}^s(t) + \mu \delta_i^s x_{\text{out},j}^s + \alpha\left[w_{ij}^s(t) - w_{ij}^s(t-1)\right] \tag{3}$$

3.2 Self-Adaptive Gaussian Mutation-Based Evol. Programming:(GmEP)

The evolution is a two-step population-based process of random variation and selection. The procedure of new solution generation can be represented as the differential equation as given in Eq. (4).

$$X(t+1) = f_s(f_V(X[t])) \tag{4}$$

where $X[t]$ is the population at time t under a representation X, $f_{\mathbf{V}}$ is a random variation function operator to provide the perturbation in the present solution, and $f_{\mathbf{S}}$ is the selection function operator which decides which part of the next generation will take place. There is a variety of possibility exist in the solution representation and definition of variation and selection methods. The efficiency of an evolutionary algorithm depends on the interplay between the operator's $f_{\mathbf{S}}$ and $f_{\mathbf{V}}$ as it defines the balance between exploration and exploitation. In EP, the primary search operator is the mutation which provides the controlled change in each parent to produce an offspring as given by Eq. (5) and standard deviation of Gaussian function controlled in a self-adaptive manner to explore the solution space faster as given by Eq. (6).

$$\overline{x}_i^{'}(j) = \overline{x}_i(j) + \overline{\sigma}_i(j) N_j(0,1) \tag{5}$$

$$\overline{\sigma}_i^{'}(j) = \overline{\sigma}_i(j) \cdot \exp(\tau' N(0,1) + \tau N_j(0,1)) \tag{6}$$

$$\forall j \in \{1, 2, 3, \cdots r\} \ldots,\ \tau = \left(\sqrt{2\sqrt{n}}\right)^{-1};\ \tau' = \left(\sqrt{2n}\right)^{-1}$$

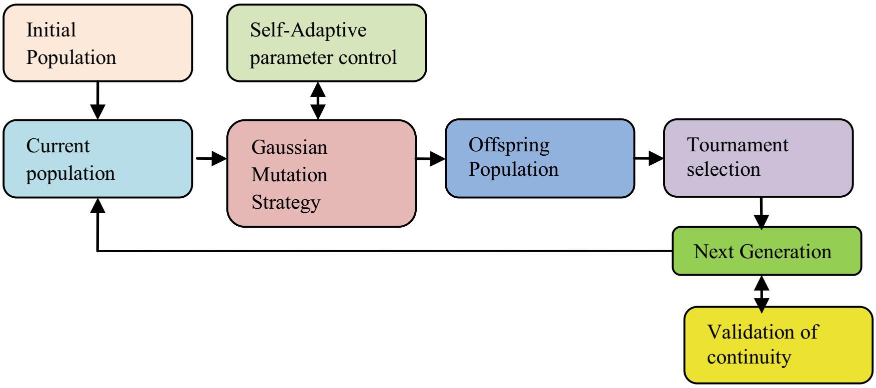


Fig. 3 Process flow of self–adaptive evolutionary programming

where $\overline{x}_i^{'}(j)$, $\overline{x}_i(j)$, $\overline{\sigma}_i^{'}(j)$ and $\overline{\sigma}_i(j)$ denote the jth component of vectors $\overline{x}_i^{'}$, $\overline{x}_i$, $\overline{\sigma}_i^{'}$ and $\overline{\sigma}_i$, respectively. $N(0, 1)$ denotes a standard Gaussian random variable. $N_j(0, 1)$ indicates that the random variable is sampled for each new value of the counter j. The scaling factors τ and $\tau^{'}$ are robust exogenous parameters and n denotes the dimension of the problem considered in hand. The working process of the proposed method has shown in Fig. 3. The tournament selection provides the opportunity for each solution to compete with randomly selected members from a pool of parents and offspring. Based on the higher value of the total winning score by the individual, decisions over next-generation members took place.

4 Experimental Work

There are 11 attributes that are generally observed by cytologists in the decision process of defining the category of the observed lesion from FNA in malignant and benign [6]. These attributes contain binary outcomes in terms of presence or absence. The considered attributes are intracytoplasmic lumina, cellular dyshesion, 3D epithelial cell clusters, bipolar naked nuclei, foamy macrophase, nucleai, nuclear pleiomorphism, nuclear size, necrotic epithelial cells, apocrine change, and the age of the patient. There were a total of 692 instances of data available for the development of the training data set; for each classifier, 199 new instances have been considered. The detailed characteristics of all the three data set have shown in Table 1. The size

Table 1 Data set characteristics

	Data set1	Data set2	Data set3
Malagin	67	71	61
Benign	132	128	138

of neural architecture has taken as 11 input nodes, 6 hidden nodes, and 1 as the output node. The parameters of learning algorithm learning rate and momentum rate have taken as 0.9 and 0.2. The allowed number of iteration for learning was 100; while for over a data set, 10 dependent trials have applied to explore the generalized behavior. The performances of all three classifiers have shown in Table 2. Sensitivity and specificity are also estimated for each case. It can observe that CLF1 and CLF2 were excellent in the training phase in comparison to CLF3, while over test data CLF3 has shown dominancy. This shows that a single classifier cannot be considered very reliable especially when the area of applications is very sensitive like health care. The experimental work has been developed in the MATLAB environment.

The mean convergence characteristics for all the three classifiers have been shown in Fig. 4. It was observed that even though there were nearly identical number of different category instances and identical architecture developed for learning environment, but possibly the differences in instances attributes were variations in the learning path.

In the development of ensemble architecture, self-adaptive GmEP and dynamic weight variation-based PSO algorithms have been applied. For both algorithms, there was a population size of 50, and the allowed numbers of iteration were 100. There were three weights have to evolve to form ensemble architecture; hence, the length of the solution in the population was three. The initial value of standard deviation in the Gaussian mutation strategy has been taken as 0.001, while tournament size is

Table 2 Mean classifier performance over training and test data

	CLF1 Tr. data/Test data	CLF2 Tr. data/Test data	CLF3 Tr. data/Test data
Mean recog. efficiency	99.5000/92.8750	99.5000/91.4250	97.9000/96.0750
Sensitivity	98.5075/87.8788	98.5915/89.0625	96.5574/93.1884
Specificity	100.0000/95.3358	100.0000/92.5368	98.4892/97.5954

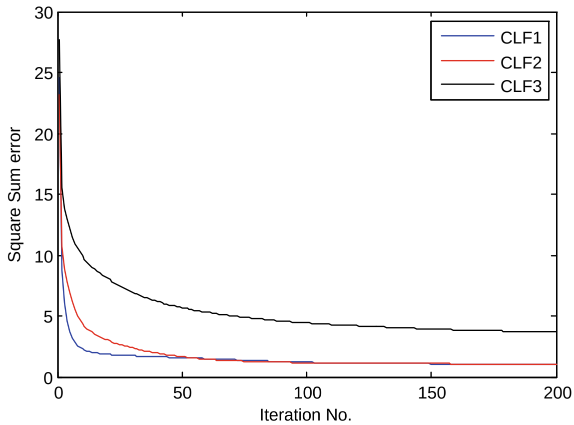


Fig. 4 Convergence characteristics of three classifiers

taken as 10. In DyPSO, the inertia weight decrease from the value of 1.2–0.1 linearly with iterations and social and cognition constant values have taken as 0.5 while the value of constriction factor was 0.7. There were ten independent trials obtained, and the performances are shown in Table 3 while the mean convergence characteristics are shown in Fig. 5.

It was observed that with DyPSO, there were high values of error as well as there was a high fluctuation in performance in different trials. The proposed GmEP has shown excellent consistency and maintains the minimum error always. Further, the experimental work shows that the GmEP is considered for ensemble development. The predicted performance by different methods over the complete data set has shown in Table 4. It was interesting to note that the majority voting decision method (MJDC) and mean value decision method (MNDC) even have the performances more inferior than classifier CLF3, and this happens because of lower performances value by other

Table 3 Performance obtained by GmEP and DyPSO under ten independent trials

Trials	GmEP			Total no. of decision error	DyPSO			Total no. of decision error
	W1	W2	W3		W1	W2	W3	
1	0.1654	0.2796	0.6345	14	0.2357	0.2562	0.5674	14
2	0.1976	0.3008	0.5463	14	0.0105	0.3776	0.9996	16
3	0.1686	0.2952	0.6361	14	0.2252	0.2862	0.5867	14
4	0.1714	0.2548	0.6452	14	0.2870	0.1061	1.0113	17
5	0.1711	0.3055	0.5886	14	0.2052	0.3123	0.6278	14
6	0.1433	0.2622	0.6064	14	0.1415	0.2677	0.7615	15
7	0.1904	0.2726	0.6182	14	0.2065	0.2965	0.6162	14
8	0.1981	0.2677	0.5939	14	0.3511	0.1857	0.6393	17
9	0.2046	0.3076	0.5296	14	0.1276	0.2852	0.7986	15
10	0.1618	0.3061	0.6076	14	0.1831	0.3055	0.6441	14

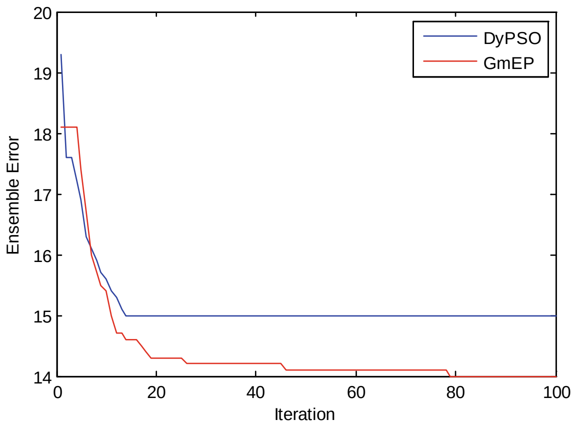


Fig. 5 Learning convergence characterize by DyPSO and GmEP

Table 4 Predicted efficiency by different method

	CLF1	CLF2	CLF3	MJDC	MNDC	EPENSM
Correct prediction	94.7977	94.0751	96.8208	96.5318	96.5318	**97.3988**

Bold value shows the betterment against the other values by different methods

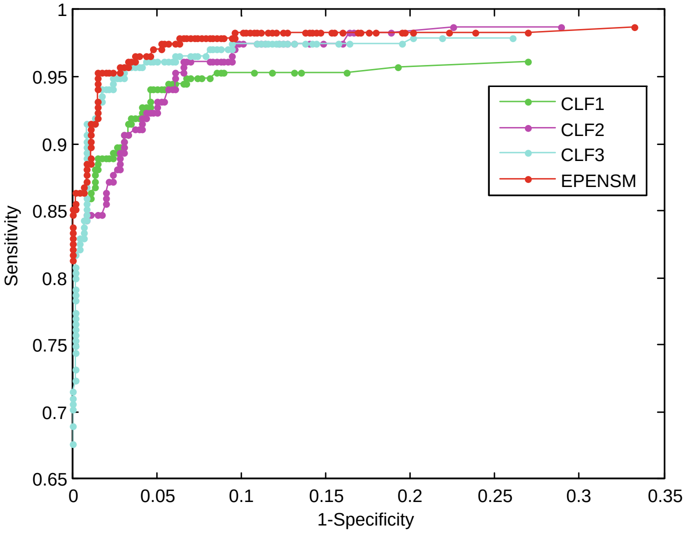


Fig. 6 Comparative ROC characteristics of ensemble approach against individual classifier

classifiers, while EP-based ensemble approach has delivered the maximum efficiency which is more than 97% which can be considered as very appreciated in the health sector. To understand the behavior of ensemble approach over the variability of the different threshold, the performance has explored over the ROC platform as shown in Fig. 6. The upper curve of EP-based ensemble (EPENSM) against the individual classifier curve ensures that under all conditions ensemble-based solution will deliver the better solution.

5 Conclusion

In the proposed work, the computational intelligence-based decision diagnosis is suggested for breast cancer detection. In the conventional form of CAD, decision process which generally occupied single classifiers lacking in generalization, hence an ensemble approach, was proposed which has shown the beneficial. The majority voting and mean value decision sustained from bias problems particularly when there is high variability exist in the individual outcomes. An intelligent ensemble was proposed which has used evolutionary intelligence to explore the optimal weights for the individual classifier in the final decision process. EP-based ensemble has outperformed the DyPSO-based model which had a higher error and less reliability in outcomes. The ROC curve ensures the proposed model of the ensemble is very valuable under different possibilities of a threshold. The proposed model is easy in implementation while having a high efficiency of recognition and can be used as a valuable decision aid for FNA-based breast cancer detection. In the future, there are different dimensions over the development of a better classifier in terms of architectural design through computational intelligence and automated analysis of mammography image and integration with FNAB outcome.

Acknowledgements This research work has been completed in Manuro tech research Pvt. Ltd, Bangalore, India under the program of Computational Intelligence in Health Care (CIHC).

References

1. World Health Organization. https://www.who.int/cancer
2. ICMR, Wed, 18 May 2016, PTI, New Delhi
3. E.D. Pisano, L.L. Fajardo, D.J. Caudry, N. Sneige, W.J. Frable, W.A. Berg, I. Tocino, S.J. Schnitt, J.L. Connolly, C.A. Gatsonis, B.J. McNeil, Fine-needle aspiration biopsy of nonpalpable breast lesions in a multicenter clinical trial. Radiology **219**(3), 785–792 (2001)
4. A. Osareh, B. Shadgar, Machine learning techniques to diagnose breast cancer. in *2010 5th International Symposium on Health Informatics and Bioinformatics (HIBIT)* (IEEE, 2010)
5. S. Silva, O. Anunciação, M. Lotz A comparison of machine learning methods for the prediction of breast cancer. in *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics. EvoBIO 2011*, ed. by C. Pizzuti, M.D. Ritchie, M. Giacobini. Lecture Notes in Computer Science, vol. 6623. (Springer, Berlin, Heidelberg, 2011)
6. UC Irvine Machine Learning Repository, https://archive.ics.uci.edu/ml/
7. R. Ramos-Pollán, M.A. Guevara-López, C. Suárez-Ortega et al., Discovering mammography-based machine learning classifiers for breast cancer diagnosis. J. Med. Syst. **36**, 2259 (2012)
8. S. Mani, Machine learning for predicting the response of breast cancer to neoadjuvant chemotherapy. J. Am. Med. Inf. Assoc. **20**(4), 688–695 (2013)
9. J. Fombellida, S. Torres-Alegre, J.A. Piñuela-Izquierdo, D. Andina Artificial metaplasticity for deep learning: application to WBCD breast cancer database classification. in *Bioinspired Computation in Artificial Systems. IWINAC 2015*. Lecture Notes in Computer Science, vol. 9108 (Springer, Cham, 2015)
10. HibaAsri, Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Comput. Sci. **83**, 1064–1069 (2016)
11. Pedro Abreu, Pridicting breast cancer recurrence using machine learning techniques: a systmatic review. ACM Comput. Surv. (CSUR) **49**(3), (2016)
12. Burcu Bektaş, Sebahattin Babur, Machine learning based performance development for diagnosis of breast cancer. in *Medical Technologies National Congress (TIPTEKNO)* (IEEE, 2016)
13. Nurdan Kolay, Pakize Erdoğmuş, The classification of breast cancer with machine learning techniques. in *Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)* (IEEE, 2016)
14. SungHwan Kim, Weighted K-means support vector machine for cancer prediction. SpringerPlus **5**(1), 1162 (2016)

15. M. Montazeri, M. Montazeri, M. Montazeri, A. Beigzadeh, Machine learning models in breast cancer survival prediction. Technol. Health Care **24**(1), 31–42 (2016). https://doi.org/10.3233/THC-151071
16. M.-W. Huang, SVM and SVM ensembles in breast cancer prediction. PLoS ONE **12**(1), e0161501 (2017)
17. L. Abdel-Ilah, H. Šahinbegović, Using machine learning tool in classification of breast cancer. in *CMBEBIH 2017*, ed. by A. Badnjevic. IFMBE Proceedings, vol. 62. (Springer, Singapore, 2017)
18. T. Saba, Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges. J. Infect. Public Health **13**(9), 1274–1289 (2020)
19. S.A. Mohammed, S. Darrab, S.A. Noaman, G. Saake, Analysis of breast cancer detection using different machine learning techniques. in: *Data Mining and Big Data. DMBD 2020*, ed. by Y. Tan, Y. Shi, M. Tuba. Communications in Computer and Information Science, vol. 1234 Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-7

Review on Augmented Reality and Virtual Reality Enabled Precious Jewelry Selling

Nishita Hada, Sejal Jain, Shreya Soni, and Shubham Joshi

Abstract In this new era of social distance, technology can help prevent the spread and build a resilient society. Technology is playing a vital role in keeping our society functioning without much disturbance. In such situations, online shopping becomes the easiest and the most suitable way for shopping. Considering India is one of the largest consumers of gold, diamonds, and other precious metals in the world. The jewelry and gem exchange of India plays a major role in the country's GDP. This paper shows jewelry shopping using augmented reality and virtual reality that will allow the users to virtually examine the jewelry in the comfort of their home and give them the joy of conventional shopping.

Keywords Virtual reality · Augmented reality · Virtual try-on · Online shopping

1 Introduction

With the rising challenges from the propagation of the novel coronavirus (COVID-19), the new and existing technologies and tools will witness rapid growth. Many people are bound to work from home and more digitalization is coming their way from basic everyday habits to important work. Just like work from home has become a part of life, online shopping has also started to gain consumers' attraction as it is safer and can be done while relaxing, according to their time and convenience. The customers check for the particular product is available or not. Once the desired product is found, it can be ordered at any time because online stores never closed. Online shopping has become much more simple and stress-free. Many shops can be visited in online shopping. It provides privilege to the customers to have access to a more personalized experience and explore the products of their choice. The prices of different e-stores can be compared and opt for the best one.

Since ancient times jewelry is considered an important part of the Indian tradition and heritage whether it be ancient civilizations like Harappa and MohenjoDaro or

N. Hada · S. Jain · S. Soni · S. Joshi (✉)
Department of Computer Engineering, SVKM's NMIMS University MPSTME, Shirpur, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_52

the gods and goddesses worshiped in every Indian household. Jewelry apart from ornamentation is considered as an investment that secures the future in times of mishappening. The purchase of gold jewelry is the preferred investment for bank deposits. Jewelry can be purchased using the conventional method from either goldsmiths or jewelry marts but now consumers have started to realize the reality of this pandemic and are trying to adapt to online shopping.

Customers hesitate to buy jewelry online as they fear that the jewelry sold online lacks quality. In online shopping, buyers cannot try or feel the material, which makes them reluctant toward it. Customer satisfaction is very low by just watching the photographs and videos of the jewelry and can be improved if a feature is introduced wherein they can try the jewelry pieces on themselves. This type of experience may propel the customer to buy jewelry. The use of technology like augmented reality can solve such problems. Digital information overspread onto an actual physical world is called augmented reality (AR). It is an enhancement of a person's actual world with some imaginary virtual objects present in it. Thus, a real environment supplemented with digital images generated by a computer.

There are various organizations which are making use of the AR and VR in many different manners. Some of them are: IKEA utilizes AR and in this way permits its clients to imagine household items in their homes continuously utilizing cell phone applications. The organization like Amazon and Target gives their clients who are iOS users visualization of the online items in their own living space and Wayfair offers similar assistance for the android clients. Converse is making use of this technology by permitting its clients to imagine shoes on their feet by using cell phone applications. Tanishq is executing MirrAR, an augmented reality software program as a team with StyleDotMe, a startup zeroed in on the imaginative utilization of augmented reality (AR) and artificial intelligence (AI) for giving a unique experience to customers who are keen on the gems and jewelry industry. Utilizing this unique platform clients can virtually try-on the jewelry, without really wearing it. Tanishq has made a major step to be more available to its clients by stepping into the augmented reality technology experience at the Bangalore and Delhi air terminals. For the first time in India, a jewelry brand is doing an augmented reality/hybrid reality (a combination of physical space augmented reality) campaign at an air terminal to draw in a massive crowd. PCJ is executing the 'AR Experience' on the iPad with 'MirrAR,' by StyleDotMe, a stage that lets clients basically try-on the jewelry without really wearing it and get expert styling views before making a buy. This would expand the customers' experience and influence their purchasing decisions as well as decrease the prerequisite for the in-store stock on a large level.

With the use of augmented reality, customers will be able to virtually try-on jewelry and this will make the customers' experience exciting and influence their purchasing decisions. It provides an integration of the physical world with the virtual world to make the experience real and adaptable for the shoppers. The following figure (Fig. 1) illustrates the use of AR and VR in online jewelry shopping.

Figure 2 explains the work model, and a customer does not need to register or login to go through the products on the website unless the customer wants to add the items in the wishlist or cart. The customers can either search for the desired product

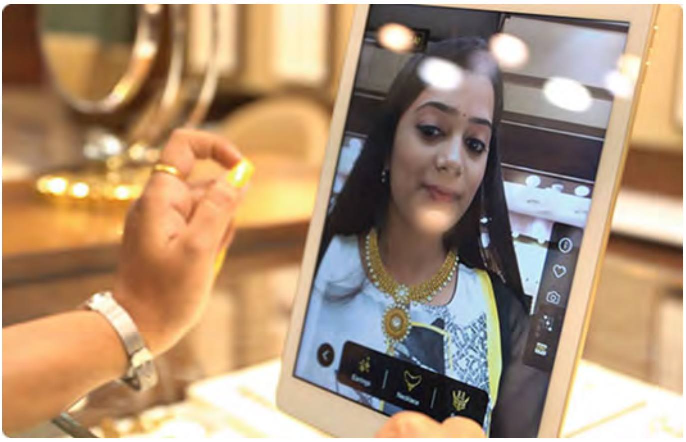


Fig. 1 Virtual jewelry try by face detection [1]

or can browse through the category section to look for the product. After selecting a product, the customer can virtually try it on, then proceed with further actions.

If the customer likes the product, then they can either add or remove the product to the wishlist or in the cart. Once the items are added to the cart, the customer can either choose to checkout or continue shopping. If the customer is already logged in, then he/she can continue shopping by entering all the necessary details and the shipping address. If the customer is a new then, he/she needs to register first and then can proceed to finish the shopping. After entering all the details for the shipping, the customer will be redirected to the payment gateway where the customer can either choose to pay using a credit card or debit card.

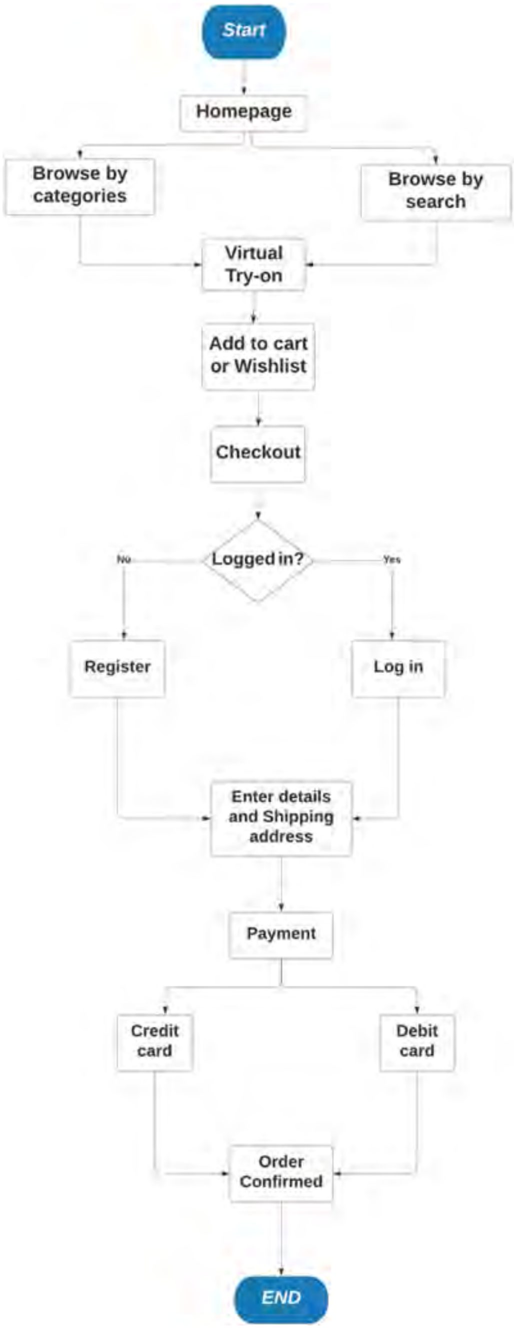
Once the payment is done, the user will receive the message of order confirmation and a receipt of the items purchased.

2 Technologies

Augmented reality (AR) is an artificial environment that contains a blend of physical elements and the digital world. AR involves mapping, computer vision as well as depth tracking to display suitable information to the user. This usefulness permits cameras to gather, send, and process information to display the appropriate content for the appropriate user digitally. Thus, this integration of real-time content with a virtual environment can be experienced by a smartphone or some special hardware and is easily accessible by all age groups. Hence, it is effective as well as feasible.

Virtual reality (VR) is a computer-generated simulation of an alternate world or reality. The user can experience this using a sensory device like headsets or gloves with a computer or mobile phones. To limit interaction with the real world and for a real experience, a VR headset is used by the user.

Fig. 2 Process flow of online jewelry shopping using virtual try



3 Literature Survey

Digital technologies have been an integral factor in removing the factors affecting the in-store purchase. Virtual reality (VR) and augmented reality (AR) are transforming the e-commerce business experience, by generating interactive data reports.

One such technology is 3-D cameras which helps in face recognition and mapping it just like the purchaser which gets showcased onto virtual mirrors [2].

In some methods when a virtual object is represented using AR in a jewelry shop application, the object is fixed on the screen and the customer has to move according to the object to try it on. But the markers on the customer's face using the HAAR algorithm, which will result in the object being managed according to the customer in real-time, this reconstruction of the 3D object is performed using the ICP algorithm [3].

With the increasing use of e-commerce, the demand for diamond jewelry appears to have some new trends. The consumers buying online diamond jewelry depicted a certain behavior which accounts for some major influencing factors in the marketing tactics of online selling [4]. The virtual jewelry system is based on AR toolkit technology which enables the customers to virtually try-on the jewelry on themselves before receiving the actual product, this helps in increasing customer satisfaction and enrich their experience bridging the gap between apparel goods and photographs [5]. The jewelry and gem industry in India operates at all levels starting from mining to sourcing, processing, designing, cutting, and polishing [6]. Industry experts say that the high prices of precious metals do not affect the sales of gold jewelry in India as it is one of the largest consumers of gold in the world. The Indian jewelry industry has been gaining prominence as an international sourcing destination for high-quality jewelry [7].

Online shopping has allowed brands and even local shops to go multi-channel and increase their reach and access the untouched markets, increasing their customer base and targeting a more specific consumer base. Online business depends mainly on how well the company adopts new technology and follows the latest trends and most importantly how well they analyze their data and optimizes according to that data. Augmented reality has introduced a new dimension to online selling platforms [8].

In the rapid development of technology, an application developed by a company allows its customers to view the company's collection of rings, from which the user can choose and customize it according to their own choice and then a fully rendered 3D model of the ring is rendered onto the user's finger [9].

E-shopping can be taken to a higher and more developed level with convenience and user-friendly interface using augmented reality in these technologically progressive times. An innovative shopping application pseudo-eye which uses augmented reality and can be used on various platforms uses the content-based image retrieval technique to search the captured image of a product and retrieve its details from the database. The product finder in pseudo-eye is implemented using the Ricoh visual search (RVS) technology. It enables the users to get an in-store experience which helps them in easing their shopping experience [10].

The proposed method of Bayesian skin classifier detects the skin pixel values and then places the jewelry virtually on the body. For example, a bangle after classifying the skin pixels fits the hand perfectly and the object will fit according to the user's movements. The particular design of a bangle can be selected using GUI buttons. This can be used in both android and iOS platforms [11].

In a method developed to focus on the emotional and sensitive perception of the users, half-tone dot effects (screen-tone effects) were created. To show the texture, shape, and appearance of the garment on the user's body contour, the image of the selected garment is created on top of the clothes that the user is already wearing. This is done using small dots [12].

Virtual trial rooms have gained a lot of importance among online shoppers as it provides a hassle-free experience, avoiding the long wait times in the queues. The Microsoft Kinect V2 sensor is used to provide the user with the in-store experience [13].

The virtual mirror technology was created to visualize customized shoes that a person can try-on virtually. It is executed by using AR techniques which combine videos that are real time with objects being displayed virtually [14].

EyeBuy is a technology developed to help people who have eyesight issues or color blindness. The application when active scans the products using the device's camera and matches the product with the application's stored database, after recognizing the product, EyeBuy in an AR environment shows the nutritional information and content of the product. The information is bigger in size and available in an easy and fast manner with appropriate contrast even for colorblind people [15].

Employing the Browser/Search framework and Java along with the JSP technology, the management of an online shopping module can be accomplished. The online shopping framework depends on the browser/search method. This module focuses on giving the best experience of online shopping to the clients in their own comfort and without any time restriction. The shoppers will be provided with all the information about the product regardless of the time [16].

4 Figures and Tables

The data involves people of various age groups, one can see the trend and shift in online jewelry shopping according to different people belonging to different age groups to provide a broader view of the upcoming trend of online shopping and integration of augmented reality.

While considering various mindsets and studying the below pie chart (Fig. 3) more percentage of people are uncertain about buying jewelry online, which in turn is an opportunity to mold their mindsets in the future, considering a good increase in buyers through online shopping. About one-fourth of people would prefer the conventional method, but the rest three-fourth of the population can engage in online

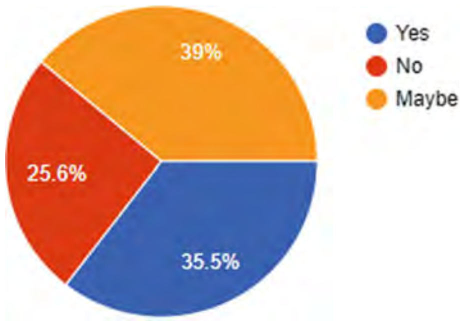


Fig. 3 Number of people preferring online jewelry shopping

buying considering proper safety measures, secure payment options, design concerns, certificates, quality, and other important aspects.

The article preferences are showcased in Fig. 4. Earrings among the top products which are likely to be purchased online. Having different designs, sizes, shapes, material, and being a handy item to purchase, it is mostly preferred by people of almost every age group, considering that earrings have less to do with the face cut and skin tone of a person, in general, is listed followed by rings. The people prefer small and handy jewelry items more than bulky or heavy ones. The least preferred items are nose rings and bangles among women, which indicates a unpopular trend in the online market.

Apart from the articles, one needs to understand the type of jewelry to regulate the purchase ranging from real (gold, silver, platinum, diamond, etc.) to artificial. The online platform is mostly preferred for the sale of artificial jewelry with about more than 60% of people engaging in its buying, while due to the various certificates, exchange value, and quality assurance the real jewelry is preferred by only 5% of the people as shown in Fig. 5.

The latest trends are shifted in the situation of pandemic wherein people prefer social distancing and staying safe at home, there has been a significant rise in the number of online buyers. Among all the customers, 70% of them prefer online mode for jewelry shopping (Fig. 6).

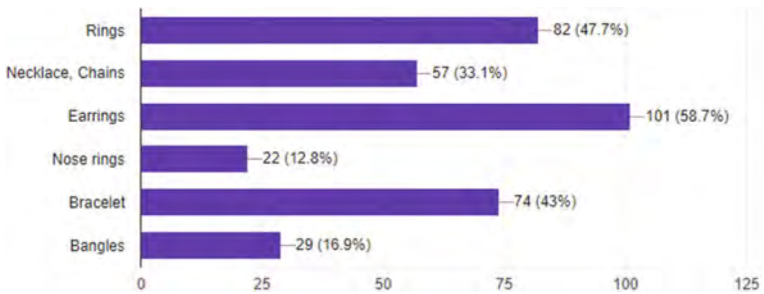


Fig. 4 Different preferences of the article in online jewelry shopping

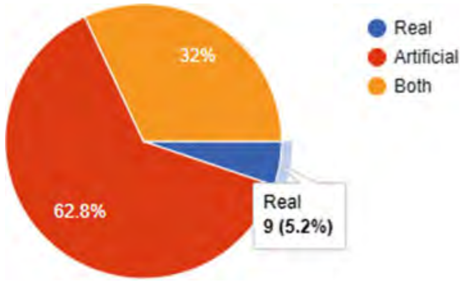


Fig. 5 Type of jewelry preferred in online shopping

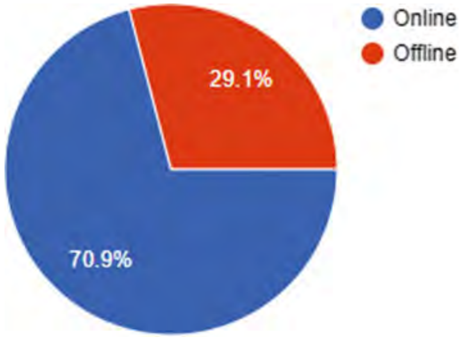


Fig. 6 Depicting shift in COVID-19

The users explore by enhancing their purchase that has led to an increase in customer-based and reduced the number of abandoned carts. Thus, to achieve such a personalized experience and to go hand in hand with the conventional buying method augmented reality integrated platform wherein you can try and buy is preferred by 80% of the people(Fig. 7). One can try the jewelry items on themselves, ask people for review by capturing their image, view the same product with the specifications at their comfort by just looking into the camera of their smartphone or laptops.

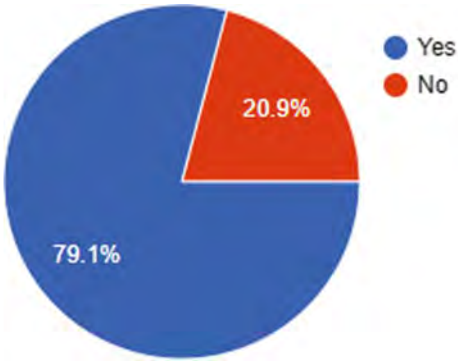


Fig. 7 Virtual try in online shopping

In the modern trends, considering the current pandemic scenario, online jewelry shopping with integrated augmented reality is the need of the hour and everybody is longing for it.

5 Conclusion

From the analysis of the concerned sector during the pandemic the shopping experience is more digitalized. To increase sales in the jewelry sector, augmented reality is introduced which will allow the customers to try-on jewelry without physical touch. Augmented reality is mixed with the real world to provide the customers the joy of shopping while sitting at their home. The new virtual jewelry experience is not only fascinating but also makes the procedure of jewelry selection and selling easier. In this era of social distancing and sanitization, to make everything contact-free and safe, augmented reality will play a major role.

References

1. https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.ammazza.me%2Ftreasure%2F&psig=AOvVaw0UkT6XexMyoxwwsma5QWHG&ust=1598636669435000&source=images&cd=vfe&ved=0CA0QjhxqFwoTCMC3h-f7u-sCFQAAAAAdAAAAABAD
2. S. Jain, N. Hada, S. Joshi, Shopping using virtual reality and augmented reality. IJAST **29**(04), 7546 (2020)
3. Himanshu More, Gaurav Salunke, Rigved Shete, Prakash Kawade, Pravin Nimbalkar, Virtual object representation using augmented reality. Int. J. Eng. Res. Technol. (IJERT) **08**(11), (2019)
4. Y. Yang, Q. Li, A quantitative research on ınfluencing factors of the behavior that consumers buying Diamond Jewelry online. in *2010 International Conference on Internet Technology and Applications* (Wuhan, 2010), pp. 1–4. https://doi.org/10.1109/ITAPP.2010.5566582
5. J. Jiang, Y. Kuang, The research of developing virtual jewelry worn system based on ARToolkit. Appl. Mech. Mater. **713–715**, 2382–2386 (2015). https://doi.org/10.4028/www.scientific.net/amm.713-715.2382
6. Ann Jacob, N.V. Sreedharan, A study on jewelry and gem industry in India. Int. J. Manage. Soc. Sci. **05**(04), (2017). ISSN: 2321–1784. Impact factor: 6.178
7. Chellam, Chitra. (2018). Jewelry industry in India. Contribution of jewelry industry in India. Int. J. Mark. Technol. **8**(2), (2018). ISSN: 2249–1058. Impact factor: 6.559
8. S. Chakraborty, D. Gupta, A study of the factors impacting the adoption of augmented reality in online purchases in India. in *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT)* (Bangalore, 2017), pp. 1526–1529. https://doi.org/10.1109/RTEICT.2017.8256853
9. Oche A. Egaji, Ikram Asghar, Will Warren, Mark Griffiths, Simon Evans, An augmented reality application for personalised diamond shopping. in *Proceedings of the 25th International Conference on Automation and Computing* (Lancaster University, Lancaster UK, 5–7 Sept 2019)
10. S. Ananda Kanagaraj, G. Arjun, T. Sree Sharmila, Pseudo eye: the next-generation shopping application using augmented reality. in *The 8th International Conference for Internet Technology and Secured Transactions (ICITST-2013)*

11. G. Radhika, T. Senthil Kumar, N. Mangayarkarasi, R. Nirmala, Virtual try of a jewelry using skin classifiers. IJCTA **9**(5), 2275–2278 (2016)
12. S. Hah, J. Park, J. Kim, Sensitivity of image based augmented reality fitting simulation. in *2011 IEEE International Symposium on Mixed and Augmented Reality—Arts, Media, and Humanities*, (Basel, 2011), pp. 107-108. https://doi.org/10.1109/ISMAR-AMH.2011.6093667
13. K.M.A. Yousef, B.J. Mohd, M. Al-Omari, Kinect-based virtual try-on system: a case study. in *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)* (Amman, Jordan, 2019), pp. 91–96. https://doi.org/10.1109/JEEIT.2019.8717498
14. A. Kusumaningsih, A. Kurniawati, C.V. Angkoso, E.M. Yuniarno, M. Hariadi, User experience measurement on virtual dressing room of Madura batik clothes. in *2017 International Conference on Sustainable Information Engineering and Technology (SIET)* (Malang, 2017), pp. 203–208
15. C.E. Forte, S. Luciano, C.V. Pazini, M. Marengoni, EyeBuy: using augmented reality in accessible interfaces for consumers with visual acuity reduction. in *2013 1st Workshop on Virtual and Augmented Assistive Technology (VAAT)* (Lake Buena Vista, FL, 2013), pp. 1-6. https://doi.org/10.1109/VAAT.2013.6786183
16. F. Wei, Q. Zhang, Design and implementation of online shopping system based on B/S model. MATEC Web Conf. **246**, 03033 (2018). https://doi.org/10.1051/matecconf/201824603033

Age and Volume Detection of Heartwood and Sapwood in Scots Pine Species Using Machine Learning

Piyush Juyal and Sachin Sharma

Abstract Pine trees are a vital asset for many companies. They are used for a broad spectrum of purposes. It is the most commercially important species of tree. It provides timber for construction, and its resin or oleoresin is used as a sealant, glue, and varnish. It is further refined into rosin, which facilitates improved grip between objects and turpentine oil. Heartwood and pine tree sapwood are used to extract timber and pine resin. Examining its wide use, through a picture containing a tree, trunk, and a guide established a method to estimate the age, radial width, and height of heartwood and sapwood. To precisely locate the relation, trunk, and the tree, the R-CNN mask is employed. The heartwood and sapwood volume are measured to generate the output.

Keywords Tree volume · R-CNN mask · Heartwood · Sapwood

1 Introduction

A pine tree is a member of the genus Pinus. It includes around 120-pine family (Pinaceae) species of evergreen conifers, which are scattered throughout the globe but mainly native to the temperate northern regions. They are resinous trees that grow up to 3–80 m in height and reach around 15–45 m in height for most species. Pines have long life spans, typically reaching 100–1000 years of age and some even more. With needles and cone scales arranged in Fibonacci number ratios, its branches can be seen to have spiral growth. Most pine trees have dense and scaly bark, but thin, flaky bark may be present in some plants. Pines are called sapwoods, but when used commercially, they can be denoted as sap or heart pines. Their stalked cones have scales that lack prickles and a small amount of resin. Their timber is

P. Juyal (✉) · S. Sharma
Department of Computer Science and Engineering, Graphic Era Deemed To Be University, Dehradun, India

S. Sharma
e-mail: sachin.cse@geu.ac.in

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_53

thin whitish sapwood and close grained. Hard pines have fairly hardwood, including Scotch, loblolly, and Corsican pines. They have cone scales with prickles and a significant amount of resin. Their timber, with pale and dense sapwood, is coarse-grained and usually dark-colored. In the building and paper-product sectors, the principal economic benefit of pines is Terpentin, rosin, oils, and wood tars which are also sources. Several species have edible pine seeds, which are sold as pine nuts or pinyons. There are two pieces of the pine tree truck: one is heartwood, and the other is sapwood. Generally, the entire pine tree is not used as lumber, but rather as heartwood. It is the portion of the sapwood that is within the tree. The pith of the tree is used to refer to the term. Heartwood, as well, the dead, central wood of the tree, is known as duramen. Tannins or other elements are usually included in the cells that give it a darker tone and often an aromatic property. It is mechanically durable, less resistant to decay, and less easily penetrated by wood-preservative chemicals than any other form of wood. The resin of the Oleo pine originates from the sapwood. It is lighter in color but has the same intensity as heartwood. Sapwood conducts sap, which strengthens the stem and acts as a food storage tank to some degree. Through the tapping process, oleoresin is extracted as gum rosin and turpentine. Pine oleoresin is a comprehensive source of useful terpenes for distillation of the oleoresin. It has two major divisions such as turpentine and volatile fraction and rosin, where the volatile fraction and rosin are a significant fraction. The basic component of tree protection is oleoresin. Due to the various traditional and potential uses of its terpenes, it is a critical non-wood forestry commodity (Fig. 1).

Our reliance on pine trees is proved by the enormous international requirement for by-products of pine resins in industry and long-term production of this feedstock by large commercial plantations. Our suggested methodology offers more in-depth

Fig. 1 An image of a pine tree

insights into a pine tree through the image and comparison of the tree. These observations will help to maximize or even inspire growers to better pick trees that will increase their yield. The insights can be obtained from the picture that are the age of the pine tree, the diameter of the tree, the height of the tree, and the radial width of the heartwood, and then the volume of the heartwood and sapwood is computed. The age of heartwood and sapwood can be measured too.

2 Literature Survey

A relationship is proposed between heartwood and age, and heartwood with growth rate. It mathematically supported their results and confirmed that they were tested against observed values [1]. This article records the pine heartwood age law, which has captured our attention in particular. The other laws are also useful, such as heartwood radial fraction and ring width in sawlogs. It has laid the stage for contemporary research on heartwood. It has been shown that the comparative quantity of heartwood produced differs widely within and among trees [2]. Regardless of tree species, he developed a pipe-model theory. In [1], authors assumed a simple ring to ring relation that proclaimed a constant number of sapwood rings beneath the living crown. However, the constant area of sapwood and the constant number of rings have not been confirmed, so age remains the most important factor in the relationship between heartwood and sapwood. In [3], the authors described the heartwood and sapwood of a tree in detail. This provides beginners with optimal insights into these subjects. In [4], the authors illustrated the amount of sapwood and heartwood. They have sample 12 trees at nine height stem levels. They have found that heartwood diameter and sapwood width were positively correlated with the radial growth of the tree. In [5], the variation of heartwood amount in Dalbergis sissoo in different zones of Punjab is discussed. They have conducted the study in three broad agro-climatic zones. These zones are viz. central plain zone and southwestern zone. They have concluded that there was no difference in heartwood amount in localities within each of these zones and the thickness of heartwood had a strong linear relationship with the total diameter in all zones. In [6], the dimensional relationship of heartwood and sapwood compared to the trunk length of 11 different timber trees is being analyzed. They have found that tree trunk shows a positive correlation with the circumference, cross-sectional area of heartwood and the sapwood, and heartwood area ratio. In [7], automated blood cell counting and its classification proposed using CNN and mask R-CNN. They have identified and counted red and white blood cells. In [8], the authors have discussed proper management and utilization of pine trees in the forest of Uttarakhand. In [9], the authors explore the difference between the strengths of sapwood and heartwood for all practical purpose. A conclusion of negligible difference has been given.

3 Proposed Methodology

Initially, the objects are detected and interested in gathering any inferences from pictures. The state-of-the-art object detector and mask R-CNN were used to solve our puzzle. The tree trunk is captured by our proposed model and identifies any white rectangular board, so that held as a guide along with it [9, 10]. Similarly, following a guide can capture the entire tree and the dimension of the relation, so that the trunk's dimensions and the height of the tree can be calculated. The basal area is measured to determine the volume and then calculate a product using basal area, height, and tree-like constants. The formula for basal area is this:

$$\mathrm{BA}(\mathrm{m}^2) = \mathrm{pi} * \mathrm{DBH}(\mathrm{cm}^2)/40000 \tag{1}$$

BA here is the basal area, and DBH denotes diameter at breast height that will determined a little further. The equation to calculate the tree volume will be:

$$V = 0.42 * \mathrm{BA} * h \tag{2}$$

V and h are volume and height of the pine tree, respectively.

For the training model, the data is obtained from the Forest Research Institute (FRI) [11] in Dehradun. The value of the height and diameter estimations is approximate, and the dataset contains images that are similar to Fig. 2 to train the network. The reference is kept in parallel with the tree. The data contains about 400 images, with 325 examples of reference and 408 tree examples. Figure 3 illustrates a proposed model that makes instance segmentation feasible. Instances segmentation is through pixel-level recognition of the object. The R-CNN is a convolutional neural network

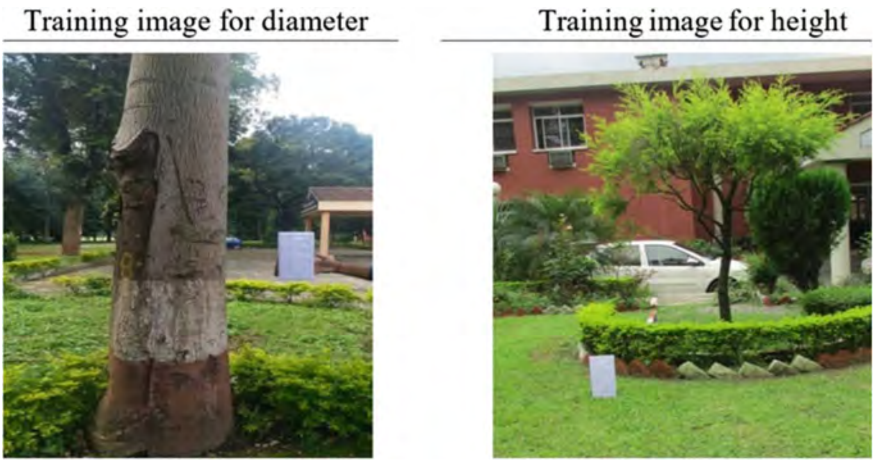


Fig. 2 Images of trees for training the proposed model

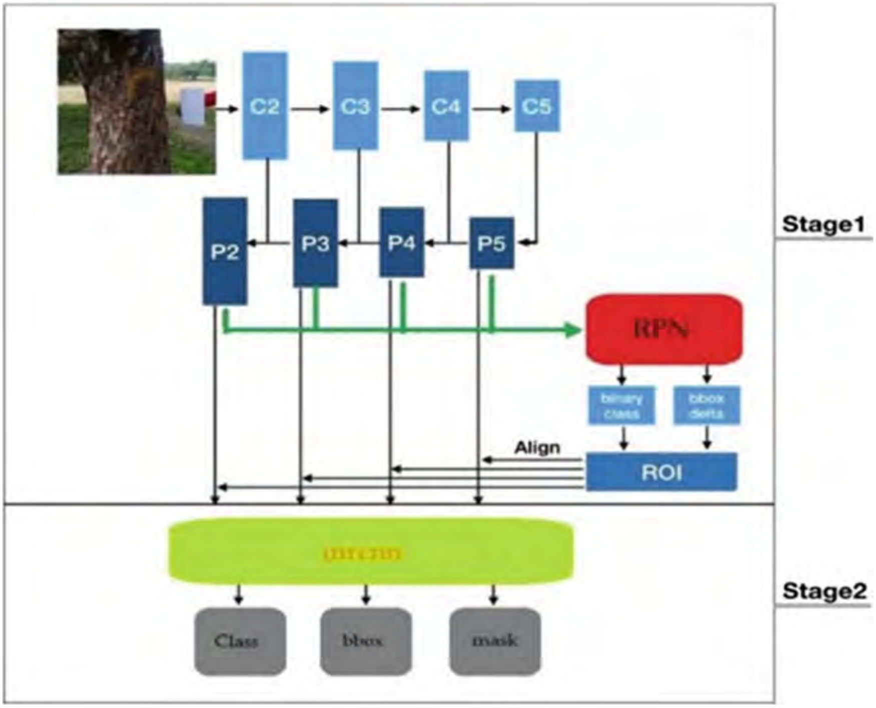


Fig. 3 Proposed methodology using mask structure of R-CNN

centered in an area. The R-CNN mask distinguishes the various objects in an image separately [12, 13]. Mask R-CNN provides us with object binding boxes, classes, and masks. It is a more responsive R-CNN architecture update [14]. The upgrade is the branch of the object that is capable of creating a binary mask. The binary mask indicates whether the pixel concerned is a part of an entity. The branch of the item is an incredibly small overhead to the quicker R-CNN, which makes it inexpensive. It consists of two stages:

Stage 1: Providing recommendations for locations where the object may be found.

Stage 2: Classifying the class of the object, refining the bounding box, and creating a mask centered on the proposed regions at the pixel level of the object.

4 Implementation of Proposed Methodology and Result Analysis

The diameter of the tree is used to calculate the age of the tree. The tree's diameter and its height are used to determine its volume. To calculate the diameter of the pine

Fig. 4 Measuring horizontal distance of a pine tree

tree at first, the trunk must be identified. The detection will cover the area in Fig. 4 labeled as horizontal distance. Simultaneously, the reference and the tree trunk are detected to find out the width per pixel. This is a two-step process. In first step, find the width per pixel by utilizing the detected coordinates of reference (X_{min} and X_{max}). And then, divide it by the actual width of the reference.

$$\text{Multiplier} = (X_{\text{max}} - X_{\text{min}})/(\text{width of the reference}) \tag{3}$$

The multiplier is the length of one pixel.

In the second step, the coordinates of the detected trunk ($X1_{\text{max}}$ and $X2_{\text{max}}$) is utilized to calculate the number of pixels horizontally.

$$\text{Horizontal distance} = (X1_{\text{max}} - X1_{\text{min}}) \tag{4}$$

The diameter of the tree is calculated from the breast height which is multiplied with Eqs. (3) and (4).

$$\text{DBH} = \text{multiplier} * \text{horizontal distance} \tag{5}$$

The plot of mask R-CNN includes class loss and bounding box loss for the model that has trained for detecting the tree trunk and the reference.

Figure 5 illustrates the training loss for diameter detection of a pine tree, and Fig. 6 explores the validation loss for diameter detection of a pine tree.

According to the International Society of Arboriculture, the growth factor for the pine tree is five. Hence, the diameter (inches) of the pine tree at breast height is multiplied by 5. A pine tree with a breast height circumference of 70 inches will be about 110 years old. It can then be utilized to find out the age of the heartwood in pine trees. Pine tree age can also be determined by counting the number of rings in the xylem which will require cutting down the tree. As there is a high correlation between the number of rings and age, it can be concluded that the age will give the number of rings; hence, the pine heartwood age rule can be utilized by the [1] number of rings in xylem. Ac, AH, As: no. of annual rings in xylem, heart and sapwood.

$$\mathrm{Ac} = \mathrm{AH} + \mathrm{As} \tag{6}$$

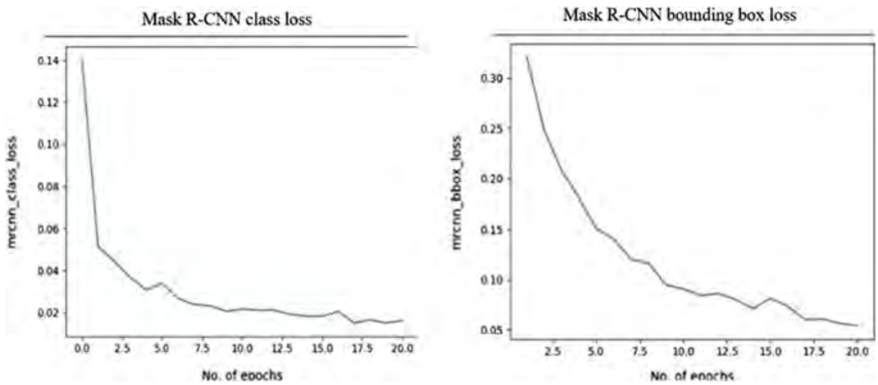


Fig. 5 Training loss for diameter detection of pine tree

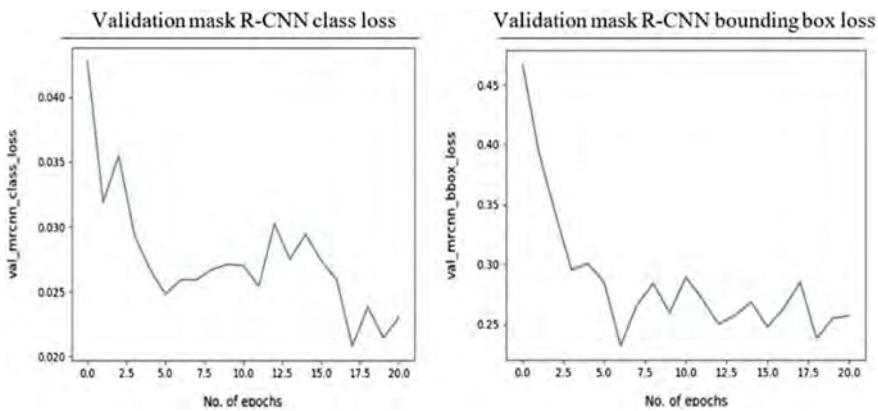


Fig. 6 Validation loss for diameter detection of pine tree

The Ac is determined by the age of the pine tree that can be calculated. The age of heartwood (AH) is calculated by the formula.

$$\mathrm{AH} = (\sqrt{\mathrm{Ac}} - 3.0)^2 \tag{7}$$

To estimate the volume of the heartwood, the diameter of heartwood is calculated. The complete diameter (Rc) of the pine tree consists of the following: Rc. RH, and Rs; where Rs is the radial width in xylem, heart, and sapwood (mm).

$$\mathrm{Rc} = \mathrm{RH} + \mathrm{Rs} \tag{8}$$

Derived variables:

$$\mathrm{RFH} = \mathrm{RH}/\mathrm{Rc} \text{ radial fraction of heartwood} \tag{9}$$

$$\mathrm{RWC} = \mathrm{Rc}/\mathrm{Ac}\text{: mean ring width for xylem} \tag{10}$$

The ring width in sawlog formula:

$$\mathrm{RWC} = \mathrm{Rc}/(2.6 + 0.013 * \mathrm{Rc} + 8.0 * \mathrm{RFH})^2 \tag{11}$$

The below-mentioned formula is framed using Eq. (8)–(11).

$$\mathrm{RH} = \mathrm{Rc}\,(\sqrt{\mathrm{Ac}} - 2.6 - 0.013 * \mathrm{Rc})/8.0 \tag{12}$$

With RH known, the value of Rs from Eq. (9) can be obtained. Now, to calculate the radial width of both heartwood and sapwood, the height of the heartwood and sapwood is required.

In this article, a different model is trained to detect the entire tree simultaneously with a reference. The detection will cover the area in Fig. 7 that illustrates the vertical distance measurement of the pine tree. Using the reference, the height per pixel can be obtained. This is a two-step process. The coordinates can be utilized on Y-axis detected by the mask R-CNN. The $Y_{\min}$ and $Y_{\max}$ can be detected by the reference to calculate the height per pixel.

$$\text{Multiplier} = (Y_{\max} - Y_{\min})/(\text{height of the reference}) \tag{13}$$

The multiplier is the height per pixel. The second task is to find the vertical distance that can be utilized by the coordinates of the detected tree.

$$\text{Vertical distance} = (Y1_{\max} - Y1_{\min}) \tag{14}$$

The result from Eqs. (14) and (15) can be multiplied to get the height

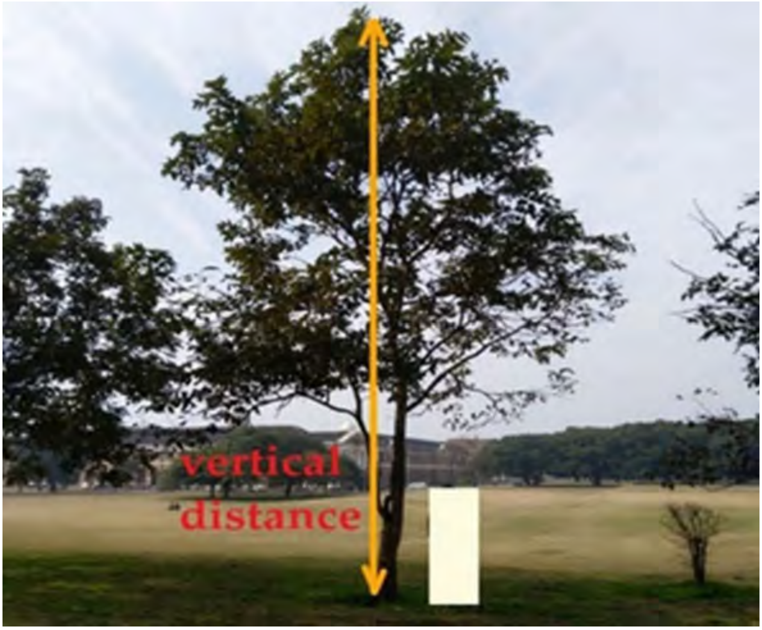


Fig. 7 Measuring the vertical distance of pine tree

$$\text{Height of tree} = \text{multiplier} * \text{horizontal distance} \tag{15}$$

The mask R-CNN class loss and bounding box loss for detecting height of a pine tree are shown in Fig. 8. Figure 9 explores the validation loss for height detection of pine tree.

Now, the height of the tree and the heartwood sapwood radial width are calculated. So, using Eqs. (1) and (2), calculate the total volume.

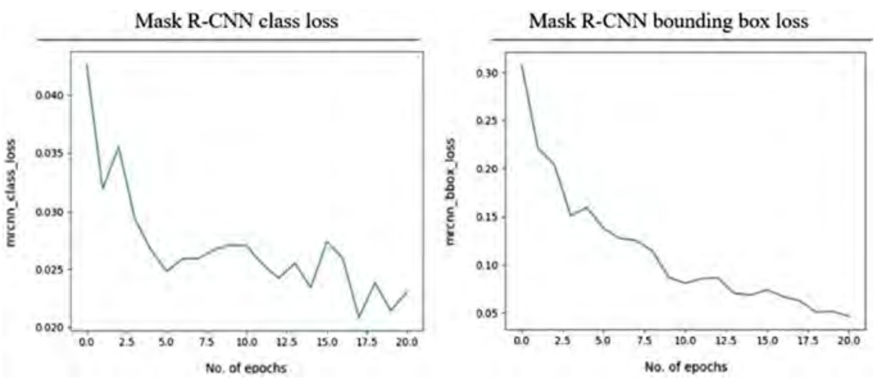


Fig. 8 Training loss for height detection of pine tree

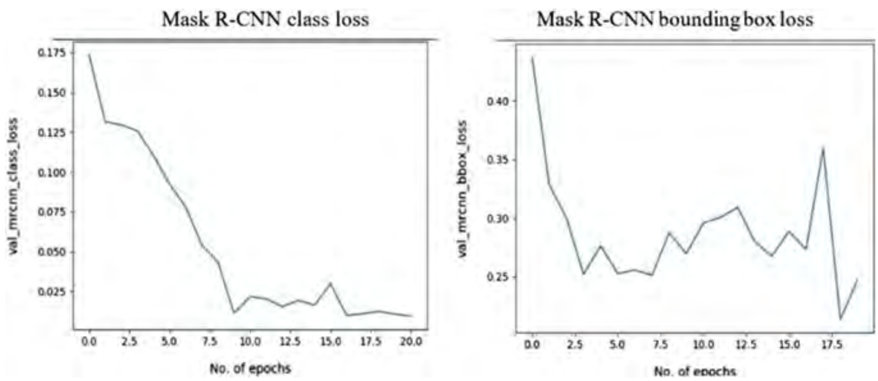


Fig. 9 Validation loss for height detection of pine tree

$$\mathrm{BA}_1(\mathrm{m}^2) = \mathrm{pi} * \mathrm{DBH}(\mathrm{cm})^2/40,000$$

$$\mathrm{Volume} = 0.42 * \mathrm{BA} * \text{height of the tree}$$

The volume of the heartwood is calculated by switching the DBH with the RH (radial width of heartwood).

$$\mathrm{BA}_2\left(\mathrm{m}^2\right) = \mathrm{pi} * \mathrm{RH}(\mathrm{cm})^2/40,000 \tag{16}$$

$$\mathrm{Volume\,H} = 0.42 * \mathrm{BA} * \text{height of the tree}$$

The volume is subtracted from the volume H to find the volume of sapwood.

$$\mathrm{Volume}\ S = \mathrm{Volume} - \mathrm{Volume}\ H$$

The results of each model are shown in Figs. 10 and 11. Figure 10 explores the age of the pine tree that is 39 years. The heartwood age is 10.5 years or 11 annual rings in the heartwood. The sapwood annual rings are 28v and the radial width of heartwood is 2.6 cm. The diameter of heartwood is 5.2 cm. Figure 11 illustrates the total basal area of the pine tree which is 0.0314 m^2. The heartwood basal area is 0.0021 m^2. The total volume is 0.098 m^3. The heartwood volume is 0.006 m^3, and the sapwood volume is 0.0092 m^3. To measure the accuracy of the model, we have calculated mAP. It is the mean average precision of the model. These metrics are extensively used for gauging the performance of an object detector. The mAP for diameter estimation is 0.92 while the mAP for height estimation model is 0.86.

Fig. 10 Age detection of pine tree

5 Conclusion

The investigation of the woods by using these techniques can bring great insights into the trees. Biomass and treetop values are clearly correlated with the tree volume; using the state-of-the-art deep learning algorithm mask R-CNN, a simple method is established for calculating treetop volume. If applied correctly, it will save time and effort and help you function effectively and rapidly. This application can become more precise and robust if more information is therefore available with more data, and a deeper analysis algorithm such as mask R-CNN would have further precise results.

Fig. 11 Basal area detection of pine tree

References

1. P. Gjerdrum, Heartwood in relation to age and growth rate in Pinus sylvestris L. in Scandinavia. Forestry **76**(4), 413–424 (2003)
2. Piyush Juyal, Sachin Sharma, Estimation of tree volume using mask R-CNN based deep learning. in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (IEEE, 2020), pp. 1–6
3. S. Kellomäki, V.-P. Ikonen, H. Peltola, T. Kolström, Modelling the structural growth of Scots pine with implications for wood quality. Ecol. Model. **122**(1–2), 117–134 (1999)
4. S. Pallardy, *Physiology of Woody Plants*, 3rd. edn. (2008)
5. M.C. Morais, H. Pereira, Heartwood and sapwood variation in Eucalyptus globulus Labill. trees at the end of rotation for pulp wood production. Ann. For. Sci. **64**(6), 665–671 (2007)
6. L.B. Rishi, D.S. Sidhu, Effect of environment on heartwood amount in Dalbergia sissoo Roxb. in different climatic zones of Punjab. Indian Forester **128**(7), 779–785 (2002)
7. M.N.B. Nair, R.R. Chavan, Dimensional analysis of some wood parameters in eleven timber trees. Indian Forester **111**(6), 410–417 (1985)

8. N. Dhieb, H. Ghazzai, H. Besbes, Y. Massoud, An automated blood cells counting and classification framework using mask R-CNN deep learning model. in *2019 31st International Conference on Microelectronics (ICM)* (IEEE, 2019) pp. 300–303
9. G. Singh, A. Giri, S. Paul, Pinus roxburghii Sarg. (Chir Pine): a valuable forest resource of Uttarakhand. Indian Forester **143**(7), 700–709 (2017)
10. H. Trotter, Note on the comparative strengths of Sapwood and Heartwood. Indian Forester **63**(9), 598–601 (1937)
11. Piyush Juyal, Chitransh Kulshrestha, Sachin Sharma, Tejasvi Ghanshala, Common bamboo species identification using machine learning and deep learning algorithms. Int. J. Innov. Technol. Explor. Eng. (IJITEE) **9**(4), (2020). ISSN: 2278-3075
12. Shuchi Bhadula, Sachin Sharma, Piyush Juyal, and Chitransh Kulshrestha. Machine learning algorithms based skin disease detection. Int. J. Innov. Technol. Explor. Eng. (IJITEE) **9**(2), (2019). ISSN: 2278-3075
13. S. Yin, H. Li, L. Teng, Airport detection based on improved faster RCNN in large scale remote sensing images. Sens. Imag. **21**(1), 1–13 (2020)
14. M. Zeybel, Y.S. Akgul, Localization and ıdentification of lumbar ıntervertebral discs on spine MR ımages with faster RCNN based shortest path algorithm. in *Annual Conference on Medical Image Understanding and Analysis*, (Springer, Cham, 2020), pp. 143–154

Multi-layer Perceptron Training Using Hybridized Bat Algorithm

Luka Gajic, Dusan Cvetnic, Miodrag Zivkovic, Timea Bezdan, Nebojsa Bacanin, and Stefan Milosevic

Abstract Neural networks (NNs) are a subset in the field of machine learning (ML) that tends to make it possible for a machine to learn and make new predictions based on previous experiences and provided data. It is important to emphasize that there is no need to program this kind of behavior since the whole process is supported by the "self-adjustment" of the algorithm, which can evaluate itself and therefore adjust its parameters to get better performance and accuracy. Neural networks are different from other types of machine learning algorithms in such a way that they do not use statistical and mathematical models to make future predictions. Instead, they replicate the structure and the processes that happen inside the human brain. However, this type of learning is very computationally expensive since there is an enormous amount of states and conditions in which the network itself can be found. Therefore, it can be said that a process of learning for neural networks relates to the collection of NP-complete problems because of the large search space of possible solutions. Swarm intelligence (SI) algorithms can help to reduce this large space of solutions by finding a solution that is not optimal but close optimal and provide satisfactory results given how much longer it would take to train a network without them. In this paper, the authors have proposed a solution for the stated problem based on hybridized bat algorithm.

L. Gajic (✉) · D. Cvetnic · M. Zivkovic · T. Bezdan · N. Bacanin · S. Milosevic
Singidunum University, Danijelova 32, 11000 Belgrade, Serbia
e-mail: luka.gajic.17@singimail.rs

D. Cvetnic
e-mail: dusan.cvetnic.17@singimail.rs

M. Zivkovic
e-mail: mzivkovic@singidunum.ac.rs

T. Bezdan
e-mail: tbezdan@singidunum.ac.rs

N. Bacanin
e-mail: nbacanin@singidunum.ac.rs

S. Milosevic
e-mail: stefan.milosevic.191@singimail.rs

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_54

Keywords Neural networks · Swarm intelligence · Artificial intelligence · Machine learning

1 Introduction

Artificial neural networks (ANNs) [1] are the advantage in terms of computational intelligence, more precisely the ability for machines to make assumptions and predictions about a certain problem based on collected and processed data. They are of great benefit in situations, where traditional statistical methods cannot produce satisfying outcomes for a realistic time frame. Generally, the more data is provided to the network, the more trained and accurate the network would be. Another important factor is computational power; hence, high-end computers with larger computing capabilities will spend less-time training a network and also can train larger and more complex networks. These kinds of systems can be employed in different tasks throughout scientific and engineering fields; especially, it has great importance in the field of medicine. For example, they can successfully diagnose patients with different types of human diseases such as diabetes [2], Alzheimer's disease [3], and tuberculosis [4] as well as certain types of plant diseases such as rice and apple leaf diseases [5, 6].

1.1 The Architecture of Artificial Neural Network

It is already stated that the artificial neural networks will simulate the structure and processes inside the brain of humans. The brain represents a complicated structure of cells [1], cells that are reactive to external stimuli and are interconnected through a series of synaptic connections. Depending on the type of external stimuli, different neurons will be activated, and the signal will propagate to different neurons based on how the neurons are interconnected. In this paper, the training and test of the proposed solution were conducted using two datasets for classification. Further, it can describe the architecture of an ANN as a directed graph, where each node in that graph has a certain weight. Each node in a graph is called a neuron and is connected to other neurons with a connection that also has a weight that describes how strong a connection between two nodes is. Important parts of an ANN are the input layer, which maps to input data provided, and an output layer, which is responsible for presenting results. Using backpropagation, a weighted sum of neurons can be calculated and then calculate an error that ANN produced. The objective is to minimize this error using some type of cost function that is evaluated multiple times during the learning process by assigning different values of weights to ANN.

1.2 Multi-layer Perceptron

Multi-layer perceptron [7] (MLP) represents feedforward artificial neural network structure. Feedforward ANNs are not cyclic, which means that there cannot be any cyclic connection between two nodes, and the information flow is unidirectional, from input through hidden layers to the output of ANN. Compared to single-layer perceptron [8] (SLP), the most basic form of the neural network, MLPs can tackle problems with nonlinearity included using one or more hidden layers. MLPs use backpropagation [9, 10] to achieve self-adjustment and increase accuracy during multiple iterations. First of all, each neuron and connection in a network has assigned values of weights and biases, real numbers that take values from an interval $[-1, 1]$. Then, as the learning process starts, the products of inputs and weights are summed up, add bias, and get the output (1). (Fig. 1).

$$S_j = \sum_{i=1}^{n} \omega_{ij} I_i + \beta_j \tag{1}$$

where the output is denoted by S_j, ω_{ij} denotes the wights, and the bias term is denoted by β_j.

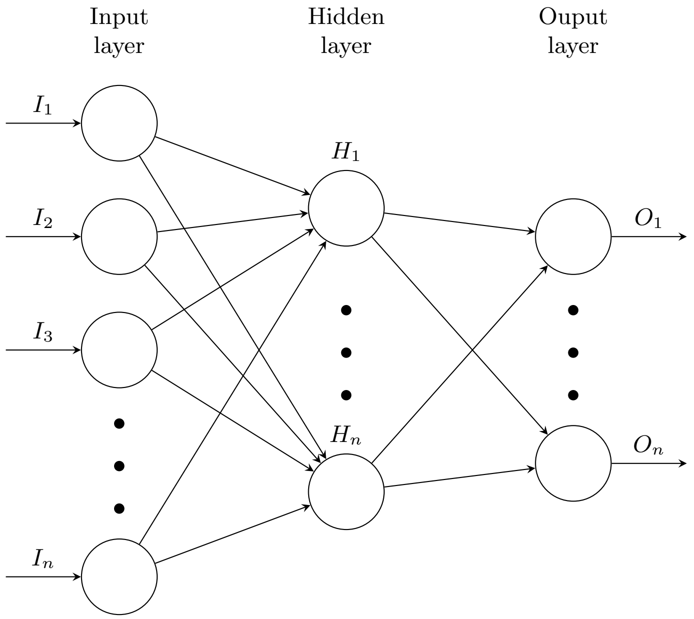


Fig. 1 The structure of multi-layer perceptron ANN

Based on the calculated error, using backpropagation, the weights of nodes are updated, and the whole method is repeated to he last iteration or some other criteria is reached. Additionally, each node in MLP can map nonlinear relationships among outputs and inputs with an activation function. One example of an activation function is S-shaped sigmoid function (2).

$$f_j(x) = \frac{1}{1 + e^{-S_j}} \tag{2}$$

The following pages are systematized in the next order: Sect. 2 provides an overview of ANN training and swarm intelligence, Sect. 3 explains the original bat algorithm as well as the proposed hybridized bat algorithm, Sect. 4 displays the proposed method organization and obtained outcomes, and Sect. 5 closes the paper.

2 ANN Training and Swarm Intelligence

The process of training an artificial neural network to make correct predictions consists of several steps that were previously described. However, all the necessary calculations performed in the process of ANN training are very computationally expensive and require a lot of computing resources as well as time to successfully finish. The search space of potentials solutions (correct weights) for a complex ANN can be extremely large and that is why this problem is often categorized as NP-hard optimization problem. The main reason for ANN training's NP-hardness is the backpropagation. Since the role of backpropagation is to adjust all weights in a network, it runs pretty slow. NP-hardness also states that while there are some instances of the problem that are solvable, the large set of instances can only be solved approximately, even if it gives them an extremely long computational time. Swarm intelligence can help us in reducing such a large search space to a subset of "sub-optimal" solutions which may not be the best in the search space but give pretty good results considering the amount of time saved, where the network would adjust the parameters by itself. This category of algorithms tackles common optimization problems using simulated behavior of many insects or animal colonies found in nature such as ants [11], bats [12], bees [13], hawks [14], and many more. Each of the algorithms has a group of agents that explore the search space individually, but together create a form of collective intelligence and can optimize a given function through multiple iterations. This kind of behavior has stimulated the scientific community to model such behaviors to implement them in problems from real life. Such problems can range from cloud computing [15–17], to the conservation of energy required for the proper and long-term functioning of wireless sensor networks [18–20], to the detection and classification of brain tumors [21]and convolutional neural network architecture optimizations [22–24]. Some papers have been researching the hybridization of SI algorithms to improve their overall performance. Algorithm hybridization blends the benefits of various algorithms, which results in the synergistic synthesis of fused algorithms. Also, some of them used SI algorithms to replace the backpropagation process in ANN training [25]. Determining the right values of the connection weights and biases of MLP is an optimization problem, and the two widely used methods are gradient-based training and stochastic optimization algorithms. The gradient-based algorithm can easily be stuck in the local optima. There are different stochastic gradient descent optimizations, such as adam, adadelta, momentum, adagrad, rmsprop, and adamax [26–28]. The objective of this paper is to propose an SI algorithm, the hybridized bat algorithm to optimize the connection weight and biases of MLP instead of using a gradient-based algorithm that has a disadvantage and which may be stuck in the local optima. The proposed method avoids this issue of getting stuck in the local optima and finds the right values of the wight and biases.

3 Original and Hybridized Bat Algorithm

Bat algorithm [12] relates to the collection of optimization algorithms, more precisely metaheuristics, and was introduced in 2010 by Xin She-Yang. The main goal of this method is to utilize bat echolocation behavior and find sub-optimal solutions in a given search space. The algorithm itself represents the behavior of microbats. They manage to distinguish insects and catch prey in total darkness. The important factor of their success is the echolocation, which helps them to go around the obstacles, detect, chase, and finally catch the desired prey in darkness.

3.1 Original Bat Algorithm

The original BA is constructed of three important rules [12] which must be followed:

1. Bats can detect prey using echolocation and can distinguish prey from an obstacle.
2. Bats are flying aimlessly with parameters v_i that represents velocity at position x_i. Also, they have a specific frequency and producing loudness that helps them in searching for prey. Frequency and wavelength serve for emitting pulses, and by adjusting frequency and wavelength, bats can control the emission rate, which is ranges from 0 to 1, depending of the distance from the location of the pray.
3. The value of the loudness varies from A_0, which is a positive large value, to the $A_{\min}$, which is a constant.

Initially, N random bats (individuals or solutions) are generated by the algorithm within the minimum (lower) and maximum (upper) boundaries as follows:

$$x_{i,j} = lb_j + \phi(ub_j - lb_j) \tag{3}$$

where a solution is denoted by $x_{i,j}$ and ϕ denotes a uniform random number. The notation lb_j refers to the lower bound, while ub_j indicates the upper bound.

The value of the frequency is updated at each iteration as follows:

$$f_i = f_{\min} + (f_{\max} - f_{\min})\beta \tag{4}$$

where f_i refers to the frequency of the ith solution, the value β is randomly chosen from the range [0, 1].

In the next step, the velocity of ith solution is calculated as:

$$v_i^t = v_i^{t-1} + (x_i^{t-1} - x^*)f_i \tag{5}$$

where the velocity is denoted by v, x^* indicates to the global best solution, and t is the iteration number.

The position of each bat, at time step t, is updated by using the following formula:

$$x_i^t = x_i^{t-1} + v_i^t \tag{6}$$

where the updated location is indicated by x_i^t, the current position is denoted by x_i^{t-1}, and v_i^t denotes the velocity at iteration t.

The local search is carried out by the random walk stochastic component, which can be defined as follows:

$$x_{\text{new}} = x_{\text{old}} + \epsilon A^t, \tag{7}$$

where A^t refers to the average value of all solutions' loudness, and ϵ denotes a random value between -1 and 1.

If a random number is less than the loudness value and if the best solution is replaced by a new solution, in other words, if a bat finds the prey, the loudness value will be smaller, and the pulse rate value will be greater:

$$A_i^t = \alpha A_i^{t-1},\;\; r_i^t = r_i^0[1 - exp(-\gamma t)], \tag{8}$$

$$A_i^t \to 0,\; r_i^t \to r_i^0,\;\; \text{while } t \to \infty \tag{9}$$

where the loudness is denoted by A_i^t, α and γ are constant values, and the pulse rate is indicated by r.

3.2 Hybridized Bat Algorithm

To avoid the low exploration ability and premature convergence of the original BA, a hybridization of the BA algorithm is proposed by implementing components from the artificial bee colony algorithm (ABC) [29], which results in a better balance of exploration and exploitation [30]. The hybrid BA incorporates the ABC's onlooker bee component.

The onlooker bee component enhances the exploitation process in the bat method, which is mathematically formulated as:

$$x_i = x_i + \phi(x_i - x_k) \tag{10}$$

where x_i denotes the new position, ϕ is a random value [0, 1], and x_k denotes the neighbor bat k.

The iteration number t will decide which component will be used in the algorithm; if the iteration t is odd, the bat search is utilized, and otherwise, the onlooker search is employed.

The exploration process remains the same as in the original bat algorithm, by employing the random walk equation (7) if a random value is greater than r_i. Otherwise, the solution is replaced by a random solution within the bounds of the minimum and maximum value of the best solution as follows:

$$x_{i,j} = x^*_{\min} + \phi(x^*_{\max} - x^*_{\min}) \tag{11}$$

The new hybridized algorithm is named as BAEABC. The algorithm can be described in detail by pseudocode 1.

4 Experiments and Discussion

In this section, it is described how the experiment is designed with the proposed BAEABC approach and also presents the setup of simulations and results that are obtained. This experiment has used an MLP with one hidden layer, where the connection weights and biases are subject to optimization, and the objective is to find such a set, which will minimize the test error rate. The value of the weights and biases is optimized with the proposed BAEABC approach, where the solutions represent the possible set of weights and biases.

The dataset features are the network inputs, which values are multiplied by the weights, the bias value is added to the weighted sum, the sigmoid activation function is applied, then the output values of the hidden layer are multiplied by weights and summed, the bias is added, and the output value defines the class. The loss function is utilized to calculate the error rate.

Algorithm 1 Pseudo-code of the proposed BAEABC

```
Initially generate the bat population randomly x_i (i = 1, 2, ..., n) and initialize the value of velocity v_i
Define loudness (A_i), pulse emission rate (r_i), and the the frequency of pulse (f_i)
Initialize the value of the iteration counter to 0
Define MaxIter (the value of maximum iteration)
while t < MaxIter do
  for i = 1 to N do
    if t%2 == 0 then
      The ABC's onlooker bee search mechanism is employed by using Eq. (10)
    else
      Employ the BA search component using Eq. (4), (5) and (6)
    end if
    if rand > r_i then
      Evaluate the solutions, sort the population and save the best location
      Perform the exploration by utilizing Eq. (7)
    end if
    Create new random individuals by utilizing 11
    if (f(x_i) < f(x_*) and rand < A_i then
      The updated position is obtained
      Update the values of A_i and r_i by using Eq. (8)
    end if
  end for
  Detect the fittest individual x*
end while
Output the fittest solution
```

Every population member is represented with a vector with only one dimension. Weights and biases are stored in each population member. And the whole population is a possible network structure. As stated before, the experiment utilized MLP with one hidden layer. The size of the one-dimensional vector is calculated using the following formula:

$$n_w = (n_x \times n_h + n_h) + (n_h \times n_o + n_o) \tag{12}$$

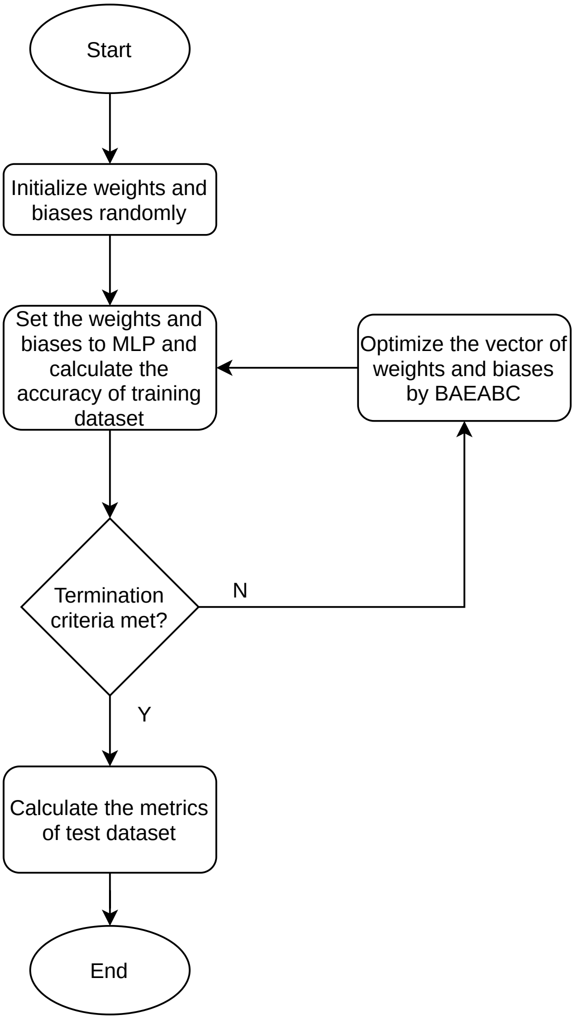
where n_w, n_x, n_h, and n_o are the vector length, the size of the input feature, hidden unit number in the hidden layer, and the hidden unit (neuron) number in the output layer, respectively.

Authors have used the mean squared error (MSE) function, which is computed as the mean of the squared differences between predicted and actual values. A mathematical perfect value is zero.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{13}$$

where y and $\hat{y}$ denote the actual value and the predicted value, and n is total number of values. The process of weight optimization is visualized in Fig. 2.

Fig. 2 MLP training process by BAEABC



The process of training an ANN using hybridized bat algorithm can be described with following pseudocode:

This experiment has utilized two well-known datasets. The first dataset, Saheart is dataset composed of 462 male samples of heart disease in South Africa. Around two controls per hearth disease case is made. The second dataset that is used is the Vertebral dataset. Dr. Henrique da Mota gathered data for the set, and it contains in total 310 patients. The first dataset has nine and the second dataset six features.

Algorithm 2 Pseudo-code of the ANN training process with hybridized Bat algorithm

```
Define population size n
Initialize the wights and biases x_i (i = 1, 2, ..., n)
Define the fitness function f(x)
Define maximum number of iterations MaxIter
Define counter t = 0
while t < MaxIter do
   for each solution do
      Set weights and biases of an ANN with solution's decision variables
      Calculate fitness for current solution
      Optimize decision variables using Hybridized Bat algorithm
   end for
   Sort population by fitness function
   Save the current best solution
end while
Output the solution with minimal fitness function value from all iterations and test it on test dataset
```

The datasets are split into train and test sets. The train test covers two-thirds of the total samples, and the test set covers one-third of the total number of samples. The features are normalized utilizing the following formula:

$$X_{\text{norm}} = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \tag{14}$$

X_{norm} represents the normalized value, and the ith input feature can be labeled as X_i, while $X_{\min}$ and $X_{\max}$ represent the minimum and the maximum value of that particular feature, respectively.

Each element of the vector is initialized between -1 and 1. The number of population members is set to 50, and the whole process is iterated for 250 iterations. 30 independent runs were made to obtain unbiased results.

The control parameters of BAEABC are outlined in Table 1.

Table 1 Control parameters of BAEABC

Parameter	Notation	Value
Size of the population	N	50
Maximum number of iteration	MaxIter	250
Minimum frequency	$f_{\min}$	0
Maximum frequency	$f_{\max}$	2
Loudness	A	0.9
Pulse rate	r	0.5
Loudness adaption parameter	α	0.9
Pulse rate adaption parameter	γ	0.9

For performance measurement, five different metrics are used:

1. Area under the curve (AUC): $\text{AUC} = \frac{1}{(\text{TP} + \text{FP})(\text{TN} + \text{FN})} \int_0^1 \text{TP d FP}$
2. Accuracy: $\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$
3. Specificity: $\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$
4. Sensitivity: $\text{sensitivity} = \frac{\text{TN}}{\text{FN} + \text{TP}}$
5. Geometric mean (g-mean): $\text{g-mean} = \sqrt{\text{specificity} \times \text{sensitivity}}$

In the metrics equations, TN represents the true negative and TP the true positive value, while the false negative and false positive are denoted by FN and FP.

The collected empirical results are matched to other metaheuristic methods (GOA, PSO, GA, BAT, ABC, FF, MBO, BBO, and FPA), and the results of these algorithms are taken from [25]. To make the right comparison, the simulation configuration is done similarly as it is described in the [25]. The results of Saheart dataset are presented in Table 2, and the results of Vertebral dataset are presented in Table 3.

Figure 3 presents the comparison of the proposed BAEABC and other metaheuristic methods best results on five metrics.

The accuracy convergence graph is depicted in Fig. 4, which shows the accuracy over the course of iterations. The best accuracy on the Saheart dataset and Vertebral dataset is achieved after 100 iterations.

Taking into account, the simulation outcomes the suggested BAEABC method achieved high performance and accuracy compared to other metaheuristic approaches. The BAEABC performed with the best statistical result on all five metrics on the test with the Saheart dataset, while in the case of the Vertebral dataset test, the algorithm resulted in the highest AUC and accuracy. In the event of worst and average statistical results on the Saheart dataset evaluation, the proposed method has the highest results on the accuracy and AUC metrics. The Vertebral dataset evaluation resulted in the highest values on all metrics in case of the worst and average statistical results. The proposed method has 81.312% accuracy on the Saheart dataset 95.658% accuracy on the Vertebral dataset.

5 Conclusion

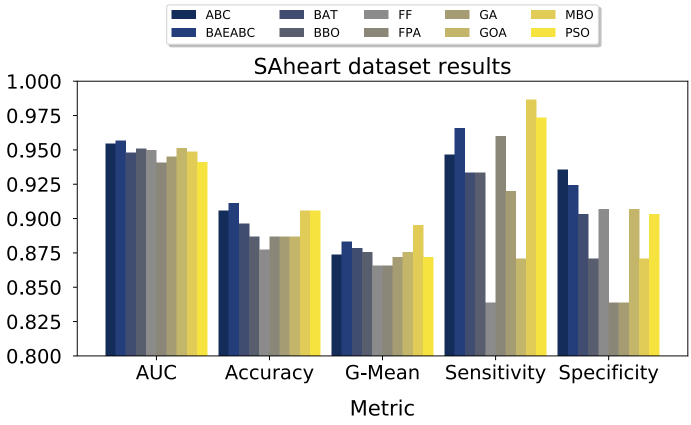
In this work, the authors wanted to show that a combination such as a hybridized algorithm together with a neural network is very competitive compared to other optimization systems. The comparative analysis suggests that such a system provides

Table 2 MLP training results on Sahart dataset

Algorithm	Metric	AUC	Accuracy	Specificity	Sensitivity	G-Mean
BAEABC	Best	0.81312	0.79125	0.95198	0.92645	0.72217
	StdDev	0.00385	0.07168	0.02798	0.01685	0.03179
	Mean	0.78112	0.73853	0.83712	0.85254	0.65425
	Worst	0.76128	0.79382	0.77982	0.78152	0.78066
GOA	Best	0.78793	0.79114	0.57407	0.91346	0.71238
	StdDev	0.01379	0.02378	0.03809	0.02645	0.02879
	Mean	0.75555	0.73122	0.49383	0.85449	0.64913
	Worst	0.72685	0.67722	0.42593	0.79808	0.58653
PSO	Best	0.79897	0.77848	0.95192	0.61111	0.68990
	StdDev	0.02382	0.02768	0.05845	0.06018	0.02874
	Mean	0.76022	0.72658	0.85096	0.48704	0.64126
	Worst	0.71546	0.64557	0.70192	0.38889	0.57689
GA	Best	0.78241	0.75949	0.94231	0.55556	0.67792
	StdDev	0.01242	0.01757	0.03645	0.03587	0.01702
	Mean	0.76671	0.71814	0.82372	0.51481	0.65030
	Worst	0.73326	0.68354	0.75962	0.40741	0.61048
BAT	Best	0.78846	0.75949	0.89423	0.59259	0.67779
	StdDev	0.01025	0.01840	0.03292	0.03248	0.01780
	Mean	0.76642	0.72405	0.83654	0.50741	0.65086
	Worst	0.74252	0.68987	0.76923	0.42593	0.61715
ABC	Best	0.81250	0.76582	0.95192	0.61111	0.71909
	StdDev	0.03388	0.02120	0.05373	0.07355	0.03546
	Mean	0.74454	0.71160	0.82276	0.49753	0.63644
	Worst	0.66560	0.67722	0.72115	0.29630	0.53109
FF	Best	0.77902	0.74051	0.55556	0.84615	0.68172
	StdDev	0.00402	0.01259	0.01709	0.01650	0.01288
	Mean	0.77276	0.71730	0.51667	0.82147	0.65137
	Worst	0.76086	0.69620	0.48148	0.79808	0.62361
MBO	Best	0.79790	0.76582	0.89423	0.62963	0.70408
	StdDev	0.02234	0.02126	0.04101	0.05246	0.02611
	Mean	0.76113	0.72932	0.83782	0.52037	0.65881
	Worst	0.71599	0.68354	0.75000	0.40741	0.59706
BBO	Best	0.78775	0.75316	0.91346	0.61111	0.69696
	StdDev	0.00905	0.01395	0.02976	0.04508	0.02235
	Mean	0.77369	0.72911	0.83910	0.51728	0.65776
	Worst	0.75036	0.69620	0.78846	0.40741	0.60359
FPA	Best	0.80520	0.76582	0.93269	0.64815	0.71050
	StdDev	0.02171	0.02307	0.04802	0.05728	0.02813
	Mean	0.76480	0.72869	0.84231	0.50988	0.65336
	Worst	0.72489	0.68354	0.77885	0.38889	0.59377

Table 3 MLP training results on Vertebral dataset

Algorithms	Metric	AUC	Accuracy	Specificity	Sensitivity	G-Mean
BAEABC	Best	0.95658	0.91114	0.92428	0.96587	0.88312
	StdDev	0.00485	0.00822	0.02896	0.03585	0.05897
	Mean	0.94138	0.87195	0.88517	0.90534	0.85638
	Worst	0.93385	0.85385	0.87895	0.88568	0.84125
GOA	Best	0.95140	0.88679	0.90667	0.87097	0.87547
	StdDev	0.00501	0.00885	0.00948	0.02553	0.01307
	Mean	0.94060	0.86321	0.88444	0.81183	0.84723
	Worst	0.93204	0.84906	0.86667	0.77419	0.82540
PSO	Best	0.94108	0.90566	0.90323	0.97333	0.87203
	StdDev	0.01703	0.03349	0.07023	0.05628	0.03254
	Mean	0.91911	0.85031	0.78925	0.87556	0.82942
	Worst	0.87570	0.77358	0.61290	0.76000	0.73441
GA	Best	0.94495	0.88679	0.83871	0.92000	0.87203
	StdDev	0.00465	0.01049	0.02198	0.01706	0.01026
	Mean	0.93508	0.85786	0.82151	0.87289	0.84664
	Worst	0.92602	0.83962	0.77419	0.84000	0.82540
BAT	Best	0.94796	0.89623	0.90323	0.93333	0.87841
	StdDev	0.04299	0.05471	0.05272	0.08269	0.04071
	Mean	0.91209	0.83553	0.81075	0.84578	0.82617
	Worst	0.77247	0.68868	0.67742	0.62667	0.68533
ABC	Best	0.95441	0.90566	0.93548	0.94667	0.87375
	StdDev	0.01995	0.03265	0.05522	0.05216	0.02786
	Mean	0.92189	0.84371	0.81828	0.85422	0.83483
	Worst	0.87484	0.77358	0.67742	0.74667	0.76031
FF	Best	0.94968	0.87736	0.90667	0.83871	0.86559
	StdDev	0.00434	0.00810	0.01259	0.01794	0.00860
	Mean	0.94072	0.86384	0.87911	0.82688	0.85250
	Worst	0.92946	0.84906	0.85333	0.77419	0.82540
MBO	Best	0.94882	0.90566	0.87097	0.98667	0.89515
	StdDev	0.01766	0.02796	0.06460	0.05083	0.02750
	Mean	0.92346	0.85660	0.78387	0.88667	0.83203
	Worst	0.87957	0.76415	0.64516	0.73333	0.77598
BBO	Best	0.95097	0.88679	0.87097	0.93333	0.87547
	StdDev	0.00660	0.01106	0.02896	0.01836	0.01285
	Mean	0.93875	0.86730	0.82473	0.88489	0.85403
	Worst	0.92645	0.84906	0.74194	0.84000	0.83163
FPA	Best	0.94065	0.88679	0.83871	0.96000	0.86559
	StdDev	0.00836	0.01506	0.03425	0.02351	0.01676
	Mean	0.92433	0.86730	0.77742	0.90444	0.83816
	Worst	0.90538	0.83962	0.70968	0.86667	0.80215



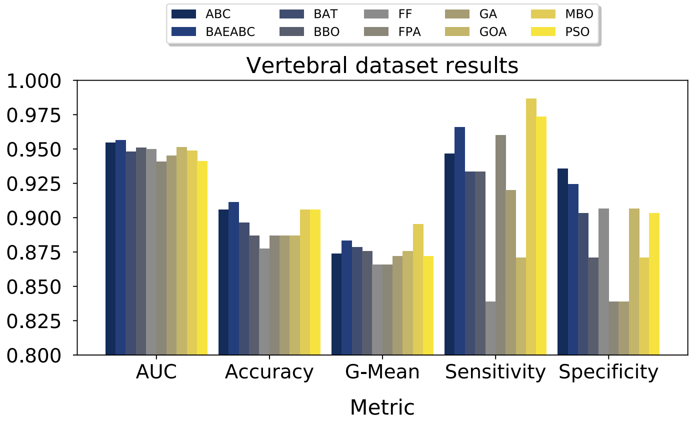


Fig. 3 Algorithm comparison of best result on five metrics

optimal results, better than the other proposed solutions. The main goal of this algorithm was to improve and speed up the process of training an ANN using calculated values of weights and biases instead of letting the algorithm itself to produce values of weights and biases using backpropagation, which is a much slower process as already mentioned in the paper. It is confirmed that swarm intelligence algorithms (metaheuristics) indeed can speed up the learning process of artificial neural networks, with an equal amount of success as in the other optimization problems.

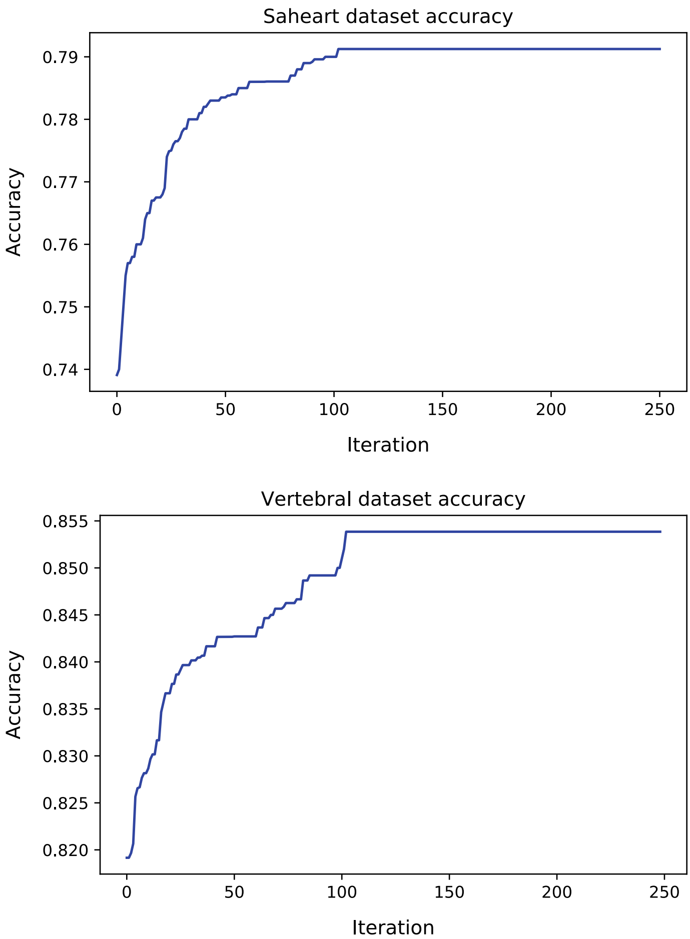


Fig. 4 Convergence graph

References

1. W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. **5**(4), 115–133 (1943)
2. M.S. Shanker, Using neural networks to predict the onset of diabetes mellitus. J. Chem. Inform. Computer Sci. **36**(1), 35–41 (1996)
3. J.J. Palop, L. Mucke, Amyloid-β-induced neuronal dysfunction in alzheimer's disease: from synapses toward neural networks. Nature Neurosci. **13**(7), 812–818 (2010)
4. O. Er, F. Temurtas, A.Ç. Tanrıkulu, Tuberculosis disease diagnosis using artificial neural networks. J. Med. Syst. **34**(3), 299–302 (2010)
5. Y. Lu, S. Yi, N. Zeng, Y. Liu, Y. Zhang, Identification of rice diseases using deep convolutional neural networks. Neurocomputing **267**, 378–384 (2017)
6. B. Liu, Y. Zhang, D. He, Y. Li, Identification of apple leaf diseases based on deep convolutional neural networks. Symmetry **10**(1), 11 (2018)

7. J. Orbach, Principles of neurodynamics. Perceptrons and the theory of brain mechanisms. Arch. General Psychiatry **7**(3), 218–219 (1962)
8. Y. Freund, R.E. Schapire, Large margin classification using the perceptron algorithm. Mach. Learn. **37**(3), 277–296 (1999)
9. R. Hecht-Nielsen, *Theory of the backpropagation neural network, in Neural Networks for Perception* (Elsevier, Amsterdam, 1992)
10. F.-C. Chen, Back-propagation neural networks for nonlinear self-tuning adaptive control. IEEE Control Syst. Mag. **10**(3), 44–48 (1990)
11. M. Dorigo, M. Birattari, T. Stutzle, Ant colony optimization. IEEE Comput. Intell. Mag. **1**(4), 28–39 (2006)
12. X.-S. Yang, *A new metaheuristic bat-inspired algorithm, in Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)* (Springer, Berlin, 2010), pp. 65–74
13. D. Karaboga, B. Basturk, *Artificial bee colony (abc) optimization algorithm for solving constrained optimization problems, in International Fuzzy Systems Association World Congress* (Springer, Berlin, 2007), pp. 789–798
14. A.A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, H. Chen, Harris hawks optimization: algorithm and applications. Future Gener. Comput. Syst. **97**, 849–872 (2019)
15. T. Bezdan, M. Zivkovic, E. Tuba, I. Strumberger, N. Bacanin, M. Tuba, Multi-objective task scheduling in cloud computing environment by hybridized bat algorithm, in *International Conference on Intelligent and Fuzzy Systems* (Springer, 2020), pp. 718–725
16. T. Bezdan, M. Zivkovic, M. Antonijevic, T. Zivkovic, N. Bacanin, Enhanced flower pollination algorithm for task scheduling in cloud computing environment, in *Machine Learning for Predictive Analysis*, ed. by A. Joshi, M. Khosravy, N. Gupta (Springer, Singapore, 2021), pp. 163–171
17. N. Bacanin, T. Bezdan, E. Tuba, I. Strumberger, M. Tuba, M. Zivkovic, Task scheduling in cloud computing environment by grey wolf optimizer, in *2019 27th Telecommunications Forum (TELFOR)* (IEEE, 2019), pp. 1–4
18. M. Zivkovic, N. Bacanin, E. Tuba, I. Strumberger, T. Bezdan, M. Tuba, Wireless sensor networks life time optimization based on the improved firefly algorithm, in *2020 International Wireless Communications and Mobile Computing (IWCMC)* (IEEE, 2020), pp. 1176–1181
19. N. Bacanin, E. Tuba, M. Zivkovic, I. Strumberger, M. Tuba, *Whale optimization algorithm with exploratory move for wireless sensor networks localization, in International Conference on Hybrid Intelligent Systems* (Springer, Berlin, 2019), pp. 328–338
20. M. Zivkovic, N. Bacanin, T. Zivkovic, I. Strumberger, E. Tuba, M. Tuba, Enhanced grey wolf algorithm for energy efficient wireless sensor networks, in *2020 Zooming Innovation in Consumer Technologies Conference (ZINC)* (IEEE, 2020), pp. 87–92
21. T. Bezdan, M. Zivkovic, E. Tuba, I. Strumberger, N. Bacanin, M. Tuba, *Glioma brain tumor grade classification from MRI using convolutional neural networks designed by modified FA", in International Conference on Intelligent and Fuzzy Systems* (Springer, Berlin, 2020), pp. 955–963
22. E.T.I.S. Nebojsa Bacanin, T. Bezdan, M. Tuba, Optimizing convolutional neural network hyperparameters by enhanced swarm intelligence metaheuristics. Algorithms **13**(3), 67 (2020)
23. N. Bacanin, T. Bezdan, E. Tuba, I. Strumberger, M. Tuba, Monarch butterfly optimization based convolutional neural network design. Mathematics **8**(6), 936 (2020)
24. T. Bezdan, E. Tuba, I. Strumberger, N. Bacanin, M. Tuba, Automatically designing convolutional neural network architecture with artificial flora algorithm, in *ICT Systems and Sustainability*, ed. by M. Tuba, S. Akashe, A. Joshi (Springer, Singapore, 2020), pp. 371–378
25. A.A. Heidari, H. Faris, I. Aljarah, S. Mirjalili, An efficient hybrid multilayer perceptron neural network with grasshopper optimization. Soft Comput. **23**(17), 7941–7958 (2019)
26. J.C. Duchi, E. Hazan, Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization. J. Mach. Learn. Res. **12**, 2121–2159 (2011)
27. M. D. Zeiler, Adadelta: an adaptive learning rate method (2012)
28. D.P. Kingma, J. Ba, Adam: a method for stochastic optimization (2014)

29. D. Karaboga, B. Akay, A modified artificial bee colony (ABC) algorithm for constrained optimization problems. Appl. Soft Comput. **11**(3), 3021–3031 (2011)
30. M. Tuba, N. Bacanin, Hybridized bat algorithm for multi-objective radio frequency identification (RFID) network planning, in *2015 IEEE Congress on Evolutionary Computation (CEC)* (2015), pp. 499–506

Bayes Wavelet-CNN for Classifying COVID-19 in Chest X-ray Images

S. Kavitha and Hannah Inbarani

Abstract In digital image processing, removing noise from the images is essential for better classification, especially for medical images. A novel hybrid approach of Bayes wavelet transform and convolutional neural network (BayesWavT-CNN) is used for classifying the chest X-ray images into ordinary and COVID-19 images. Bayes wavelet transform denoising method is used to denoise the images before classification. A simple eight-layer CNN is developed for the classification of denoised images. The proposed model achieved the highest test accuracy of 97.10% for the 20 epochs, and it is compared with the SVM and CNN developed on the same images without denoising. The results indicate that the developed hybrid model provides excellent performance in accuracy and receiver operator characteristic (ROC) curve analysis.

Keywords Denoising · COVID-19 image data · Bayes wavelet transform · Simple CNN · Support vector machine (SVM) · Receiver operator characteristic (ROC) curve

1 Introduction

The coronavirus infection is spreading very rapidly all over the world through droplets in the air. As of November 9, 2020, there have been more than 50,000,000 confirmed cases of coronavirus disease is reported globally to the World Health Organization. The increasing number of COVID-19 patient data creates a challenge in the research of medical image processing to detect the disease at the earliest. So far, several research works have been done to expose the applications of artificial intelligence techniques in the detection of COVID-19 cases [1, 2]. Recently, CNN is used in many of the research works to detect the coronavirus infections present in the clinical images [3, 4]. In this work, a deep learning model CNN is proposed which is

S. Kavitha (✉) · H. Inbarani
Department of Computer Science, Periyar University, Salem, Tamil Nadu, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_55

integrated with wavelet transform to detect the infections in the chest X-ray images with less computational power and speed.

Usually, digital images such as medical images are affected by the noise due to the environment and transmission media, etc. [5]. In digital image processing, it is required to pre-process the images to remove the noise present in the digital images. It is an important step to be taken before images are analyzed for features extraction and classification [6]. To obtain high-quality images in the process of image, denoising from the noisy images is a classical problem as it is difficult to distinguish noise, edge, and texture in the noisy images [7]. Many researchers have proposed different denoising methods such as smoothing filters, frequency domain denoising, and wavelet transform methods to remove the noise from different types of image datasets [8–11]. This article presents the Bayes wavelet transform to remove the noise in the chest X-ray images since its performance is desirable based on the peak signal to noise ratio (PSNR), mean squared error (MSE) metrics, and visual image quality.

Deep learning in artificial intelligence plays an important role in processing images in a better way. Though many algorithms exist for classification such as logistic regression, decision tree, and support vector machines, etc. CNN is a well-known deep learning network used by many researchers for image classification and recognition problems. By applying relevant filters, CNN can capture the spatial and temporal dependencies in an image [12]. CNN is a suitable deep learning technique to deal with complex dataset compared to traditional neural network [13]. The objective of this paper is to present the novel hybrid approach of combining the Bayes wavelet transform denoising method with the simple CNN to classify the COVID-19 X-ray image dataset. The remaining part of the paper is structured as follows: Sect. 2 describes the literature study, and the next Sect. 3 is followed by the motivation of research. Section 4 explains the dataset used for analysis, and the proposed methodology is presented in Sect. 5. Section 6 discusses the findings, and finally, Sect. 7 concludes the research work.

2 Related Work

There are various studies in pre-processing as reported in Table 1 focusing on different denoising techniques which are presented in related work along with the research works done on the classification of diseases present in the medical images using machine learning and deep learning algorithms.

Table 1 Related works for this study

Authors	Techniques	Result
Vikas Gupta et al. [14]	BayesInvariant wavelet, VisuShrink and Sureshrink methods to denoise the images	BayesInvariant wavelet-based method performance is best compared with the VisuShrink and Sureshrink methods
Gurmeet Kaur et al. [15]	Wavelet transform and filtering techniques	The wavelet transform method is good compared with the filtering methods
Jyotsna Patil et al. [16]	Wavelet transform and spatial filtering	It is concluded that the wavelet transform is best suitable than spatial filters as these filters causes over smoothing and blurring of image
Suvajit Dutta et al. [23]	Feed-forward neural network, Deep neural network, CNN	DNN performs better than other networks
Nitish Srivastava et al. [24]	Neural network with dropout function	Dropout overcomes the problem of over fitting
Dominik Scherer et al. [25]	Max pooling, sub-sampling	It is concluded that max-pooling operation is better than sub-sampling operation for capturing invariance in images
Sathish et al. [26]	Fuzzy C-means clustering and CNN	Achieved 87.11% accuracy for classifying tumors in brain MRI images
El Boustani et al. [27]	Probabilistic neural network and CNN along with root mean square propagation (RMSprop) optimizer	It is concluded that the RMSprop optimizer performs well with the highest accuracy
Abbas et al. [3]	CNN with decomposition and transfer learning to classify COVID-19 in chest X-ray images	An accuracy of 93.1% is achieved
Wang et al. [4]	CNN with transfer learning to detect COVID-19 infections on CT scan images	Achieved a total accuracy of 89.5%
Liu et al. [28]	The ensemble of the bagged tree algorithm with statistical textural features to detect COVID-19	The classification accuracy is 94.16% attained
Perumal et al. [29]	Feature extraction using Haralick features with transfer learning to detect COVID-19 like pneumonia in CT images	An accuracy of 93% is achieved

3 Research Motivation

Since the COVID-19 infections are rapidly increasing every day, it is a challenge for the researchers to find a better classification approach to classify the COVID-19 image dataset. Due to the noise present in the digital images, it is required to denoise the images as a pre-processing step before using it for classification. In this work, the Bayesian approach [6] of wavelet transform denoising is used because it gives better quality images by removing the irrelevant data for feature extraction and classification. Hence, in this methodology, the Bayes wavelet model is integrated with CNN with less number of layers to classify the images in a better way with less computational power and speed.

4 COVID-19 Chest X-Ray Dataset

The data set has nearly 350 Chest X-ray scans which are categorized into two classes as COVID and normal. Among the 350 images, 234 images are normal X-ray images, and 110 images are COVID-19 X-ray images. The images are grayscale, and the dimensions of the images are 256×256. The image dataset is split into the ratio of 80 and 20% for the training and testing process, respectively.

5 Proposed Method

The novel hybrid approach of combining the Bayes wavelet transform denoising method with simple CNN is proposed in this work for the better classification of images.

Figure 1 shows the proposed hybrid model which consists of three phases:

(i) In the first phase, Gaussian noise is added in the images with the sigma value as 0.15 because it is evaluated that the wavelet-based denoising method provides high-quality images for Gaussian noise than the speckle noise.

(ii) In the second phase, a Bayes wavelet transform method is adopted to eliminate the noise added in the images. In this method, the noisy image signal is first decomposed using the discrete wavelet transform (DWT) into the various wavelet coefficients. These coefficients are threshold by using the BayesShrink thresholding function, and then, the denoised image is achieved by taking the inverse transform of the threshold coefficients.

(iii) In the third phase, a simple CNN architecture is developed with three convolutional layers, two max-pooling layers, two dense layers, and dropout function. Finally, the denoised images are classified using the CNN into normal and COVID-19 images, and the results are compared with the SVM [30] and CNN developed without applying Bayes wavelet transform.

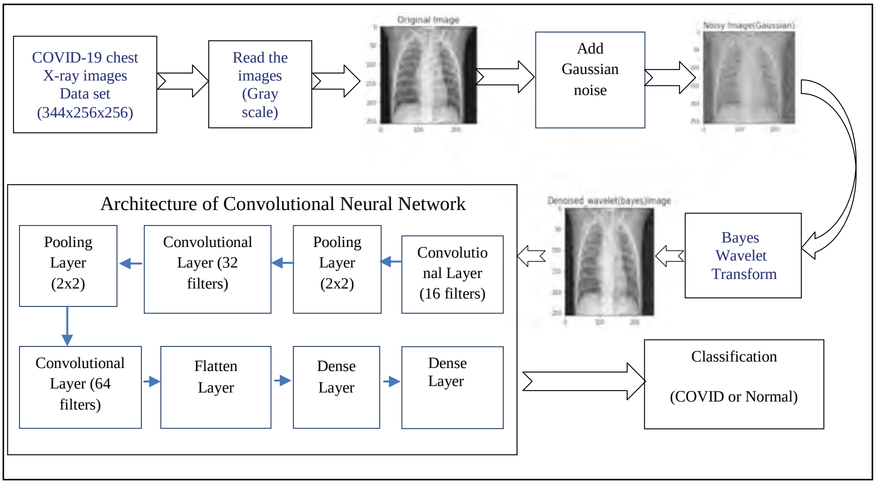


Fig. 1 Methodology of proposed method

5.1 *Adding Gaussian Noise to the Images*

During the first phase, the input images from the dataset are added with Gaussian noise. It is an additive noise which is the sum of random Gaussian distributed noise and true pixel values [17]. It is represented using the probability density function p as:

$$p(g) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(g-\mu)^2}{2\sigma^2}} \tag{1}$$

where g is the Gaussian random variable (graylevel), µ is a mean of average value of g, σ is a standard deviation, and and σ^2 is the variance of g in which the Gaussian distribution is characterized by the mean μ and the variance σ^2.

5.2 *Denoising the Images Using Bayes Wavelet Transform*

BayesShrink [18] soft thresholding data-driven method is applied with three-level wavelet decomposition which sets the different thresholds for each sub-band to denoise the chest X-ray images added with Gaussian noise. This method estimates a threshold value to minimize the Bayesian risk by assuming the Gaussian distribution for the wavelet coefficients in each detail sub-band [6]. The threshold value is given by the following equation:

$$T_b = \left\{ \begin{array}{cc} \frac{\sigma_v^2}{\sigma_x^2} & \text{if } \sigma_v^2 < \sigma_x^2 \\ \text{otherwise} & \max\{|W_j|\} \end{array} \right\} \tag{2}$$

where W_j is the wavelet coefficients in each scale j, ${\sigma_v}^2$ is the noise variance which is estimated from the sub-band by the median estimator, and ${\sigma_x}^2$ is the original image variance. The variance of the degraded image for each sub-band can be calculated as:

$$\sigma_y^2 = \frac{1}{J} \sum_{j=1}^{J} W_j^2 \tag{3}$$

where J is the total number of coefficients in the sub-band, and σ_x is calculated by the following equation:

$$\sigma_x = \sqrt{\max\left(\sigma_y^2 - \sigma_v^2, 0\right)} \tag{4}$$

5.3 Proposed Simple Convolutional Neural Network

A simple CNN is adopted in this model with three convolutional layers, two pooling layers, and two dense layers for classifying the denoised images. The convolutional layer has several filters that perform convolution operation on the input image to extract features. Then apply the activation function on the output to get a feature map [19]. The proposed CNN has 16, 32 and 64 filters in the first, second, and third convolutional layers, respectively. Each convolutional layer has a 3×3 kernel size. The proposed model uses the Relu activation function [20] in each convolutional layer. It is defined by the following equation:

$$g(x) = \max(0, x) = \left\{ \begin{array}{ll} x \text{ if } & x > 0 \\ 0 \text{ if } & x \leqslant 0 \end{array} \right\} \tag{5}$$

The above equation states that it has output zero if the input is less than or equal to zero, and it outputs a linear function if x is greater than zero.

Two max-pooling layers [21] are used in this model with 2×2 filters to down-sample the images without losing any information. Max pooling decreases the dimensions of activation maps and increases the strength of feature extraction by selecting the highest value from the group of neurons at the previous layer. A flattening layer is placed between the convolution layer and the dense layer. It transforms the two-dimensional matrix of a vector which is fed into the dense layer for classification. This model adopted two dense layers for the classification. The dense layer connects every neuron of the previous layer with every neuron of the current layer. The last fully connected layer is an output layer which gives the output from the number of classes. The Softmax regression [22] is used for the classification task as it generates a well-performed probability distribution of the outputs.

6 Result Analysis

6.1 *Findings of Bayes Wavelet Transform*

In the first and second phases of the model, the images are pre-processed for noise reduction. Figure 2 shows the sample images of the chest X-ray image dataset, Gaussian noisy image, and the denoised image using Bayes wavelet transform.

The above figure shows that the quality of the Bayes wavelet transform method is visually good which is also evaluated using PSNR and MSE, and the results are given in Table 2.

Table 2 shows the average PSNR and MSE values of Bayes wavelet transform method applied to the noisy images. Wavelet transforms denoised images to provide the higher PSNR and lowest MSE value compared with the PSNR and MSE values of original and noisy images. If the value of PSNR is high after denoising, then the quality of the reconstructed image from the noisy image is high which is good for better classification.

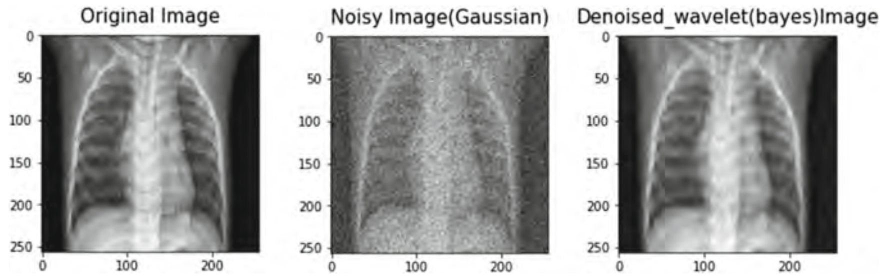


Fig. 2 Sample image of Gaussian noise and wavelet transform

Table 2 Average PSNR and MSE values of Bayes wavelet transform

Method	PSNR/dB	MSE
Before denoising (between original and noisy images)	16.47	0.0225
Wavelet transform-BayesShrink method (between original and denoised images)	29.51	0.0011

6.2 *Performance Measurements of Bayes Wavelet-CNN*

In the third phase of the model, the denoised images are classified using simple CNN, and the results are compared with the SVM and CNN developed without denoising the images. The performance of the proposed approach is analyzed through accuracy, classification report, and ROC curve. Figure 3 illustrates the graph which shows the accuracy results of simple CNN and BayesWavT-CNN.

As shown in Fig. 3, the proposed BayesWavT-CNN attained the highest test accuracy of 97.10% for the 20 epochs which represents the rate of correct classification. But the CNN developed without wavelet transform achieves the 91.30% as test accuracy for the same number of epochs. Table 3 illustrates the accuracy results of SVM, CNN before denoising, and CNN after denoising (BayesWavT-CNN).

As illustrated in Table 3, the proposed BayesWavT-CNN achieves the highest training and testing accuracy compared with the SVM and CNN developed without denoising of the images.

The metric is a classification report which is given in Table 4. It is used to measure the predictions of the classification algorithm using precision, recall, and F1-score which are calculated using the terms true positive and false positive, true negative and false negatives. For a class, the precision is computed as the number of true positives divided by the total number of items labeled as positive, whereas the recall measures the fraction of positives that are correctly identified. F1 score is the weighted average of precision and recall.

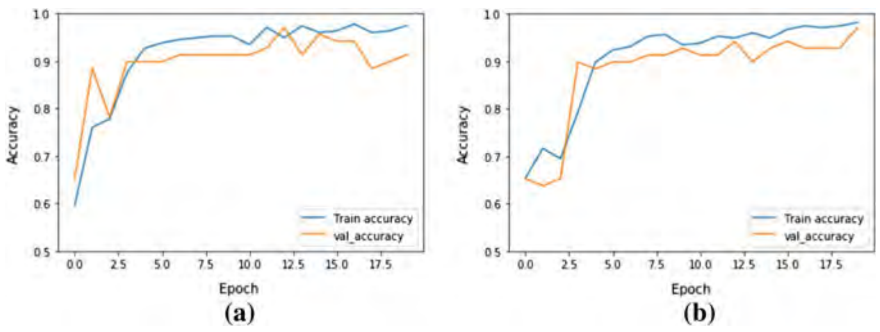


Fig. 3 **a** Accuracy results of simple CNN and **b** accuracy results from BayesWavT-CNN

Table 3 COVID-19 accuracy results (training and testing accuracy) of SVM, simple CNN and BayesWavT-CNN

Method	Train accuracy (%)	Test accuracy (%)
Support vector machine (SVM)	94.90	89.85
Simple CNN	97.45	91.30
BayesWavT-CNN (Proposed method)	98.18	97.10

Table 4 Classification report of BayesWavT-CNN, CNN, and SVM

Classification report	SVM	Simple CNN	BayesWavT-CNN (proposed method)
Class 0 (COVID)			
Precision	0.95	1.00	1.00
Recall	0.75	0.75	0.92
F1-score	0.84	0.86	0.96
Class 1 (Normal)			
Precision	0.88	0.88	0.96
Recall	0.98	1.00	1.00
F1-score	0.93	0.94	0.98

As given in Table 4, the precision value of 1.00 for a class 0 (COVID) denotes that every image labeled as class 0 belongs to class 0. The recall value of 1.00 means that every image from class 1 is labeled as belongs to class 1. Hence, the quality of the predictions made by BayesWavT-CNN is good based on precision, recall, and F1-score than the SVM and CNN.

Another parameter, ROC curve evaluation is given in Fig. 4. It shows the false positive and true positive rates of CNN, BayesWavT-CNN, and SVM. From the ROC curve analysis, the proposed method BayesWavT-CNN classifies the images well compared with SVM and CNN.

Figure 5 shows the comparisons of the proposed method with SVM and CNN. From the figure, it is shown that the performance of Bayes Wavelet-CNN is good than SVM and CNN.

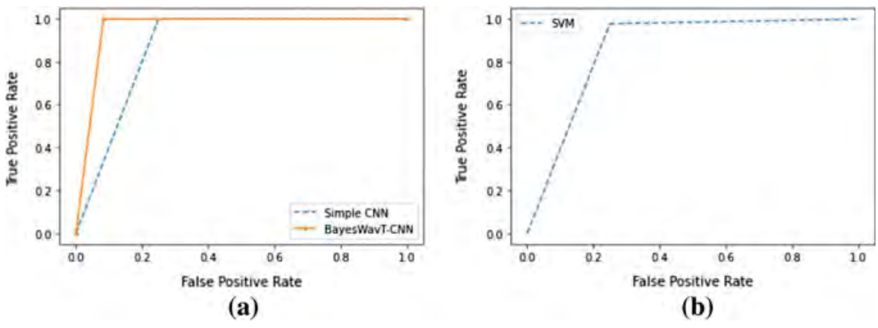


Figure 4 **a** ROC curve of BayesWavT-CNN and CNN and **b** ROC curve of SVM

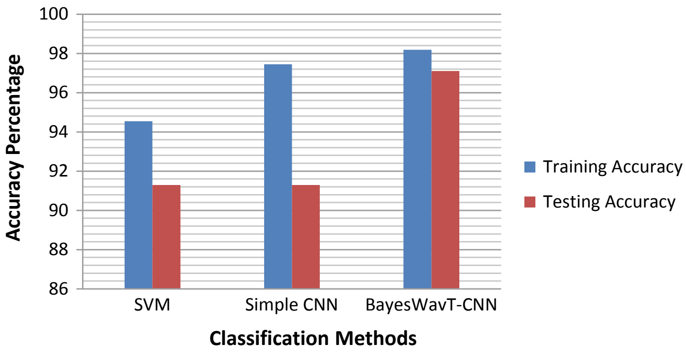


Fig. 5 Performance comparisons of SVM, CNN, and BayesWavT-CNN

7 Conclusion

In this article, the novel classification approach based on the Bayes wavelet method and simple CNN (BayesWavT-CNN) is proposed for classifying the chest X-ray images into ordinary and COVID-19 images. Initially, the images are pre-processed for noise reduction using Bayes wavelet transform. Next, the simple CNN is developed on the denoised images to create the classification model. Finally, the proposed BayesWavT-CNN is compared with the SVM and simple CNN for the performance analysis. From the experimental outcomes, it is shown that the proposed model performs well than the SVM and CNN. In the future, the model can be improved to achieve the highest accuracy.

References

1. L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. Radiology 200905 (2020)
2. F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, D. Shen, Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. IEEE Rev. Biomed. Eng. (2020)
3. A. Abbas, M.M. Abdelsamea, Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. Appl. Intell. (2020)
4. S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, A deep learning algorithm using CT images to screen for corona virus disease (COVID-19), in *MedRxiv* (2020)
5. J.M. Sanches, J.C. Nascimento, J.S. Marques, Medical image noise reduction using the SylvesterLyapunov equation. IEEE Trans. Image Process. **17**(9), 1522–1539 (2018)
6. P.B. Alisha, K. Gnana Sheela, Image denoising techniques—an overview. IOSR J. Electron. Commun. Eng. (IOSR-JECE) **11**(1) (2016). e-ISSN: 2278-2834, ISSN: 2278-8735
7. L. Fan, F. Zhang, H. Fan, C. Zhang, Brief review of image denoising techniques. Visual Comput. Ind. Biomed. Art **2**, Article number: 7 (2019)

8. S.G. Chang, B. Yu, M. Vetterli, Spatially adaptive wavelet thresholding with context modeling for image denoising. IEEE Trans. Image Process. **9**(9), 1522–1531 (2000)
9. A. Pizurica, W. Philips, I. Lemahieu, M. Acheroy, A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising. IEEE Trans. Image Process. **11**(5), 545–557 (2002)
10. L. Gondara, Medical image denoising using convolutional denoising autoencoders, in *IEEE 16th International Conference on Data Mining Workshops* (2016), pp. 241–246
11. L. Zhang, P. Bao, X. Wu, Multiscale LMMSE-based image denoising with optimal wavelet selection. IEEE Trans. Circ. Syst. Video Technol. **15**(4), 469–481 (2005)
12. S. Saha, A comprehensive guide to convolutional neural networks—the ELI5 way, towards data science (2015)
13. M. Xin, Y. Yong Wang, Research on image classification model based on deep convolution neural network. EURASIP J. Image Video Process. **40** (2019)
14. V. Gupta, R. Mahle, R.S. Shriwas, Image denoising using wavelet transform method, in *Tenth International Conference on Wireless and Optical Communications Networks (WOCN)* (2013), pp. 1–4
15. G. Kaur, R. Kaur, Image de-noising using wavelet transform and various filters. Int. J. Res. Comput. Sci. **2**(2), 15–21 (2012)
16. J. Patil, S. Jadhav, A comparative study of image denoising techniques. Int. J. Innov. Res. Sci. Eng. Technol. **2**(3) (2013)
17. R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, 2ns edn. (Pearson Education, 2005)
18. K. Tharani, C. Mani, I. Arora, A comparative study of image denoising methods using wavelet thresholding techniques. Int. J. Eng. Res. Appl. **6**(12) (2016). ISSN: 2248-9622
19. S. Albawi, T.A. Mohammed, S. Al-Zawi, Understanding of a convolutional neural network, in *International Conference on Engineering and Technology (ICET)*, Antalya (2017), pp. 1–6
20. V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines, in *International Conference on Machine Learning* (2010), pp. 807–814
21. Y. Boureau, J. Ponce, Y. Le Cun, A theoretical analysis of feature pooling in visual recognition, in *International Conference of Machine Learning* (2010), pp. 111–118
22. T. Guo, J. Dong, H. Li, Y. Gao, Simple convolutional neural network on image classification, in *IEEE 2nd International Conference on Big Data Analysis* (2017)
23. S. Dutta, B.C.S. Bonthala, S. Rai, V. Vijayarajan, A comparative study of deep learning models for medical image classification, in *IOP Conference Series: Materials Science and Engineering* 263 (2017)
24. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(56), 1929−1958 (2014)
25. D. Scherer, A. Muller, S. Behnke, Evaluation of pooling operations in convolutional architectures for object recognition, in *20th International Conference on Artificial Neural Networks (ICANN), Lecture Notes in Computer Science*, vol. 6354 (Springer, Berlin, 2010)
26. P. Sathish, N.M. Elango, V. Thirunavukkarasu, Piecewise fuzzy C-means clustering and deep convolutional neural network for automatic brain tumour classification using MRI images, Test Eng. Manage. **83**, 3729–3736 (2020)
27. A. El Boustani, M. Aatila, E. El Bachari, A. El Oirrak, MRI brain images classification using convolutional neural networks, in *Advanced Intelligent Systems for Sustainable Development (AI2SD'2019). AI2SD 2019. Advances in Intelligent Systems and Computing*, vol. 1105, ed. by M. Ezziyyani (Springer, Berlin, 2020)
28. C. Liu, X. Wang, C. Liu, Q. Sun, W. Peng, Differentiating novel coronavirus pneumonia from general pneumonia based on machine learning. BioMed. Eng. OnLine **19**, 66 (2020)
29. V. Perumal, V. Narayanan, S.J.S. Rajasekar, Detection of COVID-19 using CXR and CT images using transfer learning and Haralick features. Appl. Intell. (2020)
30. X. Sun, L. Liu, H. Wang, W. Song, J. Lu, Image classification via support vector machine, in *2015 4th International Conference on Computer Science and Network Technology (ICCSNT)*, Harbin (2015), pp. 485–489

Survey of Color Feature Extraction Schemes in Content-Based Picture Recovery System

Kiran H. Patil and M. Nirupama Bhat

Abstract The method of retrieving most visually similar pictures from a large database or group of picture files is called content-based picture or image recovery (CBIR) system. It is among one of the challenging research areas of multimedia computing and information recovery. In the past few decades, many different picture matching, indexing and recovery algorithms have been developed. In CBIR systems, recovery is based on matching visual content or characteristics of query picture with picture database using a picture—picture similarity calculation. The term "Content" in CBIR refers to the visual content of a picture that means texture, shape, color, etc., or any other feature/descriptor that can be acquired from the picture itself. In this paper, a survey of feature recovery techniques using color feature in the last 10 years is presented and summarized year wise.

Keywords Color moments · Color correlogram · Color histogram · Feature recovery

1 Introduction

Image processing is one of the significant research areas in computer science and engineering. Recently, image processing is impacting a lot during the COVID pandemic also. There is much importance achieved by image/picture/color detection, especially for the lungs images as the COVID disease spreads in the lungs. Image processing also has a number of applications such as scanners for product identification, identification of human beings using thumb impression or iris scan, etc.

Content-based picture/image recovery (CBIR) is the application of computer vision to picture recovery problems. Nowadays, there is a tremendous use and generation of digital pictures by the common user of electronic devices such as smartphones, Internet, medical devices, e-commerce, academia, etc. Therefore, searching

K. H. Patil (✉) · M. N. Bhat
CSE Department, VFSTR Deemed To Be University, Vadlamudi, Guntur, Andhra Pradesh, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_56

and recovery of desired pictures in a large database are very important for users from various fields like academia, marketing, hospitals, military tasks, crime prevention, geography, etc.

There are two main areas in CBIR systems such as computer vision and database systems. Computer vision involves image processing techniques such as obtaining the picture characteristics or descriptors and picture matching. Image processing and image transformations are used to extract picture characteristics or descriptors. There are certain steps involved like importing the image for analysis using computer tools and then extracting the necessary information. The database system includes database indexing, searching and recovery techniques.

Knowledge discovery in databases (KDD) process mainly focuses on the data cleaning, its integration and then relevant data selection tasks. Afterwards, it consolidates that data in aggregate form for a pattern evaluation, and then, based on this information, standard techniques are used to mine the data. In this article, the techniques based on content-based image retrieval are reviewed in which input image and image datasets are taken for extraction of features such as texture, color, shape, etc. Hence, the article focuses only on a set of images and it's processing, obtaining the best information from the set of images to retrieve similar images.

2 Working of CBIR

A typical CBIR system consists of following main modules feature recovery, feature storage, similarity/distance measure, indexing and recovery.

(1) **Feature Recovery**: Analyze picture database to extract feature specific information.
(2) **Feature Storage**: Provide sufficient storage for the extracted information and also help to improve the searching speed.
(3) **Similarity/Distance Measure**: It is the difference between the picture database and query picture for determining relevance between pictures.
(4) **Indexing and Recovery**: The indexing scheme gives an efficient searching method for a picture database.

The working of CBIR is shown in Fig. 1. The characteristics of the pictures in the picture database are identified and extracted, which can also be articulated as multi-dimensional feature vectors. The user supplies a query picture for the CBIR system. The characteristics of the query picture are also extracted and expressed as feature vectors. Using these feature vectors, the distance measure is calculated between the query picture and picture database. Finally, an indexing scheme [1] is used to retrieve the output pictures.

Performance Evaluation and Datasets Used by the Various Researchers

A number of datasets are available for the CBIR, it includes the Oxford building, INRIA holidays which are used for small scale datasets, and there is also the Kentucky

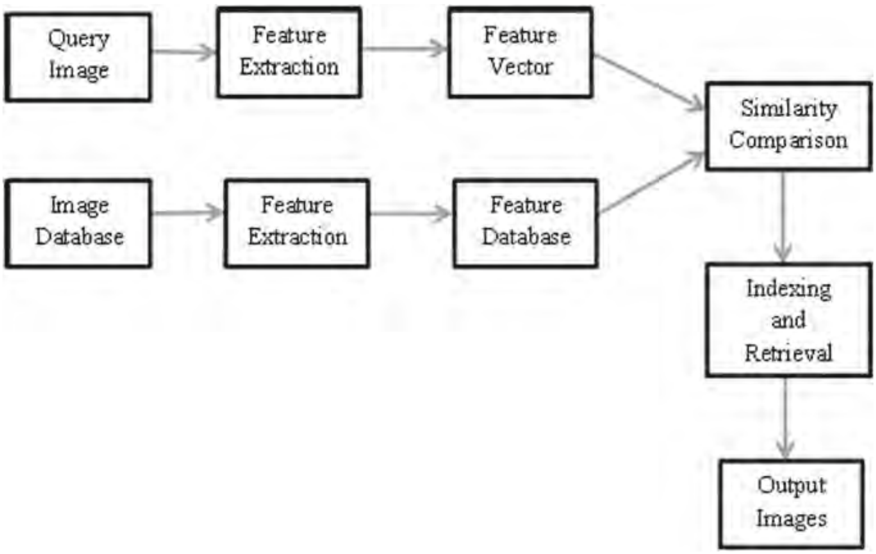


Fig. 1 Working of CBIR system

dataset which is easy to use. Tecvid dataset is used for the image and video retrievals and used for large-scale applications. In the literature review, details have been added regarding various datasets used by the corresponding researchers.

It is important to note that the performance evaluation of the images is carried out using precision and recall-based methods, which is the ratio of the number of images/figures retrieved to the total number of images taken in the data set. Further, one more important parameter is there called an error count or rate in which the image performance is evaluated using the number of correct and incorrect images retrieved to the number of total images supplied in the form of data set. Retrieval efficiency is another parameter of evaluation that indicates the ratio of the number of correctly retrieved images to the number of correct images present in the dataset. The graphical representation of the output is also used in the evaluation of the performance of various datasets.

3 Feature Recovery

Every picture has two types of content such as visual and semantic content. Further, visual content is of two types, general or domain-specific. The low-level characteristics like structure, size, color and texture, etc. are included under general visual content. The domain-specific visual content may contain face recognition, handwriting and fingerprints and is application dependent. It also involves domain knowledge. Based on visual content, the semantic content is achieved by textural notation or interference procedure [1]. Generally, low-level characteristics are more popular in describing pictures using picture signatures or descriptors. Low-level characteristics present various ways to describe any picture despite various picture classes and types [2]. Some of the descriptions of the term related to color feature recovery are given below:

(1) **Color**: For picture recovery, color is one of the most commonly used visual content. The picture comparison is based on matching pictures by their color distribution. Different techniques like color coherence vector, color histogram, color correlogram, color moments and invariant color characteristics are used in past decades. Color space must be obtained first to use any of these techniques.
(2) **Color Space**: Color space is a specific organization of colors. Each pixel of a picture can be represented as a point in 3D color space. Generally, the following color spaces are used for picture recovery such as RGB, HSV, CIE L*u*v*, CIE L*a*b, Munsell, etc. RGB color space is most commonly used for picture display. It consists of red, green and blue color components known as additive primaries because in RGB space, color is obtained by combining three components.
(3) **Color Histogram**: The color histogram represents the colors distribution in a picture. Color histogram can be applied for any type of color space like RGB or HSV, etc. It is a smooth function described over the color space that approximates the pixel counts.
(4) **Color Moments**: Color moments are used to discriminate pictures based on their color distribution same as central moments which describe a probability distribution. Once these moments are calculated, they provide a measurement for the similarity between two pictures. Generally, first three color moments are used as characteristics in the picture recovery process, mean, standard deviation and skewness.
(5) **Color Correlogram**: This feature is defined by Huang et al. [3] in 1997 for picture indexing and comparison. It is a three-dimensional table indexed by color and distance between pixels. It differentiates how the spatial correlation of color changes with the distance of a picture. Color correlogram of a table indexed by color pairs. It is easy to calculate and quite small in size.

4 Literature Review

There are two important issues that need to address in the content-based image retrieval system; first, the image dataset is quite large in number, and the second important problem is image retrieval depends on the number of uncertainties and subjects, i.e., image types. So, image processing through CBIR needs to add more features to undertake the varieties of images. Image retrieval depends on the accuracy of the extraction process. Hence, through this review article, current research trends are reviewed which involves low to high-level feature extraction through various techniques. The following literature review describes different techniques used for color feature recovery.

Duanmu [4] proposed a recovery method for color pictures using color moment invariants in 2010. Color characteristics are computed from the individual picture and it uses small picture descriptors. The proposed method was implemented on object image library, COIL-100 of Columbia University. Average precision and recall rates are better than the state-of-the-art image recovery techniques. Using the picture color distribution, Chen et al. [5] in 2010 proposed a method for the color feature recovery. To extract the color characteristics, fixed cardinality (FC) and variable cardinality (VC) recovery methods are proposed which utilizes binary quaternion-moment-preserving (BQMP) thresholding technique. They have also devised comparing histograms by clustering (CHIC) along with earth mover's distance (EMD) measures. The database of a total of 1200 pictures used for the experimentation, collected from image albums of Corel Corporation and DataCraft Corporation, respectively. The proposed color feature recovery scheme shows an improvement in the recovery results by a factor of 25% over traditional methods. Kekre et al. [6], in 2010, proposed a new algorithm using fast Fourier transform (FFT) of each R, G and B component of an image separately. Each component of an image was divided into 12 mean values, and 6 upper half sectors are used to produce the feature vectors. Euclidean distances between the database pictures and feature vectors of a query picture are determined. To evaluate the effectiveness of the proposed algorithm, a database of 249 pictures of 10 different classes is used.

Rao et al. [7] in 2011 proposed a texture and dominant color-based picture recovery system. In this method, three characteristics of a picture such as dynamic dominant color, motif co-occurrence matrix (MCM) and the difference between pixels of scan pattern (DBPSP) are considered to retrieve a color picture. Using fast color quantization algorithm, the picture is separated into eight coarse partitions, and eight dominant colors are derived from the eight partitions. MCM and DBPSP are used to represent the texture distribution of a picture, and DDC represents color characteristics of the pixels in a picture. They experimented on Wang dataset. In 2011, Yue et al. [8] proposed a method by fusion of color and texture attributes and by constructing weights of feature vectors. Initially, HSV color space is quantified to derive feature vectors. Then, based on a co-occurrence matrix, color and texture attributes are extracted. Afterward, characteristics of global, local color histogram and texture attributes are compared and analyzed for recovery. Picture recovery accuracy is improved as shown by experimentation. They used a car picture database from web sources.

Afifi et al. [9] in 2012 proposed a new CBIR system using Ranklet transform (RT) and color feature to represent an image. For image enhancement operations and image invariant to rotation, Ranklet transform is used as a preprocessing step. K-means clustering algorithm used to cluster the picture respective to their feature to enhance the recovery time. They have used Wang database for experimentation. In 2013, Subrahmanyam et al. [10] proposed a new method, which integrates modified color motif co-occurrence matrix (MCMCM) and the difference between the pixels of a scan pattern (DBPSP) with equal weights for an effective CBIR system. The effectiveness of the proposed recovery method is verified on two different benchmark databases, such as MIT VisTex and Corel-1000. Walia et al. [11], in 2014, proposed a novel fusion framework for the recovery of color pictures. The proposed framework fuses both modified color difference histogram (CDH) which extracts the color and texture characteristics of an image and angular radial transform (ART) which extracts the shape characteristics globally or locally of an image. A variety of databases are considered to test the effectiveness of the applied fusion framework. In 2015, Guo et al.[12] proposed an image recovery system by using ordered dither block truncation coding (ODBTC) to produce a picture content descriptor. These picture characteristics are generated from the ODBTC encoded data streams, color co-occurrence feature (CCF), and second is a bit-pattern feature (BPF) by involving the visual codebook. They have used 13 different databases for testing. Shama et al. [13] in 2015 proposed an efficient indexing and recovery technique for plant pictures. The 2D-OTSU threshold-based segmentation technique is used to separate the object from the background. Modified color co-occurrence matrix (MCCM) and Gabor filter are used for feature recovery. Then, Euclidean distance measure was used to find the similarity between pictures. R* tree structure is used for better indexing and fast searching of pictures. They have created a database of 300 plant pictures. The experimental results proved better recovery accuracy with reduced time of recovery.

In 2016, Li et al. [14] implemented an improved algorithm that combines the fuzzy color histogram and block color histogram. It also considers local and global color information for color feature recovery and helps in decreasing the color feature dimension. Experiments are conducted using Corel 1000 dataset to check the effectiveness of the improved algorithm. In 2016, Somnugpong et al. [15] proposed a novel method by combining color correlogram and edge direction histogram (EDH) to obtain the robustness to spatial changes in an image. Spatial color correlation information is processed by color correlogram. In the case of the same picture but different colors, EDH provides the geometrical information of the same picture having different color. Experimental result on the Wang dataset proved that using combination of spatial correlation of picture, color and texture semantic works better than the combination of the traditional histogram and its texture. To improve picture recovery precision in 2017, Fadaei et al. [16] developed a new CBIR scheme which combines the optimization of color and texture characteristics. Dominant color descriptor (DCD) characteristics are extracted using a uniform partitioning scheme applied in HSV color space. To overcome the problem of noise and noise translation, various wavelet and curvelet characteristics are characterized as texture characteristics. The color and texture characteristics of a picture optimized by particle swarm optimization (PSO) algorithm which is applied on the Corel dataset. Khwildi et al. [17] in 2018 proposed a set of global descriptors for modeling high dynamic range (HDR) pictures and displaying the results in standard dynamic range (SDR) devices. It uses a vector of characteristics which is a combination of two color attributes, color histogram based on HSV space and color moments. The measure of Manhattan distance was used to obtain the dissimilarity between pictures. The proposed method accurately retrieved similar HDR pictures and tested on the LDR version of the same dataset. In 2019, Unar et al. [18] addressed the problem of combining both low-level visual characteristics and color information. The feature vectors are obtained for the low-level visual characteristics and color information. The similarity is computed for obtained feature vectors and combined. The top-rank pictures are retrieved using a distance metric. The experimentation was performed on Corel 1 K and Oxford flowers datasets.

Ahmed et al. [19] in 2020 proposed a novel method to retrieve the pictures by representing the shapes, texture, objects and spatial color information. L2 normalization applied on RGB channel used for the spatial arrangement. For efficient recovery and ranking, combined feature vectors are transformed to bag of words (BoW). The efficiency of the proposed approach is verified on nine standard picture datasets such as Caltech-101, 17-Flowers, Columbia object picture library(COIL), Corel-1000, PictureNet, Corel-10000, Caltech-256, tropical fruits and Amsterdam Library of Textures (ALOT). In 2020, Sathiamoorthy et al. [20] proposed a new variant of multi-trend structure descriptor (MTSD). It encodes a feature matrix of color, edge orientation and texture quantized values versus orientation of equal, small and large trends. Using discrete Haar wavelet transform, the picture is decomposed into the fine level to reduce the time cost and to preserve the accuracy of the proposed variant of MTSD. To find the similarity, the Euclidean distance measure is used. Considerable improvement is achieved as compared to the state-of-the-art descriptors on seven different datasets.

Zahid et al. [21] reported the new technique to overcome the issue of the semantic gap. The author has used a combination of speedup techniques with a histogram analyzer. Alshehri [22] used aerial satellite images for getting a detailed analysis of the vegetation present in the Riyadh region. The author used the hybrid technique of fuzzy and artificial neural network for retrieval of images. The author has compared the analysis with the existing techniques. It is found that the proposed technique shows better efficiency compared to the existing available techniques. Singh [23] et al. proposed a wavelet transform-based method for the image processing of satellite-based images.

A summary of the literature review described so far is given in Table 1.

5 Conclusion and Future Directions

This article reviewed the basic information about the content-based picture recovery model. Various techniques for color feature recovery are summarized amongst color, texture and shape from the last 10 years so far with the help of the literature survey. Recovery efficiency can be further improved by combining color feature with shape, texture or any other low-level feature of a picture. Hence, combining local and global characteristics can be the topic for future research to improve the performance of recovery systems. Furthermore, machine learning and deep learning approaches can be used for feature extraction to get improved results and fast processing.

Table 1 Summary of the literature review

Sr. no.	Authors	Technique	Proposed method	Outcome/merits	Demerits/future directions	Dataset used
1	Duanmu [4]	Color moment invariant	Small picture descriptors	Better average precision and recall rates	Color moment invariant cannot deal with occlusion very well	COIL-100
2	Chen et al. [5]	BQMP thresholding technique, fixed cardinality, variable cardinality, earth mover's distance	Proposed FC and VC methods with BQMP thresholding technique using picture color distribution. Devised new distance measure, comparing histograms by comparing (CHIC)	Recovery precision rate is enhanced by 25% than traditional methods. CHIC distance method reduces execution time than EMD measure	–	Corel Corp DataCraftcorp
3	Kekre et al. [6]	Fast Fourier transform, Euclidian distance	FFT's of each R, G, B components are divided into 12 sectors and 6 upper half sectors are used to produce feature vectors	Average precision and overall average performance of precision and recall of each class ha cross-over point at 50%	–	249 picture of 10 different classes
4	Babu Rao et al. [7]	Dynamic dominant color, motif co-occurrence matrix, DBPSP picture characteristics are used	Using fast color quantization algorithm, an picture is divided into eight partitions and from which eight dominant color s are obtained	It outperforms Hung's and Jhanwar's methods	–	Wang dataset

(continued)

Table 1 (continued)

Sr. no.	Authors	Technique	Proposed method	Outcome/merits	Demerits/future directions	Dataset used
5	Yue et al. [8]	HSV color space, co-occurrence matrix and Euclidian distance	Based on co-occurrence matrix, color and texture characteristics are extracted to obtain feature vectors. HSV color space is quantified	Recovery accuracy is improved	Other low-level features like shape, spatial location can be incorporated	Car pictures database from web sources used for experimentation
6	Afifi et al. [9]	Ranklet transform, K-means clustering algorithm	Ranklet transform is used for each picture layer, i.e., R, G, B and color moments are calculated using K-means clustering algorithm	Maximum average of 0.93 at recall value of 0.05 is achieved, and it performs better than other 4 systems used for comparison	Combination of shape or texture feature along with color can give good results	Wang dataset
7	Subrahmanyam et al. [10]	Modified color motif co-occurrence matrix (MCMCM), difference between pixels of scan pattern (DBPSP)	Nine color patterns are generated from the separated R, G, B and planes of a color picture. MCMCM and DBPSP characteristics are integrated with equal weights	Improvement in average precision, recovery rate and recall on DB1 and DB2 was observed when compared with traditional methods	–	MIT VisTex(DB1), Corel-1000(DB 2)

(continued)

Table 1 (continued)

Sr. no.	Authors	Technique	Proposed method	Outcome/merits	Demerits/future directions	Dataset used
8	Walia et al. [11]	Modified color difference histogram (CDH), angular radiant transform (ART)	Modified CDH algorithm and ART methods are used to extract the color, shape and texture characteristics of a color picture	It improved average recovery accuracy by apprx. 16% and 14% over CDH and ART resp.	–	Wang;s picture database, Olivia and Torralba (OT) scene database, VisTex database
9	Guo et al. [12]	Ordered dither block truncation coding (ODBTC)	Color co-occurrence feature and bit patterns characteristics of picture are generated from ODBTC encoded data streams	It provides best average precision rate compared to earlier schemes	Can be applied to video retrieval	13 databases are used such as Corel, Vis tex-640, UKBench, etc.
10	Shama et al. [13]	2D-OTSU threshold-based segmentation, modified color co-occurrence matrix (MCCM), Gabor filter, Euclidean distance R*tree indexing	To obtain object from background, 2D-OTSU threshold-based segmentation is used. MCCM and Gabor filters are used for feature recovery	Improved recovery accuracy with reduced time of recovery	Larger datasets can be used for further experimentation	300 plant picture databases

(continued)

Table 1 (continued)

Sr. no.	Authors	Technique	Proposed method	Outcome/merits	Demerits/future directions	Dataset used
11	Li et al. [14]	Fuzzy color histogram (FCH), block color histogram (BCH)	It considers the local and global color information by combining FCH and BCH and decreasing the color feature dimension	Best recovery accuracy and low feature dimension	–	Corel 1000
12	Somnugpong et al. [15]	Color correlogram, edge detection histogram (EDH), Euclidean distance	Spatial color correlation information is processed by color correlogram and EDH provides the geometrical information	It works better than the combination of traditional histogram and its texture	Proposed technique can be practically implemented	Wang dataset
13	Fadaei et al. [16]	Dominant color descriptor (DCD), particle swarm optimization (PSO) algorithm	It extracts the dominant color descriptor characteristics and uniform partitioning scheme is used in HSV color space. Texture and color characteristics are combined by PSO algorithm	Higher average precision of 76.5% achieved when compared to state-of-the-art techniques	Segmentation is applied to whole image for feature extraction instead of main regions	Corel database

(continued)

Table 1 (continued)

Sr. no.	Authors	Technique	Proposed method	Outcome/merits	Demerits/future directions	Dataset used
14	Khwildi et al. [17]	Color histogram, color moments, Manhattan distance	For modeling of HDR picture, a set of global descriptors are used by using HSV space-based color histogram and color moment attributes	Better performance regarding accuracy and efficiency	Other descriptors can be combined on large HDR image dataset	100 picture database collected from website
15	Unar et al. [18]	K-means clustering algorithm, Euclidean distance	Color information is extracted and segmented with nonlinear L*a*b color space to get a feature vector	Improved efficiency and 85% accuracy over other methods	Machine learning technique can be applied for better results	Corel 1000, Oxford Flowers
16	Ahmed et al. [19]	Gaussian smoothing, Hessian blob detector bag of words (BoW)	Color characteristics are extracted using L2 norm and high-variance coefficients are transformed to BoW for effective recovery and ranking	Higher precision in average, mean and average recovery, and also for recall rated can be observed when tested with other picture groups of standard datasets	Convolutional neural network can be used for improvement	9 datasets: Caltech-101, Corel-1000, Columbia object picture library, 17 Flowers, Picturenet, Corel-10000, Caltech-256, ALOT
17	Sathiamoorthy et al. [20]	Multi-trend structure descriptor (MTSD), Discrete Haar wavelet transform, Euclidean distance measure	MTSD encodes orientation details of local level structures, and Haar wavelet transform is used to decompose the picture into fine level	Considerable improvement is achieved with the help of precision and recall	Proposed technique can be applied to medical images using machine learning approach	7 databases: Corel 1 K, 5 K, 10 K, Caltech 101, LIDC-IDRI

References

1. L. Fuhui, Z. Hongjiang, F. David, Fundamentals of content based picture recovery, in *Multimedia Information Recovery and Management* (Springer, Berlin, 2003), pp 1–26. ISSN: 1860-4862
2. S. Aghav-Palwe, D. Mishra, Color picture recovery using compacted feature vector with mean count tree. Procedia Comput. Sci. **132**, 1739–1746 (2018)
3. H. Jing, S. Ravi Kumar, M. Mhetre, Z. Wei-Jing, Z. Ramin, Picture ındexing using color correlograms, in *IEEE International Conference on Computer Vision and Pattern Recognition* (1997), pp. 762–768
4. D. Xiaoyin, D, Picture recovery using color moment invariant, in *Proceedings of Seventh International Conference on Information Technology: New Generations (ITNG)* (IEEE, 2010), pp. 200–203
5. C. Wei-Ta, L. Wei-Chuan, Ming-Syan, Adaptive color feature recovery based on picture color distribution. IEEE Trans. Picture Process. **19**(8) (2010)
6. H. Kekre, D. Mishra, CBIR using upper six FFT sectors of Color pictures for feature vector generation. Int. J. Eng. Technol. **2**(2) (2010)
7. M.B. Rao, B.P. Rao, A. Govardhan, CTDCIRS: content based picture recovery system based on dominant color and texture characterstics. Int. J. Comput. Appl. **18**(6) (2011)
8. J. Yue, Z. Li, L. Liu, Z. Fu, Content-based picture recovery using color and texture fused characterstics. Math. Comput. Model. **54** (2011)
9. A.J. Afifi, W,M. Ashour, Picture recovery based on content using Color feature. ISRN Computer Graphi. **2012**, Article ID 248285 (2012)
10. M. Subrahmanyam, Q.M. Jonathan, R.P. Maheshwari, R. Balasubramanian, Modified color motif co-occurrence matrix for picture indexing and recovery. Comput, Electr. Eng. **39**(3) (2013)
11. E. Walia, A. Pal, Fusion framework for effective Color picture recovery. J. Visual Commun. Picture Represent. **25**(6) (2014)
12. J.M. Guo, H. Prasetyo, Content based picture recovery using characterstics extracted from halftoning-based block truncation coding. IEEE Trans. Picture Process. **24**(3) (2015)
13. P.S.Shama, K. Badrinath, T. Anand, An efficient ındexing approach for content based picture recovery. Int. J. Comput. Appl. **117**(15) (2015)
14. M. Li, X. Jiang, An improved algorithm based on color feature recovery for picture recovery, in *8th IEEE International conference on Intelligent Human-Machine Systems and Cybernetics* (2016), pp. 281–285
15. S Somnugpon, K. Khiewwan, Content-based picture recovery using a combination of color correlograms and edge direction, in *13th International Joint Conference on Computer Science and Software Engineering (IJCSSE)* (2016)
16. S. Fadaei, R. Amirfattahi, A.M. Ahmadzadeh, New content based picture recovery system based on optimised integration of DCD, wavelet and curvelet characterstics. IET Picture Process. **11**(2) (2017)
17. R. Khwildi, A.O. Zaid, Color based HDR picture recovery using HSV histogram and color moments, in *IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)* (2018)
18. S. Unar, X. Wang, C. Wang, M. Wang, New strategy for CBIR by combining low-level visual characterstics with a colour descriptor. IET Picture Process. **13**(7) (2019)
19. K.T. Ahmed, H. Afzal, M.R. Mufti, A. Mehmood, G.S. Choi, Deep picture sensing and recovery using suppression, scale spacing and division, interpolation and spatial Color coordinates with bag of words for large and complex datasets. IEEE Access **8** (2020)
20. S. Sathiamoorthy, M. Natarajan, An efficient content based picture recovery using enhanced multi-trend structure descriptor. Appl. Sci. **2**, 217 (2020)
21. Z. Mehmood, F. Abbas, M. Mahmood, M.A, Javid, A. Rehman, T. Nawaz, T. Content-based ımage retrieval based on visual words fusion versus features fusion of local and global features. Arab. J. Sci. Eng. (2018)

22. M. Alshehri, Content-based ımage retrieval method using neural network-based prediction technique. Arab. J. Sci. Eng. (2020)
23. D. Singh, D. Garg, P.H. Singh, Efficient land satellite image fusion using fuzzy and stationary discrete wavelet transform. Imaging Sci. J. **65**(2) (2017)

A New Method of Interval Type-2 Fuzzy-Based CNN for Image Classification

P. Murugeswari and S. Vijayalakshmi

Abstract Last two decades, neural networks and fuzzy logic have been successfully implemented in intelligent systems. The fuzzy neural network system framework infers the union of fuzzy logic and neural system framework thoughts, which consolidates the advantages of fuzzy logic and neural network system framework. This FNN is applied in many scientific and engineering areas. Wherever there is an uncertainty associated with data fuzzy logic place a vital rule, and the fuzzy set can represent and handle uncertain information effectively. The main objective of the FNN system is to achieve a high level of accuracy by including the fuzzy logic in either neural network structure, activation function, or learning algorithms. In computer vision and intelligent systems such as convolutional neural network has more popular architectures, and their performance is excellent in many applications. In this article, fuzzy-based CNN image classification methods are analyzed, and also interval type-2 fuzzy-based CNN is proposed. From the experiment, it is identified that the proposed method performance is well.

Keywords CNN · FCNN · Fuzzy logic · Interval type-2 fuzzy logic · Feature extraction · Computer vision · Image classification

1 Introduction

In computer vision, image classification is that the undertaking of characterizing a given image into one among the pre-defined characterized classes. Customary image classification consists of feature extraction and classification modules. Feature extraction includes extricating a superior degree of pixel information from the raw pixel which will catch the greatness among the classifications. Normally, this process

P. Murugeswari (✉)
Department of Computer Science and Engineering, Karpagam College of Engineering, Coimbatore, Tamil Nadu, India

S. Vijayalakshmi
Department of Computer Applications, NMS S. Vellaichamy Nadar College, Madurai, Tamil Nadu, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*, Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_57

is done using an unsupervised manner, wherein the classes of the picture do not have anything to do with data extricated from pixels. A portion of the typical and generally utilized feature extractions is GIST, HOG, SIFT, LBP, and so on. After the feature extraction, a classification module is prepared with the image and their related names. A couple of tests of this module are SVM, logistic regression, random forest, choice trees, and so on.

The different types of neural network architecture, such as recurrent neural networks (RNN), long short-term memory (LSTM), artificial neural networks (ANN) and convolutional neural network (CNN), etc., were analyzed. CNNs are the most popular architecture and suitable for image database. It works excellently on computer vision tasks like image classification [1, 2], object detection [3], image recognition [4], etc. In CNN, there is no different element extractor, it is implicit, which incorporates, include extraction and characterization modules coordinated framework, and it figures out how to remove, by separating representations from the pictures and order them dependent on regulated information.

CNN is utilized in various assignments that have an extraordinary presentation in various applications. CNN has been presenting an employable class of models for better information on substance present in the image along these lines producing better image acknowledgment, segmentation, identification, and retrieval. CNN structures are productively and successfully utilized in many patterns and image recognition applications [5], for instance, motion acknowledgment [4, 6], face acknowledgment [7, 8], object characterization [9, 10], and creating scene descriptions [11].

Zadeh [12] presented the idea of fuzzy logic (type-1 fuzzy) for tackling the control framework related issues. Later, analysts have contributed many fascinating applications with regards to the field of computer vision. Thoughtfully, type-2 fuzzy set (T2FS) was presented by Zadeh in 1975[13], and further, it is created by Jerry M. Mendel. In type-1 fuzzy set (T1FS), the degree of participation is determined by a crisp number having a place with the interval [0–1], in T2FS, the degree of participation is itself fuzzy, and it is indicated by a secondary membership function. If the secondary membership function values are at its limit of 1 at each point which is called as interval type-2 fuzzy set (IT2FS) [13–15]. The T2FS incorporates a third measurement and impression of uncertainty as appeared in Fig. 1, which gives the additional level of opportunity to deal with uncertainty. This additional level of fuzziness helps an increasingly capable method of taking care of uncertainty. Figure 2 shows the secondary membership functions (MFs) (third element) of the T1FS (Fig. 2a), the IT2FS (Fig. 2b), and the general T2FS (Fig. 2c) as initiated by a similar information p as appeared in Fig. 1.

For the most part, FCM type-1 has become the most notable calculation in group investigation. Numerous analysts have demonstrated that there are imperatives in the limit of T1 FSs to show and breaking point the effect of uncertainties since its interest grades are crisp. The T2FS is addressed by membership functions (MFs) that are themselves fuzzy. The IT2FS [16], an exceptional instance of T2FS, are at present most generally utilized taking into their decreased computational expense. An IT2FS is restricted with two T1FSs above and underneath, which are called upper MF (UMF)

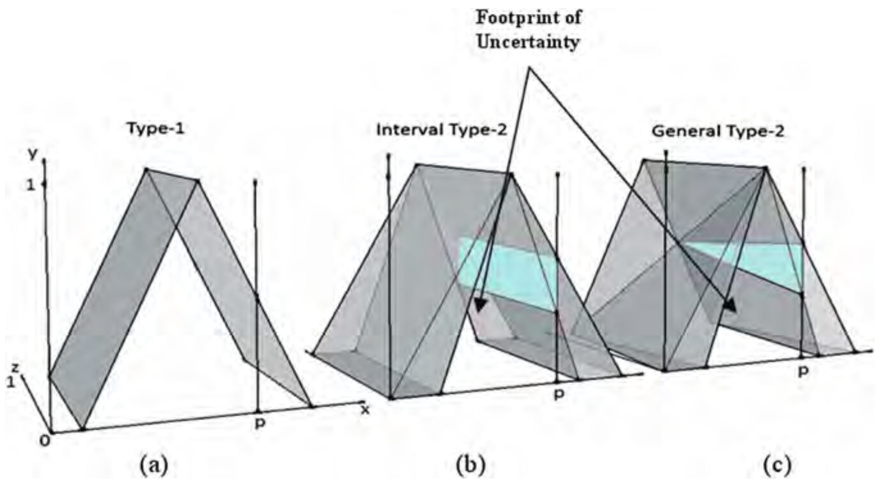


Fig. 1 A case of the three kinds of fuzzy sets. A similar information p is applied to each fuzzy set. **a** T1FS, **b** IT2FS, and **c** T2FS

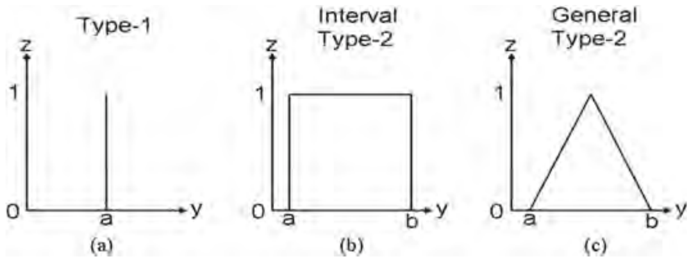


Fig. 2 A perspective on the secondary membership function (three dimensions) initiated by an information *p* for **a** T1FS, **b** IT2FS, and **c** T2FS

and lower MF (LMF) independently, and domain among UMF and LMF is footprint of uncertainty (FOU). T2FS exhibits to show various uncertainties yet it increase the computational unpredictability due to its extra component of optional evaluations of every essential enrollment. Model applications are type-2 fuzzy clustering [17], Gaussian noise filter, classification of coded video streams, medical applications, and color picture division.

2 Literature Survey

CNN is a sort of neural networks, which has indicated commendable execution on a couple of difficulties related to computer vision and image processing. A part of the invigorating application areas of CNN fuse image classification and segmentation [18], object detection [19], video processing [20], natural language processing [21, 22], and speech recognition [23, 24]. The learning limit of significant CNN is basically a result of the usage of various component extractions composes that can normally take in exposé from the data. The availability of a mass of data and improvement in the gear development has accelerated the investigation in CNNs, and starting late attractive profound CNN models have been represented. A few moving plans to get advancement in CNNs have been investigated [7], for example, the utilization of various activation and loss functions, parameter streamlining, regularization, and compositional advancements.

Mendel [25] stated that to use a T1FS to model a word is scientifically incorrect because a word is uncertain, whereas a type-1 FS is certain. Therefore, made an in-depth research in type-2 fuzzy and contributed many papers [26–28] in type-2 fuzzy logic. Based on that, many researchers have been contributed many algorithms for their applications. For example, the classification of coded video streams, diagnosis of diseases, pre-processing radiographic images, medical image applications, transport scheduling, forecasting of time series, learning linguistic membership grades, inference engine design, control of mobile robots, and so on. The computational complexity is high in type-2 fuzzy. Therefore, the type-2 fuzzy set is simplified into IT2 fuzzy which computational complexity can be significantly reduced in appropriate applications.

Recently, fuzzy logic and neural networks are widely applied to solve real-world problems. Fuzzy logic is a lot of mathematical standards for information representation dependent on degrees of participation as opposed to binary logic. It is an incredible asset to handle imprecision and uncertainty, and it was introduced to pick up robustness and low-cost resolution for real-world problems [29]. For the most part, type-1 fuzzy logic frameworks have been executed in numerous systems to a wider scale some of which include approximation and forecasting systems, control systems, databases, healthcare clinical diagnosis, and so on.

Researchers have been combined neural network and fuzzy logic and implemented successfully in intelligent systems. The fuzzy neural network (FNN) system framework implies together of fuzzy logic and neural network system ideas, which incorporates the benefits of fuzzy logic and neural network system. This FNN is applied in many scientific and engineering areas text sentiment analysis [30], object classification with small training database [19], emotion features extraction from text [31], emotion understanding in movie [32], real-world objects, and image classification[20, 33], to recognize Marathi handwritten numerals[34, 35], to predict the traffic flow [36], electric load forecasting [37], and to recognize handwritten digits [38]. Keller et al. [39] proposed a hierarchical deep neural network fuzzy system which obtains information from both fuzzy and neural representations. Price et al. [40] proposed introducing the fuzzy layers for deep learning, and fuzzy methodologies taken to deep learning have experienced the use of different combination procedures at the choice level to total yields from best in class pre-prepared models, e.g., AlexNet, VGG16, GoogLeNet, Inception-v3, ResNet-18, and so on.

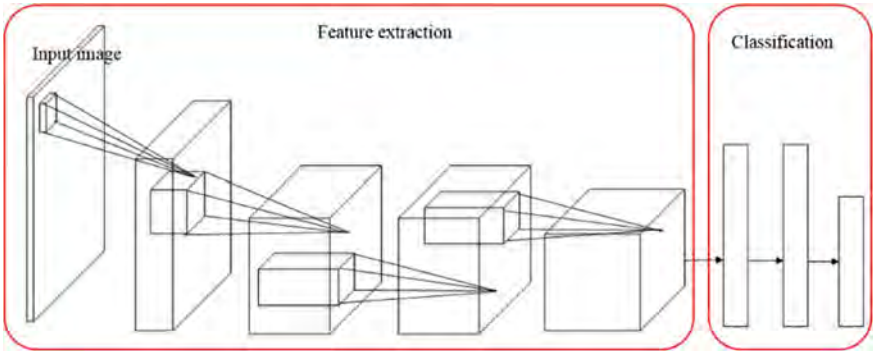


Fig. 3 Structure of a convolutional neural network (CNN)

Generally, CNN architecture consists of two phases; there are feature extraction and classification. The FCNN is the combination of CNN and fuzzy logic; therefore, the fuzzy logic may include either feature extraction phase or classification phase. Dependence on the application, the researchers have proposed various FCNN architectures including fuzzy logic in the feature extraction phase or classification phase. Here, the two FCNN architectures have been compared for image classification. In these two architectures, fuzzy logic is included in the classification phase. Hsu et al. [19] have approached integrated a convolutional neural network with a fuzzy neural network (FCNN Model 1), where the FNN summarizes the feature information from every fuzzy map. Korshunova [9] (FCNN Model 2) have proposed CFNN architecture which includes the fuzzy layer, which situated in between the convolutional network and classifier (Fig. 3).

3 Proposed Method

The new inter type-2 fuzzy CNN architecture integrates the features from CNN and the FNN. A new architecture integrates the interval type-2 fuzzy rectifying unit (IT2FRU) [41] activation function in convolution for features extraction in CNN and interval type-2 fuzzy-based classification in the fuzzy layer. This method combines the advantages of both network architectures and interval type-2 fuzzy logic. IT2FCFNN architecture has four types of layers: (i) convolutional layer with IT2FRU; (ii) pooling layer; (iii) fuzzy layer; and (iv) a fuzzy classifier.

The convolutional neural framework takes a data image and performs a course of action of convolutional and pooling layers. The fuzzy layer performs grouping using the interval type-2 fuzzy clustering algorithm. The yields of the fuzzy layer neurons speak to the estimations of the participation capacities for the fuzzy clustering of input data. The information point’s cluster is chosen relies on their participation grade. These characteristics go to the promise of a classifier. Its yield is the full

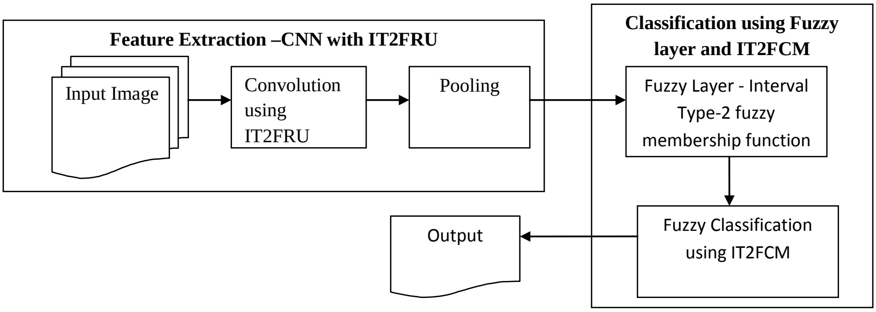


Fig. 4 Outline of the proposed method IT2FCNN

IT2FCNN and is the class scores for the picture. Leave C alone the number of neurons of the fuzzy layer (the number of clusters). The neurons of the fuzzy layer commencement limits are IT2FRU showing the interest of the information vector x to all of the L groups (Fig. 4).

IT2FRU employs the following equalities; $Z = 0$ to guarantee that $\sigma = 0 \Rightarrow \varphi o = 0$. Additionally, the height of the LMFs is employed as $m2 = \alpha$, $m1 = m3 = 1 - \alpha$ as suggested in [26]. The resulting IT2-FM ($\varphi o\ (\sigma)$) for $\sigma \in [0, 1]$ can be formulated

$$\varphi_o(\sigma) = P\sigma k(\sigma) \tag{1}$$

where $k(\sigma)$ is defined as

$$k(\sigma) = \frac{1}{2}\left(\frac{1}{\alpha + \sigma - \alpha\sigma} + \frac{-1 + \alpha}{-1 + \alpha\sigma}\right) \tag{2}$$

Similarly, for the input interval $\sigma \epsilon [-1, 0]$, the IT2FM can be derived as

$$\varphi_o(\sigma) = N\sigma k(-\sigma) \tag{3}$$

The activation unit can be formulated by arranging Eqs. (1) and (3) as following:

$$f(\sigma) = \begin{cases} P\sigma k(\sigma), & \text{if } \sigma > 0 \\ N\sigma k(-\sigma), & \text{if } \sigma \le 0 \end{cases} \tag{4}$$

the parameter P controls the incline of the capacity in the positive quadrant, while the parameter N controls the slant of the capacity in the negative quadrant. The resulting output of the IT2-FRU could be a linear or nonlinear activation depending on the selection of the parameters. IT2FRU has three learnable parameters P, N, and α.

The vector $= \left[x_1, x_2 \ldots x_j \ldots x_n\right]$ is dealt with to the commitment of the framework, and the fuzzy layer formed a vector involving the degrees of having a spot x with the specific cluster territories: $\left[v_i v_2 \ldots v_j\right]$. The parts $(\bar{u}_j(x_i), \underline{u}_j(x_i))$ are determined utilizing Eq. (5) to fulfill the standardization condition utilizing Eq. (6) for each preparation test vector $x^{(k)}$, $k = 1, \ldots K$ K, where K is the number of vectors in preparing ready to set. The yields of neurons of the fuzzy layer are used as commitments of the classifier.

$$\tilde{\mu}_i\left(x^k\right) = f\sum_{j=1}^{n} x_j^k \tag{5}$$

$$\sum_{i=1}^{L} \tilde{\mu}_i\left(x^k\right) = 1 \tag{6}$$

(a) The interval type-2 fuzzy membership becomes

$$\overline{u}_j(x_i) = \begin{cases} \frac{\sum_{k=1}^{C}\left(\left(d_{ji}/d_{ki}\right)+\alpha\left(d_{ji}/d_{ki}\right)\delta\right)^{2/(m_1-1)}}{\sum_{k=1}^{C}\left(\left(d_{ji}/d_{ki}\right)+\alpha\left(d_{ji}/d_{ki}\right)\delta\right)^{2/(m_1-1)}}, & \text{if } \frac{1}{\sum_{k=1}^{C}\left(d_{ji}/d_{ki}\right)} < \frac{1}{C} \\ \frac{\sum_{k=1}^{C}\left(\left(d_{ji}/d_{ki}\right)+\alpha\left(d_{ji}/d_{ki}\right)\delta\right)^{2/(m_2-1)}}{\sum_{k=1}^{C}\left(\left(d_{ji}/d_{ki}\right)+\alpha\left(d_{ji}/d_{ki}\right)\delta\right)^{2/(m_2-1)}}, & \text{otherwise} \end{cases} \tag{7}$$

$$\overline{u}_j(x_i) = \begin{cases} \frac{\sum_{k=1}^{C}\left(\left(d_{ji}/d_{ki}\right)+\alpha\left(d_{ji}/d_{ki}\right)\delta\right)^{2/(m_1-1)}}{\sum_{k=1}^{C}\left(\left(d_{ji}/d_{ki}\right)+\alpha\left(d_{ji}/d_{ki}\right)\delta\right)^{2/(m_1-1)}}, & \text{if } \frac{1}{\sum_{k=1}^{C}\left(d_{ji}/d_{ki}\right)} < \frac{1}{C} \\ \frac{\sum_{k=1}^{C}\left(\left(d_{ji}/d_{ki}\right)+\alpha\left(d_{ji}/d_{ki}\right)\delta\right)^{2/(m_2-1)}}{\sum_{k=1}^{C}\left(\left(d_{ji}/d_{ki}\right)+\alpha\left(d_{ji}/d_{ki}\right)\delta\right)^{2/(m_2-1)}}, & \text{otherwise} \end{cases} \tag{8}$$

(b) Updating cluster centers

$$v_j = \frac{v_L + v_R}{2}. \tag{9}$$

(c) Type reduction and hard partitioning can be acquired as follows:

$$u_j(x_i) = \frac{u_j^R(x_i) + u_j^L(x_i)}{2}, \quad j = 1, \ldots, C \tag{10}$$

$$u_j^R(x_i) = \frac{\sum_{l=1}^{M} u_{jl}(x_i)}{M} \tag{11}$$

where $u_{jl}(x_i) = \begin{cases} \overline{u}_j(x_i), \text{ if } x_{il} \text{ uses } \overline{u}_j(x_i) \text{ for } v_j^R \\ \underline{u}_j(x_i), \text{ otherwise} \end{cases}$ and

$$u_j^L(x_i) = \frac{\sum_{l=1}^{M} u_{jl}(x_i)}{M} \tag{12}$$

where $u_{jl}(x_i) = \begin{cases} \overline{u}_j(x_i), \text{ if } x_{il} \text{ uses } \overline{u}_j(x_i) \text{ for } v_j^L \\ \underline{u}_j(x_i), \text{ otherwise} \end{cases}$.

The processes of IT2CFNN is isolated into three phases: The information design (picture) gets through a progression of changes; subsequently, a vector of significant level attributes is framed; further, fuzzy layer plays out groundwork dissemination of the information into fuzzy groups; the last totally related layers play out the plan, consigning the result class name to each get-together of clusters.

4 Experimental Result

4.1 Dataset and Experimental Setup

The various datasets are available to apply the neural networks. The most popular datasets are CIFAR-10, Caltech101, and ImageNet. The CIFAR-10 dataset consists of 60,000 images in 10 classes, with 6000 images per class. The caltech101 dataset consists of 101 classes with 40 to 800 images per class. The ImageNet dataset has more than 14 million images, with 20,000 categories. The experiments have been executed in Windows 7 with 64 pieces working framework, and the principle memory and capacity limit of the PC frameworks are 8 GB RAM and 1 TB separately. The kind of the CPU utilized is Intel(R) Core(TM) i5-4590 CPU @3.30 GHz, and the design of the illustrations card is NVIDIA GeForce GT 705. The product utilized in this investigation is Python 3.6 and MATLAB. The PyCharm IDE is utilized, and organize is fabricated utilizing the Keras libraries on the PyCharm exploratory stage.

4.2 Training the Architecture

The training of IT2FCNN is the foremost step, which includes three autonomous steps of the three components of the net. First, the model is prepared to utilize the theoretical properties of the input image by the back-propagation model. In the second part of the model, the fuzzy layer is tuned using the competitive learning scheme, which means choosing the parameters of the membership function for setting the cluster centers. Various fuzzy clustering algorithms are available; here, IT2FCM is used for clustering. Finally, the classifier is trained using the weights tuning in the fully connected layers. When the training part is completed now the IT2FCNN becomes ready for implementation, now the image pixel cluster is taken care of to the CFNN, the yield of the system is input image p class scores, and image is allocated to the class max score esteem class.

4.3 Comparative Analysis

In this article, the AlexNet, ZFNet, GoogleNet, VGGNet16, ResNet50 pre-trained on CIFAR, ImageNet and Caltech101 datasets have been chosen for the experiment. The CFNN model has fine-tuned the AlexNet, ZFNet, GoogleNet, VGGNet16, ResNet50 to classify the images. Here, 2, 5, and 7 epochs have been taken for training the models. In the fuzzy layer, IT2FCM clustering has been used to cluster the set of data several times with a different number of clusters. When the fuzzy partition coefficient is maximized, the number of clusters has been chosen for the experiment. Adam is the stochastic optimization method that is used to classifier training (to tune weight) for the fully connected layer.

Table 1 shows the model performance compared with existing CNN and fuzzy-based CNN architectures. Figure 5 shows that the fuzzy-based CNN architecture increases the performance accuracy compared with traditional CNN architecture. The investigation distinctly shows that while remembering the fuzzy layer for the CNN which gives the high caliber of exactness contrasted with comparing customary CNN.

4.4 The Comparison of RMSE, MSE, and MAE

The percentage of error (%Error), mean squared error (MSE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) are the performance criteria for image classification; the corresponding calculation method is defined as follows:

$$\%\text{Errr} = \frac{\left|y_i - \overline{y}_i\right|}{y_i} \times 100 \tag{13}$$

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y^x - y\right)^2 \tag{14}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y^x - y)^2} \tag{15}$$

$$\text{MAPE} = \frac{100}{N} \times \sum_{i=1}^{N}\left|\frac{y - y^x}{y}\right| \tag{16}$$

Table 2 shows the comparison of FCNN models with IT2FCNN-based MSE and RMSE. The result shows that proposed to produce a good result.

Table 1 Performance comparative analysis with various fine-tuning epochs (3, 5, and 7, respectively)

Model	Fine-tuning epochs	Dog versus cat				Lion versus tiger				Horse versus donkey			
		Regular	FCNN Model 1	FCNN Model 2	IT2FCNN	Regular	FCNN Model 1	FCNN Model 2	IT2FCNN	Regular	FCNN Model 1	FCNN Model 2	IT2FCNN
AlexNet	3	40	54	58	60	56	58	60	64	54	58	60	64
	5	51	60	61	65	61	64	65	68	60	62	64	68
	7	54	65	68	72	66	68	72	74	65	68	71	74
ZFNet	3	41	53	56	61	56	58	61	64	53	58	61	64
	5	53	62	61	65	62	63	65	68	62	64	68	70
	7	54	64	67	73	68	69	73	76	64	69	72	76
GoogleNet	3	42	56	57	61	58	61	61	64	56	61	64	64
	5	54	61	64	68	64	68	68	72	61	68	70	73
	7	57	68	70	74	68	70	74	78	68	70	74	78
VGGNet16	3	44	58	57	61	58	57	61	65	58	57	62	65
	5	53	62	64	68	64	66	68	72	62	66	69	72
	7	55	70	72	76	70	72	76	78	70	72	74	78
ResNet50	3	43	56	56	63	59	61	63	65	56	61	64	65
	5	54	61	62	67	63	62	67	69	61	62	65	69
	7	56	69	71	78	68	72	76	79	69	72	74	79

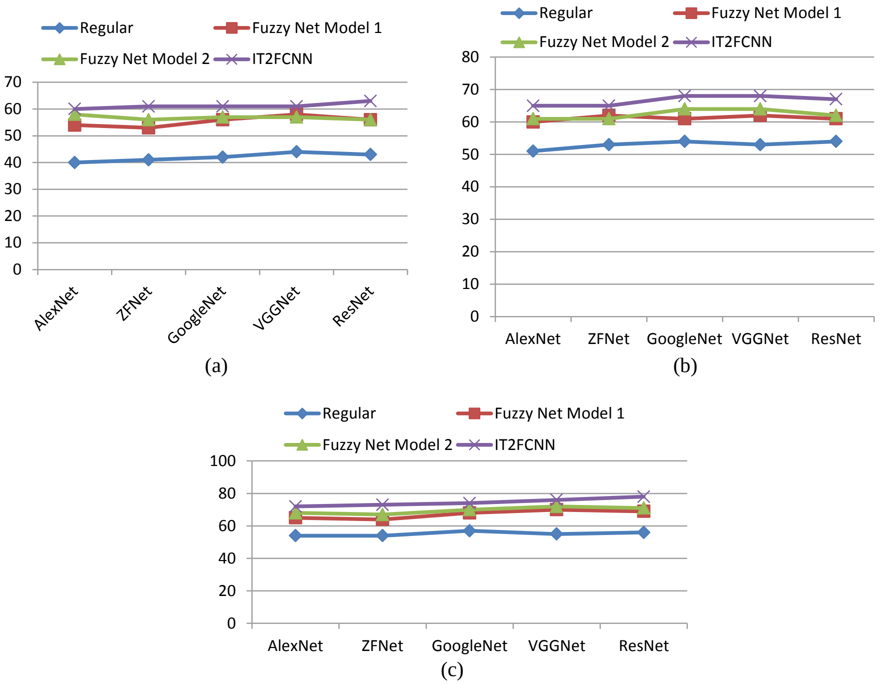


Fig. 5 Performance comparative analysis for dog versus cat with fine-tuning epochs 3, 5, and 7—(**a**, **b**, and **c**), respectively

Table 2 Comparison of FCNN models with IT2FCNN based on MSE and RMSE

Models		MSE	RMSE	MAPE
AlexNet	FCNN Model1	0.00245	0.052	4.4
	FCNN Model 2	0.00183	0.043	3.2
	IT2FCNN	0.00123	0.035	2.4
ZFNet	FCNN Model1	0.00254	0.054	4.2
	FCNN Model 2	0.00143	0.045	3.1
	IT2FCNN	0.00134	0.037	2.1
GoogleNet	FCNN Model1	0.00249	0.053	4.2
	FCNN Model 2	0.00197	0.041	3.2
	IT2FCNN	0.00123	0.036	2.0
VGGNet16	FCNN Model1	0.00244	0.056	4.2
	FCNN Model 2	0.00158	0.047	3.0
	IT2FCNN	0.00198	0.039	2.2
ResNet50	FCNN Model1	0.00268	0.058	4.1
	FCNN Model 2	0.00139	0.046	3.1
	IT2FCNN	0.00132	0.037	2.1

5 Conclusion

The exploratory outcomes state that the fuzzy neural networks articulate to a groundbreaking and reasonable option in contrast to regular arrangement strategies. The merging of fuzzy logic with neural network applications is progressively experts of decision marking systems. In the proposed method, CNN is used to extract the features and integrating the interval type-2 fuzzy to classify the images, which increases the accuracy of the experiment. Besides, our test results demonstrate that it is conceivable to improve testing exactness by watching the conveyance of pixels it includes maps and modifying the membership function. This method gives a better solution, and it has more advantages than the other existing methods. Although the results are more optimistic, image classification based on interval type-2 fuzzy logic still requires more future research.

References

1. A. Howard, Some improvements on deep convolutional neural network based image classification. ICLR (2014)
2. Nadia Jmour,Sehla Zayen, Afef Abdelkrim, Convolutional neural networks for image classification, International Conference on Advanced Systems and Electric Technologies, 2018
3. M.-J. Hsu, Y.-H. Chien, W.-Y. Wang, C.-C. Hsu, A convolutional fuzzy neural network architecture for object classification with small training database. Int. J. Fuzzy Syst. **22**, 1–10 (2020)
4. X. Yan Wu, A hand gesture recognition algorithm based on DC-CNN. Multimedia Tools Appl. **79**, 9193–9205 (2020)
5. L. Li, S. Qin, Z. Lu, K. Xu, Z. Hu, One-shot learning gesture recognition based on joint training of 3D ResNet and memory module. Multimedia Tools Appl. **79**, 6727–6757 (2020)
6. Z. Zeng, Q. Gong, J. Zhang, CNN model design of gesture recognition based on tensorflow framework, in *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)* (2019)
7. C.-J. Lin, C.-H. Lin, C.-C. Sun, S.-H. Wang, Evolutionary-fuzzy-integral-based convolutional neural networks for facial image classification (2019)
8. T. Ahonen, A. Hadid, M. Pietikinen, Face description with local binary patterns: Application to face recognition. Pattern Anal. Mach. Intell. 2037–2041 (2016)
9. K.P. Korshunova, A convolutional fuzzy neural network for image classification, in *3rd Russian-Pacific Conference on Computer Technology and Applications (RPC)* (2018)
10. W. Wang, Y. Yang, X. Wang, W. Wang, J. Li, Development of convolutional neural network and its application in image classification: a survey. Opt. Eng. **58**(4) (2019)
11. R. Chauhan, K.K. Ghanshala, R.C. Joshi, Convolutional neural network (CNN) for image detection and recognition, in *First International Conference on Secure Cyber Computing and Communication (ICSCCC)* (2018)
12. L.A. Zadeh, Fuzzy sets. Inf. Control **8**, 338–353 (1965)
13. L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, Part I. Inform. Sci. **8**, 199–249 (1975)
14. J.M. Mendel, R.I. John, F. Liu, Interval type-2 fuzzy logic systems made simple. IEEE Trans. Fuzzy Syst. **14**(6) (2006)
15. J. Zeng, Z.Q. Liu, Type-2 fuzzy sets for handling uncertainty in pattern recognition, in *Proceeding of FUZZ-IEEE*, Vancouver, Canada (2006), pp. 6597–6602

16. C. Hwang, F. Chung-Hoon Rhee, Uncertain fuzzy clustering: interval type-2 fuzzy approach to C-means. IEEE Trans. Fuzzy Syst. **15**(1) (2007)
17. N.N. Karnik, J.M. Mendel, Applications of type-2 fuzzy logic systems to forecasting of time-series. Inf. Sci. **120**, 89–111 (1999)
18. A.D. Kulkarni, C.D. Cavanaugh, Fuzzy neural network models for classification. Appl. Intell. **12**, 207–215 (2000)
19. M.-J. Hsu, Y.-H. Chien, W.-Y. Wang, C.-C. Hsu, A convolutional fuzzy neural network architecture for object classification with small training database
20. T. Guo, J. Dong, H. Li, Y. Gao, Simple convolutional neural network on image classification, in *IEEE 2nd International Conference on Big Data Analysis (ICBDA)* (2017)
21. E. Cambria, B. White, Jumping NLP curves: A review of natural language processing research. IEEE Comput. Intell. Mag. **9**(2), 48–57 (2014)
22. W. Wang, J. Gang, Application of convolutional neural network in natural language processing, in *International Conference on Information Systems and Computer Aided Education (ICISCAE)* (2018)
23. Md. Amaan Haque, A. Verma, J. Sahaya Rani Alex, N. Venkatesan, Experimental evaluation of CNN architecture for speech recognition, in *First International Conference on Sustainable Technologies for Computational Intelligence* (2019), pp. 507–514
24. Z. Huang, M. Dong, Q. Mao, Y. Zhan, Speech recognition using CNN. IEEE/ACM Trans. Audio Speech Language Process. **22**(10), 1533–1545 (2014)
25. N.N. Karnik, J.M. Mendel, Q. Liang, Type-2 fuzzy logic systems. IEEE Trans. Fuzzy Syst. **7**, 643–658 (1999)
26. J.M. Mendel, R.I.B. John, Type-2 fuzzy sets. IEEE Trans. Fuzzy Syst. **7**(6) (1999)
27. J.M. Mendel, R.I.B. John, Type-2 fuzzy sets made simple. IEEE Trans. Fuzzy Syst. **10**, 117–127 (2002)
28. J.M. Mendel, Uncertainty, fuzzy logic, and signal processing. Signal Process. **80**, 913–933 (2000)
29. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in *ICLR* (2015)
30. T.-L. Nguyen, S. Kavuri, M. Lee, A fuzzy convolutional neural network for text sentiment analysis. J. Intell. Fuzzy Syst. **35**(6), 6025–6034 (2018)
31. S. Alizadeh, A. Fazel, Convolutional neural networks for facial expression recognition. Computer Vision Pattern Recogn. Cornell University Library (2017). ArXiv:1704.06756v1
32. T.-L. Nguyen, S. Kavuri, M. Lee, A multimodal convolutional neuro-fuzzy network for emotion understanding of movie clips. Neural Network **118**, 208–219 (2019)
33. S.-J. Lee, T. Chen, L. Yu, C.-H. Lai, Image classification based on the boost convolutional neural network, IEEE Access **6** (2018)
34. D.T. Mane, U.V. Kulkarni, A novel fuzzy convolutional neural network for recognition of handwritten Marathi numerals. Int. J. High Performance Comput. Network. **15**(3/4) (2019)
35. E.A. Popko, I.A. Weinstein, Fuzzy logic module of convolutional neural network for handwritten digits recognition. J. Phys. Conf. Ser. **738**(1) (2016)
36. J. An, L. Fu, M. Hu, W. Chen, J. Zhan, A novel fuzzy-based convolutional neural network method to traffic flow prediction with uncertain traffic accident information. IEEE Access **7** (2019)
37. U.A. Umoh, Interval type-2 fuzzy neural networks for short term electric load forecasting: a comparative study
38. Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard et al., Back-propagation applied to handwritten zip code recognition. Neural Comput. **1**(4), 541–551 (1989)
39. J.M. Keller, D.J. Hunt, Incorporating fuzzy membership function into the perceptron algorithm. IEEE Trans. Pattern Anal. Mach. Intell. **7**(6), 693–699 (1985)
40. S.R. Price, S.R. Price, D. T. Anderson, Introducing fuzzy layers for deep learning, in *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (2019)

41. A. Beke, T. Kumbasar, Interval type-2 fuzzy systems as deep neural network activation functions, in *Proceedings of the 11th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2019)* (2019)
42. V.V. Borisov, V.V. Kruglov, A.S. Fedulov, *Fuzzy Models and Nets* (Telekom Publication, Moscow, Goryachaya Liniya, 2012), 284p. (in Russian)
43. R. Fuler, Neural Fuzzy Systems (Publishing House Abo Akademi University, Abo, 1995)
44. Q. Liang, J.M. Mendel, Interval type-2 fuzzy logic systems: theory and design. IEEE Trans. Fuzzy Syst. **8**, 535–550 (2000)
45. S. Mitra, S.K. Pal, Fuzzy multi-layer perceptron, inferencing and rule generation. IEEE Trans. Neural Networks **6**, 51–63 (1995)
46. Y. Le Cun, L. Bottou, J. Huang Fu, Learning methods for generic object recognition with invariance to pose and lighting, in *CVPR* (2004)

Validating Retinal Color Fundus Databases and Methods for Diabetic Retinopathy Screening

S. Anitha and S. Madhusudhan

Abstract In recent days, there is an exponential increase in diabetic patient and increase in eye-related problems, the major eye problems boost due to diabetes, the diabetic retinopathy screening can be performed by image processing methods using fundus images which is an active research area, For this reason, a deep knowledge of various databases for diabetic retinopathy, various eye diseases characteristics, fundus image processing methods and analysis are necessary. This article provides an overview of twelve retinal fundus image datasets for diabetic retinopathy and its validation and presents a collective discussion on twelve recent diabetic retinopathy fundus image processing methods and different stages of diabetic disease. This article will provide insight into databases classified with several images, image quality, camera involved to capture images, image resolution, and field of view.

Keywords Diabetic-Retinopathy (DR) · Age-Related-Macular-Degeneration (AMD) · Field of view (FOV) · Micro-aneurysms (MA) · Optical nerve head (ONH) · Data set (DS) · Image processing

1 Introduction

The eye plays important role in almost all human activities. The eyes are an important organ of the human visual system. The human eye suffers from Several diseases from infant to old age, as age increases these diseases become very common and also difficult to diagnose, further it requires more analysis and resources to diagnose. The survey conducted by the ministry of union health in India national diabetes and diabetic retinopathy survey reveals that there is about 16.9% of Non-proliferative [1] and 3.6% of Proliferative retinopathy [1, 2] active cases and increasing day by day. The survey also reveals that Mild Non-proliferative Retinopathy is about 11.8%.

S. Anitha
Department of Bio-Medical Engineering, ACS College of Engineering, Bengaluru, India

S. Madhusudhan (✉)
Department of Electronics and Communication Engineering, AIeMS, Bidadi, Bengaluru, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_58

Moderate Non-proliferative Retinopathy is 8.0%. The world health organization estimates that global diabetes among 18-year adults in 2014 is about 8.5% and in India, it is around 72.96 million cases. According to the literature survey, people with diabetes will suffer more eye-related problems than non-diabetic people. The blood vessels, veins and other sensitive parts of the eye will get damage due to diabetes, which in turn damages the retina which consists of cones, rods and also damages tissue which supplies nutrients to the eye and tissues which are responsible for clear vision. The diabetic patient will suffer problems with blood vessels [3–5] in the entire body and also the eye. The blood vessels in the back of the eye i.e. retina will get affected, and both eyes may get affected due to diabetics, it will be easier to diagnose if it is treated in the initial stage. These eye disease analyses of diabetics are called Diabetic Retinopathy [1] and the processing of the fundus images is called fundus image processing [1, 3, 6–8]. Modern technology applies new techniques and methods to diagnose and a fundus image database is formed which contains information about the diabetic retinopathy images [3]. The database is called a fundus database, which is a collection of fundus images. The fundus images are the output of the fundus camera, this camera is specially designed to acquire the front side as well as the rear side of the eyeball, thus fundus camera provides complete eyeball image and the captured images are of different qualities which depend on the quality of the camera. These captured images of different individuals with healthy as well as diseased or infected images are formed in a single dataset.

Validation of these database gives information about the number of fundus images in the particular database, different sets of the image depending on the diabetic and non-diabetic retinopathy. The dataset is validated for the number of images, camera employed to capture the images, the resolution considered, the field of view [4], and image mask [9, 10] used. This information of different data sets provides insight knowledge to a researcher or a medical specialist to disorganize and for comparison of novel methods with the existing method. Validation of fundus data sets helpful in selecting particular data set and image with required resolution and parameters. Since the ratio of several eye specialties to the number of patients is very low, the patients from a remote location can be diagnosed by considering images of the patient, which sent to an eye hospital or eye specialist [11] and compared with the data set image which is a type Telemedicine [12]. Telemedicine is a real-time application that uses a fundus image dataset. For a researcher, it will help apply image processing methods like segmentation, disc extraction, vessel segmentation, exudates detection, fovea detection, and other image processing technique on images in the dataset.

The DR-diabetic retinopathy analysis is classified as non-proliferative and non-proliferative types, similarly, fundus images are also classified for micro-aneurysms, macular edema [1, 6], or macular ischemic, ischemia, neo-vascularisation, and vitreous hemorrhage. The application of image processing techniques helps in diagnosing these decreases. Further, the selection of a database plays an important role in DR analysis. The database like Messidor database: Messidor-2, REVIEW and DRIVE provide more number of images and images with various DR diseases which gives more options while selecting images. Selection of images with the required problem is easier while considering a larger database.

2 Classification of Diabetic Retinopathy Stages

2.1 *Non-proliferative Diabetic Retinopathy*

2.1.1 Mild Non-proliferative Retinopathy

It is considered as the initial stage of diabetic retinopathy, in this stage, some small balloon-like swellings will appear in the blood vessels or microscopic blood-filled bulges in the artery walls called micro-aneurysms [1, 2].

2.1.2 Moderate Non-proliferative Retinopathy

In this stage, the blood vessels called choroid [1] which supplies nutrition to the retina and sclera gets blocked and a significant increase in the number of micro-aneurysms spots and formation of hard exudates [1] due to accumulations of fluid that has leaked from blood vessels, macular edema [1, 6] or macular ischemia [1, 6] may cause rapid vision loss.

2.1.3 Severe Non-proliferative Retinopathy

The number of new blood vessels will increase due to blockage of a considerable number of blood vessels in the retina and signals sent by the retina to the brain to compensate for the loss of blood supply and nourishment to the retinal called ischemia and trigger progression to sight-threatening proliferative disease [1].

2.2 *Proliferative Retinopathy*

The growth of new blood vessels increases in the retina. These vessels may blast and bleed which in turn affects the center part of the eye contains clear jelly or fluid called vitreous. The proliferative stage is characterized by neo-vascularisation [1, 2] also called vitreous hemorrhage [1] and this vitreous humor may shrink and lead to a detached retina and lead to distorted vision.

The database should include the fundus images of non-proliferative and proliferative stages and its disease type, some analysis may require not only fundus images but also considering images of eyelashes and other parts of the eye. This can be achieved by considering digital images captured by a digital camera (Table 1).

Table 1 Comparison of Non-proliferative and proliferative DR [1–10]

Sl. No.	DR type	Related diseases	Severity	Stage	Characteristics	Effect
		Micro-aneurysms	Mild	Initial stage of diabetic retinopathy	Small balloon-like swellings	Blurring vision
1	Non-proliferative	Hard exudates and Macular edema	Moderate	Increase in number of micro-aneurysms spots	Accumulations of fluid that has leaked from blood vessels	Rapid vision loss
		Macular ischemia	Sever	The number of new blood vessels will increase	Blockage of a considerable number of blood vessels in the retina	Sight-threatening
2	Proliferative	Neovascularisation and Vitreous hemorrhage	Sever	Detached retina	Effects center part of the eye contains clear jelly or fluid called vitreous	Distorted vision

3 Details of DR Databases

The Diabetic Retinopathy (DR) database is classified according to the number of fundus images available like healthy images, images of diabetic background and its stages, image quality, resolution, camera employed, FOV and grouping of images for different image processing techniques.

3.1 ARIA Database

The ARIA is an automatic retinal image analysis database, it is a color fundus image database formulated by 143 images with a size of 768 × 576 picture elements, these images are organized into three types: AMD [3, 9] of 23 subjects ($n = 23$), control-group with healthy 61 subjects ($n = 61$), and subjects with the diabetic background ($n = 59$). [1] further, these classified groups are divided by applying vessel segmentation algorithms, Vessel Classes, Vessel GUI, Vessel Library, Vessel Processors and presented in zip format as aria a markup vessel, aria a markups, aria c markup vessel, aria c markup disc fovea [3], aria c markups, aria d markup vessel, aria d markup disc fovea [9], aria d markups. All images are having a TIFF format with ground truth which is used for the segmentation of vessels [3–5], and the images are obtained using the fundus camera Zeiss FF450 + with a 50-deg field of view (Figs. 1, 2 and 3).

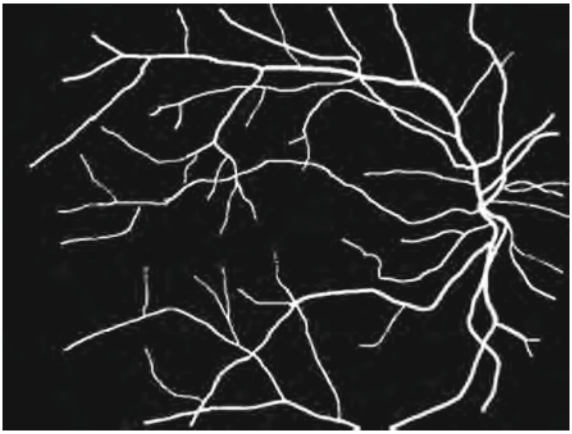


Fig. 1 Vessel mark up [1, 2]

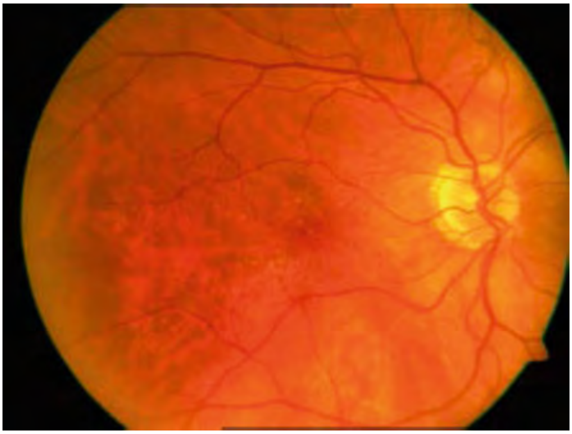


Fig. 2 Fundus [1–3]

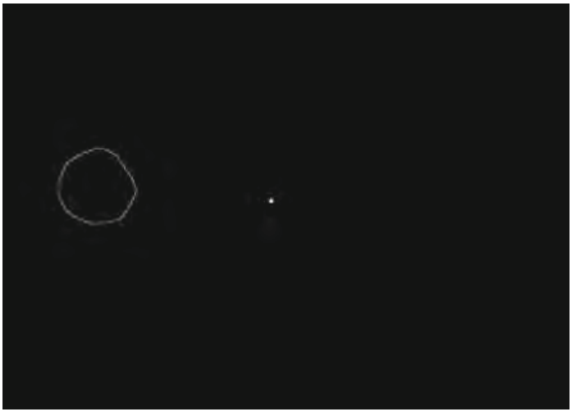


Fig. 3 Markup disc fovea [1–3]

3.2 DRiDB: Diabetic Retinopathy Image Dataset

DR Dataset consisting of a total of 216 images and the automated fundus imaging requires automatic assessment [4] of image quality for analysis and to obtain accurate results for retinal diseases [4, 10]. This article uses good images, bad images, and outlier image classes.

3.3 DRIONS DB

The Dataset is of Vascular Networks taken from Retinal photographs of Hypertensive retinopathy subjects and Glaucoma Patients [5], the database is of the high resolution of blood vessels taken from a database called optic nerve segmentation db (DRIONS-DB). The database is of 110 blood vessels [5, 7] images in which 23% are patients with long-standing glaucoma also remaining 77% are hypertensive retinopathy patients.

3.4 DRIVE: Digital Retinal Images for Vessel Extraction Database

The DRIVE database basically designed for segmenting retinal blood vines and marking of structural attributes, like shape, width, position, length [2], patterns, screening angle, and analysis of different heart and eye diseases [2, 11], data set of screening population consists of 400 diabetic subjects of age group range 25 to 90 years. The Canon-CR5 series non-mydriatic 3CCD camera 45-degree FOV circular with a circle diameter of 540 pixels is applied to take images with 768×584 pixels resolution. The training set is of 40 images for Segmentation of vasculature testing [2].

3.5 E-Optha

E-ophtha data set have two subsets called e-ophtha-EX for Exudates [2] and e-ophtha-MA [2] for Micro Aneurysms, E-ophtha-Exudate [13] is used when images having exudates. DS is of 47 fundii images of exudates, 35 fundii images of no lesion traces. E-ophtha Micro aneurysms [8, 13] are the DS with fundus images of MA-micro-aneurysms. The DS is of 148 fundii images of microaneurysms or also called as small or less hemorrhage or damage and 233 fundii images having non-lesion [13] (Figs. 4, 5 and 6).

3.6 HEI-MED

The full form is "Hamilton eye institute macular-edema database" it's a combination of 169 photographs for training as well as testing of flow methods for exudates identification [14] and also for DR macular-edema [14]. 169 JPEG format are of high quality and compressed, mainly used to analyze exudates, bright lesions, cotton wool spots, and fundus drusens [6, 14, 15].

Fig. 4 Healthy image [3–5]

Fig. 5 Image exudates [3–5]

Fig. 6 Macula edema [5]

3.7 HRF-High-Resolution Fundus

HRF-High-Resolution Fundus supports study on automated segmentation. Dataset is of 15 photographs under a healthy base, 15 pics of DR subjects also 15 data for glaucoma [16]. localize the macula, OD, and analysis between arteries and veins (A/V) [16] are made available for the images mentioned with masks indicating FOV [5, 16] are also made available, the input data contains the Field of View (FOV) masks images, the Vessel Segmentation Gold standard images, the Optic Disk Gold standard [16, 17] containing Center Point and Radius of the retina the primary Dataset consists of 18 pairs of the particular retina collected from 18 individuals taken by

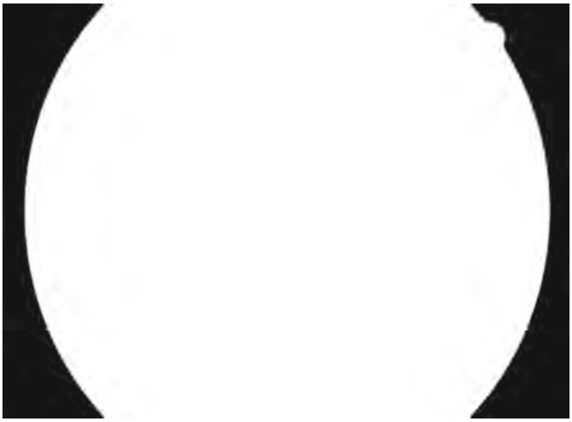


Fig. 7 Image mask [5, 6]

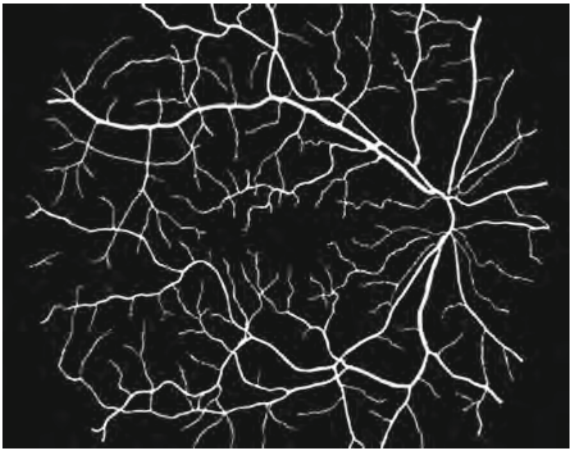


Fig. 8 Vessel extraction [5, 6]

fundus photograph cam series Canon CR-1 45° [16] and various setting (Figs. 7 and 8).

3.8 Imageret Database

DIARETDB0 is a Standard D-R-D-B of Calibration level 0, of 130 color fundii pictures, 20 are non-chronic [18] and 110 of traces of DR like microaneurysms, hemorrhages, hard-exudates, soft-exudates and neo-vascularization [13, 18, 19]. 50-degree field-of-view is applied (Figs. 9 and 10).

Fig. 9 Fundus image [5–10]

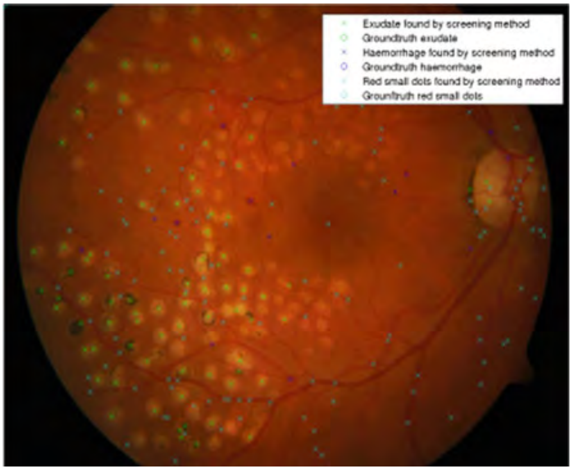


Fig. 10 Ground truth and binary mark [5–10]

3.9 *DIARETDB*

DIARETDB is a Standard D-R Dataset of eighty nine-89 color pictures in this 84 of at least Lenient non-proliferative indications [6] also called Micro aneurysms particular DR and 5 of non-DR indications. Soft exudates, red small dots, hemorrhages, hard exudates [1, 4–6, 11] (Figs. 11 and 12).

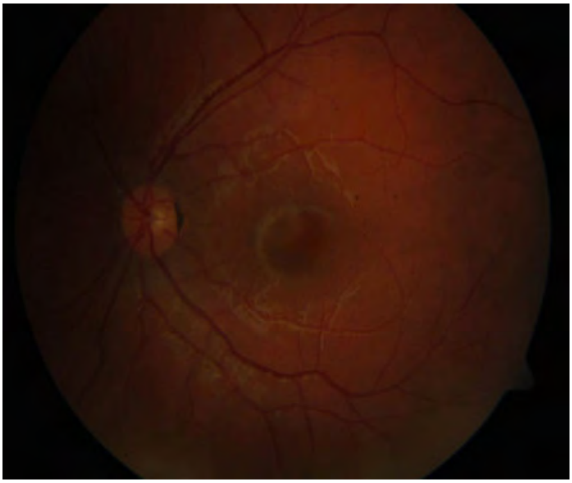


Fig. 11 Fundus image [4–11]

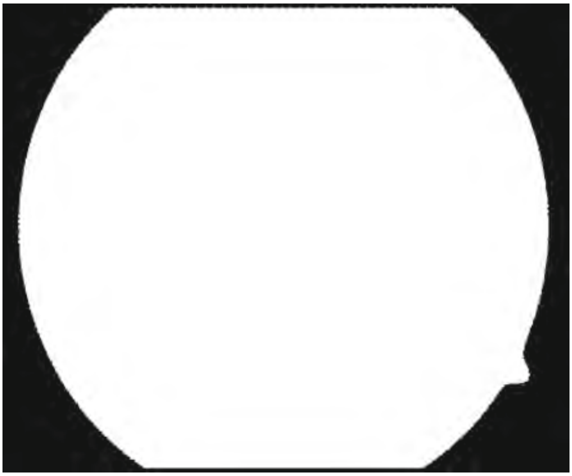


Fig. 12 Mask applied [11]

3.10 Messidor Database: Messidor-2 Dataset

The database consists of 2 macula-centered [20] eye fundus images 1 for each eye. The Messidor-Original data set contains all image pairs [20] of images, which contains 529 tests and 1058-PNG images. Here no pharmacological dilation is applied; Top-con TRC NW6 non-mydriatic-fundus-camera is used [20] with a 45^0 fov. The Messidor-Extension data set contains 345 tests which include 690 JPG-photographs; M-2 DS is of 874 tests with 1748 pictures [20, 21].

3.11 REVIEW

The full form is “Retinal Vessel Image set for Estimation of Widths”, this set contains subsets which includes JRD–Junction Resolution dataset based on the DRIVE database [20], MMRBF–Manual Measurement of Retinal Bifurcations Features [22], ONHSEG–Optic Nerve Head Segmentation database [22], LEDGE–Lesion (edge) segmentation database [20], The original color photos for the four sets, and also the edge points of 3 observers for 4 sets. MMRBF [20]–Manual Measurement of Retinal Bifurcation Features, ONHSD [20]–99 images taken from 50 patients for diabetic retinopathy screening [1, 22], 96 images have discernible ONH. The fundus is recorded using Canon camera series CR6 45MNf [22], with a field angle lens of 45^0, 640 $\times$ 480pels. Finally transformed to grey-scale images by getting HSI component [22, 23] (Figs. 13, 14, 15 and 16).

Fig. 13 Vessel in disc [20–22]

Fig. 14 Vessel extraction [22]

Fig. 15 Exudates [22]

Fig. 16 Optical nerve head [22]

3.12 DR HAGIS

DR HAGIS—automatic extraction [5, 6] of FI–fundii image or vessels in eye retina or on the surface from DR subjects, this database contains 39 HD color fundus images

[9]. 10 images of glaucoma, 10 images of hypertension, 10 images of DR, and 10 images of AMD are arranged in order. The data set is taken from Canon CR DGi, Top-con TRC-NW6s series cam and Top-con TRC-NW8 cam [9, 24] along with parallel FOV of 45-deg. The photographs are with particular resolution according to cam selected are of 4752 × 3168 pels, 3456 × 2304 picture element, 3216 × 2136 pixels, 2896 × 1944 cell, 2816 × 1880 p [9] (Figs. 17, 18 and 19).

The selection of database plays a major role in retinal fundus image processing. Most of the database which is publicly available are open source, means it's of free cost and downloadable. And some of the database will have access to only registered users and some database will charge to download. The database contains a fundus image of the retina in two forms, it may be of grayscale [13] or color fundus images. Many manufacture supplies fundus camera with various resolution and characteristics. Fundus generally refers to the complete picture of the eyeball, which includes the front position as well as the rear position of the eyeball. The fundus

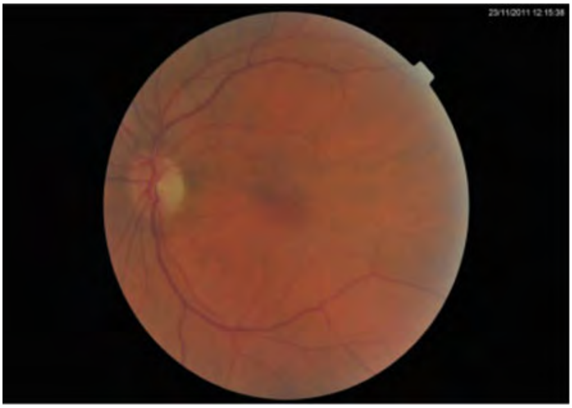


Fig. 17 Fundus image [24]

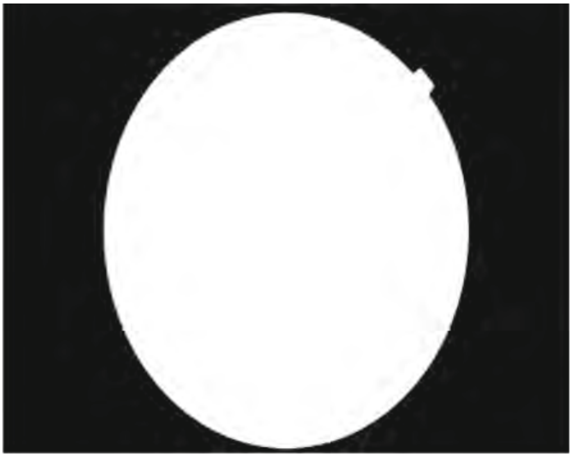


Fig. 18 Mask generated [24]

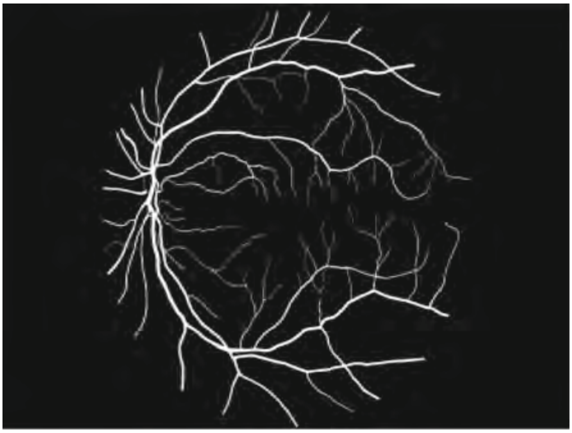


Fig. 19 Manual segmented image [24]

is captured with pre-defined angle called FOV-field of view [1, 2], once the image is captured it will be stored in digital form in the computer memory. All databases will have separate categories for different types of fundus images. Once the image is found the next step is to check for the required resolution, some database will contain high-resolution images and some will have low-resolution images. According to the requirement, these images are selected and processed (Table 2).

4 Recent Methods in DR Analysis

See Table 3.

5 Discussion and Summary

Table 3 provides a comparison of various recent methods applied for diabetic retinopathy screening. MA- Micro Aneurysms and dot HM which come under non-proliferative DR are processed using Pixel clustering, KNN, Radon transform, wavelet preprocessing, Wavelet, genetic algorithm, Wavelet template matching, 2-D adaptive filtering, region growing, Pixel classification, Morphological, region growing and Manual rule-based classifier. Exudates are processed by Machine learning Dynamic clustering with domain knowledge The sensitivity and specificity are the two major parameters considered for comparison. In some methods, accuracy is also considered. The widely applied parameters are: mean Truly-Positive-Rate (TPR), mediocre Falsely Positive Rate (FPR), moderate-Sensitivity, mean-Specificity, mean-Accuracy, and exactness [12, 21, 23–32]. Sensitive and specific are the most commonly used grades in ophthalmology [12, 21, 23–32]. Higher the

Table 2 Details of D-R database [1–6, 13, 14, 16, 18, 20, 22]

Sl. No.	Name of database	Camera applied	Total images under consideration	Image resolution	DR and its diseases
1	ARIA database [1]	Zeiss-series-FF450 + fundus camera with a 50-deg fov	143 color fundus images	768 × 576 pixels	ARMD, diabetic-Retinopathi
2	DRiDB: Diabetic retinopathi image dataset [3]	45^0 fov camera arbitary	216 images	High resolution	Diabetic Retinopathi
3	DRIONS DB [4]	45^0 fov Canon-CR5-non-mydriatic-3CCD camera	110 blood vessel images	Multi-resolution images	Diabetic-retinopathi
4	DRIVE-digital retinal images for vessel extraction data set [5]	Canon CR5-non-mydriatic 3CCD camera with a 45^0 (FOV)	40 images	768 × 584 pels	Diabetic, hyper-tension, arterio -sclerosis and chorodial-neo-vascularization
5	E-optha [2]	45^0 fov camera arbitary	47 exudates image and 35 normal and 148 micro aneurysms and 233 fundus images with no lesion traces	High resolution	Micro Aneurysms and Exudates
6	The H-Hamilton E-Eye I-Institute M-Macular E-Edema D-Dataset HEI-MED [13]	50^0 fov camera arbitary	169 images	Highest quality and compressed	Exudates and also for diabetic macular edema.

(continued)

Table 2 (continued)

Sl. No.	Name of database	Camera applied	Total images under consideration	Image resolution	DR and its diseases
7	HRF-Image database [14]	CR-1 Canon fundus imaging camera with 45°	15 from healthy subjects, 15 diabetic retinopathy and 15 images of normal subject	Multi-resolution images	Diabetic, glaucoma
8	Imageret database DIARETDB0 [16]	50^0 fov camera arbitary	130 color fundus images of which 20 are normal and 110 of traces of diabetic retinopathy	High-resolution	Hard exudates, soft exudates, micronaneuyrysms, hemorrhages and neovascularization.
9	DIARETDB1 [18]	50^0 fov camera arbitary	89 color fundus photographs	High-resolution	Micro aneurysms particular for diabetic retinopathy
10	Messidor database: M-2 [6]	Top con TRC NW6 non-mydriatic with a 45^0 fov	Messidor-2 is of 874 checks with totally 1748 images.	Multi-resolution images	Diabetic Retinopathy (DR)
11	REVIEW Database [20]	Canon CR6 45MNf, with 45^0	99 images from fifty subjects, 96 images have ONH	640 × 480	Lesion, Optic Nerve Head for diabetic retinopathy screening
12	DR HAGIS [22]	Canon CR DGi, Topcon TRC-NW6s and Topcon TRC-NW8. With 45 deg	39 HD color fundus images	4752 × 3168 pixels, 3456 × 2304 pixels, 3216 × 2136 pixels, 2896 × 1944 pels,2816 × 1880	10 images of glaucoma, 10 images of hypertension, 10 images of DR, and 10 images of AMD

Table 3 Comparison of various methods for DR diseases with aria database [12, 21, 23–32]

Sl. No.	Author	Method	DR type	Number of images	Sensitivity in %	Specificity in %	Merits of method	Limitations of method
1	Abramoff MD Reinhardt JM Russell SR et al. [25]	Pixel clustering, kNN	Micro Aneurysms, dot HM	16,670	47.7	90	Generalized to any fundus image, High accuracy and simple	Sensitivity to noise, huge storage needed
2	Giancardo L Meriaudeau F Karnowski et al. [21]	Radon transform, wavelet preprocessing	MA -Micro Aneurysms	100	50	>10 false positive per image	Line parameter extraction, signogram	Blur image
3	Quellec G Lamard M Josselin et al. [26]	Wavelet, genetic algorithm	MA- Micro Aneurysms	1115	90.24	89.75	Flexible and applied for non stationary data	Computationally intensive
4	Niemeijer M van Ginneken B Russell et al. [23]	Machine learning	Exudates	430	95	88	Multi dimensional and multi variant	Pre-processing needed and huge memory
5	Quellec G Lamard M Josselin PM et al. [24]	Wavelet template matching	MA- Micro Aneurysms	995	87.90	96.20	Accurate and simple De-noising	Content of image degrades
6	Serrano C Acha B Revuelto S et al. [27]	2-D adaptive filtering, region growing	MA- Micro Aneurysms	11	90.72	82.35	Choice of similarity and adaptive	Less efficient and slow process

(continued)

Table 3 (continued)

Sl. No.	Author	Method	DR type	Number of images	Sensitivity in %	Specificity in %	Merits of method	Limitations of method
7	Niemeijer M van Ginneken B et al. [28]	Pixel classification	Micro Aneurysms, dot HM	240	100	87	High sensitive and compression	Large memory space required
8	Hsu W Pallawala PMDS Lee ML Kah-Guan AE et al. [29]	Dynamic clustering with domain knowledge	Exudates	543	100	74	High accuracy and simple algorithm	Prior data knowledge required for pre processing
9	Yang G Gagnon L Wang S Boucher MC et al. [30]	Morphological, region growing	MA- Micro Aneurysms	46	90	80	Flexible, systematic alteration	Pre processing needed and slow
10	Wang H Hsu W Goh KG Lee ML. et al. [12]	Statistical classification with local window-based verification	Exudates	200	100	70	High accuracy and high sensitive	Sensitivity to noise, huge storage needed
11	Hipwell JH Strachan F Olson JA et al. [31]	Manual rule-based classifier	MA- Micro Aneurysms	3885	85	76	Accurate and simple De-noising	Ground truth is required,
12	Gardner GG Keating D Williamson TH Elliott AT et al. [32]	Neural network	HM, exudates	480	73.80	73.80	Simple and efficient	More memory required

values better the method and analysis. The sensitive parameter gives the ability to perceive eye vessels element; the specific parameters provides the extraction of no-vessel area. Specificity [12, 21, 23–32] reflects the headlining of breakthrough which is related to accurate value in fundus image processing.

Sensitivity = TPTP + FN

Specificity = TNTN + FP

Accuracy = TP + TNTP + FN + TN + FP

where, TP: True Positives: Correctly classified vessel pixels.

FP, False Positives: Wrongly classified non-vessel pixels as vessel pixels.

TN: True Negatives: non-vessel pixels that are correctly classified.

FN: False Negatives: vessel pixels which are wrongly classified as non-vessel pixels.

Sensitivity: Ratio of correctly classified vessel pixels and the total number of vessel pixels.

Specificity: Ratio of correctly classified non-vessel pixels and the total number of non-vessel pixels.

Accuracy: Extent to which algorithm works correctly.

6 Conclusion

The various diabetic retinopathy fundus image database are analyzed with various DR phases images like Non-proliferative Retinopathy images Age-Related-Macular-Degeneration (AMD), Micro-aneurysms (MA), Optical nerve head (ONH) and other Severe proliferative Retinopathy images. The database presented provides fundus images of high quality and resolution which are helpful in image processing application and analysis. It has been examined that the database with more fundus images are necessary and provides more information. The database like Messidor, REVIEW and DRIVE provides 1750, 250 and 400 images respectively. Further, the DR diseases are screened according to the method or the algorithm for Sensitivity, Specificity. It is seen that the wavelet and template matching method gives 90.24% and 89.75% and the wavelet with genetic algorithm gives 87.90% and 96.20% of Sensitivity and Specificity respectively as compared to other methods. The merits and demerits of various image processing methods are also discussed.

References

1. Guidelines for the management of diabetic retinopathy [online]. Canberra, ACT: National Health and Medical Research Council, 2008 [cited 28 February 2011]. Available from: URL link Spurling G, Askew D, Jackson C. Retinopathy: Screening recommendations. *Aust Fam Physician.* **38**(10), 780–783 (2009)
2. Z. Yavuz, C.K. Karadeniz, Blood vessel extraction in color retinal fundus images with enhancement filtering and unsupervised classification. https://drive.grand-challenge.org/

3. Y. Zheng, M.H.A. Hijazi, Frans Coenen automated "Disease/no disease" Grading of age-related macular degeneration by an image mining approach. Investigative Ophthalmol. Visual Sci., December 2012, 53, 8310–8318. https://doi.org/10.1167/iovs.12-9576
4. P. Prentašić, S. Lončarić, Z. Vatavuk, G. Benčić, M. Subašić, T. Petković, L. Dujmović, M. Malenica-Ravlić, N. Budimlija, R. Tadić, Diabetic retinopathy image database (DRiDB): a new database for diabetic retinopathy screening programs research, in *Proceedings of 8th International Symposium on Image and Signal Processing and Analysis (ISPA 2013)* (Trieste, 2013), pp. 704–709
5. A.N. Oluwatobi, A.O. Christianah, A. Joyce, M. Adebiyi. I.C. Obagbuwa, A. Peace, O.R. Oluwaseun, O.A. Elisha. https://zenodo.org/record/1410497#.Xql70WgzbIU
6. T. Kauppi, V. Kalesnykiene, J.K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kälviäinen, J. Pietilä, DIARETDB1 diabetic retinopathy database and evaluation protocol, Technical report (PDF)
7. A. Hoover, et al., locating blood vessels in retinal images by piece-wise threhsold probing of a matched filter response, in IEEE Trans. Med. Imaging **19**(3), 203–210 (March 2000)
8. P. Cabaleiro, et al., Automatic identification and representation of the cornea–contact lens relationship using as-oct images. Sensors **19**, 5087 (2019)
9. Jayanthi Sivaswamyy S.R. Krishnadas, Gopal Dutt Joshiy Madhulika Jainy Ujjwaly Syed Tabish A.y y IIIT, Hyderabad, India Arvind Eye Hospital, Madurai, India
10. A. Budai, R. Bock, A. Maier, J. Hornegger, G. Michelson, Robust vessel segmentation in fundus images. Int. J. Biomed. Imaging, vol. 2013 (2013)
11. M. Ortega, N. Barreira, J. Novo et al., SIRIUS: a web-based system for retinal image analysis. Int. J. MedicalInform. **79**(10), 722–732 (2010)
12. H. Wang, W. Hsu, K.G. Goh, M.L. Lee, An effective approach to detect lesions in color retinal images, in Proceedings from the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Hilton Head, SC, 2000), pp. 2181–2186
13. P. Khojasteh, B. Aliahmad, D.K. Kumar, Fundus images analysis using deep features for detection of exudates, hemorrhages and microaneurysms. BMC Ophthalmol. **18**, 288 (2018). Published online 2018 November 6. https://doi.org/10.1186/s12886-018-0954-4, http://www.adcis.net/en/third-party/e-ophtha/
14. L. Giancardo, F. Meriaudeau, T.P. Karnowski, Y. Li, S. Garg, K.W. Tobin, E. Chaum, Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. Med. Image Anal. **16**(1), 216–226 (2012). https://github.com/lgiancaUTH/HEI-MED
15. J. Sivaswamyy, S.R. Krishnadas, G.D. Joshi, M. Jain, A.U.S. Tabish, DRISHTI-GS: retinal image data set for optic nerve head–ONH segmentation, IIIT, Hyderabad, India, Arvind Eye Hospital, Madurai, India
16. T. Köhler, A. Budai, et.al., Automatic no-reference quality assessment for retinal fundus images using vessel segmentation, in 26th IEEE Internatioal Symposium on Computer-Based Medical Systems 2013, Porto
17. J. Zhang, B. Dashtbozorg, et.al., Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores. IEEE Trans. Med. Imaging **35**(12), 2631–2644 (2016)
18. T. Kauppi, V. Kalesnykiene, J.K. Kamarainen, L. Lensu, I. Sorri, H. Uusitalo, H. Kälviäinen, J. Pietilä, DIARETDB0: Evaluation Database and Methodology for Diabetic Retinopathy Algorithms, Technical report [10]. http://www.adcis.net/en/third-party/messidor/
19. Q. Hu, M.D. Abràmoff, M.K. Garvin, Automated separation of binary overlapping trees in low-contrast color retinal images. Med Image Comput. Comput. Assist. Interv. **16**(Pt 2), 436–443 (2013). PubMed PMID: 24579170
20. R. Murugan, An automatic detection of hemorrhages in retinal fundus images by motion pattern generation. Biomed. Pharmacol. J. 12(3) (2019). http://biomedpharmajournal.org/?p=23708, http://www.adcis.net/en/third-party/messidor/
21. L. Giancardo, F. Meriaudeau, T.P. Karnowski, Y. Li, K.W. Tobin Jr., E. Chaum, Microaneurysm detection with Radon transform-based classification on retina images. Conf. Proc. IEEE Eng. Med. Biol. Soc. **2011**, 5939–5942 (2011)

22. K. Acharjya, G.S. Sahoo, S.Kr. Sharma, An extensive review on various fundus databases use for development of computer-aided diabetic retinopathy screening tool. http://cvit.iiit.ac.in/projects/mip/drishti-gs/mip-dataset2/Home.php
23. M. Niemeijer, B. van Ginneken, S.R. Russell, M.S.A. Suttorp-Schulten, M.D. Abramoff, Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. Invest. Ophthalmol. Vis. Sci. **48**, 2260–2267 (2007)
24. G. Quellec, M. Lamard, P.M Josselin, Detection of lesions in retina photographs based on the wavelet transform, in Proceedings from the IEEE EMBC International Conference on Engineering in Medicine and Biology Society (New York, 2006)
25. M.D. Abramoff, J.M. Reinhardt, S.R. Russell, Automated early detection of diabetic retinopathy. Ophthalmology **117**, 1147–1154 (2010)
26. G. Quellec, M. Lamard, P.M. Josselin, G. Cazuguel, B. Cochener, C. Roux, Optimal wavelet transform for the detection of microaneurysms in retina photographs. IEEE Trans. Med. Imaging **27**, 1230–1241 (2008)
27. C. Serrano, B. Acha, S. Revuelto, 2D adaptive filtering and region growing algorithm for the detection of microaneurysms, in SPIE Med. Imaging: Image Process. **5370**: 1924–1931 (2004)
28. M. Niemeijer, B. van Ginneken, J. Staal, M.S.A. Suttorp-Schulten, M.D. Abramoff, Automatic detection of red lesions in digital color fundus photographs. IEEE Trans. Med. Imaging **24**, 584–592 (2005)
29. W. Hsu, P.M.D.S. Pallawala, M.L. Lee, A.E. Kah-Guan, The role of domain knowledge in the detection of retinal hard exudates, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 246–251 (2001)
30. G. Yang, L. Gagnon, S. Wang, M.C. Boucher, Algorithm for detecting micro-aneurysms in low-resolution color retinal images, in Proceedings of Vision Interface, pp. 265–271 (2001)
31. J.H. Hipwell, F. Strachan, J.A. Olson, Automated detection of microaneurysms in digital red-free photographs: a diabetic retinopathy screening tool. Diabetic Med. **17**, 588–594 (2000)
32. G.G. Gardner, D. Keating, T.H. Williamson, A.T. Elliott, Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. Br. J. Ophthalmol. **80**, 940 (1996)

An Investigation for Interpreting the Epidemiological Occurrence of COVID-19 in India Using GP-ARIMA

K. M. Baalamurugan, Tanya Yaqub, Akshat Shukla, and Akshita

Abstract There is awareness towards the coronavirus 2019 (2019-nCoV) as a worldwide public health threat. COVID-19 is an outbreak caused due to severe acute respiratory syndrome coronavirus 2(SARS-CoV-2) which initiates from china and rapidly moves to epidemiological data which is required for intervention and awareness strategies. The current outbreak for this disease commences from Wuhan, China caused by 2(SARS-CoV-2) which emphasizes the analysis for epidemiological data for this virus and predicts the risk for infected people over the entire world. The enormous amount of cases are generally increasing day-by-day which is flattened out. In this study, Tamil Nadu state data is considered, Genetic Programming based ARIMA model (GP-ARIMA) test is carried out to determine the relationship among the gender (female and male), age group (>18–40 and 41–65 or greater than that) where the current status is deceased, hospitalized, and recovered with spreading mode linkage. This study analyzes gender independently for the various current situations. Based on the age, the current status, age group, and gender relation are dependent; however, linkage problems are extremely significant over spreading ratio.

Keywords Epidemiological data · SARS-CoV-2 · Lethality · Galton · GP-ARIMA

1 Introduction

SARS-COVID-19 is contamination from the Nido virus that comprises of Roiniviridae, Coronaviridae, and Artieviridae which is responsible for respiratory sickness over a human community that origins from cold to more serious disease like Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS) [1]. The common symptoms of COVID-19 are fever, cold, dry cough, nasal blocking, tiredness and pain, tender throat, and running nose. When the person gets infected and does not get any symptoms and feel unwell. The people of all age groups with medical history like blood pressure, cardiovascular disease, diabetes

K. M. Baalamurugan (✉) · T. Yaqub · A. Shukla · Akshita
School of Computing Science and Engineering, Galgotias University, Uttar Pradesh, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_59

are prone to infection and found to have fever, cough, breathing issues needs to seek immediate medical attention. It is a transmissible disease and passes over droplets from the mouth and nose when infected people respire or cough. The individuals need to maintain a distance of 1 m from the affected person. Various investigations related to COVID-19 is spread through transmission from the air.

The same number of individuals just experienced gentle side effects so it is a high likelihood to get COVID-19 from an individual who doesn't feel sick. Assurance from and anticipation of COVID-19 spreading is limited and simple to receive safety measures [2] in the day by day propensities which incorporate altogether cleaning hands with hand rub liquor or washing them with water and cleanser, try not to contact nose, eyes and mouth as hands contacts surfaces and hands are transporter for COVID-19 and infected over the body, remain at home on the off chance that you feel unwell and in particular abstain from going however much as could reasonably be expected. Follow National and neighborhood specialists just as they have forward-thinking data about the circumstance. India revealed the first Covid case in Kerala accepted back from Wuhan (the focal point of Covid) and till then the number of cases has been expanding dramatically [3].

Recently, there is no immunization or medication accessible especially for COVID-19 treatment under scrutiny. This work breaks down the COVID-19 pattern dependent on standard utilizing Exploratory Data Analysis.GP-ARIMA is the best approach to investigate information by extricating valuable data. It is a brilliant advance in any sort of examination.

2 Data Analysis

In data analysis, a COVID-19 data report is considered from Tamil Nadu, India. From this analysis, the dataset is downloaded from Kaggle. Here, a list of COVID-19 cases of all states is provided, however, most of the attribute values are missing. From this analysis, three attributes are concentrated with age, gender, and present state of the patient (deceased, hospitalized, and recovered). This dataset comprises of various missing values, and analysis is applied directly to the dataset. However, it does not provide appropriate outcomes.

Henceforth, to carry out data pre-processing, initial validation is performed for missing values based on the state and validation. There are some missing values for certain time intervals. Based on these conditions, the Tamil Nadu region is considered. When the target data is chosen, the data is thoroughly analyzed for captivating the research problems.

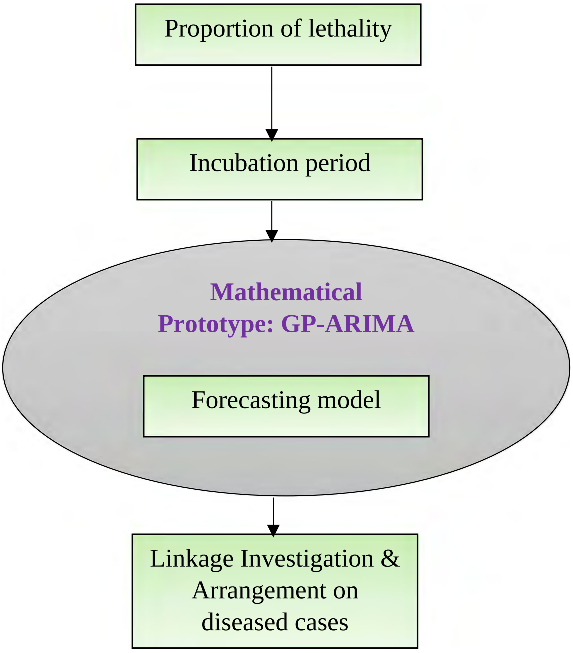
2.1 Dataset

Here, the Kaggle dataset is considered with a sum of 17 attributes in the dataset. The attributes are strings except for age, however the SPSS model cannot carry out any kind of analysis on string data type. Therefore, values of transmission type, gender, and status are substituted by nominal data list.

3 Background Preliminaries for Data Interpretation

The COVID-19 outbreak stimulates data analysis over available datasets, scraped from different sources like "Ministry of Health and Family Welfare", "COVID-19 India website", "Kaggle", "Worldometer" and "Wikipedia" by "Python" [4] and analysis the spread and COVID-19 trends over India. The comparison is done with the analysis from neighborhood countries all over the work. Here, the Kaggle dataset is used for the normalization process for selecting appropriate columns for filtering, column derivation, and data visualization over graphical format. This work considers MATLAB for web scrapping and pre-processing. Some libraries are included for processing and extraction of information from the available dataset. The graphs are generated for the finest visualization using MATLAB tools. The flow of the survey analysis is portrayed in Fig. 1.

Fig. 1 Flow of data interpretation



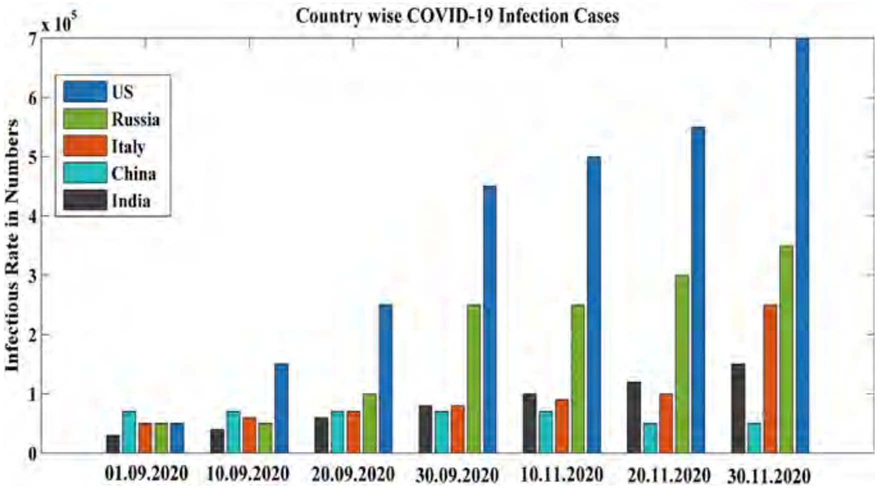


Fig. 2 Depiction of Country-wise COVID-19 Infection Cases

3.1 Proportion of Lethality

It is depicted as the ratio among the number of deaths to the number of diagnosed cases [5]. Therefore, an offset needs to fit with the Poisson regression model by facilitating over-dispersion as the logarithm of diagnosed cases. It is expressed in Eq. (1):

$$\text{Log}(M(d_t)) = \mu_0 + \mu_1 t + \mu_2 t^2 + \mu_3 t^3 + \mu_4 t^4 + \log(g_t) \tag{1}$$

Here, d_t is the number of deaths/day, and g_t is the number of diagnosed cases/day. The lethality proportion is computed with people of the same age group. It is not probable to compute the estimation proportion of lethality because of the cases under-reporting the official statistics [6]. However, monitoring and estimation of lethality proportion monitoring of current epidemic scenario. Country-wise COVID-19 Infection Cases are depicted in Fig. 2 based on the proportion of lethality.

3.2 Incubation Period

The COVID-19 incubation period is estimated with an interval of SARS-CoV-2 and diagnosis data is evaluated for the successive approaches as in [7], who recently examines the incubation period of COVID-19 with a group of symptomatic patients. For all patients, the interval exposure to SARS-CoV 2 and the appearance of the data of the symptoms are evaluated. The assumption is made by the fact that incubation time is followed by viral respiratory tract infections which are expressed using Log-normal distribution as in Eq. (2):

$$\text{Galton}\left(\text{mu}, \text{omega}^2\right) = \text{Galton}(1.621, 0.418) \tag{2}$$

The group distribution is replicated for the diagnostic cases for approximating the exposure data to SARS-CoV-2 as in Eq. (3).

$$p(j) = \sum_{t=1}^{14} Q(i) \times C_{i+j} \tag{3}$$

Here, 'Q' is the number of diagnosed cases/day; $p(j)$ is the probability of infected cases/day; $i = 1, 2, \ldots, 14(2weeks)$ is the maximal time which is considered for disease progression; and $Q(i)$ is the probability of symptoms identified/dy based on Galton probability distribution with the parameters explained above. For estimating the occurrence in 14 days, the information on diagnosed cases is not accessible for successive days. Here, a 4-degree of the polynomial model is used for diagnosed cases. These are shown in different colors for displaying the applications.

3.3 Forecasting Model

The graph shown in Fig. 3 is attained with the analysis of dataset over a series of time from Genetic Programming based ARIMA (GP-ARIMA) model which helps

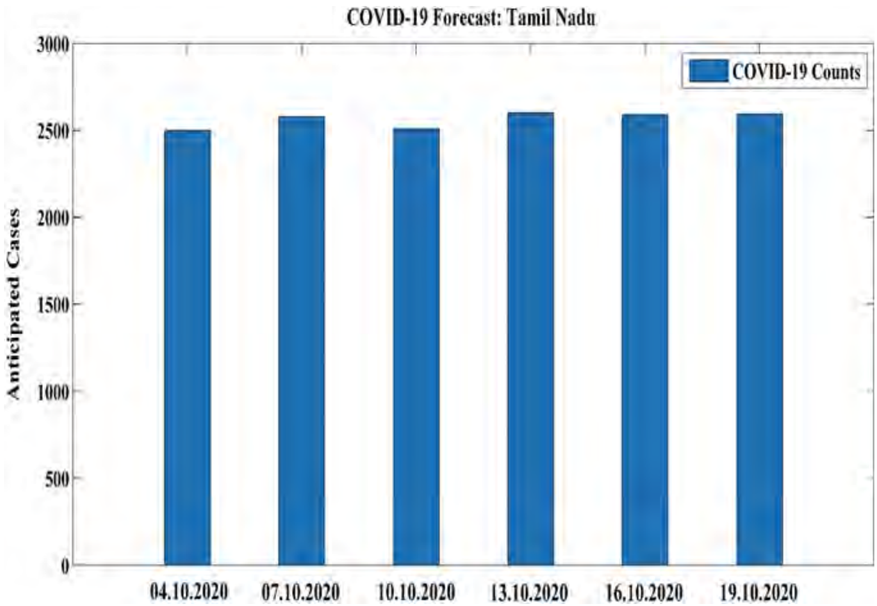


Fig. 3 Computation of COVID-19 Forecast: Tamil Nadu

Table 1 Computation of COVID-19 forecast: Tamil Nadu

Date	Anticipated Cases
04.10.2020	2500
07.10.2020	2580
10.10.2020	2510
13.10.2020	2600
16.10.2020	2590
19.10.2020	2595

in predicting the spread of COVID-19 [8] and the number of increased cases over India is provided with the equalized bar as in Fig. 3. The dataset holds the report till 30 October 2020. Therefore, the predicted graph is shown in November 2020. The *X*-axis depicts the three days interval and *Y*-axis gives the anticipated cases in 1000s predicted till November, 2020 approximately. However, 5000 cases are reported in Tamil Nadu. When the comparison is made with the report till October 2020, i.e. roughly 2500 cases are predicted with the interval of 3 days until November 2020. Simultaneously, the number of reports reaches 2600 in Tamil Nadu which is a nightmare for Tamil Nadu but not for other countries over the world. The prediction with GP-ARIMA is anticipated in Table 1.

Algorithm: Forecast Depiction using GP-ARIMA Model
D ← days between November 4 and November 19
C ← cases on November 4
Prototype ← GP based ARIMA (a,g,c,D)
Predict ← Prototype (f)
If Predict (1) > C true
then
coef ← C—Predict (1)
else false
coef ← C—Predict (1)
return Predict

3.4 Linkage Investigation and Arrangement on Diseased Cases

In light of the information accessible from the publicly supported data set at COVID-19-India dab organization, patients organization and segment subtleties were made. Patient ID, nations of movement, and mass functions were determined as hubs and relationships among patients' and voyaging functional history which is determined as organization edge. The organization comprises 551 hubs which are not detailed in this work; however, the degree of centrality of significant hubs was determined in Table 2 and shown in Fig. 4. Centrality hubs degree is determined as association's quality

Table 2 Degree Centrality based on nodal

Nodal depiction	Degree centrality
India: Religious event in Delhi	0.102775941
India: Mumbai	0.005948447
Italy	0.030072703
Gulf	0.011566642
United Kingdom	0.007270324
Saudi Arabia	0.004296100

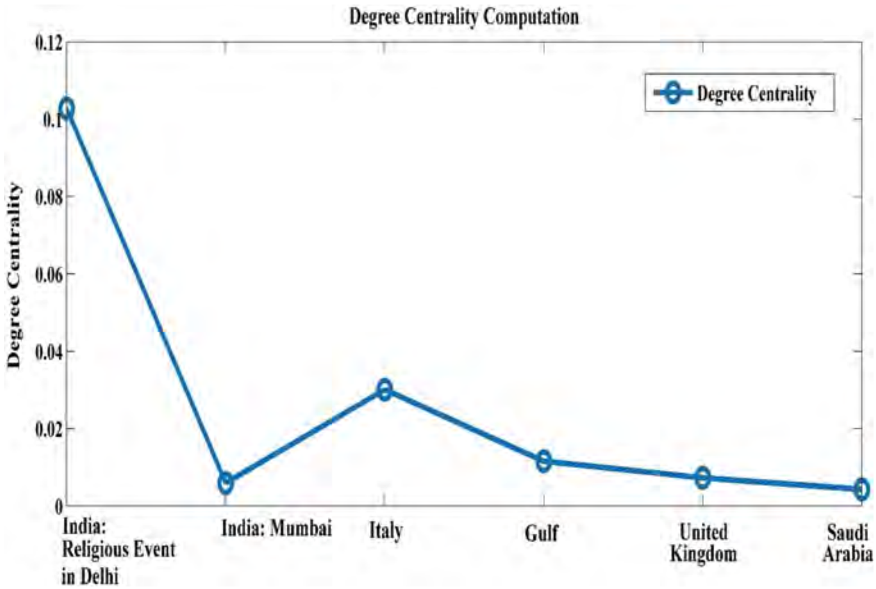


Fig. 4 Computation of degree centrality

with specific hub separated by an absolute number of edges over the organization. The main 7 hubs dependent on hubs centralities were Religious function in Delhi, Gulf, Italy, UK, Mumbai and Saudi Arabia. It is discovered as a significant hotspot for the quick spread of Covid [9].

In view of the patient information database, a characterization model (GP-ARIMA) is created to observe that based on the segment feature, can a patient have the likelihood to perish. Different examples were dug for COVID-19 patients; anyway, it is not critical for positive cases whereas information of negative cases is not accessible. The acquisition of COVID-19 contaminated patient's discovered that greater part cases rely on over 31–40 years in India, as appeared in Fig. 5. The patients are significantly over 60 year's old gathering and the greater part of individuals contaminated went back to Italy.

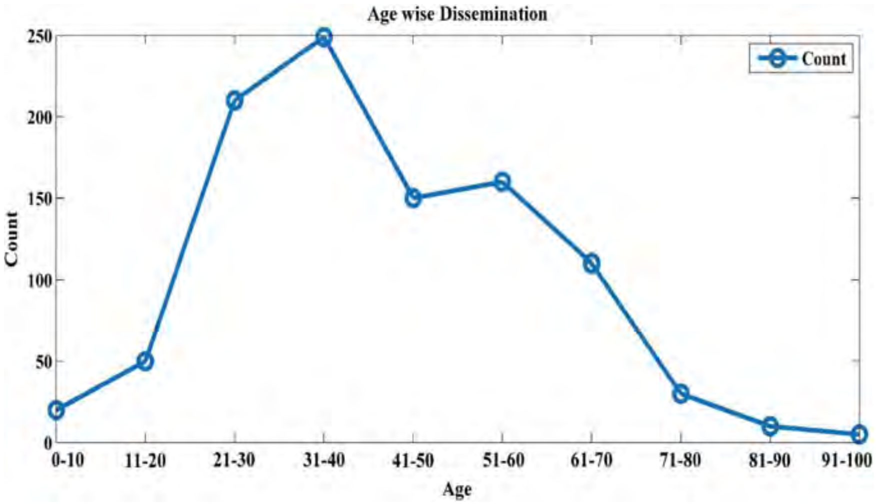


Fig. 5 Age wise dissemination

There is a characterization model for patients during the contamination from COVID-19. Choice Tree order model [10] was utilized to the group of COVID-19 contaminated patients before they face crucial conditions. The features discovered are very critical in gathering information's like patient's age, patient's gender, and patient's state. Similarly, 5000 patients are contaminated from COVID-19 infection up until this point, yet a segment subtlety of the apparent multitude of patients isn't known. Table 3 depicts the state shrewd foreseen development rate till the mid of November 2020.

4 Conclusion

The proposed work contributes an extensive analysis with COVID-19 outbreaking conditions over India. The infectious cases are rising gradually and the country needs aggressive control over the condition from India's administrative unit. There are diverse factors that deal with this condition. It is related to the growing trends of infectious cases over India. The impact of mass events over a number of cases is provided based on linkage analysis and pattern excavation who suffers from the coronavirus. It leads to the uplifting condition of complete lockdown towards the country. The present study executes various approaches to offer an analysis and the outcomes need to bridge the gap between the existing drawbacks. This work is essential for the Indian Government and various states of India, scientists, researchers, healthcare sectors of India, and the administrative unit of India. This work is completely favorable for the administrative unit to deal with the factors related to COVID 19 control in the corresponding regions.

Table 3 State wise AGR till November 15th, 2020

States	Number of infected cases as of 30th October 2020	Anticipated growth rate (AGR) till 15th November 2020
Tamil Nadu	738	215.38
Delhi	576	278.95
Maharashtra	1135	238.81
Telangana	453	256.69
Rajasthan	363	202.50
Uttar Pradesh	361	208.55
Andhra Pradesh	348	213.51
Kerala	345	30.19
Madhya Pradesh	290	195.92
Gujarat	186	126.83
Karnataka	181	64.55
Haryana	167	288.37
Jammu and Kashmir	158	154.84
Punjab	106	140.91
West Bengal	99	266.67
Odisha	42	950.00
Bihar	38	65.22
Uttarakhand	33	371.43
Assam	28	75.00
Himachal Pradesh	27	575.00
Chandigarh	18	38.46
Ladakh	14	7.69
Andaman and Nicobar Islands	11	10.00
Chhattisgarh	10	-44.44
Goa	7	16.67
Puducherry	5	66.67
Jharkhand	4	300.00
Manipur	2	100.00
Arunachal Pradesh	1	Nil
Dadra and Nagar Haveli	1	Nil
Mizoram	1	Nil
Tripura	1	Nil
Total	5749	180.85

References

1. M. Varshney, J.T. Parel, N. Raizada, S.K. Sarin, Initial psychological impact of COVID-19 and its correlates in Indian Community: an online (FEEL-COVID) survey. PLoS ONE **15**(5), e0233874 (2020)
2. J. Qiu, B. Shen, M. Zhao, Z. Wang, B. Xie, Y. Xu, A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. Gen. Psychiatry **33**(2) (2020)
3. B.L. Zhong, W. Luo, H.M. Li, Q.Q. Zhang, X.G. Liu, W.T. Li, Y. Li, Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. Int. J. Biol. Sci. **16**(10), 1745 (2020)
4. J.Z. Huang, M.F. Han, T.D. Luo, A.K. Ren, X.P. Zhou, Mental health survey of 230 medical staff in a tertiary infectious disease hospital for COVID-19. Zhonghua lao dong wei sheng zhi ye bing za zhi, Zhonghua laodong weisheng zhiyebing zazhi. Chin. J. Ind. Hyg. Occup. Dis. **38**, E001-E001 (2020)
5. Press Release. https://pib.gov.in/PressReleasePage.aspx?PRID1539877
6. M. Battegay, R. Kuehl, S. Tschudin-Sutter, H.H. Hirsch, A.F. Widmer, R.A. Neher, 2019-novel coronavirus (2019-nCoV): estimating the case fatality rate—a word of caution. Swiss Med Wkly. **150**, w20203 (2020)
7. S.A. Lauer, K.H. Grantz, Q. Bi, F.K. Jones, Q. Zheng, H.R. Meredith et al., The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann. Intern. Med. **172**(9), 577–582 (2020)
8. A.S.S. Rao, J.A. Vazquez, Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone–based survey when cities and towns are under quarantine. Infect. Control Hosp. Epidemiol. **41**(7), 826–830 (2020)
9. T. Singhal, A review of coronavirus disease-2019 (COVID-19). Indian J. Pediatr. 1–6 (2020)
10. WHO World Health Organization. https://www.who.int/emergencies/diseases/novel-coronavirus-2019 on 31st March 2020 (2020)

Artificial Intelligence and Medical Decision Support in Advanced Healthcare System

Anandakumar Haldorai and Arulmurugan Ramu

Abstract The improvements in deep learning (DL) and machine learning (ML) centered on enhancing the availability of medical information have stimulated renewed interests in computerized clinical decision support systems (CDSSs). These systems have indicated significant capability to enhance medical provisions, patient privacy, and service affordability. Nonetheless, the usage of the systems does not come without problems, since faulty and inadequate CDSS might deteriorate the quality of medical provisions and put the patients at potential risks. Moreover, CDSSs adoption might fail due to projected patients ignoring CDSS outputs as a result of lack of action, relevancy, and trust. The main purpose of this research is to provide the required guidance centered on the literature done for various aspects of CDSS adoption with a critical focus on DL and ML-centered systems: quality assurance, commission, acceptance, and selections.

Keywords Clinical decision support systems (CDSSs) · Deep learning (DL) · Machine learning (ML)

1 Introduction

Machine learning (ML) and artificial intelligence (AI) prominence, over the past few decades, has been connected to the advancing volume of the medical data. As such, this has amounted to the enhancement in AI applications in general which includes computing clinical decision support systems (CDSSs). The computing CDSSs are a particular application formulated to assist patients and clinicians in medical decision-making illustrated as an active-knowledge framework. These systems utilize more than two items of patient information to produce case certain advice about Spiegel

A. Haldorai (✉)
Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India

A. Ramu
Presidency University, Yelahanka, Bengaluru, Karnataka, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_60

halter and Wyatt. CDSSs might apply expert knowledge or frameworks mastered using ML and statistics from the clinical sets of data.

Over the past few decades, CDSSs were regarded as being capable of replacing the decision-making process in the medical sector. Nuanced which represents the modernized perspective of the CDSSs purpose aids the medical practitioners to make good decisions compared to clinicians or CDSS. The system makes this possible by processing the massive amount of data available in the medical databases. Normally, modernized CDSS makes the required recommendations to medical practitioners and the experts are expected to initiate their own decisions and overrule CDSSs projections that are considered inappropriate. Computing CDSSs has significantly evolved drastically, since their first advent that featured the computer-assisted diagnosis in the rule-centered MYCIN, HELP-alert systems and the Leed-Abdominal Pain framework. One way in which these systems have transformed is their incorporation into the medical workflow and other medical information systems. At the start, they were considered a standalone framework whereby medical practitioners had to compute the patient's data before interpreting and reading the results. Starting from 1967, CDSSs began to be incorporated into the medical information management frameworks hence assuring two major advantages: Users might not necessarily have to compute data and CDSSs might be considered proactive which means recommending or alerting the required actions without the users necessarily seeking assistance from the CDSSs [1].

In earlier research the enhancements and adoptions of standards to share, store and represent medical information which permits the separation of knowledge contents from the codes of software in CDSS. As of 2005, the medical information systems began offering application programming interfaces (APIs) whereby they might potentially interact with CDSSs, hence permitting less standard and dynamic relationships [2]. The transformation of CDSS has amounted to a high variety of CDSS forms that might be categorized according to the number of characteristics. CDSSs can provide the required support on unprompted or demand data refer to information from the alert systems. Moreover, CDSSs might be categorized underlying technologies and reinforcement learning, genetic algorithms, probabilistic models, deep learning and rules among others. According to their functions, CDSSs are categorized to supportive diagnosis, treatment planning, outcome projections, management of medications, chronic disease management, image representation, image segmentation, pathological detection and preventive care [3].

Systematic literature suggests that the usage of CDSS minimizes unwarranted practice variation, waste in the medical frameworks, enhances the quality of medical care, minimizes the risks of burnouts and overloads among the medical practitioners. However, CDSSs have fundamental negative results, since the CDSS is faulty or its ineffective usage might potentially amount to the deterioration of the quality of healthcare provisions. Critical ethical queries and the safety of patient's concerns are still pending concerns. The obligation of CDSSs has initially been to support and promote users (patients and clinicians) who are significantly liable for medical decisions. Considering the invention of DL, CDSSs are attaining ultimate human performance dimensions for different tasks, mostly for the analysis of images, acting as black-boxes whereby the aspect of reasoning for the projected tasks has not been confirmed [4]. As such, this raises some novel questions concerning the liabilities and responsibilities. The regulatory procedures are adapting particularly, categorizing the CDSSs as clinical devices (legalized effects), whereas eliminating the definition of some other CDSSs like the ones that do not evaluate images or the ones that permit the users to evaluate the recommendation basis. Nonetheless, not even the regulatory approvals are assured about the positive complications.

CDSSs inadvertently advance the workloads of medical practitioners. For instance, the well-known effects of the CDSSs alert system in patient monitoring are the alert fatigue which happens when medical practitioners start ignoring the messages as a result of the overwhelming false-alarm frequencies. Another possible risk amounting from the CDSSs adoption is medical practitioners losing the capacity to make proper decisions on their own or to effectively determine whenever it is effective to override CDSS. The present gains in AI make it critical for CDSS to be used in the process of making decisions for humans which possibly make these risks considerably pertinent [5]. This might become crucial in case of the computing system downtime or when patients with informal medical conditions are admitted for diagnosis and treatment. In that case, it is fundamental to establish the alerts to both negative and positive potential implications of CDSSs on medical decision-making processes. Several CDSSs have been in application for a few decades. However, their application has not yet been spread widely as a result of the number of issues connected to the implementation and design process such as medical practitioners not utilizing them as a result of lack of time or confidence in CDSS output.

However, there exists an immense possibility requirement for CDSS as a result of the increased volume of the present information, advancing the diversity of treatment options and the rapid transformation of clinical technologies. CDSSs might be considered valuable as a form of delivering clinical care that has been tailored to the preference of patients and the biological features. The patients might possibly benefit from the general accumulation of human skills and the clinical professionalism leading in condition monitoring, treatment and diagnosis. There exists an advancing international need for quality custom medicine which is meant to enhance patient results, minimizing financial burdens and eliminating unwarranted practice deviations. Machine learning-centered CDSSs are projected to assist in the alleviation of a number of the present knowledge and the connected quality of healthcare variation over regions and countries [6]. Therefore, the query of maintaining, evaluating, implementing, presenting, developing and designing all forms of medical decision support capacities for consumers, patients and clinicians remains a crucial segment of research in the field of modern medicine.

The purpose of this research is to provide an explanation and guidance on the various stages for successful and safe adoption of CDSS as shown in Table 1. This paper also explains how CDSS can be selected and also provides recommendations for acknowledging CDSS commissioning or testing. Moreover, the paper also describes how clinicians can rollout CDSS and provide the required guidelines for CDSS quality assurance. The rigorous process of doing the selections will aid in the identification of CDSS which fits the requirements and preference of localized

Table 1 Summarization of the stages—CDSSs adoption

Stages	Purpose
Selection	Select the best CDSS based on matched targets and the medical workflows, five rights, user acceptability and performances
Acceptance testing	Tests that CDSSs satisfy safety, privacy and security requirements applied onto clinical devices covering normal error unforeseen, exceptions and scenario situations.
Commissioning	Satisfying CDSS tests for optimized application of the medical facility (incorporating possible personalization of the safety tests). This also includes the performances in the localized context.
Implementation	CDSSs rollout the transitions from the initial workflow to the novel one after training the users and controlling their expectations.
Quality assurance	Ensure that the quality of CDSS is fit and monitoring can be done for both the external and internal updates, including the context drifts.

medical sites. Acceptance testing might ensure that the chosen CDSS fulfills the illustrated specifications and satisfies the safety guidelines. The process of commissioning will prepare CDSS for encrypted medical utility at the localized site [7]. The effective application phase has to amount in the rollout of CDSS to the trained end-users whose projections have properly been managed. Quality assurance will be utilized to ensure that the performance of CDSS is properly maintained and any issues are quickly noticed and mitigated. In this article, it is concluded that a systematic approach to adopt CDSS will aid in the elimination of pitfalls, enhance the safety of patients and promote the projections to success.

2 Selection Methods

The dimension of the commercialized CDSSs for medical applications has been advancing over the past few decades. In that case, choosing an effective CDSS from those present is not easier, yet it is considered a critical step in the implementation of effective CDSS. User acceptance CDSS is fundamental and various implementation researches have indicated how CDSS is beneficial in perceiving the concept significantly. The usages of these are determined by the medical practitioners and allied medical professionals. In that case, the projected initial step in the entire process might be to formulate the multi-disciplinary steering board which includes key stakeholders, for example, patient representation, clinician champions, IT experts and department administrators who might be willing to be accountable and make decisions for CDSSs implementation. For over three decades now, studies indicate that the likelihood of user acceptance enhances whenever CDSS implementation includes the end-users other than forcing CDSS on the end-users [8]. For CDSSs to be considered effective, CDSS might be conceived as a segment of a wider, department-wide and coherent quality enhancement strategy whereby the medical quality gap between the present patient process and the desired end-state has been identified and measured critically.

Two fundamental aspects to investigate when choosing CDSSs include CDSSs quality and how effective CDSSs fits with tackling the medical quality gap. The CDSSs quality requires to be considered at about two levels: technological level platforms and that of the knowledge or data utilized to structure it. The CDSSs framework serves as an application that is possibly a medical device to be implemented, documented, tested and designed based on the application of recognized quality assurance techniques for the development of software application applicable in clinical domains. The clinical knowledge utilized in the construction of CDSSs might not be proved medically or objectively precise. However, it should attempt to captivate the present condition of scientific or professional opinions. Moreover, it should be possible to effectively verify that the essential clinical skill set satisfies some requirements such as being considered biased, completed and consistently interpreted. As for the CDSSs-centered models mastered using statistical evaluation or through ML, the evaluation of the quality of source information is fundamental. Information quality is essential since the principle of garbage in, garbage out applies to ML. Information is principally defined to be of high quality if it can fit close to the projected purpose and certainly, it has to be unbiased and a representative sample of the medical domain (medical or patient conditions) considerably being modeled. The appropriate procedures for anomaly detections dealing with incomplete data and data cleansing have widely been applied in databases and the presence of the potential biases rectified and assessed [9].

The major indicator of the CDSSs quality is its performance matrix. Performance measures vary depending on various forms of CDSSs. For instance, in CDSSs undertaking, the outcome predictions, the segment under the receivers operating characteristics (ROCs) curve and the c-index are certainly utilized as a performance metric. In other instances, the performance aspect is evaluated in terms of saved timeframe. Nonetheless, the evaluations of performance have to be complicated, mostly when the standard of medical performance is not present such as in therapy advice frameworks, whereby the medical practitioners might disagree. In the end, the most challenging to evaluate yet a valuable performance metric is the influence of CDSS on the medical processes and outcomes. Systematic reviews by CDSSs vendors of illustrative evaluations of usability and efficiency of CDSSs application might facilitate the execution of decisions. However, this has to be mastered that the experiments done by CDSSs developers might overestimate their potential benefits whereby a third party is necessary. A critical hazard evaluation that amounts to the exhaustive list of the possible risks and consequences alongside the mitigation plans for the mentioned risks is a segment of the regulatory procedure and might provide fundamental insights into CDSSs desirability [10].

During the process of selection, CDSSs acceptability has to be considered and evaluated over the performance matrix. For users to effectively accept the CDSS output, the evidence strongly supports the medical recommendations delivered by CDSSs have to be user transparent. The dimensions of the frameworks based on the hand-engineered elements and simplified frameworks (decision trees) are typically greater compared to the ones based on advanced techniques such as DL and random trees. As indicated earlier, it is fundamental to choose a CDSS that fits the necessities of the localized site. Foremost, the following population, intervention, comparison and outcome (PICO) framework and the 'choice' procedure has to be restricted to CDSS which targets the suitable population based on relevant comparators and interventions focusing on the outcomes of interests [11].

When choosing CDSS, it is based on five essential rights that CDSSs has to accomplish, namely: what (delivery of the right data, who (right people), how (correct format, where (correct channel) and when (the best timeframe. Delivering the correct set of data also means that the CDSSs output (clinical assessments and recommendations) has to be medically actionable, unambiguous, brief and relevant. CDSS has to fit the present workflow of the users as closely as possible. For instance, incorporated in the electronic health records (HER) limiting the efforts necessary for users to accomplish and act of systematic recommendations. For CDSSs to categorically fit the workflow of a certain clinic, CDSS customization is fundamental. In that case, functionality customization provided by every CDSS has to be considered during the process of selection. Another consideration connected to the localized workflow is whether the essential information for the precise functions of CDSS is present in a certain workflow point. Another element to consider whenever choosing CDSSs is its utility, more certainly depending on how training is required to be capable of using CDSSs [12]. The vendors require being clear concerning the expertise necessary for utilizing the framework.

An essential consideration when choosing CDSSs has to be measured, contrasted to the alternate CDSS or other clinical devices (new equipment). Nonetheless, it is challenging to showcase the returns on investments of CDSSs, mostly against a lot of competing priorities evident at the delivery system framework. A comprehensive evaluation of the overall costs included in CDSS acquisition had been undertaken before its purchase which includes a single form of cost release (implementation, training and purchase, among others) but the expenses incurred over a specific timeframe such as resource utilization and costs of maintenance (time of users). The expenses have to be weighted not just over the estimated enhancement in medical outcomes, but also the projected savings as a result of the efficiencies that have been facilitated by CDSSs. Other considered factors include compatibility with CDSS maturity, legacy application and the upgraded availability.

3 Acceptance Testing

As for acceptance testing, CDSSs might effectively be seen as a clinical tool in which a lot of procedures are normally in place with medical providers. Acceptance tests for clinical tools assure that the all-defined specifications are accomplished and that the clinical devices satisfy pertinent privacy requirements. These tests are typically defined based on the CDSSs vendors but have to be operated in the availability of localized site representatives. On successful attainment of acceptance tests, the reports will thus be signed and facilitate the approval of the payments. Consequent to that, the collection of test cases has to be understandable which includes covering the edge cases of the CDSSs domains that are considered as projected cases. The technical factor of the acceptance test has to be done by the technological representatives whereas tests projected on clinically oriented tests and usability tests have to be done by a sub-group of users which include the representative samples of the expected end-user populations. The acceptance test plans have to tackle the aspects below:

(a) Setup and installation of devices.
(b) The effective functioning of APIs provided by CDSSs.
(c) Complete walkthroughs of user interface, operating CDSSs as a segment of the prevailing workflow.
(d) Medical completeness, comprehensibility, consistency, repeatability and relevance of the CDSSs output.
(e) Security, privacy and auditing functions.
(f) Normal error cases such as incorrect, incomplete, unexpected input information and closure cases (such as the outage of power) amounting to incomplete transactions. CDSSs need not output ineffective projections in the aspect of inaccurate or incomplete data. Additionally, CDSSs are projected to deal with the conditions by placing internalized consistency, giving appropriate error tests and if essential proceeding to unconsiderable shutdowns.

On top of the aspect mentioned above, the acceptance testing of CDSSs has to evaluate the CDSSs accuracy and its recommendations, as an ineffective projection that might endanger the privacy or condition of patients. The tests have to contrast the potential results of CDSSs to project outcomes on restricted, small and fixed but representative examples of the actual cases. The projected accuracy centered on acceptance test findings needs to be contrasted against the accuracy aspect claimed by statistical tests and vendors whether it is in the specified error tolerance aspect or now. The sample is applied to the other qualitative and quantitative approximations provided by the vendors. To test whether the actual accuracy of CDSS (various parameters) is in a certain error tolerance centered on the test samples, statistical tests have to be applied to evaluate the probability that the precision seen in the samples belongs to the probability distributions resolute by claimed error tolerance and accuracy. The probability is evaluated below in a particular fundamental threshold to eliminate the hypothesis that there is real accuracy in error tolerance. Lastly, an assessment for accessibility and completeness of CDSSs user manual is a segment of the acceptance test which would be fundamental for novice users or unusual or emergencies.

4 Commissioning

Commissioning represents the procedure of preparing CDSSs for safe medical use in localized sites, meeting more considerable requirements and the end-user expectations. In that case, commissioning tells if CDSSs have been incorporated in the localized site based on the agreed requirement, effectively handover the vendors and see significantly if it operates properly. It is considerable to prepare for this stage through the creation of a commissioning structure that defines the tasks, plans and necessary equipment resources which include the support necessary for CDSSs vendors. The foremost step in the commissioning structure is its incorporation in the localized site that in case CDSS inevitably necessitates some customization and configuration. Customization might be essential for safety or technical purposes. For instance, to ensure that CDSSs parameters are precisely connected to localized EHRs and that the explanation of medical terms is in sync between localized EHRs and CDSSs. Customization is considered as a powerful framework to ensure that CDSSs output is safe, useful and relevant for users. The qualitative evaluation found that the essential sites devoted enough staff timeframe for CDSS customization. A sample of customization is capable of improving and assessing the effectiveness of alerts to eliminate alert fatigue.

To evaluate that the installed CDSSs operates effectively in the localized site, the test structure has to be created and executed. To start with, the CDSSs implementation is likely to accept some transformations in the workflow of the end-users. In that regard, the data essential to support the futuristic workflow requires identification which also includes the testing of the novel workflows. Once the novel workflow is structured, the purpose is to make sure that CDSSs are operating effectively by proper testing of fundamental medically relevant cases. The steering members formed by IT experts, administrators and clinicians have to be included in the identification of essential cases and corner cases whereby the CDSSs which have been installed might fail in the localized site ecosystem and amounts to poor reliability and quality. A rare and difficult cases alongside the representative samples of localized case populations which might retrospectively be tested when the databases with past scenarios exist. In this scenario, the recommendations of CDSSs are particularly evaluated by the board of medical experts in particular blind research where medical practitioners have ignored the CDSS output and contrasted it over the decisions which have been taken over the past few decades. Nonetheless, it is fundamental that CDSS has to be evaluated and the actual-world scenarios from users' personal medical practices before implementation might be considered as well.

The option is to evaluate CDSSs prospectively through the process of running a piloted program whereby CDSSs are utilized in analogous to prevailing workflow or whereby CDSSs are utilized with the supervisions of the present workflows as a fall-back. Techniques to cover the represented samples of rare and usual cases incorporate random sampling, control flow testing and input selection. In the process of plotting, it is essential to perform an initial evaluation of the medical relevance of CDSSs as user acceptance, considering CDSS projections and its effects on medical decisions and certainly on the health and patient results. Fundamental deviations on the approximation performance of CDSSs during this stage as contrasted with that in the vendors' performance claim and acceptance testing, including error tolerance have to be assessed with the vendors. Failure mode evaluation is a fundamental segment of commissioning tests, whereby errors in information entry are categorically simulated and CDSSs are tested and analyzed for consistency. Testing in commissioning is fundamental to grow the confidence of localized physicians so that the support framework operates in a localized setting [13].

5 Implementation

The process of implementation is a fundamental factor in the CDSSs success and includes the designing and execution of the wide-range rollout plans in which transition occurs from old workflow to novel ones. This includes CDSSs and the CDSS deployment in the localized site. Effective CDSSs implementation necessitates the preparation of both the localized site infrastructure and the individualized users, for global usage of CDSSs. The analysis of infrastructure will depend on localized sites and CDSSs. However, there are common frameworks in the manner in which users can get prepared for the application of novel CDSSs. It is fundamental to educate and communicate with users [14].

Good training of the relevant stakeholders, unexpected CDSS uses critical care for the successful implementation and has to include various aspects of the time to use or eliminate it [15]. Apart from that, various aspects of how it has to be used, how the interpretation of the CDSS output has to be done and how the system can be overridden are some of the factors to consider. This also incorporates assisting users to understand how the CDSS can affect normal activities in the modern age and how responses can be provided [16]. It is critical to consider this as a major segment of the training to effectively manage expectations of the users in terms of effectiveness and efficiency which also includes ensuring that users that comprehend the weaknesses and strengths of CDSS. Various key stakeholders might have various expectations [17] and several primarily considered CDSS can be used as a channel for promoting standardization, safety and quality where medical practitioners might consider it as different. The process of training should also consider the major purpose of getting the users ready for the essential process. In this process, CDSS shall only be utilized if it is considered essential by the people using it. The hands-on training aspect is considered as an essential tool which means that users might choir hand-holding at first since the on-site support from the relevant vendors is required to aid in the process of mitigating immediate problems that might be evident. The rollout or deployment of CDSS might be done incrementally. For instance, rolling out one post or facility to retrieve the Kinks might require prompt the effective preparation [18].

6 Quality Assurance

Before the process of deploying CDSS fully, it is fundamental to formulate the quality assurance system which will be obliged to ensure that the performance and CDSS safety is effectively maintained. This process ensures that quality is maintained throughout the life cycle. As a fundamental segment of quality assurance program in CDSS, the aspect of performance has to be illustrated based on the application of a wide range of metrics that focuses on efficacy and efficiency. These allow the effects of CDSS to be evaluated over a particular time frame. The basis of efficacy can be categorized to CDSS functioning which includes specific and sensitivity of agnostic frameworks, and pretty generic aspects such as the safety of patients are transforming potential patient outcomes, e.g., life expectancy. The aspect of efficiency can be evaluated as resource-based, such as productivity and costs. To evaluate the performance of CDSS, it is fundamental to quantify the baseline levels of performance which include CDSS implementations beforehand and evaluation of the targeted performance upfront.

The quality assurance plan should guarantee that the various malfunctions are resolved after not being still in the shortest duration possible. To make the discovery of CDSS issues possible, techniques should be identified through the use of quantitative and qualitative analysis such as creating alert frameworks through our overriding recommended CDSS. Visual evaluation, statistical procedures and control evaluation have indicated good results in the process of detecting the possible malfunctions. Adding to these malfunctions, it is essential to track or log the various cases where CDSS has not been fulfilled; i.e., whenever the alerts are being ignored or whenever the recommendations are being overridden. This is the case to tell if CDSS has been overridden and the reason behind why valuable and urgent insights amount to the identification of malfunctions that have not been noticed. Evaluations are tracking down the processed application installed. CDSS is essential since it might amount to minimized performance.

Models can be upgraded using various means such as baseline shift levels (for binary models) and the cut-off framework for the binary results; computing of novel values for the present parameters or doing framework training on expanded information which amounts to novel framework parameters, coefficients and potential cut-offs for the novel binary outcomes. The best safeguarding for external and internal drift is to create and significantly analyze logs for incorrect and inappropriate feedbacks from CDSSs. At the same moment, repeated localized validation cohorts have been being gathered from one time to the other or to fundamentally reevaluate the kind of tests conducted during the commissioning phase. This form of repetition might aid in ensuring that CDSSs remains are medically valid irrespective of the transformations of localized practices or the evidence-centered guidelines. This form of continuous localized validation system will be essential when giving updates to CDSSs. Lastly, it is essential to refocus that no CDSSs can be considered perfect. However, at a minimal measure quality assurance framework will report that the CDSSs performance has met the projected criterion on the commission findings as benchmarked.

Among the CDSSs top priorities, steering up the members might be purposed to update and establish the management protocols. CDSSs with clinical software generally are purposed to be updated even in offline modes. This means that through the user instigated or vendor instigated transition CDSS is considered as temporary, it takes out the medical application and is included in the maintenance condition. The subsequent transformations are done in the maintenance conditions, such as the application of software version corrections and updating of the error function. As for analogy with the various maintenance aspects and quality assurance of medical systems, the medical handover which is the recognition of the system into the medically operational condition based on these upgrades has to be only permitted after several CDSSs performance and verification done on the transformation system.

The essential tests have to be specified by the vendors or the maintenance manual-centered risk evaluation, but it might be necessary to incorporate several tests from the acceptance testing process. This is to certify that the fundamental functionalities of CDSSs have been updated. With the shift of clinical software frameworks to the cloud system, enhancing system automations and the mathematical algorithm that is capable of learning should countenance the chance of CDSS updating in an online mode. These CDSSs might be permitted to transform the actual time centered on the kind of interactions between the users and system projections, such that the condition of CDSSs might significantly transform some forms of interactions. An upgraded management protocol might explicitly allow online updates that pose interesting and novel challenges which are based on the search idea trade-offs between the possibility of continuous enhancement of performance against the risk of undetected performances and degradations such as systematic biases in inputs.

Another essential priority is based on the implementation of a routine quality assurance plan which specifies the kind of tests that have to be undertaken, the time they should be undertaken and to who they are meant for. As a major segment of the quality assurance test, different functionality aspect of CDSSs is evaluated over the acknowledged ground truth. As a rule, the forms of quality assurance obligations are retrieved from the various checks as in the commissioning stage. In that regard, the documentation findings of commissioning might be utilized at specified timeframe intervals, to certify that CDSSs performance has not transited over a considerable timeframe. A lot of statistical anomaly determination frameworks applied to the CDSSs detection over a considerable timeframe are compared and described in research that will vary depending on the condition of CDSSs. The frequency and condition of CDSSs quality assurance tests depend on the chances of unwanted deviations of CDSSs performances and the possible effects.

The quality assurance tests have to be undertaken frequently for either the potential failures or non-conformance events which amount to severe requirements. Unlikely events and failures which do not have fundamental medical results require checking more infrequently. An essential effort has to be directed to the procedural mitigation of the failures carrying some consequences because this might be challenging to intercept in the routine quality assurance system. To effectively create and execute the quality assurance plan, it is projected that the users in charge are in acquisition for the tasks based on data evaluation methods through the process of training. As such, bearing in-depth knowledge on the manner in which CDSS underlying technologies operating will permit clinicians to identify the potential malfunctions and comprehend their root cause. This form of training can be provided by the vendors themselves or any third party providing some customer training services. Eventually, AI and ML aspects will be tackled in the medicinal and physics field that will therefore amount to a deeper and wider understanding of the various computing systems.

7 Conclusion and Future Scope

CDSS has shown significant potential for enhancing medical services and the safety of patients which also incorporates the minimization of unwarranted variation, resource usage and costs. AI-centered CDSS in recent times has been recognized for their capability to leverage the enhancing availability of medical information which is purposed to help patients and medical practitioners in various hospital conditions (such as providing personalized evaluations of medical results or proposing the correct diagnoses) which are based on both structured data (i.e., EHR) and unstructured data (i.e., clinical imaging). However, inappropriate and inaccurate CDSS might potentially deteriorate the quality of medical services and place the patients at possible risks. AI-centered CDSSs have reported some pitfalls which have to be covered in future research. In the future, researchers should focus on over fitting the limitations on the kind of information utilized in AI training. These measures might amount to CDSS failure in generalizing from the training information and to determine the performances at localized sites. In that case, considerable precautions have to be put in a place to control the possible CDSS effects. It is fundamental to choose with care the CDSSs matching the medical requirements of the localized sites. With any clinical device, CDSS necessitated rigorous acceptance testing, quality assurance and commissioning by the localized sites. Apart from that, effective implementation plans are key to overcome the potential barriers for successful CDSSs.

References

1. C. Bennett, K. Hauser, Artificial intelligence framework for simulating clinical decision-making: a Markov decision process approach. Artif. Intell. Med. **57**(1), 9–19 (2013). https://doi.org/10.1016/j.artmed.2012.12.003
2. P. Lucas, Dealing with medical knowledge: computers in clinical decision making. Artif. Intell. Med. **8**(6), 579–580 (1996). https://doi.org/10.1016/s0933-3657(97)83108-9
3. H.N. Dinh, Y.T. Yoon, Two novel methods for real time determining critical clearing time: SIME-B and CCS-B based on clustering identification, in 2012 Third International Conference on Intelligent Systems Modelling and Simulation, February 2012
4. Z. McLean, Formulating the business case for hospital information systems—analysis of Kaiser permanente investment choice. J. Med. Image Comput. 42–49 (2020, July)
5. B. Srpak, N. Campbell, Analysis of biological framework and incorporating physiological modelling, J. Med. Image Comput. 34–41 (2020, July)
6. W. Horn, Artificial intelligence in medicine and medical decision-making Europe. Artif. Intell. Med. **20**(1), 1–3 (2000). https://doi.org/10.1016/s0933-3657(00)00049-x
7. I. Rábová, V. Konečný, A. Matiášová, Decision making with support of artificial intelligence. Agric. Econ. (Zemědělská Ekon.) **51**(9), 385–388 (2012). https://doi.org/10.17221/5124-agricecon

8. R. Yager, Generalized regret-based decision making. Eng. Appl. Artif. Intell. **65**, 400–405 (2017). https://doi.org/10.1016/j.engappai.2017.08.001
9. N.M. Hewahi, A hybrid architecture for a decision-making system. J. Artif. Intell. **2**(2), 73–80 (2009). https://doi.org/10.3923/jai.2009.73.80
10. C. Gonzales, P. Perny, J. Dubus, Decision making with multiple objectives using GAI networks. Artif. Intell. **175**(7–8), 1153–1179 (2011). https://doi.org/10.1016/j.artint.2010.11.020
11. R. Degani, G. Bortolan, Fuzzy decision-making in electrocardiography. Artif. Intell. Med. **1**(2), 87–91 (1989). https://doi.org/10.1016/0933-3657(89)90020-1
12. P. Giang, P. Shenoy, Decision making on the sole basis of statistical likelihood. Artif. Intell. **165**(2), 137–163 (2005). https://doi.org/10.1016/j.artint.2005.03.004
13. D. McSherry, Conversational case-based reasoning in medical decision making. Artif. Intell. Med. **52**(2), 59–66 (2011). https://doi.org/10.1016/j.artmed.2011.04.007
14. T. Leong, Multiple perspective dynamic decision making. Artif. Intell. **105**(1–2), 209–261 (1998). https://doi.org/10.1016/s0004-3702(98)00082-4
15. R. Chan, A. Morse, Artificial intelligence-enabled technologies and clinical decision making. Univ. West. Ont. Med. J. **87**(2), 35–36 (2019). https://doi.org/10.5206/uwomj.v87i2.1425Mallapur
16. P. Mallapur, J. Shiva Krishna, U. Hosmani, K. Kodancha, Design of gear tooth rounding and chamfering machine. Trends Mach. Des. **4**(3), 38–44 (2018)
17. T.K. Araghi, Digital image watermarking and performance analysis of histogram modification based methods, Intelligent Comput. 631–637 (2018, November)
18. G. Yu, Z. wei Xu, J. Xiong, Modeling and safety test of safety-critical software, in 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems (2010, October)

Survey of Image Processing Techniques in Medical Image Assessment Methodologies

Anandakumar Haldorai and Arulmurugan Ramu

Abstract Medical image assessment represents the procedure utilized in the formulation of pictures of different parts of the body to study or identify a particular disease. There are many medical image assessment procedures carried out every week in the entire globe which means that the sector is rapidly growing as a result of the constant advancement of image processing methodologies such as Image Enhancement, Analysis, and Recognition. This article, presents a critical survey of medical image assessment based on the application of image processing methodologies. It also provides a summary of how image interpretation issues can be exemplified using various image analysis algorithms such as RIO-centered segmentation, *k*-means and watershed methodologies.

Keywords Image processing methodologies · Medical image assessment · Medical image processing

1 Introduction

Medical image assessment represents the procedure of projecting visible pictures in the inner body structure for medicinal, scientific evaluation and treatment to get a clear view of the interior parts. This procedure follows the management and identification disorders of the body and also creates an information bank of regular functions and structures of the body organs for ease recognition of body anomalies. This procedure incorporates both the radiological and organic pictures that utilized the electromagnetic energy (Gamma rays), magnetic scopes, sonography, isotope and thermal pictures. There are various technologies utilized in the process of recording data on the functions and location of certain body parts. In a single year, thousands of pictures are done in the whole world for various diagnostic aims and basically, half

A. Haldorai (✉)
Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India

A. Ramu
Presidency University, Yelahanka, Bengaluru, Karnataka, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*, Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_61

of them utilize non-ionizing and ionizing radiation modules. The process of medical image assessment projects the image of internalized structures of the internal structures without the invasion procedures. These body pictures have been projected by the rapid processors as a result of the conversion of energies which are projected logically and arithmetically into signals. The signals as a result are transformed into digitalized pictures to represent various forms of body tissues.

The healthcare image assessment processing denotes how digitalized pictures are handled in a computer system. This procedure incorporates various forms of operations and techniques which include image storage, communication, presentation and gaining. The image assessment element which represents the measure of features such as colour and illumination is considered as a viewed element. The digitalized picture includes several advantages such as affordability and speedy processing which further helps to cut on the cost, easy storage, quality evaluation and adaptability. The demerits of the digitalized pictures include the exploration of copyrights, the inability of resizing and quality preservation including the necessary massive storage volumes. The image processing methodology is the application of the computer in the process of manipulating digitalized pictures. This methodology incorporates advantages such as communication, information storage, adaptability and elasticity. With the advancement of the various protocols image assessment resizing methodologies, the pictures are maintained as efficient. This methodology incorporates several protocols meant to synchronize pictures. Both the 3D and two-dimensional pictures can effectively be processed in multi-dimensional aspects.

The methodologies of image processing were established in the early 1960s and utilized in various fields such as clinical purposes, TV image enhancement, space and artwork. In the late 1960s and early 1970s with the rapid advancement in computer technology, the image processing cost is gradually becoming rapid and less. In the 2000s, the processing of pictures was made faster, cheaper and simple to use. Our human visual framework is one of the most credible schemes that have once existed in history. It is a framework that allows organisms to comprehend and organize the various complex aspects of the exterior ecosystem. Apart from that, the visual framework includes the eye which is capable of transmitting light to neural elements and signals before relating them to the brain system and excerpts the required information. The human eye is considered as a bilateral cylinders framework, which is found in the anterior segment of the skulls and is 2.5 cm in length and crosswise diameter [1].

At the center of the human eyeball, a blackened structure known as the pupil is Doi. This part allows light to penetrate the eye and narrows whenever exposed to too much light. This diminishes the amount of light to the human retina and boosts the process of visual image assessment. There are a lot of muscles around the human eye which are obliged to control the pupil’s widening. In that case, the eye includes some supportive elements, typically known as the sclera. Its lens acts as a ligamentous section found around the cornea and with a shape that transforms considerably as a result of the contraction of muscles. Light is concentrated in the middle section of the human eye to focus from the lens and cornea onto the retina. The Fovea is Doi to concentrate the pictures into the human retina [2]. Lastly, the brain system is there to formulate the colours and details through the application of multiple image assessment processes. In that case, this paper evaluates the image processing methodologies, image classification and segmentation, with a critical focus on the human eye.

2 Image Classification and Image Segmentation

The methodologies of image segmentation can be categorized as both the techniques and features utilized. The aspect of features includes Edge Information, Pixel Intensity and Texture. The techniques are the center of the features and can also be grouped into structural methods and statistical methods.

2.1 Structural Methods

In this section, spatial elements of the image: regions and edges are considered. Different edge detection methods and algorithms have been related to the extraction boundaries between the complete brain systems. Nonetheless, these algorithms are considered sensitive to noise and artifacts. The growth of regions is another key to structural methodology. With this technique, an individual starts by grouping the pictures into smaller segments that can be viewed as seeds. The boundaries between the various adjacent parts are therefore evaluated. Firm boundaries considering several properties are maintained, whereas rare boundaries are eliminated and the adjacent parts are joined. The procedures are done iteratively to eliminate the boundaries that are weak and worth the elimination. Nonetheless, the performance methodology is dependent on the selection of seeds and whether certain parts or regions have been defined or considered robust.

2.2 Statistical Methods

The statistical method is considered as pixels about probability figures and is determined about the intensified image distribution aspect. In this methodology, the image structures are labelled and assigned through the comparison of the Gray level figures to certain intensified thresholds. One threshold acts on the image segments in two separate parts such as foregrounds and backgrounds. In other times, the job of choosing the thresholds is easier since there is a variation between the Gray level of these objects requiring segmentation.

2.3 Mathematical Methods

Mathematics methods and frameworks establish the foundation of medical image processing. Based on these methodologies on the extraction of information of the pictures fundamental techniques are presented for attaining scientific development in behavioral, biomedical, clinical and experimental research. Currently, medical/healthcare pictures are attained by applying wide-range methodologies over various biological scales that are beyond simply visible microscope and photograph lights used in the twentieth Century. Modernized medical/healthcare pictures might be viewed to be geometrically featured as a collection of data samples that quantify the diverse physical aspects and the time varied for the tissue hemoglobin. The widened scope of the dramatic illness enhances the capacity to apply novel processing methodologies and link a lot of sample channels of information into complex and sophisticated mathematical frameworks of both physiological dysfunction and function.

3 Key Issues in Image Classification and Image Segmentation

The key issues are well comprehended to play a fundamental role of image assessment in the process of therapy analysis and to appreciate the present application of pictures after, during and before treatment. Our evaluation focuses on four essential elements of Image-Guided Therapy (IGT) and Image-Guided Surgery (IGS): control, monitoring, targeting and localization [3]. Certainly, in the sector of medical image assessment, there are four significant issues identified:

(a) Segmentation—Automated methodologies that form the patient certain frameworks of essential anatomies are from the pictures.
(b) Registration—Automated methodologies that are in accordance with the multiple sets of data with each other.
(c) Visualization—The tech ecosystem that image-guided processes are displayed.
(d) Simulation—The software which might be utilized to plan, experiment processes and evaluate certain accessibility techniques and simulate projected treatments.

3.1 Modality of Pictures

Various forms of image assessment techniques are been evaluated to suit the clinical application. Practically, these are complementary as technologies that provide more insight into underlying realities. In the sector of medical image assessment, these various image assessment methodologies are referred to as modalities. The anatomic modalities provide more insight into the anatomic morphology and include Magnetic Resonance Image assessment (MRI), Ultrasound (US), Radiography and Digitalized Subtraction among others. The functional modality represents forms of metabolism of the various underlying organs and tissues. They incorporate the three various nuclear medical modalities which include Scintigraphy, Positron Emission Tomography (PET) and the Single Photon Emission Computer Tomography (SPECT), including the Functioning Magnetized Resonance Image assessment (FMRI) [4]. The mentioned modalities are not exhausted since other novel methodologies are being introduced every year. Actually, different pictures have now been acquired digitally and incorporated in computerized image archives and the various communication frameworks.

3.2 Issues in Medical Image Assessment

There are a significant number of issues in medical image assessment and processing. These incorporate:

(a) Image restoration and enhancement.
(b) Accurate and automated segmentation of feature elements.
(c) Accurate and automated fusion and registration of multimodalities of pictures.
(d) Categorization and picture features and the categorization of typing structures.
(e) Quantitative evaluation of image assessment features and data interpretation measures.
(f) Enhancement of integrated frameworks in the medical sector.

Categorization of digitalized pictures is done based on two broad forms of pictures. The raster pictures illustrate the four-sided categorization of typically grouped sampled figures referred to as pixels. The digitalized pictures are typically inaccessible and include multifaceted colour variations. The digitalized pictures include fixed resolutions as a result of the size of the pixels. The digitalized pictures have the capacity to lose their quality in the process of resizing as a result of some form of missing information. Digitalized pictures are utilized significantly in photography and image assessment as a result of effective shading of image colours. Image-gaining tools are obliged to control image resolution. Digitalized pictures incorporate several formats such as Tag Inter-leave Formats (TIFF), Portable Network Graphic (PNG) and Paintbrush (PCX) among others. In image assessment, the vectors are described as the bent and wrinkles which illustrate the precision of computers. These vectors incorporate multiple qualities such as hue, dimension and line width. Apart from that vectors can typically be scalable and the reproduced pictures are considered to be in various magnitudes without transforming or altering their quality. The vectors are also effective for use in diagrams, line paintings and design.

4 Digitalized Image Processing Application

The technology of digital image processing is applied in various medical sectors. These include:

4.1 Medicine

In the medical sector, a lot of methodologies are utilized in the process of segmentation and texture evaluation that is utilized for disorder identification and cancer treatment. Image assessment registration and fusion techniques are incredibly utilized in the modern age mostly in novel modalities i.e. PET MRI and PET CT [5]. In the sector of bio-data, telemedicine and format compression methods are utilized to link to the pictures remotely.

4.2 Forensics

The most common methodology utilized in the sector includes pattern matching, edge detection, de-noising, biometric purposes and security purposes such as the face, fingerprint documentation and personal identification. The aspect of forensics is centered on the database aspect related to an individual. The forensics is capable of matching input information (photo, eye and fingerprints) with the potential databases to effectively define individual identity.

Several diagnostic healthcare image assessment modalities are applied in the probing of the human body. Moreover, it involves the interpretation of image results that necessitate sophisticated image assessment and processing of methodologies that have the capacity to enhance the interpretation and analysis of pictures hence providing both fully-automated and semi-automated detection frameworks of tissues. Moreover, it also enhances the characterization and measurement of pictures. Generally, several transformations are required to effectively extract information based on the interest of pictures and hierarchy of steps for data enhancement that facilitates feature extraction, image classification and preceding analysis [6]. Typically, these are done sequentially for more sophisticated obligations that necessitate parameter feedback concerning steps to initiate iterative loops. Various ongoing research areas have been chosen to highlight new enhancement in the display and analysis of data in the projection that these methodologies will be considered to other selected applications.

4.3 Mammography

Mammography represents one of the most essential techniques in the process of analyzing and identifying breast cancer which is a form of malignancy in women. This application is capable of detecting illness at its initial stages whenever surgery and therapy are considered effective. Nonetheless, the screening mammogram and its interpretation are considered a repetitive task that includes subtle suffers and signs from the high rates of false positive and false negative. The Computer-Aided Diagnosis (CAD) focuses on the increment of predictive values of the methodologies through the identification of mammograms to show the areas of suspicious body abnormalities and evaluate their features, as an assistance framework of radiologists. Approximately 90% of breast cancer disorders are reported in the cell lining the milky duct of the human breast and referred to as the Ductal Carcinoma in Situ. Whenever tumour surpasses the duct lining, it is considered invasive and might metastasize to some other body parts. The radiographic indication is grouped into two broad sections: Lesions and Micro-calcification. Micro-calcification is considered a major segment of detecting the 'in situ' carcinomas which are the ones found in the milk ducts and the order of minor diameter and microns. Several lesions are ill-defined to shape and typically with speculations and strands that radiate from tissues and the same in radio-opacity to the normalized tissues. The image assessment necessity in mammography is considered stringent in both contrast and spatial resolution.

The reliability and performance of CAD are dependent on several elements including feature selection, computational efficiency and lesion segmentation optimization, computational efficacy and image relationship between the healthcare relevance and image visual similarity of the CAD findings [7]. Image segmentation of the breast part acts as a limit of the search region for micro-calcification and lesion. Moreover, this aspect is essential to transform the grey valuations of the pictures to compensate for the varying thickness of tissues. One of the best methods to execute this is to incorporate grey values about the Euclidian Distance Map which is capable of mapping distances to the skin lines in a more smoothened aspect of a mammogram.

The noises in these pictures might be minimized based on the application of median filtration. However, it might disturb the contrast and shapes of minor body structures. An enhanced methodology links the findings of morphological dilation and erosion using the multi-structured elements. To enhance the reliability and accuracy of mass body part segmentation, a wide-range computing algorithm have been projected, tested and developed. These algorithms include the active contour model algorithm, adaptive part growth algorithm, multiple layer topographical algorithms and the vibrant programming boundary algorithms. As a result of the diversified breast mass and the overlapping of breast tissues in the two-dimensional projection of pictures, it is challenging to contrast the robustness and performance of the image segmentation techniques.

The features that are essential for characterizing lesions incorporate the measure of speculation, its texture and shape. Spiculation elements are typically incorporated in the calculation of image assessment gradient since Sobel Masks. The edge gradient which is cumulative from the Sobel magnitude of image edges might be plotted as histograms of the radial angles from the Sobel edge phases to determine the speculation degree. The Full Width of Half Maximum (FWHM) of the image gradient is capable of distinguishing the speculated masses of the smooth mass. Other features have utilized the multiple-scaled line detector to measure and detect the speculated mass [8]. The middle of the mass lesion acts as circular to specified filters while the boundaries of the lesion might be unwrapped which means that it is varied from the smoothed version utilized to feature the spiculations degree. Other essential features incorporate the symmetry that might include the automated registration of both the right and left breast pictures and transform with time. Gabor and Wavelets filters have exclusively been compared and investigated in literature and the Gabor filter has more performance and corresponds to the human vision, particularly for edge detection sensitivity. Other texture elements retrieved from the co-occurrence transformation have been also been tested.

Over the past few years, fractal dimension has been indicated to be efficient and effective as a metric for evaluating the texture in the classification and detection of suspicious breast mass parts. The fractal dimensions might be utilized to compare benign and malignant breast mass and incorporates the high connections with visualized similarities. Since the fractal dimension is considered as a feature computed in the domain frequency it incorporates the merits of invariant lesion position to scale and rotation. Several researchers have extracted various characteristics and utilized the principle element evaluation to detect the most essential combinations. Various methodologies might be analyzed using the Receiver Operating Characteristics (ROC) evaluation. However, this might not be contracted with one another unless one image database is utilized. ROC curves show the performance of the PC categorization radiologists and methodologies in the job of contrasting between the benign lesion and malignant. ANN shows the artificial neural networks utilizing the cumulative edge gradients elements and the hybrid framework utilizing several features.

Micro-calcification might be analyzed using the morphology aspect (brightness, area and shape) of personal spatial distribution, heterogeneity and calcification of personal calcification in a specific cluster. They might be enhanced through the thresholding of pictures and morphological opening based on the application of structural elements to effectively eliminate minor objects whereas preserving the shape and sizes of calcification. The isolated form of calcification incorporates minor clinical relevance which means that researchers and investigators applied the clustered algorithms to the categorized system, whereby the clusters include more than the chosen micro-calcification in the parts of the sizes chosen. These frameworks are curtained and implemented with the use of *k*-Nearest Neighbour (k-NN) algorithms [9].

Both the spatial heterogeneity distributed with these body features in a single cluster can be utilized in the evaluation of the malignancy possibility, discriminant evaluation, Bayesian methodology, rule-centered methodology and the genetic algorithms, essential in classification. CAD system does not necessarily have to be flawless since they are typically utilized with radiologists only. Due to the fact that the cost of the missing cancer is significant compared to benign misclassification results, it is essential to develop and minimize false negatives which are those that incorporate high levels and sensitivity for false positives.

4.4 Bone Osteoporosis and Strength

Osteoporosis represents the prevalent bone illness features by the absence of bone firmness and the consequent fractured risks. Due to fact that it tends to be asymptomatic until the body structures happen, a few individuals are essentially diagnosed at a go for effective therapy analysis and administration. Medically, bone minerals and densities are utilized to effectively assess and diagnose osteoporosis. Transformation in bone masses is typically utilized as a surrogate for mitigating future risks. Even though Bone Mineral Density (BMD) is considerably utilized clinically, it has graduated the internalized bone system which is a fundamental determinant of the healthcare strength of the bones and might possibly amount to accurate osteoporosis diagnosis [10]. The limited resolutions of the commercialized CT scanners preclude effective resolution of the trabecular features. Nonetheless, CT pictures retain a number of architectural data. However, it is diminished by insufficient Modulation Transfer Function (MTF) of the image assessment framework which is featured by the fractal element of the trabecular bone and lacunarity that represents the measure of gap distribution in pictures.

Fractal dimension illustrates the manner in which objects occupy sufficient spaces and are connected to the structural complexity. The dimension of fractal structures is connected to the radial Fourier energy spectrum of pictures as a result of applying the Fraction Brownian Motion as a framework of the natural fractal. The fractal signature estimates that are CT independent as scanners utilized a number of settings and might be retrieved through the correction of the energy spectrum for the degradation of pictures as a result of noises and image blurring using MTF scanners. Nonetheless, transformation in fractal dimensions requires to be interpreted carefully. The globalized fractal dimension should not transform monotonically with reference to decalcification.

$$L(r) = 1 + \left\{ variation \frac{r}{mean^2(r)} \right\} \tag{1}$$

Lacunarity evaluates the gap evaluation and sizes of information which means that the higher the heterogeneity the higher the form of lacunarity. In this case, a firm algorithm for evaluating lacunarity evaluates the form of deviation from the translating invariance of the image brightness and distribution utilizing the gliding box sample. Lacunarity might be defined based on the application of localized moments which is a measure for every neighbouring size of pixel image i.e. the mean of (r) and the variation of (r) that represent pixel values for the nearest sizes of (r). Lacunarity value mean can be evaluated and calculated over a wide scaled range for the bone pictures to show the average marrow sizes and the heterogeneity degree.

5 The Medical Image Assessment System

The medical image assessment systems utilized the signals from the various patients to project pictures. The systems involve both non-ionizing and ionizing sources.

5.1 X-Ray Image Assessment System

From the advent of the X-ray system by Roentgen, a German scientist, this system has been utilized to picture the various body parts for the purposes of diagnostic. In the X-ray tubes, an electron is produced in the cathode via the thermal processes and effectively accelerated via a Potential Difference (PD) of 50–150 kV. An electron hits an electron to effectively produce the X-rays. However, only 1% of power is transformed to X-rays while the remaining energy is transformed to heat [11].

5.1.1 X-Ray Tube

In an X-ray machine, the pictures are projected based on the two-dimensional design for the evaluated part of the body. The fluoroscopy framework is utilized to effectively scan any moving organs. As such, the obtained image can be communicated, stored and displayed via various machines. Computing Radiography (CT) utilizes image receptors to effectively project pictures. X-ray includes a screen enclosed with housing phosphor devices. Mammography pictures are utilized to compare the various illnesses in the breast tissue. The image assessment of mammography makes use of minimal energy contrasted to bony structures image. PD ranges applied include 15–40 kV.

5.2 Computing Tomography (CT)

In modality, pictures are projected in multi-dimensional aspects instead of conventional radiography. The CT scanner projects multiple body slices of tissues of the body in various directions. In the CT scanners, patients are included in scanners and apertures through the rotation of an X-ray tube in various directions in Fig. 1.

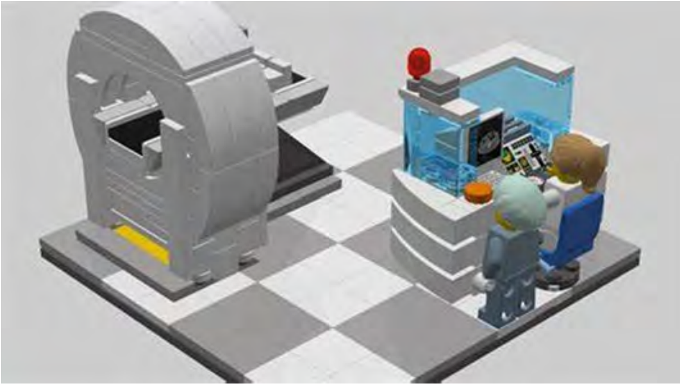


Fig. 1 Image of the computer tomography scanner

5.3 Nuclear Medicine

This form of image assessment modality makes use of radioisotopes to project pictures of functions of various structures like liver, kidney and heart. Radioisotopes are labelled by the pharmaceutical segments and materials that are used to control the various body organs. The photons produced by patients are collected in the detectors and changed into signals which are then converted to interpretable digitalized pictures. There are various forms of nuclear medicine scanner modality such as positron, tomographic and planar emission. The planar emissions project two-dimensional image for both positron and tomographic emission [12] shown in Fig. 2.

5.4 Ultrasound

Ultrasound represents the method that uses high-frequency sound waves to effectively produce an image of internal body structures from the returning echoes. This method is the same as the region detection method that is used by animals such as whales and bats in the ecosystem. With this method, a high-frequency pulse is transmitted throughout the body using transducers as a travelling wave via the tissues of the body. A number of these waves are therefore reflected and absorbed back. The reflected form of a wave is accepted by these transducers and changed to electrified signals. These electric signals are changed into a digitalized aspect and transformed into the main computer system. The computer systems utilize logic and arithmetic calculation to project two-dimensional pictures of every scanned structure. In the ultrasound system, a wide-range of pulses is communicated per millisecond [13]. There are various image assessment methodologies utilized to boost the ultrasound pictures shown in Fig. 3.

Fig. 2 Image nuclear medicine

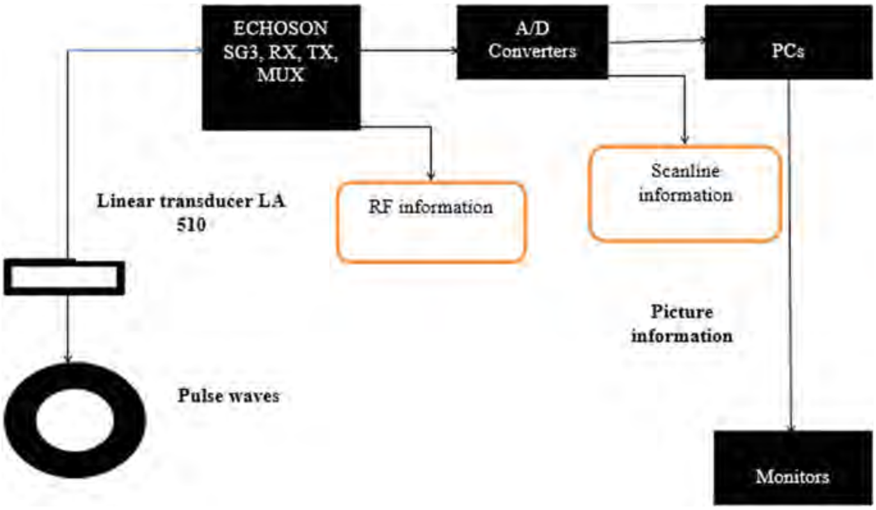


Fig. 3 Ultrasound image diagram

6 The Methodological Basics of Digital Processing of Pictures

Digitalized pictures are categorized in reference to their individual qualities such as signal-to-noise ratios, entropy, contrast, illumination ratio. In this case, a histogram provides a simpler method of processing pictures. With this method, pictures are displayed and do not change the quality of projected pictures. Grayscale histogram considers the most basic form of pictures which are therefore utilized to enhance the quality of the projected pictures. The method produces a scheme that indicates the values of the pixels and their various regions. As such, the Grey-level indicates if the images are either bright or shady. As for the mean pixel figure, it is possible to find it using the summation of the projected pixel values and the bin altitudes which are therefore subdivided by the definite base. The method is effective through the transformation of the histogram to being balanced, identical and smooth. Mean valuation of the centralized pixel intensity is effectively designated to an ideal form of brightness. Image intensity either below or above, makes all the pictures to be brighter or darker. The Signal-to-Noise Ratio (SNR) of the pictures is utilized to connect the various levels of the anticipated forms of signals to the various contextual levels of image signals [13]. SNR is considered as a ratio of signals and intensity to the intensity of noises. It calculates pictures in an up-front methodology. The intensity mean of pictures is calculated as the mean square of the pixel values, as seen in the expression below.

$$SNR = PSignal Pnoise \tag{2}$$

Whereby P is considered as the mean power.

6.1 Image Assessment Enhancement

The enhancement of pictures represents a technique utilized to enhance the perceptibility and quality of pictures through the application of computer-aided application software. The methodology incorporates both subjective and objective advancements. As such, the technique includes localized operations and points. In the localized operation, there are district input pixels and values. The enhancement of pictures is determined in two various forms: transform and spatial domain technologies. The spatial methodology operates more directly on the level of pixels whereas transformation methodology operates on Fourier and the spatial methods.

6.2 Image Assessment Segmentation

The segmentation of pictures represents the methodology for segregating pictures into various segments. The basic purpose of this form of segregation is to execute pictures in a manner that can easily be understood and interpreted while maintaining its quality. This methodology labels the pixels in reference to the characteristics and intensity. These segments signify complete original pictures and attain their features such as similarity and intensity. The segmentation of pictures and techniques are utilized in the creation of 3D contours in the structure of the body for medical purposes. The aspect of segmentation is utilized in the perception of machines, tissue volume analysis, malignant illness analysis, anatomic and functional evaluation, virtual reality, 3D rendered methods and anomaly evaluation in the detection of objects and pictures. The segmentation of pictures is grouped into Localised Segmentation and Globalized Segmentation. In the localized segmentation, there is a single subdivision of pictures. This methodology incorporates a few pixels that are contrasted to the globalized form. The globalized segment operates in the complete image as a single unit. This methodology includes more number of pixels that can effectively be manipulated. Segmentation is classified into methods: Boundary, Edge and Regional Methods.

6.3 Image Segmentation Considering a Threshold

The threshold segment is dependent on the value of thresholding to transform the grey colour to either white or black. There are various techniques functional in radiology to effectively replace and rebuild pictures like *k*-means and Otsu's methodologies. The threshold aspect is utilized for the establishment of borders of solidified objects in the dark foreground and background. The threshold technique is essential for the establishment of the solid object borders in the darkened background. These techniques require the availability of the variations between the backgrounds and object intensities [14]. Threshold techniques appear in three forms: histogram-built, adaptive and global-selection which is broader and utilized for the various segmentation methodologies. The globalized threshold Ø is calculated using the binary process according to the expression below.

$$f(x) = \{1\mathrm{iff}(m, n) \geq \text{Ø}o\,\mathrm{f}x = 1\,\mathrm{iff}\,mn \geq \text{Ø}o\} \tag{3}$$

In the fixed or adaptive threshold segments, an image is faster compared to the parts of interest and includes novel intensity that explains the variation from the image background. The limitation of this technique is its incapability to process or simplify multiple channels of pictures.

6.4 Image Segmentation Considering Edge Detection

Edge detection defines the technique of segmentation which utilizes the recognition of borders of strictly connected regions and objects. This method denotes the discontinuity of regions and objects. It is used critically in image evaluation and the identification of various image parts where a significant variation of image intensity happens.

6.4.1 Types of Edge Detection

Robert's Kernel

Robert's Kernel is a methodology utilized to determine the variation between two image pixels. Precisely known as forwarding variation and used to identify edges in highly noised pictures. It is evaluated using the first-order fraction derivative and the cross gradient operators as shown in the equation below.

$$\partial f \partial = f(i, j_{-f(i+1,j+1)\partial f \partial x} = fij - fi + 1j + 1 \tag{4}$$

$$\partial f \partial x = f(i, 1, j) - f(i, j+1)\partial f \partial x = fi + 1j - fij + 1 \tag{5}$$

The fraction derivative can also be applied in the two 2 × 2 matrices. In this condition, Robert's masks are evaluated, as shown in the equation below:

$$Gx = \{-1001\} \text{ and } Gy = \{01 - 10\}\; Gx = -1001 \text{ and } Gy = 0 - 110 \tag{6}$$

Prewitt Kernel

This methodology is centered on the ideology of centralized variation. This methodology is more effective compared to Robert's operators. Assuming that the matrices are categorically arranged in pixels, Eq. (7) will be produced.

$$\equiv a0a7a6a1[i, j]a5a2a3a4 \equiv a0a1a2a7[i, j]a3a6a5a4 \tag{7}$$

The fraction derivative of the Prewitt operators is expressed as shown in the equation below.

$$Gx = (a2 + ca3 + a4) - (a0 + ca7 + a6)Gx = a2 + ca3 + a4 - a0 + ca7 + a6 \tag{8}$$

In the above equation, c is a constant and shows the pixel encrypted at the image center. $GyGy$ and $GxGx$ represent the expression at $[I, j]$. Whenever c is equated with 1, the Prewitt operator is expressed as the equation below.

$$\begin{aligned} Gx &=\equiv -101 - 101 - 101 \equiv \\ &\text{and } Gy =\equiv -1 - 1 - 1000111 \\ &\equiv Gx = -1 - 1 - 1 - 111 \\ &\text{and } Gy = -101 - 101 - 101 \end{aligned} \tag{9}$$

Sobel Kernel

The Sobel Kernel method can be reliant on a centralized variation that acts on the centralized pixel in the image average. This method can be expressed as a 3 by 3 matrices, to an initial Gaussia Kernel derivative. This methodology is calculated as Eqs. (10), (11), and (12).

$$
\begin{aligned}
Gx &= (a2 + 2a3 + a4) - a0 + 2a7 + a6 \\
Gx &= a2 + 2a3 + a4 - a0 + 2a7 + a6
\end{aligned} \tag{10}
$$

$$
\begin{aligned}
Gy &= (a6 + 2a5 + a4) - (a0 + 2a1 + a2) \\
Gy &= a6 + 2a5 + a4 - a0 + 2a1 + a2
\end{aligned} \tag{11}
$$

The Sobel mask is expressed as shown below.

$$
\begin{aligned}
Gx &=\equiv -101 - 201 - 101 \\
&\equiv \text{and}\, Gy = -1 - 2 - 1000121 \\
&\equiv Gx = -1 - 2 - 1000111 \\
&\text{and}\, Gy = -101 - 202 - 101
\end{aligned} \tag{12}
$$

The Sobel model is considered effective compared to Prewitt in the reduction aspect of noise. This methodology is utilized in the functional aspect of image modality i.e. nuclear medicine [15]. In the research of red blood cells and pictures, the unveiling of near neighbouring cells is termed challenging as a result of background noises. This influences the interpretation procedure which makes it challenging to effectively diagnose physicians. The process of segmentation might also mitigate the challenges identified in the red blood cells.

7 Conclusion and Future Scope

The proposed work provided an analysis of medical image assessment based on the application of image processing methodologies. The pictures provide a method of expression of information in a pictographic manner. They consist of different minor elements known as image pixels whereby every pixel includes a certain value and position. Geometric pictures represent a picture arithmetically with the geometrical primitives like the lines. Every image has been saved in a certain file format that incorporates two segments namely, data and heading. Image processing methods represent a classification of techniques used in the process of image handling using a computer system. The main purpose of image segmentation describes the partition of pictures into essential image portions. Localized segments handle the partition of these pictures into minor segments in the individual image whereas globalized segmentation handles the assembly of the partitions. In the future, medical experts and researchers should focus more on applying artificial intelligence in the image assessment of body part parts. This technological aspect is useful for mitigating a wide range of image assessment problems and enhancing the delivery of patient results. Sophisticated software should be used to evaluate significant amounts of pictures and data to create algorithms based on image segmentation and processing. Artificial intelligence can effectively be applied to retrieve novel patient data and predict potential body disorders in patients. As a result, it allows radiologists to easily identify the body organisms that require further analysis.

References

1. K. Aoyagi, Medical image assessment apparatus, ultrasonic image assessment apparatus, magnetic resonance image assessment apparatus, medical image processing apparatus, and medical image processing method. J. Acoust. Soc. Am. **133**(5), 3220 (2013). https://doi.org/10.1121/1.4803793
2. B. Lelieveldt, N. Karssemeijer, Information processing in medical image assessment 2007. Med. Image Anal. **12**(6), 729–730 (2008). https://doi.org/10.1016/j.media.2008.03.005
3. C. Hung, Computational algorithms on medical image processing. Curr. Med. Image Assess. Former. Curr. Med. Image Assess. Rev. **16**(5), 467–468 (2020). https://doi.org/10.2174/157340561605200410144743
4. F. Aubry, V. Chameroy, R. Di Paola, A medical image object-oriented database with image processing and automatic reorganization capabilities. Comput. Med. Image Assess. Graph. **20**(4), 315–331 (1996). https://doi.org/10.1016/s0895-6111(96)00022-5
5. Y. Chen, E. Ginell, An analysis of medical informatics and application of computer-aided decision support framework. J. Med. Image Comput., (July) 10–17 (2020)
6. M. Heng Li, M. Yu Zhang, Computational benefits, limitations and techniques of parallel image processing. J. Med. Image Comput., (July) 1–9 (2020)
7. A. Khusein, U.A, Clinical decision support system for the activity of evidence based computation. J. Med. Image Comput., (September) 50–57 (2020)
8. P.L. Aaron, S. Bonni, An evaluation of wearable technological advancement in medical practices. J. Med. Image Comput., (September) 58–65 (2020)
9. K.K, P.E.P, Web based analysis of critical medical care technology. J. Med. Image Comput., (September) 66–73 (2020)
10. A. Haldorai, S. Anandakumar, image segmentation and the projections of graphic centred approaches in medical image processing. J. Med. Image Comput., (September) 74–81 (2020)
11. P. Jannin, J. Fitzpatrick, D. Hawkes, X. Pennec, R. Shahidl, M. Vannier, Validation of medical image processing in image-guided therapy. IEEE Trans. Med. Image Assess. **21**(12), 1445–1449 (2002). https://doi.org/10.1109/tmi.2002.806568
12. K. Drukker, Applied medical image processing, second edition: a basic course. J. Med. Image Assess. **1**(2), 029901 (2014). https://doi.org/10.1117/1.jmi.1.2.029901
13. P. Jannin, Validation in medical image processing: methodological issues for proper quantification of uncertainties. Curr. Med. Image Assess. Rev. **8**(4), 322–330 (2012). https://doi.org/10.2174/157340512803759785
14. M. Goris, Medical image acquisition and processing: clinical validation. Open J. Med. Image Assess. **04**(04), 205–209 (2014). https://doi.org/10.4236/ojmi.2014.44028
15. T. Aach, Digital image acquisition and processing in medical x-ray image assessment. J. Electron. Image Assess. **8**(1), 7 (1999). https://doi.org/10.1117/1.482680

An Analysis of Artificial Intelligence Clinical Decision-Making and Patient-Centric Framework

Anandakumar Haldorai and Arulmurugan Ramu

Abstract The smart decision-making support framework is typically referred to as artificial intelligence (AI). The clinical decision framework can transform the process of decision-making using various technologies. These technologies incorporate framework engineering and information technology. The vital centered ontology-centered automatic reasoning which is incorporated in machine learning methodologies has been established in the present patient databases. The approach evaluated in this paper is in the support of the interoperability between various health information systems (HIS). This has been evaluated in sampling implementations that link up to three separate databases: drug prescriptions guidelines, drug-to-drug interaction and patient information which are databases used to showcase the efficiency of an algorithm used to provide effective healthcare decisions. Generally, the possibility of artificial intelligence was evaluated in the process of supporting tasks that are essential for medical experts including the aspect of coping up with noisy and missing patient information and enhancing the utility of various healthcare datasets.

Keywords Artificial intelligence (AI) · Health information systems (HIS) · Medical decision support system (MDSS) · Intelligent decision support system (IDSS)

1 Introduction

The medical decision support system (MDSS) maps the data for patients in the process of determining diagnostic and treatment causeway. The technological segment of this system has been showcased in different medical settings which will be discussed in detail in this research paper. The aspect of smart decision-making is gradually expanding because of advances evident in artificial intelligence and

A. Haldorai (✉)
Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India

A. Ramu
Presidency University, Yelahanka, Bengaluru, Karnataka, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_62

system-centric ecosystems which have the capacity to deliver fundamental technologies required in decision-making. The process of coordination and communication between dispersed frameworks can be capable of delivering actual-time information, time processing, collaborative ecosystems and globally updated data to enhance the process of decision-making. During the duration, the technology of artificial intelligence and fundamental techniques has proved to be vital in the process of ensuring computational aid to users of practical applications. In this analysis, the system that can initiate the medical decisions based on the application of partial information about artificial intelligence. The properties of the information which includes heterogeneity, representation, interoperability and availability are significant in the process of assuring the application of MDSS [1]. The process of decision-making must apply various essential datasets from the various distributed frameworks other than one form of information source meant to maximize its efficiency. However, actual-time healthcare decisions are normally centered on incomplete data because of the potential issues posed by the features or properties of data synthesis.

Various artificial intelligence (AI) methods which include learning-centered methods and knowledge-centered methods have been applied to handle any form of data challenges which potentially creates a practical and robust MDSS. Even though the past methods have been reported to have partial significance, just a few of them have been deemed successful in the actual-world healthcare setting. The knowledge-centered system is vulnerable to losses when it comes to the evaluation of patients' performance since the data might be considered incomplete. For instance, the elimination of patient data or accessibility restrictions from visualizing healthcare records stops contrary to the decision of the learning-centered system which cannot be easily explained. There are some issues to differentiate the causation and correlation of these systems when making proper healthcare decisions [2]. The system analyzed in this research is a crucial format for the advantages of machine learning, logic-centered inferences and structural representation. Apart from that, the actual-world data which is essential for the provision of intelligent and robust decisions in the healthcare sector irrespective of the complexity over healthcare relationship and the dependency evident in healthcare decisions. Although the past form of machine learning has been deemed to have some disadvantages, our hybrid form of machine learning architecture provides decisions that can be explained and verified critically based on missing data. To evaluate the framework effectively, the raw patient data has been represented based on the application over ontological aspects replaced by structured triple stores. Inference guidelines have been established by the domain professionals and incorporated based on the application of semantic reasoning meant to develop proper decisions in the healthcare sector. As a result, the decisions produced by the framework are easy to explain and validate. However, the resultant knowledge-centered framework necessitates complete data that limited its effectiveness in the actual world.

The limitations are overcome through the administration of the semantic reasoning using machine learning methods that compute values of missing information. The imputation concepts are produced in the pre-processing stages before being integrated into the ontological framework, hence allowing the performance of the system during an actual time. In that case, patient-centric and evidence-centered decisions are made in the healthcare support system. The proof of the concepts and implementation system incorporates three vital sources of data, namely massive actual-world datasets for patient healthcare data, medicinal interaction registry and the gathering of healthcare prescription protocols [3]. The preliminary findings evaluated in this paper ascertain that practical healthcare cases, where the data from patients are incomplete or missing, are based on the knowledge-centered methods and the hybrid design system dependent on machine learning.

The process of decision-making in the healthcare sector is as critical as a human procedure in the process of interacting with patients in the actual world. The healthcare practitioners initiate both poor and good decisions which make it essential for researchers to debate the most applicable means of supporting the healthcare sector in the process of making effective decisions. One of the best ways of characterizing decisions is to focus on comprehending the manner in which healthcare practitioners will be assisted to effectively categorize decisions as unstructured, structured or semi-structured. Structural decision issues are known to have a specific optimal remedy which means that they do not require a specifically designed decision framework. For instance, an exact remedy can be applied by practitioners to effectively make a proper decision regarding the most resized causeway between two points as for the unstructured decision issues. There are no certain solutions or criteria to apply which means that the kind of preferences considered is entirely upon the decision-makers. For instance, determining an individual's partner might be deemed to be an unstructured decision. Visualizing between the two forms of problems, it is possible to encounter a wide range of semi-structured issues that are known for their acknowledged parameters that might as well require the preference and input of humans in the process of making proper decisions using a specific method. For instance, an organizational decision concerning the choice of expanding the organization to an international market might be considered as a semi-structured decision. Semi-structured decision issues are therefore amenable to the decision-making framework that is attached to the interconnection of user's interaction with the analytical techniques meant to develop alternatives concerning the optimal remedies and methodologies. When artificial intelligence methods are applicable in the enhancement of fundamental healthcare alternatives, the resultant system is considered as an IDSS. Data analysts have also reminded us that comprehensive acknowledgment of decisions is required for the efficient utility of artificial intelligence in the healthcare sector. Moreover, AI is focused on limiting human decisions to develop them in some capacity and fundamental advances.

With that regard, this paper will focus on decision-making protocols and DSS which are centered on the comprehension and concomitant applications of artificial intelligence methodologies to establish firm IDSS. Several literature assumptions have been made about the human decision-making process which critically recognizes and explains the keepers in the process of decision-making. In the DSS, effective decisions are featured by an element of critical reasoning based on the application of distinctive human characteristics for selecting optimal alternatives based on a specific criterion. Many forms of reasoning are being showcased through analytical methodologies, and in that regard, it might be embedded into the system of intelligent decision support system (IDSS). Actually, not all segments of decision-making frameworks are embedded in the IDSS system [4]. In some other segments, the stimuli of recognition alternative might result in a specific decision action as a mastered response without any form of identifiable reasoning framework. The process of comprehending the decisions in response to some healthcare systems does not entirely focus on the aspect of reasoning. Several research findings done by researchers have concluded that decision-making is mostly initiated by emergency response departments or healthcare emergency services.

2 Literature Review

The literature works focus on application open matching conditions, where the experience of human beings and their immediate responses are required to enhance effective decisions in the healthcare sector. In such scenarios, the decision support in case applicable should provide essential data that could focus on the critical element of human processing for decision-making. Physiologically, the capability to enhance the process of decision-making is termed focus on the preferred lobes in the brain system, where critical decisions are made. Any faults in this area amount to irrational decisions and ineffective evaluation of risks. The process of decision-making is considered to be affected by emotions inscribed in the neural system which might be both unconscious and conscious. In the most recent analysis, IDSS has shunned the capacity to model effective features like the emotions required to enhance the process of decision-making, even though effective inclusion of such emotions in machine learning is a future concern. Some researchers have also pointed out that operating the memory for cognitive functions is essential in the process of decision-making. The intelligent data processing frameworks such as an IDSS harbor stone memory technologies, symbol reasoning, and the capability to interpret and capture stimuli, IDSS has the fundamental capacity to copy the decisions made by humans [5].

2.1 System Evaluation and Implementation

The system of human decision-making underpins fundamental frameworks that have been proposed for IDSS and DSS. An initial framework for decision-making has also been offered by several researchers which are mostly referred to as the subjective expected utility framework or the expected utility framework. Researchers assume that in case decision-makers focus on several outcomes preferences or steps when satisfying a certain postulated decision, then the decision-makers can lead the functionality of outcomes being maximized for a specific form of decision probability.

The researcher's theory typically is utilized to suggest actions and decisions that will be maximized in the application of decisions provided that probabilities have been provided for specific events.

In that case, the researcher's advanced framework of decision-making has some uncertainties. There are a lot of criticisms about such theories; certainly, the conclusion that decision-makers might evaluate critical consequences of their actions acquires knowledge about the future happenings without any form of connected probabilities. Nonetheless, one of the contributions of the theory was to enhance the separation of actions, outcomes and events. To effectively evaluate the recommended system formulated a proof of concepts and implementation concentrated on the knowledge management wiry and component execution from the present ontological decision systems. The insomnia is cured by selecting the channel of inquiry and utilized some actual-world datasets.

1. The records of patients extracted from the center of disease control are based on the behavioral risk factors (BRF) and surveillance system for 2010. The behavioral risk factors include a wide assumption of respondent data such as location sex, rest and age including information about medical statuses such as asthma, diabetes, mental illness and cancer. Several behavioral risk elements including the consumption of alcohol, sleep deprivation and drug usage were evaluated. All the crucial data stored and recorded were done in our structured format in a relation and database.
2. The application protocols extracted from the Mayo clinic were also utilized as professional decision-making protocols that correspond to the prescription protocols for various medical issues such as sleep disorders.
3. Drug interaction registration was done to identify the interaction of drug-to-drug medical application. Practitioners use an ontological graph based on the record and information of patients. The figures included in the ontological framework represent the vital concept used in the evaluation of the patient relationships. The figures also demonstrated the application of inference protocol to effectively map raw information into ontological concepts.

2.2 Skill Management Concept

Researchers argue that to effectively instantiate the skill management concept in the designing of the decision-making system, it is critical to establish a simplified ontological model to define essential key concepts and different patient relationships. The formulated inference protocols about the BRF codes that are defined by the semantic values of various data attributes in the process of transforming numerical BRF information into the concept of patients. The protocols have been applied in the process of recording data to formulate semantic knowledge storage of BRF data.

2.3 Query Execution Concept

To effectively instantiate the query execution concept, researchers have linked a semantic reasoning segment known as "Euler Proof Mechanism" together with the WEKA machine learning system to establish the execution component. Semantic form of reasoning has provided a critical mechanism element for the logical best decision-making aspect in the system, where WEKA has enhanced the process of supporting the mandate to impute missing information. A subset of sleeping aids was identified for the researchers to be applied in the Mayo clinic protocol to effectively establish the condition that should be focused on when prescribing medicines.

Utilizing the ontological framework, this data has been transformed into inference protocol for the process of decision-making. A family physician from proximity aided in the selection of different drugs which have been validated in the Mayo clinic sleeping aid prescription framework [6]. Even though the inference protocol has been maintained as simple, this does not reflect the actual healthcare consideration for the sleeping aid prescription. The generics segment of the resultant protocol has been explained below. However, several interactions have been verified by global healthcare practitioners. These include:

1. Drug-to-drug interaction protocol: In case patients are presently on a specific drug D1 and that drug cannot be provided with the drug D2, patients cannot be provided with D2.
2. Drug-to-condition interaction protocol: In case patients are reported with a specific medical condition C, under specific drug D reported to contradict a condition, patients cannot be provided with drug D.
3. Drug–to-disease interaction protocol: In case patients are reported with the illness E and drug D is reported to have some contradictions to the disease, patients might not be given drug D.

To effectively mitigate the problem of missing values in the records of patients, the formulated classifiers based on the application of machine learning in the process of predicting values for the missing information fields. Apart from that, these classifiers were trained effectively to predict every attribute considering all the complete information from the BRS datasets in the process of training. For instance, in case of the sleeping aid, Estazolam is not capable of prescribing elderly patients healthcare practitioners need to determine the ages of patients. To focus on this principally, all the codes for the patients were extracted in the BRF set, where the ages have been confirmed before partitioning the subsets into training data and validated data [7]. The classifiers are then built based on the application of validated data which can be used to provide an estimated valuation of the classifiers for the process of initiating decision-making for future works. In the future instances, where the age of patients has been missing, practitioners will apply classifiers to effectively label the ages of patients as young or old. The projected value is then substituted into the records of patients based on semantic reasoning which is done progressively. The decision-making confidence through semantic reasoning is centered on the point estimates evaluated in this article.

3 Experiment Evaluation and Comparison

To effectively evaluate the operation of the system, the proposed hybrid decision-making framework was reviewed and experimented. Patients who are provided with sleeping aid at the Mayo clinic prescription center have been provided whether positive exemplars, such as labels against the negative exemplars. Whenever a specific system has been labeled, patients can effectively respond to specific queries that are either true negative (TN) or true positive (TP). Moreover, the false negative (FN) and false positive (FP) are some of the cases produced by the system. The findings are evaluated as follows:

(a) Sensitivity: This applies to the rate of positive exemplars that have been labeled as positive.
(b) Specificity: This applies to the rates of negative exemplars that have been labeled as negative.
(c) Balance form of accuracy: This principle computes simple averages of sensitivity and specificity as shown in Eq. 1.

$$\text{Specificity} = \frac{tn}{tn + fn}\text{Sensitivity} = \frac{tp}{tp + fn}\text{BalAcc} = \frac{\text{Specifity}}{\text{Sensitivity}} \tag{1}$$

3.1 Machine Learning-Centered Framework

To effectiveness of our hybrid framework focused against the learning-centered framework, and it used to evaluate the condition and performance of various machine learning algorithms of the BRF datasets. Using the four algorithms AdaBoost, bagging, C4.5-R8 and decision stamp, it was possible to formulate 50 various random chosen training sets with various sizes of 2500–5000 exemplars. This data was utilized to evaluate the framework based on the elementary selection of algorithms to minimize different sets for data attributes. For every algorithm, a predictive classifier were trained and the model for each sleeping aid in this case which is meant to effectively predict if the patients can be provided with sleeping aid or not. The patients have been trained to execute outputs of skill-centered systems concerning the datasets of patients.

The fundamental truth concerning every record of data based on the application of semantic reasoning since skill-centered decisions can be considered as 100% accurate compared to predictive accuracy for every ML algorithm. The AdaBoost algorithm is known for its best performance over the entire four algorithms. In that case, a system of algorithms is compared to the AdaBoost-derived classifier. In this case, the general accuracy of the algorithm when projected with the most precise medical decision is minimal from the scale of zero to one which is just the same as the learning-centered system, projecting that there might be some forms of missing data in the system.

Other than the poor form of performance in the machine learning segment, it can be projected and be tolerant of missing information. The implication of the missing data concerning the performance of ML through the process of removing known figures from the records of patients was analyzed [8]. Moreover, the £ as the average figure of the attributed values was evaluated to eliminate the patients' records and varied £ from the average of 1 which is the eliminated value in a single record of 6 that has been removed in each record. As for every £ record, analysts have trained the AdaBoost-centered classifiers based on the application of 50 sets of five thousand exemplars from the partially—DOI information. As a result, the implication of the missing £ was evaluated concerning the performance of the algorithm based on the aspect of machine learning.

3.2 Skill-Centered and Hybrid-Centered Systems

Finally, a comparison is performed to measure the hybrid frameworks with the machine learning systems, including the skill-centered system which is free from the imputation capacity. The process of the skill-centered reasoning is performed using Eulersharp, and the algorithm-based classifier is used for the process of machine learning and the evaluation of critical elements. There are four datasets are chosen in the process of evaluating the critical values of £. For every £, we evaluated and measured the skill-centered decision process and trained the algorithm-based classifier to effectively predict the values of patients projected through the process of machine learning. The semantic form of reasoning is therefore re-evaluated before initiating any forms of decisions in the system. Throughout this process, the hybrid decision-making process experiences some form of performance degradation in the process of balancing the level of accuracy since £ increases (which represents an increment of £ to affect the decreasing performance that is less than 1% a point) [9]. Nonetheless, the performance of the skill-centered decision-making framework model degrades with the same range as £ typically increases by 0.5 in £ due to the diminishing performance of approximately 4%. Generally, the hybrid framework attains a significant form of balancing accuracy which shows that its recommendation for the healthcare decision-making process can be considered to be effective.

3.3 Standardized Imputation Techniques

The evaluation of datasets with unverified or missing information is a common issue that is globally studied in the field of statistics. In this field, multiple imputation (MI) methods are normally applied. During the process of performing MI, every unverified or missing data is imputed several times through the process of drawing featured values from the predictive form of distribution. As such, this amounts to the collection of imputed sets of information. Every imputed dataset in a single shape is considered as an original form of datasets, whereas its non-missing figures are identical to the original forms of information since the missing and unverified datasets are computed differently for every imputed version. The grouped imputed datasets might be utilized to project unbiased estimations of summarized statistics such as regression and means coefficients, including the statistically verified confined datasets in the field of statistics. It should be noted that the main obligation of producing an accurate summary of these datasets might vary from our mission of accurately projecting the unverified and missing data of individuals in the information set.

However, for completeness, this article has evaluated the application of MI in the decision-making system. Here, the famous MI method, namely Bayesian MI is known to assume a particular joint probability framework over some form of feature values before projecting imputed information off the posterior form of distribution of these sets of missing information for the observed sets of data [10]. This methodology was applicable in the medical survey evaluation over the past few decades. The mix opening is used as the sourced package for the R software ecosystem to effectively test the off-the-shelf capacity of these techniques. The mixture is composed of several limitations which affect the overall performance: It might not be capable of using more than 30 features considering that BRF is composed of more than 400 features, it operates generally slow, and it might not be capable of using the features characterized by massive data missing in the set. To effectively test the mix, it can be decided to choose features to structure imputed dataset.

It should be noted that due to the modeling assumption which is inherent in the Bayesian MI, it might be difficult to facilitate the separation element of selection during the process of predicting various features. It is possible to execute the mix on various dataset portions, and the procedure is not considered that straightforward since it had multiple flaws in translation between various formats of datasets. As a result, it was difficult to effectively explain the implications of utilizing the mixes for imputation in the ecosystem, other than noting that this mix might be projected in the future. As such, future researches might focus on the evaluation of the constructions of effective issue-specific versions of these mixes which are essential for application in a secure decision-making ecosystem. Moreover, our projected hybrid framework provides substantial performance merits over other alternatives in the absence of unverified and missing datasets in machine learning frameworks and the availability of skill-based systems in MI. Consequent to that, the framework shows a robust remedy to the issues of partially unverified and missing data in the decision-making systems DOI in the healthcare sector.

4 Discussion

In this article, the actual-world issue was analyzed such as helping medical practitioners to execute proper decisions about the present patients' data and the application of best practices encoded in the protocol-based case on unverified and missing data. Healthcare experts consider this issue to be a prevailing concern for solutions to be adopted (especially with patients misrepresenting and omitting their present medical profiles). As a result, the AI methodologies can be a significant advantage to effectively address this prevailing concern to categorically yield accurate healthcare advice effectively for patients dependent on traditional probability reasoning in isolation and delivery of tasks.

To initiate hybrid construction, a certain hybrid construction framework was evaluated and presented for healthcare decision support engines. Our projected framework is capable of processing the queries of users critically based on the application of logic-centered reasoning and utilizing ML inference frameworks to handle unverified and missing data. This technique has distinctive merits with different findings depending on the end users which might also be verified for correctness by third parties because answers are focused on logic assumption. Even though our validation approach utilized a certain sleeping aid prescription case, the system is more generic to be applied in other healthcare facilities.

To effectively build a remedy around this issue serving various issue domains, an issue specific to the ontological framework for information representation has to be evaluated, including the professional inference protocols for the processes of decision-making. After that, the ML algorithm which operates effectively with certain datasets can be utilized to predict unverified and missing data direct from raw information. Once the founding primitives have been evaluated, the framework construction can be considered uniform to the kind projected in this paper. In other related works, there is a significant deal of interest in the application of ML methodologies for medical decision-making support framework. For this aim, other methodologies are ontological to our own. These works project a comparative assumption of two ML methods over the present decision-making procedures based on the application of the medical assessment protocol [11]. Their findings show definite merit of utilizing ML algorithms. Nonetheless, they have showcased that ML methods alone presented a significant amount of false negatives and false positives.

4.1 The Decision Support System (DSS)

DSS refers to the wide-range interactive computing frameworks which aid decision-makers and experts to use models, knowledge and data to mitigate structured, semi-structured and unstructured issues. The individuals initiating the decisions are a part of the system. In that case, DSS incorporates the capacity to permit these decision-makers to choose single or multiple input selections, drilling explanations, querying systems, examining outputs and interacting with the various networking devices. Since a lot of DSS systems have been established to mitigate a certain issue or a specific segment of issues, there are various forms of DSS that have been specialized for various forms of problems and users.

The system can also be formulated for a single or multiple decision-makers which might be applicable in the process of supporting crucial decisions made in the range of creative to managerial problem-solving. Different terms have been incorporated to DSS such as group DSS, collaborated DSS, expert DSS, medical DSS, intelligent DSS and adaptive DSS among others in the process of attempting to capture the merits of DSS or the individuals requiring it. The single-user DSS is typically considered as a procedure of making proper decisions and incorporates components for designing, imputing and choosing specific decisions. The collaborative or grouped DSS is still in their early development because theories of collaborative decision-making processes for humans are still an emergent issue. As a matter of fact, DSS to support idea generation and system innovation is still a prevailing concern with fundamental theories of human creativity still in progress.

The schematic of a typical DSS structure shown in Fig. 1 incorporates the decision-making process as a key segment of the system. Inputs incorporate the skill base, model base and database aspect. In the database, there is data that is essential for the process of decision-making, whereas the skill base incorporates the guidance required in the selection of proper alternatives. The model base considers the formal framework, techniques and algorithms that possibly establish potential outcomes and identifies the most effective remedies under a specific constraint. The responses from the processing stage might showcase more inputs which are typically updated in actual time to enhance the process of solving potential issues. Outputs might also be generated to project explanations and forecasts meant to justify crucial recommendations that project potential advice. As such, results can be given to decision-makers that possibly use the system to project essential data or queries. Over the past few decades, the terminology "decision support" has been used widely to incorporate other decision-making assistance aspects such as skill management systems, organizational intelligence and data analytics which consider the essential interaction of decision-makers. AI characteristics are typically utilized and initiate observations from a massive, distributed aspect of big data. These frameworks can therefore incorporate the personalization of decision-makers and preferences which emulate the decision of humans. Moreover, the frameworks project firm novel tools meant to deal with complex issues and emergent trends for the future of AI.

4.2 Evaluation of IDSS

To effectively evaluate IDSS, or typically any form of DSS, it is essential to comprehend the opportunities and advantages for boosting the performance of the system. In this attempt, a primary assumption of the research concerning the fundamental purpose of decision-making was evaluated and presented in a specific healthcare

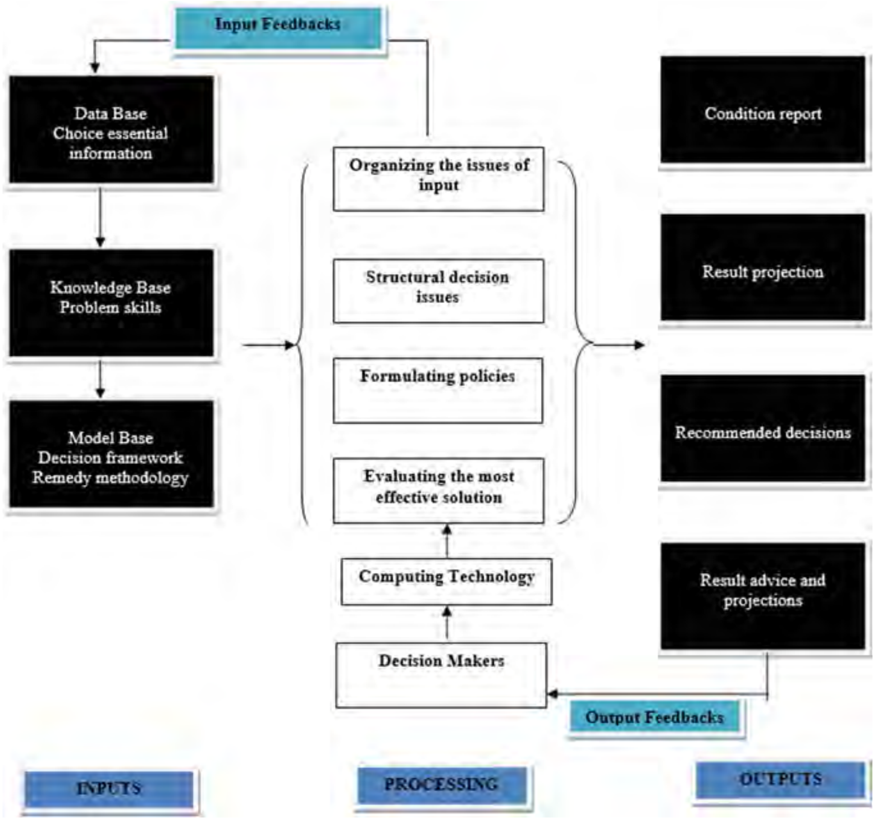


Fig. 1 Arrangement of DSS

facility. According to the literature research, one system feedback criterion to evaluate the DSS success is based on the enhancement of speed and boosting the efficiency aspect of decision-making. The potential outcomes might be linked to tangible advantages such as the reduction in revenue and the increment in costs. Nonetheless, a closer evaluation of results showed that the process feedback has also been considered as the merits of IDSS and DSS. Moreover, according to the theory of Simson, the procedure of decision-making includes implementation, choice, design and intelligence [12]. The system of IDSS might, for instance, perform database evaluation and assessment in reference since the systems perceive the requirement of decision-makers or aid the users in the process of choosing proper variables in the designing phase shown in Fig. 1.

As for the instance, where the general outcomes are not transformed, the decision-makers have an effective comprehensive ability of the decision-making issue based on the application of IDSS. In that case, the evaluation of IDSS and DSS might be considered as a multiple format and criterion, based on its assessment to improve design and guide decision-makers on how systems should be utilized. Multiple criteria assessment of data systems has to be evaluated further in the future works since several DSS and IDSS have been deemed to report some essential challenges. The users of these systems follow the application of trade-offs based on the performance and goals of utilizing multiple programming techniques. Several researchers have applied these approaches to DSS and professional systems with multiple criterion approaches that include technical, empirical and subjective techniques.

The subjective methods incorporated sponsors and users' perspectives, technical methods, analytical techniques and the effectiveness of the system while the criterion method incorporated the performance of the system with or without the involvement of humans. Apart from that, other researchers have evaluated numerous data systems with two essential categories of performance measurement aspects: efficiency and effectiveness. The three crucial methodologies and criteria that are evaluated by other researchers are as follows: the overall system protocol meant to assess how output adds to the accomplishment of fundamental organizational goals, effectiveness to evaluate the efficiency of resources, inputs in attaining outputs and efficacy in the process of evaluating how systems produce the potential results. In this case, efficacy represents the value-centered perspective.

4.3 The Intelligent Decision Support System (IDSS)

The intelligent DSS typically known as IDSS makes use of AI methodologies to effectively improve and enhance the process of decision-making in the healthcare sector. Tools used in AI include fuzzy logic, evolutionary computing, case-based reasoning, intelligent agents and artificial neural network. All these tools when incorporated with DSS are a powerful aid for the healthcare sector and can be assured to categorically solve problems of massive datasets and complex reasoning in actual time [13]. IDSS is gradually turning up to be significantly important and also the practical application since it applies essential AI methods. The applications range from the medical support system to initiating proper decisions with the capacity to enhance proper decisions.

Researchers have pointed out that DSS influence both the outcomes and processes of decision-making, where DSS applications are based on varied sources of systems that potentially evaluate the process of decision-making. The process criterion enhances how decisions are measured and executed in a qualitative approach to boost the results and speed of decisions. The outcome criterion has been assessed based on quantifiable aspects such as the decrement in costs, enhanced profits, prediction accuracy and prediction success or failure. In that case, multiple procedures of IDSS have been justified both in practice and theory [14]. As such, IDSS can be evaluated and assessed based on the outcomes and processes of decision-making. The most crucial manner to develop the quantitative framework for the decision values in the system is to apply the analytical hierarchy process (AHP) [15].

AHP is advantageous and allows the individuals to add on various components that are acceptable by the system. The stochastic development of AHP is critical and permits the statistical merit of contributions which can be applied by the system. Apart from that, AHP projects a methodology that will be used to compare various alternatives through the process of structuring methodologies into hierarchical system relevance for the process of making a proper decision [16]. The system hierarchy is formulated into a hierarchy essential to the system and is broken into levels as a reporting structure in an organization. The assessor typically requires supplying paired comparisons of these alternatives at a low level, where AHP is capable of computing intermediate comparisons before combining them to a specific decision value comparing different alternatives in this level. The criterion has been measured to project an eigenvalue remedy which will be utilized in the execution of judgment, where two alternatives are fundamental. AHP is therefore essential in the process of mitigating decision issues which can be applied in evaluating IDSS [17].

5 Conclusion and Future Scope

The emergent research in the use of artificial intelligence is to initiate decisions in the healthcare sector, and smart adaptive frameworks are meant to deal with complex issues. The systems are linked to the preferences of the decision-makers and interfaces of virtual humans to facilitate effective interaction of both machines and humans. Throughout the process of making critical decisions, techniques are utilized for difficult issues such as big data. Machine learning is integrated with the ontological reasoning for the inherent merits of decision-making approaches to provide accuracy for medical domains. The proposed system operates with the actual-world datasets and protocols. This system can be typically applied to any medical context to make decisions. Since the IDSS is considered to have some prevailing issues for users, further research on the same is fundamental. Future works should focus on the projection of an effective interface for users of DSS with the projection of advancing from the emergent application to the kind that is deployed. These works in smart decision-making advancements project to the fundamental challenges and opportunities in the medical field. Opportunities are viewed in the enhanced process of decision-making which is typically essential for handling complex issues that are beyond the capacity of humans to perceive different variables vibrant in nature. On the other hand, challenges are considered for the wide-range IDSS in the process of designing systems that have clear returns on investment and can interact with humans to gain trust in system usage and deal with potential issues in the system. As a result, in the future, IDSS can lead to a novel wave of a wide and sophisticated framework of decision-making.

References

1. C. Bennett, K. Hauser, Artificial intelligence framework for simulating clinical decision-making: a Markov decision process approach. Artif. Intell. Med. **57**(1), 9–19 (2013). https://doi.org/10.1016/j.artmed.2012.12.003
2. P. Lucas, Dealing with medical knowledge: Computers in clinical decision making. Artif. Intell. Med. **8**(6), 579–580 (1996). https://doi.org/10.1016/s0933-3657(97)83108-9
3. W. Horn, Artificial intelligence in medicine and medical decision-making Europe. Artif. Intell. Med. **20**(1), 1–3 (2000). https://doi.org/10.1016/s0933-3657(00)00049-x
4. P.L. Aaron, S. Bonni, An evaluation of wearable technological advancement in medical practices. J. Med. Image Comput. 58–65 (2020)
5. Web based analysis of critical medical care technology. J. Med. Image Comput. 66–73 (2020)
6. A. Haldorai, S. Anandakumar, Image segmentation and the projections of graphic centered approaches in medical image processing. J. Med. Image Comput. 74–81 (2020)
7. I. Rábová, V. Konečný, A. Matiášová, Decision making with support of artificial intelligence. Agric. Econ. (Zemědělská Ekonomika) **51**(9), 385–388 (2012). https://doi.org/10.17221/5124-agricecon
8. R. Yager, Generalized regret based decision making. Eng. Appl. Artif. Intell. **65**, 400–405 (2017). https://doi.org/10.1016/j.engappai.2017.08.001
9. N.M. Hewahi, A hybrid architecture for a decision making system. J. Artif. Intell. **2**(2), 73–80 (2009). https://doi.org/10.3923/jai.2009.73.80
10. C. Gonzales, P. Perny, J. Dubus, Decision making with multiple objectives using GAI networks. Artif. Intell. **175**(7–8), 1153–1179 (2011). https://doi.org/10.1016/j.artint.2010.11.020
11. Y. Chen, E. Ginell, An analysis of medical informatics and application of computer-aided decision support framework. J. Med. Image Comput. 10–17 (2020)
12. M. Heng Li, M. Yu Zhang, Computational benefits, limitations and techniques of parallel image processing. J. Med. Image Comput. 1–9 (2020)
13. R. Degani, G. Bortolan, Fuzzy decision-making in electrocardiography. Artif. Intell. Med. **1**(2), 87–91 (1989). https://doi.org/10.1016/0933-3657(89)90020-1
14. P. Giang, P. Shenoy, Decision making on the sole basis of statistical likelihood. Artif. Intell. **165**(2), 137–163 (2005). https://doi.org/10.1016/j.artint.2005.03.004
15. D. McSherry, Conversational case-based reasoning in medical decision making. Artif. Intell. Med. **52**(2), 59–66 (2011). https://doi.org/10.1016/j.artmed.2011.04.007
16. T. Leong, Multiple perspective dynamic decision making. Artif. Intell. **105**(1–2), 209–261 (1998). https://doi.org/10.1016/s0004-3702(98)00082-4
17. A. Khusein, Clinical decision support system for the activity of evidence based computation. J. Med. Image Comput. 50–57 (2020)

A Critical Review of the Intelligent Computing Methods for the Identification of the Sleeping Disorders

Anandakumar Haldorai and Arulmurugan Ramu

Abstract The intelligence computing techniques and the knowledge-centered systems are considered in the process of identifying the different complications in a clinical setting. In this article, a critical review of the different intelligent computing techniques, which are utilized in the detection of the sleeping disorders, will be analyzed. The core issue in this contribution is centered on the identification of the sleeping disorders such as snoring, parasomnia, insomnia, and sleep apnea. The mostly used diagnostic techniques by medical researchers are centered on the knowledge-based systems (KBSs), rule-based reasoning (RBR), the fuzzy logic (FL), case-based reasoning (CBR), artificial neural networks (ANNs), multi-layered perceptron (MLP), genetic algorithm (GA), neural networks (NNs), k-nearest neighbor (K-NN), data mining (DM), Bayesian network (BN), and the support vector machine (SVM), including other many methods integrated with the medical sector. As for the ancient methods, questionnaires are utilized for the identification of different forms of disorders, which have now been replaced with the above methods. This is meant to enhance sensitivity, specificity, and accuracy.

Keywords Knowledge-based systems (KBSs) · Rule-based reasoning (RBR) · The fuzzy logic (FL) · Case-based reasoning (CBR) · Artificial neural networks (ANNs) · Multi-layered perceptron (MLP)

1 Introduction

In the field of ancient neuroscience, sleep is considered a fundamental therapeutic factor. This field becomes considerably essential because of the prevailing familiarity among humans. The research done in [1] has shown that approximately 40% of the considered medical themes had about a single sign and syndrome that was projected

A. Haldorai (✉)
Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India

A. Ramu
Presidency University, Yelahanka, Bengaluru, Karnataka, India

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_63

to be disrupted by sleep. The young people mentioned that adults (i.e., one out of five adults) are normally distressed from sleepiness during the day and in that matter, narcolepsy is an issue due to extreme sleep during the day. The two forms of disorders include a vital, but confusing effect posed on regular activities of the day. Like that, sleep disturbances are formed by sleep-connected breathing disorders, which are known as sleep apnea. Moreover, the sleep disorders have a long duration and concise duration of dreadful factors. The short duration of effects that are more direct to impaired sleep attention potentially affect the quality of human life, which increases the chances of mishaps. The lengthy effects of the sleep distress shift over to the increment of the mortality and morbidity rate from the increment of mishaps, high blood pressure, cardiovascular illnesses, learning disabilities, and the bulkiness alongside the human discouragement.

Some disorders of sleep, according to the researchers in [2], are considered severe enough and hamper the psychological, cognitive, physical and motor functioning of a person. Normally, humans pay minimal attention to sleep apnea, snoring, parasomnia, and insomnia, but the persistence of these disorders can be considered to be incredibly serious. In that regard, early identification of these issues becomes a fundamental task. In the initial stages of identifying these disorders, the medical researchers utilize the quantitative methods in the manner of the questionnaire for the identification of the different sleeping disorders. However, the accuracy evaluations are always an issue in this method due to the complete questionnaire review being dependent on the complete number of medical participants, including the questions designed for the medical survey. In that case, the researchers shift toward the enhancement of the various forms of accuracy based on the application of different intelligent methods. Artificial intelligence (AI) is vital when applied in these methods. The intelligent computing methods such as BN, DM, MLP, NN, SVM, GA, FL, and ANN are all information-controlling techniques, other than knowledge. However, a lot of researchers have implemented different integration methods of knowledge controlling, which are applied in the clinical domain. The basic issue-solving paradigms in the segment of AI are CBR and RBR.

The researchers in [3] projected a more rigid protocol for the patients' sleep macro-structure specification that is mutated recently by the Academy of Sleeping and Medicine in the SA. Human snoring is considered extravagant and tedious. In that case, the frequent attempts have been structured to formulate such frameworks, which take the count of records in a more automated manner. The R-K method utilized in the construction of the rule-centered sleep staging framework is based on the application of the multi-rule decision tree. The tree framework and technique were utilized in the improvement of the accuracy over one decision tree system and was considered that the multiple decision tree system has about 7% form of accuracy over one model. The RBR techniques are also time consuming as they necessitate signal data, identification of certain patterns such as k-complexes, rapid eye shifts in EOG and the sleep spindles in EEG.

Canisius utilized the bio-signal processing algorithm for the identification of the sleep problems that are based on the ECG signals with an approximate accuracy of about 77%. A lot of time is essential to formulate a framework from RBR that retrieves the elements from the initial dataset before constructing the rules according to the human brain system whereas in opposite classification numerical does not require any complicated rule-centered features extracted by the power spectrum. The chromosomes with variable structures and fitness elements are utilized in the process of finding optimal input elements and the network recognition specifications. In numerical categorization techniques, no human knowledge and protocols are essential. The improvements of the portable microcontrollers based on the medical devices are effectively enhanced for home monitoring for long-term monitoring. This form of the medical device might potentially provide different forms of outputs as overall snoring count, medium numbers of snores in every hour, and irregular numbers of snoring. The success aspect of the devices was about 85% in a laboratory ecosystem and about 70% in the home-based [4]. CBR is capable of utilizing the exact knowledge of the experienced and confirmed issue cases. It considerably favors learning based on experience, since this is easier to master from the real issue-solving experience.

The first framework, which is the case-centered reasoning, was considered as the CYRUS framework that was formulated by researchers in [5]. CYRUS was a showcasing of Schank's dynamic aspect of the memory framework. As a result of the contrasted merit and demerit of CBR and RBR, it is considerably challenging to mitigate the issue independently. However, in case their advantages are noted and the demerits are ignored, their various junctions present more considerable advantages such as BOLERO and MIKAS. These are some of the systems that can be integrated with CASEY, PROTOS, CBR, and RBR. MBR and CBR are meant to integrate GREBE, MBR, CBR, and RBR with a knowledge base that includes the cases and rules connected with the kind of laws for the injuries to the workforce. The transformation of the hidden skill set into the more precise protocols would amount to the irregularity and loss of knowledgeable contents.

An alternative of these forms of inferences in Baye's theory, which fit the probabilistic figure for the measured outputs such as MES and ES, was analyzed in 2007. SAMOA represents an automated sleeping apnea sign diagnostic framework. All of these systems are effective for the special illnesses and the self-governing signs but can also fail when someone is having more than a single system or disease, which gives a disease a reason to advance. Moreover, the rule-centered expert framework has two potential demerits:

(a) All situations might not be illustrated by protocols for different conditions
(b) Experiences collected by errors and trials might not be contained by the knowledgebase with no human input. As a result of this, low forms of agreements are reported at a rate of 83% [6].

In that case, ANNs have entertained to apply some form of human intelligence. ANNs have considerably been utilized and acknowledged as a method for the treatment and diagnosis of sleep disorders at the various stages of the human developmental stages. In some instances, GA is utilized to evaluate the neurons number of the potential hidden layers. There is a difficulty in the advancement of the automatic framework as a result of several uncertainties that arose as issues arise day in day out. To mitigate this form of problem, FL is analyzed and applied as a suitable framework such as the ISSSC version of the 1.0 system that is applicable for the clinical treatment and diagnosis. DM is considered as a proficient technique and tool for the creation of novel knowledge set from the wide-range databases. Different techniques of DM are termed in the collective diagnosis measure of wide-range illnesses such as the prediction protocol for the obstructive disorder known as sleep apnea. This research is centered on the analysis of the various techniques of the diagnosis and detection of sleep illnesses such as snoring, parasomnia, insomnia, and apnea [7].

Various techniques centered on the wide-range intelligent computing methods and their connections such as DM, FL, GA, ANN, CBR, and RBR/KBS were discussed. The connected techniques are GA-FL, ANN-DM, ANN-GA, ANN-BN, ANN-FL, RBR-ANN, CBR-FL, and RBR-CBR [8]. The DM techniques and the proficient systems projected for the identification of sleeping disorders have widely been evaluated in this research article. In the wide research of sleep disorders, wireless technology is applicable where patients are allowed to acquire the merits of treatment and diagnosis, even with not formulation of any forms of disturbances in the patients' normal sleep patterns. In this process, healthcare practitioners can obtain the necessary information about the patients [9]. The remaining section of the article is divided into the following sections. Section 2 explains the various intelligent computing methods. Section 3 handles the obtained findings from the reviews. Section 4 explains the results and discussions. Finally, Sect. 5 concludes the research work along with the future scope.

2 Intelligent Computing Methods

2.1 Knowledge-Based Systems (KBSs) and Rule-Based Reasoning (RBR)

Knowledge-based systems (KBSs) represent the AI device that provides smart decisions purposed for validation. In this, the skill representation and acquisition are being structured based on various scripts, frames, and rules. Knowledge is thus represented by the use of different cases: RBR and CBR. The core elements of RBR are the rule-based and the inference engines. The rule based includes several standards termed as the knowledge based. The inference engine is capable of inferring data about the interaction of the rule-based input. The match-resolve act cycle is structured to effectively execute the construction framework program. The core merit of RBR is the definition of the data in the manner of protocols, compressing the representation of the modalities, and the protocols [10]. The R-K method is basically centered on the marking of events such as k-complex, sleep spindle, slow delta waves, and the rapid eye movement instead of the background signal activities.

In case no marking instances or events are identified at the sleeping epoch, then the event-centered smoothing protocols and classification rules affect the ANNs with minimal performance. Moreover, the rule-centered proficient framework has three fundamental limitations:

(1) All the situations might not be defined by the protocols
(2) The experiences retrieved by the errors and trials might easily be contained by the knowledge-based with no human efforts
(3) Misconception to properly comprehend the protocols.

The neural network systems included the shortcomings of the protocol-based expert framework. The limited reliability of the automatic sleep scorings has been replaced by the demerits of the hybrid neural networks and the rule-centered proficient framework with the agreement rates of about 85.9%. CBR represents the procedure of mitigating the novel issues based on the remedies of the same past issues. CBR-based frameworks are structured to provide a remedy on the novel issues by applying the four R's: Retain, Revise, Reuse, and Retrieve. CBR has the capability to effectively convey a certain update and knowledge of KBS when novel cases are reported. This also aids in the management of the unpredicted input. However, there are some demerits or limitations such as the acquisition of knowledge and the issues of the limited or unDoi cases, where the efficiency of inference is not desired based on the straightforward provision of explanations [11]. Several research evaluations have utilized the connected or integrated technique of RBR and CBR for the implementation of the frameworks in the process of detecting sleeping disorders.

2.2 *Artificial Neural Networks (ANNs)*

ANNs commonly utilized for the pattern classification and recognition include the collection of the kind of perceptron linked with the layers by the data-controlling techniques and connectionist in the clinical domains. This possesses the adaptive condition to transform its structure during the learning stage, in that regard, utilized in the mitigation of the actual-world complicated connections. It relates to the issues in which the training information assimilates to the complex and noisy sensor information, including the issues which are more symbolic when illustrated. BP algorithms are majorly utilized in ANN study methods that have been utilized by a lot of researchers in the classification and detection of hypopnea and SA events. ANN is known for its few merits over KBS in the exhibition of the complementary approaches to RBR in terms of skill set illustration that necessitates a lengthy duration to construct such a framework from the rule-centered techniques meant to extract features from the initial forms of recordings such as PSG and EEG then structure the protocols about the human knowledge. ANNs incorporates attractive properties for the automatic recognition of the sleeping EEG pattern that does not require any more elaborated categorization protocols or complex domain skill set.

ANNs agreed with a more manual scoring of about 93% for the wide-range scoring epoch to a more manual scoring of time consuming and arousal procedure. The semi-automated arousal detection framework has been implemented by a system of

Sorensen applying FFNN meant to overcome the kind of limitations of the manual framework [12]. The repetitiveness and abstraction of the tasks are more direct to the low inter-scorer and inaccuracies agreements, whereas no skill of probability distribution is necessary. Neural network is more proficient in the evaluation of the posterior probability, giving the basis of implementation of the category protocols. The automatic categorizer was projected to identify the arousal, less time consuming, and inexpensive procedure as contrasted to other methods.

In the automatic sleeping spindle detection research, cited from the visual detection findings range from 70 to 99% and the false-positive ranging from 3% to 4.7. ANNs is effective in the categorization of the non-periodic and nonlinear forms of signals such as sleeping EEG pattern. In another critical analysis, two forms of ANNs: LVQ and MLP are utilized to categorize the sleeping stages for babies. Automated sleeping stage and scoring in humans were projected based on the application of the multi-layered FFNN, which includes the rates of recognition that varies from 82 to 90%. The merits of bio-sleep TM utilize the automatic neural network techniques for the sleeping stage for speed and ease of analysis. Not entirely bounded to R-K technique, it provides the merits over the manual evaluation. ANNs insensitive to the data distribution indicates the same results being obtained with transformed and raw data. ANNs are considered deft in handling the non-Gaussian probabilities and density function that includes extreme values. A transformation is ineffective based on its capacity to separate the spaces into subspaces. The k-NN function is non-parametric as a method that can be used for categorization, which is also assumed to have no apriori parameter knowledge on the probability framework of data samples being used when the best performances are being provided. This is done for transformed and homogeneous data for a considerable number of cells with minima sizes that can be accomplished by LDA as contrasted to ANNs.

LVQ, which represents the network system with supervised training, indicated an outstanding adaption to the training sets with an enhanced number of neurons attaining maximum categorization. The normalized readings are considered separate completely from the apneic analyzed recordings from NN and k-NN supervised learning classifiers. The merits of RBF networks include architectural simplicity, minimization in the training timeframe, and the capacity to handle the unseen information. The RBF networks are applicable for fault identification, face identification, or clinical diagnosis. RBF-FCM networks assure the best categorization accuracy with minimized network complexities whereby RBF-KM networks assure the most effective categorization performance as contrasted to the RBF-OLS and RBF-FCM networks. With the mentioned networks, performance is considered somewhat minima compared to the RBF-KM. Irrespective of these merits, ANNs are known to have some drawbacks, such as the organization of NN, which is considered more ambiguous.

The priori knowledge/data utilized for the initialization aim might not be handled for better initialization of networking parameters and the minimization of the learning timeframe. It is included as a rule-centered proficient framework that might not contain all the essential protocols as a result of the minimal agreement rate of 55%

as contrasted to the rule-based proficient framework of 83%. The visual identification and counting of the spindles are considered laborious and takes time for the complete sleeping EEG recordings. As such, there is a problem connected with the mimicking of the human score in the automatic sleeping spindle detection framework. ANN-centered analysis framework is considered not sufficiently effective and accurate for the sleeping research that utilizes the R-K categorization framework. Bio-sleeping RM is constrained to the visualized inspection between the pseudo-R-K hypnogram and the scores manually. Eliminate a single scorer out and the classical cross-validation methods might not be applicable whenever operating the massive database.

The categorization errors reach about 30% and might not be enhanced whenever the number of information in the sets is more than 500 different samples that use the k-NN classifier or the Parzen estimators. K-NN and MLP gave incredible results as contrasted to other methods; however, k-NN is known for its majority votes over the remainder of the classes. The choice of the most effective NN and the optimization structure of layers are time consuming and challenging to consider. K-NN necessitates spacing for the massive amounts of study vectors in the span of memory. The unsupervised GCS and SOM techniques were considerably poorer. In the wide-range apnea-screening techniques, it is seen that about 90% of correct categorization on timely analysis can be obtained for patients. It potentially extracts the spectral elements via the Fourier transform of both ECG and RR series morphological features.

In our perspective, its major demerit is in the dimensional characteristic spaces, i.e., 88 various features. The selection of a simplified topology amount in a network system that is not capable of studying the complex characteristics whereas the complex topology amounts to the generalization loss and its capacity that amounts to overfitting the studying dataset. In a considerably complex architecture, NN might have the capacity to master the trained sets amounting to the inaccurate projection of futuristic samples. In that case, early stopping methods are known to be an alternative to make this possible. It includes the validation sets meant to halt the trained algorithms before the networks begin mastering the noises in the information set as a segment of the model amounting to the evaluation and estimation of the generalization errors. The most effective generalization evaluation is attained based on the networks whose level of complexity is considered too large or too low.

The RBF-KM network from the increment in the sizes of the hidden layers was not known to substantially enhance the varied form of accuracy. Apart from that, studying the MLP network with BP necessitates a significant number of the user-based parameter, which will be a priori-specified training epoch, momentum and a learning rate. SVM, as a result of its generalization capacity, is utilized in the process of solving the supervised categorization, binary categorization, and regression problems, which include the tasks of non-parametric elements applicable in statistics. This maximizes the margins identified between the decision boundaries and the training sets of information. As a result, this can cast as the quadratic optimization issue. Machine learning technique is projected by researchers in 1995. The ideology of SVM is to structure the optimized separation hyperplanes. An optimization method of SVM represents the width of margins for different categories, i.e., the spaces around the decision boundaries illustrated by the prevailing distance to the closest training pattern.

SVM is a supervised learning model connected to the learning algorithm meant to evaluate data and identify the potential patterns and map information into the high-dimensional space meant to identify the separated hyper-plane with more maximize margin. The merits of SVM are that it might potentially mitigate the issue of nonlinear categorization and there is no need for speed clamping for the constriction framework. The kernel element is applied in the process of mitigating the issue of the inner products and its calculations in the high-dimensions; hence, an effective technique for the nonlinear categorization is considered from it. The kernel elements have to be selected to accomplish the most effective categorization accuracy for unidentified samples. The cross-independent and validation test accuracy of the apneic events' identification are noted to be about 92%–93%, respectively. For hypopnea events, these two forms of accuracies are considered to be 90 and 89%, whereby sensitivity was utilized to effectively optimize SVM parameters.

After evaluating three various kernel elements such as sigmoid, polynomial, and RBF, it was identified that the polynomial kernel indicates high performance compared to the rest. In contrast to GA, PSO is known to have lesser complicated operations. In that case, lesser parameters might be coded based on the stochastic procedures to mitigate it. SVM has the capacity to effectively minimize both the empirical and structural risks amounting to the effective generalization of novel forms of data categorization. Rapid convergence is a single limitation of PSO. The self-advising SVM fundamentally gives better findings compared to the ancient SVM. Self-advising SVM purposed to handle with the ignoring of skill sets retrieved from the miscategorized information. SVM represents the approximate application technique of the structural minimization of risks to accomplish low forms of probability based on generalized errors. The categorization performance of k-NN, linear discriminant, and PNN categorizer on test information was minimal compared to SVM. Three various kernel functions are utilized such as radial basis, polynomial, and linear function, whereby 100% accuracy is gotten based on the polynomial kernel element with four various features, the same as the linear function with just two different feature subsets. PNN and k-NN indicate poor categorization performance, i.e., 70% and 83%, respectively, on the tested information.

2.3 Fuzzy Reasoning/Logic

Fuzzy sets theories play a fundamental role in handling the complexities whenever drafting the required decisions in the clinical domain. Fuzzy logic is a form of probabilistic logic that handles logic, which is considered appropriate instead of being accurate and constant. The fuzzy logic variable might have the truth value, which ranges in degrees between 0–1. It is considered expanded to effectively manage the theories of partial truth, whereby the truth value might range between the completely true and false. In the linguistic variable, the degrees might be dealt with based on certain functions. The fuzzy rule-based framework attained accurate results for several samples, but it still requires being improved based on performance in the various samples of the recorded information. It also had to overcome the potential limitations of the epoch-centered sleeping staging by accessing more continuous transformation of the sleeping habit in patients. The elimination of binary decisions assures soft transformation and allows concurrent characterization of the various sleeping states. The usage of the Mamdani fuzzy protocols allows knowledge to effectively be applied in the manner of linguistic protocols, next to the human language, that effectively facilitates the acquisition of knowledge, understandability, and also permits explanatory capacity.

The limitation of the R-K protocols was based on the unnatural assignment of discrete stage apart from producing it more continuously. Receiver operating curves (ROC) index of 1 was received for the categorization of such events as hypopneas and apneas. The manual sleeping categorization of patients was structured by various experts with the inter-raters reliability of approximately 70%. The categorization procedure has considerable results compared to the processes of discretization. It has wide-range advantages such as easy reformation of the rule-base or the fuzzy datasets, ease of comprehension as a result of the output presence in the linguistic formation, ease of designing as a result of minimal costs, and the provisions meant to permit the conflicting inputs arrives at a lesser time interval. Contrary to that, fuzzy includes the limitation such as the challenging construction of the models from the fuzzy framework.

2.4 Genetic Algorithms (GA)

The genetic algorithms represent the search heuristic optimization method that mimics the process of natural advancement. These algorithms have been exposed to identify the optimal remedy for various challenging issues as a suspected or effective tool centered on the principles of evolutionary strategies. This is also possible to effectively transfer GA to the present simulation framework. A stylish genetic algorithm requires two terms: robustness element meant to evaluate the remedy domain and the inherited demonstration of the remedy domain. This provides the automated score about the sleeping stages [13]. The major merit of GA is that no considerable condition has the knowledge of mathematical evaluation for the purpose of comprehension. It is more effective to utilize the case of a complex and huge search domain. In these instances, the localized minima price element might be retrieved from the gradient optimization technique. There are several limitations of GA such as the no assurance of international optimum results, no firm optimization response timeframe, limited controls of GA application and the accurate optimization issues that might not be mitigated by GA.

2.5 Data Mining (DM)

Data mining represents the procedure utilized in the discovery of knowledge and the process of pulling previously unforeseen connections, layouts, and unidentified relationships. DM gives offers meant to effectively distinguish productive chunks of data covered in huge and expensive databases. Statistics provided from DM packages include associated rules, data segmentation, k-nearest neighbors, rule induction, and the decision tree. The tree-like architecture is utilized to enhance the accuracy level over the single form of the tree structure. In case the neuron variation is enhanced in the hidden layers, the mean accuracy is considerably enhanced and thus the standardized deviation is minimized. The most effective performance is attained based on the application of LMBP (trainlm) and the Gradient Descent based on the mean momentum (traingdx) and the adaptive study rate BP learning element to study the ANNs. However, trainlm provides effective results in the testing and training process contrasted to traingdx. The insomnia prevalence happened more frequently in OSAS patients and was systematically connected with poor sleeping quality does not affect the long-term complications about the kind of patients suffering from insomnia ranging from moderate to more severe syndrome. The rate of agreement was thus extended through the extraction of features from EMG and EOG signals ranging from 71 to 80%. An automated detection framework was thus proposed through the avoidance of a time-consuming manual arousal procedure.

2.6 Bayesian Network (BN)

Bayesian network provides a framework for undetermined reasoning, which is capable of dealing with diagnostic issues. BN is healthy and strong as a system formalism, which permits reasoning under some forms of uncertainties, recommending a more graphic representation of more statistical dependency between the variables of the domain. BN is considered a more directed cyclic graph, which means it is a combination of different nodes that represents more random variables linked by different edges that feature the conditional probabilistic dependency between the various vertices. The categorization of different nodes is dependent on the parent nodes and Node Y is considered conditionally independent of X in case there are no straight paths from X–Y. In that case, BN is structured to represents not just correlate the causalities. It thus facilitates the various forms of visualization of firm links.

3 Computational Analysis

The single form of KBS framework has individualized merits and demerits such as inference issues, and skillset acquisition issues. As dealt with in the above segment, the demerits of the single form KBS are limited through the integration of ICT and KBS methods. In this segment, the merits incurred from the integrated frameworks alongside their applications from treatment and diagnosis of sleeping disorders. The interlinkage of CBR-RBR amounts to the ease of the acquisition of knowledge, enhancement attained based on accuracy, efficiency, and performance. The major merits prevailing from the connected methods incorporate high performances on sleeping disorders, sleep justification, and learning capacity. The segment of sleep stages and its scoring is considered to necessitate hybrid reasoning due to human experts using both rule-centered experiences and knowledge.

To mention the merits of CBR in the treatment and diagnosis of OSA, researchers have considerably formulated the minor prototype framework known as Somnus that provides case retrieval and storage in the sleeping disorder. To handle with complexity and diversity of information, researchers have utilized the model that includes the fuzzy logic method from modeling the case elements and the semiotic method for the model of the wide-range measures. The user interfaces structured in the prototype of the Somnus framework are entirely restricted to the SQL statement hence providing more quick access to information; however, the system is now limited to deal with the minor groups of users that are familiar with the database schema. The usage of the semi-fuzzy method provides users a more uniform representation of subjective, objective, qualitative, and quantitative measures. Even the more limited CBR cases in Somnus, it is essential to consider that different healthcare providers in the training field use the measures. A lot of computational frameworks are recommended in the segment of sleep to entirely handle the detection framework. The first application of NN is the time of sleep has widely been evaluated by researchers [14].

The BP technique is termed in this instance but the categorization rate does not exceed 60%. The researchers in [15] had tried about three various supervised learning techniques such as the scaled conjugate gradients, Bayesian approaches, and the regularly scaled conjugate gradients. SCG was chosen as a studying algorithm for the networks as a result of the prompt convergence speed and minimal memory requirement and the mean squared errors (MSE) as a more cost function. However, the second approach is more opposite to SCG whereby MSE with more regularization terminology term known as weight decay is utilized as a cost element. Whereby only the most effective results were accomplished based on the application of the Bayesian model and the regular cross-entropy element through the reduction of the error function.

4 Results and Discussion

Discrete wavelet transformation was utilized as a processing phase to fix and minimize the input number of the classifiers. The choice of Bayesian network has fundamental advantages that the various learning parameters are auto-adaptive or non-external human action required. In past research, a lot of emphasis has not been done on the categorization and detection of the nonlinear features of the EEG signal. The major causal factor might be the challenging mathematics and require some expertise meant to interpret. However, many researchers have projected the connected interlinkage of the bispectral evaluation of ANNs. Contrary to the energy spectrum, the bio-spectrum presents the non-Gaussian and the nonlinear data that permits the identification of the nonlinear features and the featuring of nonlinear approaches. The quadratic phase coupling (GPC) is considered a unique feature of bispectrum, which is utilized to differentiate, quantify and detect the normalized QSAS patients. Training is effectively controlled through the cross-validation method.

In the training process, chromosomes have been structured based on the variable-length architecture and the fitness element, which is utilized to identify the optimal input character and the identification of the network specifics. The integrated framework base on feed-forward NN had considerably transformed concerning genetic algorithms, to identify the input features and optimal structures. The massive amount of information, complexities of categorization evaluation, and the variability in the human expert present the sole reason to formulate the automated sleeping categorization framework. The neuro-fuzzy model was selected for the structuring of the protocols, for the categorization procedure and to identify the kind of parameters, which illustrate the degrees of absence and presence of patterns. The findings from this analysis were from the multilayer perceptron NN and were recorded at 87.3% and based on the professional's rules of sleeping categorization of 86.7% whereas the accomplishment of the findings in NF categorizer was about 88%. More time is essential to structure the model from the rule-centered method for the feature extraction from the more initial recordings. The R-K method determines the sleeping and waking stage has to last for about a single minute.

In this research, an attempt was structured to evaluate the different methods adopted for the sleeping disorders such as BN, DM, MLP, NN, SVM, GA, FL, and ANN are all information controlling techniques such as GA-FL, ANN-DM, ANN-GA, ANN-BN, ANN-FL, RBR-ANN, CBR-FL, and RBR-CBR. This paper uses and shows the overall cases in which different techniques are utilized for diagnosis, classification, and detection of sleeping disorders. From the research, it is examined that the standalone class, from the 79 different cases, 5 BN cases, 7 Fuzzy cases, GA cases, and ANN major cases are utilized in the timeframe of sleeping disorders: its diagnosis, classification, and detection. A few instances of the incorporated methods were witnessed in this duration such as GA-FL (3), ANN-DM (1), ANN-GA (4), ANN-BN (5), ANN-FL (2), RBR-ANN (2), CBR-FL (1), and RBR-CBR (2). It is considered that out of the 97 cases connected to the practice of the above methods in sleeping disorders, several approaches, i.e., 59 are used by a single form of methodology whereas the remainder of the 20 cases is positioned by the connected techniques.

From Figs. 1 and 2 of this research, the relative application of every technique about the overall cases were identified based on the application of a single methodology that is represented by (*m*, *p*%). From the representation, m shows the number of the cases based on the application of a certain methodology whereby p is the percentage ratio of m (41) to the overall cases based on the application of the singular methodology ($41 + 5 + 2 + 7 + 4 = 59$) shown in the table (7th row). In that case,

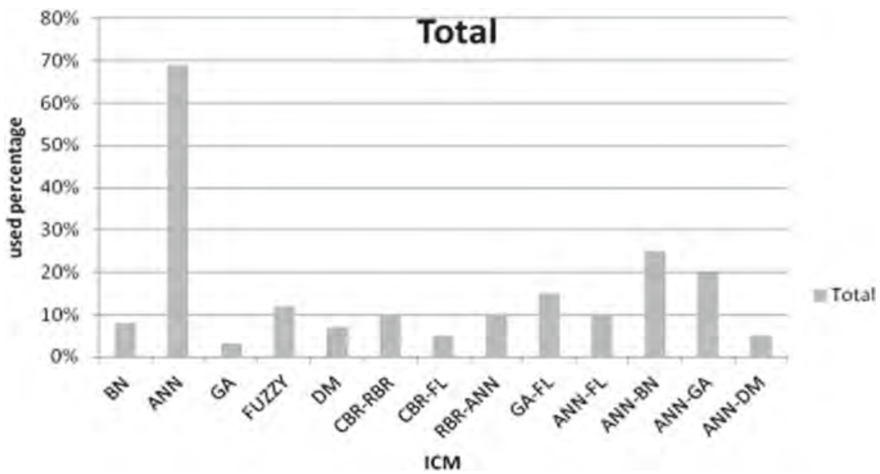


Fig. 1 Comparison of computing methods against usage

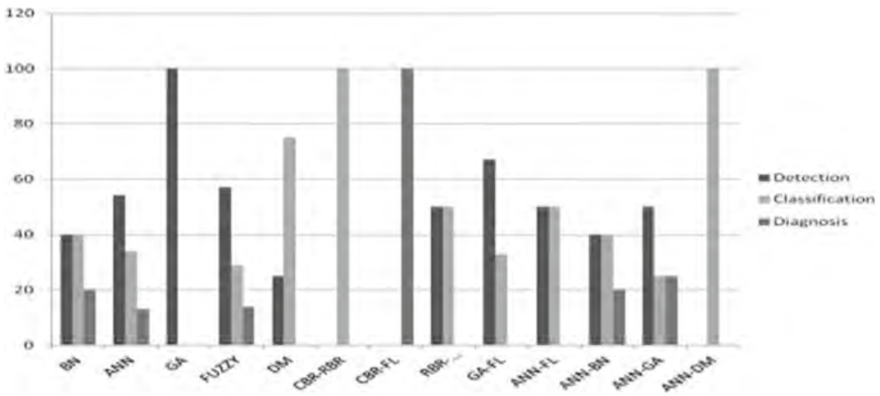


Fig. 2 Comparison of computing methods against percentage usage

the relative application of ANN methodology is given as FL (12%), DM (7%), BN (8%), and GA (3%) which depicts a total of 69 as shown in Fig. 1. As for the incorporated methods, the same calculations are utilized and shown by (*i*, *q*%). From the representation, *i* represents the overall cases based on the application of certain connected methods and *q* represents the percentage ratios of i to the overall number of the connected methods utilized. In that case, the relative usage of CBR-RBR is exactly 10%, GA-FL (15%), ANN-DM (20%), ANN-GA (20%), ANN-BN (25%), ANN-FL (10%), RBR-ANN (10%), CBR-FL (5%), and RBR-CBR (10%).

5 Conclusion and Future Scope

The core rationale of this article is delivering a critical instance in the deployment and development of different ICM in the duration of sleeping disorders. Various literature sources were analyzed in the domain of sleeping and walking. It is evaluated that ANN methods are considerably utilized in the sleeping disorders domains as contrasted to a single methodology where minimal reasoning. The connected methods are also utilized in the classification and detection of sleeping disorders and contrasted to single methods, whereby their numbers are minimal. Out of the eight connected methods, ANN-GA and ANN-BN have widely been utilized percentagewise compared to other connected methods, i.e., 20 and 25%. Over the past few decades, the trend of utilizing fuzzy logic and DM has increased considerably. ANN, fuzzy, and BN are mainly utilized for diagnosis, categorization, and detection purposes; GA is utilized for detection whereas DM is for classification and detection. As such, future research might find our research relevant for novices, which might emerge in the medical field.

References

1. K. Aoyagi, Medical image assessment apparatus, ultrasonic image assessment apparatus, magnetic resonance image assessment apparatus, medical image processing apparatus, and medical image processing method. J. Acoust. Soc. Am. **133**(5), 3220 (2013). https://doi.org/10.1121/1.4803793
2. B. Lelieveldt, N. Karssemeijer, Information processing in medical image assessment 2007. Med. Image Anal. **12**(6), 729–730 (2008). https://doi.org/10.1016/j.media.2008.03.005
3. P.L. Aaron, S. Bonni, An evaluation of wearable technological advancement in medical practices. J. Med. Image Comput. 58–65 (2020)
4. Web based analysis of critical medical care technology. J. Med. Image Comput. 66–73 (2020)
5. A. Haldorai, S. Anandakumar, Image segmentation and the projections of graphic centered approaches in medical image processing. J. Med. Image Comput. 74–81 (2020)
6. C. Hung, Computational algorithms on medical image processing. Current Med. Image Assess. Form. Current Med. Image Assess. Rev. **16**(5), 467–468 (2020). https://doi.org/10.2174/157340561605200410144743
7. F. Aubry, V. Chameroy, R. Di Paola, A medical image object-oriented database with image processing and automatic reorganization capabilities. Comput. Med. Image Assess. Graph. **20**(4), 315–331 (1996). https://doi.org/10.1016/s0895-6111(96)00022-5
8. P. Jannin, J. Fitzpatrick, D. Hawkes, X. Pennec, R. Shahidl, M. Vannier, Validation of medical image processing in image-guided therapy. IEEE Trans. Med. Image Assess. **21**(12), 1445–1449 (2002). https://doi.org/10.1109/tmi.2002.806568
9. K. Drukker, Applied medical image processing, second edition: a basic course. J. Med. Image Assess. **1**(2), 029901 (2014). https://doi.org/10.1117/1.jmi.1.2.029901
10. P. Jannin, Validation in medical image processing: methodological issues for proper quantification of uncertainties. Current Med. Image Assess. Rev. **8**(4), 322–330 (2012). https://doi.org/10.2174/157340512803759785
11. M. Goris, Medical image acquisition and processing: clinical validation. Open J. Med. Image Assess. **04**(04), 205–209 (2014). https://doi.org/10.4236/ojmi.2014.44028
12. T. Aach, Digital image acquisition and processing in medical x-ray image assessment. J. Electron. Image Assess. **8**(1), 7 (1999). https://doi.org/10.1117/1.482680

13. H. Barrett, A. Gmitro, Information processing in medical image assessment. Image Vis. Comput. **12**(6), 315 (1994). https://doi.org/10.1016/0262-8856(94)90055-8
14. T.K. Araghi, Digital image watermarking and performance analysis of histogram modification based methods. Intell. Comput. 631–637 (20180
15. G. Yu, Z. Wei Xu, J. Xiong, Modeling and safety test of safety-critical software, in *2010 IEEE International Conference on Intelligent Computing and Intelligent Systems* (2010)

Review on Face Recognition Using Deep Learning Techniques and Research Challenges

V. Karunakaran, S. Iwin Thanakumar Joseph, and Shanthini Pandiaraj

Abstract In the research area of object recognition, last few decades many of the researches did research on face recognition and the research is still active because of its application and challenges present in the real world. In the constrained environment most of the recent face recognition techniques offer a better result but it fails in an unconstrained environment. An unconstrained environment such as the images were captured during various environments like different resolutions, in a different pose with various expression, illumination, and occlusions. In this article, various deep learning techniques used for face recognition have been discussed.

Keywords Face recognition · Deep learning techniques · Object recognition · Constrained environment · Unconstrained environment

1 Introduction

The researchers who belong to computer vision had more attention toward deep learning techniques. Most of the researches carried out with various deep learning techniques for face recognition. The reasons for most the researches have more attention toward deep learning techniques can easily deal with a huge amount of data and it provides better classification accuracy. It is heavily dependent on high-end machines. Here, the problems are not divided into a small problem, and it solves the problem as end to end. It takes a long time to train the system and it takes

V. Karunakaran (✉) · S. Iwin Thanakumar Joseph
Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India
e-mail: karunakaran@karunya.edu

S. Iwin Thanakumar Joseph
e-mail: iwinjoseph@karunya.edu

S. Pandiaraj
Department of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India
e-mail: shanthini@karunya.edu

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_64

only a small amount of time for testing the data. Face recognition research is still alive due to its following applications, such as prevent retail crime, unlock phones, smarter advertising, find missing persons, help the blind, protect law enforcement, aid forensic investigations, identify people on social media platforms, track school attendance, and so on.

The rest of the paper is organized as follows: Section 2 presents challenges in face recognition, Sect. 3 explains the face recognition using deep learning algorithms. Finally, Sect. 4 concludes the research work.

2 Challenges in Face Recognition

Challenges in face recognition are as follows: pose variations, presence or absence of structuring elements or occlusions, facial expression changes, aging of the face, various illumination conditions, image resolution and modality, and availability and quality of face dataset. In this section, the challenges of face recognition were discussed.

2.1 Pose Variations

In face recognition, one of the main challenges is posed variations. In practical applications, the head pose plays an important role in face recognition. Many researchers were handled pose variations problem using the following three algorithms such as

1. Invariant feature extraction-based approach
2. Multi-view based approach
3. 3D image-based approach.

Invariant feature extraction face recognition was carried out with invariant pose changes. This approach is further classified into

1. Appearance-based algorithm
2. Geometric model-based algorithm.

If the face image dataset is sufficient, then appearance-based algorithm is performed well. If the input face image dataset is insufficient, the appearance-based algorithm will not be performed well.

This problem was overcome by face synthesis, it will create additional face images from the existing ones, and it will improve the accuracy of the model. The geometric-based model provides a promising result in the situation of various face pose changes and with an insufficient face image dataset but it acquires more computational cost when compared to the appearance-based algorithm [1]. The same person's face image with a different pose is shown in Fig. 1.

Fig. 1 Same person face image with different pose

2.2 Presence or Absence of Structuring Element or Occlusion

When the face recognition algorithm is tested with occlusions image, definitely there will be a small drop in the algorithm performance. Occlusions mean the person is wearing sunglasses, a cap or hat, a beard, a scarf, and so on. It will degrade the performance of the algorithm during face recognition. Many researchers have handled this problem by using texture-based algorithms [2, 3]. Sample occlusion image is shown in Fig. 2 [4].

Fig. 2 Sample occlusion image

Fig. 3 Same person with different expressions **a** anger, **b** disgust, **c** sadness and **d** happiness

Fig. 4 Same person image with various illumination conditions

2.3 Facial Expression Changes

The automatic face recognition algorithm will be tested with various facial expressions definitely there is a small drop in the algorithm performance. Various facial expressions such as anger, disgust, happiness sadness, and so on. Figure 3 shows a sample of various facial expression images of the same person [5].

2.4 Various Illumination Conditions

There is a large variation in the illumination present in the image will definitely drop down the performance of the face recognition algorithm. The various illumination conditions such as low-level lighting in the background or foreground of the image and high-level of lighting in the foreground or background of the image. Various illumination conditions image is shown in Fig. 4.

In the next section, the researchers handled the above-mentioned challenges by using different approaches and algorithms.

3 Face Recognition Using Machine Learning and Deep Learning Techniques

Masi I et al. proposed pose aware model for tackling pose variation problems by using a convolutional neural network with specific several poses. In this article, the

3D rendering is used to synthesize multiple face poses obtained from the input image. If the model is trained with multiple faces to pose obtained from the input image will help to achieve the best accuracy result in the test phase. The obtained result clearly shows that the proposed method provides better accuracy and proved the method is capable of tackling pose variation problems [6]. In this article, the face images were processed by various poses using deep convolution neural network. Deep convolution neural network layers and various pose model selection had improved the performance of the proposed system during the recognition phase.

The proposed system provides better results in the aspect of verification and identification tasks compared to the state-of-the-art methods [7]. Chen et al. evaluated the performance of deep convolution neural network for a new dataset such as IARPA benchmark A and Janus Benchmark A (IJB), additionally with traditional data set such as Labeled Face in the Wild (LFW). The experiment was conducted with two methods: one is DCNN and another one is a Fisher Vector method. The experiment result clearly shows deep convolution neural network model was performed well on both new datasets such as IARPA benchmark A and IBJ A when compared to the fisher vector method in the identification and verification tasks [8].

Su et al. proposed a model for detecting sunglasses and scarves. If the image is present with sunglass and scarf, the support vector machine is used for detecting the occlusion in the image and the regression analysis will be used to remove the sunglass and scarf from the image, that image is called a reconstructing image. The experiment was conducted for both the reconstructed image and the original image. The experimental results concluded when compared to the reconstructed image, the non-occluded part for the face recognition image provides a better result [9].

Wang et al. [10] give an overall view about the exploration and content creation of virtual reality with deep learning methodology. A rapid growth in the deep learning techniques and its advantages over all applications energized the involvement of machine learning paradigms in virtual reality methods for an intelligent approach. In general, the content creation and exploration of virtual reality directly map to

(a) Analysis of image and video
(b) Appropriate synthesis as well as editing.

Usually, the general adversarial networks are mostly used and modeled for a specific application to manage

(a) Panoramic images
(b) Videos
(c) Virtual 3D scenes.

Golnari et al. [11] developed DeepFaceAR which is used for the recognition of deep faces and display of personal information through the concept of augmented reality. The popular research topic in machine vision is biometric recognition. Here, deep learning methodologies combined with augmented reality logic are used to recognize the individual faces and also listed about the person utilizing the concept of augmented reality. The dataset consists of 1200 facial images of approximately 100 faculty teachers belong to the Shahrood University of Technology.

Mostly, augmented reality based works follow three approaches, namely location based, marker based, and motion based.

Shaul Hammed et al. [12] give an insight into the facial recognition system. The major form of non-verbal communication is facial expression. It expresses

1. Person feeling
2. Judgment of an individual.

The general facial expression system consists of the following four steps:

1. Signal acquisition
2. Preprocessing
3. Feature extraction
4. Classification.

Hbali et al. [13] developed an augmented reality system for face and dual eyes based on the feature of the histogram of gradients used for object detection. Appropriate machine learning algorithms are utilized for

(a) The detection of the face and eyes of a human using that application.
(b) Tracking of eyes in a real time.
(c) Eye and face positions are used for embedding the image of glasses upon the face.

These types of systems are very well utilized for the application of checking the quality of glasses without directly checks in into the shops. The accuracy of the system is highly improved due to the utilization of HOG features. Based on the reduction of computational complexity, these kinds of virtual reality-based applications can be effectively implemented on smart gadgets thus upgrade the shopping behavior of customers through e-commerce.

Yampolskiy et al. [14] developed a system for recognition of face in the virtual world especially, avatar face recognition. The major problem in the virtual world is several types of criminal activities enabled the forensic communities to track accurately the user in an automotive manner. A COTS FR methodology provides the best accuracy of identification and also avatar face recognition mainly introduced for the authentication purpose. FERET-to—avatar face dataset is used in this system for testing the efficiency of the algorithm.

The psychological, social, and economic position of the user and their relevant avatar in the virtual world gives that avatars mostly mapped with their owners instead of being as a fully virtual design and implies high stability.

The general template for recognizing avatar face requires the following three basic steps

1. Detection of the face and image normalization
2. Representation of face
3. Matching.

Yong et al. [15] developed deep learning-based recognition of emotions for who wear head-mounted displays. Appropriate training to the convolutional neural network (CNN) is given by concealing eyebrows and eyes of the face image wears head mounted display (HMD) in an available dataset. This gives an excellent performance in the estimation of emotions from images of a person who wears HMD. Lang et al. [16] investigated the algorithms for face detection in the real-time environment and introduced an adaboost-based face detection algorithm which categorizes under multi-classifier model. This methodology generates a cascade basis in the training phase and the approach is comparatively strong to illumination, pose, and highly applicable for real-time systems (Table 1).

Table 1 List of deep learning-based face recognition in virtual reality

Authors and year	Description	Inference
Mesi et al. [6]	The pose aware model is used for tackling pose variations problem by using a convolutional neural network with specific several poses	Better in terms of accuracy during the recognition phase
Almageed et al. [7]	Deep convolutional neural network layers and various pose model selection is used in face recognition to solve pose variations problem in face recognition	Better in terms of verification tasks and identification tasks
Chen et al. [8]	DCNN performance was evaluated on two new datasets such as IARPA benchmark A and IJB-A	DCNN is performed better in terms of both the verification task and identification task for both new datasets when compared to the fisher vector method
Su et al. [9]	SVM is used for detecting an occlusion in the image. Regression analysis is used for removing occlusion from the image	The experimental results concluded when compared to the reconstructed image; the non-occluded part for the face recognition image provides a better result
Golnari [11]	Combination of deep neural network with augmented reality to recognize the individual face	Recognition accuracy improved
Hbali [13]	Hybrid of augmented reality with HOG features for detection of face and eye	Virtual eyeglass try on system
Yampolskiy et al. [14]	COTS FR algorithm to verify and recognize avatar faces	The algorithm performs 99.58% accuracy
Yong et al. [15]	CCN is trained to measure the emotions from facial image wearing HMD	Accuracy of estimating emotion improved

4 Conclusion

This research article gives insight into the facial detection system based on machine learning approaches in virtual reality applications. This new research direction paves the way to further research enhancement on visual media. This research challenges and future scope in the research field of facial recognition were discussed. The combined approach of the deep learning algorithm and augmented reality enhanced the identification of the face of an individual with improved accuracy.

References

1. S. Du, R. Ward, Face recognition under pose variations. J. Franklin Inst. **343**(6), 596–613 (2006)
2. R. Min, A. Hadid, J.L. Dugelay, Efficient detection of occlusion prior to robust face recognition. Sci. World J. (2014)
3. R. Singh, M. Vatsa, A. Noore, Recognizing face images with disguise variations. Recent Adv. Face Recogn. 149–160 (2008)
4. A.A. Yusuf, F.S. Mohamad, Z. Sufyanu, A state of the art comparison of databases for facial occlusion. Jurnal Teknologi**77**(13) (2015)
5. F. Prikler, Evaluation of emotional state of a person based on facial expression, in *2016 XII International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH)* (IEEE, 2016), pp. 161–163
6. I. Masi, F.-J. Chang, J. Choi, S. Harel, J. Kim, K. Kim, J. Leksut et al., Learning pose-aware models for pose-invariant face recognition in the wild. IEEE Trans. Pattern Anal. Mach. Intell. **41**(2), 379–393 (2018)
7. A. Almageed, Y.W. Wael, S. Rawls, S. Harel, T. Hassner, I. Masi, J. Choi et al., Face recognition using deep multi-pose representations, in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (IEEE, 2016), pp. 1–9
8. J.C. Chen, V.M. Patel, R. Chellappa, Unconstrained face verification using deep cnn features, in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (IEEE, 2016), pp. 1–9
9. Y. Su, Y. Yang, Z. Guo, W. Yang, Face recognition with occlusion, in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)* (Kuala Lumpur, 2015), pp. 670–674. https://doi.org/10.1109/ACPR.2015.7486587
10. M. Wang, et al.,VR content creation and exploration with deep learning: a survey. Comput. Visual Media 1–26 (2020)
11. A. Golnari, H. Khosravi, S. Sanei, DeepFaceAR: deep face recognition and displaying personal information via augmented reality, in *2020 International Conference on Machine Vision and Image Processing (MVIP)* (IEEE, 2020)
12. Hammed, S. Shaul, A. Sabanayagam, E. Ramakalaivani, A review on facial expression recognition systems. J. Crit. Rev. **7**(4) (2019)
13. Y. Hbali, M. Sadgal, A.E. Fazziki, Object detection based on HOG features: faces and dual-eyes augmented reality, in*2013 World Congress on Computer and Information Technology (WCCIT)* (IEEE, 2013)
14. R.V. Yampolskiy, B. Klare, A.K. Jain,Face recognition in the virtual world: recognizing avatar faces, in *2012 11th International Conference on Machine Learning and Applications*, vol. 1 (IEEE, 2012)

15. H. Yong, J. Lee, J. Choi, Emotion recognition in gamers wearing head-mounted display, in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (IEEE, 2019)
16. L. Yong, W. Gu,Study of face detection algorithm for real-time face detection system, in *2009 Second International Symposium on Electronic Commerce and Security*, vol. 2 (IEEE, 2009)

Steganalysis for Images Security Classification in Machine Learning Using SVM

P. Karthika, B. Barani Sundaram, Tucha Kedir, Tesfaye Tadele Sorsa, Nune Sreenivas, Manish Kumar Mishra, and Dhanabal Thirumoorthy

Abstract Grouping is one of the most important errands for different applications, such as text order, tone recognition, image characterization, articulation of miniature cluster efficiency, protein function forecasts, and classification of details. A significant portion of the current controlled order modulation formats on traditional measurements that can give ideal results when the test size is maintained. Nonetheless, only minimal examples can be obtained through and through. A new learning technique, support vector machine (SVM), is used in this paper on different data (diabetes knowledge, heart data, satellite data, and shuttle data) that have two or multi-groups. SVM, an impressive computer technique created from observable learning, has achieved important achievements in some fields. Introduced in the mid-nineties, they inspired a blast of excitement for AI. The SVM establishments were founded by Vapnik and, due to numerous attractive highlights and promising precise execution, are gaining ubiquity in the field of AI. The SVM technique does not withstand information computational complexity constraints and minimal examples.

Keywords Support vector machine · Classification · Machine learning · Security

P. Karthika (✉)
Kalasalingam Academy of Research and Education, Krishnankoil, India

B. Barani Sundaram
Computer Science Department, College of Informatics, Bule Hora University, Bule Hora, Ethiopia

T. Kedir · T. T. Sorsa
College of Informatics, Bule Hora University, Bule Hora, Ethiopia

N. Sreenivas
School of Electrical and Computer Engineering, Addis Ababa Institute of Technology, Addis Ababa University, Addis Ababa, Ethiopia

M. K. Mishra
Department of Computer Science, University of Gondar, Gondar, Ethiopia

D. Thirumoorthy
Blue Hora University, Blue Hora, Ethiopia

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*,
Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0_65

1 Introduction

Vapnik first suggested the support vector machine (SVM) and has since been seriously enthusiastic about the AI research network [1]. A few ongoing investigations have detailed that the SVM (uphold vector machines) is often designed to convey better arrangement accuracy than the other calculations of the information order [2–5]. In a wide range of genuine problems, Sims have been used, such as text order, manually written digital recognition, tone recognition, image characterization, object identification, miniature cluster quality articulation data analysis, and information arrangement. It has been proposed that Sims is better than other controlled learning strategies in a reliable way [6–9]. For certain datasets, however, the SVM display is responsive to how the cost boundaries and portion boundaries are set. Accordingly, in order to figure out the optimal boundary environment, the client typically needs to perform broad cross-approval. As model determination, this loop is typically referred to. One practical problem with model selection is that this cycle is rather repetitive. We also explored numerous avenues for different limits related to the use of the SVM calculation that can affect the results [10]. These limits include bit job decision, the standard deviation of the Gaussian section, relative loads associated with slack factors to reflect the non-uniform transmission of marked information, and the amount of model preparation.

For example, we have taken four specific information index applications, for analysis, diabetes information, heart information, and satellite information, all of which have different highlights, grades, number of preparation information, and distinctive number of testing information.

2 Support Vector Machine Algorithm

Backing vector machine or SVM is one of the most famous supervised learning calculations, and it is used for image classification techniques. In machine learning, classification techniques play a vital role to solve the regression issues. The SVM approach helps to identify the best line or n-dimensional space to isolate the classes, and it will make a new information region without any stretch points. Hyperplane is defined as choice limit. Extraordinary focus in SVM picks the hyperplane, so extraordinary focus called as help vectors and this estimation is known as support vector machine (SVM) [11]. Consider the underneath outline in which there are two distinct classifications that are characterized utilizing a choice limit or hyperplane as shown in Fig. 1.

Model: SVM can be perceived with the model that we have utilized in the KNN classifier. Assume that we see an abnormal feline that additionally has a few highlights of canines, so on the off chance that we need a model that can precisely recognize whether it is a feline or canine, so such a model can be made by utilizing the SVM calculation [12]. We will initially prepare our model with bunches of pictures of

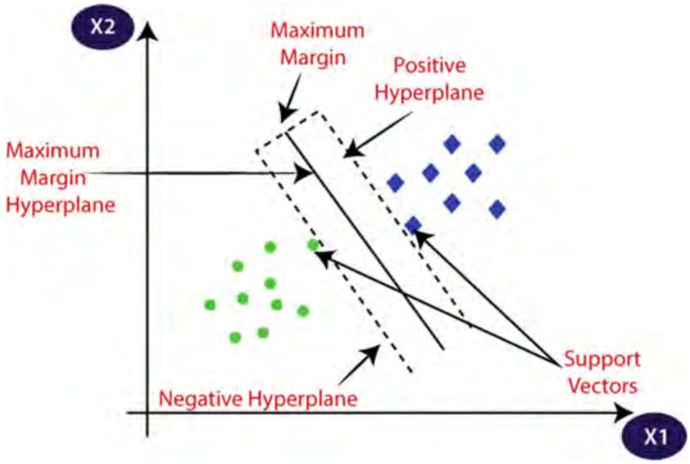


Fig. 1 SVM classified using a decision boundary or hyperplane

felines and canines so it can find out about various highlights of felines and canines, and afterward, we test it with this odd animal. So, as help vector makes a choice limit between these two information (feline and canine) and pick extraordinary cases (uphold vectors), it will see the outrageous instance of feline and canine [13]. Based on the help vectors, it will arrange it as a feline. Consider the chart is shown in Fig. 2:

SVM algorithm can be used for face detection, image classification, text categorization, etc.

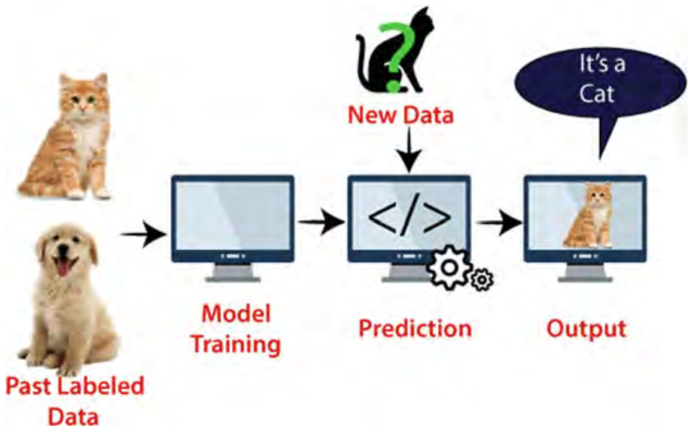


Fig. 2 Example of SVM classifier identifier original or copy image

3 Implementation of Hyperplane for Linear and Nonlinear SVM

3.1 Linear SVM

The working of the SVM calculation can be perceived by utilizing a model. Assume that we have a dataset that has two labels (green and blue), and the dataset has two highlights x1 and x2. We need a classifier that can group the pair (x1, x2) of directions in either green or blue. Consider the beneath picture as shown in Fig. 3.

So, as it is 2D space so by utilizing a straight line, we can without much of a stretch separate these two classes. In any case, there can be various lines that can isolate these classes. Consider the picture as shown in Fig. 4.

Thus, the SVM calculation assists with finding the best line or choice limit; this best limit or locale is called hyperplane. The closest purpose of the lines from both classes is found by the SVM calculation. This focus area is referred to as support vectors. An edge is called the distance between some of the vectors as well as the hyperplane. What is more, SVMs target is to extend this edge. The highest edge hyperplane is recognized as the perfect hyperplane as shown in Fig. 5.

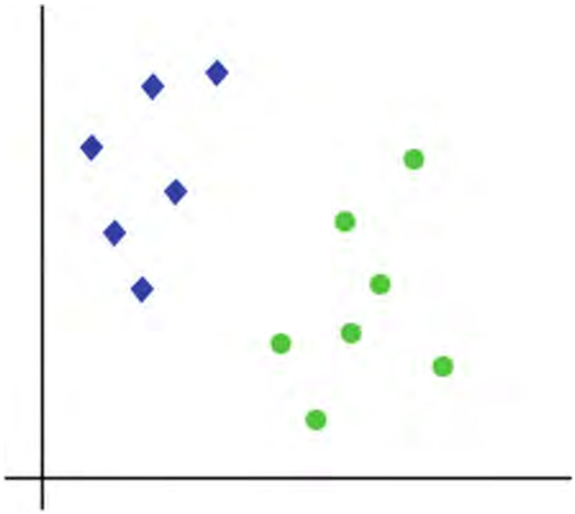


Fig. 3 Working of SVM algorithm

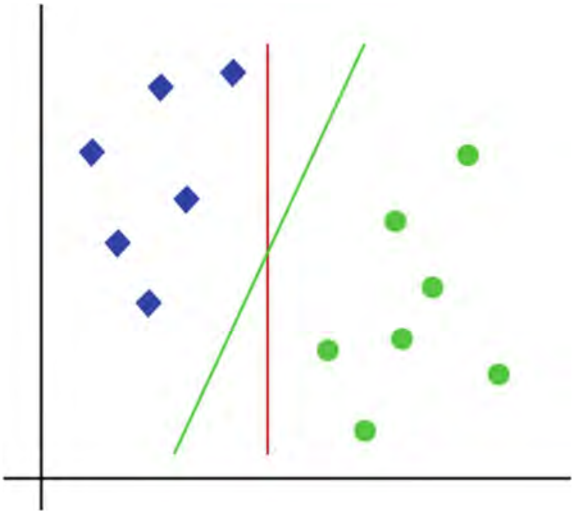


Fig. 4 SVM algorithm using line or decision boundary

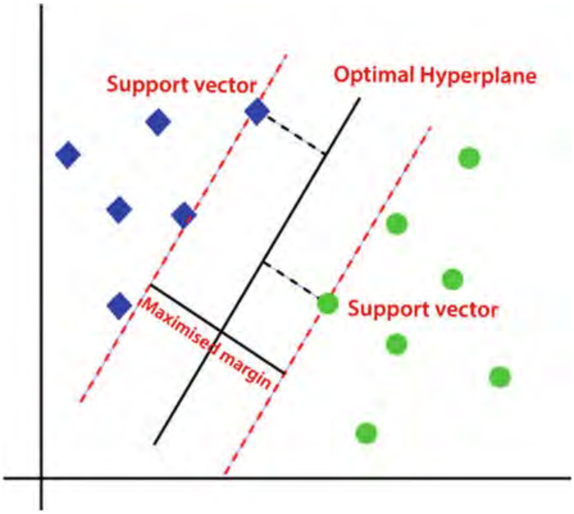


Fig. 5 SVM algorithm goal to maximize margin

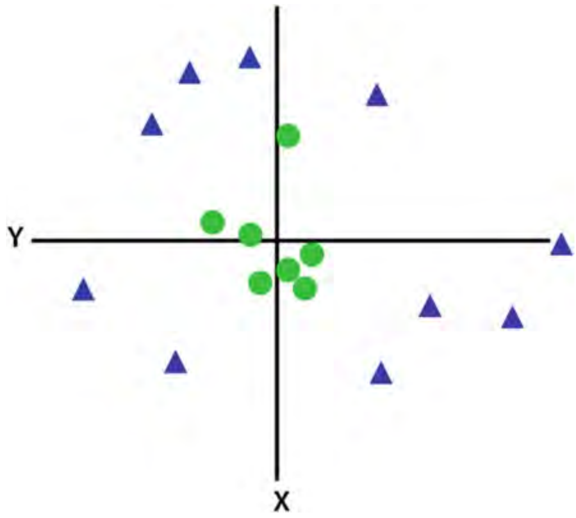


Fig. 6 SVM algorithm using nonlinear data

3.2 Nonlinear SVM

In the event that information is directly masterminded, at that point, we can isolate it by utilizing a straight line, yet for non-straight information, we cannot draw a solitary straight line. Consider the picture as shown in Fig. 6.

So, to isolate these information focuses, we have to include one more measurement. For straight information, we have utilized two measurements *x* and *y*, so for non-direct information, we will include a third measurement z. It very well may be determined as:

$$z = x^2 + y^2 \tag{1}$$

The input vector of the SVM maps to a larger dimensional space, where a maximum isolating hyperplane is created. On each side of the hyperplane, two equal hyperplanes are generated that distinguish the details. The isolating hyperplane is the hyperplane that increases the gap between the two equal hyperplans. The presumption is that the greater the edge or separation between these equal hyperplanes, the better the classifier's speculation blunder would be [2]. We recognize the structure's knowledge purposes

$$\{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4) \ldots, (x_n,\ y_n)\}.$$

where $y_n = 1/{-1}$, a steady sense of the class, where x_n has a position for that point. $N =$ the test number. Each x_n is a true p-dimensional vector. In order to prepare for vector (ascribes), the scaling is imperative with greater variance. This training knowledge can be seen by methods for isolating (or separating) the hyperplane, which involves

$$w \cdot x + b = o \tag{2}$$

On the off chance that the preparation information is straightly distinguishable, we can choose these hyperplanes so that there are no focuses among them and afterward attempt to expand their separation.

$$\begin{aligned} w \cdot x \, + \, b \, &= \, 1 \\ w \cdot x \, + \, b \, &= -1 \end{aligned}$$

We find the separation between the hyperplane is 2/|w| by mathematics. So, we need to restrict |w|. In order to energies the focus of knowledge, we have to guarantee that either for all *I*.

By math, we discover the separation between the hyperplane is 2/|*w*|. So, we need to limit |*w*|. To energize information focuses, we have to guarantee that for all *I* either

$$w \cdot x_i - b \geq 1 \text{ or } w \cdot x_i - b \, \leq \, -1$$

This can be written as

$$y_i\,(w \cdot x_i - b) \geq 1, \; 1 \leq i \leq n \tag{3}$$

Tests are called support vectors (SVs) along the hyperplane in Fig. 7 isolating hyperplane with the biggest edge characterized by $M = 2/|w|$ that is indicates uphold vectors implies preparing information guides storage rooms toward it.

$$y_j\left[w^T \cdot x_j + b\right] = 1, \; i = 1 \tag{4}$$

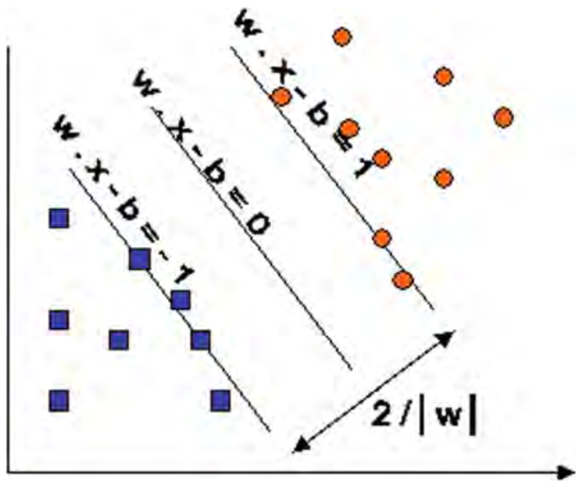


Fig. 7 Maximum margin hyperplanes with samples from two groups trained for an SVM

Ideal canonical hyperplane (OCH) is an authoritative hyperplane having a greatest edge. For all the information, OCH ought to fulfill the accompanying limitations

$$y_i\left[w^T \cdot x_i + b\right] \geq 1;\ i = 1, 2 \ldots l \tag{5}$$

to locate the ideal isolating hyperplane having a maximal edge, a learning machine ought to limit $\|w\|^2$ subject to the disparity requirements.

This enhancement issue illuminated by the seat purposes of the Lagrange's function l

$$\begin{aligned} L_P = L_{(w,b,\alpha)} &= 1/2\|w\|2 - \sum_{i=1}^{1} \alpha_i (y_i (w^T x_i + b) - 1) \\ &= 1/2 w^T w - \sum_{i=1}^{1} \alpha_i (y_i (w^T x_i + b) - 1) \end{aligned} \tag{6}$$

where α_i is a Lagranges multiplier. The quest for an ideal seat focuses (w_0, b_0, α_0) is essential on the grounds that Lagranges must be limited concerning w and b and must be boosted as for nonnegative ($\alpha_i \geq 0$).

4 Analysis for Steganography Method and SVM Classification

To build the security and the size of put away information, another versatile LSB method is utilized. Rather than putting away the information in each most unnoteworthy piece of the pixels, this method attempts to utilize more than the slightest bit in a pixel so that this change would not influence the visual appearance of the host picture. It utilizes the side data of neighboring pixels to gauge the quantity of touch which can be conveyed in the pixels of the host picture to shroud the mystery information.

Step 1: Binaryfunction

$$f(y_i, y_{i+1}) = LSB([y_i / 2] + y_{i+1})$$

Primary1: $f(L-1, R) \neq f(L+1, R)$

Primary2: $f(L, R) \neq f(L, R+1) \neq f(L, R-1)$

$$f(129{,}140) = 1$$
$$f(131{,}140) = 0$$
$$f(130{,}140) = 0$$
$$f(130{,}140) = 0$$
$$f(130{,}141) = 1$$
$$f(130{,}139) = 1$$

$$f(y_i, y_{i+1}) = LSB([y_i / 2] + y_{i+1})$$

$$f(130{,}140) = 0$$
$$f(130{,}140) = 0 \qquad f(130{,}141) = 1$$
$$f(130{,}139) = 1$$

Step 2:

$m_i = LSB(y_i)$ $\qquad$ $m_{i+1} = f(y_i, y_{i+1})$

130	140
x_i	x_{i+1}

10000010	10001100

$$f(130{,}140) = 0$$

0	0
m_i	m_{i+1}

→

130	140

Step3:

$m_i = LSB(y_i)$ $\qquad$ $m_{i+1} = f(y_i, y_{i+1})$

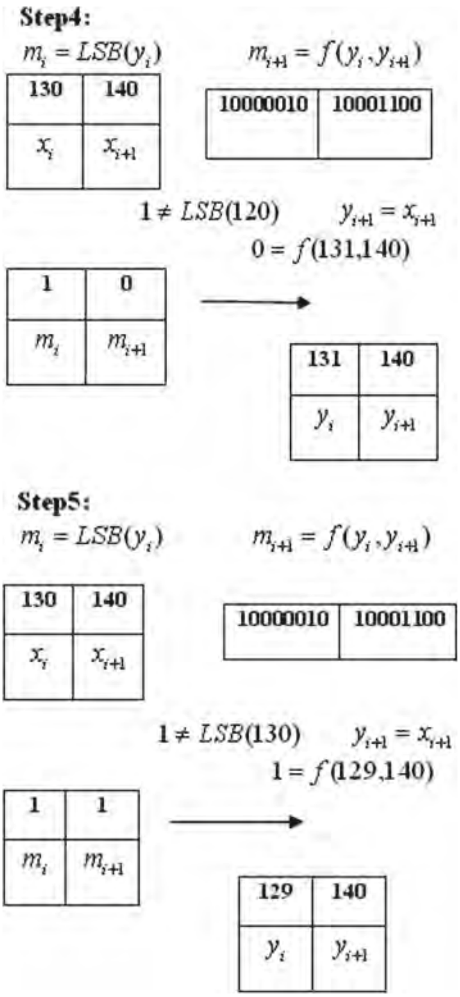
130	140
x_i	x_{i+1}

10000010	10001100

$$f(130{,}140) \neq 1$$
$$y_{i+1} = x_{i+1} \pm 1$$

0	1
m_i	m_{i+1}

130	141
	or139



The ongoing writing accentuates on the impact of AI procedures for the steganalysis applications, yet at the same time just, a couple of writing has examined this technique, leaving a road to investigate the examination, in view of AI. It is discovered that steganalysis utilizing factual methodology would offer promising results when the techniques for AI for picture order is applied. The ensuing segments of this paper talk about the system and the test results for the proposed technique.

The following three tables relate to the different outcomes of the exams. The best estimate of different RBF boundary esteem (C, γ) and cross-approval rate with 5-crease cross-approval using the matrix search method [5], 6 is shown in Table 1. The cumulative execution time for all data to anticipate accuracy in a moment or two is shown in Table 2.

Table 1 Best bargain for various RBF value parameters (C, γ)

Applications	Training data	Testing data	Best g and c with five crease (C, γ)	Cross-validation rate
Diabetes data	600	300	$2^{12} = 3048$, $2^{*8} = 0.008048$	85.7
Heart data	300	70	$2^{6} = 45$, $2^{*8} = 0.008048$	92.6
Satellite data	5435	3000	$2^{1} = 2$, $2^{1} = 2$	95.768
Shuttle data	5350	15,435	$2^{16} = 33{,}748$, $2^{1} = 2$	99.72

Table 2 Execution time using SVM in seconds

Applications	Complete time to predict execution	
Diabetes data	SVM	RSES
	81	24
Heart data	33	7.5
Satellite data	84,849	95
Shuttle data	262,234.1	320

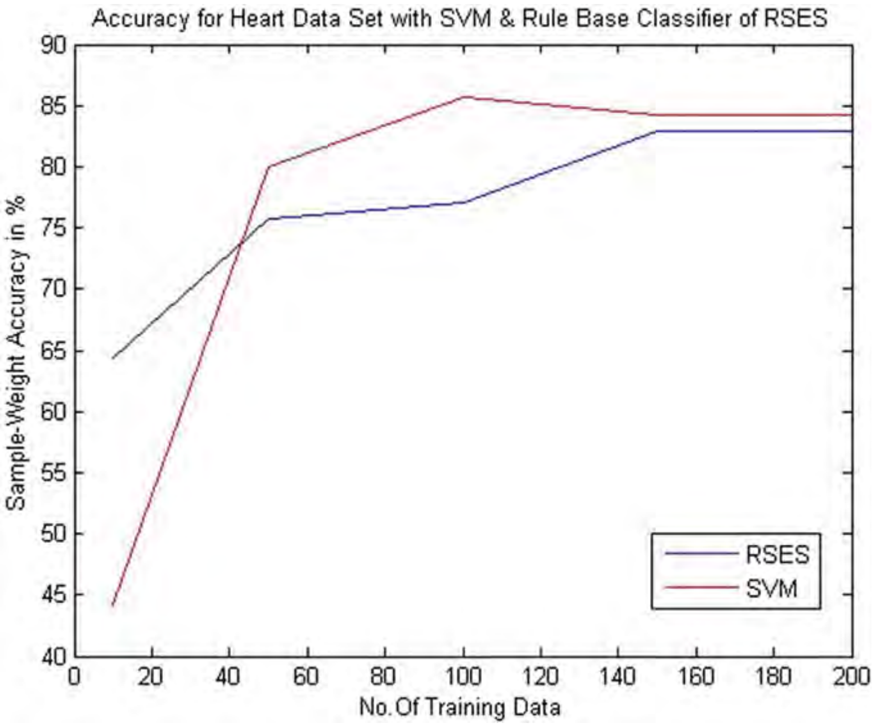


Fig. 8 Heart data accuracy with SVM

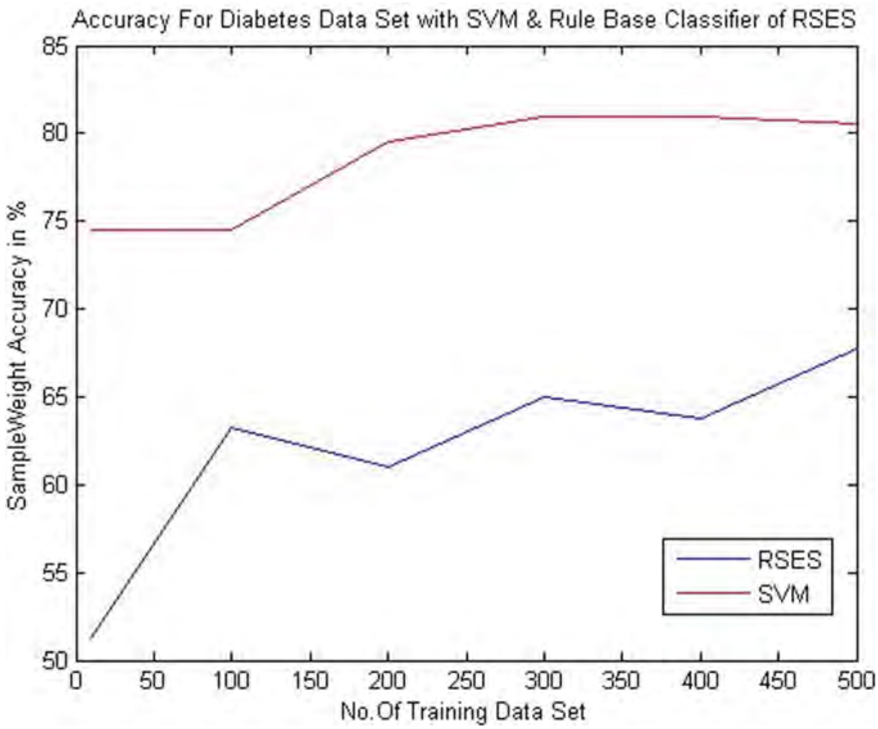


Fig. 9 Accuracy diabetes data with SVM

Figures 8 and 9 indicate the accuracy of diabetes information index correlation following the adoption of diverse preparation set but all tests set for both procedures (SVM and RSES) using RBF component and rule base classifier for work in SVM.

5 Conclusion

The online characterization task is a lot costlier utilizing a straight and non-direct classifier contrasted and a various leveled SVM. Truth be told, each order task requires the Euclidian separation estimation to all focuses in the dataset, which would be a costly expense to cause within the sight of a huge dataset. In this paper, the comparable findings were shown using distinctive bit capacities. The related after effects of different data samples using different sections of direct, polynomial, sigmoid, and RBF are shown in Figs. 8 and 9. The results of the trial are empowering. It can be seen that the choice of portion ability and best approximation of precise bit boundaries are important for a given knowledge measure.

References

1. U.H.-G. Kressel, Pairwise classification and support vector machines, in *Advances in Kernel Methods: Support Vector Learning*, ed. by B. Schölkopf, C.J.C. Burges, A.J. Smola (Massachusetts Institute of Technology Press, Cambridge, MA, 1999), pp. 255–268
2. P. Karthika, P. Vidhya Saraswathi, IoT using machine learning security enhancement in video steganography allocation for Raspberry Pi. J. Ambient Intell. Human. Comput. (2020). https://doi.org/10.1007/s12652-020-02126-4
3. S. Koknar-Tezel, L.J. Latecki, Improving SVM classification on imbalanced data sets in distance spaces, in *ICDM '09: Proceedings of the Ninth IEEE International Conference on Data Mining*, ed. by W. Wang, H. Kargupta, S. Ranka, P.S. Yu, X. Wu (Institute for Electrical and Electronics Engineers, Piscataway, NJ, 2010), pp. 259–267
4. X. Liu, Y. Ding, General scaled support vector machines, in *ICMLC 2011: Proceedings of the 3rd International Conference on Machine Learning and Computing* (Institute of Electrical and Electronics Engineers, Piscataway, NJ, 2011)
5. P. Karthika, P. Vidhya Saraswathi, Content based video copy detection using frame based fusion technique. J. Adv. Res. Dyn. Control Syst. (JARDCS) **9**, Sp-17/2017, 885–894 (2017). Online ISSN 1943-023x
6. Y. Rizk, N. Mitri, M. Awad, A local mixture based SVM for an efficient supervised binary classification, in *IJCNN 2013: Proceedings of the International Joint Conference on Neural Networks* (Institute for Electrical and Electronics Engineers, Piscataway, NJ, 2013), pp. 1–8
7. P. Karthika, P. Vidhya Saraswathi, A survey of content based video copy detection using big data. Int. J. Sci. Res. Sci. Technol. (IJSRST) **3**(5), 01–08 (2017). Print ISSN : 2395-6011, Online ISSN: 2395-602X. https://ijsrst.com/ICASCT2519.php
8. Y. Rizk, N. Mitri, M. Awad, An ordinal kernel trick for a computationally efficient support vector machine, in *IJCNN 2014: Proceedings of the 2014 International Joint Conference on Neural Networks* (Institute for Electrical and Electronics Engineers, Piscataway, NJ, 2014), pp. 3930–3937
9. P. Karthika, P. Vidhya Saraswath, Digital video copy detection using steganography frame based fusion techniques, in *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB). ISMAC 2018. Lecture Notes in Computational Vision and Biomechanics*, ed. by D. Pandian, X. Fernando, Z. Baig, F. Shi, vol. 30 (Springer, Cham, 2019). https://doi.org/10.1007/978-3-030-00665-5_7
10. M. Stockman, M. Awad, Multistage SVM as a clinical decision making tool for predicting post operative patient status, in *IKE '10: Proceedings of the 2010 International Conference on Information and Knowledge Engineering* (Athens, GA, CSREA, 2010)
11. P. Karthika, P. Vidhya Saraswathi, Image security performance analysis for SVM and ANN classification techniques. Int. J. Recent Technol. Eng. (IJRTE) **8**(4S2), 436–442 (2019) (Publisher: Blue Eyes Intelligence Engineering & Sciences Publication)
12. Tax, M.J. David, R.P.W. Ruin, Support vector domain description. Pattern Recogn. Lett. **20**, 1191–1199 (1999)
13. P. Karthika, P. Vidhya Saraswathi, Raspberry Pi—a tool for strategic machine learning security allocation in IoT. Apple Academic Press/CRC Press (A Taylor & Francis Group). Proposal has been accepted (provisionally) for the book entitled "Making Machine Intelligent by Artificial Learning", to be published by CRC Press (2020)

Author Index

S. Smys et al. (eds.), *Computational Vision and Bio-Inspired Computing*, Advances in Intelligent Systems and Computing 1318,
https://doi.org/10.1007/978-981-33-6862-0

A Survey on Graphical Authentication System Resisting Shoulder Surfing Attack

S. Arun Kumar, R. Ramya, R. Rashika, and R. Renu

Abstract In today's modern world, security is a major concern. To provide security, the most widely recognized authentication methods are credentials, OTP, LTP, etc, and these methods are more prone to brute-force attack, shoulder surfing attack, and dictionary attack. Shoulder surfing attack (SSA) is a data theft approach used to obtain the personal identification numbers or passwords by looking over the user's shoulder or by external recording devices and video-capturing devices. Since SSA occurs in a benevolent way, it goes unnoticed most of the time. It is one of the simple and easy methods for hackers to steal one's sensitive information. The hacker has to simply peek in while the user types in the password without any much effort involved. Therefore, this phenomenon is widely unknown to people all over the world. Textual passwords are a ubiquitous part of digital age. Web applications/mobile applications demand a strong password with at least one capital letter and a special letter. People tend to give easy passwords in order to remember them which can be easily shoulder surfed. To overcome this, graphical password techniques are used to provide a more secure password. In the graphical authentication system, the users click on target images from a challenge set for authentication. Various graphical systems have been proposed over the years which are shown to be more secure when compared to other authentication systems. In this paper, an overview of various graphical authentication systems is presented.

Keywords Shoulder surfing attack · Textual password · Graphical authentication system

S. Arun Kumar (✉) · R. Ramya · R. Rashika · R. Renu
Sapthagiri College of Engineering, Bangalore, India
e-mail: aarunkumar889@gmail.com

R. Ramya
e-mail: ramyahai01@gmail.com

R. Rashika
e-mail: rashikarajaraman@gmail.com

R. Renu
e-mail: renurudramurthybt@gmail.com

N. N. Chiplunkar and T. Fukao (eds.), *Advances in Artificial Intelligence and Data Engineering*, Advances in Intelligent Systems and Computing 1133,
https://doi.org/10.1007/978-981-15-3514-7_57

1 Introduction

Security plays a vital role in any organization. Data protection is one of the main challenges faced in any business environment. In order to protect any resources, the companies undertake various security measures. However, security has become a worldwide problem as Web sites have become an integral part of everyone's life [1]. The uncompromising security issues that have to be addressed in Web sites occur during the user authentication phase. In today's computer world, authentication is very important in order to keep the unauthorized users from accessing the protected resources. Authentication is a process that allows a device to verify the identity of a person who connects to a network resource. In order to keep the users' data private, authentication mechanism is used, wherein the user types in the username and password to access their private account. However, people performing authentication process in public results in shoulder surfing attack [2].

Shoulder surfing attack is a direct observation approach where the shoulder surfer steals the user's personal identification number (PIN) and passwords by looking over his shoulder [2, 3]. It commonly happens in public transports while the victim is commuting which involves a smart phone in almost all cases. A good example is shoulder surfing at ATMs, a crime in which a suspect watch over the victim's shoulder as he punches in his PIN number. The ATM screen asks for another transaction when the customers complete theirs. Some customers fail to notice the prompt and walk away leaving it on the screen. In this way, the thief enters the stolen PIN and pretends to be the user. But the phenomenon of shoulder surfing is not widely known [4]. Users tend to use the strategies such as hiding the device screen and shielding the device with their hand. However, by observing, one cannot get a hold with most of the victim's detailed biodata such as information about his relationships, sexual preferences, interests, hobbies, and login data. Hence, the damage shoulder surfing can cause is widely unknown [5].

Textual password approach is used tremendously all over for authentication. During the authentication phase, Web sites demand strong passwords with at least six to eight characters comprising of uppercase and lowercase alphabets, numbers, and special characters. Such passwords are believed to prevent brute-force attacks [6]. A password cannot be remembered if its strength is more. In [7] today's digital age, Web sites play a major role in one's life. People are part of an enormous amount of such Web sites with each containing the authentication phase where the user validates him by entering a password to access their private data. In order to remember all such passwords, the user tends to choose the same password for multiple Web sites which makes the password unprotected for the hackers to break. A more complex password is shoulder surfing resistant. Thus, these passwords can be easily revealed if the shoulder surfer peeks or uses video-recording devices [7].

Graphical authentication systems are used in order to overcome the disadvantages of textual password systems. Here, images are used as the password instead of a string of characters. These graphical passwords are expected to be stronger and safer than textual passwords [8]. Several studies prove that a human brain has a better ability to memorize and recollect images easily when compared to a string of characters. Since it is easy to recollect the password, the user need not choose the same password for multiple Web sites. It makes it hard for the assailants to break the password if the user prefers to use a strong graphical password. This, in turn, increases the security level during the authentication phase. A strong graphical authentication system not only safeguards the password from brute-force and dictionary attacks, but also from shoulder surfing attacks. Since shoulder surfing can create damage to the user during authentication, a strong graphical password is preferred over a textual password according to the studies conducted [9, 10].

2 Technologies for Graphical Authentication

Password-based authentication schemes have been most commonly used on many smart devices when compared to other authentication schemes. The lower complexities in implementation, computation, processing requirements, and so forth have led to the use of a password-based authentication system. Again, text-based passwords are more commonly used when compared to other existing authentication systems. However, various vulnerabilities were discovered by several cryptanalysts in text-based systems like brute-force attack, guessing attack, dictionary attack, and social engineering attack. In smart phones, the tiny screen size imposes some more constraints such as limited password length and implementation of easier authentication systems to increase performance. Moreover, the small on-screen keyboard makes typing inefficient and less precise. Consequently, the users tend to use a smaller password which makes it even more vulnerable. Since the size of smart devices is getting smaller and smaller; few authentication systems cannot be implemented in it due to its size [11].

The invention of graphical password authentication systems was triggered by the well-known limitations of textual password authentication systems. The graphical authentication systems have been generally categorized into draw metrics, loci metrics, and search metrics systems. In draw metrics-based systems, the users will have to recall and reproduce the predefined pattern on a canvas to use the system. In loci metrics-based systems, the users will have to recall and select the previously defined points in an image in order to login the system. In search metrics-based systems, the users will have to choose the predefined target images from the displayed challenge set. During the login phase, the system throws in with entirely the same images or with a few different images which were displayed to the user during the registration phase. The selection of correct target images will let the user access the system. Shoulder surfing has always been a problem in these systems because of the use of the graphical interface [12].

Many authentication systems have been evolved over the years. Today, biometric authentication system holds a prominent place as many users utilize them over textual- or graphical-based authentication systems [13]. However, one study showed that for mobile authentication, 70% of the users preferred PIN or android graphical pattern even though they are more prone to attacks. The users tend to opt for the textual-based method as they do not care about the security but the ease with which they can simply get over with the login phase. Thus, knowing this fact, the attacker will try to break into those systems which use textual- or graphical-based systems. Besides, biometrics would not be the one used for authentication if the users give more priority to the ease of use when compared to other technologies. Biometrics also lacks privacy, reliability, and security. Thus, the existence of PIN and pattern approaches is present even in the overexposure of biometrics [14].

Shoulder surfing can be done in a place jammed with people because it is comparatively easy to stand behind or beside someone and look over his shoulder for information. So, it does not need any technical skills to find the password. In [15], a dynamic pin system is used so that it becomes difficult for the attacker to break even if he observes while the user types in the password. This also requires less memory. It deals with the new method for authentication. It dynamically changes the PIN of the device. The date and time are used as a password. So, we use four digits PIN for authentication. The keylogger which is used for the finding of a password also fails to provide the attacker the ways to authenticate to the user's system. Different format of the pin is also used based on the user's preference. The user can change PIN format like it maybe h1:h2:m1:m2, m1:m2:h1:h2, h2:m1:h1:m2. The system cannot be taken down by the brute-force as the PIN changes from time to time.

Different graphical password authentication schemes are used to mark the defects of a textual password. Humans have more capacity to remember the pictures for longer duration. Since an image-based password is easier to remember, a secured graphical authentication system named pass matrix is implemented [16]. Pass matrix protects the user suffering from shoulder surfing in public places through one-time login indicator. The login indicator which is generated randomly during each phase for pass images will be unused after the session ends. A better security is provided by the login indicator in opposition to shoulder surfing attack, because a dynamic pointer is used by the user to identify the location of their password rather than selecting the password directly. In pass matrix, a part of every image is used as a password from a sequence of n images. In this, the first square is located in the first image, second square in the second image, and so on. In pass matrix, user chooses one grid from each image instead of choosing n grid in the same image. The Cued Click Point (CCP) helps the user to remember and recall their password. If the user clicks on an improper password area within the picture, the login will be failed [16].

Physical QWERTY keyboards are the most commonly accepted input device for personal and movable computing systems. This keyboard is also one of the reasons for shoulder surfing attack. The randomized keyboard then expects the user to type in something which is incorporated with an augmented reality wearable device. The user can see the keys on the randomized keyboard through augmented reality device which is commercially feasible. Different keyboard layout is made visible to the shoulder surfer, wherein he cannot deduce the actual keyboard pattern. It is important to make sure that the keystrokes done by the user cannot be easily identified by the shoulder surfer. Even if he does so, the different keyboard pattern misleads the shoulder surfer from knowing the actual password. An algorithm called individual key randomization (IKR) is used to randomize the keys on the keyboard. An algorithm called row shifting (RS) is used to shuffle the keys row-wise, whereas column shifting (CS) is used to shuffle the keys column-wise. This method overcomes the disadvantages of having a shoulder surfer peak in while the user types in the password. The above three algorithms help the user to efficiently type in the password by misleading the shoulder surfer [12].

On the basis of authentication, graphical password techniques are classified into three major types, namely draw metric, search metric, and loci metric techniques. Searchmetric technique is followed in this paper. This technique aimed at mitigating the obstacles faced in the graphical password schemes. This [17] technique consists of two phases, namely registration and an authentication phase. During the registration phase, the user is expected to enter his valid username and select images from the given set as his password. Every image is associated with three-digit code where this code has to be entered by the user to choose his image along with direction, and the same has to be remembered by the user for the entire process. During the authentication phase, the user is expected to identify the password images and the random code is associated with the images. However, for every authentication session the images will be randomized. This technique uses indirect selection such as choosing the image next to password image called the subordinate image. This subordinate image is decided based on the direction chosen initially during the registration process. The correct identification of the subordinate image for every password image from the given set leads to successful authentication else it directs the user to start the whole process again from the beginning [17].

The proposed ColorPass [18] technique follows the concept of partially observed attacker model where the user can view only the response provided by the system but not the challenges values. Here, the user chooses four pin colors. In the login procedure firstly, the user has to enter his login id and then when the system authenticates the login id, it will generate the feature table on the system that throws some challenge values in the range 1–10 to the user. The feature table can be selected depending upon the challenge values, and further, the color pin has to be selected depending upon the feature table that exactly indicates the color cell. The digit in the color cell has to be identified and submitted as a response to the given challenge by the user. The login process will be completed only after responding similarly to all the other remaining three given challenges. The response given by the user will be evaluated by the system which then the system finally decides if the user is a legitimate user or not [18].

In [19], the phase of registration, from the given various categories the user selects few images, and then in the authentication step, the user is expected to select the correct images which were used in the phase of registration. Registration phase: During this phase, the user has to enter a valid username and is required to select a minimum of 5 images and maximum of 8 images from the given categories. In this case, depending upon the length of the password the user can select only one image from each category on each page, and this process is done by typing the character not by just pointing the mouse toward the image and the same has to continue until the final category. In addition to this, the alphanumeric characters present in each image and the position of images present in each category are randomly organized. Hence, a graphical password consisting of sequence of images will be stored in the database. Authentication phase: During this phase, the user is expected to enter the registered username first and then the image categories will be retrieved by the system that is selected by the user during the phase of registration. The user chooses his image by entering alphanumeric character attached to each image. Here in each page, a fake image named NOT MY PASSWORD is automatically added to provide more security from the attacker. Any one of the images that present in each category replaces the fake image per page. Since the user knows the image selected by him in each category, the user will select the correct image; if the known image is available otherwise, a fake image will be selected in order to ignore that particular category. And also, to make the process more complex, a random category is inserted between selected categories to confuse the attacker who watches this phase of authentication [19].

Among the many shoulder surfing-resistant techniques, loci metric scheme is one among them. This CRASH technique [20] deals with loci metric scheme such as Cued Click Point. In this scheme, there are two phases such as registration phase and login phase. In the phase of registration, the user is expected to enter the credentials such as user id, username, and email id. Then further in the subsequent page, the required category and the number of images in each category will be selected by the user. Based on the selected category, images will be projected. Then an image has to be selected by the user from the given collection which will be followed by a selected image displayed in the grid format. A grid square will be selected by the user as his click point, and this process has to be repeated for certain number of images defined. After selecting the required number of predefined images, registration process gets completed. In the next phase, i.e., login phase based on the order of selection the system starts displaying all images. The user has to click on the same grid square that he clicked during phase of registration for each image. If the point clicked by the user in this phase matches with the registration phase, then the system displays the consecutive image else a wrong image may be displayed, and this process has to be repeated for certain number of images defined. If all the click points entered by the user were correct, then a random string or word will be received via email. In the subsequent page, a different set of images with completely different meaningful words will be displayed. The user has to click on the same string or word that he received via email to successfully complete the login process else the process has to repeat from the beginning. After the failure of three login attempts, the login session expires [20].

Traditional PIN-based authentication schemes are still in use in budget touch screen devices as they can be easily implemented when compared to embedded fingerprint sensor which makes the effortless login. However, this method is less resistant to shoulder surfing attack. In [21], a concept based on merging images called hybrid images is used, wherein this technology simply fools the eyes of the shoulder surfer. The core idea is on the simple observation of the variation in the distance between the screen and the user with that of the screen and the shoulder surfer. The user views the screen from the lesser distance when compared to the shoulder surfer who is at least 0.9 m away from the screen. Taking this into account, a hybrid keypad is implemented. The keypad consists of numbers with each button being the combination of two digits. The shoulder surfer is misled since the button is totally viewed differently by him with varied layouts. Consequently, the extraction of user's PIN becomes difficult. The shuffling of digits is performed in every authentication. This helps in knowing the spatial arrangement of the digits pressed. The hybrid keypad consists of two keypads. One keypad is viewed only by the user called user's keypad and the other which is visible to the shoulder surfer called shoulder surfer's keypad to confuse him. This hybrid keypad is created by using low-pass and high-pass filter parameters. This filtering helps in creating two images. The spatial frequency of both the keypads varies which differentiates the keypad layouts. The algorithm used called visibility algorithm helps to find the minimum safety distance from the user's keypad to the shoulder surfer. Therefore, a false PIN layout is created in order to protect the shoulder surfing attack [21].

This paper [22] presents a more secure pattern-key-based password authentication system where these grid points form the pattern and these grid points only point to the location of number in an integer matrix. The pattern key being the first level is followed by the secret key and then the dummy values at the last. During the registration phase, user will be given a 5 × 5 block grid numbered from 1 to 25. Firstly, the user types in the location number of the pattern. Then in addition to the pattern, the user registers a key for numbers 0–9. A key value ranging from "0 to 9" maps to any integers or characters or to any special characters. Followed by this, the user needs to type in the number of dummy values in this phase where dummy values precede as well as succeed the real password values. These dummy values are named as left and right dummy values. During login phase, after entering the username a next screen appears containing 5 × 5 grid block will appear with randomly generated numbers. The pattern choose in the registration phase has to be remembered by the user to map the key values to the selected password values along with left and right dummy values and then enter the password. If the password matches, it authenticates the user and is able to login successfully else fails in the login process [22].

The user cannot easily recall their password since he has various login ids to remember and might forget his password if he does not use them frequently. Every user chooses an easy password so that they can remember easily. So, the user tends to choose the same password login to all of his accounts which is easy for the hacker to guess. Text-based or alphanumeric passwords are difficult to remember. So, in order to overcome the disadvantages of having an alphanumeric password, graphical passwords are used. In [23], pictures are used as password as the human brain has a capacity to remember hundred images with detail. At the beginning, users are exposed to 50 images out of 70 images, wherein each character is assigned to an image. The shoulder surfer cannot easily identify which character is assigned to which image. These images will be from the random art gallery. Here, the user chooses images that are difficult to relate and are colorful. So, for every 10 characters the user selects a picture that signifies a character. The user's pass images are these 10 images. Also, the user should enter the username. The login phase takes 5 columns and 14 rows of the 70 images. The images displayed during login phase help him to recognize the password character. The account will be blocked after 3 trials. Therefore, [23] provides a secure medium for an authentication system.

3 Comparison Results

The previous works are summarized in Table 1.

Table 1 Summary of prior works

Sl. No.	Title	Approach	Pros	Cons
1	Counterfeit shoulder surfing attack using random pin [15]	Dynamic pin generation	– A good solution for shoulder surfing attack because of dynamic PIN generation – Provides more security – Safeguards from shoulder surfer	– If this method is universally accepted, then the shoulder surfer can easily deduce the password
2	Shoulder surfing-resistant graphical authentication system [16]	Pass matrix – A graphical-based password scheme	– Images are used as password which are more effective compared to textual passwords – Easy to recall and reproduce	– Long process – The user might get frustrated to undergo three phases for the password entry level – Shoulder surfer can deduce the password through concealed cameras
3	Preventing shoulder surfing using randomized augmented reality keyboards [12]	Randomized keyboard	– Additional errors during typing are identified	– The user should always wear the augmented reality devices or glasses
4	PassNeighbor: a shoulder surfing-resistant scheme [17]	Search metric style of authentication	– Shoulder surfer cannot fetch the password images by interacting with user	– The order in which the selection of password images is done as well as the direction chose has to be remembered
5	Color Combo: an authentication method against shoulder surfing attack [18]	Observable attacker model	– Shoulder surfing attack and password guessing attack can be minimized – User-friendly – Consumes less time for login process	– This system does not work for fully observable attacker model
6	Graphical password: shoulder surfing-resistant using falsification [19]	Falsification	– The false image inserted between the real images confuses a hacker while trying to capture the password	– Long process – Several steps should be undertaken during login phase
7	CRASH-Cued recall-based authentication resistant to shoulder surfing attack [20]	Locimetric-based scheme of authentication	– Balanced security and usability features – There exists downtime in both the phases	– Long process – The user might get frustrated to get over with login phase

(continued)

Table 1 (continued)

Sl. No.	Title	Approach	Pros	Cons
8	Illusion PIN: shoulder surfing-resistant authentication using hybrid images	Holographic blending	– Provides high security and can be implemented in budget touch screen devices	– It is too complicated when compared to fingerprint scanner authentication scheme – Third-party applications already use a shuffling scheme which is nearly as complex as Illusion PIN concept
9	Secure pattern-key-based password authentication system	Secured pattern – Key approach	– Highly protected password space – Impervious to shoulder surfing attack and hidden camera	– Three secured features are time-consuming – Remembering a lot of key values during authentication may frustrate the user
10	A graphical password against spyware and shoulder surfing attacks	Graphical password approach	– The images that are hard to explain – Has more password space	– Very complicated images and could not remember easily – More number of images

4 Conclusion

In this paper, the cause for shoulder surfing attack and the prevention methods is put forth. An attempt has been made to contemplate the significance of various graphical authentication systems that have been proposed over the years to overcome shoulder surfing attacks. The methods employed to overcome the disadvantages of textual passwords are presented. The system's advantages and disadvantages are presented for each paper that has been surveyed. The need for graphical authentication systems is emphasized. How textual passwords are more prone to attacks are also looked into. Furthermore, this overview will help various analysts and researchers who are keen on creating graphical authentication systems.

References

1. Gawandi M, Pate S, Snehal P, Said SK (2017) A survey on resisting shoulder surfing attack using graphical password. Int J Adv Res Comput Eng Technol (IJARCET) 6(10):1557–1561
2. Sonar AN, Suryavanshi PD, Navarkle PR, Kukre VN (2018) Survey on graphical password authentication techniques. Int Res J Eng Technol (IRJET) 05(02):26–28
3. Eiband M, Khamis M, von Zezschwitz E, Hussmann H, Alt F (2017) Understanding shoulder surfing in the wild: stories from users and observers. In: The CHI conference on human factors in computing systems (CHI 2017), At Denver, CO, USA
4. Divyapriya K, Prabhu P (2017) Image based authentication using illusion pin for shoulder surfing attack. Int J Pure Appl Math 119(7):835–840 (2018)

5. Choi D, Choi C, Su X (2017) Invisible secure keypad solution resilient against shoulder surfing attacks. In: International conference on innovative mobile and internet services in ubiquitous computing (IMIS). https://doi.org/10.1109/imis.2016.77
6. Nimbalkar P, Pachpute Y, Bansode N, Bhorde V (2017) A survey on shoulder surfing resistant graphical authentication system. Int J Sci Eng 2 (Online). ISSN 2456-3293
7. Rodda V, Kancherla GR, Bobba BR (2017) Shoulder-surfing resistant graphical password system for cloud. Int J Appl Eng Res 12:6091–6096. ISSN 0973-4562
8. Thirupathi J (2015) A comprehensive survey on graphical passwords and shoulder surfing resistant technique analysis. IJISET Int J Innov Sci Eng Technol 2(4):1130–1136
9. Chaudhar D (2015) A survey on shoulder surfing resistant text based graphical password schemes. Int J Sci Res (IJSR) 4(11) (Online). ISSN 2319-7064
10. Pawar M, Mate GS, Sharma S, Gole S, Patil S (2017) A survey paper on authentication for shoulder surfing resistance for graphical password using cued click point (CCP). Int J Adv Res Comput Commun Eng 6(1). ISO 3297:2007 Certified
11. Ranak MSAN, Azad S, Nor NNHBM, Zamil KZ (2017) Press touch code: a finger press-based screen size independent authentication scheme for smart devices. PLoS ONE 12(10):e0186940
12. Ho PF, Kam YHS, Wee MC, Chong YN, Por LY (2014) Preventing shoulder-surfing attack with the concept of concealing the password objects' information. Sci World J vol 2014:1–13
13. Su X, Wang B, Zhang X, Wang Y, Choi D (2014) User biometric information-based secure method for smart devices. Concurr Comput Pract Exp 30(3):e4150. https://doi.org/10.1002/cpe.4150
14. Davin JT, Aviv AJ, Wolf F, Kuber R (2017) Baseline measurements of shoulder surfing analysis and comparability for smartphone unlock authentication. In: CHI 2017. Denver, CO, USA
15. Yogadinesh S, Sathishkumar R, Akash L, Aakash V, Kishore Kumar K, Harichander S (2018) Counterfeit shoulder surfing attack using random pin. Int J Pure Appl Math 118:1757–1761
16. Kannadasan M, Amarnadha Reddy J, Venkat Raman K (2017) Shoulder surfing resistant graphical authentication system. Int J Sci Eng Res 8(5):194–198
17. Saeed S, Umar MS (2016) PassNeighbor: a shoulder surfing resistant scheme. In: International conference on next generation computing technologies. Dehradun
18. Anil G, John CM, Mathew PP (2016) ColorCombo: an authentication method against shoulder surfing attack. Int J Comput Sci Inf Technol Res 4(2):142–147
19. Yeung ALC, Wai BLW, Fung CH, Mughal F, Iranmanesh V (2015) Graphical password: shoulder-surfing resistant using falsification. In: 2015 9th Malaysian software engineering international conference (MySEC), Kuala Lumpur, pp 145–148. https://doi.org/10.1109/MySEC.2015.7475211
20. Sruthi PV (2015) CRASH—Cued recall authentication resistant to shoulder surfing attack. In: 2015 Online international conference on green engineering and technologies (IC-GET), Coimbatore, pp 1–4. https://doi.org/10.1109/GET.2015.7453834
21. Papadopoulos A, Nguyen T, Durmus E (2017) IllusionPIN: shoulder-surfing resistant authentication using hybrid images. IEEE Trans Inf Forensics Secur vol 12, 2875–2889
22. Zaki MH, Husain A, Umar MS, Khan MH (2017) Secure pattern-key based password authentication scheme. In: International conference on multimedia, signal processing and communication technologies (IMPACT), Aligarh, pp 171–174. https://doi.org/10.1109/MSPCT.2017.8363998
23. Darbanian E, Fard GD (2015) A graphical password against spyware and shoulder surfing attacks. In: International symposium on computer science and software engineering (CSSE), Tabriz, pp 1–6. https://doi.org/10.1109/CSICSSE.2015.7369239

Design of 8 Bit Vedic Multiplier Using Urdhva Tiryagbhyam Sutra With Modified Carry Save Adder

CHANDRASHEKARA M N
Department of Electronics and Communication Engineering
Nagarjuna College of Engineering and Technology
Bengaluru, Karnataka, India
chandupesit.te@gmail.com

ROHITH S
Department of Electronics and Communication Engineering
Nagarjuna College of Engineering and Technology
Bengaluru, Karnataka, India
rohithvjp2006@gmail.com

***Abstract*—This paper mainly describes the design of 8-bit Vedic multiplier and its performance comparison with existing multiplier such as i) Booth multiplier ii) Array multiplier iii) Wallace tree multiplier. Vedic calculations are the olden scheme of mathematics, which has a procedure of mathematical calculations to compute the multiplication of two 8-bit number. In this work Urdhva Tiryagbhyam (vertical and crosswise) Vedic sutra is used for multiplier design which provides better performance and consumes lesser time for computation. The Urdhva Tiryagbhyam is the finest sutra and universal one among additional sutras and which represents the different multiplication process compared to normal multiplication. In this work, Modified Carry Save Adder (MCSA) is used to compute the sum of partially generated products. It reduces the computational delay towards the addition of unfinished products. The proposed design uses the Verilog HDL to develop the algorithm. The XILINX 14.7 software tool is used to simulate and synthesize the code. The proposed design is also verified on Spartan-6 Field Programmable Gate Array (FPAGA). Finally, the proposed 8-bit multiplier design is compared with 8-bit Booth multiplier, Array multiplier and Wallace tree multiplier in terms of Area, Memory and Delay. The result shows proposed 8-bit Vedic multiplier is efficient and consumes 14.219ns time for the multiplication process which is better compared to 8-bit, Booth multiplier, Array multiplier and Wallace tree multiplier.**

Keywords*— Urdhva Tiryagbhyam, DCT, FFT, Booth multiplier, Array multiplier, Wallace tree multiplier, MCSA.

I. Introduction

Rapid increase of digital devices, processing of digital data using Digital Signal Processing (DSP) unit is most common. The digital data may be in the form of text, audio, image or video. To process the digital data, multiplication is significant and important arithmetic procedure. The variety of applications related to Discrete Wavelet Transforms (DWT), Discrete Cosine Transforms (DCT), Fast Fourier Transforms (FFT), Convolution and as well as Arithmetic and Logical Unit (ALU), the multiplier is used as a basic block. As the multiplier decides the performance of the system and it provides the serious delay path in signal processing blocks. The multiplier circuit consumes relatively more power in computation. Hence reducing the time delay and power consumption of the multiplication unit is more important for Digital signal processing applications.

Many new multiplier designs are introduced in recent years, namely Array multiplier, Booth's multiplier, Braun multipliers and Wallace tree multiplier etc. In these multiplier algorithms the unfinished products are generated by addition and several other comparisons to obtain the final product. However, the computation speed is low. To improve the computation speed, in this work Vedic mathematics-based multiplier is proposed.

Sri Bharti Krishna Tirthaji (1884-1960) anticipated Vedic arithmetic later in his eight years of investigation on Atharva Vedas [12]. The Vedic mathematics is an attention-grabbing field and presenting several successful algorithms that can be useful to a range of engineering branches. The multiplication processes with Urdhva Tiryagbhyam mathematics algorithm would result in the dropping of computational delay.

In this work, the multiplier utilizes the Urdhva-Tiryakbhyam sutra [12] for multiplication of binary numbers. The major consideration of the design is to improve the speed of multiplier to produce the unfinished products. The unfinished products generated by using Vedic multiplier are subsequently added by MCSA to produce the final product. The estimated multiplier leads to an improved speed, and reduces area occupied by multiplier design.

The rest of the paper is organized as follows. Literature survey is given in Section II. The traditional multiplier designs are discussed in Section III. The Section IV constitutes the detail design of proposed Vedic Urdhva Tiryagbhyam multiplier. The Section V discusses about result analysis of the proposed scheme. The conclusion is given in Section VI.

II. Literature Survey

Many researchers have developed algorithms for multiplication using Vedic multiplier [1]-[6][7] [12]. In [1] realization of high-speed Vedic multiplier by means of Vedic mathematics sutra was discussed. They used Urdhva Tiryagbhyam sutra for the design of the 8-bit multiplier and Carry Save Adder (CSA) used to add the unfinished products to obtain the resultant product. The result shows multiplier utilizes 15.41ns time to multiply the two 8-bit numbers. In [2] Design of Vedic multiplier using higher order compressor to increase the speed and area is discussed. They used compressors to reduce the area and delay. In [3] Design of area and time delay efficient multiplier to obtain better performance of the multiplier is given. The result shows scheme consumes 44.358 ns to produce the final product of the given two 8-bit input data. In [4] Area competent modified Vedic multiplier is discussed. They used Vedic sutras to design the multiplier.

978-1-7281-0630-4/19/$31.00 ©2019 IEEE

The Ripple carry adder used to add the partial products. The result shows it requires 23.18 ns to compute the final product. In [5] design and implementation of 16 x 16 multiplier using Vedic mathematics is discussed. The result shows, it requires 25.159 ns for 8×8 Vedic multiplier. In [6] design and implementation of Vedic multiplier, they proposed new multiplier architecture using the Read Only Memory (ROM) and constant co-efficient multiplier approach, the result shows, it requires 23.87 ns time to compute the multiplication of two 8-bit number.

In this work, Urdhva Tiryagbhyam Vedic sutra used for multiplication of two 8-bit binary numbers. The proposed design intended to improve the speed of 8-bit multiplier to produce the partial products. The modified carry save adder is used to produce the final product by adding unfinished product. Also, the performance of the proposed design is compared with existing multiplier such as i) Booth multiplier ii) Array multiplier iii) Wallace tree multiplier.

III. DESIGN OF DIFFERENT MULTIPLIERS

A. Booth Multiplier [8]

In 1950, Andrew Donald Booth invented the booth algorithm [8] for his cryptography research in London. This multiplication technique involves encoding of multiplier data bits and creation of unfinished products. Booth multiplier architecture is given in Fig.1.

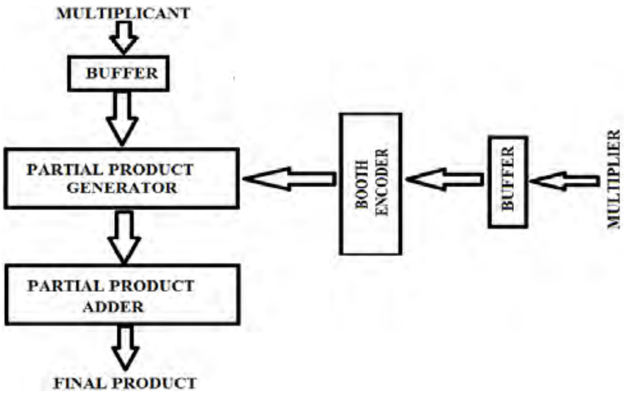


Fig.1. Booth multiplier.

This multiplier converts the negative numbers by 2's compliment method and the encoded binary number is multiplied with the multiplicand and produces the partial products. The unfinished products are added by using the adder. The booth multiplier requires more delay to produce the partial products. This multiplier mainly used to multiply the signed numbers.

B. Array Multiplier [9]

The array multiplier [9] uses the add and shift algorithm. The structure of the 8-bit array multiplier is shown in Fig.2.

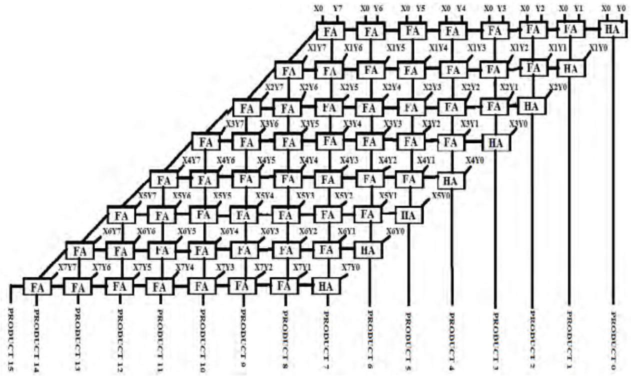


Fig.2. Traditional 8- bit Array multiplier.

The array multiplier can generate the incomplete products by using AND gates, full adder and half adder. For N×N array multiplier, N×N AND gates produce the incomplete products, these can be added by (N-2) × N full adders. The array multiplier performance is degraded by its Area and Delay.

C. Wallace Tree Multiplier

The traditional Wallace tree multiplier [10] structure illustrated in Fig.3. This multiplier uses column compression technique and hence computation speed is more compared to the array multiplier. The total delay is logarithmic of length of binary data. The performance of this multiplier is better with respect to the array multiplier because time delay is linearly varying in array multiplier, where as it varies logarithmically in Wallace tree multiplier. The N-bit multiplier requires N^2 AND gates to generate the partial products. The performance is degraded by its Area.

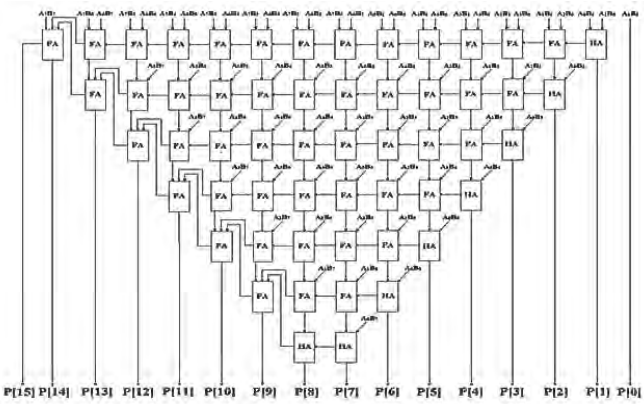


Fig.3. Wallace tree multiplier.

IV. VEDIC MULTIPLIER

A. Vedic Maths

The Vedic math is element of four Vedas. The sound "Vedic" is arrived from the word "Veda" which gives the meaning storehouse of all information [12]. It describes simplified steps in arithmetic, trigonometry, factorization, geometry, quadratic equation and calculus. As earlier discussed, Sri Bharati Teerthaji Maharaja who discovered the Vedic mathematics from Vedas. The swamiji developed sixteen sutras for Vedic maths. Vedic multiplication is not just an arithmetical wonder, and it is a logical too. Because of this reason, this multiplier has such an extent of importance it cannot be able to disapprove. Due to this, the Vedic maths became an important and interesting topic in the mathematics. Vedic maths is particularly simple and very powerful, and which deals with the basic and complex arithmetic operations. The Vedic algorithm depends on the 16 formulae out of which, Urdva Tiryagbhyam Sutra and Nikhilam Navatascaramam Sutra are most widely used for multiplication. This paper mainly concentrates on the Urdva Tiryagbhyam Sutra for the proposed multiplier design. The Table 1 shows most widely used Sutras and their meanings.

TABLE 1. MOST WIDELY USED SUTRAS IN VEDIC MATH [5].

Sl. No	Sutra	Meaning
1	Ekadhikena Purvena	More than one that of previous (division)
2	Nikhilam Sutra	Total from 9 and final from 10 (multiplication)
3	UrdhvaTiryagbhyam	Vertical and Crosswise (multiplication)

B. *Proposed Vedic Multiplier Design Flow*

Vedic multiplier is designed by using Urdhva Tiryagbhyam vertically and crosswise sutra. With the use of this multiplier, the multiplication method can be performed on different kinds of 8-bit numbers. In this work, initially 2×2 Vedic multiplier is designed. Further 2 × 2 multiplier is used as basic block to design 4-bit Vedic multiplier. Then finaly 4-bit multiplier is used to design 8-bit multiplier.

C. *2 × 2 Vedic Multiplier*

The 2-bit Vedic multiplier structure is illustrated in the Fig.4. The vertically and crosswise algorithm is used to design the 2-bit multiplier which requires the basic AND gates and Half adder. Let A and B are the two numbers which are given to the 2×2 multiplier, where A= $\{a_1, a_0\}$ and B= $\{b_1, b_0\}$. The design of 2-bit Vedic multiplier requires four 2-input AND gates and two half adders. The algorithm uses vertical and crosswise multiplication. The Table 2 shows the illustration of steps involved in vertical and diagonal multiplication and its equivalent expression for Urdhva Tiryagbhyam sutra.

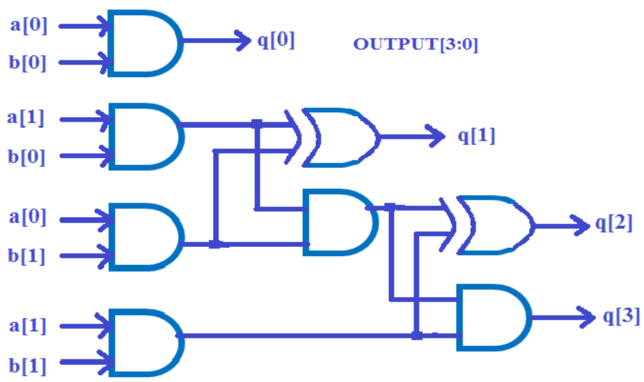


Fig. 4. Two-bit Vedic multiplier using Urdhva Tiryagbham sutra.

TABLE 2. ILLUSTRATION OF TWO-BIT VEDIC MULTIPLICATION ALGORITHMS.

Sl.No.	Pictorial representation of the steps	Equations
1	Step 1: a (1) a (0) b (1) b (0)	q (0) = a (0) *b (0)
2	Step 2: a (1) a (0) b (1) b (0)	q (1)= [a (1) *b (0)] + [a (0) *b (1)]
3	Step 3: a (1) a (0) b (1) b (0)	q (2) = [a (1) *b (1)] + C0

This algorithm entirely different from other traditional algorithms such as Booth multiplier, Array multiplier and Wallace tree multiplier.

D. *4×4 Vedic Multiplier*

The 4-bit Vedic multiplier construction is given in Fig.5. The 2×2 Vedic multiplier is used as a basic component for designing 4-bit Vedic multiplier, in which the partial products are added by a single 6-bit Modified Carry Save Adder. In this multiplier, the concatenation is used to combine the partial product q3 [3:0] with the q0 [3:2] in order to avoid the number of adders so, that the remaining partial products q2 [3:0] and q1 [3:0] are given to the adder.

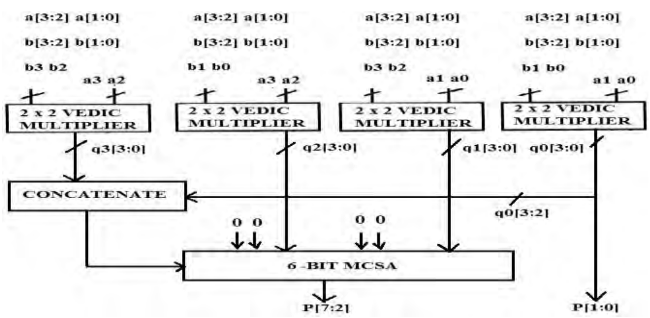


Fig.5. 4-bit Vedic multiplier using Urdhva Tiryagbhyam sutra.

E. *Design of 8 × 8 Vedic Multiplier*

The 8-bit Vedic multiplier using Urdhva Tiryagbhyam sutra is illustrated in Figure 6. The design 4-bit Vedic multiplier using Urdhva Tiryagbhyam sutra is used as a fundamental block for the design of 8-bit Vedic multiplier. In this multiplier, the unfinished products addition is carried out by using 12-bit MCSA adders.

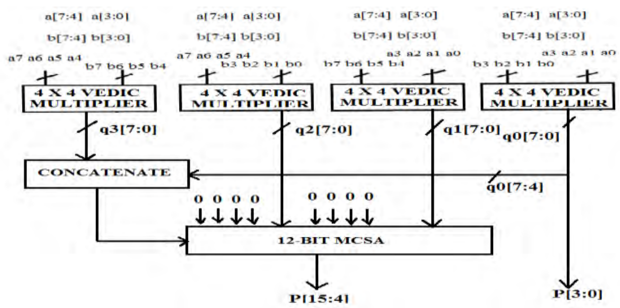


Fig.6. 8-bit Vedic Multiplier using Urdhva Tiryagbham sutra.

F. Modified Carry Save Adder

In any multiplier, the adder acts an extremely vital role in designing the multipliers. The Fig.7 shows the design of MCSA. In this work, the MCSA is used to compute the addition of partial product. The CSA is used to perform 3-bit addition at one time. Here, the three-bit input is processed and transformed to 2-bit output sum and carry at initial stage. In initial stage, outcome carry is not propagated during addition process. In order to generate carry, ripple carry adder is used in second stage for carry propagation.

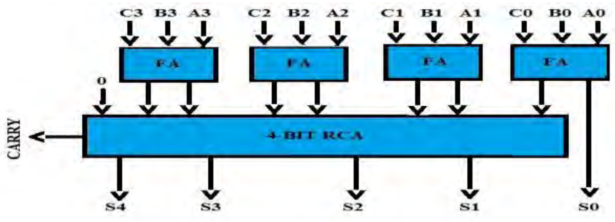


Fig. 7. Design of modified carry save adder.

V. RESULTS ANALYSIS

In this work 8-bit Vedic multiplier, using Urdhva Tiryagbhyam sutra is proposed. The Verilog Hardware Description Language is used to develop the code. The proposed Vedic multiplier is compared with i) Booth multiplier ii) Array multiplier iii) Wallace tree multiplier. To develop, simulate and synthesize the proposed design XILINX ISE 14.7 software is used. In addition, the design is verified on Spartan-6 FPGA hardware. To study the performance of the proposed design, parameters such as area, time and memory is computed. Further design is compared with traditional multiplier.

A. Booth Multiplietr Results

The Fig.8 illustrates the simulation results of 8-bit booth multiplier. To multiply the two 8-bit signed number it requires 25.860 ns computation time.

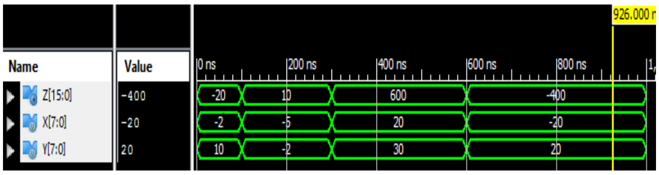


Fig.8. Simulation results of 8-bit Booth multiplier.

B. Simulation Results of Traditional Array Multiplier

The Fig.9. Shows the simulation results of 8-bit Array Multiplier, in which two 8-bit inputs are provided and the corresponding 16-bit outputs are generated. It is observed that for a given input, output is verified. The total computation time required to generate the output is 23.106 ns

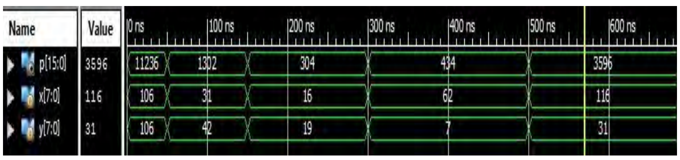


Fig.9. Simulation results of 8-bit Array multiplier.

C. Wallacetree Multiplier Results

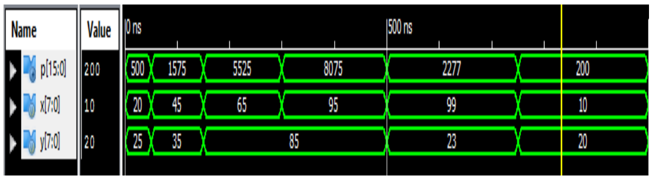


Fig.10. Simulation result of 8-bit Wallace tree multiplier.

The Fig.10 illustrates the simulation result of 8-bit Wallace tree multiplier as discussed in Section III. Simulation results verified for the different 8-bit inputs. It is also observed that to obtain the product of two 8-bit numbers, it requires 16.478 ns computational time.

D. Simulation Results Of Vedic Multiplier

In this section the Technology schematic and simulation results of proposed Vedic, multiplier is discussed. The Fig.11 shows the technology schematic of proposed 8-bit Vedic multiplier.

The Fig.12 shows the simulation results of proposed 8-bit Vedic multiplier. Two 8-bit numbers are given as inputs to the multiplier, which gives the corresponding 16-bit product. In addition, the simulation results of the proposed scheme are mathematically verified for the various inputs. In this work, the Vedic multiplier requires less numbers of adders, to add the partial products which interns reduces the computation delay. To compute the product, the proposed scheme requires 14.219 ns time. The time require to obtain the product is less compared to other multipliers such as i) Booth multiplier ii) Array multiplier and iii) Wallace tree multiplier.

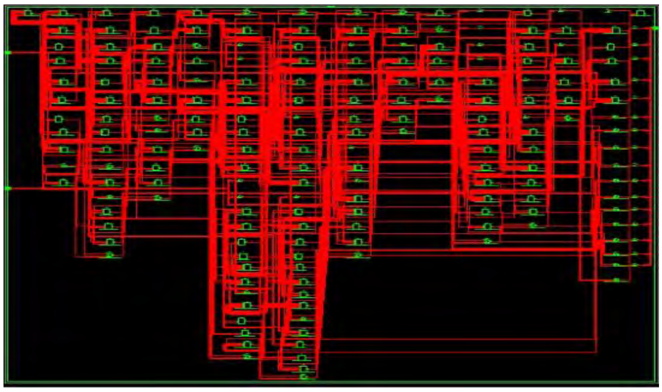


Fig.11. Technology schematic of 8-bit Vedic multiplier using Urdhva Tiryagbhyam multiplier.

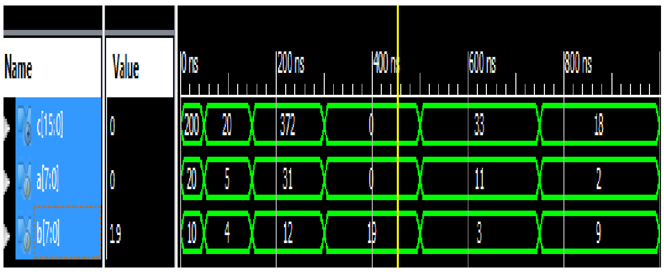


Fig.12. Simulation result of proposed 8-bit Vedic multiplier using Urdhva Tiryagbhyam multiplier.

E. Comparision of Vedic Urdhva Tiryagbhyam Multiplier With Traditional Multipliers.

The proposed Vedic Urdhva Tiryagbhyam multiplier is compared with the traditional multipliers like Booth multiplier, Array multiplier and Wallace tree multiplier with respect to parameter such as Area, Delay and Memory. The Table 3 illustrates the comparison of area occupied, speed and memory utilization of various multiplier architectures.

TABLE 3. COMPARISON OF VEDIC MULTIPLIER WITH ARRAY MULTIPLIER, BOOTH MULTIPLIER AND WALLACE TREE MULTIPLIER

Sl. no	Multiplier used	Lut's used	%of area occupied	Memory in (kb)	Delay in(ns)
1	Booth	216	9%	296640	25.860
2	Array	130	5%	280448	23.106
3	Wallace tree	145	6%	280448	16.478
4	Vedic(proposed)	128	4%	295744	14.219

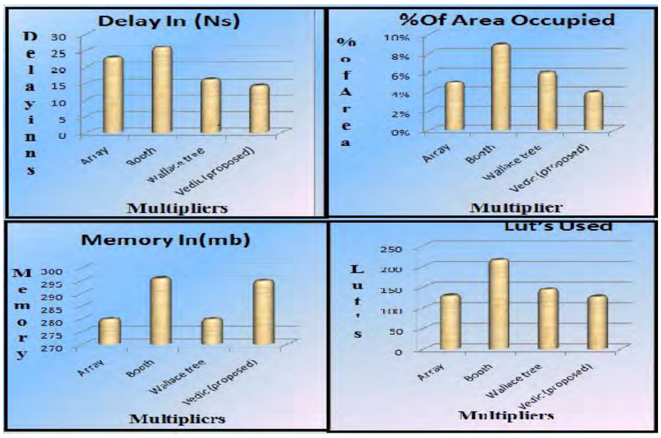


Fig.13. Plot of performance comparison of Vedic multiplier with different multipliers.

The Fig.13 shows the plot of performance comparison of 8-bit Vedic multiplier using Urdhva Tiryagbhyam multiplier with i) Booth multiplier ii) Array multiplier and iii) Wallace tree multiplier for the parameter such as delay, area, memory and number of Look Up Table (LUT). Out of these multipliers Booth multiplier requires more time for product generation among the traditional Array multiplier. The Wallace tree multiplier requires comparatively less time delay for product generation compared to Array and Booth multiplier. Whereas, the proposed Vedic multiplier with Tiryagbhyam sutra requires comparatively less time to compute the product of two 8-bit binary numbers compared to Booth multiplier, Array multiplier and Wallace tree multiplier. The number of LUT's used in all the four cases is relatively same. However, the area consumed in proposed Vedic multiplier is relatively low.

VI. CONCLUSION

In this paper, the design of 8-bit Vedic multiplier using Urdhva Tiryagbhyam sutra and Modified Carry Save Adder is discussed. The proposed multiplier is simulated and synthesized by Xilinx 14.7 on the device Spartan-6, its performance is compared with existing multiplier such as i) Booth multiplier ii) Array multiplier and iii) Wallace tree multiplier. The results show 8-bit Vedic multiplier using Urdhva Tiryagbhyam sutra with MCSA provides high speed in computing products of two 8-bit binary numbers with compared to Booth multiplier, Array multiplier and Wallace tree multiplier and it consumes 14.219 ns to compute the product of two 8-bit numbers. Hence, the proposed 8-bit Vedic multiplier is more efficient than other type traditional multipliers. Hence, the proposed multiplier can be used for the Digital signal processing applications where the speed is significant.

REFERENCES

[1]. G. Ganesh Kumar, V. Charisma. "Design of High-Speed Vedic Multiplier using Vedic Mathematics Techniques" International Journal of Scientific and Research Publications, Volume 2, Issue 3, March 2012, 2 ISSN 2250-3153.

[2]. S.K. Peda Roshan Jameer, V. Pratap Reddy, "Design of Vedic Multiplier Using Higher Order Compressor to Increase the Speed and Area", international journal of innovative Research in Science and Engineering, vol. No. 2 Issue 01, January 2016.

[3]. G. R. Gokhale , S. R. Gokhale "Design of Area and Delay Efficient Vedic Multiplier Using Carry Select Adder", Proceedings of IEEE International Conference on Information Processing (ICIP). Pune, India, 16-19 Dec. 2015

[4]. G. Challa Ram ; Y. Rama Lakshmanna ; D. Sudha Rani ; K. Bala Sindhuri "Area Efficient Modified Vedic Multiplier" Proceedings of IEEE International Conference on Circuit, Power and Computing Technologies (ICCPCT), Nagercoil, India, 18-19 March 2016. 978-1-5090-1277-0/16/$31.00@2016IEEE.

[5]. S.P. Pohokar, R.S. Sisal, K.M. Gaikwad, M.M. Patil, Rushikesh Borse, "Design and Implementation of 16 x 16 Multiplier Using Vedic Mathematics", Proceedings of IEEE International Conference on Industrial Instrumentation and Control (ICIC), Pune, India, 28-30 May 2015

[6]. C.Sheshavali, K Niranjan Kumar, "Design and Implementation of Vedic Multiplier", International Journal of Engineering Research and Development e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com Volume 8, Issue 6 (September 2013), PP.23-28

[7]. Dravik KishorBhai Kahar, Harsh Mehta, "High speed vedic multiplier used vedic mathematics", Proceedings of IEEE International Conference on Intelligent Computing and Control Systems ICICCS, Madurai, India, 15-16 June 2017.

[8]. Andrew D. Booth. "A signed binary multiplication technique" The Quarterly Journal of Mechanics and Applied Mathematics, Volume 4, Issue 2, 1951, PP 236–240.

[9]. Baugh, Charles Richmond; Wooley, Bruce A. "A Two's Complement Parallel Array Multiplication Algorithm". IEEE Transactions on Computers, Volume. 22, Issue.12, 1973, PP 1045–1047.

[10]. C. S. WALLACE, "A Suggestion for a Fast Multiplier", IEEE Transactions on Electronic Computers, Volume EC-13 , Issue: 1,1964 , PP 14 – 17.

[11]. Dravik kishorbhai Kahar, Harsh Mehta," High Speed Vedic Multiplier Used Vedic Mathematics", Proceedings of IEEE International Conference on Intelligent Computing and Control Systems, Madurai, India, 15-16 June 2017 PP 356 – 359.

[12]. W.B. Vasantha Kandasamy; Florentin Smarandache "Vedic Mathematics: Vedic Or Mathematics: A Fuzzy and Neutrosophic Analysis "American Research Press. ISBN 978-1-59973-004-2. Retrieved 23 May 2013.

An Novel Hand Gesture System for ASL using Kinet Sensor based Images

[1]Manoj H. M., , [2]Pradeep Kumar B.P., [3]Anil Kumar.C.[4] Rohith.S
, 2nd author's name and surname[2], etc
{manojhm@bmsit.in[1], pradi14cta@gmail.com[2], canilkumarc22@gmail.com[3], rohithvjp2006@gmail.com[4]}

[1] Assistant Professor, Dept of CSE, BMSIT&M, Bangalore-64., [2] Associate Professor, Dept of ECE, HKBKCE,Bangalore-45, [3] Associate Professor & HoD, Dept of ECE., RLJIT, Doddaballapur, [4]Associate Professor, Dept of ECE.NCET, Bangalore

Abstract. The prime goal of this study is to perform resolution enhancement using experimental analysis. The framework presents hand image resolution enhancement techniques based on multi scale decomposition and edge preservation smoothing. The proposed technique Dual Tree complex wavelet transform (DT-CWT) and edge preservation smoothing (EPS) algorithm images are decomposed into different sub bands and interpolated, after which sub bands are reconstructed to achieve the enhanced image. The study outcome of this phase is studied with respect to PSNR and RMSE for all the letters of ASL.

Keywords: American Sign Language, Camera, Gesture, Wavelet transform, kinet sensor.

1 Introduction

Gestures used for communicating between person and machines also between persons using sign language [5][6]. Gestures can be static (posture / certain pose) or dynamic (series of postures). The static-gestures require less computational complexity whereas dynamic-gestures are more complex. Various techniques have been developed for acquiring necessary information for gestures recognition system [4][6]. Some methods are utilized for external hardware-devices such as data-glove devices and colour-markers which can easily extract the comprehensive-description of Gesture-features. Other methods are based on the appearance of hand which segments the hand and extracts the essential features, these methods are considered as natural, easy and less cost effective than other [7][8].

For most hard of hearing people in the United States, American Sign Language (ASL) is the preferred dialect. For everyday terms, ASL employs approximately 6,000 gestures, with finger spelling for conveying dark words or structured objects, locations, or items. In ASL, communication is often based on available shapes placed in or transferred crosswise over various areas of the endorser's body, despite changes in the head and arm, as well as physical appearance [12]. In any case, proper names and words without a unified sign are spelled in English letter by letter, and ASL understudies often begin their studies by learning the 26 hand shapes that make up the manual letter range [12, 13].

I3CAC 2021, June 07-08, Chennai, India

DOI 10.4108/eai.7-6-2021.2308609

1.1 Hand Gesture Recognition System

The vision is one of the physical senses which computer is instantiated perceptibly during the communication with humans. Thus, the vision based mechanisms are considered more in HGR. The HGR system based on computer vision consists of three different steps and is shown in Figure.1. The typical process includes a) image enhancement to remove noises as a pre-processing step, b) segmenting the palm portion from the captured image of either human body or hand ,c) feature extraction and finally d) classification [14, 15].

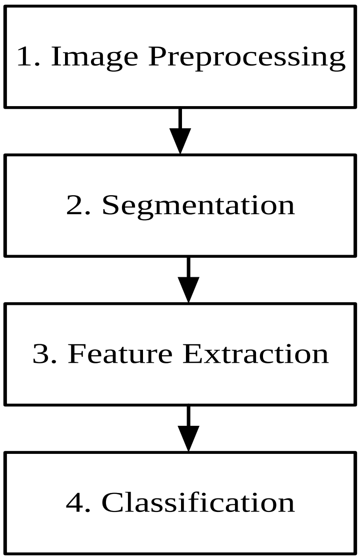


Figure.1. Main Steps of HGR System based on computer vision

1.2. Modeling Hand Gesture

The 2D modelling of hand can be mentioned with motion, shape and deformable templates. The shape based hand modelling can be classified as non-geographic and geographic models [2][3]. The non-geographic models consider the shape based features to model the hand like edges, contour, Eigen vectors etc, which are used for feature extraction and perform analysis too. The flexible/deformable models give an object shape changing flexibility level to pass the little variation of hand shape.

The 3D model of the hand can be represents and classified into skeletal, geometrical and volumetric models. The geometric models can be utilized in the real-time applications and hand animation. The skeletal models needs less parameters to structure the hand shape. The volumetric models are very complex and needs more parameters to shape the hand [14]. The geometric surfaces significantly performs the simulation of the visual hand image but it needs more parameters and is more time consuming process. The visual shapes like cylinders and ellipsoids are the alternative mechanism of geometric shape [15]. The Figure.2, describes the hand modelling mechanisms

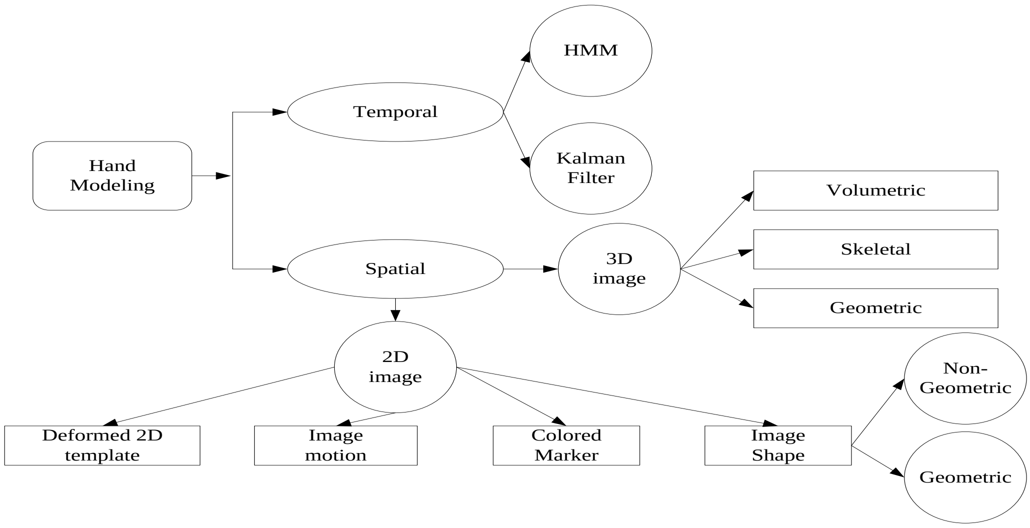


Figure.2. Hand Modelling

1.3 Research Aim And Objectives

A novel system for contact-less Hand gesture recognition using Microsoft Kinect for Xbox is described, and a real-time Hand gesture recognition arrangement is modelled and simulated into numerical computing platform (Matlab). The arrangement permits the user's to choose a situation, and it is clever to notice hand motions prepared by users. To recognize fingers, and to identify the meanings of gestures, and to show the meanings and pictures on screen. The prime aim of the proposed research work is to design a simple framework that can offer enhanced performance of hand gesture recognition system in effective manner considering ASL. In order to accomplish this research goal, following objectives were set:

- ***Preprocessing:*** To design a model that can offer enhanced resolution for input images
- ***Feature Extraction:*** To develop a simple modeling for hand gesture recognition emphasizing on an efficient feature extraction.
- ***Classification:*** To develop a hybridized scheme of hand gesture recognition for increasing recognition performance.
- ***Optimization***: To apply optimization for enhanced performance of hand gesture recognition system in cost effective manner.

2 Resolution Enhancement Of Hand Gesture Images

The hand gesture recognition (HGR) system which mainly emphasizes on the limitations of traditional hand gesture recognition techniques. This section mainly argues two-level architecture for the real-time hand gesture recognition scheme using only one camera as the input device. second describes the resolution enhancement problem and technique for hand gesture images using four different algorithms namely 1) The Nearest Value Algorithm, 2)

The Bilinear Algorithm , 3)The Bi-cubical Algorithm , 4) Dual Tree Complex Wavelet Transformation (DTCWT), further Bilateral Filter is used in order to obtain enhanced performance.

2.1 Skeleton Identification Of Kinect Camera

In this proposed system, "Kinect camera" plays the major role to gather the depth information from the skeleton. The new version of Kinect with its SDK (Software Development Kit) containing the skeleton tracking tool. This unique tool provides the system to collect the 20 joint information of the human body. For each frame, the positions of 20 points are estimated and collected. The 20 joints which is taken as an reference points is as shown in Fig 3

The kinect device provides the both RGB and D-image. This camera utilizes a structured light method to generate the real time depth information which consists of discrete measurements of physical scene. In this study, first creating the depth images of human in different sizes, and shapes, and generate the big dataset. The RGB and D-image are the input images of the system for the recognition of different ASL alphabetic symbols. This skeletal tracker device able to track the skeleton image of one or more persons moving within the kinect area view for gesture driven applications i.e. this tool able to collect 20 joint information about the skeleton. From the skeleton tracked by the Kinect first it extracts the feature of joint positions. Since, each joint has 3 values and also 3 coordinates and the detected skeleton has of 20 joints. So, the feature vector has 60 dimensions. The position of 20 points identified by kinect afterwards it will segment the right hand posture to recognize and stored in the database. From this sets can select the wanted joints of images for representing the postures.

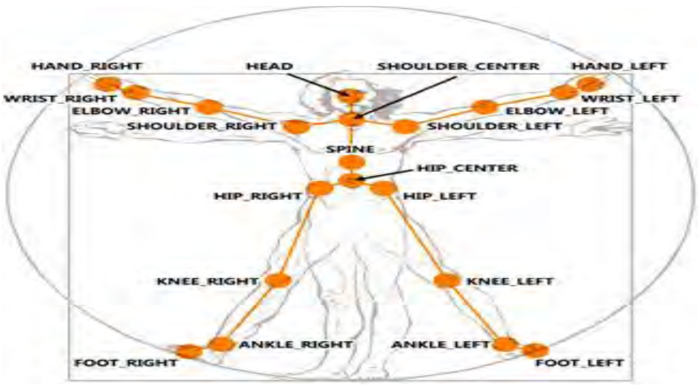


Figure. 3 Human skeleton joints as reference points

2.2 Algorithm: Skeleton identification from Kinect-Sensor.

Input: one or more people image, Output: 20 joint images

a. start
b. capture the image from kinect camera
c. segment the image from skeleton viewer
d. for reference points in body portion finds the depth-information
e. segmented moving body portion is mapped to the skeletal co-ordinates
f. if more than one moving body portion presents
g. calculate skeleton connection map for x,y coordinate
h. display the multiple skeleton moving body
i. end of predefined frames
j. stop.

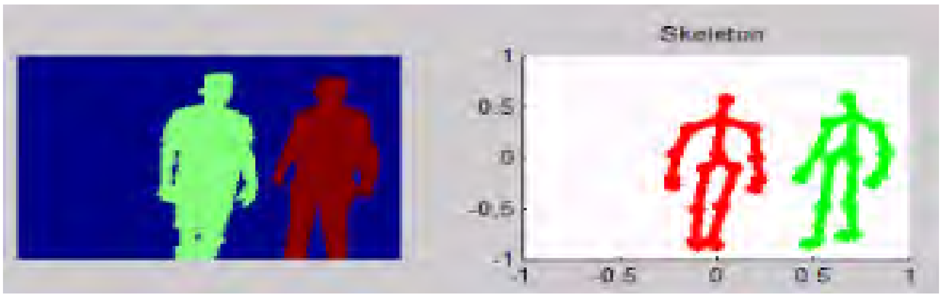


Figure.4. Skeleton-image identification from kinect sensor

Figure 4 shows the outcomes of identifying multiple skeletal signs under different distance using kinet-sensor device which represents the colour and depth image. Figure 5 shows in multiple user environments also our proposed system is identifying the user hand with respect to the distance from camera and extract the hand sign clearly for storing in to database.

Figure. 5. Multiple skeletal sign recognized under different distance

2.3. Multiple Depth Recognition:

(i) To calculate centroid: Here system is considering the three co-ordinates X, Y and Z. Calculating the centroid for each axis independently the mathematical interpretation for centroid is given by

X= sum (x) / length(x)..................... ....… 1

Y= sum (y) / length (y)………………………………….....2

Z= sum (z) / length (z)…………………………..............3

(ii) **To calculate mean:** The system will calculate the mean for the entire segmented region calculated by the centroid of the body part by pixel basis. The mathematical interpretation for mean for pixels is given by:

$\bar{x}=\frac{1}{K}\sum_{i=0}^{K} x(i)$ and $\bar{y}=\frac{1}{K}\sum_{i=0}^{K} y(i)$……………………......4

The flow graph in figure 6 shows for the multiple depth recognition using Kinect sensor is as follows

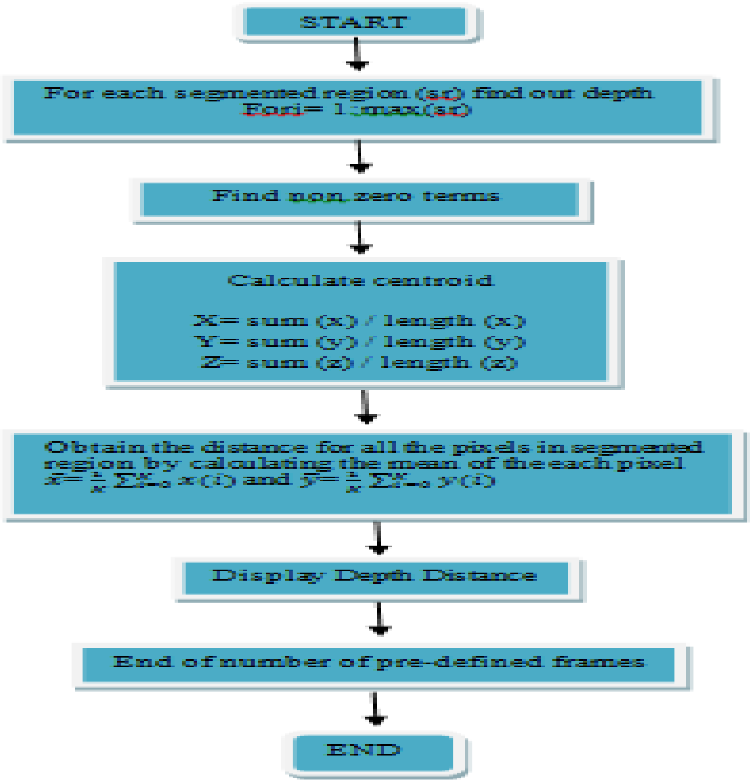


Figure. 6 The flow graph for the multiple depth recognition

2.4. Data Acquisition and Pre-Processing

The raw information acquired from the Kinect sensor via the Natural User Interface (NUI) contained 512×424 depth data, 1920×1080 RGB data, and 26-joint body skeleton data. The hand region in the depth image was illustrated using spatial thresholds in X-axis direction [Tx_min, $Tx\_$], Y-axis direction [Ty_min, Ty_max] and Z (depth)-axis direction [$TDepth\_min$, $TDepth\_max$]. As demonstrated in Figure 6, the Kinect depth sensor placed at position $\boldsymbol{S}$ has angles of view α (horizontal) and β (vertical). The declaration of the depth image is $RxbyRy$ pixels. The position of the “hand” joint (x, y, D) in the depth image can be attained from the Kinect skeleton data (Figure 8a). Thus, the spatial thresholds are illustrated as:

$$[T_{x\_min}, T_{x\_max}] = [x - \frac{d_x}{2}\frac{\frac{R_x}{2}}{Dtan\frac{\alpha}{2}}, x + \frac{d_x}{2}\frac{R_x/2}{Dtan} \qquad ....5$$

$$[T_{y\_min}, T_{y\_max}] = [y - \frac{d_y}{2}\frac{R_y/2}{Dtan\frac{\beta}{2}}, y + \frac{d_y}{2}\frac{R_y/2}{Dtan} \qquad ....6$$

$$[T_{Depth\_min}, T_{Depth\_max}] = [D - \frac{d_z}{2}, D + \frac{d_z}{2}] \qquad ....7$$

where dx , dy and dz are stable dimensions (in millimetres) of the hand's region. The hand's region in the depth image is revealed in Figure 7. The hand's region in the color image can also be gained by mapping the hand's region on top of the color image (Figure 8 a).

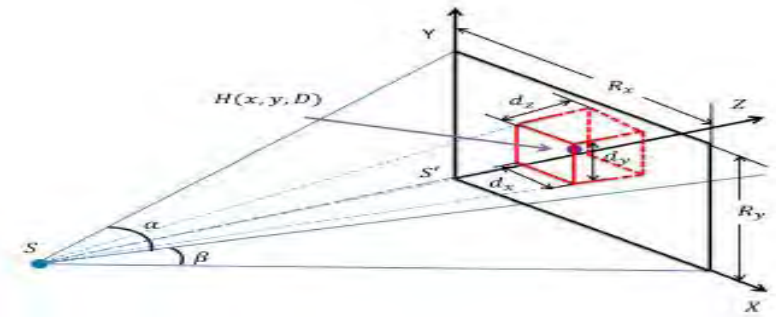


Figure. 7.. Illustration of the hand region segmentation: the$dx \times dy \times dz$hand region at
(x, y, D) was segmented from the$Rx \times Ry$depth image obtained using a depth sensor located at the position S.

(a) (b)

Figure. 8. Illustration of data obtained using Kinect.
(a) RGB Color image of the hand region. (b) Depth image of the hand region.

The following table illustrates the result analysis of proposed system can measure the recognition accuracy in different distance ranges. As sample we are experimenting skeletal signs as 10 times with different distance like 850mm to 1000mm …3000mm to 3500 mm and evaluating the recognition accuracy with time. The recognition accuracy is calculated in terms of percentage like for first experiment we are considering the (skeletal and camera) distance as 850mm to 1000mm and getting the 70% of recognition accuracy. Like this from the experiment analysis results, we can get the following results, which is shown is following table 1.

Table. 1 analysis results for recognition accuracy calculation

Distances from Kinect In mm	Number of times checked	Number of times recognised	Recognition Accuracy in %
850-1000	10	7	70 %
1000-1500	10	10	100 %
1500-2000	10	10	100 %
2000-2500	10	10	100 %
2500-3000	10	10	100 %
3000-3500	10	8	80 %

3. Framework For Resolution Enhancement

In this research methodology using experiential analysis image enhancement is executed as shown in fig 9. The techniques used here are Edge preservation smoothing and multi-scale decomposition. To obtain the enhanced image, the method of DT-CWT and EPS algorithm images are decomposed to a different sub band. At a regular interval of 5 minutes, samples of sign language are recorded from Kinect camera at a distance of 1500-2000mm. Values of PSNR, RMSE, are used to further proceed with quantitative analysis. The output of this phase is an enlarged image.

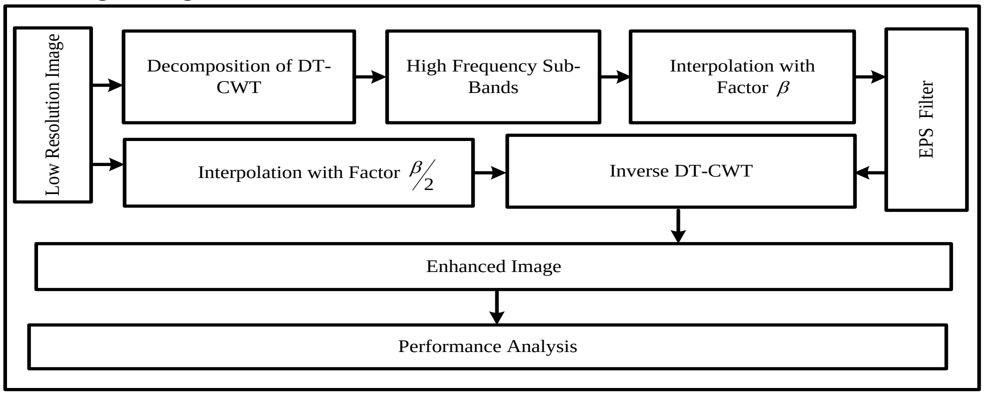


Figure.9. Block diagram representing framework for resolution enhancement

3.1. The Nearest Value Algorithm

Nearest Neighbour interpolation or proximal interpolation is way by which multiple dimensions can be interpolated. For a random point in space surrounding the nearest point (Neighbouring) would have its value approximated leading to the interpolation problem. Hence, the algorithm of the nearest neighbour selects the nearest point and neglects the value of points around it.

The above mentioned is the nearest neighbour algorithm which selects the nearest point eliminating the value associated to the random points around it. The original disk on file (I) is initialized and the size of the resized image is computed (I_{new}).Now, the size of the original file on disk is calculated. Further to check a condition that size of I and I_{new} match or not the number of rows and columns are compared respectively. If the resized image has higher number of rows then its new value will be ranging from 1-to-rows of original image considering first row and first column. If the original image has higher number of rows then its new value will be ranging from 1-to-rows of resized image considering first row and first column. Same procedure is applied to check the condition for the columns. The value of PSNR is calculated to be 23.3591, RMSE is 17.3214 for the input image as shown in fig 10.

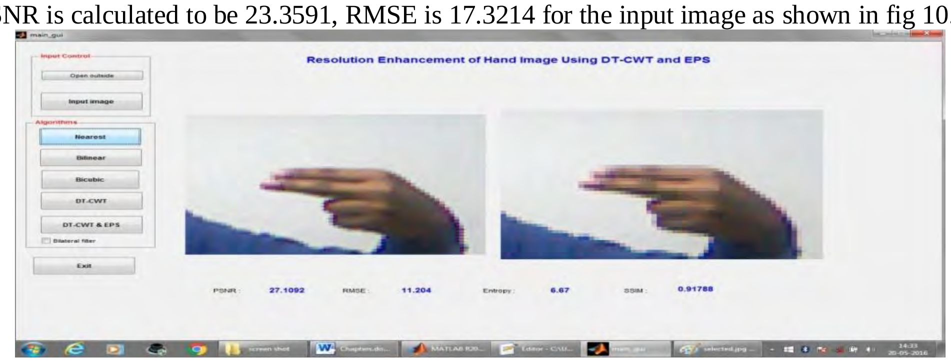


Figure.10. Application of Nearest Neighbor Transformation Function

3.2. Bilinear Interpolation

An image transformation process used in cases where pixel matching is impossible is called as bilinear interpolation. When compared to other methods of transformation bilinear interpolation considers closet 2x2 neighbourhood of known pixel values surrounding the unknown pixel's computed location.

Computation of PSNR $(\beta)=10\log_{10}\left(\frac{255}{\alpha}\right)$; Computation of RMSE $(\gamma)=\sqrt{MSE}$

The above mentioned is the bilinear interpolation method which considers 2x2 neighboring pixels. The original disk on file (I) is initialized and the size of the resized image is computed (I_{new}).Now, the size of the original file on disk is calculated. Further to check a condition that size of I and I_{new}match or not the number of rows and columns are compared respectively. If the resized image has higher number of rows then its new value will be ranging from 1-to-rows of original image considering first row and first column. If the original image has higher number of rows then its new value will be ranging from 1-to-rows of resized image considering first row and first column. Same procedure is applied to check the condition for the columns. The value of PSNR is calculated to be 23.591, RMSE is 17.3214 for the input image as shown in fig 11.

Figure. 11. Application of bilinear Transformation Function

3.3. Bicubic Interpolation.

This technique is implemented using Lagrange polynomials cubic splines or cubic convolution algorithm. Bicubic interpolation can be chosen over other methods if speed is not a constraint. Smoother images are obtained as output with lesser interpolation artifacts. The above mentioned is the bicubic interpolation method which is based on cubic convolution algorithm, Lagrange polynomials. The original disk on file (I) is initialized and the size of the resized image is computed (I_{new}).Now, the size of the original file on disk is calculated. Further to check a condition that size of I and I_{new}match or not the number of rows and columns are compared respectively. If the resized image has higher number of rows then its new value will be ranging from 1-to-rows of original image considering first row and first column.If the original image has higher number of rows then its new value will be ranging from 1-to-rows of resized image considering first row and first column. Same procedure is applied to check the condition for the columns. The value of PSNR is calculated to be 24.953, RMSE is 17.3214 for the input image as shown in fig 12.

Figure.12. Application of bicubic Transformation Function

3.4. Bilateral Filtering Process

Bilateral filter is a noise reducing filter for images with the property of non-linearity and preserving the edge. Each pixel has an intensity value a picture restored by weighted average from nearby pixels. Weight can be based on Gaussian distribution and the sample output is shown in fig 13.

The bilateral filter is defined as:

$$I^{filtered}(x) = \frac{1}{W_p} \sum_{xi \in \Omega} I(x_i)\, fr(\| I(x_i) - I(x) \|) g_s(\|x_i - x\|) \qquad \ldots\ldots 8$$

Where the normalization term;

$$W_{p} = \sum_{xi \in \Omega} fr(\| I(xi) - I(x) \|) gs (\| xi - x \|) \quad \ldots\ldots 9.$$

Ensures that the filter preserves image energy and

- $I^{filtered}$ is the image after filtration: Iis the original input image
- xare the directs of the current pixel to be filtered: Ω is the window centered in x
- f_r is the range kernel for smoothing differences in intensities. This utility can be a Gaussian function
- $g_s$is the spatial kernel for smoothing differences in coordinates. This function can be a Gaussian function.

	
Input Image for Bilateral filter	Output Image for Bilateral filter

Figure 13.Application of bilateral filtering process

3.5. Dual Tree Complex Wavelet Decomposition Transformation (DTCWT)

Using the dual tree complex wavelet transformation method the image is decomposed. This happens with respect to Discrete Continuous Wavelet Transformation (DWT) and Continuous Wavelet Transformation (CWT) .In DWT the basis function used is symlet mother wavelet. Image will be decomposed into two parts, the approximation coefficients and detailed coefficients. Only approximation coefficients are considered and similar mechanism is implemented for CWT. The algorithm of DTCWT is being performed above for the resolution enhancement for a hand gesture by splitting the image mainly into real and imaginary parts. To evaluate the analysis and synthesis parameter taken into consideration, the dual filter function is worked on. The dual cell structure is determined via cplx dual 2D function. Frequency values are divided into lower and higher components. The higher frequency components are normalized to nullify the effect of frequencies lying outside the desired range of detection. The lower frequencies have their highest value used in the algorithm among all of them. Further lower frequency image is converted into original image with the inclusion of the new dual tree cell structure. for the input image as shown in fig 14.

Figure.14. Application of DTCWT and EPS algorithm

Table.2. Comparison of algorithms for sign H

Algorithms	PSNR	RMSE
Nearest	26.24	12.37
Bilinear	28.08	10.01
Bicubic	28.11	10.11
DT-CWT	28.50	13.8
DT-CWT & EPS	29.07	10.006

Table 3 showing the detailed tabulated parametric values for all the signs using DT-CWT conditions respectively considering performance parameters of PSNR, RMSE

Table.3 Numerical Outcome of Performance parameters of DT-CWT

ALPHABET	A	B	C	D	E	F	G	H	I
PSNR	28.5	29.61	30.35	28.96	29.02	29.14	28.84	26.24	30.77
RMSE	13.8	8.66	8.63	10.49	10.9	10	12.38	12.37	8.44

ALPHABET	J	K	L	M	N	O	P	Q	R
PSNR	28.82	29.45	28.45	28.33	29.43	29.77	30.12	29.25	29.43
RMSE	12.15	12.15	11.15	9.6	11.56	10.32	7.85	11.9	8.43

ALPHABET	S	T	U	V	W	X	Y	Z
PSNR	29.59	29.89	30.45	27.78	27.32	27.55	28.56	26.72
RMSE	8.8	8.99	9.2	12.73	9.34	13.92	12.44	14.6

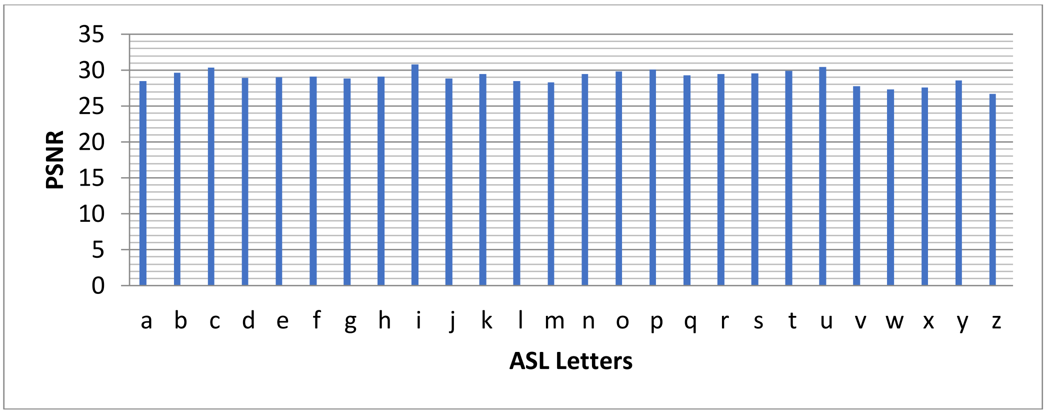


Figure.15. Illustration of PSNR values for the algorithm DT-CWT and EPS.

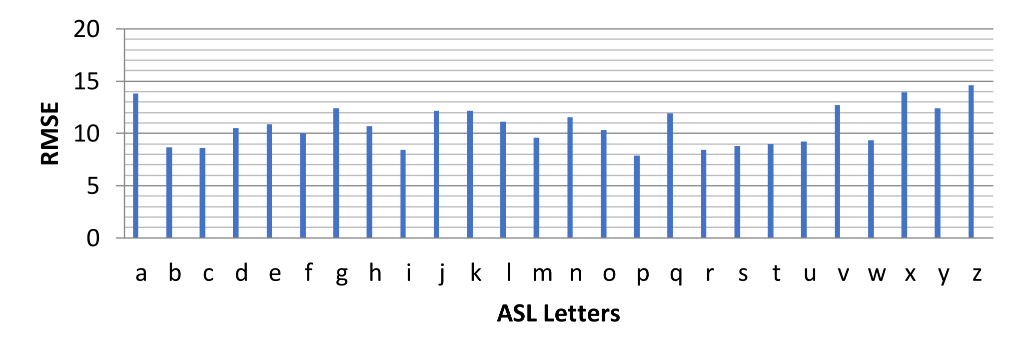


Figure.16. Illustration of RMSE, values for the algorithm DT-CWT and EPS.

4. Conclusion.

Thus, this study quickly abridges about every thing of the calculations being executed for the upgrade of a hand signal acknowledgment framework alongside the examination procedures required behind it. A novel picture determination improvement procedure in view of DT-CWT and EPS channel. The method breaks down the LR input picture utilizing DT-CWT. EPS (Bilateral) sifting is utilized to safeguard the edges and de-noising the picture and to additionally improve the execution of the proposed method as far as RMSE, PSNR

References

[1] A.R. Sarkar, G. Sanyal, and S. Majumder, "Hand gesture recognition systems: a survey", International Journal of Computer Applications, 71.15, 2013

[2] N. Neverova, Natalia, "A multi-scale approach to gesture detection and recognition", Proceedings of the IEEE International Conference on Computer Vision Workshops, 2013

[3] S. Yang, P. Premaratne, and P. Vial, "Hand gesture recognition: An overview", Broadband Network & Multimedia Technology (IC-BNMT), 5th IEEE International Conference, 2013

[4] T. Osunkoya and J-C. Chern, "Gesture-based human-computer-interaction using Kinect for Windows mouse control and Powerpoint presentation", Department of Mathematics and Computer Science, Chicago State University, Chicago, IL 60628, 2013

[5] J. Katkar, Jayshree, "Hand Gesture Recognition and Device Control", Hand, 2017
[6] J.G. Kyle and B. Woll,“Sign language: The study of deaf people and their language”, Cambridge University Press, 1988
[7] M.C. Thomas and A. P. M. S. Pradeepa,“A comprehensive review on vision based hand gesture recognition technology”, International Journal 2.1, 2014
[8] H. Zhouand H. Hu,“A survey-human movement tracking and stroke rehabilitation”, University of Essex, Colchester United Kingdom, 2004
[9] M. Turkand M.Kölsch,“Perceptual user interfaces”, Emerging Topics in Computer Vision, Prentice Hall, 2004
[10] K.K. Vyas, A. Pareek, and S. Tiwari, "Gesture Recognition and Control", International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169, Retrieved on 16th August, 2017
[11] D.H. Stefanov, Z. Bien, and W-C. Bang,“The smart house for older persons and persons with physical disabilities: structure, technology arrangements, and perspectives”, IEEE transactions on neural systems and rehabilitation engineering, vol.12.2, pp.228-250, 2004
[12] D.C. L-Martin, “Universal grammar and American sign language”, Universal Grammar and American Sign Language, pp.1-48, 1991
[13] M.L. McIntire, “The acquisition of American Sign Language hand configurations”, Sign Language Studies, 16.1, pp.247-266, 1977
[14] M. Alsheakhali, “Hand gesture recognition system”, Computer Engineering Department, The Islamic University of Gaza, Gaza Strip, Palestine, 2011
[15] S.D. Badgujar, “Hand Gesture Recognition System”, International Journal of Scientific and Research Publications, 4.2, 2014

International Journal of Engineering & Technology, 7 (1.9) (2018) 273-278

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Extraction of melanoma 3d features from tensor representation

Sushmita N S [1] *, Gururaj Murtugudde [2], T Y Satheesha [3]

[1] *Dept. of ISE, NCET Bengaluru, India*
[2] *Dept. of ISE, SVCE Bengaluru, India*
[3] *Dept. of ECE, NCET Bengaluru, India*
**Corresponding author E-mail: sushsushmita@gmail.com*

Abstract

Melanoma is one of the unsafe growth to be dealt with too to recognize in introductory stage. Here we take the skin sore by ROI and after that we take out highlights of it then it should be sectioned whether the specific picture is malignant or not. In the event that it is destructive at that point group the extricated includes and examine about kind of stages. This paper presents a non-obtrusive electronic dermoscopy framework that considers the evaluated profundity of skin sores for determination. For test assessments, the PH2 and ATLAS dermoscopy datasets is considered. A novel 3D remaking calculation from 2D dermoscopic pictures is proposed. Here we remove the 3D highlights from tensor portrayal. The discovery of 3D picture shape and RGB are to be done. In this paper, we have proposed this work for 3D profundity parameter, which will improve the grouping rate.

Keywords: *Preprocessing; Segmentation; Feature Extraction; Melanoma.*

1. Introduction

The number of melanoma dye due to continuous improvement of the results of a number of diseases. This disease is the main cause is due to ultraviolet light exposure. This is a change in color of the skin resulting in skin pigmentation. It can occur in any part of the body. This is a melanoma has spread it deep inside the skin layer needs to be discovered and cured before. This can be treated with chemotherapy. It is common, but when appropriate treatment is not very dangerous.
Melanoma is usually based on the ABCD rule, (asymmetry, irregular border, color changes, and prison), 7-point checklist of diagnosis [1] and texture. In this paper, we took a skin image for our segment. A portion of the segment are then extracted from the various features, and finally, we go for the divided. We classified [2] includes RGB images for the individual pixel data 3d Register. Here, a higher rate of discrimination is part of the feature extraction depth parameter is proposed.
The remaining paper is presented below. Section II and Section III, we carried out the background of work related to the proposed system. DISCUSSION specifies. Section IV is about the test results. Reported Finally, part V Conclusion home.

2. Background

ABCD features for melanoma skin cancer [3] Total Dermatoscopic Value (TDV) is used to calculate.
Asymmetry features of the lesions consist of information asymmetry and long index. Border irregular feature lesions Compact index, fractal dimension from the edge of the sudden transition coloring consist of information. Color homogeneous feature of homogeneous lesions with color photometry of the relationship between geometry and consist of information. The diameter of the cells from the lesions. In this research, the effect on melanoma, disease, doubt and benign skin lesions [4]. The paper "Use of Texture and color features for Dermoscopy for melanoma detection systems [5] the author of" In the first two isolated lesions of the skin and the second one to decide about the objectives of the two is to determine the color and texture. Presentation features.
The only drawback of the document classification accuracy rate, we can increase the rate of discrimination as a result of the depth of the feature as they entered our proposed system less.

3. Proposed system

Part of the skin lesions detection and feature extraction of important aim of our paper. This involves the following

a) Preprocessing:

The preprocessing stage in the first example, as shown in Figure 1 (a) of red, green and blue color components are converted into the image by adding gray. The gray image as shown in Figure 1 (b) by a Gaussian filter needs to be filtered. The image shown in Figure 1 (c) with the aid of 0s and 1s of the binary image needs to be converted by the Gaussian filter is filtered.

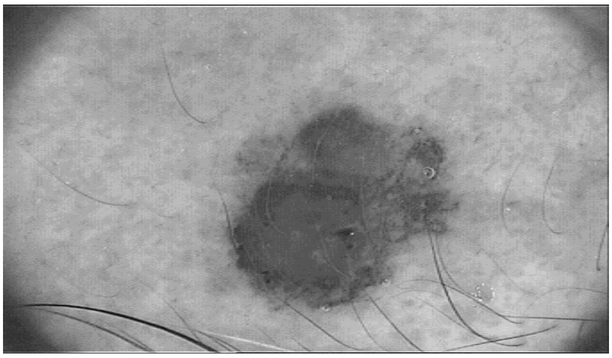


Fig. 1: (A): Grey Scale Image.

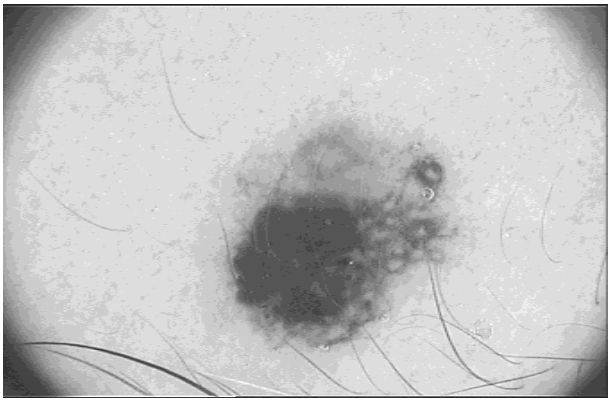


Fig. 1: (B) Gausian Filtered Image.

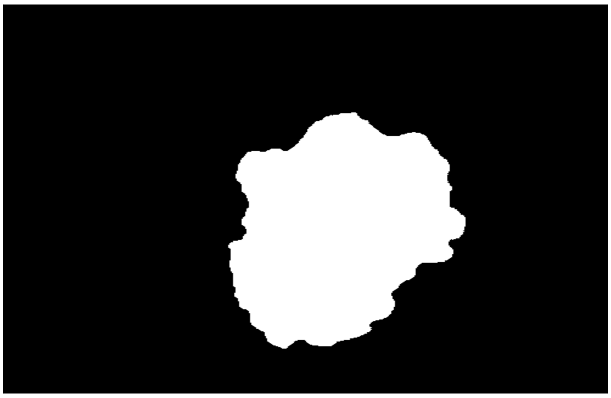


Fig. 1: (C) ROI Image.

Here, the threshold value for those with more color pixels are replaced with the binary 0 and less and the value of those with a brightness equivalent to the threshold holding 0 black color with black and white pixels are replaced by the value of an image resulting 1. white and holding a value of 1. Next, we will correct imag Acquiring e black color pixels are replaced by white pixels. The lesions using a circular pattern of periodic lines. Held by the collapse of the radial approach. In addition, in the black and white images are recycled in construction. For the next.
step is to engage in a gray scale image of the ascension for the segment based on that final mix named as the algorithm using an algorithm. [6] proposed a protected hash message verification code. A safe hash message confirmation code to stay away from authentication denial list checking is proposed for vehicular impromptu systems (VANETs). The gathering mark conspire is generally utilized as a part of VANETs for secure correspondence, the current frameworks in view of gathering mark plot gives confirmation delay in authentication repudiation list checking. With a specific end goal to conquer this postpone this paper utilizes a Hash message verification code (HMAC). It is utilized to stay away from tedious CRL checking and it likewise guarantees the honesty of messages. The Hash message verification code and computerized signature calculation are utilized to make it more secure [10]. In this plan the gathering private keys are appropriated by the roadside units (RSUs) and it additionally deals with the vehicles in a restricted way. At last, agreeable message validation is utilized among substances, in which every vehicle just needs to check few messages, therefore significantly easing the verification load.

b) Contour Algorithm for segmentation:

Most of the far edge of the artificial skin lesions from the boundary trace actively applied. Active is displayed under the image of a noisy image. These interfere with the texture of the skin, hair or changes. So the explosive effects, in order to remove the disruption, we named as the algorithm using a very powerful algorithm. They are the first to discover the skin in order to remove the stroke. These lesions are connected together around the edges of corners to make an average of at least getting out of them using the technique.

At the beginning of the associated color space using versions to match the strength of the stroke is found to have done. Then the corners are simple to connect using a standard technique called. Patterns and colors are selected. Therefore, from a part of the stroke, a confidence degree, consider using the EM method.Add roughly in order to be able to access the information contained in the model, let us consider. 1st stroke feature to extract a lot of features in the image, where the vault.
Let the black and white binary labels of the strokes be b.

c= argmaxclog (a, b, c)

To find the value of b, we replace the value log (a, b, c) by its expected value for b with Λ is the value estimated C from c

P (c; ^ =esti

{log (abc)|a,^ }

C jc b c

P(c; ^) = D+Σ PFe (cj,a)+IE(c)

With is a constant, is internal energy PF e (cj, a) is a potential function.

PFe (cj, a) = $-\Sigma\iota v^{l}$ PF^{l} (c)

With

PFl (cj) is the PF (potential function) associated with l th stroke and

vl is the probability of the

validity of the l th stroke

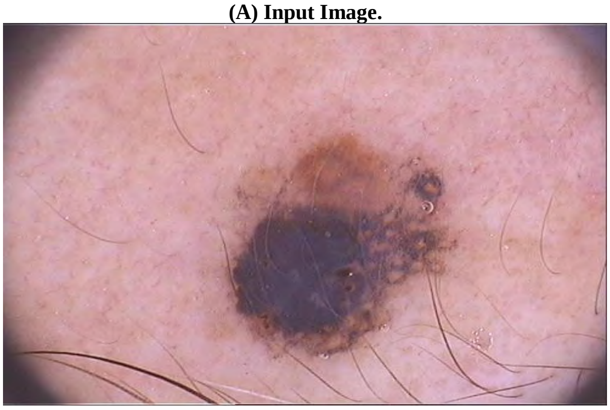
v^{l} =pro(b^{l} =1|a^{l} ,^)

The PF is so neat, he called the potential function Since keep on changing. There are two ways to estimate
these steps. The possibility of the validity of the first stroke is updated. Then secondly maximum value [7]. Thus the extra disturbances are removed from the outlier strokes.

P(c; c^) is increased to its

As a result, the output in Figure 2. 30 segment next process those features ABCD criteria are extracted in order to fulfill the feature extraction .A total. As a result, the output shown in Figure 2.

(A) Input Image.



(B) Contour Output

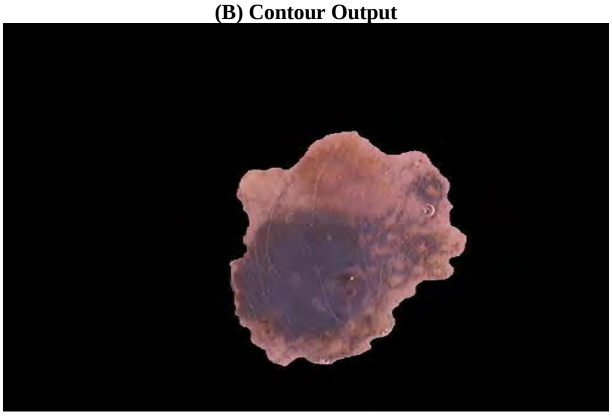


Fig. 2: Contour EM Algorithm Image for Segmentation.

Feature extraction:
Feature extraction based on texture:
Contrast: The features extracted based on contrast are:

$$\text{Contrast} = \sum_{x,y} (x - y)^2 \, pd(x, y)$$

Correlation:
The features extracted based on correlation is done using the formula:

$$correlatoion = \frac{\Sigma x, y\,(x - mx)(y - my)\,pd(x, y)}{dxdy}$$

Energy:
Features extracted based on energy is done by the formulae:

$$\text{Energy} = \sum_{x,y} pd\ 2\ (x, y)$$

Homogeneity:
The features extracted based on homogeneity is given by

21 03

homogeneity= pd(x,y) $\sum$
x, y 1+|x-y|

x, y

$$area \sum\sum (pxy\ m)$$

Features extracted based on colors:

_ 2

] 1

m x L m y N

Where m is the mean of the channels of colors red, green, blue L Is the lesion and N is the normal skin patch. These color features thus satisfy the colour aspects of the ABCD rule. The features of the color channels are2:RedRed,3:RedGreen, 4:RedBlue,5:GreenRed,6:GreenGreen, 7:GreenBlue ,8:BlueRed , 9:BlueGreen, 10: BlueBlue

30 12 30 12
21 03

Features extracted based on variability:
These are features extracted based on color and shape variability. The formulae used for it is given by:

dc L d c N

Where d c

Is the standard deviation of the channel C, L is the lesion and N is the normal skin patch. These features are serialized as follows: 11: Red, 12: Green, 13: Blue, 14: Zero.
Features based on peak and pit densities :
We take 8 to 9 nearest pixel values and the largest among them is considered as the peak density. The features based on the peak pixel are obtained using the following formulae:

nc, d=

No of peak in channel c+no of pits in the channel c area of the lesion
Where c is the channel belonging to {r, g, b, z} and d is the standard deviation of the Gaussian filter given by 0. 5, 1, 1. 5. Here also the features attained can be numbered as : 15:Red, 0.5, 16:Green, 0.5, 17:Blue, 0.5, 18:Z, 0.5, 19:Red, 1.0, 20:Green, 1.0, 21:Blue, 1.0, 22:Z, 1.0, 23:Red, 2.0, 24:Green, 2.0, 25:Blue ,2.0, 26:Z, 2.0.
Out of a total of 30 features each one of them after selecting them sequentially selected, and then sorted. As a result, segment and feature extraction of Classified training [7] are trained by then. Color distribution can be determined by their moments. The first time each color channel can be used in the middle of the second and third. The first image of the average color. In the center of the second image of the image that he is now the variance of the third central moment in each color channel asymmetry. A color space [8] the color of the characteristics to extract choice. We have an image with a pixel and prison AXB Let's consider adding the function.
Then the color moments is given by:

$$\text{Moment}_5 = (\text{eff}_{30} - 3\text{eff}_{12})$$

$$\text{Moments} = \Sigma\, \Sigma^{f}\, (i, j) i^{x} i^{y}$$

30 12 30 12

$$(\text{eff} - \text{eff})[(\text{eff} + \text{eff})^2 + (\text{eff} - \text{eff})^2]\ x, y\ _I\ j$$

$$m_{x,y} = \Sigma\, \Sigma^{f}\, (i,j)(i-ib)^{x}\,(j-jb)^{y}$$

The first and the second moment is mainly concentrated for 3D objects.

moment invariants
For an image function the 2D moments (x + y) are given as

Moments x, y =òòixiyf (i, j) didj

Where x, y is 0, 1, 2…

If the image is a piecewise function then it will be bounded and has real values and thus all moments exist. in order and the image function uniquely determines the sequencial moment
The geometry and central moments are given by :

Moment 5 = (eff 30 -3eff 12)

Moments= Σ Σf (i,j)ixi y

(eff -eff)[(eff +eff) 2 + (eff -eff)2]
mx,y= Σ Σf (i,j)(i-ib)x (j-jb)y

x, y i j

i j

Where ib and jb are the centre of gravities of the image.
The normalization of the moments is given by
m x, y

g Normalization xy =
m
00
2 2

Where

g □ (x □ y □ 2) / 2

Thus these moments' invariants are used for discriminating the shape of the objects. 7

moment =eff +eff
1 20 02

moment =(eff -eff)2 +4eff 2
2 20 02 11

moment =(eff -3eff)2 +(3eff -eff)2

3 30 12 21 03

moment
4 30 12 21 03

(eff □eff)2 □ (eff □eff)2

moment5 = □eff30 +3eff12 □ □eff30 +eff12 □

(eff +eff)2 +(eff -eff)2 □

□ 30 12 21 03 □

moment = □eff +eff □ □eff +eff □2 -(eff +eff)2]

6 20 02 30 12 21 03

+4eff 11 (eff 12)(eff 21 +eff 03)

The above invariants cannot separate the 3D features perfectly. Suppose to overcome this another 3D moment invariants can be used these moment invariants for the first 3 orders are:

mom =(mm -m2)/m 4
1 02 20 11 00

mom =(mm -6m m m m

2 30 03 30 21 12 03

3 3 2 2

+4m m +4m m -3m m)/m

30 12 21 03 21 12
00

mom =(m (m m -m 2)-m 2 –m

3 20 21 03 12 12 11

(m m -m m)+(m (m m -m 2))/m 2

30 03 21 12 02 30 12 21 00

4. Experimental result

If we want to represent the complex object we have to choose bag of feature (BoF) method. It is a collection of non-global model, e.g., lesion part of the image. In this method primarly, a set of keypoints, i.e., Pimg □ R 2, is defined in the image. In the case of lesion images, the essence are point of a regular
Network of size k x k defined in the image domain. Secondly, each essence is defined by a detail vector x □ Rn ,which shows local clue carry in the image patch of size K x K centered at the it h essence. The content of the lesion part is equal to the content of the network to ditch super-imposed. Found in the area outside of the PSL (Pigment skin lesions), more than 50% are forsaken. Each sample lesions in other words, our individual lesions found in the sample using the Histogram to set the global classification system so the right color features. As a result, the red, green and blue separately added features are shown in (3 - 6).

(A) Depth Estimation **(B) XYZ 3D Projection**

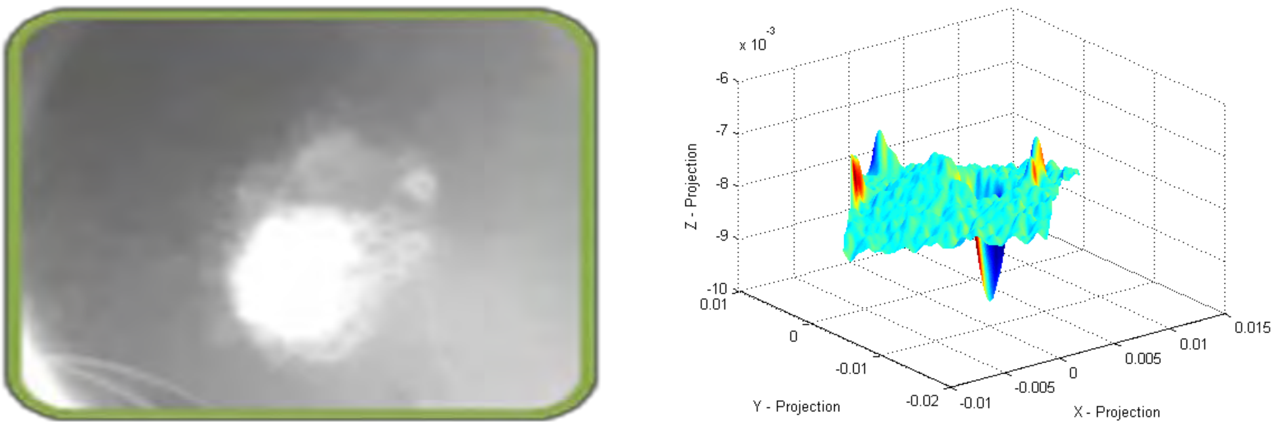


Fig. 3: (A): Depth Estimation. (B): Xyz 3d Projection.

(A) Red Localized Feature Extraction Image **(B) Red Localized Image**

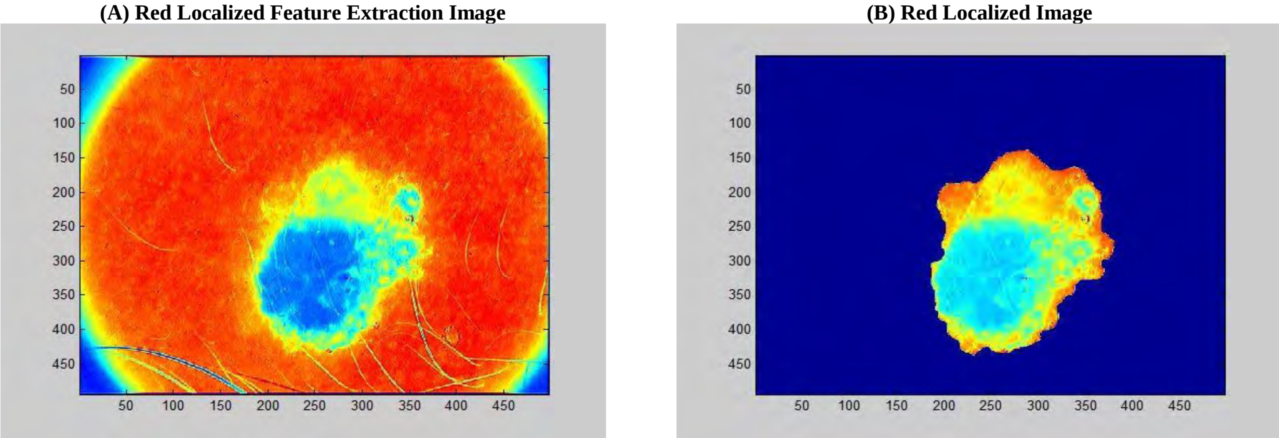


Fig. 4: (A): Red Localized Feature Extraction Image. (B): Red Localized Image

(A) Green Localized Feature Extraction Image **(B) Green Localized Image**

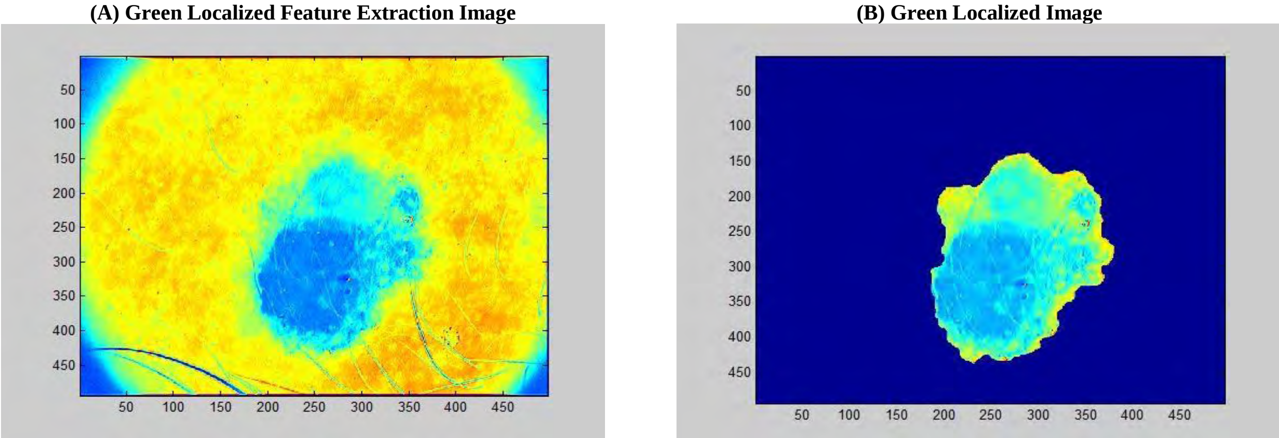


Fig. 5: (A): Green Localized Feature Extraction Image .(B): Green Localized Image

(A) Blue Localized Feature Extraction Image **(B) Blue Localized Image**

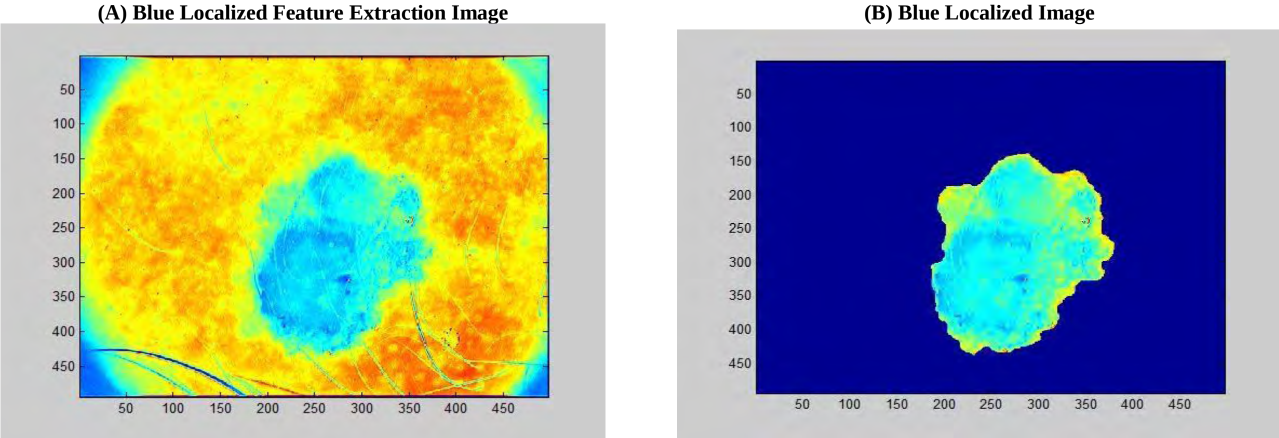


Fig. 6: (A): Blue Localized Feature Extraction Image. (B): Blue Localized Image.

5. Conclusion

A novel computer dermoscopy system proposed in this paper. melanoma skin lesions, highlighting the importance of the deep. To extract the depth of a novel 3D reconstruction algorithm is proposed. segment neat mix of success by using the algorithm. Therefore, the depth of features in addition allowed to increase rates in separate declaration. Since many features of melanoma can be used for practical assistance to accelerate melanoma are also used for detecting. Here for larger features are considered for melanoma detection. Figure 7(a-c) shows the 3D feature from Red, Green and Blue localized channels. Therefore, the rate of accuracy and discrimination both in a better performance.

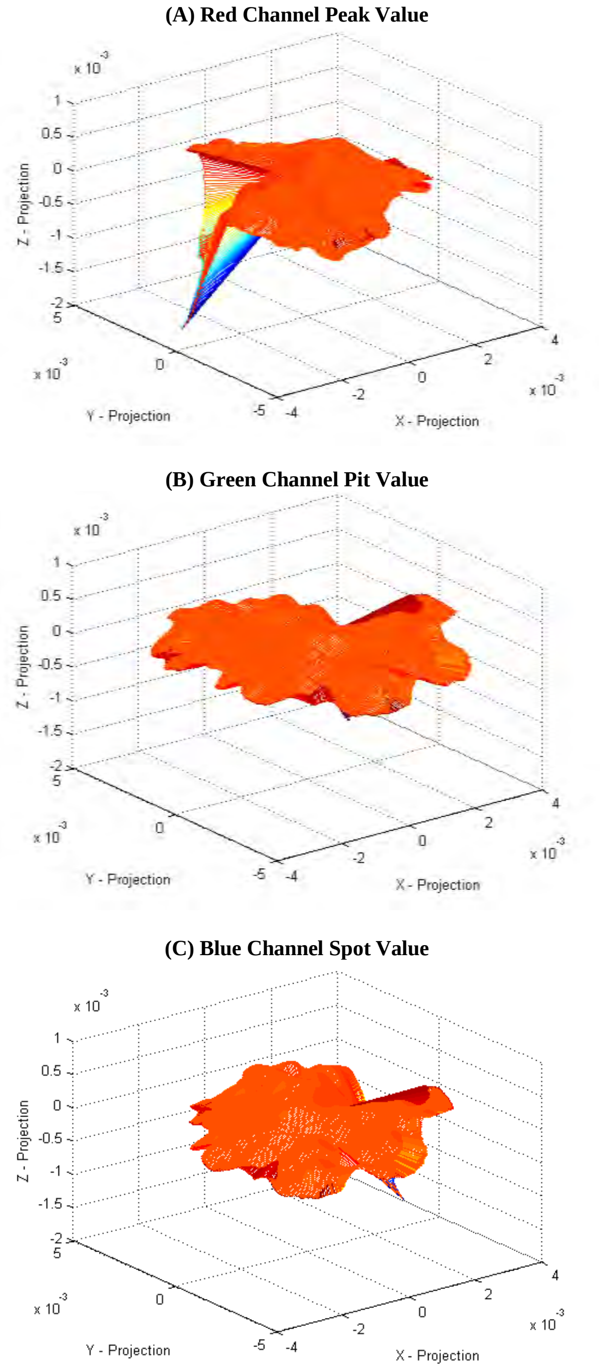


Fig. 7: (A): Red Channel Peak Value. (B): Green Channel Pit Value. (C): Blue Channel Spot Value.

References

[1] G. Betta, G. Di Leo, G. Fabbrocini, A. Paolillo, M. Scalvenzi, ," Automated Application of the "7-point checklist" Diagnosis Method for Skin Lesions: Estimation of Chromatic and Shape Parameters" IMTC 2005 – Instrumentation and Measurement Technology Conference Ottawa, Canada, 17-19 May 2005.

[2] Brian D'Alessandro, Atam P. Dhawan, "3D Volume econstruction of Skin Lesions for Melanin and Blood Volume Estimation and Lesion Severity Analysis" 2011 IEEE.

[3] K. Madhankumar, P. Kumar, "Characterization of Skin Lesions" Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering, March 21-23, 2012. https://doi.org/10.1109/ICPRIME.2012.6208362.

[4] Madhankumar, K., and P. Kumar. "Characterization of skin lesions", International Conference on Pattern Recognition Informatics and Medical Engineering (PRIME-2012), 2012. https://doi.org/10.1109/ICPRIME.2012.6208362.

[5] Scott E. Umbaugh, Randy H. Moss William V. Stoecker, Gregory A. Hance, "Automatic Color Segmentation Algorithms With Apphtion to Skin Tumor feature Identifiction" IEEE Engineering In Medicine And Biology, 1993.

[6] Christo Ananth, M.Danya Priyadharshini, "A Secure Hash Message Authentication Code to avoid Certificate Revocation list Checking in Vehicular Adhoc networks", International Journal of Applied Engineering Research (IJAER), Volume 10, Special Issue 2, 2015,(1250-1254).

[7] Ruela, Margarida, Catarina Barata, Jorge S. Marques, and Jorge Rozeira. "A system for the detection of melanomas in dermoscopy images using shape and symmetry features", Computer Methods in Biomechanics and Biomedical Engineering Imaging & Visualization, 2015.

[8] Sheng Xu. "3D object recognition using multiple features and neural network", 2008 IEEE Conference on Cybernetics and Intelligent Systems, 09/2008. https://doi.org/10.1109/ICCIS.2008.4670860.

[9] K.Sathesh Kumar, K.Shankar, M. Ilayaraja and M. Rajesh, "Sensitive Data Security In Cloud Computing Aid Of Different Encryption Techniques, Journal of Advanced Research in Dynamical and Control Systems, vol.18, no.23, 2017.

Implementation of High Speed and Low Power 5T-TSPC D Flip-flop and Its Application

Ashwini. H, Rohith. S, Sunitha. K. A

***Abstract*—True Single Phase Clock (TSPC) is a general dynamic flip-flop that operates at high speed and consumes low power. This paper describes the design and performance analysis of 5 transistor (5T) TSPC D Flip-flop in comparision with different TSPC D Flip-flops such as; (i) MS-Negative-edge triggered TSPC D Flip-flop, (ii) Positive-edge triggered TSPC D Flip-flop with (a) 13 transistors, (b) 11 transistors, (c) 9 transistors, (d) 8 transistors, (e) 6 transistor TSPC D Flip-flops with respect to transistor density, power and delay. Finally Charge Pump with PFD is designed using 5T TSPC D Flip-flop method and functionality of the circuit is verified through simulation. A Layout of 5T TSPC D Flip-flop and Charge Pump with PFD are designed. DRC, ERC, LVS are verified with gpdk 180nm technology. All the circuits used in this paper are designed and simulated using Cadence Virtuoso Platform, with gpdk 180nm CMOS process using 1.8V supply voltage.**

***Index Terms*—TSPC D Flip-flop; Phase Frequency Detector (PFD); Charge Pump; Power; Delay Element;**

I. INTRODUCTION

Achieving high performance in any Very Large Scale Integration (VLSI) systems is the most important part and the demand had increased with the growth of the semiconductor technology. As technology advances, a Systems-On-a-Chip (SOC) design contains more number of components that lead to a higher transistor density and increased power consumption [1-5]. As well as it needs a faster clock for its operation, that consumes more power. To improve the frequency of operation and integration of components of the Very Large Scale Integration (VLSI) Integrated Circuits (ICs) has been increased with the advance of Complementary Metal Oxide Semiconductor (CMOS) technology. The distribution of the global clock input and the inverse of it results in the clock skew problems in relation to each other, and also this consumes more power. Thus to overcome this problem Single Phase Clocking is very advantageous [6-10]. The True Single Phase Clock (TSPC) is a broad dynamic flip-flop that operates at high speed and with low power consumption. It uses one phase clock for synchronization. Many researches has shown TSPC method has small area, no clock skew problem and even higher clock frequencies can be achieved thereby improving the performance of the digital systems. It is used in various applications like digital VLSI clocking system, microprocessors, buffers, wireless communication systems etc [7].

In many applications such as wireless communication systems, digital circuits, and receivers, Phase Lock Loop (PLL) is one of the most important blocks. The intention of PLL is to generate a signal, such that the phase is equal to the phase of the reference signal. Some of the applications of PLL are frequency synthesis, clock generation, carrier (clock) recovery, and skew reduction, jitter and noise reduction. Phase Frequency Detector (PFD) / Charge-Pump and Voltage Control Oscillator (VCO) are the important components present in PLL. PFD is a circuit that can detect both phase and frequency errors, based on this it will generate output signals “UP” and “DOWN”. A PFD is usually built using a state machine with memory element such as D Flip-flop. In order to make the PFD circuit simpler and reduce the dead zone there are several methods. TSPC D-Flip-Flop is used more as a D-Flip-flop in designing PFD, which provides low area and high speed of operation [5].The main principle of Charge Pump is to change the logic states of the phase frequency detector into analog signals that is appropriate to control the VCO.

In [1], the design of a 3-V 300MHz Low-Power 8-b×8-b Pulse triggered TSPC Flip-flops, with the concept of True Single Phase Clocking, a new pulse triggered TSPC Flip-flop (PTTFF) that comprised of five transistors is implemented. Using this PTTFF and with the 14-transistor pseudo- NMOS full adder an 8-b×8-b pipelined multiplier is designed using 0.6μm CMOS process. In this, power analysis is performed for low power pipelined multiplier. Using TSPC D Flip-flop method the multiplier is designed and analysis is performed on power consumption with different supply voltage. The design and implementation of A 5.8 GHz Wideband TSPC Divide-by-16/17 Dual Modulus Prescaler discusses about improving the speed by incorporating a new pseudo divide-by-2/3 Prescaler and reducing delay to increase speed of operation. By using the concept of TSPC D Flip-flops the power and delay analysis is carried out using 0.18μm CMOS technology with 1.6 supply voltage and frequency range of 2MHz in [2]. The structures of different dividers for high frequency wireless

Ashwini.H, Department of Electronics & Communication, Nagarjuna College of Engineering and Technology, Bangalore, Karnataka, India-562110 (ashwinih32693@gmail.com)

Rohith.S, Department of Electronics & Communication, Nagarjuna College of Engineering and Technology, Bangalore, Karnataka, India-562110 (rohithvjp2006@gmail.com).

Sunitha.K.A, Department of Electronics & Communication, Nagarjuna College of Engineering and Technology, Bangalore, Karnataka, India-562110 (sunithaka1990@gmail.com)

978-1-5090-0396-9/16/$31.00 ©2016 IEEE

systems are discussed in [3]. A CMOS divider family for high frequency wireless localization systems. In this, using TSPC and CMOS logic the analysis on low power consumption and high speed of operation for various dividers in comparision with other dividers is performed using 130nm CMOS technology with 10GHz to 22GHz maximum operating frequency and also it is shown that TSPC logic based divider is better in performance in comparision with CML logic. It is discussed in [6], the design of A new N-Fold Flip-flop exploiting the clock gating technique for both output enabling and power saving, using TSPC logic an octal flip-flop was built and compared to the main octal flip-flop with respect to performance in terms of power, area and digital noise. Implementation and simulations is performed using STMicroelectronics 65nm technology process with 1.1V power supply.

In this paper the performance comparision among different TSPC D Flip-flops with respect to transistor density, power and delay is done for the following TSPC D Flip-flops; MS-Negative-edge triggered D Flip-flop, Positive-edge triggered 13 transistor, 11 transistor, 9 transistor, 8 transistor, 6 transistor, and 5 transistor TSPC D Flip-flops. Out of these TSPC D Flip-flops one flip-flop was found better in performance i.e. 5T TSPC D Flip-flop [11-13]. This 5T TSPC D Flip-flop method is used for the implementation of Charge Pump with PFD. And also 5T TSPC D Ff and Charge Pump layout is designed and verified with the schematic view. All the circuits used in this paper are designed and simulated and the respective output of each circuit is taken using Cadence Virtuoso platform using gpdk 180nm CMOS process. The performance comparision is made for all the circuits in terms of power dissipated and delay.

Rest of the paper is discussed as follows, in Section II, the design of 5T TSPC D Flip-flop is discussed. In Section III, Application of TSPC Flip-flop in PFD and Charge Pump with PFD is discussed. In Section IV, simulation results and analysis are shown. Finally in Section V, the conclusion is given.

II. DESIGN OF 5T TSPC D FLIP-FLOP

A. 5Transistor (5T) TSPC (True Single Phase Clocking) D Flip-flop

Fig.1 and Table I shows Circuit and Truth table of positive edge triggered 5 transistors TSPC D Flip-flop respectively. 5T TSPC D Flip-flop is constructed using 3 NMOS and 2 PMOS transistors. This flip-flop uses single clock phase signal for synchronization. It consumes less area since it uses only 5 transistors, which signify low transistor count and it also consumes less power. Thus the performance of this design is improved. The working principle of this flip-flop is as follows: When clock and input signal is HIGH i.e. clk=1, D input=1 then the NMOS transistors m2 and m3 is ON while PMOS transistor m1 is OFF this in turn turns ON PMOS transistor m4 and turns OFF NMOS transistor m5 thus pulling the output to HIGH i.e. Q=1. Similarly when clk=1 and D input=0 then m1 and m2 transistors are ON while m3 is OFF this turns ON m5 producing the output LOW. That is during ON period of clock signal whatever the value of the input is, the output follows the input. On the other hand when clk=0, the output follows the previous value of the output.

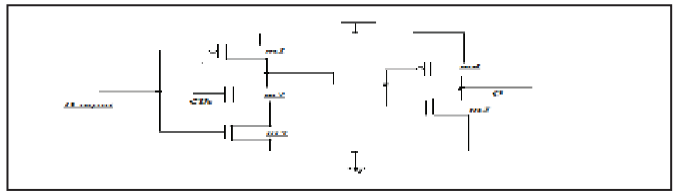


Fig. 1. Circuit of 5T TSPC D Flip-flop [6]

TABLE I

TRUTH TABLE OF 5T TSPC D FLIP-FLOP

Clk	D	M1	M2	M3	M4	M5	Q
1	0	ON	ON	OFF	OFF	ON	0
1	1	OFF	ON	ON	ON	OFF	1
0	0	ON	OFF	OFF	OFF	ON	0
0	1	OFF	OFF	ON	OFF	ON	0

III. PROPOSED WORK

Using (Master-Slave) MS-5T TSPC D Flip-flop method the Charge Pump with PFD is implemented. The Circuit of MS-5T TSPC D Flip-flop with reset input is shown in Fig. 2. It uses active high reset i.e. a high on reset pin will reset the flip-flop to the initial value. The operation of MS-5T TSPC D Flip-flop is illustrated as shown in Truth Table II.

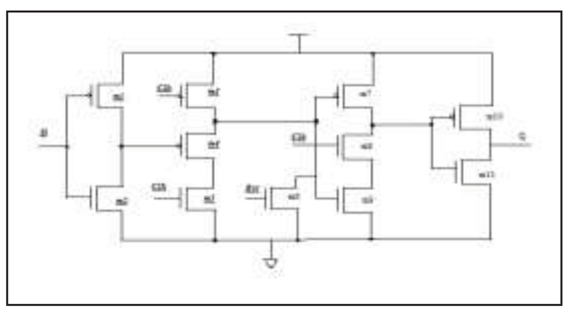


Fig. 2. Circuit of MS-5T TSPC D Flip-flop with reset input

The Circuit of Charge Pump shown in Fig. 3, it consists of two current sources idc and idc1 which are identical, UP and DOWN are the output signals of PFD which is given as an input to charge pump. Based on the three conditions of PFD i.e. based on error signal generated by the PFD the charge pump is controlled. That is whenever UP or DOWN signal of PFD is raised to HIGH it provides (+/- Vp) charge pump voltage or charge pump current (+/-Ip). When both UP and

DOWN output signals from PFD are off i.e. LOW output remains constant.

TABLE II

TRUTH TABLE OF MS-5T TSPC D FLIP-FLOP

Clk	D	m1	m2	m3	m4	m5	m7	m8	m9	m10	m11	Q
1	0	on	off	off	on	on	on	on	off	off	on	0
0	0	on	off	on	on	off	off	off	on	off	on	0
1	1	off	on	off	off	on	off	on	on	on	off	1
0	1	off	on	on	off	off	off	off	on	on	off	1

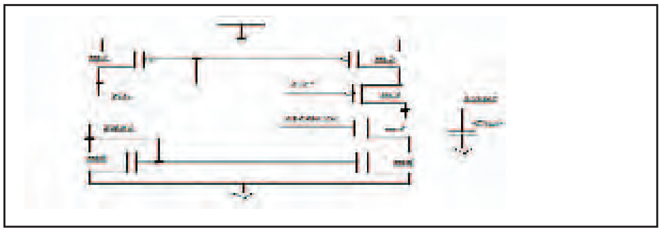


Fig. 3. Circuit of Charge Pump[11]

The block diagram of Charge Pump with PFD is shown in Fig 4. Based on the two input clock signals i.e. ref and div the output signals UP and DOWN of the PFD are generated which are then fed as an input to the Charge Pump. Based on the error signal produced by PFD the Charge Pump is controlled.

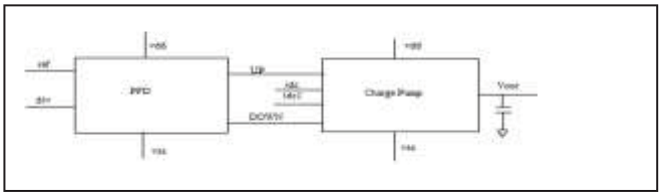


Fig. 4. Block Diagram of Charge Pump with PFD

IV. SIMULATION RESULTS AND ANALYSIS

The different TSPC D Flip-flops such as MS-Negative edge triggered D Flip-flop, Positive edge triggered 13 transistor, 11 transistor, 9 transistor, 8 transistor, 6 transistor, and 5 transistor TSPC D Flip-flops are designed, simulated and the respective output of each circuit is taken using Cadence Virtuoso 180nm CMOS technology with 1.8V supply voltage.

Fig. 5. Schematic of Positive-edge triggered 5T TSPC D Flip-flop

Fig.5, 6, and 7 show schematic, waveform and layout of positive-edge triggered 5 transistors TSPC D Flip-flop. Table III shows comparision table interms of power, delay, and no of transistors used for different TSPC D Flip-flops

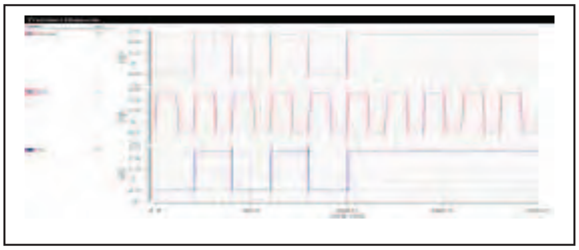


Fig. 6. Waveform of Positive-edge triggered 5T TSPC D Flip-flop

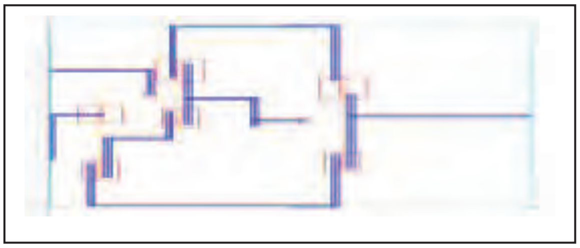


Fig. 7. Layout of Positive-edge triggered 5T TSPC D Flip-flop

TABLE III

COMPARISION TABLE OF DIFFERENTTSPC D FLIP-FLOP

Supply voltage=1.8v Width=2µm Channel length=180nm	**Power Dissipated**			**Delay (ns)**	**No of transistors used**	**Current (nA)**
	Total power (µw)	**Static power (µw)**	**Dynamic power (µw)**			
MS-Negative edge triggered D Flip-Flop	7.321	0.1303	7.19074	51.07	11-basic	72.3675
Positive-edge triggered 13T TSPC D Flip-flop	6.472	15182.0	15175.5	40.10	13	8434458.
Positive-edge triggered 11T TSPC D Flip-flop	5.064	21575.2	21570.08	0.098	11	11986195.0
Positive-edge triggered 9T TSPC D Flip-flop	3.165	32144.2	32141.1	19.91	9	17857944.0
Positive-edge triggered 8T TSPC D Flip-flop	39.21	31914.6	31875.41	22.36	8	17730345.0
Positive-edge triggered 6T DE-TSPC D Flip-flop	2.604	0.0117	2.5922	39.55	6	6.53600
Positive-edge triggered 5T TSPC D FF	1.116	0.0012	1.1148	39.54	5	0.66582

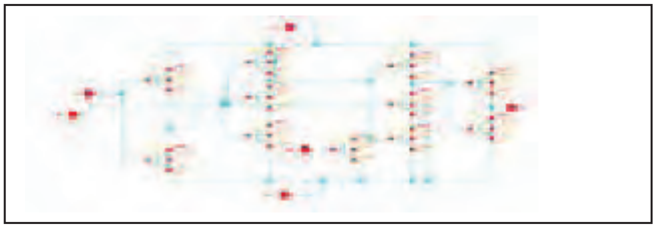


Fig. 8. Schematic of MS- 5T TSPC D Flip-flop with reset input

The schematic of positive-edge triggered MS-5T TSPC D FF is shown in Fig. 8.

The block diagram of PFD using AND-Gate and MS-5T TSPC D Flip-flop is shown in Fig 9. The circuit consists of two Positive-edge triggered MS-5T TSPC D Flip-flops with their D inputs tied to logic 1 and AND-Gate in reset Path. The Ref and Div inputs serve as clocks of the flip-flops. The UP and DOWN are the outputs of PFD which is given as input to the AND-Gate, when both inputs to AND-Gate is HIGH, its output is raised to HIGH which pre-sets the PFD to the Initial value.

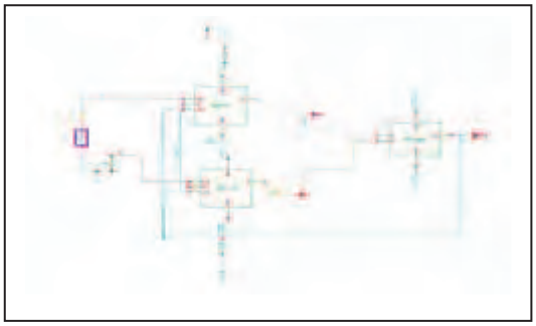


Fig. 9. Schematic of PFD using AND-Gate and MS-5T TSPC D Flip-flop

The simulated output of PFD using AND-Gate and MS-5T TSPC D Flip-flop (ref leads div) is as shown in Fig 10. From figure we can see that when ref signal f1 leads div signal f2 then the output UP signal is asserted on the rising edge of the reference signal, while the small pulse is generated at the DOWN output, whose delay is equal to the delay through reset path and logic gates.

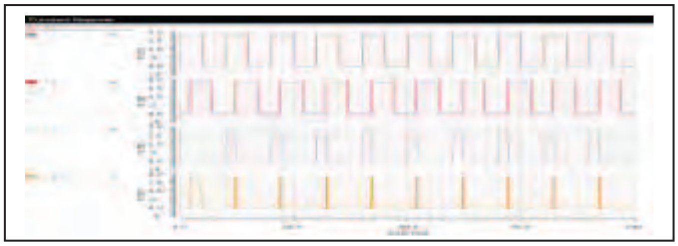


Fig. 10. Waveform of PFD using AND-Gate and MS-5T TSPC D Flip-flop (ref leads div)

In the second case, when the ref signal f1 lags the feedback signal div f2 then the DOWN output signal is produced indicating the phase difference between the two clock signals as shown in Fig. 11.

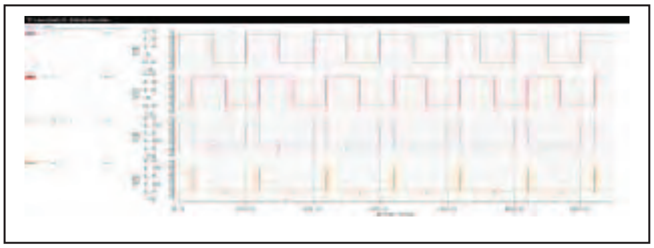


Fig. 11. Waveform of PFD using AND-Gate and MS-5T TSPC D Flip-flop (ref lags div)

In the third case as shown in Fig. 12, if the reference signal is in phase with the feedback signal the loop is in the locked state producing small pulses at the output of PFD.

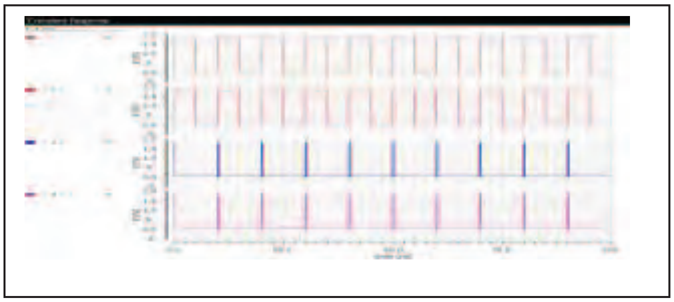


Fig. 12. Waveform of PFD using AND-Gate and MS-5T TSPC D Flip-flop (ref in-phase with div)

The schematic of Charge Pump with PFD is shown in Fig. 13. Charge pump is the subsequent block to the PFD. The output signals UP and DOWN produced by PFD are given as an input to the Charge Pump i.e. the UP and DOWN signals are combined by charge pump into the single output to drive the LPF.

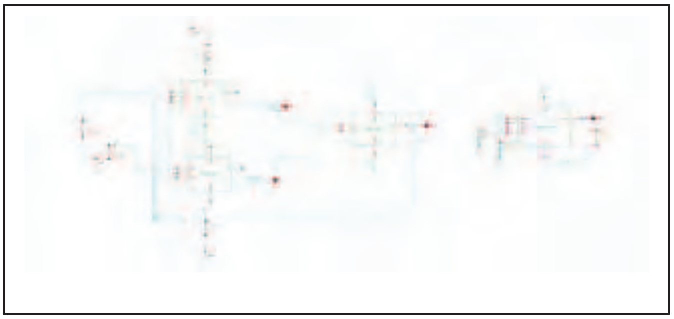


Fig. 13. Schematic of Charge Pump with PFD

The Simulated output of Charge Pump with PFD is shown in Fig. 14. Analog Design Environment in cadence is used for simulations.

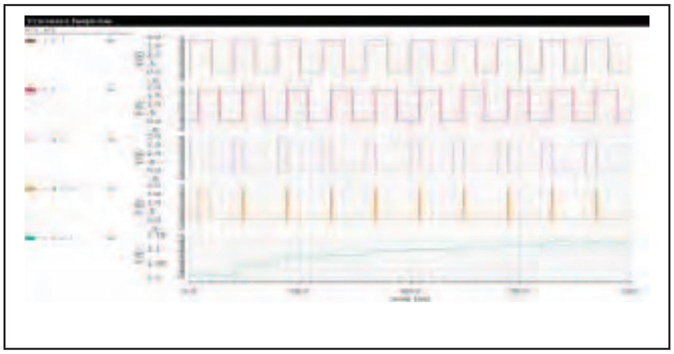


Fig. 14. Waveform of Charge Pump with PFD

The Layout of Charge Pump with PFD is as shown in Fig. 15. The DRC, ERC, and LVS are verified for the designed layout.

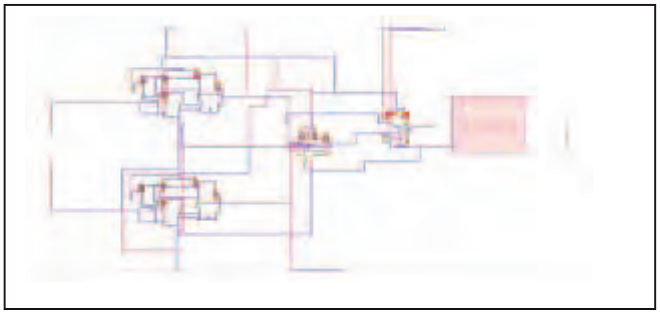


Fig. 15. Layout of Charge Pump with PFD

Table IV shows dynamic, static, total power dissipation of Charge pump with PFD. From table it is clear that it consumes less power compared to conventional Charge Pump with PFD.

TABLE IV

DYNAMIC, STATIC, TOTAL POWER AND CURRENT OF CHARGE PUMP WITH PFD

Supply voltage=1.8v Width=2µm Channel length=180nm	**Power Dissipated**			**Current (µA)**
	Total power (µw)	**Static power (µw)**	**Dynamic power (w)**	
Charge pump with PFD	99.5	10195.3	919.75	566.25

The Schematic of conventional TSPC D Flip-flop is shown in Fig 16. Using this conventional TSPC D Flip-flop in the implementation of Charge Pump with PFD the total power consumed was 99.52µw which is more than the proposed method.

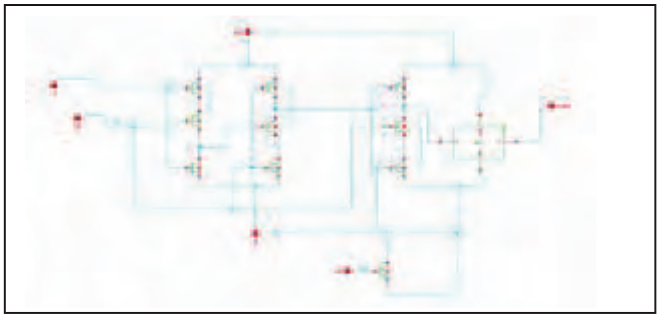


Fig. 16. Schematic of Positive-edge triggered Conventional TSPC D Flip-flop

V. CONCLUSION

This paper discusses about implementation of different TSPC D Flip-flops and its comparison in terms of transistors, power and delay, from comparison result and analysis it was seen that 5T TSPC D Flip-flop is better in performance with less power consumption compared to other flip-flops. Thus 5T TSPC Flip-flop method is used for implementation of the Charge pump with PFD and the simulated results are verified. And also layout for 5T TSPC D FF and Charge Pump with PFD is designed and extraction results are obtained. All these circuits are designed and simulated using cadence virtuoso 180nm CMOS process with 1.8V supply Voltage. The total power consumed by proposed method is 99.5µw and it less compared to the conventional method with total power consumption of 99.52µw.

VI. REFERENCES

[1] Jinn-Shyan Wang, Po-Hui Yang, and Duo Sheng, " Design of a 3-V 300 MHz Low Power 8-b×8-b Pipelined Multiplier using Pulse Triggered TSPC Flip-flops", *IEEE Journal of Solid Circuits*, Vol.35, No.4, April 2000.

[2] Wenrui Zhu, Haigang Yang, Tongqiang Gao, Fei Lui, Dandan Zhang, and Hongfeng Zhang, "A 5.8 GHz Wideband TSPC Divide-by-16/17 Dual Modulus Prescaler", *IEEE transactions on very large scale integration (VLSI) systems*, January 2015.

[3] Melanie Jung, Georg Fischer, Robert Weigel and Thomas Ussmueller, "A CMOS divider family for high frequency wireless localization systems", Institute for Electronics Engineering, *Friedrich Alexander University of Erlangen-Nuremberg, Cauerstrasse, Erlangen, Germany © IEEE 2012*.

[4] Manthena Vamshi Krishna, Manh Anh Do, Kiat Seng Yeo, Chirn Chye Boon, Wei Meng Lim, "Design and Analysis of Ultra Low Power True Single Phase Clock CMOS 2/3 Prescaler", *IEEE transactions on circuits and systems – Regular papers*, VOl. 57, No.1, January 2 2010.

[5] Behzad Razavi, "RF MicroElectronics", University of California, Los Angeles, Published by Dorling Kindersley (India) Pvt. Ltd., licensees of Pearson Education in south Asia, Copyright © 1998 by Prentice Hall PTR Prentice-Hall, Inc.

[6] Mounir Zid, Carlo Pistritto, Rached Tourki and Alberto Scandurra, "A new N-fold flip-flop with output enable", Electronics and Micro-Electronics Laboratory, Faculty of Sciences of Monastir, Tunisia, International Journal of Embedded Systems and Applications(IJESA) Vol.3, No.2, June 2013.

[7] Priyanka Sharma and Rajesh Mehra, "True Single Phase Clocking Based Flip-flop Design using Different Foundries", Department of E.C.E, International Journal of Advances in Engineering and Technology, May 2014.

[8] Sunita Arvind Rathod, Siva Yellampalli, "Design of Op-amp, Comparator and D flip-flop for Fifth Order Continuous Time Sigma-delta Modulator", Published by: Blue Eyes Intelligence Engineering and Sciences Publication Pvt. Ltd, International Journal of Innovative Technology and Exploring Engineering, Volume-4 Issue-2, July 2014.

[9] Priyanka Sharma and Rajesh Mehra, "True Single Phase Clocking Flip-flop Design using Multi Threshold CMOS Technique", Department of E.C.E, International Journal of Computers Applications Volume 96-No.11, June 2014.

[10] Melanie Jung, Joerg Fuhrmann, Alban Ferizi, Georg Fischer, Robert Weigel, and Thomas Ussmueller, "Design of a 12 GHz Low-power Extended True Single Phase Clock (E-TSPC) Prescaler in 0.13µm CMOS Technology", Institute for electronics Engineering, Friedrich-Alexander University Erlangen-Nuernberg Cauerstrasse, Erlangen, Germany, Proceedings of the Asia-Pacific Microwave conference 2011.

[11] Kashyap K. Patel1, Nilesh D. patel2, "Phase Frequency Detector and Charge Pump for DPLL using 0.18µm CMOS Technology", L.C Institute of Technology, Bhandu, International Journal of Emerging Technology and Advanced Engineering 2008.

[12] H. Shin, "A 1-V TSPC dual modulus prescaler with speed scalability using forward body biasing in 0.18 µm CMOS," *IEICE Trans. Electron.*, vol. E95-C, no. 6, pp. 1121–1124, Jan. 2012.

[13] V. K. Manthena, M. V. Do, C. C. Boon, and K. S. Yeo, "A low-power single-phase clock multiband flexible divider," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 20, no. 20, pp. 376–380, Feb. 2012.

Design of 8 Bit Vedic Multiplier Using Urdhva Tiryagbhyam Sutra With Modified Carry Save Adder

CHANDRASHEKARA M N
Department of Electronics and Communication Engineering
Nagarjuna College of Engineering and Technology
Bengaluru, Karnataka, India
chandupesit.te@gmail.com

ROHITH S
Department of Electronics and Communication Engineering
Nagarjuna College of Engineering and Technology
Bengaluru, Karnataka, India
rohithvjp2006@gmail.com

***Abstract*—This paper mainly describes the design of 8-bit Vedic multiplier and its performance comparison with existing multiplier such as i) Booth multiplier ii) Array multiplier iii) Wallace tree multiplier. Vedic calculations are the olden scheme of mathematics, which has a procedure of mathematical calculations to compute the multiplication of two 8-bit number. In this work Urdhva Tiryagbhyam (vertical and crosswise) Vedic sutra is used for multiplier design which provides better performance and consumes lesser time for computation. The Urdhva Tiryagbhyam is the finest sutra and universal one among additional sutras and which represents the different multiplication process compared to normal multiplication. In this work, Modified Carry Save Adder (MCSA) is used to compute the sum of partially generated products. It reduces the computational delay towards the addition of unfinished products. The proposed design uses the Verilog HDL to develop the algorithm. The XILINX 14.7 software tool is used to simulate and synthesize the code. The proposed design is also verified on Spartan-6 Field Programmable Gate Array (FPAGA). Finally, the proposed 8-bit multiplier design is compared with 8-bit Booth multiplier, Array multiplier and Wallace tree multiplier in terms of Area, Memory and Delay. The result shows proposed 8-bit Vedic multiplier is efficient and consumes 14.219ns time for the multiplication process which is better compared to 8-bit, Booth multiplier, Array multiplier and Wallace tree multiplier.**

Keywords—* *Urdhva Tiryagbhyam, DCT, FFT, Booth multiplier, Array multiplier, Wallace tree multiplier, MCSA.

I. Introduction

Rapid increase of digital devices, processing of digital data using Digital Signal Processing (DSP) unit is most common. The digital data may be in the form of text, audio, image or video. To process the digital data, multiplication is significant and important arithmetic procedure. The variety of applications related to Discrete Wavelet Transforms (DWT), Discrete Cosine Transforms (DCT), Fast Fourier Transforms (FFT), Convolution and as well as Arithmetic and Logical Unit (ALU), the multiplier is used as a basic block. As the multiplier decides the performance of the system and it provides the serious delay path in signal processing blocks. The multiplier circuit consumes relatively more power in computation. Hence reducing the time delay and power consumption of the multiplication unit is more important for Digital signal processing applications.

Many new multiplier designs are introduced in recent years, namely Array multiplier, Booth's multiplier, Braun multipliers and Wallace tree multiplier etc. In these multiplier algorithms the unfinished products are generated by addition and several other comparisons to obtain the final product. However, the computation speed is low. To improve the computation speed, in this work Vedic mathematics-based multiplier is proposed.

Sri Bharti Krishna Tirthaji (1884-1960) anticipated Vedic arithmetic later in his eight years of investigation on Atharva Vedas [12]. The Vedic mathematics is an attention-grabbing field and presenting several successful algorithms that can be useful to a range of engineering branches. The multiplication processes with Urdhva Tiryagbhyam mathematics algorithm would result in the dropping of computational delay.

In this work, the multiplier utilizes the Urdhva-Tiryakbhyam sutra [12] for multiplication of binary numbers. The major consideration of the design is to improve the speed of multiplier to produce the unfinished products. The unfinished products generated by using Vedic multiplier are subsequently added by MCSA to produce the final product. The estimated multiplier leads to an improved speed, and reduces area occupied by multiplier design.

The rest of the paper is organized as follows. Literature survey is given in Section II. The traditional multiplier designs are discussed in Section III. The Section IV constitutes the detail design of proposed Vedic Urdhva Tiryagbhyam multiplier. The Section V discusses about result analysis of the proposed scheme. The conclusion is given in Section VI.

II. Literature Survey

Many researchers have developed algorithms for multiplication using Vedic multiplier [1]-[6][7] [12]. In [1] realization of high-speed Vedic multiplier by means of Vedic mathematics sutra was discussed. They used Urdhva Tiryagbhyam sutra for the design of the 8-bit multiplier and Carry Save Adder (CSA) used to add the unfinished products to obtain the resultant product. The result shows multiplier utilizes 15.41ns time to multiply the two 8-bit numbers. In [2] Design of Vedic multiplier using higher order compressor to increase the speed and area is discussed. They used compressors to reduce the area and delay. In [3] Design of area and time delay efficient multiplier to obtain better performance of the multiplier is given. The result shows scheme consumes 44.358 ns to produce the final product of the given two 8-bit input data. In [4] Area competent modified Vedic multiplier is discussed. They used Vedic sutras to design the multiplier.

978-1-7281-0630-4/19/$31.00 ©2019 IEEE

The Ripple carry adder used to add the partial products. The result shows it requires 23.18 ns to compute the final product. In [5] design and implementation of 16 x 16 multiplier using Vedic mathematics is discussed. The result shows, it requires 25.159 ns for 8×8 Vedic multiplier. In [6] design and implementation of Vedic multiplier, they proposed new multiplier architecture using the Read Only Memory (ROM) and constant co-efficient multiplier approach, the result shows, it requires 23.87 ns time to compute the multiplication of two 8-bit number.

In this work, Urdhva Tiryagbhyam Vedic sutra used for multiplication of two 8-bit binary numbers. The proposed design intended to improve the speed of 8-bit multiplier to produce the partial products. The modified carry save adder is used to produce the final product by adding unfinished product. Also, the performance of the proposed design is compared with existing multiplier such as i) Booth multiplier ii) Array multiplier iii) Wallace tree multiplier.

III. DESIGN OF DIFFERENT MULTIPLIERS

A. Booth Multiplier [8]

In 1950, Andrew Donald Booth invented the booth algorithm [8] for his cryptography research in London. This multiplication technique involves encoding of multiplier data bits and creation of unfinished products. Booth multiplier architecture is given in Fig.1.

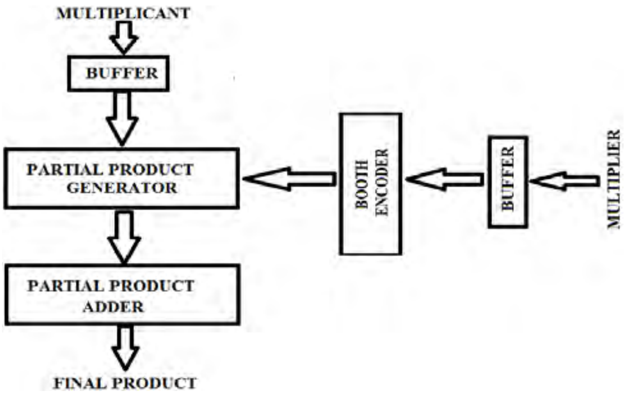


Fig.1. Booth multiplier.

This multiplier converts the negative numbers by 2's compliment method and the encoded binary number is multiplied with the multiplicand and produces the partial products. The unfinished products are added by using the adder. The booth multiplier requires more delay to produce the partial products. This multiplier mainly used to multiply the signed numbers.

B. Array Multiplier [9]

The array multiplier [9] uses the add and shift algorithm. The structure of the 8-bit array multiplier is shown in Fig.2.

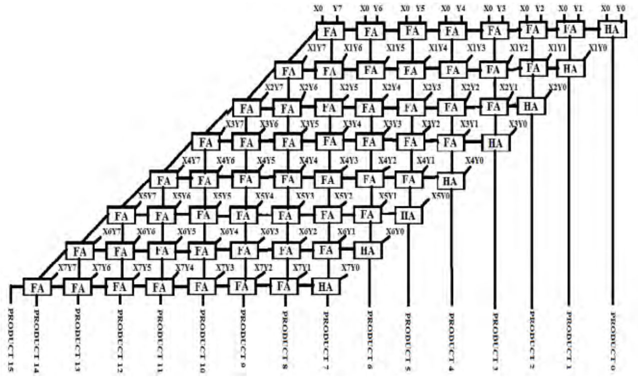


Fig.2. Traditional 8- bit Array multiplier.

The array multiplier can generate the incomplete products by using AND gates, full adder and half adder. For N×N array multiplier, N×N AND gates produce the incomplete products, these can be added by (N-2) × N full adders. The array multiplier performance is degraded by its Area and Delay.

C. Wallace Tree Multiplier

The traditional Wallace tree multiplier [10] structure illustrated in Fig.3. This multiplier uses column compression technique and hence computation speed is more compared to the array multiplier. The total delay is logarithmic of length of binary data. The performance of this multiplier is better with respect to the array multiplier because time delay is linearly varying in array multiplier, where as it varies logarithmically in Wallace tree multiplier. The N-bit multiplier requires N^2 AND gates to generate the partial products. The performance is degraded by its Area.

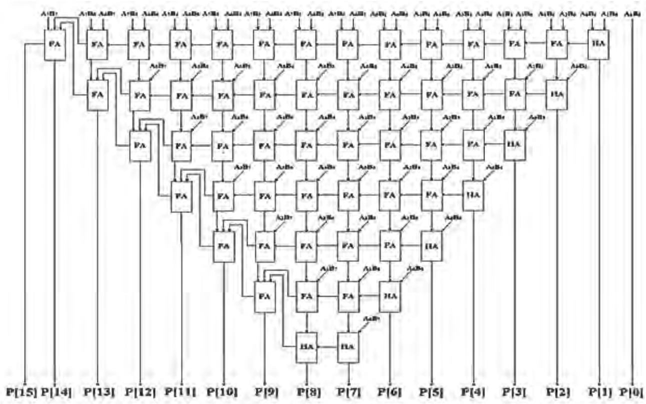


Fig.3. Wallace tree multiplier.

IV. VEDIC MULTIPLIER

A. Vedic Maths

The Vedic math is element of four Vedas. The sound "Vedic" is arrived from the word "Veda" which gives the meaning storehouse of all information [12]. It describes simplified steps in arithmetic, trigonometry, factorization, geometry, quadratic equation and calculus. As earlier discussed, Sri Bharati Teerthaji Maharaja who discovered the Vedic mathematics from Vedas. The swamiji developed sixteen sutras for Vedic maths. Vedic multiplication is not just an arithmetical wonder, and it is a logical too. Because of this reason, this multiplier has such an extent of importance it cannot be able to disapprove. Due to this, the Vedic maths became an important and interesting topic in the mathematics. Vedic maths is particularly simple and very powerful, and which deals with the basic and complex arithmetic operations. The Vedic algorithm depends on the 16 formulae out of which, Urdva Tiryagbhyam Sutra and Nikhilam Navatascaramam Sutra are most widely used for multiplication. This paper mainly concentrates on the Urdva Tiryagbhyam Sutra for the proposed multiplier design. The Table 1 shows most widely used Sutras and their meanings.

TABLE 1. MOST WIDELY USED SUTRAS IN VEDIC MATH [5].

Sl. No	Sutra	Meaning
1	Ekadhikena Purvena	More than one that of previous (division)
2	Nikhilam Sutra	Total from 9 and final from 10 (multiplication)
3	UrdhvaTiryagbhyam	Vertical and Crosswise (multiplication)

B. *Proposed Vedic Multiplier Design Flow*

Vedic multiplier is designed by using Urdhva Tiryagbhyam vertically and crosswise sutra. With the use of this multiplier, the multiplication method can be performed on different kinds of 8-bit numbers. In this work, initially 2×2 Vedic multiplier is designed. Further 2 × 2 multiplier is used as basic block to design 4-bit Vedic multiplier. Then finaly 4-bit multiplier is used to design 8-bit multiplier.

C. *2 × 2 Vedic Multiplier*

The 2-bit Vedic multiplier structure is illustrated in the Fig.4. The vertically and crosswise algorithm is used to design the 2-bit multiplier which requires the basic AND gates and Half adder. Let A and B are the two numbers which are given to the 2×2 multiplier, where A= $\{a_1, a_0\}$ and B= $\{b_1, b_0\}$. The design of 2-bit Vedic multiplier requires four 2-input AND gates and two half adders. The algorithm uses vertical and crosswise multiplication. The Table 2 shows the illustration of steps involved in vertical and diagonal multiplication and its equivalent expression for Urdhva Tiryagbhyam sutra.

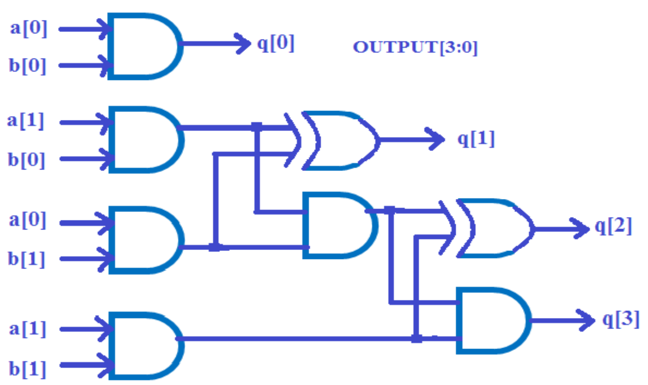


Fig. 4. Two-bit Vedic multiplier using Urdhva Tiryagbham sutra.

TABLE 2. ILLUSTRATION OF TWO-BIT VEDIC MULTIPLICATION ALGORITHMS.

Sl.No.	Pictorial representation of the steps	Equations
1	Step 1: a (1) a (0) b (1) b (0)	q (0) = a (0) *b (0)
2	Step 2: a (1) a (0) b (1) b (0)	q (1)= [a (1) *b (0)] + [a (0) *b (1)]
3	Step 3: a (1) a (0) b (1) b (0)	q (2) = [a (1) *b (1)] + C0

This algorithm entirely different from other traditional algorithms such as Booth multiplier, Array multiplier and Wallace tree multiplier.

D. *4×4 Vedic Multiplier*

The 4-bit Vedic multiplier construction is given in Fig.5. The 2×2 Vedic multiplier is used as a basic component for designing 4-bit Vedic multiplier, in which the partial products are added by a single 6-bit Modified Carry Save Adder. In this multiplier, the concatenation is used to combine the partial product q3 [3:0] with the q0 [3:2] in order to avoid the number of adders so, that the remaining partial products q2 [3:0] and q1 [3:0] are given to the adder.

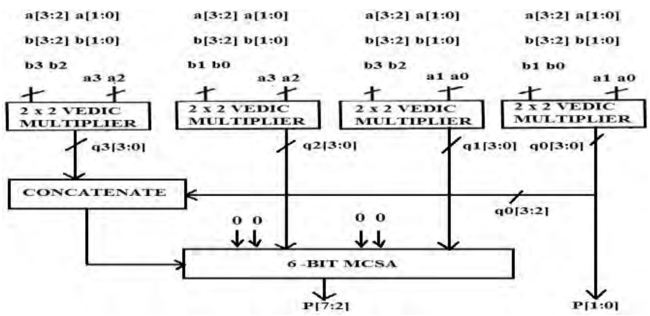


Fig.5. 4-bit Vedic multiplier using Urdhva Tiryagbhyam sutra.

E. *Design of 8 × 8 Vedic Multiplier*

The 8-bit Vedic multiplier using Urdhva Tiryagbhyam sutra is illustrated in Figure 6. The design 4-bit Vedic multiplier using Urdhva Tiryagbhyam sutra is used as a fundamental block for the design of 8-bit Vedic multiplier. In this multiplier, the unfinished products addition is carried out by using 12-bit MCSA adders.

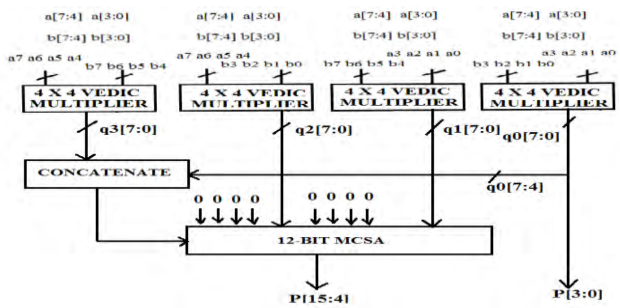


Fig.6. 8-bit Vedic Multiplier using Urdhva Tiryagbham sutra.

F. *Modified Carry Save Adder*

In any multiplier, the adder acts an extremely vital role in designing the multipliers. The Fig.7 shows the design of MCSA. In this work, the MCSA is used to compute the addition of partial product. The CSA is used to perform 3-bit addition at one time. Here, the three-bit input is processed and transformed to 2-bit output sum and carry at initial stage. In initial stage, outcome carry is not propagated during addition process. In order to generate carry, ripple carry adder is used in second stage for carry propagation.

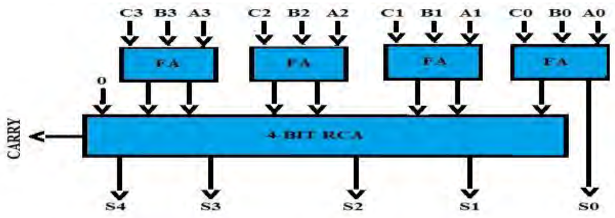


Fig. 7. Design of modified carry save adder.

V. RESULTS ANALYSIS

In this work 8-bit Vedic multiplier, using Urdhva Tiryagbhyam sutra is proposed. The Verilog Hardware Description Language is used to develop the code. The proposed Vedic multiplier is compared with i) Booth multiplier ii) Array multiplier iii) Wallace tree multiplier. To develop, simulate and synthesize the proposed design XILINX ISE 14.7 software is used. In addition, the design is verified on Spartan-6 FPGA hardware. To study the performance of the proposed design, parameters such as area, time and memory is computed. Further design is compared with traditional multiplier.

A. *Booth Multiplietr Results*

The Fig.8 illustrates the simulation results of 8-bit booth multiplier. To multiply the two 8-bit signed number it requires 25.860 ns computation time.

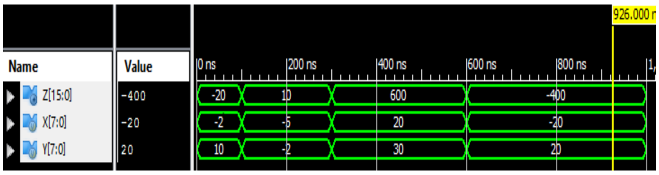


Fig.8. Simulation results of 8-bit Booth multiplier.

B. *Simulation Results of Traditional Array Multiplier*

The Fig.9. Shows the simulation results of 8-bit Array Multiplier, in which two 8-bit inputs are provided and the corresponding 16-bit outputs are generated. It is observed that for a given input, output is verified. The total computation time required to generate the output is 23.106 ns

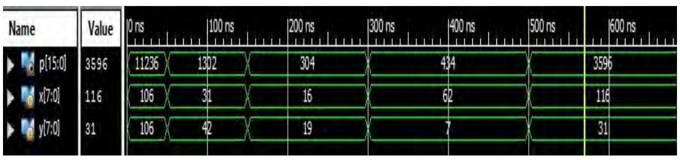


Fig.9. Simulation results of 8-bit Array multiplier.

C. *Wallacetree Multiplier Results*

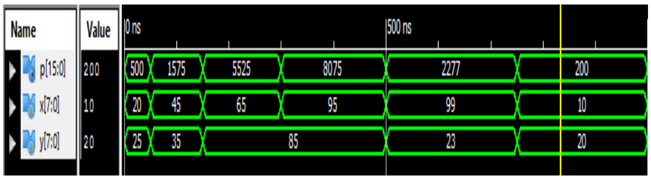


Fig.10. Simulation result of 8-bit Wallace tree multiplier.

The Fig.10 illustrates the simulation result of 8-bit Wallace tree multiplier as discussed in Section III. Simulation results verified for the different 8-bit inputs. It is also observed that to obtain the product of two 8-bit numbers, it requires 16.478 ns computational time.

D. *Simulation Results Of Vedic Multiplier*

In this section the Technology schematic and simulation results of proposed Vedic, multiplier is discussed. The Fig.11 shows the technology schematic of proposed 8-bit Vedic multiplier.

The Fig.12 shows the simulation results of proposed 8-bit Vedic multiplier. Two 8-bit numbers are given as inputs to the multiplier, which gives the corresponding 16-bit product. In addition, the simulation results of the proposed scheme are mathematically verified for the various inputs. In this work, the Vedic multiplier requires less numbers of adders, to add the partial products which interns reduces the computation delay. To compute the product, the proposed scheme requires 14.219 ns time. The time require to obtain the product is less compared to other multipliers such as i) Booth multiplier ii) Array multiplier and iii) Wallace tree multiplier.

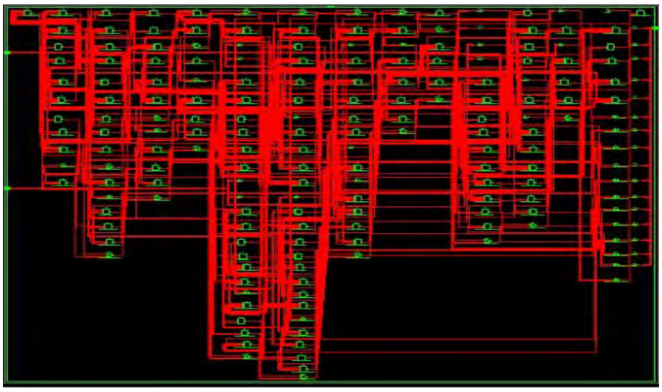


Fig.11. Technology schematic of 8-bit Vedic multiplier using Urdhva Tiryagbhyam multiplier.

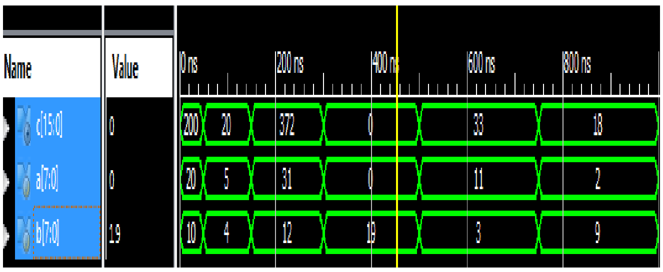


Fig.12. Simulation result of proposed 8-bit Vedic multiplier using Urdhva Tiryagbhyam multiplier.

E. Comparision of Vedic Urdhva Tiryagbhyam Multiplier With Traditional Multipliers.

The proposed Vedic Urdhva Tiryagbhyam multiplier is compared with the traditional multipliers like Booth multiplier, Array multiplier and Wallace tree multiplier with respect to parameter such as Area, Delay and Memory. The Table 3 illustrates the comparison of area occupied, speed and memory utilization of various multiplier architectures.

TABLE 3. COMPARISON OF VEDIC MULTIPLIER WITH ARRAY MULTIPLIER, BOOTH MULTIPLIER AND WALLACE TREE MULTIPLIER

Sl. no	Multiplier used	Lut's used	%of area occupied	Memory in (kb)	Delay in(ns)
1	Booth	216	9%	296640	25.860
2	Array	130	5%	280448	23.106
3	Wallace tree	145	6%	280448	16.478
4	Vedic(proposed)	128	4%	295744	14.219

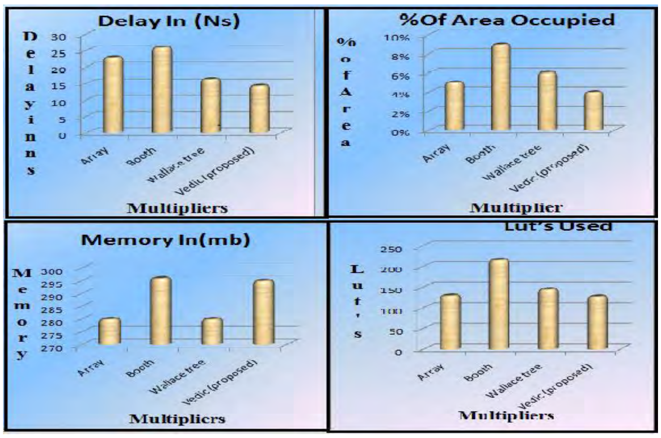


Fig.13. Plot of performance comparison of Vedic multiplier with different multipliers.

The Fig.13 shows the plot of performance comparison of 8-bit Vedic multiplier using Urdhva Tiryagbhyam multiplier with i) Booth multiplier ii) Array multiplier and iii) Wallace tree multiplier for the parameter such as delay, area, memory and number of Look Up Table (LUT). Out of these multipliers Booth multiplier requires more time for product generation among the traditional Array multiplier. The Wallace tree multiplier requires comparatively less time delay for product generation compared to Array and Booth multiplier. Whereas, the proposed Vedic multiplier with Tiryagbhyam sutra requires comparatively less time to compute the product of two 8-bit binary numbers compared to Booth multiplier, Array multiplier and Wallace tree multiplier. The number of LUT's used in all the four cases is relatively same. However, the area consumed in proposed Vedic multiplier is relatively low.

VI. CONCLUSION

In this paper, the design of 8-bit Vedic multiplier using Urdhva Tiryagbhyam sutra and Modified Carry Save Adder is discussed. The proposed multiplier is simulated and synthesized by Xilinx 14.7 on the device Spartan-6, its performance is compared with existing multiplier such as i) Booth multiplier ii) Array multiplier and iii) Wallace tree multiplier. The results show 8-bit Vedic multiplier using Urdhva Tiryagbhyam sutra with MCSA provides high speed in computing products of two 8-bit binary numbers with compared to Booth multiplier, Array multiplier and Wallace tree multiplier and it consumes 14.219 ns to compute the product of two 8-bit numbers. Hence, the proposed 8-bit Vedic multiplier is more efficient than other type traditional multipliers. Hence, the proposed multiplier can be used for the Digital signal processing applications where the speed is significant.

REFERENCES

[1]. G. Ganesh Kumar, V. Charisma. "Design of High-Speed Vedic Multiplier using Vedic Mathematics Techniques" International Journal of Scientific and Research Publications, Volume 2, Issue 3, March 2012, 2 ISSN 2250-3153.

[2]. S.K. Peda Roshan Jameer, V. Pratap Reddy, "Design of Vedic Multiplier Using Higher Order Compressor to Increase the Speed and Area", international journal of innovative Research in Science and Engineering, vol. No. 2 Issue 01, January 2016.

[3]. G. R. Gokhale , S. R. Gokhale "Design of Area and Delay Efficient Vedic Multiplier Using Carry Select Adder", Proceedings of IEEE International Conference on Information Processing (ICIP). Pune, India, 16-19 Dec. 2015

[4]. G. Challa Ram ; Y. Rama Lakshmanna ; D. Sudha Rani ; K. Bala Sindhuri "Area Efficient Modified Vedic Multiplier" Proceedings of IEEE International Conference on Circuit, Power and Computing Technologies (ICCPCT), Nagercoil, India, 18-19 March 2016. 978-1-5090-1277-0/16/$31.00@2016IEEE.

[5]. S.P. Pohokar, R.S. Sisal, K.M. Gaikwad, M.M. Patil, Rushikesh Borse, "Design and Implementation of 16 x 16 Multiplier Using Vedic Mathematics", Proceedings of IEEE International Conference on Industrial Instrumentation and Control (ICIC), Pune, India, 28-30 May 2015

[6]. C.Sheshavali, K Niranjan Kumar, "Design and Implementation of Vedic Multiplier", International Journal of Engineering Research and Development e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com Volume 8, Issue 6 (September 2013), PP.23-28

[7]. Dravik KishorBhai Kahar, Harsh Mehta, "High speed vedic multiplier used vedic mathematics", Proceedings of IEEE International Conference on Intelligent Computing and Control Systems ICICCS, Madurai, India, 15-16 June 2017.

[8]. Andrew D. Booth. "A signed binary multiplication technique" The Quarterly Journal of Mechanics and Applied Mathematics, Volume 4, Issue 2, 1951, PP 236–240.

[9]. Baugh, Charles Richmond; Wooley, Bruce A. "A Two's Complement Parallel Array Multiplication Algorithm". IEEE Transactions on Computers, Volume. 22, Issue.12, 1973, PP 1045–1047.

[10]. C. S. WALLACE, "A Suggestion for a Fast Multiplier", IEEE Transactions on Electronic Computers, Volume EC-13 , Issue: 1,1964 , PP 14 – 17.

[11]. Dravik kishorbhai Kahar, Harsh Mehta," High Speed Vedic Multiplier Used Vedic Mathematics", Proceedings of IEEE International Conference on Intelligent Computing and Control Systems, Madurai, India, 15-16 June 2017 PP 356 – 359.

[12]. W.B. Vasantha Kandasamy; Florentin Smarandache "Vedic Mathematics: Vedic Or Mathematics: A Fuzzy and Neutrosophic Analysis "American Research Press. ISBN 978-1-59973-004-2. Retrieved 23 May 2013.

2020 Third International Conference on Advances in Electronics, Computers and Communications (ICAECC) | 978-1-7281-9183-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICAECC50550.2020.9339529

Image Encryption and Decryption Using Key Sequence of Triple Logistic Map for Medical Applications

Dr. Rohith S
Department of ECE, NCET, Bengaluru, Karnataka, India
rohithvjp2006@gmail.com

Ms. Jahnavi L
Department of ECE, NCET, Bengaluru, Karnataka, India
jahnavi131@gmail.com

Ms. Bhuvaneshwari S C
Department of ECE,NCET, Bengaluru, Karnataka, India
bhuvanagowdasc@gmail.com

Mr. Supreeth S
School of C & IT REVA University, Bengaluru, Karnataka, India
supreeth1588@gmail.com

Dr. B K Sujatha
ETE Department, Ramaiah Institute of Technology,Bengaluru, Karnataka, India
bksujatha@msrit.edu

Abstract*—In the present world scenario, security of medical images from unauthorized access is one of the key challenges. This paper provides one of the solutions to solve the security of medical images. The proposed scheme uses a combination of key sequences of Triple Logistic Map (TLM) for the encryption process. To generate the key sequence Logistic function is used with initial value X_0 and r. The three key sequences$\{X_{1i}\}$, $\{X_{2i}\}$ and $\{X_{3i}\}$ with different initial value are used ($X_{10} \neq X_{20} \neq X_{30}$) with r=3.99. The key sequences $\{X_{1i}\}$, $\{X_{2i}\}$ and $\{X_{3i}\}$ are converted into discrete key sequence $\{K_{1i}\}$, $\{K_{2i}\}$ and $\{K_{3i}\}$ in the range of 0 to 255. The combined key sequence $\{K_i\}$ is obtained using bit by bit logical XOR operation between $\{K_{1i}\}$, $\{K_{2i}\}$, and $\{K_{3i}\}$. The resultant Key sequence $\{K_i\}$ is used forthe encryption process of the medical image. To evaluate the performance of the proposed scheme different grayscale images are chosen. The proposed TLM based image encryption scheme is compared with the encryption scheme using (i) Logistic Map (LM) alone ii) Double Logistic Map (DLM). To evaluate the proposed scheme parameters such as (i) Visual analysis (ii) Histogram plot (iii) Mean Square Error (MSE) iv) Correlation (v) Entropyis used. The simulation result shows that the proposed TLM based scheme provides better performance and immune to statistical attacks.

Keywords- Logistic Map, Triple Logistic Map, Combined Key Sequence, Image encryption.

I. INTRODUCTION

The rapid development of computer technology, Multimedia technology, and the Internet, production, transmission, and storage of digital images are most common. Security plays an important role in the transmission or storage of medical images of the patient such as X-Ray, MRI, etc. The security of the medical image is at most important due to sensitive information about the patient. The unauthorized alteration of the medical image by an intruder may lead to the wrong decision. Also, there may be chances of leakage of confidential health information about the patient by unauthorized parties. To avoid problems of unauthorized access or alteration, cryptography may be one of the solution.

The cryptography may be broadly classified as an asymmetric and asymmetric cryptosystem. Further symmetric cryptosystems classified as a block cipher or stream cipher. The well-known block cipher schemes are Data Encryption Standard (DES), Advanced Encryption Standard (AES), and RSA [1]. These block-based encryption techniques are complex in computation, relatively highly redundancy, and correlation between the encrypted image pixels are high and require high computing power for image encryption applications. Hence these block cipher based cryptosystems are may not be suitable for image encryption applications. However stream-cipher based cryptosystem operate with a time-varying reversible transformation on individual digits of the real number.

Chaotic maps are one example of the Stream cipher. Last two decades, manyresearchers suggested the idea of using the chaos-based cryptosystem for image encryption application[2]-[9]. The reason is, the chaotic map has the property of periodicity, sensitivity to initial conditions and parameters, high encryption rate, and simple in implementation, etc [2]-[10]. The chaotic key sequences play an important role to confuse the original to encrypted relationship imagesusing varying the position of the pixel and pixel values[4]. For encryption of the images, a wide variety of 1-Dimensional (1D), 2-Dimensional(2D) chaotic maps are available [2] [4] [5] [8][9]. Out of these schemes, many researchers have proven using a combination of the chaotic maps provide better performance than or single map[4]-[9].

In [4] Image encryption scheme using combination of two logistic map is discussed. In the first phase, they used XOR operation between the two logistic sequences to obtain the single key sequence. In the second phase, image encryption is performed using the single sequence. An image encryption system suitable for biomedical applications is discussed in [5]. They used two discrete logistic maps to generate a single pseudo-random sequence. In [6] multi-chaotic based image encryption scheme is discussed. They used a combination of chaotic sequence of 3D- Baker map, Henon map, and logistic map, for encryption of images. However, the histogram plot of the encrypted image was not uniformly distributed. In [7] a combination of a standard map and a Logistic map based image encryption scheme is discussed. In [8] they used Pixel Mapping Table (PMT) and logistic map for encryption. Firstly PMT was used to confuse and increase uncertainty in the original image. Then rows and columns replacement was done. Then logical XOR operation was applied with random sequence generated using the logistic map. In [9] Logistic map and Lozi map based combined key sequence based image encryption scheme is discussed. They used the key sequence of the Logistic map and the Lozi map is combined with logical XOR operation to obtain a single key sequence. So obtained sequence is used for encryption of the image.

In this work,Triple Logistic Map (TLM) based image encryption scheme is proposed. Initially, the three Logistic maps with different initial values generated such that all key sequences are statistically independent. These discrete key sequences are combined using simple logical XOR operation. So obtained combined key sequences are used to encrypt the medical image.

The reminder of the paper is as follows. The basics of the Logistic map, proposed TLM based image encryption and decryption algorithm discussed in Section II. Section III covers Simulation tests and an overview of the performance of the proposed image encryption scheme. The conclusion of the paper is given in Section IV.

II. PROPOSED SCHEME

A. Logistic map

The logistic map[4] is simple non-linear dynamic equations show complex chaotic behavior. The logistic map is one of the simple chaotic maps that are very useful for cryptography applications. The chaotic sequence of the logistic map is highly sensitive to the change of its initial value. Mathematically, the logistic map is defined as

$$X_{n+1}= r\, X_n(1- X_n) \qquad (1)$$

Where X_0 is initial value ranges between 0 to 1 and positive integer r is bifurcation parameter, lies between 0 to 4. Using equation (1), the logistic sequence $\{X_i\} = \{X_1, X_2,\ldots X_n\}$ are generated. The sequence of $\{X_i\}$ lies between 0 and 1 which are random.Typical plot of generated sequence $\{X_i\}$vs length t=500 with an initial value X_0 = 0.5 and r = 3.9999 are shown in Fig.(1). It is observed in the plot that the sequence is random and nonperiodic.

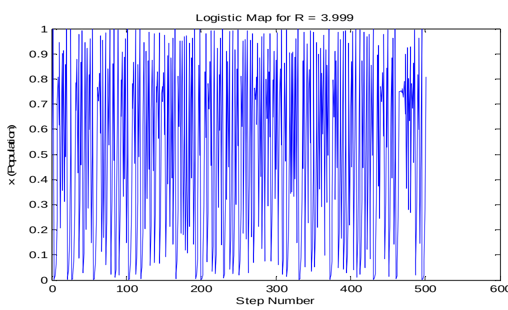


Fig.1: Sequence of Logistic Map of Length t=500 With an Initial Condition X_0=0.5 and r=3.9999.

Further, sequence of logistic map$\{X_i\}$ is converted into an Unsigned integer by elements in the range 0 to 255 of $X_i$is multiplied with 255. Round the elements of the X_i value to the nearest decimal value as defined in equation (2). So obtained sequence is further transformed into block of the 8-bit word and used in the generation of main key sequence. The obtained sequence is named as Ki.

$$K_i=Round\ (X_i*255) \qquad (2)$$

B. Tripple Logistic map Sequence

The three logistic key sequences $\{X_{1,i}\},\{X_{2,i}\},\{X_{3,i}\}$ generated using r=3.999 with different initial condition such that $X_{10} \neq X_{20} \neq X_{30}$. Where $\{X_{1,i}\},\{X_{2,i}\},\{X_{3,i}\}$ logistic sequence of first, second and third Logistic map respectively. Fig.(2) shows proposed Triple Logistic Map (TLM) based key sequence generation scheme. The sequence generated by all the three logistic map are uncorrelated in nature. So obtained three sequences are converted in to block of 8-bit sequence $\{K_{1i}\} = \{K_{11}, K_{12}, K_{13}\ldots K_{1t}\}$ $\{K_{2i}\} = \{K_{21}, K_{22}, K_{23}\ldots K_{2t}\}$ and $\{K_{3i}\} = \{K_{31}, K_{32}, K_{33}\ldots K_{3t}\}$ where 't' is size of the image using Eq (3). To obtain the proposed TLM based key sequence $\{K_i\}$ bit by bit XOR operation is performed between the three statistically independent key sequence of logistic map $\{K_{1i}\} \neq \{K_{2i}\} \neq \{K_{3i}\}$. The TLM based key sequence is obtained by

$$\{K_k\} = (\{K_{1k}\} \oplus \{K_{2k}\} \oplus \{K_{3k}\}) \qquad (3)$$

The key sequence $\{K_k\}$ obtained from Eq(3) is highly uncorrelated and ranges between 0 to 255.

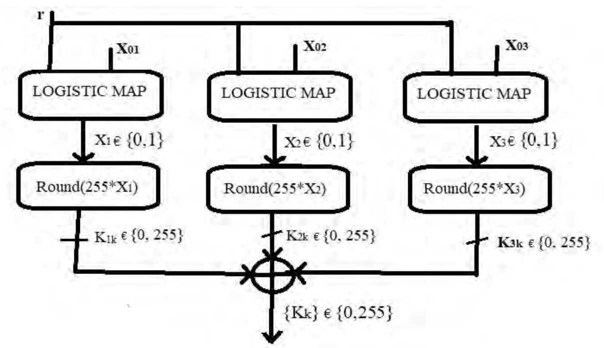


Fig.2:Proposed Triple Logistic Map (TLM) based key sequence generation scheme.

C. Proposed Image Encryption Scheme

Image encryption scheme using Triple Logistic Map (TLM) based key sequence is discussed. The steps are listed below.

Step1: First, 8 bit image pixels $I(x, y)\in(0, 255)$ of two dimension are converted into one dimension (1D) pixels $\{I_i\} = \{I_1,I_2,\ldots\ldots I_t\}$, Where total image pixels i.e t=M×N. M and N is row and column of two dimensional (2D) image pixels I(x, y) respectively.

Step2: The logical XOR operation is performed between image pixels $\{I_i\} = \{I_1,I_2,\ldots\ldots I_t\}$ and key sequence of TLM $\{K_i\}=\{K_1, K_2, K_3\ldots K_t\}\{K_i\}\in(0, 255)$. Hence encrypted image sequence $\{E_i\} = \{E_1,E_2,\ldots\ldots E_t\}$ are obtained. It is computed using equation (4)

$$\{E_i\} = [\{I_i\} \oplus \{K_i\}] \qquad (4)$$

Step 3: The encrypted image pixels$\{E_i\} = \{E_1,E_2,\ldots E_t\}$ are converted into 2D array of size M×N. The resulting image is called encrypted image E(x, y).

D. Proposed Image Decryption Scheme

In this section steps used in the image decryption process are given.

Step1: The Encrypted image pixels E(x, y) are transformed into image pixels of 1D $\{E_i\} = \{E_1, E_2, \ldots\ldots E_t\}$.

Step2: The Decryption process is applied using the same TLM based key sequence $\{K_i\}$ used during encryption. It is computed using equation (5)

$$\{D_i\} = [\{E_i\} \oplus \{K_i\}] \qquad (5)$$

Step3: The 1D decrypted imagepixels $\{D_i\}$are transformed into a 2D size of M×N. The resulting image is decrypted.

III. EXPERIMENTAL RESULTS

The performance analysis of the proposed image encryption scheme is studied using MATLAB-2017 software. For the simulation study, 8-bit gray images of 256 × 256 pixels in size are chosen as test images. In this paper, seven different images are chosen out of which six images are medical images and one image is Lena's image. The parameters such as i) histogram plot (for both original and encrypted image) ii)Mean Square Error(MSE) iii) Entropy iv) Correlation are considered to study the performance of the proposed scheme. The proposed single key sequence $\{K_i\}$ is generated from the sequence of three different logistic map $\{K_{1,i}\}\{K_{2,i}\}$ and $\{K_{3,i}\}$ wherer=3.99 and initial value $X_{01}=0.1$ $X_{02}=0.2$ $X_{03}=0.3$ are chosen.The proposed key generation scheme and image encryption and decryption process are given in Section-II. The proposed TLM based image encryption scheme is compared with image encryption using i) Logistic Map (LM) ii) Double Logistic Map (DLM).

A. Visual Analysis

To analyze the proposed TLM based encryption scheme, Lena's image, and six different medical images are considered and shown in Fig (3a) (1-7). Fig.(3a)(1-7) shows the original Lena image and different medical images. Fig.(3b)(1-7)are encrypted images using only the Logistic Map (LM)based scheme. Fig. (3c) (1-7)are encrypted images Double Logistic Map (DLM) based scheme. Fig. (3d)(1-7) shows encrypted images using $\{K_i\}$ of the proposed Triple Logistic Map (TLM) based scheme. It is observed that encrypted images will not reveal any information about the original image . However,the proposed TLM based scheme provides better performance compared to LM or DLM based scheme. To verify the sensitivity of the key in the case of proposed TLM based scheme, decryption process of first image Fig(3d)(1) a wrong key witha combination of X_{10}=0.1000000000000001, X_{20}=0.2000000000000001, and X_{30}=0.3000000000000001 with the same r=3.99 is used. After employing the wrong key the decrypted image is not the same as the original image. Hence even very small changes in the combination of the initial value of X_{10}, $X_{20,}$ and X_{30} it is highly impossible to decrypt the image. Hence it shows TLM based scheme highly sensitive to the initial value or key.

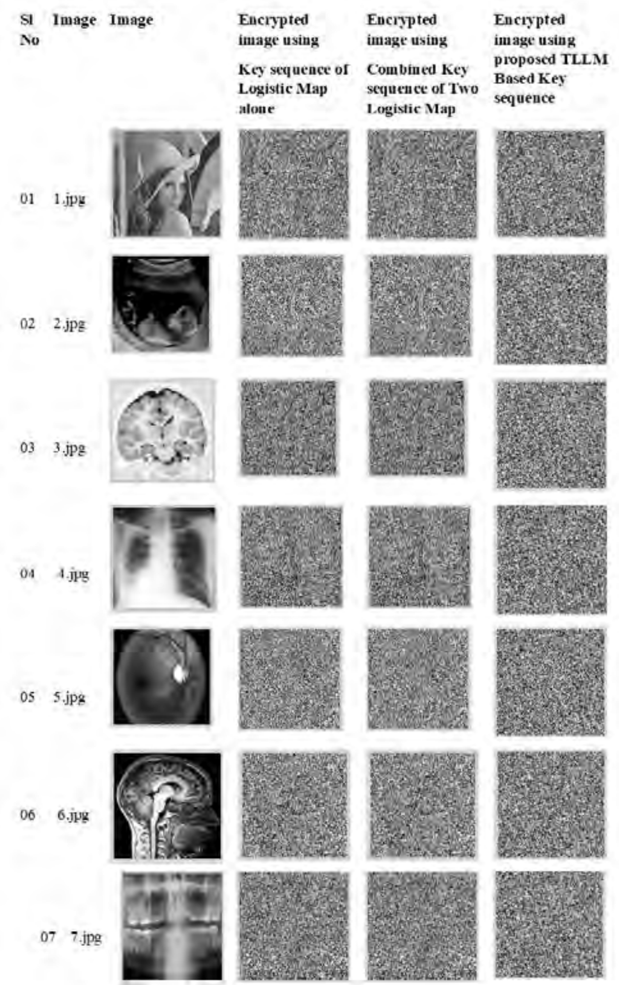


Fig.3a (1-7): Original Image. Fig. (3b)(1-7) Encrypted image X_0=0.1 r=3.99 Key Sequence of Logistic map alone Fig. (3c)(1-7)shows encrypted images with X_{10} and X_{20}=0.2 r=3.99 using the Key sequence $\{K_{2i}\}$ of Double Logistic Map (DLM) based scheme. Fig. (3d) (1-7) shows encrypted images with X_{10} =0.1, X_{20}=0.2 and X_{30}= 0.3 r=3.99 using the Key sequence $\{K_{3i}\}$ of Proposed Triple Logistic Map (TLM) based scheme.

B. Histogram Analysis

Histogram plot is a graphical representation of pixels values of an image. It is a plot of the Maximum pixel count with increasing gray level color. In this paper, histogram plot is used to study the performance of the image encryption scheme. Here histogram of the original image, encrypted image using i) Logistic Map ii) DLM based scheme, and iii) Proposed TLM based scheme are studied. Fig. 4(a)(1-7) shows a histogram plot of seven different images as shown in Fig(3a)(1-7). Similarly, Fig.4(b)(1-7). 4(c)(1-7) and 4(d)(1-7) shows a Histogram plot of the encrypted image using i) Logistic Map ii) DLM based scheme and iii) Proposed TLM based scheme respectively. From Fig.4(b)(1-7), (4c)(1-7), and (4d)(1-7), it is observed that the histogram plot of the encrypted image fairly uniform. However,the proposed TLM based scheme is flat and fairly uniform compared with encryption using an LM or DLM based scheme sequence.

Therefore the scheme proposed is highly immune from statistical attacks.

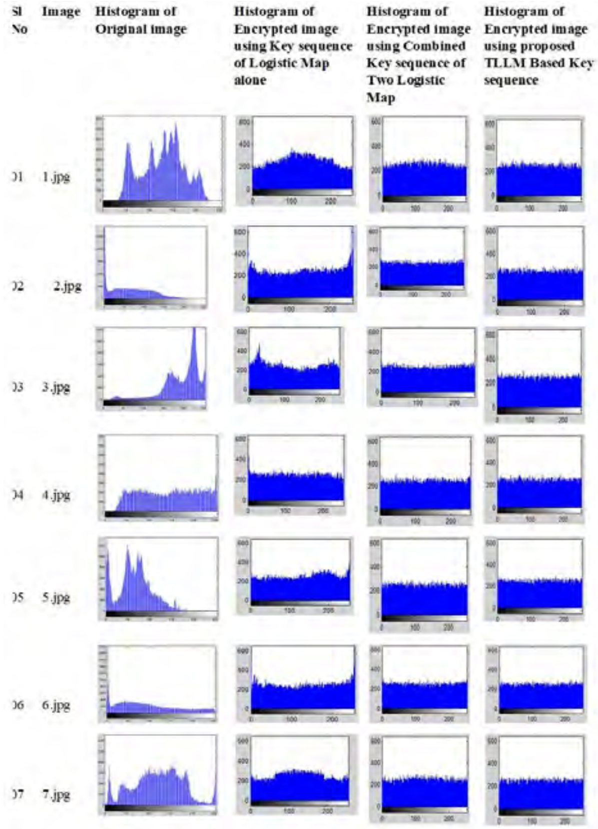


Fig. 4a: shows Histogram plot of Original image. Fig. 4b: shows Histogram plot of Encrypted image using Logistic map key sequence with X_{10} =0.1 and r=3.99 Fig. 4c: shows the Histogram plot of encrypted images with X_{10} =0.1 and X_{20}=0.2 r=3.99 using the Key sequence of Double Logistic Map (DLM) based scheme. Fig. 4d: shows Histogram plot of encrypted images with X_{10} =0.1, X_{20}=0.2 andX_{20}=0.3 r=3.99 using the Key sequence $\{K_i\}$ of Proposed Triple Logistic Map TLM based scheme.

C. Mean Square Error Analysis

The parameter Mean Square Error (MSE) can be defined as

$$\text{MSE}=\frac{1}{MN}\sum_{i=0}^{M\times N\text{-}1}[I(i,j)\text{-}E(i,j)]^2 \qquad (6)$$

Where I(i,j) is original image pixel value
E(i,j) is encrypted image pixel value,
M and N represents rows and columns of the image.

Ideally, the parameter MSE should be as high as possible in case of an image encryption scheme. It means a Higher value of MSE between original and encrypted image indicates,better immunity against statistical attacks. The MSE is computed between the images given in Fig.(3a)(1-7) and corresponding encrypted image using i) Logistic Map alone ii) DLM based scheme iii) Proposed TLM based scheme and tabulate in Table1. It is observed in Table1 that MSE value in case of proposed TLM based encryption scheme is better compared to a scheme using DLM based scheme or LM based Scheme. Similarly for all the medical image cases, MSE in the case of TLM based encryption scheme provides the best performance compared DLM or LM based scheme.

Table 1MSE VALUE BETWEEN ORIGINAL AND ENCRYPTED IMAGE

Sl no	Image	MSE (Logistic map)	MSE(Combination of two Logistic maps)	MSE(Proposed TLM based scheme)
01	1.jpg	7084.3	7300.2	7511.3
02	2.jpg	19989.00	13606.00	13562.00
03	3.jpg	14016.00	12617.00	12588.00
04	4.jpg	11683.00	10811.00	10929.00
05	5.jpg	12125.00	11144.00	11201.00
06	6.jpg	14447.00	13028.00	13031.00
07	7.jpg	9212.2	8920.9	9130.80

D. The entropy of the image

Information entropy[3][10] is used to calculate the degree of unpredictability of the content of information. Source C information entropy H(C), defined as

$$H(C)=\sum_{i=0}^{M\times N-1}P(C_i)\log_2\frac{1}{P(C_i)} \qquad (7)$$

Where $P(C_i)$ is the probability of symbol occurrence C_i
M×N is the total number of symbols.
Ideally, the maximum entropy value for an 8-bit grayscale picture is 8 bits/symbols that occur when equiprobable pixel values of image or histogram are flat. The entropy value of the image also indicates the distribution of image pixel value. It means a higher value of entropy indicates the uniform distribution of the image pixels.

Table 2Original and encrypted Entropy values of an image

Sl No	Image Type	The entropy of Original image	The entropy of encrypted image using (Logistic map	Entropy encrypted image using a Combination of two Logistic map	Entropy encrypted image using Proposed TLM based scheme
01	1.jpg	7.3837	7.9736	7.9955	7.9969
02	2.jpg	6.4531	7.9751	7.9972	7.9984
03	3.jpg	6.8764	7.9793	7.9966	7.9974
04	4.jpg	7.5701	7.9931	7.9966	7.9975
05	5.jpg	7.0459	7.9903	7.9975	7.9975
06	6.jpg	7.0666	7.9812	7.9967	7.9973
07	7.jpg	7.6101	7.9826	7.9960	7.9974

In Table 2 the entropy value of original and encrypted image using LM alone, DLM scheme, proposed TLM based scheme is computed and tabulated for the entire seven different images. For all the cases entropy of encrypted image is greater than the entropy of the original image. It is seen that entropy of Encrypted image using the proposed TLM-based scheme is high when compared with encrypted image entropyusing LM or DLM based scheme. It shows in the case of the proposed method, the distribution of pixels in encrypted image pixels is uniform compared to LM or DLM based scheme. Hence, the proposed TLM based scheme highly robust against statistical attacks.

E. Correlation analysis

Correlation coefficient [3][10] is as defined as

$$Correlation\ Coefficient = \frac{Cov(x,y)}{\sqrt{D(x)}\sqrt{D(y)}} \quad (8)$$

Where, Covariance of x, y is

$$Cov(x,y) = \frac{1}{L}\sum_{i=0}^{L}(x_i - E(x)(y_i - E(y) \quad (9)$$

Where, x and y, adjacent pixel values
L=M× N and variance of x_i is

$$D(x) = \frac{1}{L}\sum_{i=0}^{L}\left(x_i - E(x)\right)^2 \quad (10)$$

Mean of the value of x_i is

$$E(x) = \frac{1}{L}\sum_{i=0}^{L} x_i \quad (11)$$

Similarly, D(y) is the variance of y_i and E(y) is mean of y_i

Table-3 Correlation Coefficient

Sl No	Name of the image	Correlation (Logistic map)	Correlation (Combination of two Logistic maps)	Correlation(Proposed TLM based scheme)
01	1.jpg	-0.0537	0.0077	-0.0015
02	2.jpg	-0.0441	-0.0419	0.0015
03	3.jpg	-0.0347	0.00059247	0.00034
04	4.jpg	-0.0679	0.0051	-0.0052
05	5.jpg	-0.0419	0.00080854	0.0068
06	6.jpg	-0.0643	0.0042	0.00012034
07	7.jpg	-0.0725	0.0070	-0.0024

Correlation coeeficient is computed between pixels for the original and encrypted image using (i) Logistic map alone (ii) DLM based scheme and proposed TLM based scheme for seven different image cases. Ideally Correlation Coefficient between original and decrypted image ‘1’ is with the same key. It indicates both original and decrypted images are the same. However Ideally Correlation coefficient between Original and encrypted image is ‘0’. In Table, It is observed that the correlation coefficient is very low in the proposed TLM based Scheme compared to DLM based encryption scheme and encryption using a Logistic map-based scheme. It indicates pixels are highly uncorrelated and robust against statistical attacks.

IV. CONCLUSION

In this work, Triple Logistic Map based image encryption and decryption scheme for the medical image, applications is proposed. The proposed scheme is highly sensitive to the chosen initial value and r. The proposed TLM-based scheme is compared with LM and DLM based image encryption scheme. The results indicate encrypted image will not reval any visual information about original image and highly un-correlated. Also, histogram plot of encrypted image using proposed encryption flat and fairly uniform compared to encryption using DLM or LM based scheme. Also Entropy, Correlation, Mean Square Error value is computed for proposed TLM and DLM, Logistic map based scheme. It is noted that the proposed image encryption scheme using TLM provides best results compared to encryption using DLM or LM based scheme.

V. REFERENCES

[1] William Stallings, “Cryptography and network security - principles and practice”(3. ed.). Prentice-Hall 2003, ISBN 978-0-13-111502-6, pp. I-XIV, 1-681

[2] Chengqing Li, Tao Xie , Qi Li “ Cryptanalyzing image encryption using chaotic logistic map” Nonlinear Dynamics, 2014, Vol. 78, pp 1545–1551.

[3] C. E. Shannon, “A Mathematical Theory of Communication” The Bell System Technical Journal, Vol. 27, 1948, pp. 379–423.

[4] Safi, Haifaa W, Maghari, Ashraf Y. A. “Image Encryption Using Double Chaotic Logistic Map” Proceedings of IEEE International Conference on Promising Electronic Technologies (ICPET), 2017, pp. 66-70.

[5] ISMAIL, Samar M.,“Biomedical image encryption based on double-humped and fractional logistic maps”,Proceedings of IEEE international conference on modern circuits and systems technologies (MOCAST). 2017. pp. 1-4.

[6] Hong, Lianxi Li, Chuanmu. “A Novel Color Image Encryption Approach Based on Multi-chaotic System”, Proc. of IEEE 2nd International Conference on Anti-counterfeiting, Security and Identification, 2008, pp 1-5.

[7] Liu, Jin-mei Qu, Qiang, “Cryptanalysis of a Substitution-Diffusion Based Image Cipher Using Chaotic Standard and Logistic Map”, Proc. of the Third International Symposium on Information Processing, October 2010,pp 67 - 69.

[8] Al-Najjar, Hazem Mohammad, Asem Mohammad AL-Najjar, and K. S. A. Arar. "Image encryption algorithm based on logistic map and pixel mapping table" Proc. of International Arab Conference on Information Technology, (ACIT 2011), pp. 56-60. 2011.

[9] Rohith S , Sujatha B K, "Image encryption and decryption using combined key sequence of Logistic map and Lozi map," Proc. of International Conference on Communications and Signal Processing (ICCSP), Tamilnadu, 2015, pp. 1053-1058.

[10] Shu Lin, Daniel J. Costello “Error Control Coding” 2nd Edition, Pearson Education,May 28th 2004India.

2020 Third International Conference on Advances in Electronics, Computers and Communications (ICAECC) | 978-1-7281-9183-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICAECC50550.2020.9339529

Image Encryption and Decryption Using Key Sequence of Triple Logistic Map for Medical Applications

Dr. Rohith S
Department of ECE, NCET, Bengaluru, Karnataka, India
rohithvjp2006@gmail.com

Ms. Jahnavi L
Department of ECE, NCET, Bengaluru, Karnataka, India
jahnavi131@gmail.com

Ms. Bhuvaneshwari S C
Department of ECE,NCET, Bengaluru, Karnataka, India
bhuvanagowdasc@gmail.com

Mr. Supreeth S
School of C & IT REVA University, Bengaluru, Karnataka, India
supreeth1588@gmail.com

Dr. B K Sujatha
ETE Department, Ramaiah Institute of Technology,Bengaluru, Karnataka, India
bksujatha@msrit.edu

***Abstract*—In the present world scenario, security of medical images from unauthorized access is one of the key challenges. This paper provides one of the solutions to solve the security of medical images. The proposed scheme uses a combination of key sequences of Triple Logistic Map (TLM) for the encryption process. To generate the key sequence Logistic function is used with initial value X_0 and r. The three key sequences$\{X_{1i}\}$, $\{X_{2i}\}$ and $\{X_{3i}\}$ with different initial value are used ($X_{10} \neq X_{20} \neq X_{30}$) with r=3.99. The key sequences $\{X_{1i}\}$, $\{X_{2i}\}$ and $\{X_{3i}\}$ are converted into discrete key sequence $\{K_{1i}\}$, $\{K_{2i}\}$ and $\{K_{3i}\}$ in the range of 0 to 255. The combined key sequence $\{K_i\}$ is obtained using bit by bit logical XOR operation between $\{K_{1i}\}$, $\{K_{2i}\}$, and $\{K_{3i}\}$. The resultant Key sequence $\{K_i\}$ is used forthe encryption process of the medical image. To evaluate the performance of the proposed scheme different grayscale images are chosen. The proposed TLM based image encryption scheme is compared with the encryption scheme using (i) Logistic Map (LM) alone ii) Double Logistic Map (DLM). To evaluate the proposed scheme parameters such as (i) Visual analysis (ii) Histogram plot (iii) Mean Square Error (MSE) iv) Correlation (v) Entropyis used. The simulation result shows that the proposed TLM based scheme provides better performance and immune to statistical attacks.**

Keywords- Logistic Map, Triple Logistic Map, Combined Key Sequence, Image encryption.

I. INTRODUCTION

The rapid development of computer technology, Multimedia technology, and the Internet, production, transmission, and storage of digital images are most common. Security plays an important role in the transmission or storage of medical images of the patient such as X-Ray, MRI, etc. The security of the medical image is at most important due to sensitive information about the patient. The unauthorized alteration of the medical image by an intruder may lead to the wrong decision. Also, there may be chances of leakage of confidential health information about the patient by unauthorized parties. To avoid problems of unauthorized access or alteration, cryptography may be one of the solution.

The cryptography may be broadly classified as an asymmetric and asymmetric cryptosystem. Further symmetric cryptosystems classified as a block cipher or stream cipher. The well-known block cipher schemes are Data Encryption Standard (DES), Advanced Encryption Standard (AES), and RSA [1]. These block-based encryption techniques are complex in computation, relatively highly redundancy, and correlation between the encrypted image pixels are high and require high computing power for image encryption applications. Hence these block cipher based cryptosystems are may not be suitable for image encryption applications. However stream-cipher based cryptosystem operate with a time-varying reversible transformation on individual digits of the real number.

Chaotic maps are one example of the Stream cipher. Last two decades, manyresearchers suggested the idea of using the chaos-based cryptosystem for image encryption application[2]-[9]. The reason is, the chaotic map has the property of periodicity, sensitivity to initial conditions and parameters, high encryption rate, and simple in implementation, etc [2]-[10]. The chaotic key sequences play an important role to confuse the original to encrypted relationship imagesusing varying the position of the pixel and pixel values[4]. For encryption of the images, a wide variety of 1-Dimensional (1D), 2-Dimensional(2D) chaotic maps are available [2] [4] [5] [8][9]. Out of these schemes, many researchers have proven using a combination of the chaotic maps provide better performance than or single map[4]-[9].

In [4] Image encryption scheme using combination of two logistic map is discussed. In the first phase, they used XOR operation between the two logistic sequences to obtain the single key sequence. In the second phase, image encryption is performed using the single sequence. An image encryption system suitable for biomedical applications is discussed in [5]. They used two discrete logistic maps to generate a single pseudo-random sequence. In [6] multi-chaotic based image encryption scheme is discussed. They used a combination of chaotic sequence of 3D- Baker map, Henon map, and logistic map, for encryption of images. However, the histogram plot of the encrypted image was not uniformly distributed. In [7] a combination of a standard map and a Logistic map based image encryption scheme is discussed. In [8] they used Pixel Mapping Table (PMT) and logistic map for encryption. Firstly PMT was used to confuse and increase uncertainty in the original image. Then rows and columns replacement was done. Then logical XOR operation was applied with random sequence generated using the logistic map. In [9] Logistic map and Lozi map based combined key sequence based image encryption scheme is discussed. They used the key sequence of the Logistic map and the Lozi map is combined with logical XOR operation to obtain a single key sequence. So obtained sequence is used for encryption of the image.

In this work,Triple Logistic Map (TLM) based image encryption scheme is proposed. Initially, the three Logistic maps with different initial values generated such that all key sequences are statistically independent. These discrete key sequences are combined using simple logical XOR operation. So obtained combined key sequences are used to encrypt the medical image.

The reminder of the paper is as follows. The basics of the Logistic map, proposed TLM based image encryption and decryption algorithm discussed in Section II. Section III covers Simulation tests and an overview of the performance of the proposed image encryption scheme. The conclusion of the paper is given in Section IV.

II. PROPOSED SCHEME

A. Logistic map

The logistic map[4] is simple non-linear dynamic equations show complex chaotic behavior. The logistic map is one of the simple chaotic maps that are very useful for cryptography applications. The chaotic sequence of the logistic map is highly sensitive to the change of its initial value. Mathematically, the logistic map is defined as

$$X_{n+1}= r\,X_n(1- X_n) \tag{1}$$

Where X_0 is initial value ranges between 0 to 1 and positive integer r is bifurcation parameter, lies between 0 to 4. Using equation (1), the logistic sequence $\{X_i\} = \{X_1, X_2,\ldots\ldots X_n\}$ are generated. The sequence of $\{X_i\}$ lies between 0 and 1 which are random.Typical plot of generated sequence $\{X_i\}$vs length t=500 with an initial value $X_0 = 0.5$ and r = 3.9999 are shown in Fig.(1). It is observed in the plot that the sequence is random and nonperiodic.

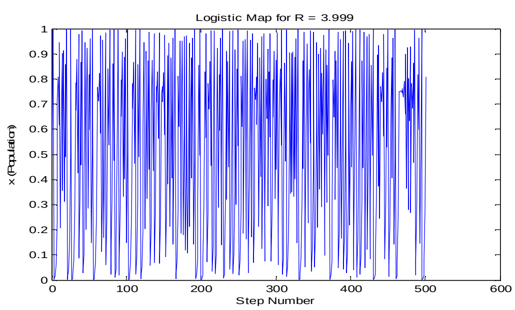


Fig.1: Sequence of Logistic Map of Length t=500 With an Initial Condition X_0=0.5 and r=3.9999.

Further, sequence of logistic map$\{X_i\}$ is converted into an Unsigned integer by elements in the range 0 to 255 of $X_i$is multiplied with 255. Round the elements of the X_i value to the nearest decimal value as defined in equation (2). So obtained sequence is further transformed into block of the 8-bit word and used in the generation of main key sequence. The obtained sequence is named as Ki.

$$K_i=Round\ (X_i*255) \tag{2}$$

B. Tripple Logistic map Sequence

The three logistic key sequences $\{X_{1,i}\},\{X_{2,i}\},\{X_{3,i}\}$ generated using r=3.999 with different initial condition such that $X_{10} \neq X_{20} \neq X_{30}$. Where $\{X_{1,i}\},\{X_{2,i}\},\{X_{3,i}\}$ logistic sequence of first, second and third Logistic map respectively. Fig.(2) shows proposed Triple Logistic Map (TLM) based key sequence generation scheme. The sequence generated by all the three logistic map are uncorrelated in nature. So obtained three sequences are converted in to block of 8-bit sequence $\{K_{1i}\} = \{K_{11}, K_{12}, K_{13}\ldots.. K_{1t}\}$ $\{K_{2i}\} = \{K_{21}, K_{22}, K_{23}\ldots.. K_{2t}\}$ and $\{K_{3i}\} = \{K_{31}, K_{32}, K_{33}\ldots.. K_{3t}\}$ where 't' is size of the image using Eq (3). To obtain the proposed TLM based key sequence $\{K_i\}$ bit by bit XOR operation is performed between the three statistically independent key sequence of logistic map $\{K_{1i}\} \neq \{K_{2i}\} \neq \{K_{3i}\}$. The TLM based key sequence is obtained by

$$\{K_k\} = (\{K_{1k}\} \oplus \{K_{2k}\} \oplus \{K_{3k}\}) \tag{3}$$

The key sequence $\{K_k\}$ obtained from Eq(3) is highly uncorrelated and ranges between 0 to 255.

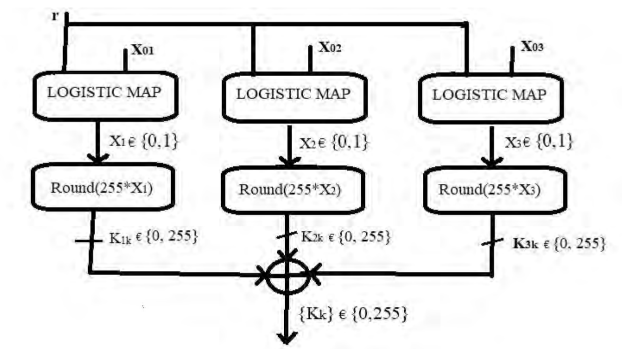


Fig.2:Proposed Triple Logistic Map (TLM) based key sequence generation scheme.

C. Proposed Image Encryption Scheme

Image encryption scheme using Triple Logistic Map (TLM) based key sequence is discussed. The steps are listed below.

Step1: First, 8 bit image pixels $I(x, y)\in(0, 255)$ of two dimension are converted into one dimension (1D) pixels $\{I_i\} = \{I_1,I_2,\ldots\ldots\ldots I_t\}$, Where total image pixels i.e t=M×N. M and N is row and column of two dimensional (2D) image pixels I(x, y) respectively.

Step2: The logical XOR operation is performed between image pixels $\{I_i\} = \{I_1,I_2,\ldots\ldots\ldots I_t\}$ and key sequence of TLM $\{K_i\}=\{K_1, K_2, K_3\ldots.. K_t\}\{K_i\}\in(0, 255)$. Hence encrypted image sequence $\{E_i\} = \{E_1,E_2,\ldots\ldots\ldots E_t\}$ are obtained. It is computed using equation (4)

$$\{E_i\} = [\{I_i\} \oplus \{K_i\}] \tag{4}$$

Step 3: The encrypted image pixels$\{E_i\} = \{E_1,E_2,\ldots\ldots E_t\}$ are converted into 2D array of size M×N. The resulting image is called encrypted image E(x, y).

D. Proposed Image Decryption Scheme

In this section steps used in the image decryption process are given.

Step1: The Encrypted image pixels E(x, y) are transformed into image pixels of 1D $\{E_i\} = \{E_1, E_2, \ldots\ldots E_t\}$.

Step2: The Decryption process is applied using the same TLM based key sequence $\{K_i\}$ used during encryption. It is computed using equation (5)

$$\{D_i\} = [\{E_i\} \oplus \{K_i\}] \tag{5}$$

Step3: The 1D decrypted imagepixels $\{D_i\}$are transformed into a 2D size of M×N. The resulting image is decrypted.

III. EXPERIMENTAL RESULTS

The performance analysis of the proposed image encryption scheme is studied using MATLAB-2017 software. For the simulation study, 8-bit gray images of 256 × 256 pixels in size are chosen as test images. In this paper, seven different images are chosen out of which six images are medical images and one image is Lena's image. The parameters such as i) histogram plot (for both original and encrypted image) ii)Mean Square Error(MSE) iii) Entropy iv) Correlation are considered to study the performance of the proposed scheme. The proposed single key sequence $\{K_i\}$ is generated from the sequence of three different logistic map $\{K_{1,i}\}\{K_{2,i}\}$ and $\{K_{3,i}\}$ wherer=3.99 and initial value X_{01}=0.1 X_{02}=0.2 X_{03}=0.3 are chosen.The proposed key generation scheme and image encryption and decryption process are given in Section-II. The proposed TLM based image encryption scheme is compared with image encryption using i) Logistic Map (LM) ii) Double Logistic Map (DLM).

A. Visual Analysis

To analyze the proposed TLM based encryption scheme, Lena's image, and six different medical images are considered and shown in Fig (3a) (1-7). Fig.(3a)(1-7) shows the original Lena image and different medical images. Fig.(3b)(1-7)are encrypted images using only the Logistic Map (LM)based scheme. Fig. (3c) (1-7)are encrypted images Double Logistic Map (DLM) based scheme. Fig. (3d)(1-7) shows encrypted images using $\{K_i\}$ of the proposed Triple Logistic Map (TLM) based scheme. It is observed that encrypted images will not reveal any information about the original image . However,the proposed TLM based scheme provides better performance compared to LM or DLM based scheme. To verify the sensitivity of the key in the case of proposed TLM based scheme, decryption process of first image Fig(3d)(1) a wrong key witha combination of X_{10}=0.1000000000000001, X_{20}=0.2000000000000001, and X_{30}=0.3000000000000001 with the same r=3.99 is used. After employing the wrong key the decrypted image is not the same as the original image. Hence even very small changes in the combination of the initial value of X_{10}, $X_{20,}$ and X_{30} it is highly impossible to decrypt the image. Hence it shows TLM based scheme highly sensitive to the initial value or key.

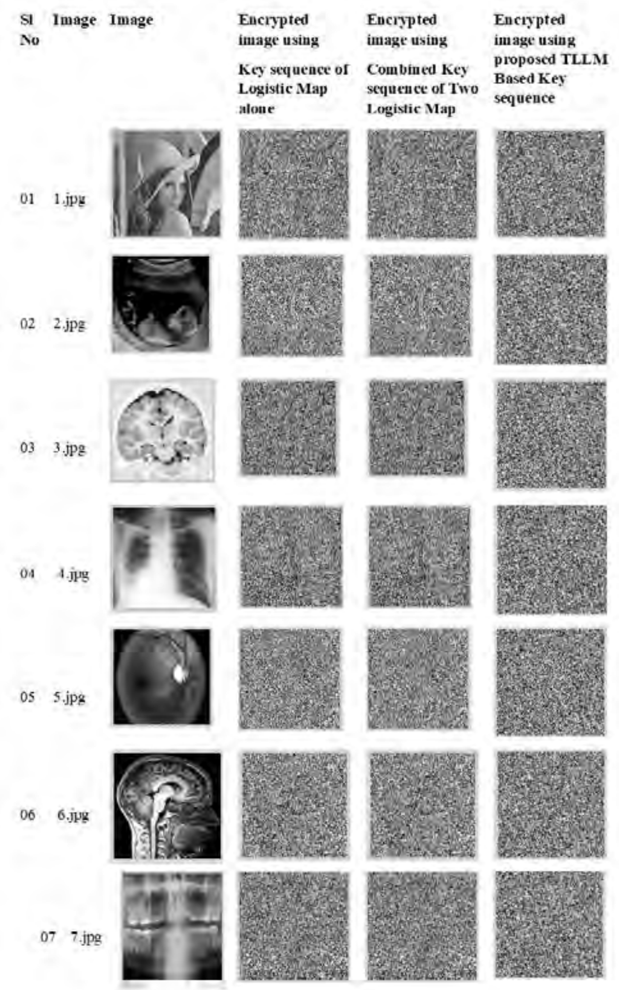


Fig.3a (1-7): Original Image. Fig. (3b)(1-7) Encrypted image X_0=0.1 r=3.99 Key Sequence of Logistic map alone Fig. (3c)(1-7)shows encrypted images with X_{10} and X_{20}=0.2 r=3.99 using the Key sequence $\{K_{2i}\}$ of Double Logistic Map (DLM) based scheme. Fig. (3d) (1-7) shows encrypted images with X_{10} =0.1, X_{20}=0.2 and X_{30}= 0.3 r=3.99 using the Key sequence $\{K_{3i}\}$ of Proposed Triple Logistic Map (TLM) based scheme.

B. Histogram Analysis

Histogram plot is a graphical representation of pixels values of an image. It is a plot of the Maximum pixel count with increasing gray level color. In this paper, histogram plot is used to study the performance of the image encryption scheme. Here histogram of the original image, encrypted image using i) Logistic Map ii) DLM based scheme, and iii) Proposed TLM based scheme are studied. Fig. 4(a)(1-7) shows a histogram plot of seven different images as shown in Fig(3a)(1-7). Similarly, Fig.4(b)(1-7). 4(c)(1-7) and 4(d)(1-7) shows a Histogram plot of the encrypted image using i) Logistic Map ii) DLM based scheme and iii) Proposed TLM based scheme respectively. From Fig.4(b)(1-7), (4c)(1-7), and (4d)(1-7), it is observed that the histogram plot of the encrypted image fairly uniform. However,the proposed TLM based scheme is flat and fairly uniform compared with encryption using an LM or DLM based scheme sequence.

Therefore the scheme proposed is highly immune from statistical attacks.

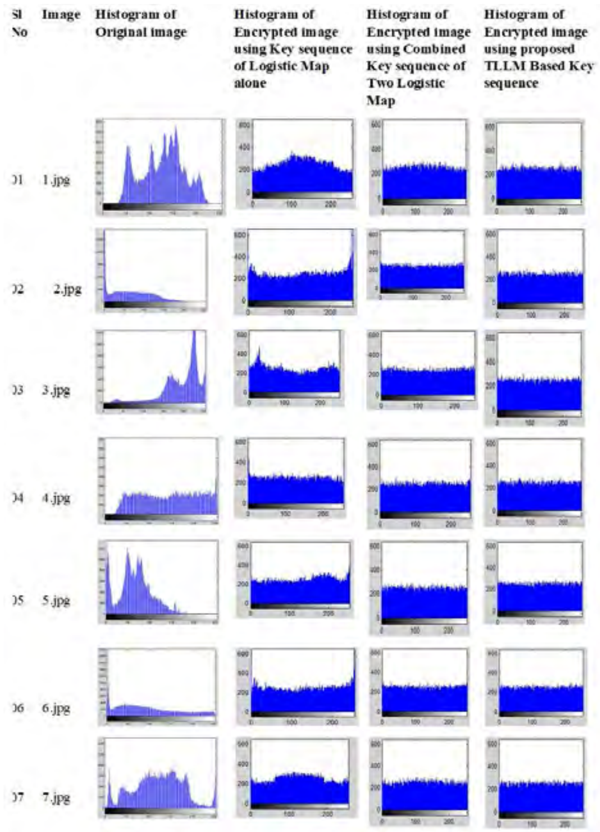


Fig. 4a: shows Histogram plot of Original image. Fig. 4b: shows Histogram plot of Encrypted image using Logistic map key sequence with X_{10} =0.1 and r=3.99 Fig. 4c: shows the Histogram plot of encrypted images with X_{10} =0.1 and X_{20}=0.2 r=3.99 using the Key sequence of Double Logistic Map (DLM) based scheme. Fig. 4d: shows Histogram plot of encrypted images with X_{10} =0.1, X_{20}=0.2 andX_{20}=0.3 r=3.99 using the Key sequence $\{K_i\}$ of Proposed Triple Logistic Map TLM based scheme.

C. Mean Square Error Analysis

The parameter Mean Square Error (MSE) can be defined as

$$\text{MSE}=\frac{1}{\text{MN}}\sum_{i=0}^{\text{M}\times\text{N-1}}[\text{I(i, j)-E(i, j)}]^2 \qquad (6)$$

Where I(i,j) is original image pixel value

E(i,j) is encrypted image pixel value,

M and N represents rows and columns of the image.

Ideally, the parameter MSE should be as high as possible in case of an image encryption scheme. It means a Higher value of MSE between original and encrypted image indicates,better immunity against statistical attacks. The MSE is computed between the images given in Fig.(3a)(1-7) and corresponding encrypted image using i) Logistic Map alone ii) DLM based scheme iii) Proposed TLM based scheme and tabulate in Table1. It is observed in Table1 that MSE value in case of proposed TLM based encryption scheme is better compared to a scheme using DLM based scheme or LM based Scheme. Similarly for all the medical image cases, MSE in the case of TLM based encryption scheme provides the best performance compared DLM or LM based scheme.

Table 1MSE VALUE BETWEEN ORIGINAL AND ENCRYPTED IMAGE

Sl no	Image	MSE (Logistic map)	MSE(Combination of two Logistic maps)	MSE(Proposed TLM based scheme)
01	1.jpg	7084.3	7300.2	7511.3
02	2.jpg	19989.00	13606.00	13562.00
03	3.jpg	14016.00	12617.00	12588.00
04	4.jpg	11683.00	10811.00	10929.00
05	5.jpg	12125.00	11144.00	11201.00
06	6.jpg	14447.00	13028.00	13031.00
07	7.jpg	9212.2	8920.9	9130.80

D. The entropy of the image

Information entropy[3][10] is used to calculate the degree of unpredictability of the content of information. Source C information entropy H(C), defined as

$$\mathrm{H(C)} = \sum_{i=0}^{\mathrm{M\times N-1}} \mathrm{P(C_i)}\log_2\frac{1}{\mathrm{P(C_i)}} \qquad (7)$$

Where $P(C_i)$ is the probability of symbol occurrence C_i
M×N is the total number of symbols.

Ideally, the maximum entropy value for an 8-bit grayscale picture is 8 bits/symbols that occur when equiprobable pixel values of image or histogram are flat. The entropy value of the image also indicates the distribution of image pixel value. It means a higher value of entropy indicates the uniform distribution of the image pixels.

Table 2Original and encrypted Entropy values of an image

Sl No	Image Type	The entropy of Original image	The entropy of encrypted image using (Logistic map	Entropy encrypted image using a Combination of two Logistic map	Entropy encrypted image using Proposed TLM based scheme
01	1.jpg	7.3837	7.9736	7.9955	7.9969
02	2.jpg	6.4531	7.9751	7.9972	7.9984
03	3.jpg	6.8764	7.9793	7.9966	7.9974
04	4.jpg	7.5701	7.9931	7.9966	7.9975
05	5.jpg	7.0459	7.9903	7.9975	7.9975
06	6.jpg	7.0666	7.9812	7.9967	7.9973
07	7.jpg	7.6101	7.9826	7.9960	7.9974

In Table 2 the entropy value of original and encrypted image using LM alone, DLM scheme, proposed TLM based scheme is computed and tabulated for the entire seven different images. For all the cases entropy of encrypted image is greater than the entropy of the original image. It is seen that entropy of Encrypted image using the proposed TLM-based scheme is high when compared with encrypted image entropyusing LM or DLM based scheme. It shows in the case of the proposed method, the distribution of pixels in encrypted image pixels is uniform compared to LM or DLM based scheme. Hence, the proposed TLM based scheme highly robust against statistical attacks.

E. Correlation analysis

Correlation coefficient [3][10] is as defined as

$$Correlation\ Coefficient = \frac{Cov(x,y)}{\sqrt{D(x)}\sqrt{D(y)}} \quad (8)$$

Where, Covariance of x, y is

$$Cov(x,y) = \frac{1}{L}\sum_{i=0}^{L}(x_i - E(x)(y_i - E(y) \quad (9)$$

Where, x and y, adjacent pixel values
L=M× N and variance of x_i is

$$D(x) = \frac{1}{L}\sum_{i=0}^{L}\left(x_i - E(x)\right)^2 \quad (10)$$

Mean of the value of x_i is

$$E(x) = \frac{1}{L}\sum_{i=0}^{L} x_i \quad (11)$$

Similarly, D(y) is the variance of y_i and E(y) is mean of y_i

Table-3 Correlation Coefficient

Sl No	Name of the image	Correlation (Logistic map)	Correlation (Combination of two Logistic maps)	Correlation(Proposed TLM based scheme)
01	1.jpg	-0.0537	0.0077	-0.0015
02	2.jpg	-0.0441	-0.0419	0.0015
03	3.jpg	-0.0347	0.00059247	0.00034
04	4.jpg	-0.0679	0.0051	-0.0052
05	5.jpg	-0.0419	0.00080854	0.0068
06	6.jpg	-0.0643	0.0042	0.00012034
07	7.jpg	-0.0725	0.0070	-0.0024

Correlation coeeficient is computed between pixels for the original and encrypted image using (i) Logistic map alone (ii) DLM based scheme and proposed TLM based scheme for seven different image cases. Ideally Correlation Coefficient between original and decrypted image '1' is with the same key. It indicates both original and decrypted images are the same. However Ideally Correlation coefficient between Original and encrypted image is '0'. In Table, It is observed that the correlation coefficient is very low in the proposed TLM based Scheme compared to DLM based encryption scheme and encryption using a Logistic map-based scheme. It indicates pixels are highly uncorrelated and robust against statistical attacks.

IV. CONCLUSION

In this work, Triple Logistic Map based image encryption and decryption scheme for the medical image, applications is proposed. The proposed scheme is highly sensitive to the chosen initial value and r. The proposed TLM-based scheme is compared with LM and DLM based image encryption scheme. The results indicate encrypted image will not reval any visual information about original image and highly un-correlated. Also, histogram plot of encrypted image using proposed encryption flat and fairly uniform compared to encryption using DLM or LM based scheme. Also Entropy, Correlation, Mean Square Error value is computed for proposed TLM and DLM, Logistic map based scheme. It is noted that the proposed image encryption scheme using TLM provides best results compared to encryption using DLM or LM based scheme.

V. REFERENCES

[1] William Stallings, "Cryptography and network security - principles and practice"(3. ed.). Prentice-Hall 2003, ISBN 978-0-13-111502-6, pp. I-XIV, 1-681

[2] Chengqing Li, Tao Xie , Qi Li " Cryptanalyzing image encryption using chaotic logistic map" Nonlinear Dynamics, 2014, Vol. 78, pp 1545–1551.

[3] C. E. Shannon, "A Mathematical Theory of Communication" The Bell System Technical Journal, Vol. 27, 1948, pp. 379–423.

[4] Safi, Haifaa W, Maghari, Ashraf Y. A. "Image Encryption Using Double Chaotic Logistic Map" Proceedings of IEEE International Conference on Promising Electronic Technologies (ICPET), 2017, pp. 66-70.

[5] ISMAIL, Samar M.,"Biomedical image encryption based on double-humped and fractional logistic maps",Proceedings of IEEE international conference on modern circuits and systems technologies (MOCAST). 2017. pp. 1-4.

[6] Hong, Lianxi Li, Chuanmu. "A Novel Color Image Encryption Approach Based on Multi-chaotic System", Proc. of IEEE 2nd International Conference on Anti-counterfeiting, Security and Identification, 2008, pp 1-5.

[7] Liu, Jin-mei Qu, Qiang, "Cryptanalysis of a Substitution-Diffusion Based Image Cipher Using Chaotic Standard and Logistic Map", Proc. of the Third International Symposium on Information Processing, October 2010,pp 67 - 69.

[8] Al-Najjar, Hazem Mohammad, Asem Mohammad AL-Najjar, and K. S. A. Arar. "Image encryption algorithm based on logistic map and pixel mapping table" Proc. of International Arab Conference on Information Technology, (ACIT 2011), pp. 56-60. 2011.

[9] Rohith S , Sujatha B K, "Image encryption and decryption using combined key sequence of Logistic map and Lozi map," Proc. of International Conference on Communications and Signal Processing (ICCSP), Tamilnadu, 2015, pp. 1053-1058.

[10] Shu Lin, Daniel J. Costello "Error Control Coding" 2nd Edition, Pearson Education,May 28th 2004India.

2018 Second International Conference on Advances in Electronics, Computer and Communications (ICAECC-2018)

Isolated Node Localization Probablity in Wireless Ad-Hoc Networks

Souparnika Jadhav
Electronics & Communication Engineering
Nagarjuna College of Engineering & Technology
Bengaluru, India
souparnika.j@gmail.com

Nagesh K. N
Electronics & Communication Engineering
Nagarjuna College of Engineering & Technology
Bengaluru, India
nageshlakmaya@gmail.com

***Abstract*— Node localization enabled wireless sensor network has gained immense popularity in real time due to its faster execution, high coverage area, performance and connectivity. It can achieve the primary objectives such as high connectivity and coverage area in wireless sensor networks. To completely exploit the wireless sensor networks, it is very essential to achieve a high connectivity between sensor nodes and gateways of sensor networks. Therefore, here, we have introduced a novel Local Isolated Node Localization in Wireless Sensor Networks based on probability analysis to locate the isolate node probability precisely, enhance connectivity between nodes and improve coverage area and scalability. A local collaborative pair method is introduced for isolated node recognition and ensures high connectivity. Experimental results verifies superiority of our proposed Isolated Node Localization model in terms of distance between sensor nodes, high connectivity and coverage area, identification of isolated sensor node location when networks are coinciding with each other at the boundaries and it also consider channel randomness while searching for connectivity between nodes.**

Keywords—wireless sensor networks; monitoring; connectivity; isolated nodes.

I. Introduction

In recent years, various technologies are highly emerged in the field of wireless communication. In those techniques, wireless ad hoc networks are highly appreciated by various researchers and organizations due to its flexibility, scalability, and wireless facility and centralized monitoring [1], [2], [3] and [13]. It is a typical network which contains numerous sensor nodes which can measure various physical quantities and transfer this essential information to concerned gateway in an effective mode [4].It consist of low processing power, low storage capacity and small passive energy. In wireless ad-hoc network, various smaller sensor nodes are present which can be deployed in different medium and large scale industries and factories to keep track of various nodes, for their observations and to control those nodes. For an instance, fault-identification, larger scale equipment monitoring and pipelining monitoring, toxic gas monitoring and leak point identification are the phenomenon's which can be controlled using WSN. WSN can be utilized for sensing, interaction between entities and for computation of components in various applications such as industrial applications, agricultural department, military applications, battlefield observation and transportation monitoring etc.

For wireless communication through WSN technique, all the sensor nodes must be setup in an organized network which can be driven by batteries with restricted energy. This setup of various smaller nodes in WSN need to maintained and controlled properly to get accurate results and high efficiency. However, if these sensor nodes are not monitored properly, they can give either faulty response or missing or partial information which can be very harmful and can give faulty results. Another essential issue is to maintain large connectivity between nodes and increase transmission range. Then, one way to avoid these inaccurate results in future interactions are to detect defected nodes and isolate those nodes which become defected. In WSN, to enhance transmission range and connectivity of the network the isolated nodes need to be identify which are unknown in the network and produces through some glitches occurs often. Moreover, isolated nodes are the nodes which are purely disconnected from the entire network and isolated group can be termed as the sensor group which is entirely disconnected from the gateway even if the sensor nodes inside that group are remains associated. Usually, node becomes isolated whenever it reaches to its dead state. These isolated nodes are unable to directly interact with the other nodes which are present near to it or the network which makes very challenging to detect their exact location and the localization problem occurs [5], [11], and [12]. The one way to identify the state (sleeping or transmitting) of all the isolated nodes in the network is to find distance between sensor nodes and sink.

However, there are numerous drawbacks which can occur in various conventional state-of-art techniques such as performance degradation due to isolated nodes, faulty results due to connectivity failure, high energy consumption in battery driven nodes and shorter transmission range etc. Therefore, to reduce these drawbacks a brief literature is presented by various researchers to locate isolated or dead nodes in the field of WSN.

In [6], a wireless sensor networks is adopted with the help of energy aware clustering technique to detect isolated nodes. This technique helps to enhance stability and

978-1-5386-3785-2/18/$31.00 ©2018 IEEE

connectivity. In [7], a life cycle of wireless sensor nodes for industrial environment is discussed. This technique helps to avoid collision and reduce energy consumption. In [8], a wireless sensor network is introduced to detect inferred and estimated energy of adjacent nodes based on the network congestion. In [9], a wireless sensor network adopted to identify the location of wireless isolated sensor nodes. In [10] and [14], a wireless sensor network is introduced to detect sensor faults and isolated nodes for wind turbine observations. This technique utilizes three-stage method for detection of faulty and isolated nodes such as SHORT fault, CONSTANT fault, and NOISE faults.

In above section, various techniques are introduced to located dead and isolated nodes which are a very challenging task due to connectivity loss between isolated nodes and neighboring nodes. However, these techniques cannot solve precisely locate the isolated nodes in a network and also very few techniques are there which tried to locate the isolated nodes considering boundary effect in the sensor network. Therefore, here, we have introduced a novel Isolated Node Localization in Wireless Sensor Networks based on probability analysis to locate the isolate node precisely, enhance connectivity between nodes and improve coverage area and scalability. This proposed technique help to reduce location and estimation problem [15] and [16] across wireless ad-hoc networks present in various conventional state-of-art-techniques

The contribution of work can be classified as follows:

In various conventional state-of-art techniques, node localization and node isolation are the biggest issues. Network or node isolation can be termed as the state they may occur when the power of few specific nodes get consumed at much quicker rate than usual and ultimately damage or break the association of nodes in the network. Thus, the global network connectivity gets affected. Therefore, in the proposed Isolated Node Localization model, some additional nodes are placed in the network to repair the damaged or isolated nodes and set up a new association with these isolated or damaged nodes. Moreover, our proposed technique also helps to recreate the existing link so that the distance between nodes and sink becomes much shorter. Therefore, the recovery of links becomes very fast.

Moreover, the proposed model adopts a Local linear prediction technique and this distributive technique helps for localization of isolated nodes. Here, we consider a metric to evaluate connectivity between nodes and coverage area, which can be termed as Node Isolation Probability. This probability helps to select random nodes which are not liked with any other nodes in the network so that isolated nodes can be determined easily. In our model, we consider the boundary effects as well. Whenever, the boundary of two networks coinciding with each other, then few nodes becomes isolated nodes which can be accessed by both the networks. Hence, the energy of that isolated nodes consume very fast and causes connectivity loss in both the networks. Therefore, these nodes need to identify in the network precisely so that connectivity between nodes can be rebuild. Therefore, here we have introduced node isolation probability to provide connectivity and coverage analysis considering various channel randomness and diversity techniques in wireless ad hoc network. We have conducted various experiments for the estimation and localization of isolate nodes considering boundary effects and results are demonstrated in graphical form for the localization and re-connectivity between nodes.

This paper is organized in following sections which are as follows. In section 2, we described our proposed methodology. In section 3, experimental results and evaluation shown and section 4 concludes our paper.

II. LOCALISOLATION NODE LOCALIZATION USING PROPOSED METHOD

In this section, we discuss about the proposed node localization method to identify the isolated nodes at the coinciding boundaries of two networks. The block diagram of proposed node localization method is demonstrated in fig1.

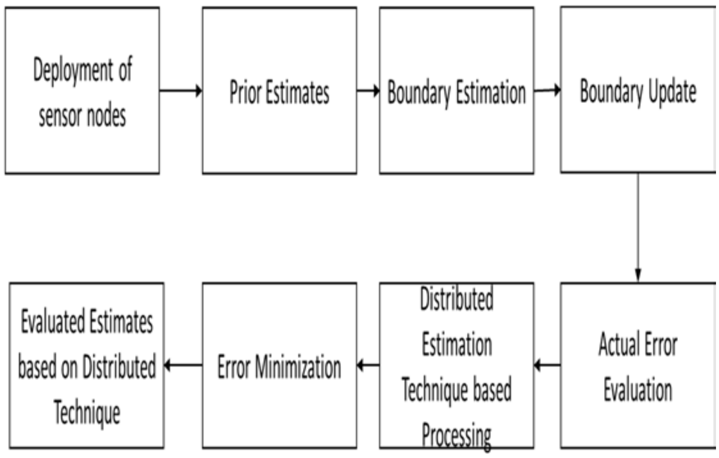


Fig1 Block Diagram of Proposed Local Isolated Node Localization technique.

Let that at the moment all the sensors know their positions well. First, the state of nodes are predicted and tracked. Here, we consider a network which consists of total number of j sensors with different states such as$\{m_1, \ldots \ldots, m_n\} \forall \, \mathbb{M} \cong \mathbb{K}^s$. Sensor states are the states which give the information about the position, alignment and operational factors of network and mainly tell about the sensor location. In the interlinked network of nodes, the sensors can be represented as $D = (W, F)$ whose vertices are $W = \{1, \ldots \ldots \ldots, j\}$ with respect to sensors and edges with respect to sensors are$|F| = i$. If two sensors $a \, and \, b$ can interact with each other, then the edges are$(a, b) \in F$. The sensor nodeswhich are linked with the sensor b can be defined asβ_b. Sensors are utilized to predict and identify the state of $g(c) \in \vartheta \cong \mathbb{K}^{s_g}$ of an isolated node where ϑ is a convex group. The isolated nodes can be defined using an isolated node motion model,

$$g(c+1) = E.g(c) + \gamma(c), \quad j(c) \sim \beta(0, Y), \qquad (1)$$

Where, process noise can be denoted as $j(c)$ and whose values can independent for two time pair. The measurement $k_b(c)$of sensor b whose location m_b and state of isolation nodes can be denoted as $g(c)$ at time c can be determined using sensor observation prototype as,

$$k_b(c) = T_b(m_b)g(c) + w_b(c, m_b), \quad (2)$$

Where,

$$w_b(c, m_b) = \beta\big(0, W_b(m_b)\big) \quad (3)$$

Where,$w_b(c, m_b)$can be defined as the measurement noise of sensor states which is independent of different time pairs of various sensors. The sensor noise $\gamma(c)$is independent of measurement noise. The measurement signal $k_b(c)$ provide partial information of only one sensor and cannot provide complete sensor state. This problem is identified as local identification error which can be encountered by every sensor. Let that all the sensor nodes are providing signals which are enough to identify the isolated sensors.

The positions of all the sensors must be known to utilize the signals $k_b(c)$for prediction. However, for larger sensor networks, only predicted rough idea of locations is enough. Assume that, the every sensor node can access the relative measurement error information of adjacent sensor node, which can helps to identify the location of specified sensors. Specifically, the sensor b receives sensor relative error measurement from its adjacent sensor a at time c,

$$h_{ab}(c) = m_a - m_b + \delta_{ab}(c), \quad (4)$$

$$\delta_{ab}(c) = \beta(0, \delta_{ab}) \quad (5)$$

Where, $\delta_{ab}(c)$ can be defined as the measurement error, which is independent of sensor pair and time pair. These measurement errors do not depend upon motion errors and sensor measurement of the required sensor. Here, assume that, every sensor decides to localize them in sensor 1 reference frame. While using the isolate node prediction method, then the prediction of locations at the place of positions of unknown sensor nodes. Each sensor b provides two types of predicted constructors. The two constructors are $\widehat{m}_b(c)$ for it own position and $\hat{g}_b(c)$ for its adjacent isolated node position in a distributed style. Using the information from its adjacent nodes and sensor measurements, which is defined as the collaborative prediction and localization error?

$$\{h_{ab}(c) | a \in \beta_b\}, \quad \{k_b(c)\} \quad (6)$$

Here, we try to find out the exact location and position of isolated mobile nodes with the help of sensor networks using range measurements. The predicted position of sensor node a can be denoted as $g_a^1, g_b^2 \in \mathbb{K}^2$ and predicted velocity of sensor node a can be denoted as $\dot{g}_a^1, \dot{g}_b^2 \in \mathbb{K}^2$ which can be driven with the help of Gaussian noise,

$$g_a(c+1) = \begin{bmatrix} \mathbb{I}_2 & \mathbb{T}\mathbb{I}_2 \\ 0 & \mathbb{I}_2 \end{bmatrix} g_a(c) + \gamma_a(t), \quad (7)$$

$$Y := Q \begin{bmatrix} \mathbb{T}^3.(3)^{-1}\mathbb{I}_2 & \mathbb{T}^2.(2)^{-1}\mathbb{I}_2 \\ \mathbb{T}^2.(2)^{-1}\mathbb{I}_2 & \mathbb{T}\mathbb{I}_2 \end{bmatrix} \quad (8)$$

Where, the sensor states of a^{th}sensor are $[g_a^1, g_b^2, \dot{g}_a^1, \dot{g}_b^2]^P$ and the sampling time can be denoted as $\mathbb{T}$ in seconds whereas the strength of diffusion can be denoted as Qin$m.(sec^2)^{-1}.Hz^{-1}$. Every sensor present in the network uses error measurements of every sensor position,

$$k_{ab}(c) = \begin{bmatrix} \sqrt{(g_a^1 - m_b^1)^2 + (g_a^2 - m_b^2)^2} \\ \text{atan}\big((g_a^2 - m_b^2).(g_a^1 - m_b^1)^{-1}\big) \end{bmatrix} + w(c, m_b, g_a). \quad (9)$$

Where, $m_b^1, m_b^2 \in \mathbb{K}^2$ represents the location of sensor and w represents error which depends on the distance between estimated isolated node and sensors. In this way, we estimated the isolated nodes in the network.

III. PERFORMANCE EVALUATION

The exploitation of wireless sensor networks is enormously enhanced in recent years due to extensive use of their precise sensor information to track different objects, nuclear reactor control, and fire identification and traffic monitoring etc. without any wiring equipment. Process Supervision, Healthcare monitoring, Environmental sensing, Observation of the environment pollution, Forest fire Identification, Landslide detection and industrial monitoring are the various application of WSN. However, WSN consist of some major drawbacks such as limited energy due to batter-constraint sensor nodes, connectivity lost due to isolate or damaged nodes, low speed interaction between the gateway and sensor nodes and comparatively little costly which may affect the performance of the WSN. Therefore, these drawbacks must be sorted out soon in order to carry out high performance in WSN in real time and there is need of finding the exact location of isolated nodes when boundaries of two different networks coincide with each other so that the connectivity between network and sensor nodes remain maintained and to reconfigure the network or to establish new network. Therefore, here we have introduced a novel Isolated Node Localization in Wireless Sensor Networks based on probability analysis to locate the isolate node probability precisely, enhance connectivity between nodes and improve coverage area and scalability. Node Isolation Probability helps to select random nodes which are not liked with any other nodes in the network so that isolated nodes can be determined easily. In proposed model, different networks are utilized for various number of sensor nodes to identify the isolate nodes in the network so the connectivity remains maintained. In the following sections connectivity range, connection between sensor nodes and isolated nodes are represented in the graphical form. The proposed Local Isolated Node Localization model simulated on 64-bit windows 10 OS with 16 GB RAM which contains an INTEL (R) core (TM) i5-4460 processor. It contains 3.20 GHz CPU. This project is simulated using $MATLAB\ 16B$ software development environment and programming language.

A. *Comparative Study*

In recent years, Wireless Sensor Networks (WSN) has received high praise from all over the world in different fields like healthcare solutions, industrial scenarios, wildlife monitoring, Landslide detection and medical applications etc. Therefore, to maintain high speed performance becomes mandatory for WSN. Furthermore, to enhance the efficiency of WSN, to identify the localization and exact position of isolated nodes becomes a high priority. However, high energy consumption and inappropriate isolated node localization techniques cause high degradation in performance of wireless sensor networks. Subsequently, to ensure proper localization technique and to reduce Estimation and Localization problem and ensure high connectivity between sensor nodes and network considering boundary conditions, we have introduced a novel Local Isolated Node Localization in Wireless Sensor Networks based on probability analysis to locate the isolate node precisely, enhance connectivity between nodes and improve scalability and range of the network. Our proposed Local Isolated Node Localization helps to boost throughput, connectivity and performance of the wireless sensor network devices. Here, we have conducted various experiments considering different networks and various sensor nodes to ensure high connectivity and identifying the precise location of sensor nodes.

Table 1

Error results considering different nodes using Local technique		
No.of nodes considered	**Actual signal error of network**	**Signal error using Local technique**
10	16.6050	0.0027
20	15.8510	0.0062
30	14.8079	0.0068
40	14.6999	0.0167
50	14.9893	0.0071

B. Graphical Representation

This section provides graphical representation of our simulated experiments considering various sensor nodes such as 10, 30, 50 and 100 etc. in terms of distance between sensor nodes, high connectivity and identification of isolated sensor node location when networks are coinciding with each other at the boundaries. Here, figure 2 shows connectivity results between sensor nodes and the network using proposed Isolated Node Localization model. In figure 2(a), the sensor node true estimates are presented considering 30 nodes. In figure 2(b), Gaussian error based sensor node estimates are presented considering 30 nodes and likewise, in figure 2(b), the proposed Local Isolated Node Localization based sensor node estimates are presented considering 30 nodes. From the results it is demonstrated that the signal error is highly reduced using our proposed Local Isolated Node Localization technique which is just 0.0071 compare to the actual error which is 14.9893.Here, figure 3 shows connectivity results between sensor nodes and the network using proposed Isolated Node Localization model considering 40 nodes. Here, figure 4 shows connectivity results between sensor nodes and the network using proposed Isolated Node Localization model considering 50 nodes.

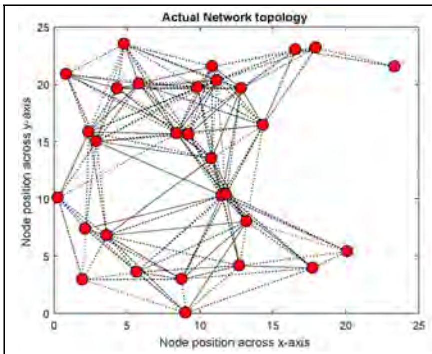


Fig. 2 (a)

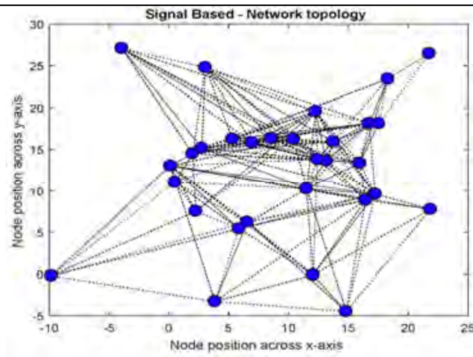


Fig. 2 (b)

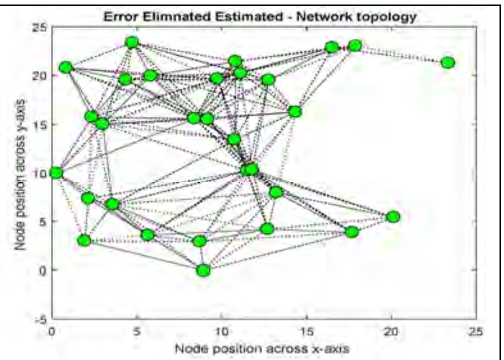


Fig. 2 (c)

Fig.2. Sensor node estimates for a sensor considering 30 nodes (a) true estimates of the sensor network (b) signal error estimates of the sensor network (c) proposed Local technique estimates for the sensor network

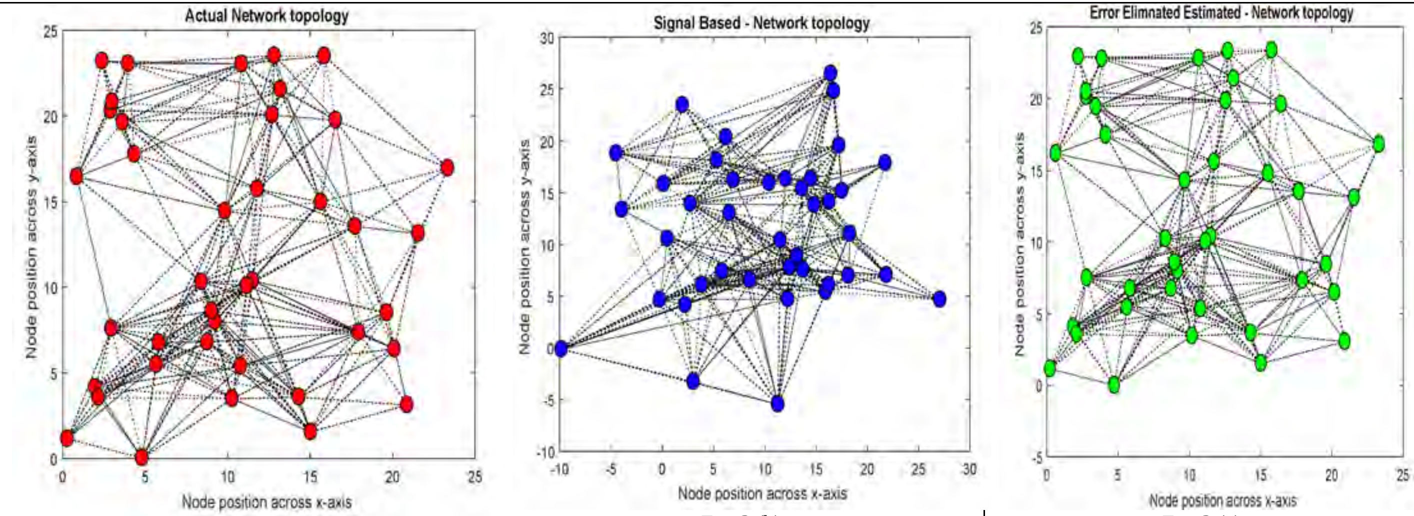


Fig. 3 (a) Fig. 3 (b) Fig. 3 (c)

Fig.3. Sensor node estimates for a sensor considering 40 nodes (a) true estimates of the sensor network (b) signal error estimates of the sensor network (c) proposed Local technique estimates for the sensor network

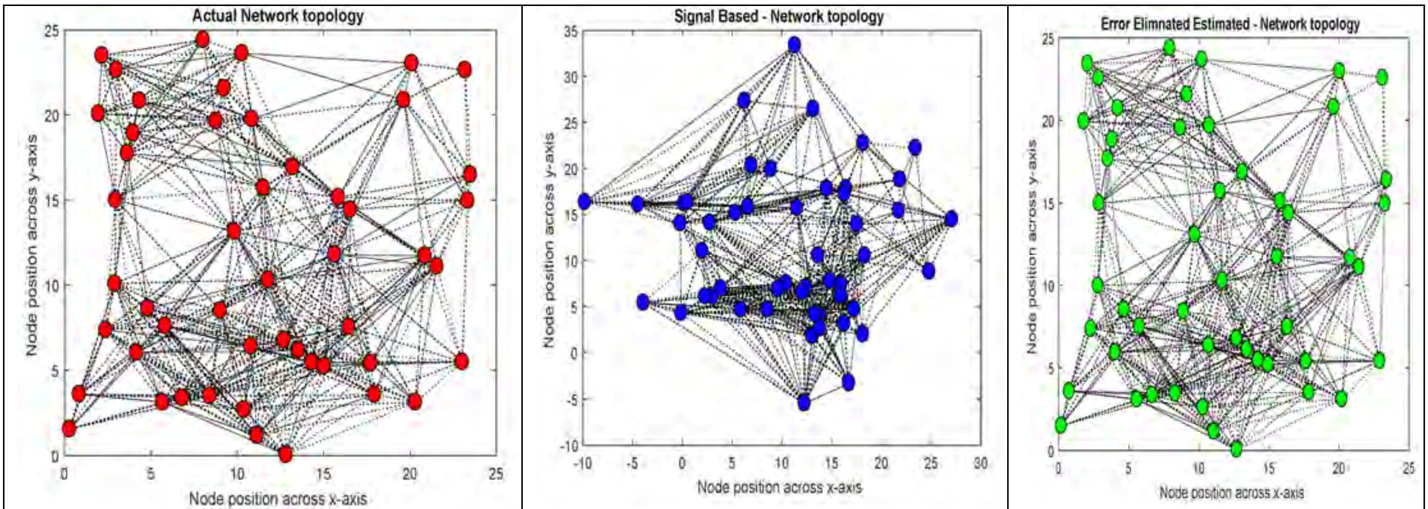


Fig. 4 (a) Fig. 4 (b) Fig. 4 (c)

Fig.4. sensor node estimates for a sensor considering 40 nodes (a) true estimates of the sensor network (b) signal error estimates of the sensor network (c) proposed Local technique estimates for the sensor network

IV. CONCLUSION

The importance of controlling high power consumption, high coverage area and achieving high connectivity between sensor nodes and sensor network and effective sensor information is very essential in wireless ad-hoc networks. Therefore, here we have introduced a novel Local Isolated Node Localization in Wireless Sensor Networks based on probability analysis to locate the isolate node precisely, enhance connectivity between nodes and improve coverage area and scalability and hence, enhance the efficiency of the system. An isolated node localization technique is adopted to encounter a collaborative Estimation and Localization problem introduced in sensor environment. A detailed modeling is presented to solve Localization problem using a Local algorithm. A Local collaborative pair method is introduced for isolated node recognition and ensures high connectivity between sensor nodes and gateway of sensor network when boundary conditions are considered. Therefore, here we have introduced node isolation probability to provide connectivity and coverage analysis considering various channel randomness and diversity techniques in wireless ad hoc network. Experimental outcomes are presented in terms of distance between sensor nodes, high connectivity and identification of isolated sensor node location when networks are coinciding with each other at the boundaries. Signal Error is highly reduced using our proposed Local Isolated Node Localization technique from 14.9893 to 0.0071. Our proposed Isolated Node Localization model ensure high connectivity and isolated node localization precisely which is much superior to any other state-of-art-techniques.

REFERENCES

[1] H. Zhang, L. Shu, J. Rodrigues, and H. Chao, “Solving network isolation problem in duty-cycled wireless sensor networks,” in Proc. MobiSys, 2013, pp. 543–544.

[2] S. M. Chung, K. T. Kim, J. Song and H. Y. Youn, "Enhancing node connectivity by utilizing RSSI for ZigBee-based WSN," 2016 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, 2016, pp. 555-560.

[3] P. Kar, A. Roy, S. Misra and M. S. Obaidat, "Energy-efficient connectivity re-establishment in WSN in the presence of dumb nodes," 2015 IEEE International Conference on Communication Workshop (ICCW), London, 2015, pp. 1485-1490.

[4] S. Rajasegarar, C. Leckie and M. Palaniswami, “Anomaly Detection in Wireless Sensor Networks,” IEEE Wireless Communications, Vol. 15, No. 4, 2008, pp. 34-40.

[5] K. Premkumar, A. Kumar and J. Kuri, "Local detection and localization of events in large ad hoc wireless sensor networks," 2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, 2009, pp. 178-185.

[6] T. H. Chiang and J. S. Leu, "Regional energy aware clustering with isolated nodes in Wireless Sensor Networks," 2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC), Washington DC, 2014, pp. 1829-1833.

[7] H. P. Bernhard, A. Springer, A. Berger and P. Priller, "Life cycle of wireless sensor nodes in industrial environments," 2017 IEEE 13th International Workshop on Factory Communication Systems (WFCS), Trondheim, 2017, pp. 1-9.

[8] J. M. Chandramouli et al., "Using network traffic to infer compromised neighbors in wireless sensor nodes," 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, 2017, pp. 1022-1023.

[9] A. Nassiri, M. A. Razzaque and A.H.Abdullah, "Isolated Adversary Zone for source location privacy in Wireless Sensor Networks," 2016 International Wireless Communications and Mobile Computing Conference (IWCMC), Paphos, 2016, pp. 108-113.

[10] Y. Peng, W. Qiao, L. Qu and J. Wang, "Sensor fault detection and isolation for a wireless sensor network-based remote wind turbine condition monitoring system," 2017 IEEE Industry Applications Society Annual Meeting, Cincinnati, OH, USA, 2017, pp. 1-7.

[11] L. Chen "Why (n + 1)th-hop neighbours are more important than nth-hop ones for localisation in multi-hop WSNs," in Electronics Letters, vol. 50, no. 22, pp. 1646-1648, 10 23 2014.

[12] A. M. Abbas, "TrustLP: A trust-based localization protocol for wireless sensor networks," 2015 13th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Mumbai, 2015, pp. 529-536.

[13] P. Corbalán, R. Marfievici, V. Cionca, D. O'Shea and D. Pesch, "Into the SMOG: The Stepping Stone to Centralized WSN Control," 2016 IEEE 13th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Brasilia, 2016, pp. 118-126.

[14] M. de Oliveira Barros and A. Weber, "System-level diagnosis for WSN: A heuristic," 2016 17th Latin-American Test Symposium (LATS), Foz do Iguacu, 2016, pp. 171-176.

[15] X. Gao, "A Novel Approximation for Multi-Hop Connected Clustering Problem in Wireless Networks," in IEEE/ACM Transactions on Networking, vol. 25, no. 4, pp. 2223-2234, Aug. 2017.

[16] L. Shu, L. Wang, J. Niu, C. Zhu and M. Mukherjee, "Releasing Network Isolation Problem in Group-Based Industrial Wireless Sensor Networks," in IEEE Systems Journal, vol. 11, no. 3, pp. 1340-1350

IEM Institute of Engineering & Management
Good Education, Good Jobs

SMART

IET The Institution of Engineering and Technology
Kolkata Local Network

CERTIFICATE

OF APPRECIATION

PROUDLY PRESENTED TO

Mallanna S D, Sreenivasa T V, Dr.Nagesh Kallollu Narayaswamy and Dr. Ajay Kumar Dwivedi

in recognition and appreciation for presenting the best paper in **Track 1**

Paper ID : 19
Paper Title : The Design and Analysis of Four Port MIMO Antenna for n77 and MTI Wireless Edge 5G Applications

Date : November 26th - 28th, 2021

PROF. (DR.) SATYAJIT CHAKRABARTI DIRECTOR, IEM KOLKATA	PROF. ANKIT RAY GHATAK CONVENOR IEMPOWER 2021	PROF. (DR.) SENJUTI KHANRA CONVENOR IEMPOWER 2021	PROF. (DR.) SUDHIR CHANDRA MURMU CONVENOR IEMPOWER 2021

ANALYSIS OF JUNCTION AND TRAFFIC DATA FOR THE IMPROVEMENT OF THE YELAHANKA POLICE STATION CIRCLE

MANDEEP B V

ASSISTANT PROFESSOR, NAGARJUNA COLLEGE OF ENGINEERING AND TECHNOLOGY, BANGALORE

E MAIL: mandeepgowda@gmail.com

ABSTRACT: The city growing radially with new towns being constructed has led to high level of traffic which has put the road network under great strain. The government and the concerned agencies such as the Bangalore Development Authority (BDA) have the task of improving the transportation infrastructure to meet the growth of traffic needs. Many schemes of grade separation facilities such as underpass and flyovers at important intersections, traffic signal systems etc. are under implementation.Yelehanka police station junction is one of the major junction. This junction serves as connection for Airport, Hebbal, Doddaballapur and Yelahanka. For Yelahanka and Doddaballapur road users, this is the only major junction which departs towards Airport and Hebbal (other than Judicial layout approach road). Hence, there is a huge flow of traffic inward and outward the junction which needs to be studied and analysed. In this paper the studies are conducted, traffic volume studies Analysis of signals at the junction, Analysis of Traffic volume flow, and spot speed study are conducted at yalahanka police station circle all the data should be entered in different tabular column and graphs should be plotted Considering the route wise traffic volume for the taken 5 days, a flyover can be proposed from NES towards Jakkur and vice-versa as there is no conflict point in this route compared to other routes. A service road should be provided on both sides of the flyover for easy movement of vehicles travelling from NES to Yelahanka Old Town and those travelling from Yelahanka Old Town to NES below the flyover, reducing the congestion rate at the junction.

Key words: traffic volume, route wise traffic, signals, dimensions of road.

Introduction

"Traffic engineering is the science of measuring traffic and travel, the study of the basic laws relating to traffic flow and generation and application of this knowledge to the professional practice of planning, designing and operating traffic systems to achieve safe and efficient movement of people and goods."

Passenger Car Units (PCUs)

Different classes of vehicles such as vans cars trucks buses auto rickshaw, pedal cycles motor cycles, etc. are found to use the common Roadway facilities without segregation The roads on most of the developing countries like India . The flow of the Traffic with unrestricted mixing of different classes on the roadways forms the mixed traffic flow or heterogeneous traffic flow. It is rather difficult to estimate the traffic volume and capacity of road way Facilities under mixed traffic flow, unless the different vehicle classes are converted in to one common practice to consider the passenger car as the standard vehicle unit is to convert the other vehicle classes and unit is called as passenger car unit or PCU. It is represented as PCU per hour or PCU/lane/hr.

PCU values suggested by the IRC

Sl. No.	VEHICLE CLASSES	EQUIVALENCY FACTORS
1	Passenger cars, Tempo	1.0
2	Bus truck	0.5
3	Auto rickshaw, agriculture tractor	3.0
4	Motor cycle pedal cycle and scooter	1.5
5	Cycle rickshaw	4.0
6	Horse drawn vehicle small bullock cart and hand cart	6.0
7	Large bullock cart	8.0

OBJECTIVES

To conduct traffic volume studies,Analysis of signals at the junction,Analysis of traffic volume flow, To conduct spot speed study,To give out proper conclusion and recommendations for junction improvement

LITERATURES

Satish C.Sharma,et.al (1996), " state wide traffic volume studies and precision of AADT estimates" the statistical precision of annual average daily traffic(AADT) estimates resulting from short period traffic counts(SPTC). The large number of automatically traffic recorder(ATR) sites located on Minnesota's highway system of the study. The study result suggest that highway agency should put more Emphasis on sample site assignment to correct ATR Groups than on durations of The count.

A.K.M Abir and Md. Sami Hasnaine (1995), "traffic volume study" In this study traffic volume with volume, speed and density and the analysis is carried out through primary traffic flow survey AUST- flyover junction to the shatrasta junction in Dhaka city. The traffic flow was studied by method of the following conclusion were drawn:

- Old buses should be replaced by new buses in order to manoeuvre easily.
- Separate lane system should be introduced for bicycles.
- Large container truck should be allowed to use this junction only at the off peak hours

Rajiv Ganguly, Ashok Kumar Gupta, et.al (2014) "Traffic volume and Accident studies on national highway- 22 between solan and shimla India."

The basic study difference Between flexile and rigid pavement based on the LCCA. Recycling durability local performance and safety features. an observation of traffic count was conducted at Tuttikandi in Shimla and Chambaghat in Solan and difference in the vehicular flow was completed. The process involved making manual count of vehicles passing by on the week days and weekends. vehicle fleet was characterized as cars buses trucks and motor cycles. The study involved in identifying the black spots which refers to those stretches on Shimla-Solan highway NH-22 reported with the most number of accidents also.

This study results gives suitable remedial measures to prevent such accidents in the future including installing various roads furniture like convex mirror light poles fluorescent sign bords.

Mr. UditBatra, Mr. Mandar V. Sarode (2013*)*, "Traffic Surveying and Analysis" In this survey, traffic survey in the specific road sections on Nagpur city has been carried out which included calculation of the present traffic density and comparison with previous year data average velocity of the traffic.Manual method of counting was used with the help of video recording The following conclusions were drawn-

i. Traffic distribution through alternative routes or construction of flyover may be envisaged.
ii. Public transport system needs to be the strengthened thereby reducing traffic density. Parking of vehicles should be prohibited on streets.

DIFFERENT STUDIES

Traffic Volume Studies

Traffic volume is the number of vehicles crossing in a section of road per unit time at any selected period of time. Traffic volume used as quantity measure of flow commonly used units are vehicles per day and vehicles per hour The objects and uses of traffic volume studies are

- Traffic volume is Generally accepted as the true measure of relative importance of the roads and in deciding the priority for improvement and expansion
- Traffic volume study is used in traffic operation planning and control of the existing facilities.
- The study is used in the analysis of the traffic patterns and trends.
- The classified volume study is useful in the structural design of the pavements In geometric design and computing roadway capacity.
- Volume distribution study is used in the planning one way streets and other regulatory measures.
- Turning movement study is used in the design of intersections in planning signal timings channelization and other control devices.
- Pedestrian traffic volume study is used for planning sidewalks cross walks subways and the pedestrian signals.

There are variations in traffic flow from time to time. In classified volume study, the traffic is classified and the volume of each class of the Traffic via, Buses, Truck, Passenger cars, other light vehicles, rickshaws, tonnages, cycles and pedestrians is found separately, The direction of each class of traffic flow is noted

DAY :	UNIT : PCU/hour				TIME :	
ROUTES	TWO WHEELER	CAR	BUS	TRUCK	AUTO	TOTAL VOLUME
YOT to Jakkur						
YOT to NES						
Jakkur to NES						
Jakkur to YOT						
NES to YOT						
NES to Jakkur						
					GRAND TOTAL	

1.Traffic Volume Study Data According to Time and Day

Speed Studies

Speed studies carried out occasionally give the general trend in speeds. There are two types of speed studies carried out,

1. Spot speed study
2. Speed and delay study.

Spot Speed Study

Spot speed study may be useful in any of the following aspects of traffic engineering:

- To use in planning traffic control and in traffic regulations
- To use in geometric design- for redesigning existing roads or for deciding design speed for new facilities.
- To use in accident studies
- To decide the speed trends
- To study the traffic capacity

The spot speeds are effected by physical features of the road like pavement width, curve, sight distance, gradient, pavement uneven intersection and road side developments. Other factors affecting spot speeds are environment conditions (like weather, visibility), enforcement, traffic conditions, driver, vehicle and motto of travel.

Traffic signals

At intersection where there is large number of the crossing and Right turn traffic there is possibility of the several accidents as there cannot be orderly movements, The earlier practice has been to control traffic police by showing stop signs alternately at the cross roads so that one of the traffic streams may be allowed to move while the cross traffic is stopped.

EXPERIMENTAL INVESTIGATION

Yelehanka Police Station Junction is the studied area. The area caters to considerable local and through amount of traffic especially at peak hours. This junction acts as a link between Yelehanka New Town / Doddaballapur and Yelahanka Old Town / Airport and Jakkur / Bengaluru city.

Manual method of counting has been used to count the traffic volume at the junction with the help of video recordings. The data were taken for 5 days (Monday, Tuesday, Wednesday, Saturday and Sunday) from 7-10 am, 1-2 pm and 4-6 pm.

Spot speed study was conducted on one working day for 4 routes, i.e., NES to Junction, Junction to NES, Jakkur to Junction and Junction to Jakkur. A distance of 150 m was fixed for all the routes. 10 four wheelers, 10 two wheelers and 10 buses were considered for all the routes. Then, time taken by different vehicles to cover the distance were recorded and speed were calculated.

METHODOLOGY

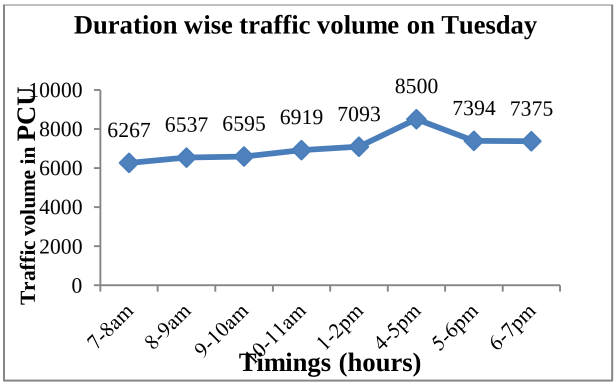
Analysis of Traffic Volume Flow

In the Civil Engineering traffic flow is study of the interactions between the travellers (pedestrians, cyclists, drivers and their vehicles also) and infrastructures (highways, signals and traffic control devices etc) with the aim of understanding and developing an optimal transport network with efficient movement of the traffic and minimal traffic congestion problems.

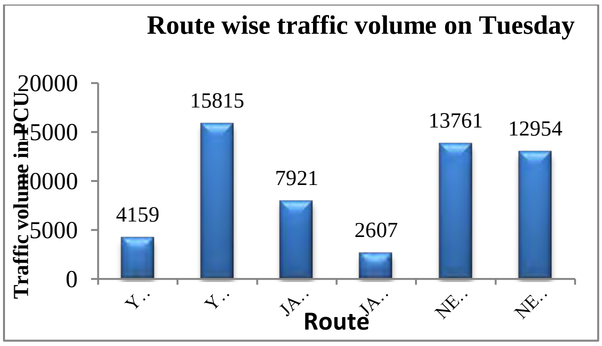
Analysis of traffic volume flow deals with segregating the traffic volume with respect to different routes, which will help us to conclude the routes in which maximum or minimum traffic flow is moving. Upon this, we can decide the improvement of junction for the exact inconvenience that are faced by the road users.

2.Traffic Volume Study Data from 8 am to 9 am on Tuesday

UNIT : PCU/hour TIME : 8:00 to 9:00 am DAY : TUESDAY							
ROUTES	TWO WHEELER	CAR	BUS	SCHOOL /CLG BUS	TRUCK	AUTO	TOTAL VOLUME
YOT to Jakkur	155	76	36	6	45	34	352
YOT to NES	263	544	330	306	153	78	1674
Jakkur to NES	199	392	264	78	81	30	1044
Jakkur to YOT	66	62	60	6	27	24	245
NES to YOT	609	482	324	114	89	114	1732
NES to Jakkur	371	736	264	27	63	30	1491
						GRAND TOTAL	6537



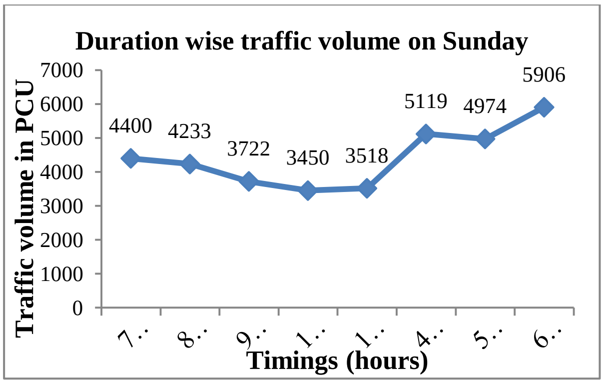
Duration wise Traffic Volume on Tuesday



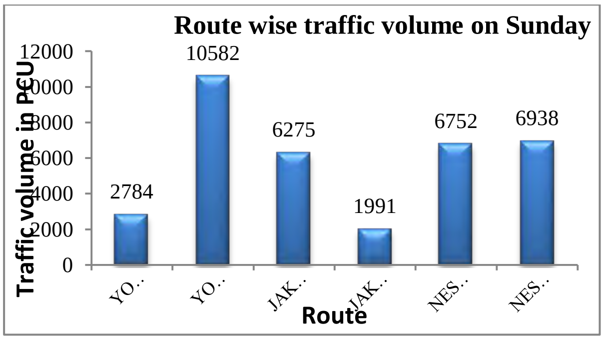
Route wise traffic volume on Tuesday

3.Traffic Volume Study Data from 8 am to 9 am on Sunday

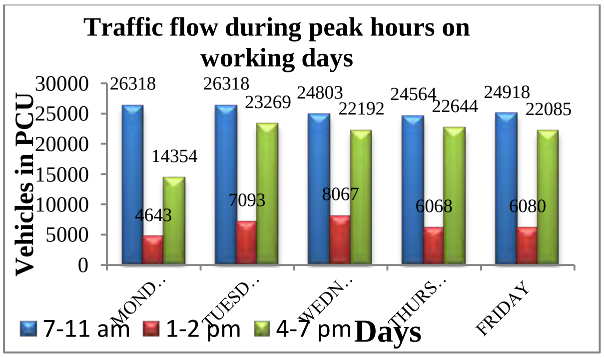
UNIT : PCU/hour TIME : 8:00 to 9:00 am DAY SUNDAY							
ROUTES	TWO WHEELER	CAR	BUS	SCHOOL /CLG BUS	TRUCK	AUTO	TOTAL VOLUME
YOT to Jakkur	75	42	0	0	90	36	243
YOT to NES	281	454	288	0	252	86	1361
Jakkur to NES	116	284	138	0	90	38	666
Jakkur to YOT	66	48	48	0	36	24	222
NES to YOT	188	294	192	0	225	72	971
NES to Jakkur	125	242	246	0	99	58	770
						GRAND TOTAL	4233



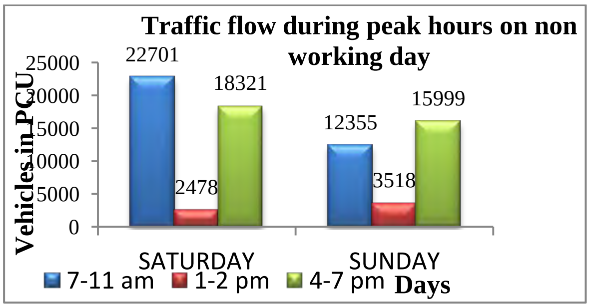
Duration wise Traffic Volume on Sunday



Route wise Traffic Volume on Sunday



Traffic flow during peak hours on working days



Traffic flow during peak hours on non - working day

Spot speed Study

Speed is the an important transportation consideration related to safety. time, comfort, convenience and economy. Spot speed studies are used to determine the speed distribution of a traffic stream at a specific location. The data gathered in Spot Speed Studies are used to determine vehicle speed percentiles, which are useful in the making speed related decisions.

To conduct spot speed study, we have selected one working day, 4 routes, i.e., NES to Junction, Junction to NES, Jakkur to Junction and Junction to Jakkur, were considered. A distance of 150 m was taken on all the routes and the time taken by different vehicles for covering the distance were recorded and speed were calculated.

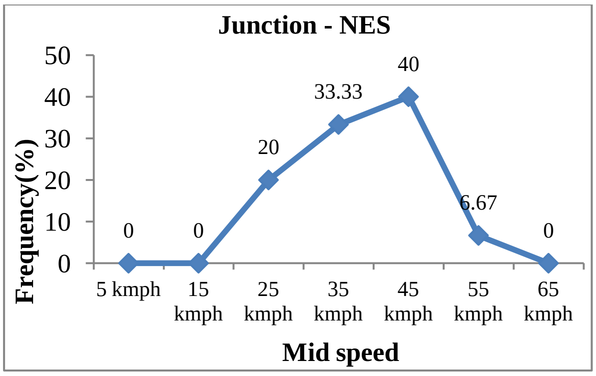
Further, frequency percentile and cumulative frequency percentile were presented in a tabular format and graphs were plotted, accordingly

4.Spot Speed Study of Stretch between NES to Jakkur

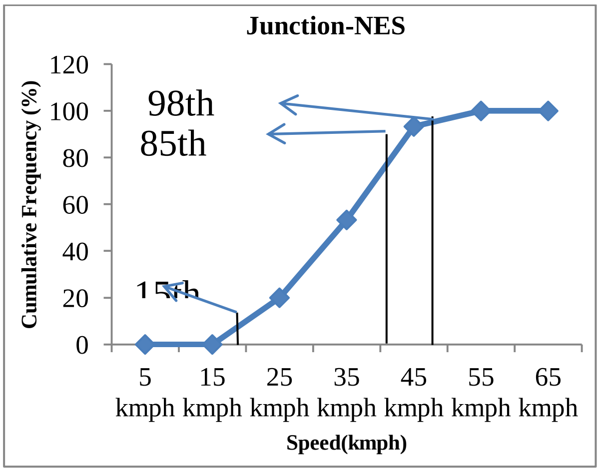
Spot speed data			Route: Junction - NES	
Vehicle	**Sl. No**	**Time (T) "sec"**	**Distance (D) "m"**	**Speed (kmph) (D/T)x3.6**
Car	1	18	150	30
	2	10		54
	3	15		36
	4	13		41.54
Motor cycle	1	13	150	41.54
	2	10		54
	3	11		49.1
	4	14		38.57
Bus and Trucks	1	20	150	27
	2	26		20.76
	3	19		28.42
	4	21		25.71

5.Spot Speed Data from Junction towards NES

ROUTE : Junction to NES			Working day	
Speed Range	**Mid Speed**	**No of Vehicles**	**Frequency**	**Cumulative frequency**
0-10	5	0	0	0
10-20	15	0	0	0
20-30	25	6	20	20
30-40	35	10	33.33	53.33
40-50	45	12	40	93.33
50-60	55	2	6.67	100
60-70	65	0	0	
Total		**30**		



Frequency Wise Spot Speed Analysis from Junction towards NES



Cumulative Frequency Wise Spot Speed Analysis from Junction towards NES

Analysis of Signals at the Junction

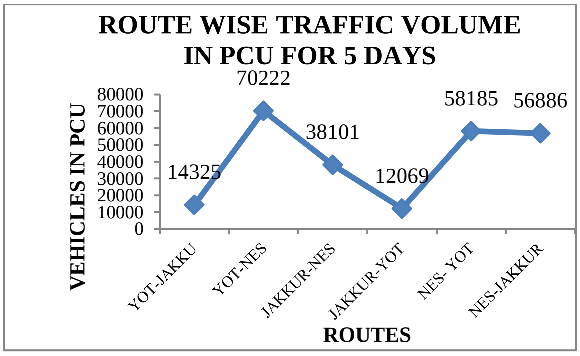
prior to initiating the design for the installation, modification, or removal of a traffic signal it is necessary to conduct an engineering study to establish need. Analysis of signal is very important to understand the signal timings and to check whether the given timings are sufficient for the present traffic volume flow.

The signal timings with respect to different routes are recorded and tabulated

Signal analysis at the junction			
Route	**Signal Timing**		
	Red Light	**Green Light**	**Pedestrian**
Jakkur-YOT	2min 23 sec	17 sec	15 sec
NES - Jakkur	1min 40 sec	57 sec	
YOT-NES	1min 30sec	1min 23 sec	
Jakkur-NES	1min 17sec	1min 10 sec	

RESULTS & DISCUSSIONS

With the present investigation, NES to Yelahanka Old Town and Yelahanka Old Town to NES is found to be the busiest routes of all having 58185 and 70222 vehicles respectively in PCU for 5 days, followed by other routes



I. At the stretch from Junction to NES for spot speed study, 15th percentile, 85th percentile, and 98th percentile of the speed are found to be 23.5 kmph, 33 kmph and 41.5 kmph respectively.

II. At the stretch from NES to Junction for spot speed study, 15th percentile, 85th percentile, and 98th percentile of the speed are found to be 22 kmph, 29 kmph and 35.5 kmph respectively.

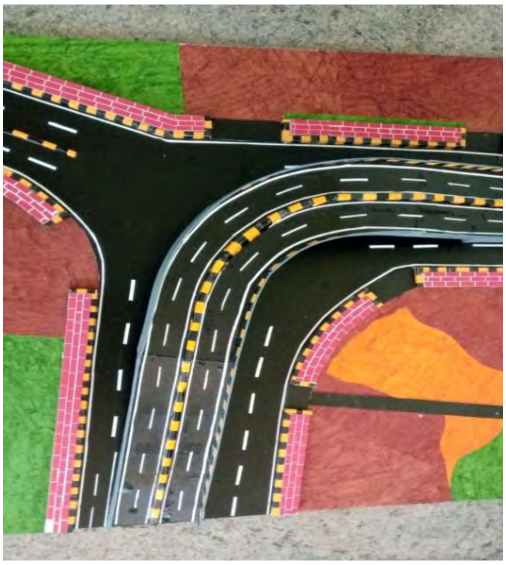
III. At the stretch from Junction to Jakkur for spot speed study, 15th percentile, 85th percentile, and 98th percentile of the speed are found to be 23.5 kmph, 32 kmph and 36.5 kmph respectively.

IV. At the stretch from Jakkur to Junction for spot speed study, 15th percentile, 85th percentile, and 98th percentile of the speed are found to be 35 kmph, 48 kmph and 62 kmph respectively.

V. As for the signal analysis the maximum Red signal time is found in JAKKUR to YOT and YOT to NES is 2 min 23 sec and 1 min 30sec respectively, by providing flyover red light can be reduced.

CONCLUSIONS

- Considering the route wise traffic volume for the taken 5 days, a flyover can be proposed from NES towards Jakkur and vice-versa as there is no conflict point in this route compared to other routes. A service road should be provided on both sides of the flyover for easy movement of vehicles travelling from NES to Yelahanka Old Town and those travelling from Yelahanka Old Town to NES below the flyover, reducing the congestion rate at the junction.
- By providing this alternative method the Red light of YOT-NES and JAKKUR-YOT will be reduced, so that we can reduce the travel time of these routes.
- The flyover proposition may not seems to be justified for now, but considering upcoming office building, a mall and apartments in the area along with other ongoing and future developments, it is very much justified as they will add much more traffic at the junction in the near future.



length

NES- junction = 220 m

junction – jakkur = 600m

radius = 60 m

width

divided two lane = 15 m

median = .5 m

height = 6 m

REFERENCES

1. "Highway Engineering" – Khanna S. K. and Justo, Nemchand and Bros.
2. "Highway Engineering" – Srinivas Kumar.
3. "Traffic Engineering and Transport Planning" – L. R. Kadiyalikhanna Publishers.
4. "Traffic Flow Theory" – Drew Mc. Graw Hill and Co.
5. "Transportation Engineering and Planning" – Papa Coastas and Prevendors PHI, New Delhi.
6. "Indian Transport Yearbook 2014-15"
7. IRC Journals : Journal of Transportation Engineering/Volume 122 issue 6-Nov 1996
8. International journal of application (web site:www.ijaiem.org)
9. Web site:www.researchgate.net

LABORATORY STUDIES ON CHEMICAL STABILIZATION OF BLACK COTTON SOIL USING COMMERCIALLY AVAILABLE STABILIZERS

Mandeep B V[1] G Kavitha[2]

[1] **MANDEEP B V**
[1]Assistant Professor, Nagarjuna College of Engineering and Technology, Bangalore
email: mandeepgowda7@gmail.com
[2] Asst Professor, RASTA center for road technology, Bangalore; email: gkavitha@rastaindia.com

Abstract

Black cotton soil is often referred as a highly expansive cohesive soil, visibly grey to black in color, found vastly in northern Karnataka region. Due to its high swelling and shrinkage characteristics, it poses huge problem in road construction industry, for use as a construction material or foundation material. Roads constructed over this soil as foundation layer, are prone to excessive cracks, rutting and potholes, because of its unstable nature. Conventional method practiced in site, is to construct subgrade layer by a stronger soil like gravel above the existing weak soil. This practice increases the project cost in terms of material cost, transportation cost, labor cost and machinery cost. If locally available weak material is strengthened using new methods and technologies that are available, total project cost might get reduced. For this to happen, first the engineering properties of stabilized soil have to be checked for its suitability as a subgrade layer. In this work an effort is made to check the engineering properties of stabilized expansive black cotton soil using commercially available chemical stabilizers and to judge its suitability as a construction material based on the results and in accordance with the standards and specifications.

Keywords: black cotton soil, expansive soil, cohesive soil, swelling, shrinkage, cracks, rutting, potholes, stabilization, chemical stabilizers

Introduction

Soil is the most important engineering construction material. It is the foundation material where all the loads from super structure gets transmitted and distributed to a greater depth thus making the super structure to remain intact and safe. In case of pavement structures, design objective is to minimize the stresses on subgrade so that pressure exerted on subgrade is well within its bearing capacity. In case of weak subgrade soil like black cotton soil, pavement crust thickness increases and so does the overall project cost. Even replacing the weaker soil with a stronger material like gravel incurs more cost.

Gravel is the most desirable construction material to be used in subgrade layers because of its high strength, low plasticity, very less shrinkage and swelling characteristics and excellent drainage property. However, not all soils satisfy the requirements for use as a subgrade material. Black cotton(BC) soil is one such material. It is a fine grained soil consisting of more than 50% particles, finer than 75micron sieve and thus having no consistent load bearing skeleton. It is highly compressible which might lead to excessive settlement and the condition is worsened by excessive shrinkage and swelling properties. During monsoons due to increased movement of water, these soils absorb water and its volume subsequently increases. This leads to swelling. During summer due to high temperature, water evaporates from soil leading to reduction of volume leading to shrinkage. This alternative volume changes is excessive and leads to disturbance in overlying layers and ultimately resulting in failure of pavement structure in terms of cracks, rutting potholes or appearance of wavy surface on top layers leading to bumpy riding surface.

This property of black cotton soil may be attributed to its mineralogy. Black cotton soils are vastly found in northern Karnataka plateau comprising of districts like Belgaum, Bidar, Bijapur and Gulbarga, the region being largely composed of Deccan trap. The Deccan trap consists of multiple layers of solidified flood basalt. This flood basalt disintegrated and weathered over a significant period of time to form various clay minerals. The mechanism of swelling and shrinkage can be well explained and solved by understanding the mineralogy and chemical composition of soil and its effect on strength properties of soil. This knowledge is used in developing stabilizers with a definite chemical composition to each type of soil. A lot of stabilizers like lime, cement, flyash and industrial wastes, have achieved limited success because improvement in engineering properties of soil depends upon chemical composition of the stabilizer used and mineralogy and composition of soil.

Objective

In the present study, two commercially available chemical stabilizers, stabilizer A and stabilizer B are used. Their effect on BC soil is studied by comparing index properties, strength and stability properties with varying dosage, curing period, soaked and unsoaked conditions with respect to that of native soil.

Literature review

K.V Madurwar et al studied the effect of RBI Grade 81 on Black cotton soil from Maharashtra region. RBI Grade 81 was applied at varying dosages ranging from 2% to 6%. Study concluded decreasing liquid limit and increasing plastic limit with increasing dosage of stabilizer and a net reduction in plasticity index. Also, decrease in free swell index was observed with increasing dosage of the stabilizer. There was increase in stability and strength properties of soil in terms of CBR and unconfined compressive strength.

In the studies conducted by Transportation engineering division, Civil engineering department, IIT-Kharagpur under title “Evaluation of subgrade soil using Stabilig modifier” the effect of Stabilig modifier at two dosages – 4% and 6% and a curing period of 28 days was studied on clayey soil. Study indicated an increase in optimum moisture content and max dry density with increasing dosage of stabilizer. Unconfined compressive strength also improved with increasing stabilizer dosage. Increasing trend of soaked CBR was observed with increasing dosage of stabilizer.

Udayashankar D. Hatari et al studied the effect of flyash on black cotton soil from Dharwad region. Flyash was applied at varying dosages ranging from 10%-60% by weight of soil. Engineering properties of blended soil was checked at three curing periods of 7 14 and 28 days. In this study both liquid limit and plastic limit were observed to have a decreasing trend with increased dosage of flyash at a certain curing period. Maximum dry density initially observed an increasing trend up to 30% dosage of flyash after which it showed a decreasing trend. Optimum moisture content was observed to have a decreasing trend with increasing flyash dosage. Unconfined compressive strength gain was high in case of 28 days curing period and was significantly high up to 20% dosage of stabilizer after 20% dosage, gain in strength was relatively less. CBR increased with increasing dosage of flyash up to 30% afterwards it was observed to have a decreasing trend.

Materials

1. Native soil

Natural soil sample was collected from Dharwad district, Karnataka. It is dark grey in color comprising of large hard lumps. It was sun dried and pulverized to make it pass through 425micron I.S Sieve prior to testing. Properties of native soil is as follows

Sl no	Property	Values
1	Gravel (%)	2.78
2	Sand (%)	6.26
3	Fines (%)	90.96
4	Liquid limit (%)	86
5	Plastic limit (%)	43.03
6	Plasticity Index (%)	43
7	Soil classification as Per IS 1498(1970)	CH/MH
8	Free swell Index (%)	56.75
9	Max Dry density (kN/cum)	1.6
10	Optimum moisture content (%)	21
11	CBR (%)	1.75
12	Unsoaked-Unconfined compressive strength (kN/sqm)	134
13	Soaked- Unconfined compressive strength (kN/sqm)	Fail

2. Stabilizers

Two commercially available stabilizers Stabilizer A and Stabilizer B are used. Both stabilizers are composites of cement, lime and naturally occurring carbonaceous polymers. However, both stabilizers have different proportions of the constituent raw materials.

Methodology and experimental program

Both stabilizers A and B were added in varying percentage of 4% and 8% by the weight of soil. Then they were cured for 7 days. Strength tests (UCS) were conducted in soaked and unsoaked conditions. Specimens were soaked 20 minutes in water prior to testing to simulate weakest state. Following tests were conducted in laboratory

1. Wet sieve analysis
2. Atterberg's limits
3. Free swell index
4. Proctor compaction test
5. CBR
6. Unconfined compression test (soaked and unsoaked condition)

Sieve analysis

Sieve analysis was conducted on native and blended soils for a curing period of 7 days. Test was conducted in accordance with I.S 2720 Part IV – 1975

Atterberg's limits

Atterberg's limits test is conducted on native and blended soils for a curing period of 7 days. Test was conducted in accordance with I.S 2720 Part V – 1970

Free swell index

Free swell index test was conducted on native and blended soils for a curing period of 7 days. Test was conducted in accordance with I.S 2720 Part XL – 1977

Proctor compaction test

Compaction test are carried out on native and blended soils. Modified compaction test was carried out. Test was conducted in accordance with I.S 2720 Part VIII – 1965

CBR

CBR test was conducted on native and blended soils for a curing period of 7 days. After 7 day curing, samples were soaked for 4 days. Test was conducted in accordance with I.S 2720 Part XVI-1979

Unconfined compression test

Unconfined compression test was conducted on native and blended soils for a curing period of 7 days and 28 days. Test was conducted in accordance with I.S 2720 Part X – 1991

Results

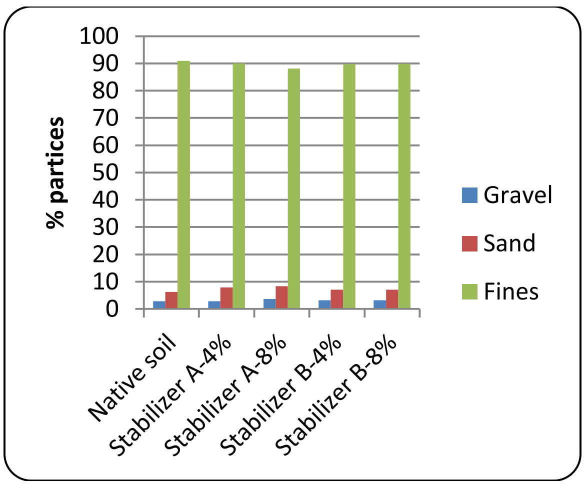


Figure 1. Sieve analysis

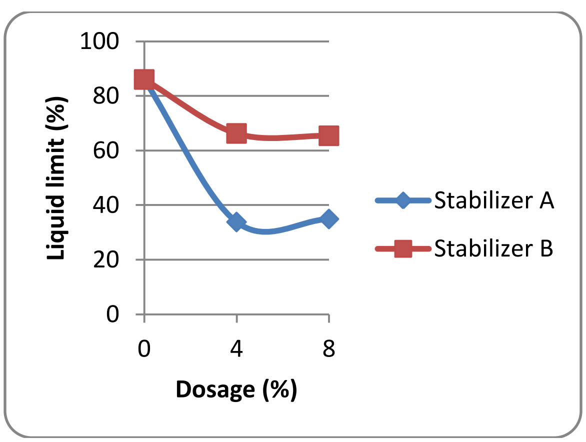


Figure 2. Liquid limit

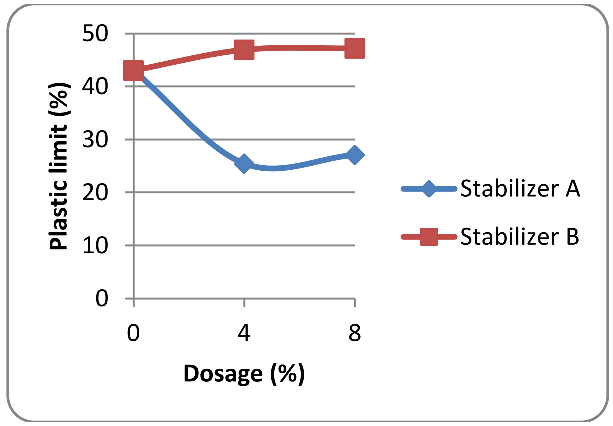


Figure 3. Plastic limit

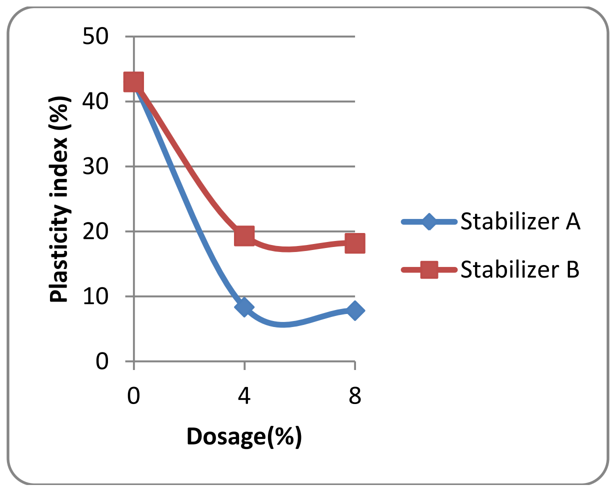


Figure 4. Plasticity index

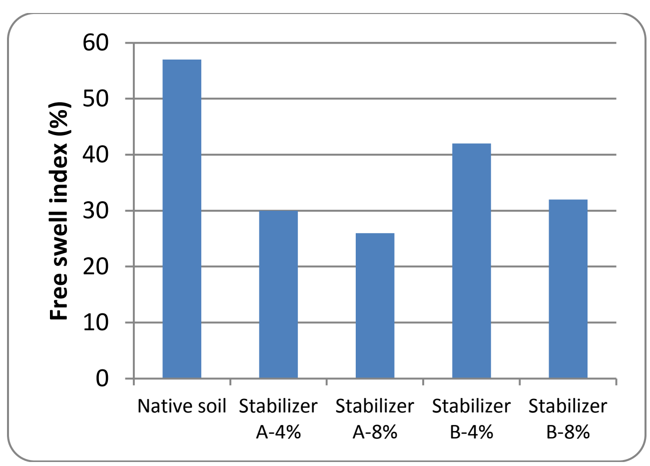


Figure 5. Free swell index

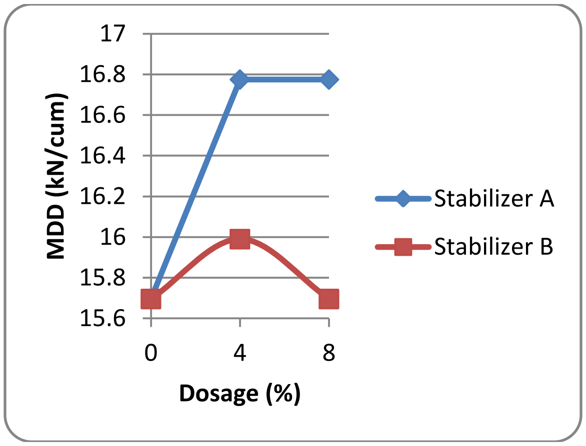


Figure 6. Maximum dry density

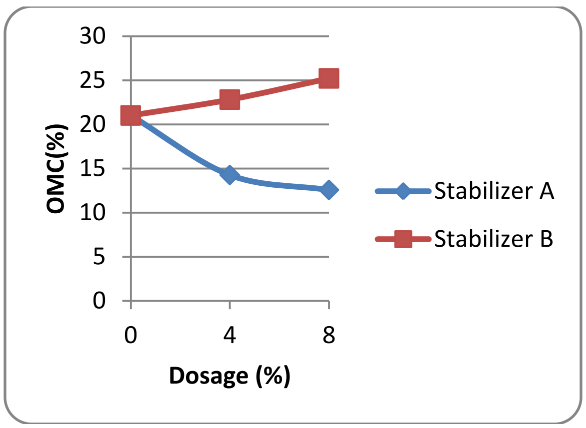


Figure 7. Optimum moisture content

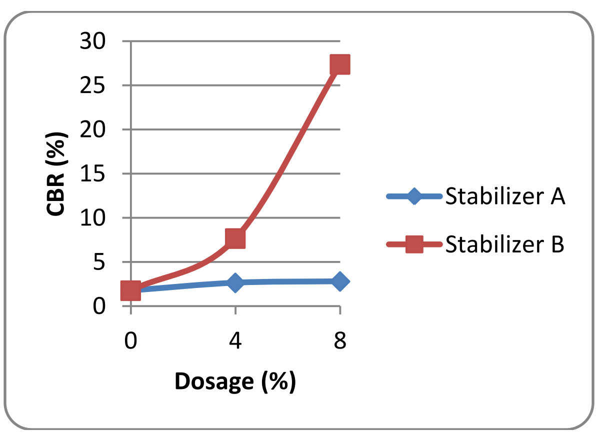


Figure 8. California bearing ratio - Soaked

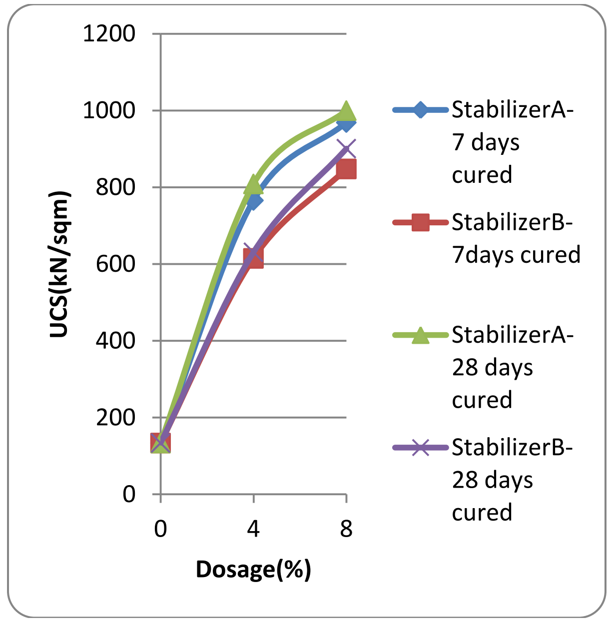


Figure 9. Unconfined compressive strength – unsoaked condition

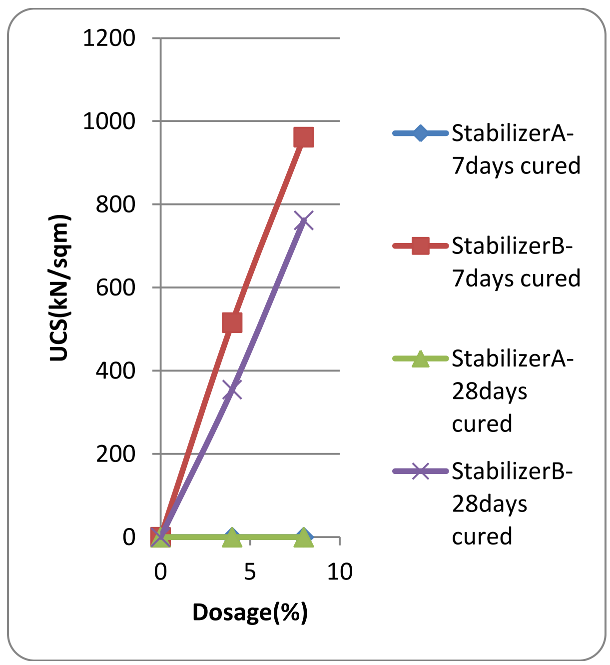


Figure 10. Unconfined compressive strength – soaked condition

Test	Native soil	Stabilizer A		Stabilizer B	
		4%	**8%**	**4%**	**8%**
Gradation					
% gravel	2.8	2.8	3.6	3.1	3.1
% sand	6.3	7.9	8.3	7.1	7.0
% fines	90.9	90	88.1	89.8	89.9
Atterberg's limits					
LL(%)	86	33.8	34.9	66.2	65.4
PL(%)	43	25.5	27.1	46.9	47.2
PI(%)	43	8.3	7.8	19.3	18.2
Freeswell index (%)	57	30	26	42	32
Compaction test					
MDD (g/cc)	1.6	1.71	1.71	1.63	1.6
OMC (%)	21	14.25	12.56	22.8	25.2
CBR (%)	1.75	2.63	2.79	7.65	27.36
UCS-7day curing-unsoaked (kN/sqm)	134	766	969	615	848
UCS-7day curing-soaked (kN/sqm)	Fail	Fail	Fail	516	962
UCS-28day curing-Unsoaked (kN/sqm)	134	809	1000	631	901
UCS-28day curing-Soaked (kN/sqm)	Fail	Fail	Fail	355	762

Observations

From above table and

1. *Figure 1* .No significant changes were observed in gradation for stabilized BC soil compared to unstabilized BC soil
2. *Figure 4* - Significant Decrease in P.I value was observed in stabilized BC soil, particularly in BC soil treated with stabilizer A, compared to the native soil. The highest reduction of 35.19 in P.I was observed in BC soil stabilized with 8 % dosage of stabilizer A, compared to BC soil stabilized with stabilizer B.
3. *Figure 5*, Free swell index is observed to have a decreasing trend due to stabilization. Highest reduction being observed in BC soil treated with Stabilizer A of 31% at 8% dosage
4. *Figures 6 and 7*, Maximum dry density has increased by 6.43% in case of BC soil treated with stabilizer A and remained constant at both dosages. OMC is observed to have a decreasing trend; highest reduction observed is 8.44% observed in BC soil treated with Stabilizer A ay 8% dosage. Not much variation is observed in case of BC soil treated with stabilizer B, also OMC is observed to have an increasing trend.
5. *Figure 8*, while CBR has been increased to 0.88% and 1.04% for BC soil stabilized with 4% and 8% dosages of stabilizer A. BC soil stabilized with stabilizer B shows a 5.9 % and 26% increase with 4% and 8% dosages respectively.
6. *Figures 9 & 10*, UCS in unsoaked condition has increased significantly in stabilized soils compared to native soil. The UCS value for unsoaked condition shows an increasing trend with increase in dosage in case of BC soil stabilized with both the stabilizers. The highest value observed is 1000 kN/sqm at 8% dosage of stabilizer A for curing period of 28days. The stabilized BC soil has shown good increase in strength in terms of UCS after a curing period of 28 days. However this is only in case of unsoaked condition. In case of soaked condition while the BC soil specimens stabilized with stabilizer B have shown increase in strength compared to the native soil, while BC soil stabilized with stabilizer A have all failed.

Conclusions

From the results and observations of laboratory tests mentioned above,

(i) Both the Stabilizers, A and B, upon stabilization, improves the engineering properties of BC soil in terms of reducing the plasticity index and free swell index.

(ii) There is significant strength gain in terms of UCS, in BC soil treated with stabilizer A in unsoaked condition than in BC soil stabilized with Stabilizer B.

(iii) Same soil treated with stabilizer A fails under soaked condition in the UCS test and also no significant increase in CBR in soaked condition.

(iv) Stabilizer B shows improvement in stability and strength in terms of both CBR and UCS even under soaked condition. This is possibly because the durability of stabilizer A may be less than the durability of stabilizer B.

(v) There is a need to conduct durability tests on the soils treated with both the stabilizers and also conduct performance studies like repeated load tests on the stabilized specimens in order to assess the best stabilizer. This is considered as the future scope of this work

References

1. K Terzaghi, R B Peck, G Mesri "Soil mechanics in engineering practice" *Wiley interscience publication*., Third edition: 3-204
2. E J Yoder and M W Witczak "Pavement analysis and design" *Wiley international publishers*., Second edition: 221-325
3. James P Ferris " Mineral catalysis and prebiotic synthesis: Montmorrilonite formation of RNA" *Cambridge university press.,* Elements, June 2005 Vol1: 145-149
4. E G Lotse et.al; "Mineral composition of some red and black soils in India" *Nuclear research laboratory, Indian agricultural research institute*., Vol 40B, No2, India: 216-226
5. Ashish Mehta et.al; "Stabilization of black cotton soil using flyash" *International journal of application or innovation in engineering and management*.,2013 ISSN 2319-4847: 1-5
6. Pankaj R Modak et.al; "Stabilization of black cotton soil using admixtures" *International journal of engineering and innovative technology*., May2012 Vol1, Issue5,: 1-3
7. B C Punmia "Soil mechanics and foundations" *Laxmi publications*., 16th edition: 9-142
8. K R Arora "Soil mechanics and foundation engineering" *Standard publishers*: 1-400
9. S K Khanna, C E J Justo "Highway engineering" *Nemchand and brothers*., 2013 edition: 1-250
10. A.O Ogundalu et.al; "Minerological and geotechnical characterization of Maiduguri black cotton soil using X Ray diffraction, X Ray photoelectron and scanni+ng electron microscopy" *International journal of Engineering and technology*., June 2014, Vol4 No6: 345-353
11. Transportation engineering division "Evaluation of subgrade soil using Stabilig modifier" *Indian Institute of Technology-Kharagpur*.,2012: 3-9
12. Bureau of Indian Standards, Relevant I.S Codes of practice

Analysis of Junction and Road User Traffic Data, to reduce the Congestion at Tin Factory Junction

MANDEEP B V
Assistant professor, Nagarjuna college of Engineering and Technology, Bangalore
E mail: mandeepgowda@gmail.com

Abstract -- A bus stand along with the steel structure skywalks designed and recommended at Tin factory junction in Bangalore for the reduction of the traffic by 80%. The skywalk is designed keeping all the norms of the traffic engineering and the structural design. After the survey and the personal investigation for 3 months with the people travelling daily, weekly, monthly in different vehicles we found out that everyone had the same problem of the delay at junction. Tin factory is a junction in Bangalore from where everybody travels to the MNC's, as the junction is the through route to K R Puram, Hoskote, Silk board and Bangalore city. Any delay at the junction causes overall wastage and loss to the capital in general and to the country at large. If a bus stop is provided far from the junction point, the buses directly moving to K R Puram will become free and will be able to pass easily. The buses travelling to Hoskote can stop at 1km away from the junction (bus stop shown in the model) so that the passengers getting down and coming in won't disturb the traffic in any ways. According to survey and conclusion it will also decrease the rate of pedestrian accident and reduce their wastage of time of walking and waiting.

I. INTRODUCTION

Bangalore city is the capital of the state of Karnataka in southern India and is one of the fastest growing cities of the country and also the world. The city is growing in its stature as a 'cosmopolitan' center and is the leading hub of Information Technology. Owing to factors such as favorable climatic condition, commercial and employment, cultural life, etc., the population of the city has been increasing by alarming proportions. Consequently, apart from increased number of public transport vehicles, there has been a tremendous increase in the number of personalized vehicles such as two-wheelers and motorcars. The city growing radially with new towns being constructed has led to high level of traffic which has put the road network under great strain. The government and the concerned agencies such as the Bangalore Development Authority (BDA) have the task of improving the transportation infrastructure to meet the growth of traffic needs.

 Definition:

"Traffic engineering is the science of measuring traffic and travel, the study of the basic laws relating to traffic flow and generation and application of this knowledge to the professional practice of planning, designing and operating traffic systems to achieve safe and efficient movement of people and goods."

 Scope of Traffic Engineering:

The basic object of traffic engineering is to achieve

- Free and rapid flow of traffic.
- Less number of accidents.
- Less delay at intersections.
- Streamline flow of traffic for less journey time.

The study of traffic engineering may be divided in to 7 sections:

- Traffic characteristics
- Traffic studies and analysis
- Traffic operations-control and regulation
- Planning and analysis
- Geometric design
- Administration and management

 Passenger Car Units (PCUs):

Different classes of vehicles such as cars, vans, buses, trucks, auto rickshaw, motor cycles, pedal cycles, etc., are found to use the common roadway facilities without segregation on most of the roads in developing countries like India. The flow of the traffic with unrestricted mixing of different classes on the roadways forms the heterogeneous traffic flow or the mixed traffic flow.It is common practice to consider the passenger car as the standard vehicle unit to convert the other vehicles classes and the unit is called passenger car unit or PCU. It is expressed as PCU per hour or PCU/lane/hour.

Sl. No.	VEHICLE CLASSES	EQUIVALENCY FACTORS
1. 2. 3. 4. 5. 6. 7.	Passenger cars, tempo, auto rickshaw, agricultural tractor Bus, truck Motor cycle, scooter and pedal cycle Cycle rickshaw Horse drawn vehicle	1.0 3.0 0.5 1.5 4.0 6.0 8.0

Factors Affecting Passenger Car Unit:

- Vehicles characteristics such as dimensions, power, Transverse and longitudinal gaps.
- Traffic stream characteristics such as composition mean speed, volume to capacity ratio, etc.
- Roadway characteristics such as road geometric including gradient, curve, etc., Regulation and control of traffic such as speed limit, one-way traffic, etc.
- Environment and climatic conditions.

Traffic Volume Studies:

Traffic volume is the number of vehicles crossing a section of road per unit time at any selected period. The objects and uses of traffic volume studies are given below:

- Traffic volume is generally accepted as a true measure of the relative importance of roads and in deciding the priority for improvement and expansion.
- Traffic volume study is used in planning, operation and control of existing facilities.
- This study is used in the analysis of traffic patterns and trends.
- Classified volume study is useful in structural design of pavements, in geometric design and in computing roadway capacity.

Methods of Volume Studies:

1. Manual method:

Manual counts are typically used when:
Small data samples are required.

Automatic equipment's are not available, or the effort and expense of using automated equipment's are not justified.
Manual counts are typically used to gather data about the following:

- Vehicle classifications
- Turning movements
- Direction of travel
- Pedestrian movements
- Vehicle occupancy

2. Automatic Count Method:

The automatic count method provides a means for gathering large amounts of traffic data. Automatic counts are usually taken in 1-hour intervals for each 24-hour period. The counts may extend for a week, month or year. When the counts are recorded for each 24-hour time period, the peak flow period can be identified.

3. Automatic Count Recording Methods:

Automatic counts are recorded using one of three methods:

- Portable counters
- Permanent counters
- Videotape

4. Portable Counters:

Portable counting is a form of manual observation. Portable counters serve the same purpose as manual counts but with automatic counting equipment. The period of data collection using this method is usually longer than when using manual counts.

5. Videotape:

Observers can record count data by videotaping traffic. Traffic volumes can be counted by viewing videotapes recorded with a camera at a collection site. A digital clock in the video image can prove useful in noting time intervals.

 Speed Studies:
Speed studies carried out occasionally give the general trend in speeds. There are two types of speed studies carried out,

1. Spot speed study
2. Speed and delay study.

1. Spot Speed Study:

Spot speed study may be useful in any of the following aspects of traffic engineering:

- To use in planning traffic control and in traffic regulations
- To use in geometric design- for redesigning existing roads or for deciding design speed for new facilities.
- To use in accident studies
- To decide the speed trends
- To study the traffic capacity
- To compare diverse types of driver and vehicles under specified condition.

The spot speeds are affected by physical features of the road like pavement width, curve, and sight distance, gradient, pavement uneven intersection and road side developments. Other factors affecting spot speeds are environment conditions (like weather, visibility), enforcement, traffic conditions, driver, vehicle and motto of travel.

There are a number of methods to measure spot speed. The spot speed may be obtained either y finding the running speed of vehicle over a short distance of less than 200m or by finding the instantaneous speed while crossing a section, depending on the method used.

2. Presentation of Spot Speed Data:

- Average speed of vehicles: From the spot speed data of the selected samples, frequency distribution tables are prepared by arranging the data in groups covering various speed ranges and the number of vehicles in such range. The arithmetic mean is taken as the average speed. The table gives the general information of the speed maintain on the section, and also regarding the speed distribution pattern.
- Modal average: a frequency distribution curve of spot speeds is plotted with speed of vehicles or average values of each speed group of vehicles on x axis and the % of vehicles in that group in the y axis. this graph is called speed distribution curve. This curve will have a definite peak value of travel speed across the section and this speed is denoted as model speed. This speed distribution curve is helpful in determining the speed at which the greatest proportion of vehicles moves, given by the model speed.

II. LITERATURE REVIEW

A. Satish C.Sharma,et.al(1996), “ state wide traffic volume studies and precision of AADT estimates”

Investigation in this paper is the statistical precision of annual average daily traffic (AADT) estimates resulting from short period traffic counts (SPTC). A large number of automatically traffic recorder (ATR) sites located on Minnesota’s highway system of the study. The study result suggests that highway agency should put more emphasis on sample site assignment to correct ATR groups than on the durations of count.

B. A.K.M Abir and Md. Sami Hasnaine(1995), “traffic volume study”

In this work emphasis was given on traffic volume with volume, speed and density and the analysis was carried out through primary traffic flow survey AUST- flyover junction to the shatrasta junction in the Dhaka city. The traffic flow was studied by method of the following conclusion were drawn:

1. Old buses should be replaced by new buses in order to maneuver easily.
2. Separate lane system should be introduced for bicycles.
3. Large container truck should be allowed to use this junction only at the off-peak hours.

C. Rajiv Ganguly, Ashok Kumar Gupta, et.al(2014),”traffic volume and accident studies on national highway-22 between solan and Shimla, India”

This paper is an attempt to study the basic difference between rigid and flexible pavements based on LCCA, durability, recycling, local performance and safety features. An observation of the traffic count was conducted at Tuttikandi in Shimla and Chambaghat in Solan and difference in vehicular flow was completed. The process involved making manual count of the vehicles passing by on week days and weekends. Vehicle fleet was characterized as cars, buses, trucks and motor cycles. The study also involved in identifying the black spots which refers to those stretches on Shimla-Solan highway of NH-22 reported with the greatest number of accidents.

The study results suggest suitable remedial measure to prevent such accidents in future including installing various roads furniture like convex mirror, fluorescent sign boards, light poles.

D. Mr. Udit Batra, Mr. Mandar V. Sarode (2013), “Traffic Surveying and Analysis”

In this survey, a traffic survey in specific road sections on Nagpur city has been carried out which included calculation of present traffic density and comparison with previous year data, average velocity of traffic. Manual method of counting was used with the help of video recording. The following conclusions were drawn-

1. Traffic distribution through alternative routes or construction of flyover may ne envisaged.
2. Public transport system needs to be strengthened thereby reducing traffic density.
3. Parking of vehicles should be prohibited on streets.

III. METHODOLOGY

1. Volume Study:

To conduct traffic volume study at the junction, we adopted manual method of counting with the help of video recordings. 3 working days and 2 non-working days were selected to conduct traffic volume study, i.e., Monday, Tuesday, Thursday, Saturday and Sunday. Morning, afternoon and evening peak hours were selected for the study, i.e., 7-10 am, 1-2 pm and 4-7:30 pm.

2. Spot Speed Study:

Speed is an important transportation consideration related to safety, time, comfort, convenience and economy. Spot speed studies are used to determine the speed distribution of a traffic stream at a specific location. The data gathered in Spot Speed Studies are used to determine the vehicle speed percentiles, which are useful in making speed related decisions. FROM BANGALORE CITY TO HOSKOTEFROM TIN FACTORY TO SILK BOARD to conduct spot speed study, we have selected a working day, i.e., Thursday. 4 routes, i.e., A distance of 150 m was taken on all the routes and the time taken by different vehicles for covering the distance were recorded and speed were calculated. Further, frequency percentile and cumulative frequency percentile were presented in a tabular format and graphs were plotted, accordingly.

3. Pedestrian Survey:

The pedestrian mode consists of travelers along a roadway or pedestrian facility making a journey (or at least part of their journey) on foot. Pedestrians walk at different speeds, depending on their age, their ability, and environmental characteristics (e.g., grades and climate).

Pedestrian survey: Pedestrian survey is conducted during morning, noon and evening peak hours. Two working days and two non-working days were selected to carry out the study i.e. Monday, Tuesday, Saturday, and Sunday. The pedestrian count was collected from 7:00 am to 9:30am in the morning, 1pm to 2pm in noon followed by 4:00 - 8:00 in the evening.

Pedestrians are the major users of any roads. Many of the pedestrians are the regular road users either using vehicle or by walk. We conducted a pedestrian survey asking them about the ease of use of the present bus stop, road, and feasibility of the future in the same condition. And then we put forth our proposals in front of them too. Most of them (around 94% of the road users were convenience and happy with our proposals. They also told us to implement our project as fast as possible.

IV. EXPERIMENTAL INVESTIGATION

Tin factory Junction is the area that is studied. The area caters to considerable local and through amount of traffic especially at peak hours.This junction acts as a link between Hebbal to Chennai, Hebbal to Silk Board, Kr Puram to Market. Manual method of counting has been used to count the traffic volume at the junction with the help of video recordings. The data were taken for 5 days (Monday, Tuesday, Thursday, Saturday and Sunday) from 7-10 am, 1-2 pm and 4-5 pm.

Spot speed study was conducted on Thursday (working day) for 4 routes, i.e., Hebbal to Chennai, Hebbal to Silk Board, Kr Puram to Market.20 four wheelers, 10 two wheelers, 10 autos and 5 buses were considered for all the routes. Then, time taken by different vehicles to cover the distance was recorded and speed was calculated. Further details have been tabulated below:

1. Edistrain Volume Survey:

The pedestrian mode consists of travelers along a roadway or pedestrian facility making a journey (or at least part of their journey) on foot. Pedestrians walk at different speeds, depending on their age, their ability, and environmental characteristics (e.g., grades and climate).

Pedestrian survey: Pedestrian survey is conducted during morning, noon and evening peak hours. Two working days and two non-working days were selected to carry out the study i.e. Monday, Tuesday, Saturday, and Sunday. The pedestrian count was collected from 7:00 am to 9:30am in the morning, 1pm to 2pm in noon followed by 4:00-8;00 in the evening.

Time	Number of Pedestrian				
	ELDERS		STUDENTS	HANDICAP	TOTAL
	Gents	Ladies			
7.00-7.30 am	236	113	113		517
7.30-8.00 am	284	103	103		504
8.00-8.30 am	112	125	125	2	311
8.30-9.00 am	206	112	112		402
9.00-9.30 am	212	81	81		385
1.00-1.30 pm	206	23	23	2	255
1.30-2.00 pm	137	46	46		228
4.00-4.30 pm	55	126	126		215
4.30-5.00 pm	65	124	124		126
5.00-5.30 pm	63	95	95	1	186
5.30-6.00 pm	87	15	15		178
6.00-6.30 pm	155	55	55		312
6.30-7.00 pm	126	19	19		241
7.00-7.30 pm	210	10	10		355
7.30-8.00 pm	194	3	3		287
				Total	4502

Manual Survey with People

S.N O	TYPES OF PEDESTRI AN	ORIGIN	DESTINATION	USAGE OF JUNCTION	PURPOSE OF JUNCTION	EASE OF USAGE	RECOMME NDATION FOR BUS STOP	CASE OF DELAY
1	STUDENT S	KR PURAM	SILK BOARD	DAILY	OFFFICE	Not at all convenient	100%	No proper Bus stop
2	STUDENT S	TIN FACTORY	KR PURAM	DAILY	SCHOOL	Not at all convenient	100%	No proper Bus stop
3	STUDENT S	MAJESTIC	KOLAR	HOLIDAYS	COLLEGE	Not at all convenient	100%	No proper Bus stop
4	STUDENT S	MAJESTIC	NANDAGUDI	DAILY	COLLEGE	Not at all convenient	100%	No proper Bus stop
5	STUDENT S	MALLUR	KALYANNAGA R	DAILY	COLLEGE	Not at all convenient	100%	No proper Bus stop
6	STUDENT S	MULBAGILU	SARJAPUR	DAILY	COLLEGE	Not at all convenient	100%	No proper Bus stop
7	STUDENT S	HEBBAL	SILKBOARD	DAILY	COLLEGE	Not at all convenient	100%	Heavy Traffic due to buses
8	STUDENT S	KR PURAM	SILK BOARD	Twice a week	COLLEGE	OK	100%	Heavy Traffic due to buses
9	STUDENT S	MAJESTIC	KALYANNAGA R	Twice a week	COLLEGE	Not at all convenient	100%	Heavy Traffic due to buses
10	STUDENT S	MURABAGILU	SARJAPUR	DAILY	SCHOOL	NOT GOOD	100%	No proper Bus stop
11	JOB HOLDERS	ALLAMUDN NAGAR	CV RAMAN NAGAR	DAILY	OFFFICE	Not at all convenient	100%	No proper Bus stop
12	JOB HOLDERS	KALYANNAGAR	KR PURAM	DAILY	OFFFICE	Not at all convenient	100%	No proper Bus stop
13	JOB HOLDERS	SARJAPUR	SARJAPUR	DAILY	OFFFICE	Not at all convenient	100%	No proper Bus stop
14	JOB HOLDERS	KR PURAM	SILK BOARD	DAILY	OFFFICE	Not at all convenient	100%	Heavy Traffic due to buses
15	JOB HOLDERS	TIN FACTORY	KR PURAM	DAILY	OFFFICE	Not at all convenient	100%	Heavy Traffic due to buses
16	JOB HOLDERS	MAJESTIC	KALYANNAGA R	DAILY	OFFFICE	Not at all convenient	100%	Heavy Traffic due to buses
17	JOB HOLDERS	MALLUR	SARJAPUR	DAILY	OFFFICE	Not at all convenient	100%	No proper Bus stop
18	DAILY WORKERS	MAJESTIC	SARJAPUR	DAILY	WORKING	Not at all convenient	90%	Traffic due to Pedestrain
19	DAILY WORKERS	MALLUR	SILK BOARD	DAILY	ONCE IN WEEK	not convenient	90%	Traffic due to buses
20	DAILY WORKERS	HEBBAL	KR PURAM	DAILY	ONCE IN WEEK	OK	90%	Heavy Traffic due to buses
21	TRAVELL ERS	MURABAGILU	MAJESTIC	once in week	ONCE IN MONTH	OK	90%	Heavy Traffic due to buses
22	TRAVELL ERS	ALLAMUDN NAGAR	TIRUPATI	RARE	ONCE IN MONTH	not convenient	80%	Heavy Traffic due to buses
23	TRAVELL ERS	HEBBAL	SILKBOARD	RARE	ONLY DURING MEETINGS	not convenient	80%	Traffic
24	TRAVELL ERS	HEBBAL	SILK BOARD	depends on rquirement	RARE	Not at all convenient	50%	Traffic
25	DRIVERS	KR PURAM	SILK BOARD	DAILY	1 IN WEEK	Not at all convenient	50%	Traffic due to buses
26	DRIVERS	TIN FACTORY	KRPURAM	RARE	RARE	Not at all convenient	70%	Traffic due to Pedestrain

V. RESULT AND DISCUSSION

1. Providing the Bus stop far from the Junction along with the sky walk

At intersection where there is large number of crossing and right turn traffic, there is possibility of Several accidents and delay as there cannot be orderly movements. The main problem occurring there is the buses comes and stop at the junction where the passengers get in and come out of the bus. Because of this the vehicle behind get stuck and are not able to move freely. If we provide the bus stop away from the junction, by proving the extra sky walk, i.e. at 1.3 kms away from the time factory junction, the buses will go and stop there and the other vehicles that need to move can move n through traffic directly. Hence this will reduce the large amount of traffic and reduce the delay (which is 2hrs and more as of now).

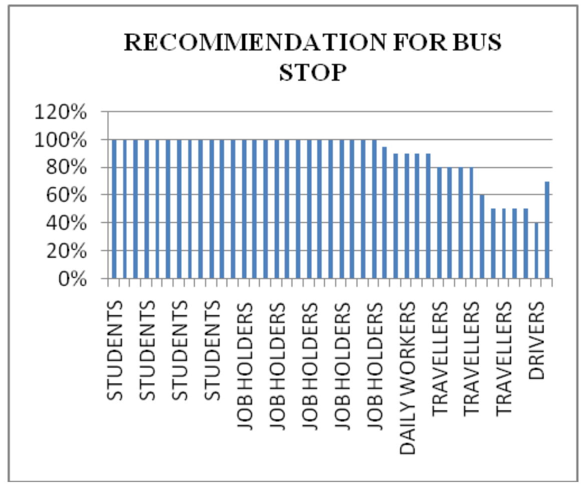
2. Advantages of Bus stop:

Properly designed Bus stop have the following advantages:

- They provide orderly movement of traffic and increase the traffic handling capacity of most of the intersections at grade.
- They reduce certain types of accidents, notably the right-angled collisions.
- Pedestrians can cross the roads safely at the signalized intersection.
- The passengers getting down the bus and climbing the bus will be comfortable.
- When the bus stop is properly coordinated, there is reasonable speed along the major road traffic.

VI. CONCLUSIONS AND RECOMMENDATIONS

- With the present investigation, the TIN FACTORY to KRPURAM route is found to be the busiest route of all having 58,215 vehicles in PCU for 4 days, followed by TIN FACTORY to SILK BOARD route and TIN FACTORY to HOSKOTE route.
- The major type of vehicle which is dominating this junction are cars followed by two wheelers at the 2nd rank and thereby the Bus.
- If a proper bus stop is provided, the buses will go and stop there, thereby allowing the other vehicles to move easily in the required direction.
- We have also recommended our project paper to the Bangalore east traffic police, and also to the deputy director of Bangalore east. The liked our project and told us that they will get back to us by discussing it with the higher authorities and implement is as fast as possible.



- DIMENSIONS of the sky walk designed
- Foot path - 4.4 feet
- Handrail width - 1 feet
- Handrail height - 3.8feet
- Kerb height - 0.7 feet
- Lane width each of 4 meters
- Median - 5 feet
- Road width @ proposed sky walk site - 15.5 meter
- Drainage width @ proposed sky walk site - 1.32 meter
- Sky walk landing site length (Availability)- 28 meter
- Sky walk site landing width(Availability) - 20.5 meter
- Distance to existing bus stand from sky walk landing site - 52 meter
- Road width @ Existing bus stand - 20 meter
- Drainage width @ Existing bus stand - 1.3 meter
- Kerb height for drainage - .3 meter
- Height of flyover @ proposed sky walk site - 7.6 meter

REFERENCES

[1] "Highway Engineering" – Khanna S. K. and Justo, Nemchand and Bros.

[2] "Highway Engineering" – Srinivas Kumar.

[3] "Indian Transport Yearbook 2014-15"

[4] "Traffic Engineering and Transport Planning" – L. R. Kadiyalikhanna Publishers.

[5] "Traffic Flow Theory" – Drew Mc. Graw Hill and Co.

[6] "Transportation Engineering and Planning" – Papa Coastas and Prevendors PHI, New Delhi.

TRAFFIC AT TIN FACTORY JUNCTION

A Proposed Model Plan for Bus Stand and Also the Sky Walk

Optical and Streakline flow based crowd estimation for surveillance system

Basavaraj G M, Ashok Kusagur

Abstract—**Recent advancement in computer vision system and growing population in the world has increased the consideration of the computer vision based system for the surveillance application in the real world scenario i.e. shopping malls, sport ground etc. The approach of surveillance requires the tracking of the persons and analysis of their motion. Existing approaches fail to address the problem of tracking and analysis due to the more density of the crowd,more occlusion in the video datasets or scenes. As the density or the occlusions increases, the quality of tracking and analysis also decreases which is a challenging task for the researchers. In this work we propose a new approach for computer vision based surveillance system which considers the video data as input by converting the frames to analyze the motion. We compute the flow and path by computing the streakline of the given frame sequence. This process results in the representation of the flow of the object to recognize the changes in the motion in spatial and temporal domains. This process is gives the better motion segmented results and outperforms the state-of-art techniques for the surveillance system**

Keywords—Crowd, Survillance system, motion, flow , fluid mechanics.

I. INTRODUCTION

In last decade the role of computer vision system is growing dramatically in the surveillance system. These systems suffer from the handling dense crowded scenarios i.e. sports ground, shopping malls, streets markets etc. Increasing the density of the crowded scenes affects the performance of the system in terms of detection of objects, tracking the objects etc. in the surveillance system analysis of the video, modelling of the crowd motion are the crucial tasks which induces various challenges for the researchers [1]. In this field of crowd scene analysis various researches have been done which are mainly classified into three main categories (a) structure based crowd behaviour analysis (b) machine learning approach for crowd behaviour analysis (c) threshold based approach for crowd behaviour analysis.

N. Li et al. [2] proposed a new approach which is based on the structure based approaches. This approach follows the topological method for tracking and analysis of crowd. Another method is machine learning approach which includes social force model [3], local pressure model [4] and Bayesian approach [5]. Threshold based approach is used in [6].P. Molnar et al [7] proposed interaction based model for crowd behaviour analysis.

This model is called social force model which is based on the interaction force state among crowd. By using this approach crowd models are developed which bifurcates the social model from the crowd. This technique is used for the crowd behaviour analysis but the disadvantage of this approach is that it works for the low dense, clutter and occlusions.

To overcome these Mehran et al. [2] proposed an improved approach of social force model by considering a holistic approach by including particle computation method is combined to classify the orientation of the data of the group of the crowd. In this method particles are placed on the image grid which moves according to the flow of the crowd. For high density crowd scene Ali et al. proposed segmentation based approach for the crowd data. Segmentation is carried out by applying coherent principle structure in fluid dynamics.Segmentation based approach for dense crowd scenes using dynamic texture is proposed in [8]. Another similar approach is given by Sand et. al., [9] which considers particle based computation of flow and motion of the crowd.

The main aim of this work is propose a streakline flow based approach for the crowd behaviour analysis to overcome the existing research issues. This approach is well considered for the flow visualization and dynamics of fluid in the computer vision applications [11]. This approach utilizes traces of the colored material to define the visualization of the object.

The paper organization is as follows: The proposed localizationmodels are presented in Section two. The results and the experimental study are presented in the section three. The concluding remark is discussed in the last section.

II. PROPOSED MODEL

In this section we discuss the proposed model.

A. Streak Flow

It is assumed that the pedestrian in crowd tend to follow the similar pathway as followed by the trailing pedestrians. If a human passes a point, there is a probability that other person behind him or her passes the same point by following the similar path. This is considered as a social behaviour of the crowd, but due to zig-zag motion of the crowd there are gaps in the flow of the crowd. These gaps need to be filled by considering the trajectory of the crowd motion.

In order to compute this we formulate a flow model which is denoted as $\alpha_s = (u_s, v_s)^T$. This flow is computed by computing the optical flow of the given scene and by forming the image models which are given as

$$P_i = \{x_i(t), y_i(t), u_i, v_i\} \tag{1}$$

$$u_i = \left(x_i^p(i), y_i^p(i), i\right)\text{and}v_i = v\left(x_i^p(i), y_i^p(i), i\right)$$

Position of the object in the frame is achieved by

$$x_{i\_pos}^p(t+1) = x_{i_{pos}}^p(t) + u\left(x_{i_{pos}}^p(t), y_i^p(t), t\right) \tag{2}$$

Similarly y –coordinates are given as

$$y_{i\_pos}^p(t+1) = y_{i_{pos}}^p(t) + v\left(y_{i_{pos}}^p(t), y_i^p(t), t\right) \tag{3}$$

Basavaraj G M, Associate professor Dept. of ECE, Nagarjuna College of Engineering and Technology, Bangalore, India.(baswaraj832@gmail.com)

Ashok Kusagur, Associate Professor, Dept. of E&E, University BDT College of Engineering, Davangere, Karnataka-5677002, India. (ashok.kusagur@gmail.com)

978-1-5090-0774-5/16/$31.00 © 2016 IEEE

Let $\left(x_{i_{pos}}^{p}, y_{i_{pos}}^{p}\right)$ is the x and y location of the object in the given frame at time t.

By using equation (2) and (3) we compute the sub-pixel accuracy. During next stage triangulation of the pixel is computed by considering the three neighbouring pixels. This is achieved by considering the interpolation method which is represented as

$$u_i = \mathcal{A}_1 \mathcal{U}_s(\mathcal{K}_1) + \mathcal{A}_2 \mathcal{U}_s(\mathcal{K}_2) + \mathcal{A}_3 \mathcal{U}_s(\mathcal{K}_3) \qquad (4)$$

Neighbouring pixel index is given by $\mathcal{K}$, $\mathcal{A}_j$ is the triangulation basis function, by using this equation a linear system is formed to achieve the data points which is given as

$$\mathcal{A}\mathcal{U}_s = \mathcal{U} \qquad (5)$$

The novelty of the proposed approach is given by performing the segmentation of each frame of the given video sequence into motion variation based regions. This can be achieved by computing the trajectories of the neighborhood pixels which is similar trajectories of particles passing from neighboring pixels over a period of time interval and it captures the affinity of current and previous motions at these pixels.

Computation of the flow is done by performing this operation on each frame and based on the streak flow is achieved. These quantities are used to compute similarity between the connectivity neighborhoods. For every pair of pixel value, the similarity is computed in terms of streak line streak flow. Each pixel is associated with a streakline of length 1 and the streakline similarity is computed using the sum of thenormalized projections of internal vectors as R_t(k,c)=$\sum_{m=0}^{l-1} prc(X_m^k, X_m^c)$.The streakline similarity defined asR_t (k,c)=(cos ($\angle\Omega_t^k$)-cos($\angle\Omega_t^c$)),where $\angle\Omega_t^k$ Is the angle of the streamline flow vector at pixel k. To define the boundaries of the regions we can be computed the similarity map at every pixel using $H(k) = \sum_{j\epsilon N(i)} \alpha R_t(k,c) + \beta R_\Omega$Where α and β are weights regulating the share of streakline and streak flow similarity is introduced in final segmentation. Since similar motions overtime builds similar storylines and streak flows, and boundaries of different motions form values in the similarity map. Using the negative of the similarity map we can segment the crowd into regions of similar motion with watershed segmentation.

III. SIMULATION RESULT AND ANALYSIS

In this section we discuss about the results evaluation and performance of the proposed model. The proposed approach is implemented using MATLAB tool by developing user interface. This approach is implemented and tested on the data set from University of Minnesota. Initially video sequence is taken and converted into frames. These sequences have normal and abnormal activities. In our simulation we have considered the HIT and Run sequence for performance evaluation.

In the below given section we describe the results of the proposed approach. Initially video frames are given as input. To show the performance we choose ground sequence. In the next stage each frame is taken as input in a given time interval. On the image the preprocessing steps are applied. After preprocessing steps, streaklines flow is computed which is sown if figure 4 and 5.

In the preprocessing steps it evolves grayscale conversion, binary conversion, noise removal and enhancement of the image.

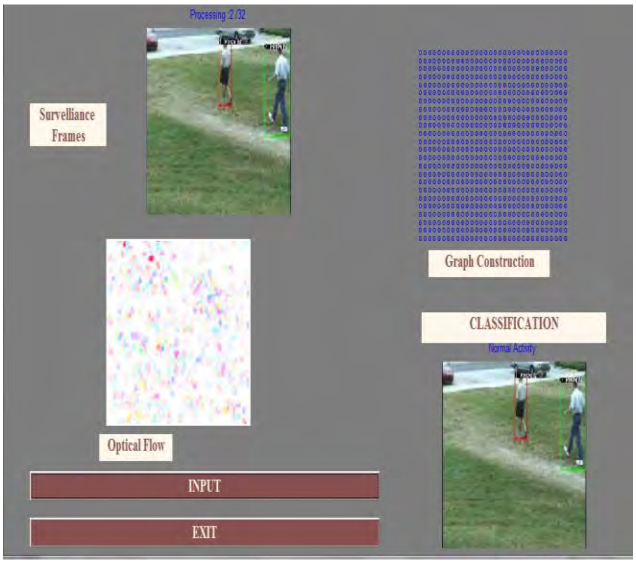


Fig 4: Optical Flow computations and normal activity classification

Once the frames are initiated, it is shown in the user interface. Input image is sown in the Input image section, next stage is to compute the streamline flow which is given in the corresponding section in the user interface. Optical flow computation is shown in the corresponding section in the user interface, finally the classification of the crowd behaviour is displayed with the indication of normal and abnormal event.

Fig5: Optical flow computation and abnormal activity classification

In the above given figure 5 detection of the abnormality in the crowd behaviour is presented.

Proposed approach is compared with the Spatio-Temporal Viscous Fluid Field (VFF) method. Experimental study shows that the proposed approach outperforms. The comparative study is done in terms of classification error and area under curve.

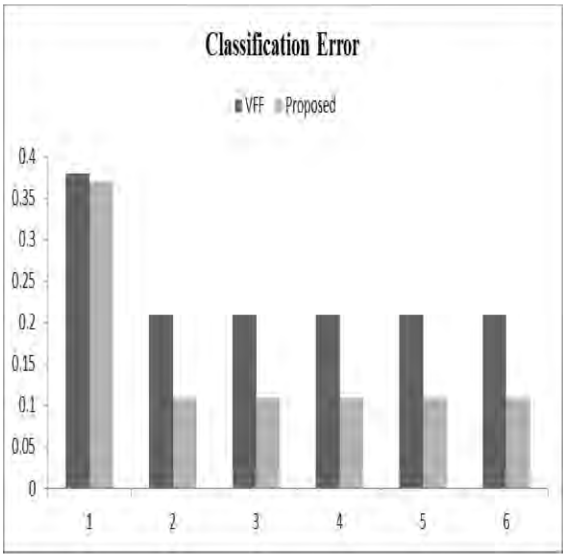


Fig.6: Classification error rate

Above figure 6 shows the performance of the proposed and existing approach in terms of classification error rate. Figure 7 shows the performance in terms of area under curve.

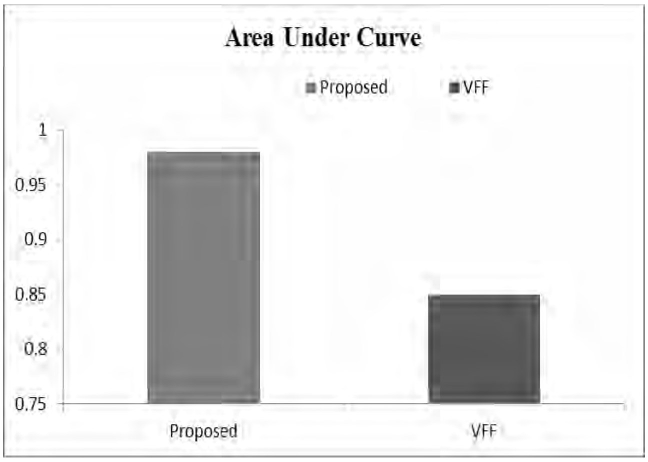


Fig.7: Area Under curve performance of the proposed and existing approach.

IV. CONCLUSION

In this work we propose an improved approach for surveillance system for computer vision systems. Here we use fluid mechanics based dynamic parameters computation of the object which helps us to find the flow of the moving object with path lines. This can be used for the surveillance system application. The proposed approach is implemented and results are carried out using MATLAB tool and benchmark datasets are considered for the study.

REFERENCES

[1] H. Su, H. Yang, S. Zheng, Y. Fan and S. Wei, "The Large-Scale Crowd Behavior Perception Based on Spatio-Temporal Viscous Fluid Field," in IEEE Transactions on Information Forensics and Security, vol. 8, no. 10, pp. 1575-1589, Oct. 2013.

[2] N. Li and Z. Zhang, "Abnormal crowd behavior detection using topological methods", IEEE Computer Society, 2011.

[3] R. Mehran, A. Oyama and M. Shah, "Abnormal crowd behavior detection using social force model", IEEE 2009

[4] H. Yang, Y. Cao, S. Wu, Weiyao Lin, S. Zheng, "Abnormal crowd behavior detection based on local pressure model," Asia Pacific Signal and Processing Association Annual Summit and Conference (APSIPA ASC), 2012

[5] S. Wu, H. Wong and Z. Yu, "A bayesian model for crowd escape behavior detection", IEEE Transaction on circuits and systems for video technology, 2013.

[6] G. Xiong*, X. Wu*, Y. Chen* and Y. Ou*, "Abnormal crowd behavior detection based on the energy model", IEEE, 2011

[7] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics". Physical Review E, 51:4282, 1995

[8] A. B. Chan et. al., "Mixtures of Dynamic Textures", IEEE ICCV, 2005.

[9] P. Sand et. al., "Particle Video: Long-Range Motion Estimation using Point Trajectories", IEEE CVPR, 2006

[10] S. C. Shadden et. al, "Definition and Properties of Lagrangian Coherent Structures from Finite Time Lyapunov Exponents in Two Dimensional Aperiodic Flows", Physica D, 212, 2005.

[11] Wijk, V., Jarke, J.: Image based flow visualization. In: SIGGRAPH. (2002)

Certificate

Department of Electronics and Communication Engineering
Motilal Nehru National Institute of Technology Allahabad
Prayagraj-211004, Uttar Pradesh, India

Springer

4th International Conference on VLSI, Communication and Signal Processing (VCAS 2021)

(September 24-26, 2021)

This is to certify that Mr./Ms./Dr./Prof. **Komal Srivastava** *from* **Motilal Nehru National Institute of Technology Allahabad** *has participated and presented paper titled* **Dual Band, Dual Port Dielectric Resonator Based MIMO Antenna with Bi-directional Pattern Diversity using PRS** *in 4th International Conference on VLSI, Communication and Signal Processing (VCAS 2021), Organized by Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Prayagraj (UP), India, held during September 24-26, 2021.*

Prof. R. A. Mishra	**Prof. R. K. Nagaria**	**Prof. Amit Dhawan**
Program Chair	Conference Chair	General Chair

Certificate

Department of Electronics and Communication Engineering
Motilal Nehru National Institute of Technology Allahabad
Prayagraj-211004, Uttar Pradesh, India

Springer

4th International Conference on VLSI, Communication and Signal Processing (VCAS 2021)

(September 24-26, 2021)

This is to certify that Mr./Ms./Dr./Prof. .... **Aditya Kumar Singh** .... *from* .... **J.K. Institute of AP&T, University of Allahabad** .... *has participated and presented paper titled* .... **Modeling and Investigation of Novel Two-Port UWB- MIMO Antenna with Enhanced Isolation** *in 4th International Conference on VLSI, Communication and Signal Processing (VCAS 2021), Organized by Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Prayagraj (UP), India, held during September 24-26, 2021.*

Prof. R. A. Mishra	Prof. R. K. Nagaria	Amit Dhawan Prof. Amit Dhawan
Program Chair	Conference Chair	General Chair

Gmail

Pinky L <pinkydevi@gmail.com>

Fwd: Acceptance of Paper for ICCRIP - 2019

mohd hussain <mohdhussain393@gmail.com> Tue, Nov 19, 2019 at 8:10 PM
To: pinkydevi@gmail.com

---------- Forwarded message ---------
From: <conference@nicmar.ac.in>
Date: Tue, 19 Nov, 2019, 2:22 PM
Subject: Acceptance of Paper for ICCRIP - 2019
To: <mohdhussain393@gmail.com>

Dear Author/s,

Greetings!

I am pleased to inform you that your paper titled **'Implementation of lean concepts in concreting work'** has been accepted for presentation in the 'International Conference on Construction, Real Estate, Infrastructure and Project Management (ICCRIP– 2019)' to be held during **'December 13 & 14, 2019'** at National Institute of Construction Management and Research (NICMAR), Pune. Please complete your registration for the conference through the online portal by paying the conference registration fees. The last date of registration is **'November 24, 2019'** .

You are also requested to upload the following by using your login details on or before **November 20, 2019**.

1.Copyright form in PDF format
2.PPT of your paper/s

Note: A link to upload copyright form and PPT will be made available upon successful payment of Registration Fee.

In the conference for paper presentation purpose, you will be given total 20 minutes, out of which 15 minutes will be given for the presentation of your paper and another 5 minutes for discussion.

For any further information, you may visit our website www.nicmar.ac.in. OR email to conference@nicmar.ac.in

Thank you and we look forward to having you with us at ICCRIP – 2019.

With Regards,

Dr. Jonardan Koner
Conference Convener, ICCRIP – 2019
Professor and Dean- Research & Publications,
NICMAR, Pune

--

Implementation of Lean Concepts in Concreting Work

Author 1 Name: Mohammed Hussain Shariff R
Department: Civil Engineering
University/Institution/Organization: Nagarjuna College of Engineering and Technology
Town/City: Bangalore
Country: India

Author 2 Name: Meghana S
Department: Civil Engineering
University/Institution/Organization: Nagarjuna College of Engineering and Technology
Town/City: Bangalore
Country: India

Author 3 Name: Dr. L Pinky Devi
Department: Civil Engineering
University/Institution/Organization: Nagarjuna College of Engineering and Technology
Town/City: Bangalore
Country: India

Corresponding author: Mohammed Hussain Shariff R
Corresponding Author's Email: mohdhussain393@gmail.com
Alternate Email: pinkydevi@gmail.com

Abstract

Construction industry is one among the prominent sectors which is growing worldwide. To improve the performance of the construction industry, the new technologies are being adopted. Implementations of new technologies are deviating from traditional to modern approach. One of the modern approaches which is gaining acceptance worldwide is lean construction. Lean construction is a new concept to minimize the waste and to increase the value of the end product. Lean also plays an important role in reduction of various factors like time, cost and energy. There are different lean tools and methodologies that can be implemented in construction site. One such methodology is value stream mapping (VSM). Value stream mapping helps in analysing the current state of any process by identifying the waste and improvising it in the future state. Two case studies A and B are selected for the implementation of lean methodology. In both the case studies initial data for the various factors like time and cost for the construction process of columns, beams and slab for the current state is analysed. The non-value added activities for the construction of columns, beams and slab in the current state are identified. Future value stream mapping is carried for both case studies. From the case studies the comparison is made between current and future state. It is observed that there is a reduction of time and cost in both the case studies. The implementation of lean methodology improved the performance and increased the efficiency of the construction sector.

Keywords: Lean; Value stream mapping; Time; Cost; Concreting.

Implementation of Lean Concepts in Concreting Work

1. Introduction

Construction is basically a process of constructing an infrastructure or a building, the major difference between construction and manufacturing is that the manufacturing process involves mass production of same item(s) without a consumer or purchaser at hand while construction typically takes place on location for a known client. Construction is an industry which comprises a 6 to 7 % of the gross domestic products of many developed countries, construction as a process starts with planning, then designing and lastly financing; this process continues until the project is complete and ready to occupy or use. Large-scale construction requires collaboration across multiple disciplines. A project manager normally manages the job, and a construction manager, design engineer, construction engineer or architects supervise it. The people involved in the designing and execution must consider the zoning requirements, impact on the environment while carrying out the job, scheduling, forecasting a budget, following construction site-safety precautions, transportation of building materials post confirming their availability, two way logistics, also factoring possible delays and bidding which could cause inconvenience to the general public.

Construction sector produces more waste; nowadays it seems very important to manage different wastes in different parts of the industry. Most of the modern western counties all over the world set different regulations in order to decrease and also manage the amount of waste in different parts of their industries. No construction site is waste less or rather the production of waste is inevitable in this industry, the fastest developing country with quite many active construction projects is Malaysia, the waste growth in such countries with a huge demands is drastic and sometimes can be taxing on the environment. Wastes can be categorised into a lot of broad categories and each category can carry its own set of impacts on the society. It seems crystal clear that in such a situation there is a great need of managing the wastes in any country.

Lean construction is “a way of designing production systems to lower the wastage of materials, time and effort consumed in order to generate the maximum amount of value possible”. Designing a production system to achieve the stated ends is only possible through the collaboration of all project participants (Owners, Architects/Engineers, Contractors, Facility Managers, and End-user) at early stages of the project. This goes beyond the contractual arrangement of design/build or constructability reviews where contractors, and sometimes facility managers, merely react to designs instead of informing and influencing the design.

Lean construction recognizes that desired ends affect the means to achieve these ends, and that available means will affect realized ends, Also the aim of lean construction is the embodiment of the benefits reaped by the Master builder concept. One can think of lean construction in a way similar to macroeconomics. Lean construction is based on the principles of project management and the principles that govern production management. Lean construction signifies that interaction between project and production management will undoubtedly result into a successful project undertaking.

Getting work to flow reliably and predictably on a construction site requires the impeccable alignment of the entire supply chain responsible for constructed facilities such that value is maximized and waste is minimized. With such a broad scope, it is fair to say that tools found in Lean Manufacturing and Lean Production, as practiced by Toyota and others, have leveraged lean construction principles. Total quality management, Statistical Process Control and Six-sigma have found their way into the lean construction. Similarly, tools and methods found in other areas such as in social science and business are used where they are applicable. The tools and methods in construction management such as CPM and Work Breakdown Structure, etc., are utilized in lean construction implementations. The three unique methods and tools that were specifically formulated for lean construction are the Last Planner System (LPS), Target Value Design (TVD), and the Lean Project Delivery System (LPD).

2. Literature Review

Evaluating two construction firm are compared with help of data collected by interviews and questionnaires and meetings. Improving the process by lean methodology and implementing in construction (Lia, S. et. al. 2016). Last planner system and Target value design which is a tool of lean used in understanding the existing cost and delay in projects. Lean concepts, is promoted as an alternative approach for improving the performance of the construction industry. The comparison of current state with the future state lean helped in reduction of waste such as cost within the activities. Software development, last planner and linear scheduling methods which is a lean tool to identify major factors affecting the construction projects, most of the construction project is delayed due to lack of coordination among contractors and clients (Song, L. et. al. 2008). Lean design approach not only helps in correcting the design process of a construction sector but also allows the team members to learn the errors from it. Value stream mapping can help in identify the waste and improvise the design process, inefficiencies that are hidden in a value stream mapping are identified and future state map is executed for showing the direction and improvise the system (Alves.et. al 2005).

In most of literature, it is studied that reduction of waste is the important criteria for any construction sector. Lean is a new concept of reducing waste. Some of the benefits of lean concepts are discussed. Lean concepts are implemented in some public sector projects where the cost overrun is managed by lean (Monyane, T.G. 2017). In some of the cases, the delay and rework are minimized using lean application (Howell, G and Ballard, G. 1998). A lean concepts namely percentage plan completion is applied to concerting process so that there is a co-ordination between the general contractor and subcontractor (Salem, O. et. al. 2006).

In some literatures the last planner and linear scheduling method helps in better coordination of architecture, engineer and contractor to follow up design changes and delivery project on time (Song, L. et. al. 2008). The lean design describes the lean productivity which exhibits value stream mapping where the work flow of the entire project is mentioned with time. Usually delay in project in current practice is due to some design error or lack of coordination in a firm. To avoid delay and wastage in a construction firm the lean tools such as project delivery system and six sigma techniques are implemented. The objective of this study is to analyse how lean concept can be applied for building construction. The scope of this work is limited to building construction located in Bangalore.

3. METHODOLOGY

Literature reviewing is beneficial to understand the importance of the study. The related literature and brief study of the methods helps in knowing the techniques used to simplify the process for the future and present research. In the literatures the implementation of lean concept for optimizing waste in construction and production unit are studied. The comparison of cost and time will help in better output towards building construction. The case studies are common research technique to analyse by various data collected from the source as much as possible for the study. Two case studies of building construction around Bangalore location are examined and each building construction is studied in detail. The building construction is inspected for 15 days to collect the data required for columns, beams and slab installation and concreting work. Preparation of data template for all the case studies is to collect the time required to complete each single activity in the construction process. The cost required for the material and equipment is also recorded in all case studies. The data collected is analysed with the help of value stream mapping with respect to time and cost. Each activity is analysed and the process is modified for better results. The analysed results are compared with the current and future process with respect to cost and time which leads to the efficient construction process.

4. Case Study

Data collected with regard to construction procedure from two different construction firms are listed in Table 1. The analysis is carried for both the firms to give a better outcome for efficient construction process.

Table 1: Details of Case Study

Description	**Case Study A**	**Case Study B**
Project name	ABC	XYZ
Location	Thirimenahalli	H.B.R layout 2^{nd} block
Total no. of floor	13	4
Built-up area	765.5 m^2	929.36 m^2
Total no. of blocks	9	1
Material supply 1. Steel 2. Ready mix concrete	By Client	By Contractor

4.1 Case Study A

Case study A is a residential construction based project. The built-up area is 765.5 m^2 with thirteen floors and the blocks built are nine and one block is chosen to carry out the study. The Block D is selected for study where the materials like steel and concrete is supplied by client. The concrete grades used in columns are M35 and for slab & beams M25 is used. Ready Mix Concrete is used at the site for all concreting work.

4.2 Case Study B

Case study B is a commercial construction based project. The built-up area is 929.36 m^2 with four numbers of floors only one block. The materials like steel and concrete is supplied by contractor. The concrete grades used in columns are M35 and for slab & beams it is M25. The concrete is of Ready Mix Concrete and for some of the concrete work site mix concrete is used.

5. Results and Discussion

5.1 Analysis of Case Study A

The Figure 1 represents the step by step construction process practiced in the Case Study A. The different activities like columns, beams and slab concreting are analysed and detailed discussions are carried out.

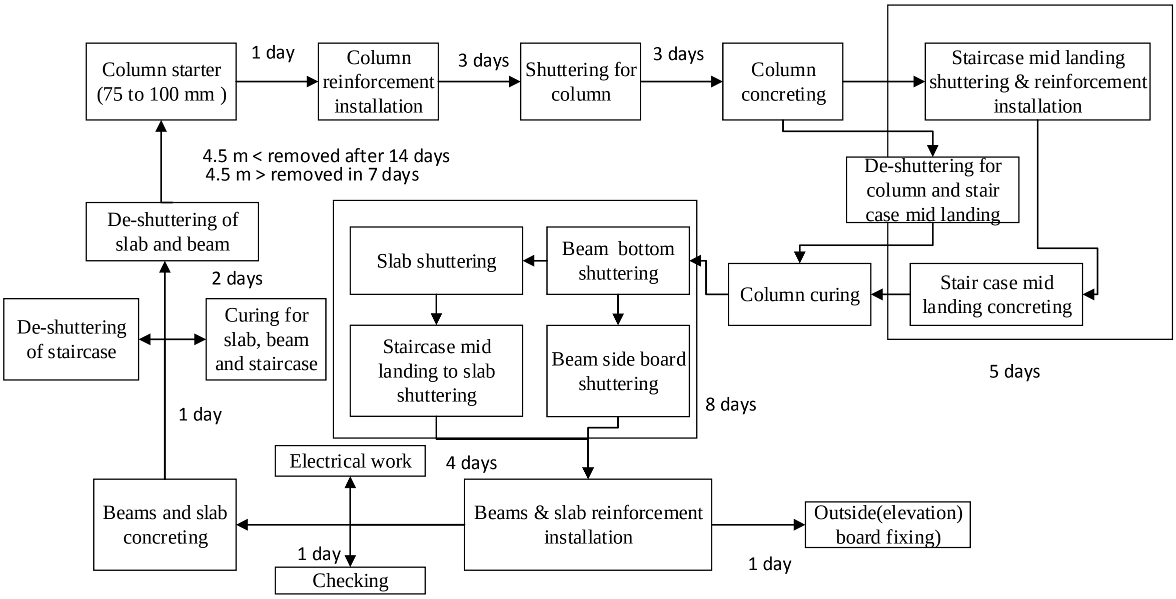


Figure 1: Construction Process in Case Study A

5.1.1 VSM for Column Execution Work

The Figure 2 represents the current practice of column execution work at the site. The time required for completing of one column is recorded which includes different activities like starter installation, reinforcement of column, fixing of shuttering board and finally with column concreting is recorded. In concreting process of column, concrete placing is done in two methods one is through crane bucket and the other one is with the help of labours. The time taken for concrete placing through crane bucket is less when compared to placing of concrete through labours.

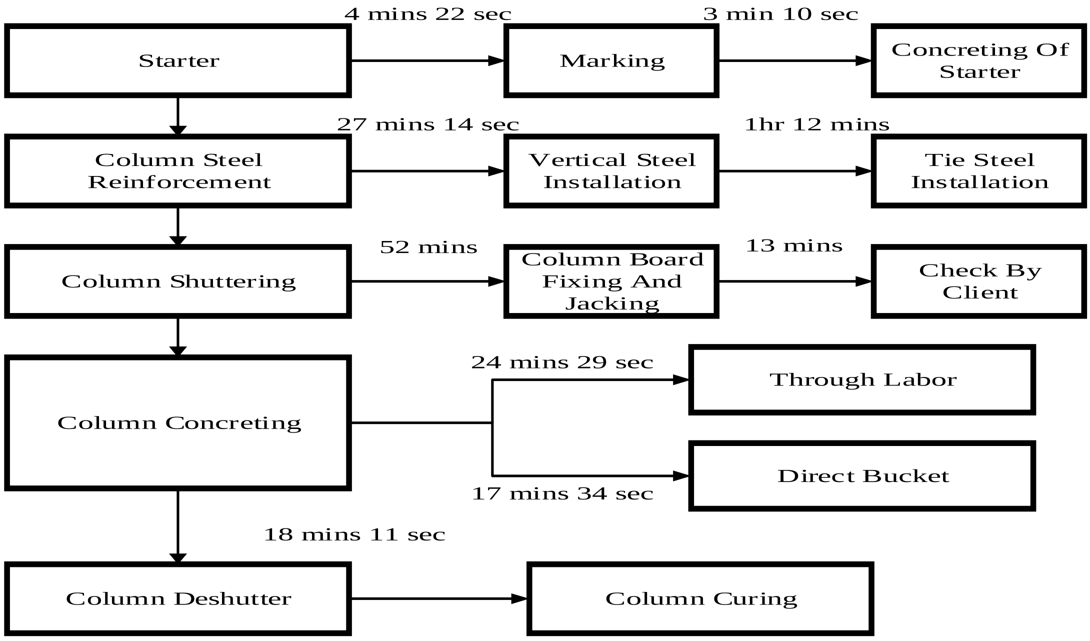


Figure 2: Current Practice of Column Execution Work

In Figure 3 the modified column execution work is mapped where the column starter is ignored and mapping is started from reinforcement of column by using clear cover block and check for plumb as it takes lesser time when compared to the current process and it is followed by column shuttering and concreting of column. It is observed that the modified column execution process takes less time compared to current practice. The labour required for column reinforcement can be increased for speeding up the work and to complete the work in an average time. Column concreting work is done by labour requires long duration when compared to concreting work by buckets using tower crane. Though only one column can be poured using tower crane bucket the time consumed for the process is less.

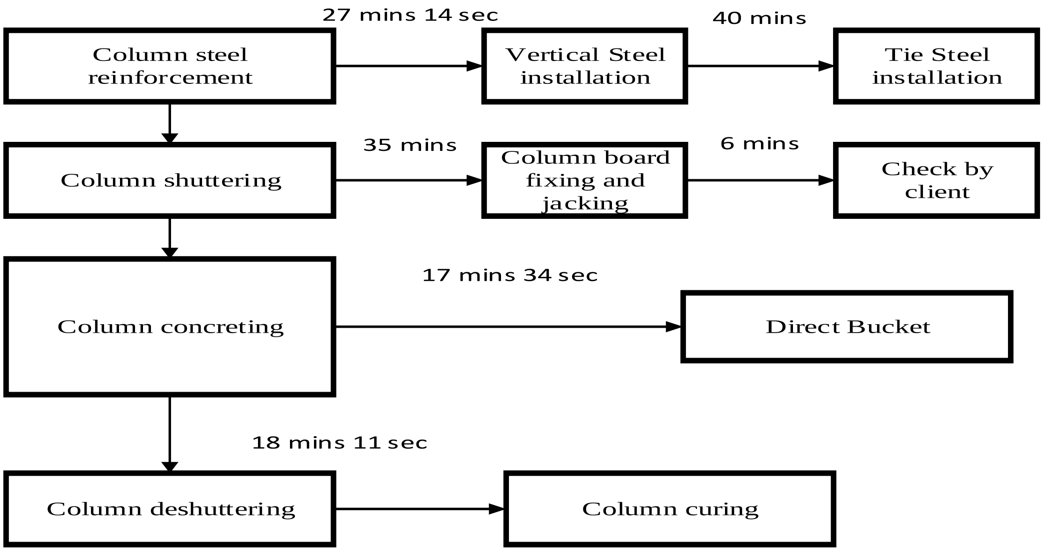


Figure 3: Modified Column Execution Work

5.1.2 VSM for Slab and Beams

The Figure 4 represents the flow diagram of beams and slab execution work with respect to one beam bottom shuttering, installation of reinforcement, shuttering of slab and slab reinforcement work along with concreting of beams and slab. The workings of the activities are recorded with respect to time taken for the execution. Block work at the site starts as soon as the second floor slab is finished.

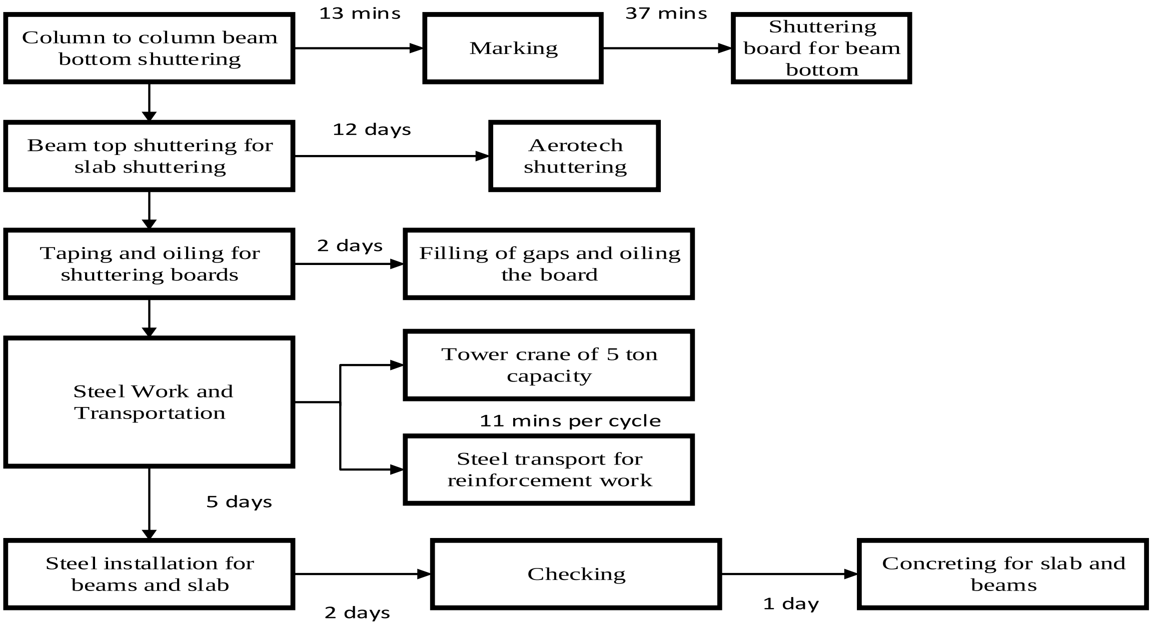


Figure 4: Beams and Slab Execution Work

The Figure 5 represents the current process of concreting work of slab and beams at site, where the concrete is placed by concrete stationary pump. For setting up the pipes to pump concrete takes almost one hour, instead it can be replaced by boom pump as it takes fifteen minutes to setup and for placing concrete. The rate of pumping concrete in boom pump is more than stationary concrete pump; as it is fully mechanized and easier for placing concrete where required. In the Figure 6, the concreting is done using boom pumping method and it helps labours work without any difficulty, more over the time and material wastage is reduced.

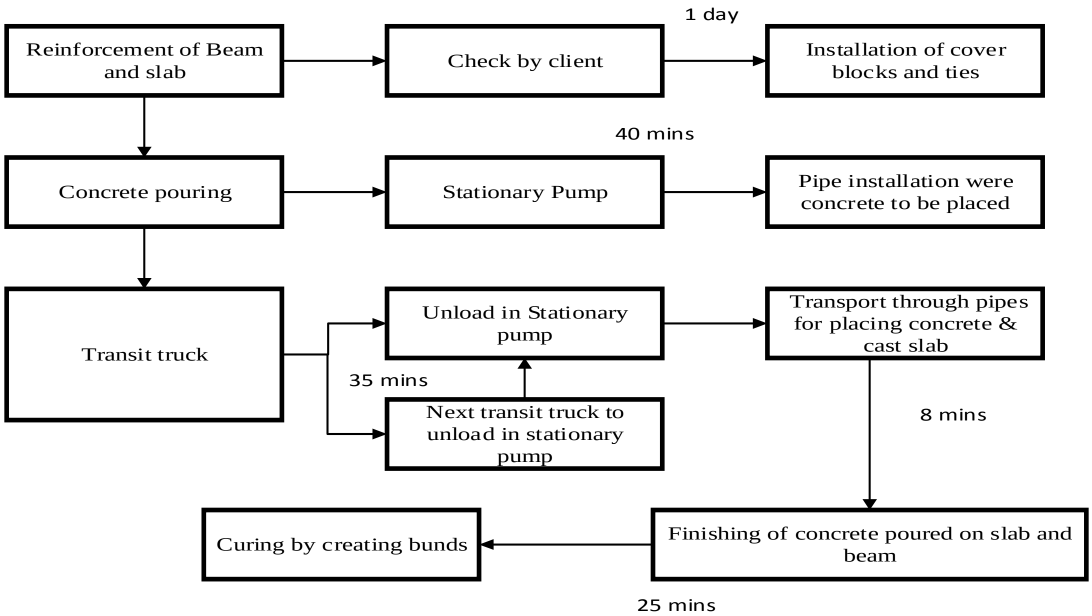


Figure 5: Current Practice of Slab and Beams Concreting Work

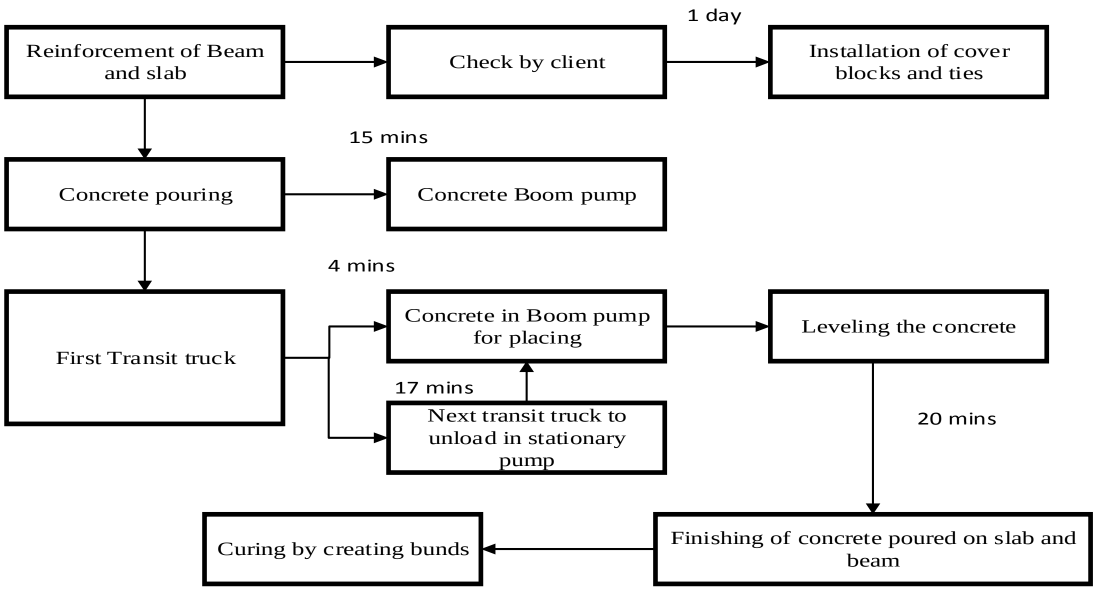


Figure 6: Modified Slab and Beams Concreting Work

5.1.3 Cost for Concreting Process

The Table 2 represents the cost of current construction process which includes labour cost for starter installation in column and it also includes cost of concreting by stationary concrete pump which is hired for two months. The modified construction cost is considered where the labour required for starter installation is deducted and the cost of equipment i.e. stationary concrete pump is replaced with concrete boom pump for concreting. The equipment is hired when huge volume of concrete is to be placed.

Table 2: Construction Cost in Case Study A

Sl no.	Equipment /Material	Quantity	Cost of Material /Equipment	Cost (in Rupees) Current cost	Modified cost
1	Concrete	115 m^3	4,950/m^3	5,69,250	5,69,250
2	Steel	11271.96 kg	55/kg	6,19,957	6,19,957
3	Tower crane (5 tons)		4,00,000/month	8,00,000	8,00,000
4	Bar bending machine		40,000/month	80,000	80,000
5	Backhoe		800/day	48,000	48,000
6	Tippers		1,000/day	60,000	60,000
7	Stationary Concrete pump		1,80,000/month	3,60,000	-
8	Concrete Boom Pump		18,000/day	-	36,000
9	Manpower			1,13,520	1,10,220
Total cost (in Rupees)				**26,50,727**	**23,23,427**

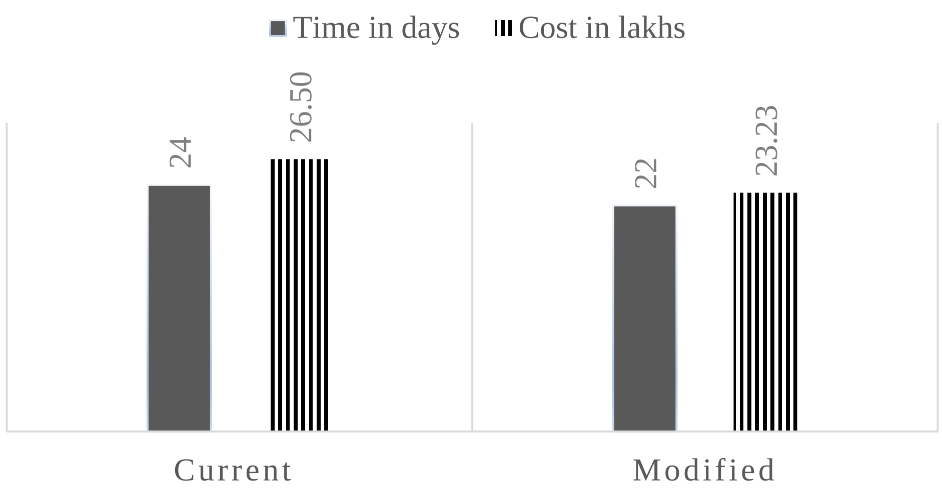


Figure 7: Comparison of Current & Modified Construction Time and Cost for Case Study A

The Figure 7 represents the comparison of the current and modified construction time and cost and the observations shows the reduction of time and cost. There is reduction in cost by 12% and time by 8% which results for efficient construction. In the present study cost and time are co-related to each other so before execution of the work proper planning is important and also managing the planned schedule as per the plan leads to efficient construction.

5.2 Analysis of Case Study B

The Figure 8 represents the step by step construction process practiced in the Case Study B. The different activities such as columns, beams and slab concreting are analysed and detailed discussions are carried.

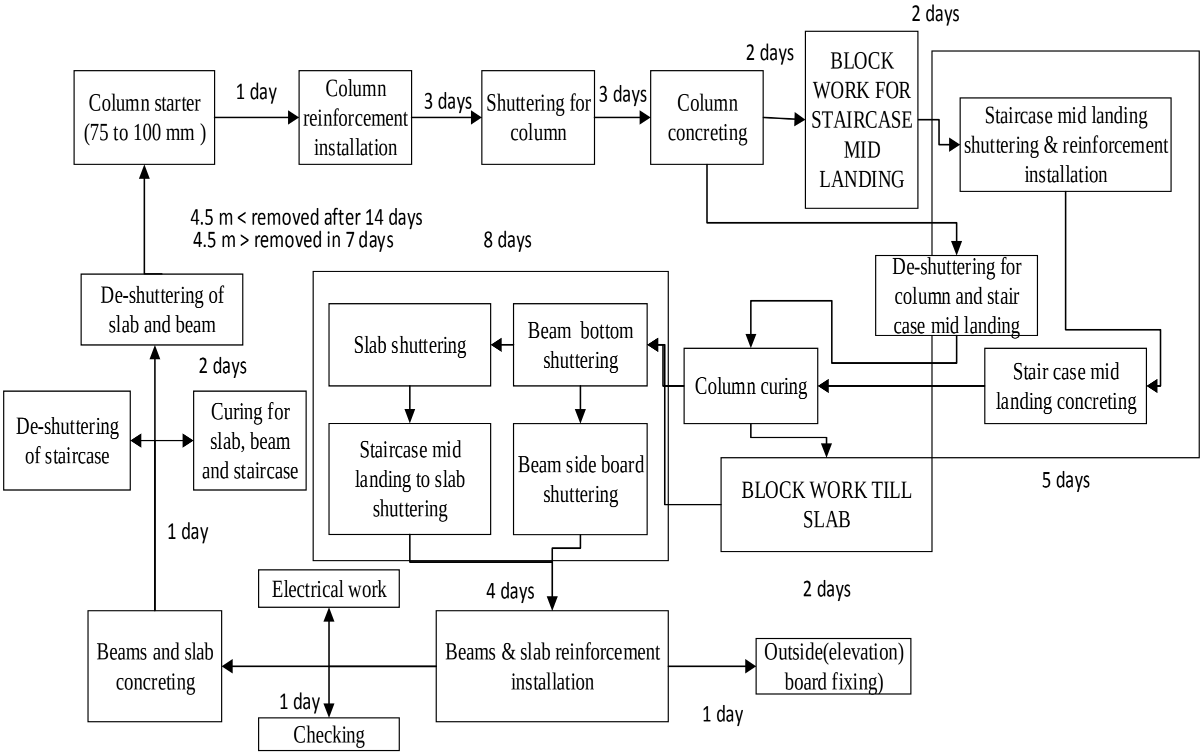


Figure 8: Construction Process in Case Study B

5.2.1 VSM for Column Execution Work

The Figure 9 represents the current practice of column execution work at the site. The time required for completing of one column is recorded which includes different activities like starter installation, reinforcement of column, fixing of shuttering board and finally with column concreting is recorded. In concreting process of column, method of placing concrete is done through labours due which the time consumption is more. To reduce cost and time a modified VSM is done.

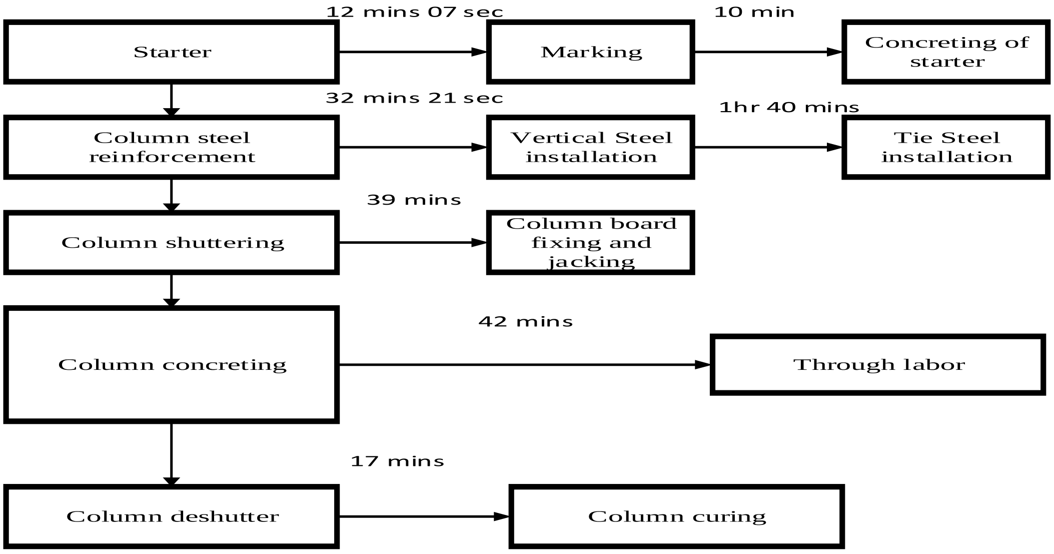


Figure 9: Current Practice of Column Execution Work

In Figure 10 the modified column execution work is mapped where the column starter is ignored and mapping is started from reinforcement of column by using clear cover block and check for plumb as it takes lesser time when compared to the current process and it is followed by column shuttering and concreting of column. It is observed that the modified column execution process take less time compared to current practice. The labour required for column reinforcement can be increased for speeding up the work and to complete the work in an average time. Column concreting work is done by labour takes long duration as compared to concreting work using a concrete pump and concrete placing is easier and time consumption is less by using concrete pumps. In this Case Study B the tower crane is unavailable because of less number of floors. So concreting of column is done using stationary concrete pump and it results in reduction of cost and time.

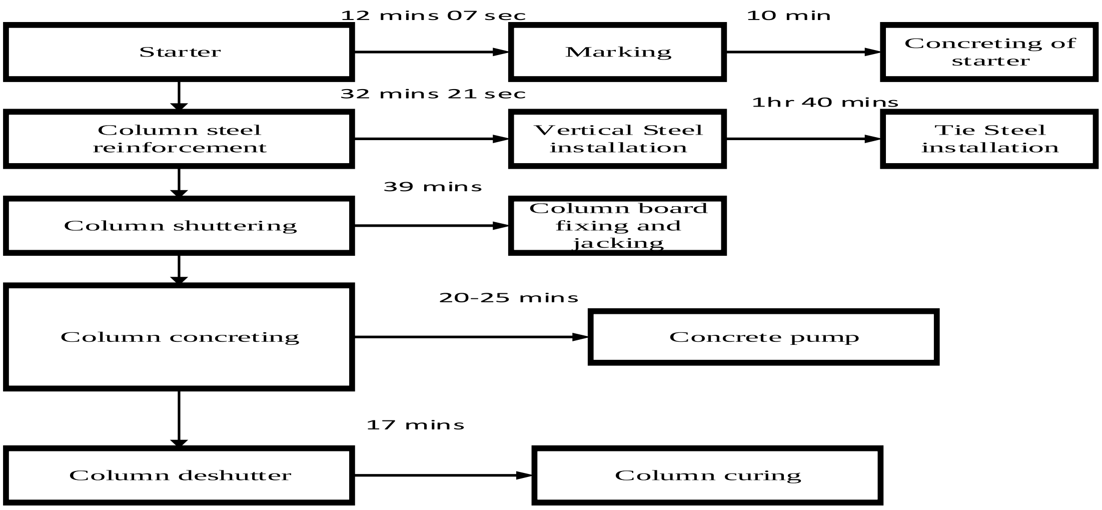


Figure 10: Modified Column Execution Work

5.2.2 VSM for Slab and Beams

The Figure 11 represents the actual process of slab and beams execution work carried out at the site where the block work is being finished till the beam bottom and then the beam bottom shuttering is done followed by the slab shuttering and reinforcement work. In this process the time taken to finish the block work depends on the number of labours working, it can speed up the work and finish the project on schedule. The site is located near residential area where traffic flow is more, so getting a boom concrete pump is difficult so it is better to use concrete pump for slab and beam concreting.

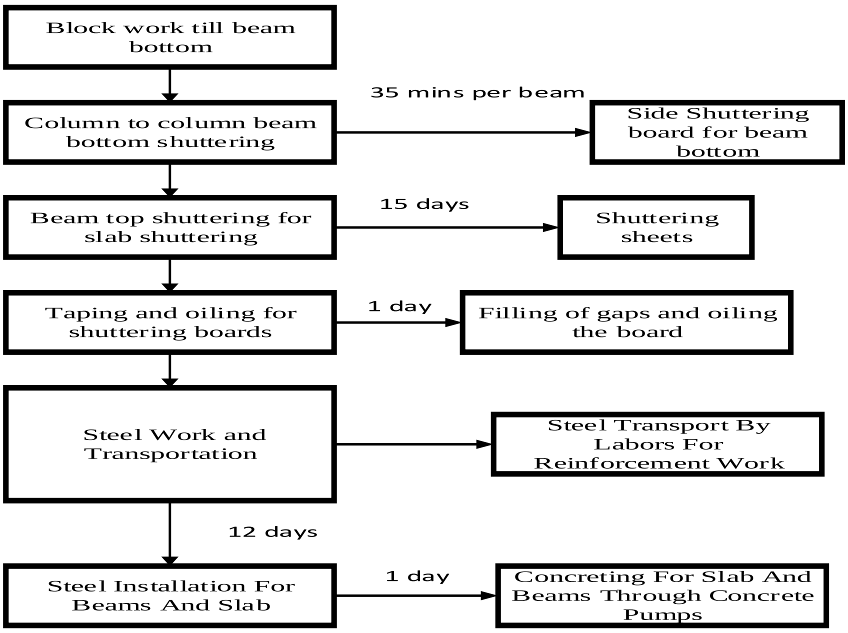


Figure 11: Current Practice of Slab and Beams execution Work

5.2.3 Cost for Concreting Process

The Table 3 represents the cost of current construction process which includes the labour cost for starter installation in column execution work. In the modified construction process the time taken for starter installation and the cost of the labour required for it is deducted in column execution work. Concreting is done through stationary concrete pump for columns, beams and slab concreting work which results in reduction of cost and time.

Table 3: Construction Cost in Case Study B

Sl no.	Equipment /Material	Quantity	Cost of Material /Equipment	Cost (in Rupees)	
				Current cost	**Modified cost**
1	Concrete	140 m^3	4800/m^3	6,72,000	6,72,000
2	Steel	14779.57 kg	52/kg	7,68,537	7,68,537
3	Bull dozer		900/day	54,000	54,000
4	Tippers		1000/day	60,000	60,000
5	Stationary Concrete pump		1,80,000/month	3,60,000	3,60,000
6	Concrete blocks		32/block	64,000	64,000
7	Manpower			1,58,400	1,02,460
Total cost (in Rupees)				**21,36,937**	**20,80,997**

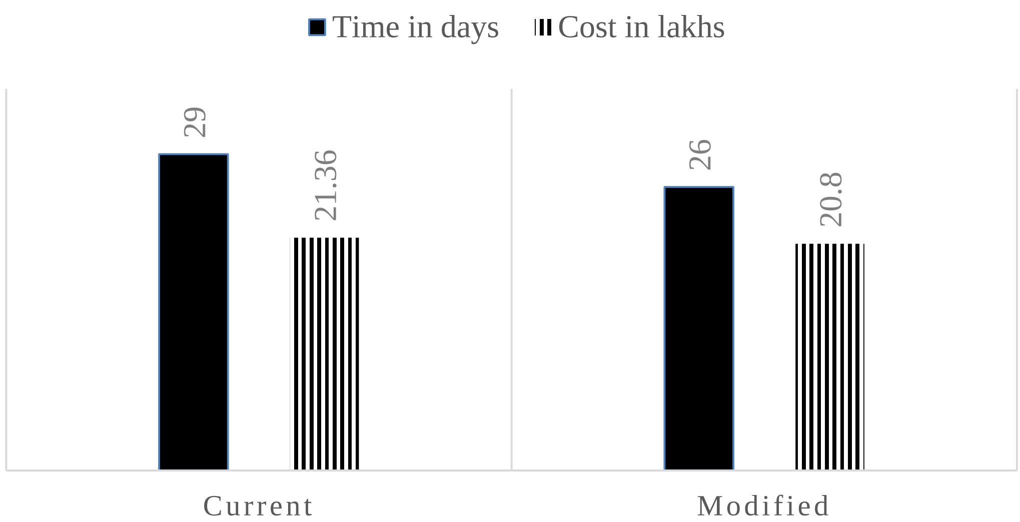


Figure 12: Comparison of Current & Modified Construction Time and Cost for Case Study B

The Figure 12 represents the comparison of the current and modified construction with respect to time and cost and the observations shows the reduction of time and cost. There is reduction in cost by 4% and time by 11% which results for efficient construction. In the present study cost and time are co-related to each other so before execution of the work proper planning is important and managing the planned schedule as per the plan leads to efficient construction.

5.3 Discussions and Comparison of Case Studies

In the two case studies it is observed that each construction sector have different construction procedure, in which the waste is taken as an important criteria with respect to time and cost. Even though the waste is generated in one way or the other, to avoid such waste in construction, implementation of new technology such as lean concepts is beneficial. It is observed that case study A and B follow different construction process. Each case study exhibits different methods in placing concrete for slab, beams and column. In case study A, to speed up the construction process, concrete pump and tower crane are used for placing of concrete. This can help in reducing the time and cost. Whereas in case study B where column concreting work is done by labours, more time is required which results in increase in cost. The shuttering material used for beams, columns and slab is either of wood or iron sheets. Iron sheets gets corroded quickly due to which it has less number of repetition is observed in case study A and B. Both the case studies are in different location, case study A is a residential construction project and it is located far from the city with large built-up area, whereas case study B is located in city where the traffic flow is more. Case Study B should consider regards with the work timing for labours while concreting as it disturbs the traffic flow and the resident located near the site. For each Case Study value stream mapping and future value stream mapping is carried out for column, beam and slab concreting work. It is observed that future value stream mapping reduces the time by 7% and cost on an average by 9% respectively. Using value stream mapping and lean concept not only identifies the constraint in a process but also results in efficient construction.

Comparison of the Case Studies is done with the current and modified state with respect to time and cost. The Case Study A which is a residential project where the use of the material is of good in quality as the company is well known in the construction sector where as in Case Study B which is a commercial project where the quality of the material is average compared with Case Study A, both the construction firm use same grade of concrete but different in quality. The difference in cost depends on the material used at the site i.e. quality and brand of concrete. The cost in the Case Study A is higher than the Case Study B where the volume of the concrete is less in Case Study A compared to Case Study B, the time between the two Case Studies A and B is five days in the current practice but in the modified process it has the difference of four days. The comparison of the Case Studies with current and modified time and cost helps in improvising the construction process for efficient construction.

6. Conclusion

Some of the major findings from the case study are listed below

- ✓ Construction firms generates various type of wastes such as extra processing, lack of system discipline, waiting time, transportation and lack of skills and technology use.
- ✓ Value stream mapping visuals the representation of waste generation.
- ✓ Value stream mapping and use of lean tool can improvise the process to reduce waste in the construction firm.
- ✓ With help of modified process the time duration of the process can be reduced by 7% on an average compared to current practice.
- ✓ Cost of the modified process reduces on an average of 9% from the current process.
- ✓ Use of equipment at construction can speed up the process which leads to minimization of duration, resulting in good quality.

The study, application on lean concepts in building construction has concentrated on the waste management with regard to time and cost, it also conveys that all site conditions does not witness the same cause of action. From the two case studies, it is observed that the techniques uses for the construction are different. The study is important to identify the time and cost consumption for the project implemented in the current state for which it can be improvised by implementing new technique and eliminating the wastes in each process of the construction. For any construction firm its main aim is to finish the project before schedule, which is not achieved in most of the construction firm due to lack of coordination, lack of management skills and shortage of labours in any project. If any construction project is to be started it should think of minimizing the waste by either time or cost. Using of technology in site with proper training for labours can make a construction firm more efficient. It is known that in any construction sector client, architects, engineers and contractors play a major role to run construction smoothly and there should always be a better relationship between architects, engineers and contractors. So that any change in design or detailing gives awareness among them. From the case studies it is known that equipment used at site can reduce efforts of labour and make work faster. A person should know how to reduce waste by proper planning and use of new concepts such has lean tool. The lean concepts are not only for the construction sector it can be used for any production firm. Thus it can reduce the time and cost which results in eliminating waste in building construction.

References

A. Genaidy; and I. Minkarah, M.ASCE, (2006) *Lean Construction: From Theory to Implementation* - O. Salem, M.ASCE; J. Solomon;.

Algan Tezel, Ph.D.; and Yasemin Nielsen (2013).*Lean Construction Conformance among Construction Contractors in Turkey*.

Thanveer M. Beary ; and Tariq S. Abdelhamid (2005) *Production Planning Process in Residential Construction Using Lean Construction and Six Sigma Principles.*

R. Sacks and M. Goldin (2007) *Implementing Lean Construction Theory into Construction Processes' Waste Management* - A. Abbasian Hosseini ; A. Nikakhtar ; *Lean Management Model for Construction of High - Rise Apartment Buildings*.

Chien-Ho Ko and Neng - Fu Chung (2014) *Lean Design Process.*

H. Randolph Thomas; Michael J. Horman, M.ASCE; R. Edward Minchin, Jr., M.ASCE; and Dong Chen, M.ASCE (2003) *Improving Labor Flow Reliability for Better Productivity as Lean Construction Principle.*

Gregory A. Howell, M.ASCE; Glenn Ballard, M.ASCE; Iris D. Tommelein, A.M.ASCE; and Lauri Koskela (2003) *Discussion of "Reducing Variability to Improve Performance as a Lean Construction Principle"* by H. Randolph Thomas, Michael J. Horman, Ubiraci Espinelli Lemes de Souza, and Ivica Zavřski.

Vitaliy Priven and Rafael Sacks (2015) *Effects of the Last Planner System on Social Networks among Construction Trade Crews.*

Xiaoming Mao and Xueqing Zhang (2008) *Construction Process Reengineering by Integrating Lean Principles and Computer Simulation Techniques*.

Luis F. Alarcón, Ph.D., A.M.ASCE; and Vicente A. González, Ph.D (2014) *Implementing Lean Production in Copper Mining Development Projects: Case Study - Gustavo Castillo*.

Ana Catarina V. M. F. Pestana, S.M.ASCE; Elaine Gilbert; and Farook Hamzeh (2016) *Lean Principles for the Management of Construction Submittals and RFIs* - Thais da C. L. Alves, A.M.ASCE.

Sergio Maturana; Luis Fernando Alarcón; Pedro Gazmuri; and Mladen Vrsalovic (2007) *On-Site Subcontractor Evaluation Method Based on Lean Principles and Partnering Practices.*

Bradley Hyatt (2014).*Evaluating the Implementation of Lean Construction into Construction Education*.

Isabelina Nahmens and Michael A. Mullens (2011).*Lean Homebuilding: Lessons Learned from a Precast Concrete Panelizer*.

An Novel Hand Gesture System for ASL using Kinet Sensor based Images

[1]Manoj H. M., , [2]Pradeep Kumar B.P., [3]Anil Kumar.C.[4] Rohith.S
, 2nd author's name and surname[2], etc
{manojhm@bmsit.in[1], pradi14cta@gmail.com[2], canilkumarc22@gmail.com[3], rohithvjp2006@gmail.com[4]}

[1] Assistant Professor, Dept of CSE, BMSIT&M, Bangalore-64., [2] Associate Professor, Dept of ECE, HKBKCE,Bangalore-45, [3] Associate Professor & HoD, Dept of ECE., RLJIT, Doddaballapur, [4]Associate Professor, Dept of ECE.NCET, Bangalore

Abstract. The prime goal of this study is to perform resolution enhancement using experimental analysis. The framework presents hand image resolution enhancement techniques based on multi scale decomposition and edge preservation smoothing. The proposed technique Dual Tree complex wavelet transform (DT-CWT) and edge preservation smoothing (EPS) algorithm images are decomposed into different sub bands and interpolated, after which sub bands are reconstructed to achieve the enhanced image. The study outcome of this phase is studied with respect to PSNR and RMSE for all the letters of ASL.

Keywords: American Sign Language, Camera, Gesture, Wavelet transform, kinet sensor.

1 Introduction

Gestures used for communicating between person and machines also between persons using sign language [5][6]. Gestures can be static (posture / certain pose) or dynamic (series of postures). The static-gestures require less computational complexity whereas dynamic-gestures are more complex. Various techniques have been developed for acquiring necessary information for gestures recognition system [4][6]. Some methods are utilized for external hardware-devices such as data-glove devices and colour-markers which can easily extract the comprehensive-description of Gesture-features. Other methods are based on the appearance of hand which segments the hand and extracts the essential features, these methods are considered as natural, easy and less cost effective than other [7][8].

For most hard of hearing people in the United States, American Sign Language (ASL) is the preferred dialect. For everyday terms, ASL employs approximately 6,000 gestures, with finger spelling for conveying dark words or structured objects, locations, or items. In ASL, communication is often based on available shapes placed in or transferred crosswise over various areas of the endorser's body, despite changes in the head and arm, as well as physical appearance [12]. In any case, proper names and words without a unified sign are spelled in English letter by letter, and ASL understudies often begin their studies by learning the 26 hand shapes that make up the manual letter range [12, 13].

I3CAC 2021, June 07-08, Chennai, India

DOI 10.4108/eai.7-6-2021.2308609

1.1 Hand Gesture Recognition System

The vision is one of the physical senses which computer is instantiated perceptibly during the communication with humans. Thus, the vision based mechanisms are considered more in HGR. The HGR system based on computer vision consists of three different steps and is shown in Figure.1. The typical process includes a) image enhancement to remove noises as a pre-processing step, b) segmenting the palm portion from the captured image of either human body or hand ,c) feature extraction and finally d) classification [14, 15].

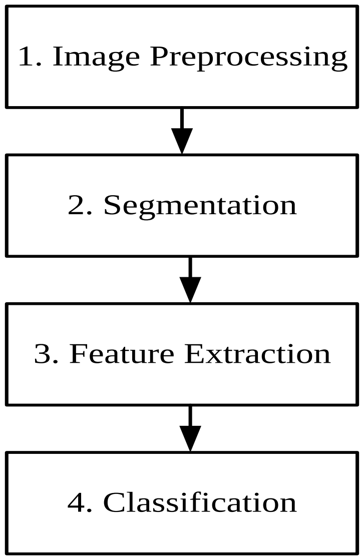


Figure.1. Main Steps of HGR System based on computer vision

1.2. Modeling Hand Gesture

The 2D modelling of hand can be mentioned with motion, shape and deformable templates. The shape based hand modelling can be classified as non-geographic and geographic models [2][3]. The non-geographic models consider the shape based features to model the hand like edges, contour, Eigen vectors etc, which are used for feature extraction and perform analysis too. The flexible/deformable models give an object shape changing flexibility level to pass the little variation of hand shape.

The 3D model of the hand can be represents and classified into skeletal, geometrical and volumetric models. The geometric models can be utilized in the real-time applications and hand animation. The skeletal models needs less parameters to structure the hand shape. The volumetric models are very complex and needs more parameters to shape the hand [14]. The geometric surfaces significantly performs the simulation of the visual hand image but it needs more parameters and is more time consuming process. The visual shapes like cylinders and ellipsoids are the alternative mechanism of geometric shape [15]. The Figure.2, describes the hand modelling mechanisms

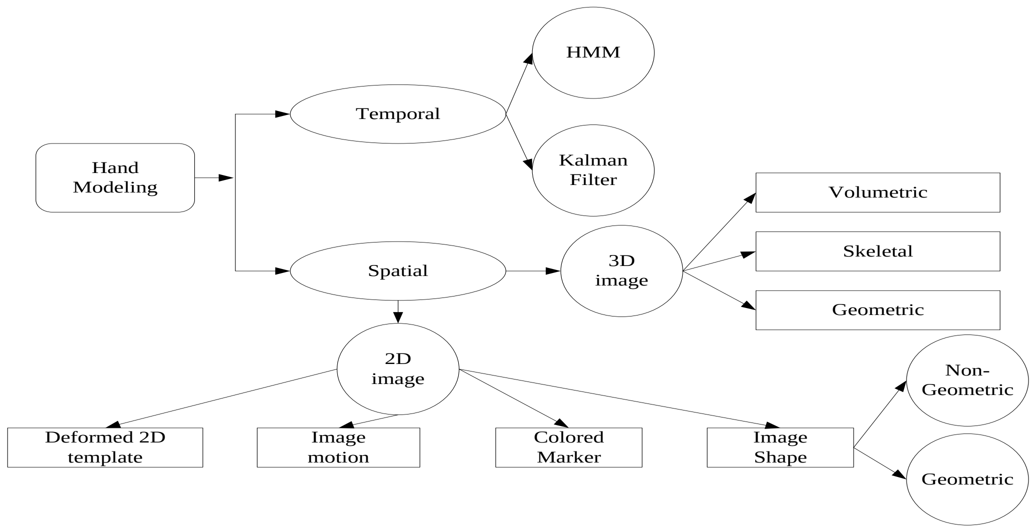


Figure.2. Hand Modelling

1.3 Research Aim And Objectives

A novel system for contact-less Hand gesture recognition using Microsoft Kinect for Xbox is described, and a real-time Hand gesture recognition arrangement is modelled and simulated into numerical computing platform (Matlab). The arrangement permits the user's to choose a situation, and it is clever to notice hand motions prepared by users. To recognize fingers, and to identify the meanings of gestures, and to show the meanings and pictures on screen. The prime aim of the proposed research work is to design a simple framework that can offer enhanced performance of hand gesture recognition system in effective manner considering ASL. In order to accomplish this research goal, following objectives were set:

- ***Preprocessing:*** To design a model that can offer enhanced resolution for input images
- ***Feature Extraction:*** To develop a simple modeling for hand gesture recognition emphasizing on an efficient feature extraction.
- ***Classification:*** To develop a hybridized scheme of hand gesture recognition for increasing recognition performance.
- ***Optimization***: To apply optimization for enhanced performance of hand gesture recognition system in cost effective manner.

2 Resolution Enhancement Of Hand Gesture Images

The hand gesture recognition (HGR) system which mainly emphasizes on the limitations of traditional hand gesture recognition techniques. This section mainly argues two-level architecture for the real-time hand gesture recognition scheme using only one camera as the input device. second describes the resolution enhancement problem and technique for hand gesture images using four different algorithms namely 1) The Nearest Value Algorithm, 2)

The Bilinear Algorithm , 3)The Bi-cubical Algorithm , 4) Dual Tree Complex Wavelet Transformation (DTCWT), further Bilateral Filter is used in order to obtain enhanced performance.

2.1 Skeleton Identification Of Kinect Camera

In this proposed system, “Kinect camera” plays the major role to gather the depth information from the skeleton. The new version of Kinect with its SDK (Software Development Kit) containing the skeleton tracking tool. This unique tool provides the system to collect the 20 joint information of the human body. For each frame, the positions of 20 points are estimated and collected. The 20 joints which is taken as an reference points is as shown in Fig 3

The kinect device provides the both RGB and D-image. This camera utilizes a structured light method to generate the real time depth information which consists of discrete measurements of physical scene. In this study, first creating the depth images of human in different sizes, and shapes, and generate the big dataset. The RGB and D-image are the input images of the system for the recognition of different ASL alphabetic symbols. This skeletal tracker device able to track the skeleton image of one or more persons moving within the kinect area view for gesture driven applications i.e. this tool able to collect 20 joint information about the skeleton. From the skeleton tracked by the Kinect first it extracts the feature of joint positions. Since, each joint has 3 values and also 3 coordinates and the detected skeleton has of 20 joints. So, the feature vector has 60 dimensions. The position of 20 points identified by kinect afterwards it will segment the right hand posture to recognize and stored in the database. From this sets can select the wanted joints of images for representing the postures.

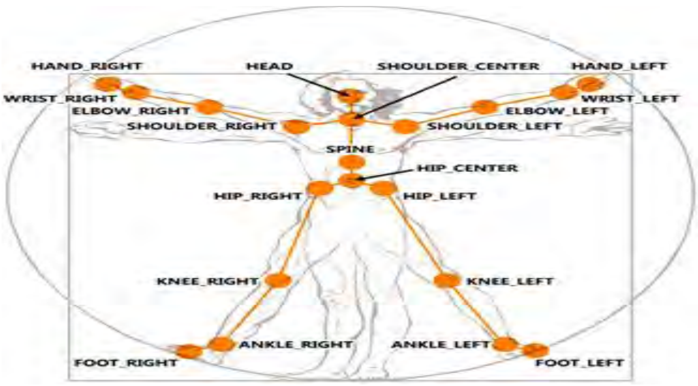


Figure. 3 Human skeleton joints as reference points

2.2 Algorithm: Skeleton identification from Kinect-Sensor.

Input: one or more people image,   Output: 20 joint images

a. start
b. capture the image from kinect camera
c. segment the image from skeleton viewer
d. for reference points in body portion finds the depth-information
e. segmented moving body portion is mapped to the skeletal co-ordinates
f. if more than one moving body portion presents
g. calculate skeleton connection map for x,y coordinate
h. display the multiple skeleton moving body
i. end of predefined frames
j. stop.

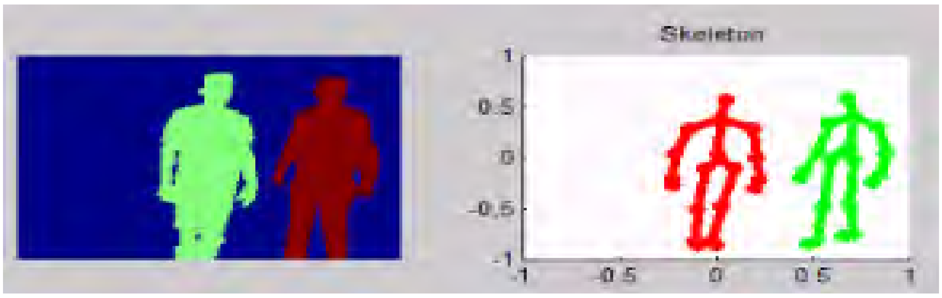


Figure.4. Skeleton-image identification from kinect sensor

Figure 4 shows the outcomes of identifying multiple skeletal signs under different distance using kinet-sensor device which represents the colour and depth image. Figure 5 shows in multiple user environments also our proposed system is identifying the user hand with respect to the distance from camera and extract the hand sign clearly for storing in to database.

Figure. 5. Multiple skeletal sign recognized under different distance

2.3. Multiple Depth Recognition:

(i) To calculate centroid: Here system is considering the three co-ordinates X, Y and Z. Calculating the centroid for each axis independently the mathematical interpretation for centroid is given by

X= sum (x) / length(x)..................... ....… 1

Y= sum (y) / length (y)………………………………….....2

Z= sum (z) / length (z)………………………….............3

(ii) **To calculate mean:** The system will calculate the mean for the entire segmented region calculated by the centroid of the body part by pixel basis. The mathematical interpretation for mean for pixels is given by:

$\bar{x}=\frac{1}{K}\sum_{i=0}^{K} x(i)$ and $\bar{y}=\frac{1}{K}\sum_{i=0}^{K} y(i)$……………………......4

The flow graph in figure 6 shows for the multiple depth recognition using Kinect sensor is as follows

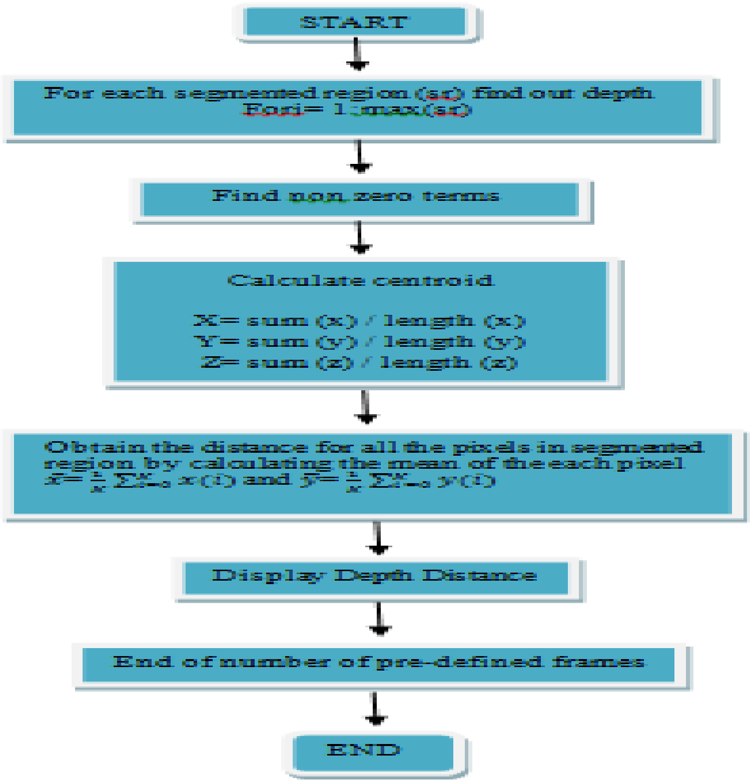


Figure. 6 The flow graph for the multiple depth recognition

2.4. Data Acquisition and Pre-Processing

The raw information acquired from the Kinect sensor via the Natural User Interface (NUI) contained 512×424 depth data, 1920×1080 RGB data, and 26-joint body skeleton data. The hand region in the depth image was illustrated using spatial thresholds in X-axis direction [Tx_min, $Tx\_$], Y-axis direction [Ty_min, Ty_max] and Z (depth)-axis direction [$TDepth\_min$, $TDepth\_max$]. As demonstrated in Figure 6, the Kinect depth sensor placed at position $\boldsymbol{S}$ has angles of view α (horizontal) and β (vertical). The declaration of the depth image is $RxbyRy$ pixels. The position of the “hand” joint (x, y, D) in the depth image can be attained from the Kinect skeleton data (Figure 8a). Thus, the spatial thresholds are illustrated as:

.

$$[T_{x\_min}, T_{x\_max}] = [x - \frac{d_x}{2}\frac{\frac{R_x}{2}}{Dtan\frac{\alpha}{2}}, x + \frac{d_x}{2}\frac{R_x/}{Dtan}$$ ……………………………….…...5

$$[T_{y\_min}, T_{y\_max}] = [y - \frac{d_y}{2}\frac{R_y/2}{Dtan\frac{\beta}{2}}, y + \frac{d_y}{2}\frac{R_y/}{Dtan}$$ ……………………………….…...6

$$[T_{Depth\_min}, T_{Depth\_max}] = [D - \frac{d_z}{2}, D + \frac{d_z}{2}]$$ ……………………………….…...7

where dx , dy and dz are stable dimensions (in millimetres) of the hand's region. The hand's region in the depth image is revealed in Figure 7. The hand's region in the color image can also be gained by mapping the hand's region on top of the color image (Figure 8 a).

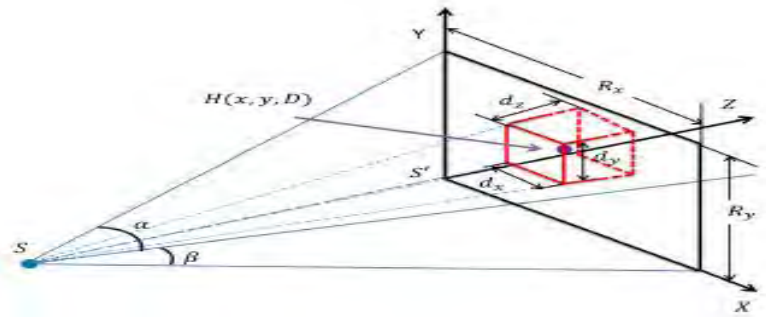


Figure. 7.. Illustration of the hand region segmentation: the $dx \times dy \times dz$ hand region at
(x, y, D) was segmented from the $Rx \times Ry$ depth image obtained using a depth sensor located at the position $\boldsymbol{S}$.

(a) (b)

Figure. 8. Illustration of data obtained using Kinect.
(a) RGB Color image of the hand region. (b) Depth image of the hand region.

The following table illustrates the result analysis of proposed system can measure the recognition accuracy in different distance ranges. As sample we are experimenting skeletal signs as 10 times with different distance like 850mm to 1000mm …3000mm to 3500 mm and evaluating the recognition accuracy with time. The recognition accuracy is calculated in terms of percentage like for first experiment we are considering the (skeletal and camera) distance as 850mm to 1000mm and getting the 70% of recognition accuracy. Like this from the experiment analysis results, we can get the following results, which is shown is following table 1.

Table. 1 analysis results for recognition accuracy calculation

Distances from Kinect In mm	Number of times checked	Number of times recognised	Recognition Accuracy in %
850-1000	10	7	70 %
1000-1500	10	10	100 %
1500-2000	10	10	100 %
2000-2500	10	10	100 %
2500-3000	10	10	100 %
3000-3500	10	8	80 %

3. Framework For Resolution Enhancement

In this research methodology using experiential analysis image enhancement is executed as shown in fig 9. The techniques used here are Edge preservation smoothing and multi-scale decomposition. To obtain the enhanced image, the method of DT-CWT and EPS algorithm images are decomposed to a different sub band. At a regular interval of 5 minutes, samples of sign language are recorded from Kinect camera at a distance of 1500-2000mm. Values of PSNR, RMSE, are used to further proceed with quantitative analysis. The output of this phase is an enlarged image.

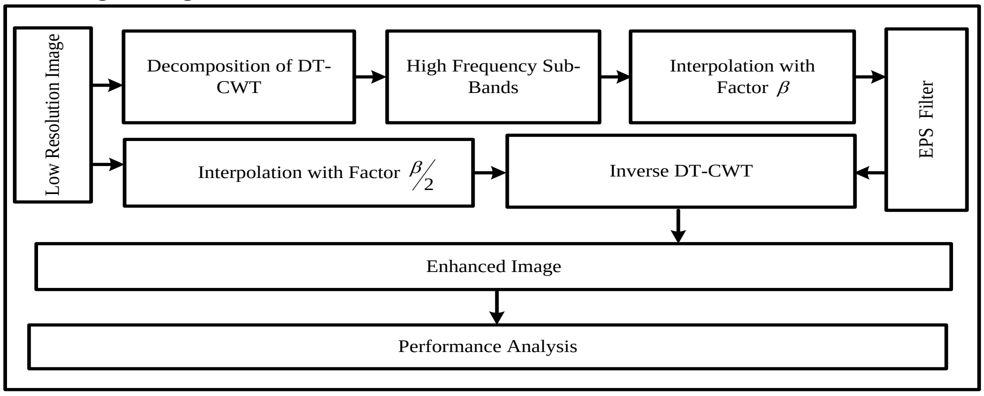


Figure.9. Block diagram representing framework for resolution enhancement

3.1. The Nearest Value Algorithm

Nearest Neighbour interpolation or proximal interpolation is way by which multiple dimensions can be interpolated. For a random point in space surrounding the nearest point (Neighbouring) would have its value approximated leading to the interpolation problem. Hence, the algorithm of the nearest neighbour selects the nearest point and neglects the value of points around it.

The above mentioned is the nearest neighbour algorithm which selects the nearest point eliminating the value associated to the random points around it. The original disk on file (I) is initialized and the size of the resized image is computed (I_{new}).Now, the size of the original file on disk is calculated. Further to check a condition that size of I and I_{new} match or not the number of rows and columns are compared respectively. If the resized image has higher number of rows then its new value will be ranging from 1-to-rows of original image considering first row and first column. If the original image has higher number of rows then its new value will be ranging from 1-to-rows of resized image considering first row and first column. Same procedure is applied to check the condition for the columns. The value of PSNR is calculated to be 23.3591, RMSE is 17.3214 for the input image as shown in fig 10.

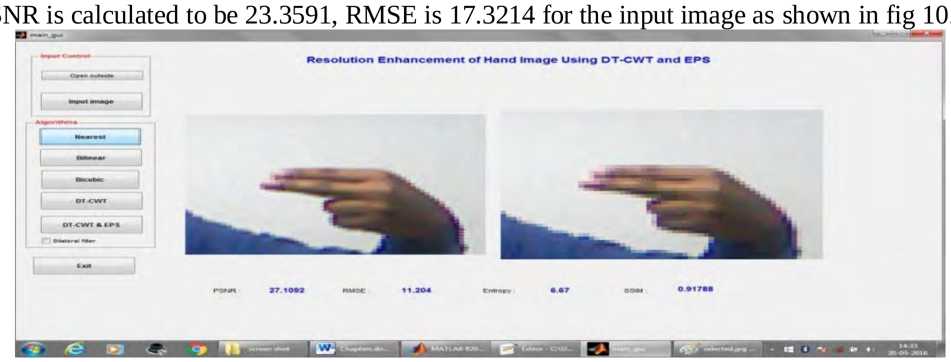


Figure.10. Application of Nearest Neighbor Transformation Function

3.2. Bilinear Interpolation

An image transformation process used in cases where pixel matching is impossible is called as bilinear interpolation. When compared to other methods of transformation bilinear interpolation considers closet 2x2 neighbourhood of known pixel values surrounding the unknown pixel's computed location.

Computation of PSNR $(\beta) = 10\log_{10}\left(\frac{255}{\alpha}\right)$; Computation of RMSE $(\gamma) = \sqrt{MSE}$

The above mentioned is the bilinear interpolation method which considers 2x2 neighboring pixels. The original disk on file (I) is initialized and the size of the resized image is computed (I_{new}).Now, the size of the original file on disk is calculated. Further to check a condition that size of I and I_{new}match or not the number of rows and columns are compared respectively. If the resized image has higher number of rows then its new value will be ranging from 1-to-rows of original image considering first row and first column. If the original image has higher number of rows then its new value will be ranging from 1-to-rows of resized image considering first row and first column. Same procedure is applied to check the condition for the columns. The value of PSNR is calculated to be 23.591, RMSE is 17.3214 for the input image as shown in fig 11.

Figure. 11. Application of bilinear Transformation Function

3.3. Bicubic Interpolation.

This technique is implemented using Lagrange polynomials cubic splines or cubic convolution algorithm. Bicubic interpolation can be chosen over other methods if speed is not a constraint. Smoother images are obtained as output with lesser interpolation artifacts. The above mentioned is the bicubic interpolation method which is based on cubic convolution algorithm, Lagrange polynomials. The original disk on file (I) is initialized and the size of the resized image is computed (I_{new}).Now, the size of the original file on disk is calculated. Further to check a condition that size of I and I_{new}match or not the number of rows and columns are compared respectively. If the resized image has higher number of rows then its new value will be ranging from 1-to-rows of original image considering first row and first column.If the original image has higher number of rows then its new value will be ranging from 1-to-rows of resized image considering first row and first column. Same procedure is applied to check the condition for the columns. The value of PSNR is calculated to be 24.953, RMSE is 17.3214 for the input image as shown in fig 12.

Figure.12. Application of bicubic Transformation Function

3.4. Bilateral Filtering Process

Bilateral filter is a noise reducing filter for images with the property of non-linearity and preserving the edge. Each pixel has an intensity value a picture restored by weighted average from nearby pixels. Weight can be based on Gaussian distribution and the sample output is shown in fig 13.

The bilateral filter is defined as:

$$I^{filtered}(x) = \frac{1}{W_p}\sum_{xi\in\Omega} I\,(x_i)\; fr\,(\|\, I(x_i)- I(x)\,\|)g_s(\|x_i-x\|)\ldots\ldots\ldots 8$$

Where the normalization term;

$$W_{p} = \sum_{xi \in \Omega} fr(\| I(xi) - I(x) \|) gs (\| xi - x \|) \quad \text{.......9.}$$

Ensures that the filter preserves image energy and

- $I^{filtered}$ is the image after filtration: Iis the original input image
- xare the directs of the current pixel to be filtered: Ω is the window centered in x
- f_r is the range kernel for smoothing differences in intensities. This utility can be a Gaussian function
- $g_s$is the spatial kernel for smoothing differences in coordinates. This function can be a Gaussian function.

	
Input Image for Bilateral filter	Output Image for Bilateral filter

Figure 13.Application of bilateral filtering process

3.5. Dual Tree Complex Wavelet Decomposition Transformation (DTCWT)

Using the dual tree complex wavelet transformation method the image is decomposed. This happens with respect to Discrete Continuous Wavelet Transformation (DWT) and Continuous Wavelet Transformation (CWT) .In DWT the basis function used is symlet mother wavelet. Image will be decomposed into two parts, the approximation coefficients and detailed coefficients. Only approximation coefficients are considered and similar mechanism is implemented for CWT. The algorithm of DTCWT is being performed above for the resolution enhancement for a hand gesture by splitting the image mainly into real and imaginary parts. To evaluate the analysis and synthesis parameter taken into consideration, the dual filter function is worked on. The dual cell structure is determined via cplx dual 2D function. Frequency values are divided into lower and higher components. The higher frequency components are normalized to nullify the effect of frequencies lying outside the desired range of detection. The lower frequencies have their highest value used in the algorithm among all of them. Further lower frequency image is converted into original image with the inclusion of the new dual tree cell structure. for the input image as shown in fig 14.

Figure.14. Application of DTCWT and EPS algorithm

Table.2. Comparison of algorithms for sign H

Algorithms	PSNR	RMSE
Nearest	26.24	12.37
Bilinear	28.08	10.01
Bicubic	28.11	10.11
DT-CWT	28.50	13.8
DT-CWT & EPS	29.07	10.006

Table 3 showing the detailed tabulated parametric values for all the signs using DT-CWT conditions respectively considering performance parameters of PSNR, RMSE

Table.3 Numerical Outcome of Performance parameters of DT-CWT

ALPHABET	A	B	C	D	E	F	G	H	I
PSNR	28.5	29.61	30.35	28.96	29.02	29.14	28.84	26.24	30.77
RMSE	13.8	8.66	8.63	10.49	10.9	10	12.38	12.37	8.44

ALPHABET	J	K	L	M	N	O	P	Q	R
PSNR	28.82	29.45	28.45	28.33	29.43	29.77	30.12	29.25	29.43
RMSE	12.15	12.15	11.15	9.6	11.56	10.32	7.85	11.9	8.43

ALPHABET	S	T	U	V	W	X	Y	Z
PSNR	29.59	29.89	30.45	27.78	27.32	27.55	28.56	26.72
RMSE	8.8	8.99	9.2	12.73	9.34	13.92	12.44	14.6

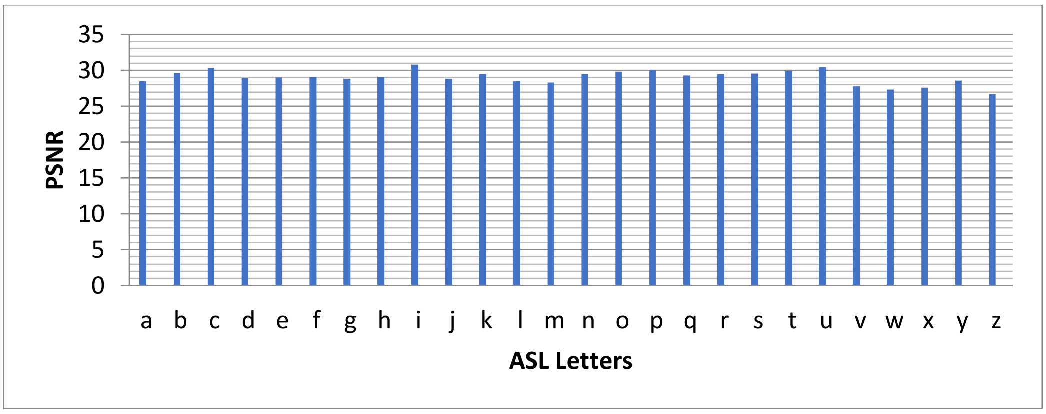


Figure.15. Illustration of PSNR values for the algorithm DT-CWT and EPS.

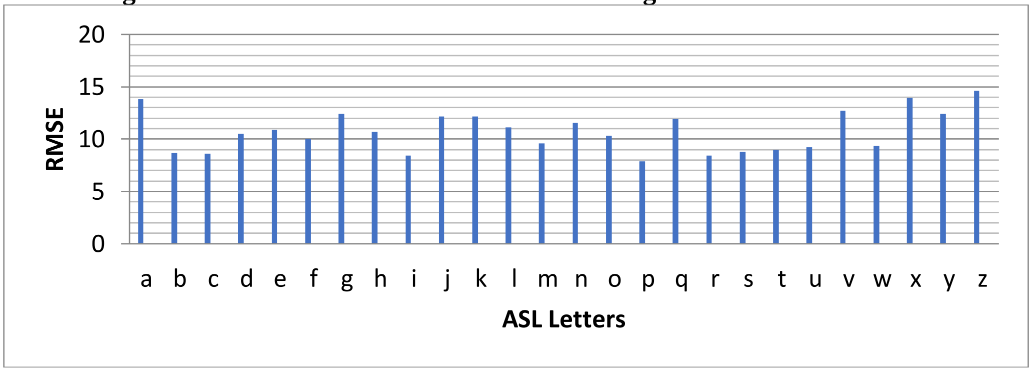


Figure.16. Illustration of RMSE, values for the algorithm DT-CWT and EPS.

4. Conclusion.

Thus, this study quickly abridges about every thing of the calculations being executed for the upgrade of a hand signal acknowledgment framework alongside the examination procedures required behind it. A novel picture determination improvement procedure in view of DT-CWT and EPS channel. The method breaks down the LR input picture utilizing DT-CWT. EPS (Bilateral) sifting is utilized to safeguard the edges and de-noising the picture and to additionally improve the execution of the proposed method as far as RMSE, PSNR

References

[1] A.R. Sarkar, G. Sanyal, and S. Majumder, "Hand gesture recognition systems: a survey", International Journal of Computer Applications, 71.15, 2013

[2] N. Neverova, Natalia, "A multi-scale approach to gesture detection and recognition", Proceedings of the IEEE International Conference on Computer Vision Workshops, 2013

[3] S. Yang, P. Premaratne, and P. Vial, "Hand gesture recognition: An overview", Broadband Network & Multimedia Technology (IC-BNMT), 5th IEEE International Conference, 2013

[4] T. Osunkoya and J-C. Chern, "Gesture-based human-computer-interaction using Kinect for Windows mouse control and Powerpoint presentation", Department of Mathematics and Computer Science, Chicago State University, Chicago, IL 60628, 2013

[5] J. Katkar, Jayshree, "Hand Gesture Recognition and Device Control", Hand, 2017
[6] J.G. Kyle and B. Woll,“Sign language: The study of deaf people and their language”, Cambridge University Press, 1988
[7] M.C. Thomas and A. P. M. S. Pradeepa,“A comprehensive review on vision based hand gesture recognition technology”, International Journal 2.1, 2014
[8] H. Zhouand H. Hu,“A survey-human movement tracking and stroke rehabilitation”, University of Essex, Colchester United Kingdom, 2004
[9] M. Turkand M.Kölsch,“Perceptual user interfaces”, Emerging Topics in Computer Vision, Prentice Hall, 2004
[10] K.K. Vyas, A. Pareek, and S. Tiwari, "Gesture Recognition and Control", International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169, Retrieved on 16th August, 2017
[11] D.H. Stefanov, Z. Bien, and W-C. Bang,“The smart house for older persons and persons with physical disabilities: structure, technology arrangements, and perspectives”, IEEE transactions on neural systems and rehabilitation engineering, vol.12.2, pp.228-250, 2004
[12] D.C. L-Martin, “Universal grammar and American sign language”, Universal Grammar and American Sign Language, pp.1-48, 1991
[13] M.L. McIntire, “The acquisition of American Sign Language hand configurations”, Sign Language Studies, 16.1, pp.247-266, 1977
[14] M. Alsheakhali, “Hand gesture recognition system”, Computer Engineering Department, The Islamic University of Gaza, Gaza Strip, Palestine, 2011
[15] S.D. Badgujar, “Hand Gesture Recognition System”, International Journal of Scientific and Research Publications, 4.2, 2014

Sponsored By

GLOBAL CONFERENCE ON
ADVANCED SMART & SUSTAINABLE TECHNOLOGIES IN ENGINEERING

GCASSTE-2020

Organized by

MANGALORE INSTITUTE OF TECHNOLOGY & ENGINEERING

(An ISO 9001:2015 Certified Institution, Accredited by NAAC)

(A Unit of Rajalaxmi Education Trust®, Mangalore)

PAPER PRESENTATION CERTIFICATE

This is to certify that **Vidyashree M** has presented a paper entitled ***Experimental Studies on Structural Characteristics of Solid Concrete Block Masonry with Partial Replacement of Sand by Quarry Dust in Mortar*** in the Global Conference on Advanced Smart & Sustainable Technologies in Engineering held at MITE, Moodabidri, India on January 30th & 31st, 2020.

Rajashekar

Dr. Rajashekar C. R
Co-ordinator

Dr. G L Easwara Prasad
Principal

Fwd: Weekend Curfew impose | Criterion - III - Google Sheets | Recent - Google Drive | (4) WhatsApp | Efficient Feature Extraction fo

sciencepubco.com/index.php/ijet/article/view/19471

Apps | Reading list

Home | About | Login | Register | Journals

Browse | Current Issue | Archives | Join Us | Conferences | Contact Us

ARTICLE TOOLS

- Print this article
- Indexing metadata
- How to cite item
- Finding References
- Email this article (Login required)
- Email the author (Login required)

ABOUT JOURNAL

- Aims and Scopes

Home > Vol 7, No 3.34 (2018) > Channagoudar

Efficient Feature Extraction for Face Recognition with Combined Method of PCA and GMM

Savita Channagoudar, Dr Srikanta Murthy k

Abstract | Keywords | References | PDF

Abstract

In the course of recent decades, various face recognition techniques have been proposed in PC vision, and the majority of them utilize all encompassing face pictures for individual ID. In some true situations particularly some unconstrained conditions, human appearances may be impeded by different articles, and it is hard to acquire completely all encompassing face pictures for acknowledgment. To address this, we propose another halfway face recognition way to deal with perceive people of enthusiasm from their fractional appearances. Given a couple of exhibition picture and test confront fix, we

2:50 PM
6/29/2021

3.4.4 - Books or Chapters Publis ×   A SYSTEMATIC AND COMPOSED ×   +

Not secure | ijarcs.info/index.php/ijarcs/article/view/6250

Apps   Download latest Hi...   Blogs about OA - O...   185.105.101.95 - /S...   Bibliography.com -...   YU YU5040 (Chrom...   Download Torrents   BERT for dummies...   SemEval 2014 Task...   Untitled Document...   Reading list

International Journal of Advanced Research in Computer Science

A Peer Reviewed, Bimonthly, Online Journal

E-ISSN:0976-5697

Starting Year: 2010

HOME   ABOUT   LOGIN   REGISTER   SEARCH   CURRENT   ARCHIVES   EDITORIAL BOARD   CONTACT US

Home > 2018 > Rajan

A SYSTEMATIC AND COMPOSED BIG DATA ENTRY RESTRICTION SCHEME WITH ISOLATION-PRESERVING POLICY

Abraham Rajan, Venkatesh Prasad

ABSTRACT

: In modern world of technologies, each and every object like smart phones, computers connected to internet generates large amount of data. It becomes a very challenging issue for data to be stored in structured formats, specifically when it is stored to storages such as Cloud. It is also challenging that the privacy of data is maintained. We have an encryption technique known as Cipher text-policy attribute-based encryption (CP-ABE) which can be used by the users to encrypt their own data using attribute values which is defined over some access policies. If the attribute values of data consumers are matched with access policies of the data owners then such users are allowed to decrypt the data. In CPABE, we have access policies attached to the encrypted data in plain text formats which may contain some private information regarding the end-users. These methods will only partially conceal the private information while the attribute values are still exposed. In this paper, we propose data access control which also ensures privacy of the data owners and data consumers. And also we have bloom-filter which is used for attribute based decryption. It is used to assess whether the attribute is in the access policies defined by end-users. It can also find the exact location of the attribute if it is present in the access-policy. It has been assessed by many security analyst and execution performance evaluators that our proposed technique can prevent private information from any linear secret-sharing strategic access policies without engaging much overhead. In our project we provide two login constraints, one as data owner and the other as data consumer. The data owner uploads a file and generates a tag number for each file once the encryption is done. This will protect the attribute values such as file name from leaking. File name may contain some information regarding the attribute values which can be used by the intruders to decrypt the files. Therefore, by providing tag number it is able to protect privacy of the data owners.

KEYWORDS

AUTHOR DESK

Submission Schedule
Call for paper
Author Guideline
Manuscript Template
Copyright Form
Publication Fee
Indexing
Impact Factor

OUR ASSOCIATE JOURNAL

Journal of Global Research in Mathematical Archives
Journal of Global Research in Computer Science

OPEN JOURNAL SYSTEMS

JOURNAL HELP

USER

Username
Password
Remember me
Login

Waiting for www.ijarcs.info...

Type here to search   30°C Sunny   ENG   10:40 01-01-2022

CiiT bringing the world locally
Programmable Device Circuits and Systems

IOT, GPRS, Raspberry PI Based Global Industrial Process Monitoring

Abraham Rajan, Venkatesh Prasad

Abstract

In modern life of technology there is lot of scope for automation which leads to more use of modern software, hardware and sensors. Usage of many big hardware and software equipment leads to power consumption. We all know power is not a natural source of energy, this is why there is need to monitor and control the system to minimize the power consumption. The Internet of Things (IoT) is an will be a technology of future industries where lives of people will be dependent on Internet Of Things for day today activities. Majority of people in the world are connected to internet and it has become very easy and natural to use smart phones. Using smart phones users can monitor their health using smart healthcare mobile apps, book a ride to home, control remote machine etc. which is making peoples life easier. With the development of infrastructure such as high speed railways and smart cities, more and more safety infrastructure needs to be monitored. In the meantime, a lot of old infrastructure such as pipelines, bridges and buildings needs to be monitored for lifecycle assessment, improving safety and security. New infrastructure is required for low cost reliable monitoring. Here we are creating an android application which receives all data from hardware and displaying it, and we providing buttons to control the bulb and motors and that storing in database, for that we providing two type of login's (manager login, senior manager login). And that received data storing in database by using web server concept. Once the manager login he can capable to see the related information in his time period, and other one senior manager login where he can see all stored information from database.

Keywords

Monitoring, Controlling, Internet of Things (IOT), Android Application, Database.

Full Text:

PDF

References

Alfredo Gardelvicente, Ignacio Bravio Munoz Jose Luis LazaroGalilea and Pedro A.Revengadel Toro, "Remote Automation Laboratory Using a Cluster of Virtual Machines," IEEETransactions on Industrial Electronics, vol. 57, no. 10,pp. 3276-3283, 2010.

Fwd: Weekend Curfew impose | Criterion - III - Google Sheets | Recent - Google Drive | (4) WhatsApp | Computer Peripherals and its

amazon.com/Computer-Peripherals-its-Organization-Handbook/dp/B086P9BD9Z

Apps | Reading list

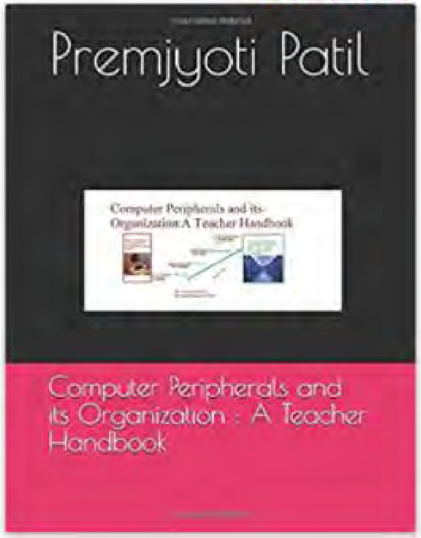
amazon Deliver to India | Books | Hello, Sign in Account & Lists | Returns & Orders | 0 Cart

All | Today's Deals | Customer Service | Gift Cards | Registry | Sell | Amazon's response to COVID-19

Books | Advanced Search | New Releases | Best Sellers & More | Children's Books | Textbooks | Textbook Rentals | Best Books of the Month

Books › Education & Teaching › Higher & Continuing Education

Look inside



Computer Peripherals and its Organization : A Teacher Handbook Paperback – April 4, 2020

by Dr. Premjyoti G Patil (Author), Dr Shantakumar B Patil (Author)

See all formats and editions

Paperback
$16.00

1 New from $16.00

The authors have seen the complexity of the subject as a student as well as a teacher. In the last 17-18 years they are teaching this subject to the different categories of students. In this journey they understood the pulses of the students. Keeping this in mind, the authors published a book called Computer Peripherals and its Organization: A Teacher Handbook. An imagination animated into an image to understand the subject easily and effectively. Numerous problems are solved with color images to avoid the confusion and complexity of the problems and create interest among student community towards the subject.

Language | Publication date | Dimensions

Buy new: $16.00

+ $17.37 Shipping & Import Fees Deposit to India Details

Arrives: **July 13 - 30**

Deliver to India

In Stock.

Qty: 1

Add to Cart

Buy Now

Secure transaction

Ships from Amazon.com
Sold by Amazon.com

Return policy: This item is returnable

Add a gift receipt for easy returns

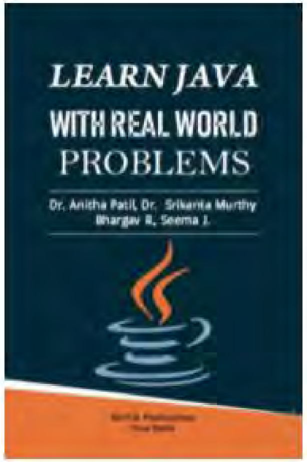
Waiting for fls-na.amazon.com...

2:52 PM
6/29/2021

Fwd: Weekend Curfew impose | Criterion - III - Google Sheets | Recent - Google Drive | (4) WhatsApp | Learn Java with Real World Pr

akinik.com/products/891/learn-java-with-real-world-problems

Apps | Reading list



Learn Java with Real World Problems

₹ 990

Author(s)	: Dr. Anitha Patil, Dr. Srikanta Murthy K, Mr. Bhargava R and Mrs. Seema J
ISBN	: 978-93-89680-89-8
Publisher	: AkiNik Publications
Language	: English
Pages	: 341
Publication Year	: 2020
Binding	: Paperback
Seller	: AkiNik Publications

PRICE DETAILS

AMOUNT	₹ 990
AVAILABILITY	IN STOCK
PAYABLE AMOUNT	₹ 990

ADD TO CART

Conferences > 2017 International Conference...

A novel approach for detection of breast cancer at an early stage using digital image processing techniques

Publisher: IEEE

R. Sangeetha ; K. Srikanta Murthy   All Authors

Abstract

Abstract:

Breast Cancer is highly predominant in women in today's world. It can start in the breast and can spread to ...her areas of the body in the course of time. Breast cancer is the second largest disease leading to the ... curable if detected early enough. A lot of research is being done to detect

Fwd: Weekend Curfew impose | Criterion - III - Google Sheets | Recent - Google Drive | (4) WhatsApp | A novel approach for detectic

ieeexplore.ieee.org/document/8226198

Apps | Reading list

A novel approach for detection of breast cancer at an early stage by identification of breast asymmetry and microcalcification cancer cells using digital image processing techniques

Publisher: IEEE

Cite This   PDF

R Sangeetha ; K Srikanta Murthy   **All Authors**

2:47 PM
6/29/2021

As per New CBCS Syllabus for 1st Semester, BBA, Bangalore University w.e.f. 2020-21

7 Lecturers'

ACCOUNTING FOR BUSINESS

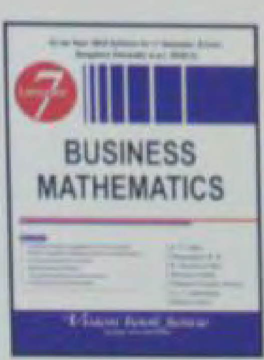
INCLUDES

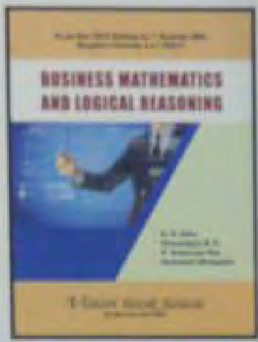
- Practical Problems highlighted with the concept
- Review Questions highlighted with year identification
- Free solution book in e-format
- Model Question Papers
- Comprehensive for student learning
- A lecturer friendly presentation

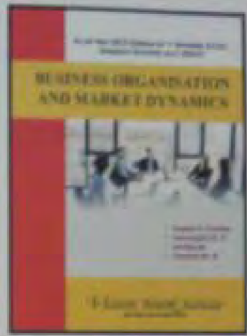
R. G. Saha
Raghavendra B. N.
Vidya Shivannavar
Zameer Pasha
Chaithra N.
Shalini Rawani
Geethanjali G.

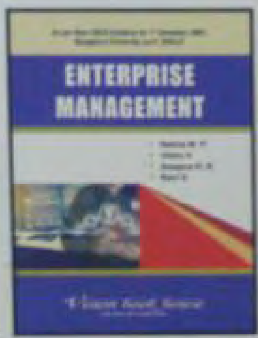
Vision book house
ISO 9001:2015 CERTIFIED

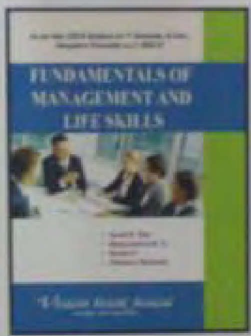
OUR OUTSTANDING PUBLICATIONS

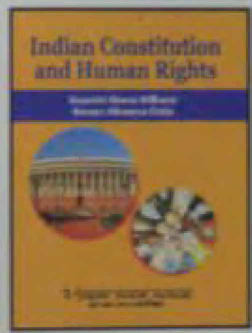












ISBN: 978-93-90515-10-3

ISBN: 978-93-90515-10-3 PAB 0166 ₹ 225/-

Vision Based Surveillance System for Detection of Human Fall

Basavaraj G M
Department of Electronics and communication , Nagarjuna College of Engineering and Technology,
Bangaluru, 562164 ,Karnataka, India
baswaraj832@gmail.com

Ashok Kusagur
Dept. of Electrical and Electronics Engineering, University BDT College of Engineering, Davangere-577004
Karnataka, India
Ashok.kusagur@gmail.com

Abstract-The intention of this paper is to present a novel method for real- time detection of human fall from the real-time video which is taken from the static digital camera which is fixed in the indoor that provides a secure environment and to improve the quality of life of the old person, children, patients and elderly. The proposed work is based on two techniques, an Ellipse approximation and Motion History Image (MHI).The novel work includes removal of shadows for best detection of human in an indoor environment. In this work human fall detection by considering ellipse approximation, Motion history image and combining both the techniques. Results were compared all techniques for the different possible position of human and it shows that the combined technique gives better accuracy and efficiency of human fall detection compared individuals techniques.

Keywords— Fall detection, surveillance, ellipse approximation, motion history image.

I INTRODUCTION

Now a days Fear offallis the main reasons for older peoples not living alone. The risk of serious injuries is increased if the person remains unconscious or unbalanced after the fall because of their inability to call for someone to help this is the reason now days automatically devices that detect the human fall are in the focus of interest. The main aim is to reduce the time between the fall and the arrival of aid, reducing the cost of treatment. Currently, the standard answer to detect falls is by means of wearable sensors [2], [3]. The afore mentioned sensors are typically attached beneath the armpit, around the wrist, behind the ear's lobe or over the waist. These devices incorporate inclinometer and accelerometer sensors and has the capability to monitor acceleration and velocity, vertical posture toward lying posture. Though, the difficulty of such detectors is that older people often tend to forget wearing them, indeed their productivity relies on the person's ability and inclination towards wearing them. However,incase of non-contact sensors, they frequently provide fairly basic data that's hard to understand. In order to overcome these difficulties, we can use computer based vision techniques that does not require that the person has to wear anything. Additional reason for using such system is that a camera can give more data and accuracy on the motion of a person actions than an accelerometer.

Therefore, huge amount of research workhas been carried out in Computer Vision-based techniques. The suggested research work is our contribution towards enhancing the knowledge in the area of tracking and activity recognition. The vision based Surveillance systems have been providing applications for human tracing, activity monitoring, fall detection and so on. Vision-based surveillance systems are getting a huge amount of interest specifically in the fields of security and assistance. Such systems are built in order to achieve several tasks from detection of human presence to identification of irregular activities. In the past few decades, Vision- based surveillance has been broadly applied in industrial inspection, traffic control, security systems, medical and scientific research.

In this paper present detection of human fall from the real-time video which is taken from camera which is fixed in the room. The techniques used here detect fall efficiently are ellipse approximation of human and motion history image and combined both the techniques.

II RELATED WORK

With the swift growth of population of the elderly in the world, the demand for the health care system is increased accordingly. The different types of fall and approaches to develop the system is described by Xinguo Yu[1]. Initially to detect fall of human using sensors are the normal method and these devices are called accelerometer or inclinometer [3] which calculate the velocity of a human. Based this information the activities human are analyzed but provided by data sensors may not accurate. Many human detection algorithms make use of human skin coloring [4,5,6]. These algorithms are not accurate due to the difference in skin color among human beings. Quming Zhou and J.K. Aggarwal[7] describes moving object tracking in the outdoor environment which provides solution for detecting and segment the moving object in the video using background subtraction but it fails in the presence of shadows. Zhuo-LIN JIANG [8] portrays an operational solution to correctly classify people from other moving objects in indoor environments, including shadows, sudden changes of light and movements caused by the wind.

There are various techniques to extract the features of a human, the best method is ellipse approximation. Ellipse has several characteristics based on that fall can be estimated [9]. But this is not efficient to get more accuracy and in turn improve the accuracy of fall detection and one more technique is applied in the proposed system is MHI[10,11]. Here proposed work by combining both techniques to improve fall detection efficiency and accuracy

The paper organization is as follows: The proposed system is presented in Section three. The results and the experimental study are presented in section four and concluding observation is discussed in the final section

978-1-5090-3704-9/17/$31.00 © 2017 IEEE

III PROPOSED SYSTEM

The objective of proposed work is to detection of humanfall usingvision-based surveillance system. The general framework of proposed system as depicted in figure1.

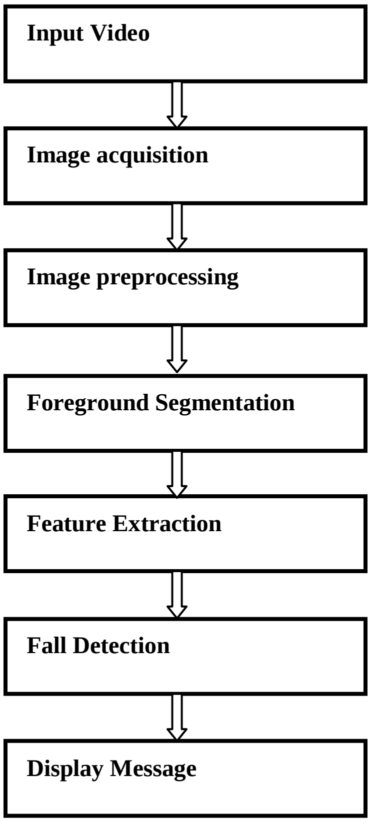


Fig1. General frame work of proposed system

A. *Image acquisition*

Image acquisition is the first and foremost step, here the video is captured at a resolution of 480 X 640 and maximum of 30fps using the digital static camera.

B. *Image pre-processing*

The goal of pre-processing is an enhancement of the captured video or image data that over whelms undesired distortions or augments some image features for further processing. Here 5x5 Gaussian low pass filter is used for undesired noise reduction which resulting in a better processing the captured video for human detection. After Gaussian low pass filter in preprocessing stage median filtering often used for background estimation Median filter are a statistical non-linear method for image smoothness by utilizing the median of the neighborhood.

The medianfilter task is to find each pixel values from the each frame at point(m,n) by ordering the values in a sorted way and selecting the middle one the processed image,the median of the sorted value is first computed and then the middle one is selected as the pixel value for the processed image. The background estimation for pixel $X_t(m,n)$ at frame t using a length L is calculated by median filter and is given by [3].

$$X_t(m,n) = \text{median}_L(x_{t - 0.5\,L}(m,n),\ldots\ldots\ldots,x_{t +-0.5\,L}(m,n))$$

C. *Foreground Segmentation*

The Separation of Foreground and Background is done by moving object detection method. Here Foreground is moving, and the background is stable. This moving object is segmented by background subtraction method followed by division method. In background subtraction method,the moving object is obtained by subtracting background image from the current image in each RGB color channel and then takes the maximum absolute value

$\text{Diff}_c = \text{Max}\,\{|R_f - R_b|, |G_f - G_b|, |B_f - B_b|\}$.This difference image is converted into binary image to clearly distinguish moving object is done by thresholding and is given by at pixel (x,y)

$$F(x,y) = \begin{cases} 1 & if\,Diff\,c \geq T \\ 0 & otherwise \end{cases}$$

After thresholding, the image contains shadows since the shadows are connected to the moving object if the object moves the shadow also moves in order to remove shadows here division method is used [12].

Let the current image I(x,y) gray value and background image B(x,y) grayvalue ratio is calculated and this ratio changes only at moving object position. The ratio is given by

$$R(x,y) = \frac{I(x,y)}{B(x,y)}$$

and to obtain binary object the thresholding is used and segmented moving object in binary is given by[12].

$$SB(x,y) = \begin{cases} 1 & R(x,y) \geq T \\ 0 & other\,wise \end{cases}$$

The segmented moving object contains unwanted blobs these blobs are removed by using mathematical morphological operations are dilation and erosion.

D. *Feature Extraction*

In order to recognize and track the moving object some of the features are extracted from the object moving by considering ellipse approximation and MHI

- Ellipse approximation

Let f(x,y) ≥ 0 be real bounded continuous image with support for a finite region R, we define its $(p+q)^{th}$ order moments is given in equation 1
where

$$\mu_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(x-\bar{y}\right)^p \left(y-\bar{y}\right)^q f(x,y)dxdy ----- (1)$$

$$\bar{x} = \frac{\mu_{10}}{\mu_{00}} \qquad \bar{y} = \frac{\mu_{01}}{\mu_{00}}$$

The two main characteristics of the ellipse which are necessary to detect human fall are orientation and. eccentricity The orientation of the ellipse is defined as the angle amid themajor axis of the person and the horizontal axis is obtained by reducing with respect to Θ and is given by equation 2 and this ranges from -90^0 to $+90^0$.

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2\mu_{11}}{\mu_{20}-\mu_{02}}\right) ------ (2)$$

The eccentricity is the ratio of major axis to the minor axis. The value of eccentricity is between 0 and 1.

The I_{max} and I_{min} are the highest moments of inertia that can be obtained by computing the Eigen values of the matrix and is given by

$$J = \begin{pmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{pmatrix} --- (3)$$

$$I_{max} = \frac{\mu_{20}+\mu_{02}+\sqrt{(\mu_{20}-\mu_{02})^2+4\mu_{11}^{\ 2}}}{2} --- (4)$$

$$I_{min} = \frac{\mu_{20}+\mu_{02}-\sqrt{(\mu_{20}-\mu_{02})^2+4\mu_{11}^{\ 2}}}{2} --- (5)$$

The major semi-axis a and minor semi axis b of the ellipse are given by

$$a = \left(\frac{4}{\pi}\right)^{\frac{1}{4}}\left[\frac{(I_{max})^3}{I_{min}}\right]^{1/8} \qquad b = \left(\frac{4}{\pi}\right)^{\frac{1}{4}}\left[\frac{(I_{min})^3}{I_{max}}\right]^{1/8} -- (6)$$

Where a and ***b*** are calculated in order to detect changes in the human shape to distinguish an abnormal behavior from normal activities. The sudden changes is considered based on standard deviation of Orientation and Eccentricity[10]

- *Motion History Image*

In order to represent how image moving, we use motion history image here. Let H(x,y,t) at time t and position (x,y) is computed by

$$H_\tau(x,y,t) = \begin{cases} \tau & \text{if } D(x,y,t)=1 \\ \max(0, H_\tau(x,y,t-1)-1) & \text{otherwise} \end{cases}$$

Where, x, y and t indicate the position and time, D(x, y, t) is the difference binary image constructed by sequential frame difference, is the maximum duration a motion is stored. The Fig2 shows example of different motions and its motion history image (MHI). can be calculated by using is

$$V_{motion} = \frac{\text{Number of gray pixels}}{\text{Number of white pixels} + \text{Number of gray pixels}}$$

Gray pixels are in the range if $1 < H_\tau(x,y,t) < \tau$
white pixels are in the range if $H_\tau(x,y,t) = \tau$

The represents how the motion has occurred fast and it is used to measure fall of the human body because during fal l the motion is very high [13].

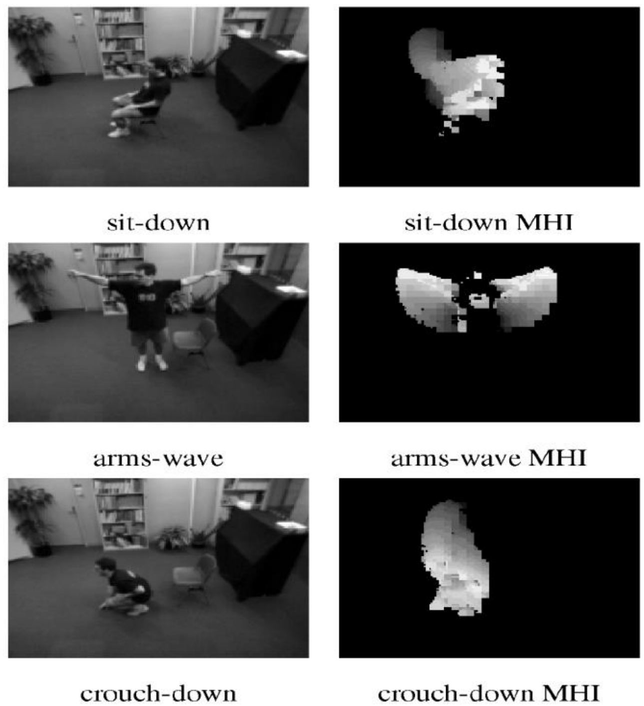


Fig2:Different motions and its motion history image[7]

IV RESULT AND DISCUSSION

The proposed works on real time videos for detecting an old person or patients living in the room. The static camera is fixed in the room, it captures the entire room. The Fig3, Fig 4 and Fig5 below shows the standard deviation of Eccentricity(S_{ec}),the standard deviation of orientation(S_{ori}) and motion velocity(V_{motion}) for the real-time video for three different threshold valuesT1,T2 and T3.The fall is detected if only if the standard deviation of Eccentricity ≥ T1 logical and standard deviation of Orientation ≥ T2 logical and motion history velocity ≥ T3 conditions must satisfied then only fall is detectetd.Table1,Table2, and Table3 below show the different techniques for different positions of a human.

.

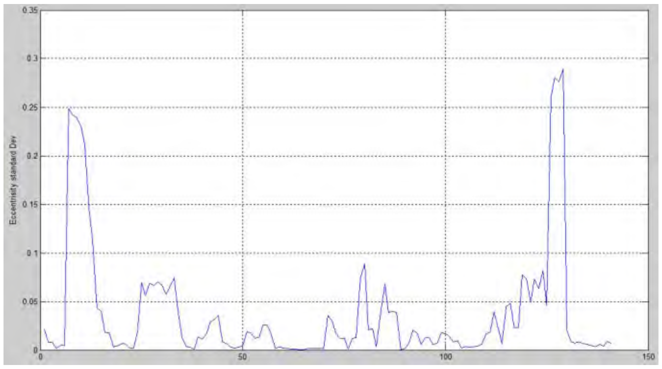


Fig3. Standard deviation eccentricity

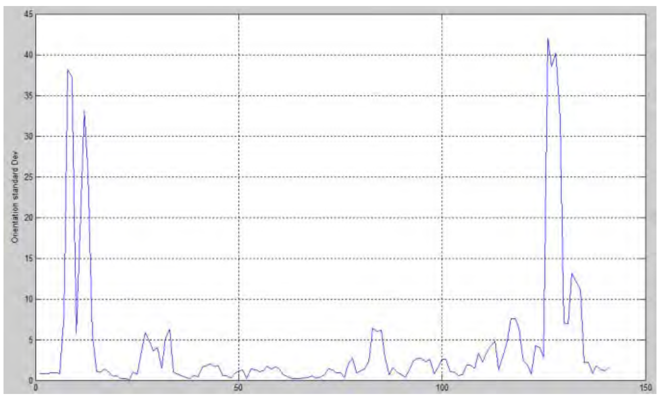


Fig4. Standard deviation of orientation

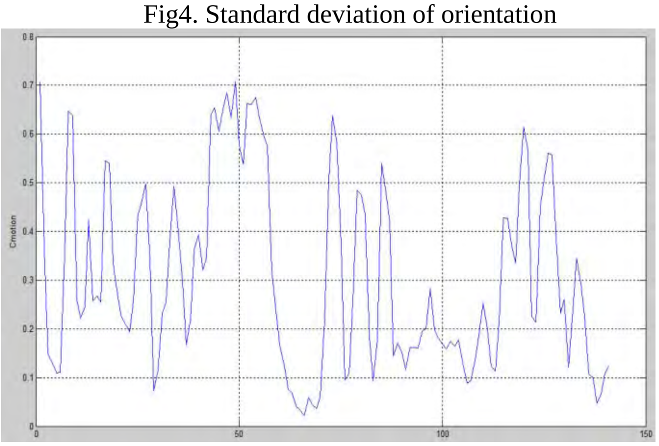


Fig5. Standard deviation of orientation

Table1:Ellipse approximation for different position of human

Ellipse parameter	Different position of Human				
	Standing position	Side fall	Front fall	Bend down position	sitting
S_{ecc}	0.001	0.2	0.18	0.01	0.05
S_{ori}	0.15	8	15	5	6

Table2:Motion history image for different position of human

MHI Parameter	Different position of Human				
	Standing	Side fall	Front fall	Bend down	Sitting Down
V_{motion}	0.84	0.5	0.6	0.47	0.17

Table 3: Combined both Ellipse approximation and MHI

Ellipse and MHI parameter	Different position of Human				
	Standing position	Side fall	Front fall	Bend down	Sitting down
S_{ecc}	0.002	0.18	0.1	0.01	0.05
S_{ori}	0. 5	30	7	5	6
V_{motion}	0.6	0.65	0.8	0.35	0.4

The comparison result of different techniques for different human activities of the above methods is shown in Table 4.The Ellipse approximation technique gives less accuracy reason is that the negative fall detection occurs in case 4 and case 5 even though the person is not falling but it is detected as a fall. This can be overcome byusing motion history image technique which gives No fall detection in case 4 and case 5.

Table 4: Comparison of the result of different techniques for different human activities.

cases	Human different position	Ellipse approximation	Motion history image	Both Ellipse approximation and MHI
1	Normal walking	No fall detected	Fall detected	No fall detected
2	Side fall	Fall detected	Fall detected	Fall detected
3	Front fall	Fall detected	Fall detected	Fall detected
4	Bend down	Fall detected	No Fall detected	No fall detected
5	Sitting down	Fall detected	No Fall detected	No Fall detected

The performance of the proposed system is calculated by taking both fall and non-fall videos and find recognition of system by positively and negatively. The Table 5 shows evaluation result of a system which uses ellipse approximation and motion history image and combined both ellipse approximation and motion history image. The combined techniques gives better accuracy compared to individual techniques.

Table 5: Evaluation of recognition system

System recognition	Ellipse approximation		Motion History Image		Both Ellipse approximation and MHI	
	Fall	Non fall	Fall	Non fall	fall	Non fall
Total videos	30	20	30	20	30	20
positive	28	4	20	12	26	18
Negative	2	16	10	8	04	02
Accuracy	93.3%	20%	6.6%	60%	86.66%	90%

CONCLUSION

In this novel work, an efficient vision-based surveillance system is proposed for detection of a human fall old person or patients. The human detection is carried out using subtraction method and division method In subtraction method there is a presence of shadow in the image and shadow image is removed using long division method and for the recognition of human fall, the different techniques are used. The comparison of fall detection systems which uses ellipse approximation, MHI, and the combination of both MHI and ellipse approximation for number of videos were done and it shows that the combined technique gives better accuracy 86.66% for fall and 90% for non-fall compared to the individual techniques.

REFERENCES

[1] Xinguo Yu " Approaches and principles of fall detection for elderly and patient" 10th IEEE Intl. Conf. on e-Health Networking, Applications and Service pp.42-47,August 2008

[2] N Noury, A Fleury, P Rumeau, A Bourke, G Laighin V Rialle, J Lundy “Fall detection - Principles and Methods” Conf Proc IEEE Eng Med Biol Soc,pp.1663- 1666. 2007

[3] James T. Perry, Scott Kellog, Sundar M. Vaidya“ Survey and evaluation of real-time fall detection approaches”.

[4] J. Yang and A. Waibel, “A real time face tracker”, In Proc. of WACV, pp. 142-147, 1996.

[5] D. Ayers and M. Shah, “Recognizing human actions in a static room”, Fourth IEEE Workshop on Applications of computer Vision, pp. 42-47, 1998.

[6] R. Kjeldsen and J. Kender, “Finder skin in color images”, Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, pp. 312-317, 1996.

[7] Quming Zhou and J.K.Aggarwal ,“Tracking and Classifying Moving Objects from Video”, Computer and Vision Research Center.

[8] Zhuolin Jiang, Shaofa Li, Dongfa Gao. “A Ti Saving Method for Human Detection in Wide Angl Camera Images” International Conference of Machine Learning and Cyber- netics, 13-16 August 2006.

[9] HomaForoughi,Baharak shakeri Aski,and Hamidreza Pourreza “Intelligent video surveillance for monitoring fall detection of elderly in home environments” Proc .of International 11th Conference on ICCIT, pp.25-27 December 2008.

[10] Abdunnasar Dief and Rachid Benlamri “An effective view based motion representation for human motion recognition"

[11] Bobick, A., Davis, J.: The recognition of human movement using temporal templates. IEEE Transactions on Pattern Analysis and Machine Intelligence ,pp 257-267 March 2001.

[12] S.A Lokhotiya, M.D Ingole, M.S.Joshi “Human object tracking using background subtraction and shadow removal technique”. International Journal Of Computer Science And Applications vol.6, No.2, Apr 2013.

[13] Y.T. Ngo, H.V. Nguyen, T.V. Pham, “ Study On Fall Detection Based On Intelligent Video Analysis”, The 2012 International Conference on Advanced Technologies for Communications (ATC), pp.114- 117 2012.