Shruti Jain
Nikhil Marriwala
C. C. Tripathi
Dinesh Kumar   *Editors*

Emergent Converging Technologies and Biomedical Systems

Select Proceedings of the
2nd International Conference, ETBS
2022

Springer

Lecture Notes in Electrical Engineering

Volume 1040

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. **

Shruti Jain · Nikhil Marriwala · C. C. Tripathi ·
Dinesh Kumar

Editors

Emergent Converging Technologies and Biomedical Systems

Select Proceedings of the 2nd International
Conference, ETBS 2022

Springer

Editors
Shruti Jain
Jaypee University of Information
Technology
Solan, India

C. C. Tripathi
NITTTR Bhopal
Bhopal, India

Nikhil Marriwala
University Institute of Engineering
and Technology
Kurukshetra University
Kurukshetra, Haryana, India

Dinesh Kumar
Department of Electrical and Computer
System Engineering
RMIT University
Melbourne, VIC, Australia

Preface

Jaypee University of Information Technology Waknaghat has been known for excellence in academics, research, and distinguished faculty since its inception. Jaypee Group is always supportive of the initiatives of bringing together academicians and industry professionals to produce quality and productive research. As a technical education university, JUIT is committed to being a leader in adopting technological advancements to train future engineers.

The second Emergent Converging Technologies and Biomedical Systems (ETBS 2022) was being organized by the Department of Electronics and Communication Engineering and Department of Physics and Materials Science, Jaypee University of Information Technology, Waknaghat, Himachal Pradesh, India, from September 23–24, 2022. The conference is sponsored by the *Council of Scientific and Industrial Research (CSIR) and the Biomedical Engineering Society of India (BMESI).* The ETBS aims to serve researchers, developers, and educators working in the area of signal processing, computing, control, and their applications to present and future work as well as to exchange research ideas. 2022 ETBS invites authors to submit their original and unpublished work that demonstrates current research in all areas of signal processing, computing, control, and their applications. ETBS 2022 solicits full-length original and unpublished papers, based on theoretical and experimental contributions, related, but not limited to the following tracks, are solicited for presentation and publication in the conference: engineering in medicine and biology, signal processing and communication, soft computing technologies, and theme-based special sessions. This conference is one of the premier venues for fostering international scientific and technical exchange across research communities working in multiple domains. The second ETBS 2022 technical program committee put together a program, consisting of 12 technical sessions and six invited talks.

We are thankful to our Chief Guest Prof. Mahavir Singh, HPU, Shimla, and Guest of Honor Prof. Alpana Agarwal, TIET, Patiala, for gracing the inaugural session of ETBS 2022. We are also thankful to the speaker Prof. (Dr.) Dinesh Kumar, Electrical and Computer System Engineering, RMIT, University, Melbourne, Australia, Prof. (Dr.) Celia Shahnaz, Department of EEE, Bangladesh University of Engineering and Technology, Prof. (Dr.) Dilbag Singh, Department of Instrumentation and Control

Engineering, National Institute of Technology Jalandhar, India, Prof. (Dr.) Mayank Dave, Department of Computer Engineering, NIT Kurukshetra, Prof. (Dr.) Pradeep Naik, HoD, Biotechnology Department, School of Life Sciences, Sambalpur University, Sambalpur, Odisha, and Ms. KamiyaKhatter, Editor—Applied Sciences and Engineering who spared time to share their knowledge, expertise, and experience despite their busy schedules.

Despite being ETBS series conferences to be in a fully online mode, ETBS 2022 has been able to garner an overwhelming response from researchers, academicians, and industry from all over the globe. We have received the papers from the USA, Nigeria, Iran, Bangladesh, China, etc., making it truly international. We received papers pan-India from Tamil Nadu, Andhra Pradesh, Raipur, Kerala, Uttarakhand, Delhi, Pune, Hyderabad, Rajasthan, Guwahati, and not to mention our neighboring states. The authors are from premium institutes IITs, NITs, Central Universities, PU, and many other reputed institutes.

We have received over 195 research papers, out of which 56 papers are accepted and registered, and presented their research papers during a two-day conference, acceptance ratio is 28.7%. We truly believe the participation of researchers from different universities and institutes, working on applications of thematic areas of the conference across the domains in this conference and deliberating their research issues and outcomes, resulted in fruitful and productive recommendations.

We sincerely hope you enjoy the conference proceedings and wish you all the best!!!

Solan, India	Shruti Jain
Kurukshetra, India	Nikhil Marriwala
Bhopal, India	C. C. Tripathi
Melbourne, Australia	Dinesh Kumar

About This Book

This book will provide a platform and aid to the researches involved in designing systems that will permit the societal acceptance of ambient intelligence. The overall goal of this book is to present the latest snapshot of the ongoing research as well as to shed further light on future directions in this space. The aim of publishing the book is to serve for educators, researchers, and developers working in the area of recent advances and upcoming technologies utilizing computational sciences in signal processing, imaging, computing, instrumentation, artificial intelligence, and their applications. As the book includes recent trends in research issues and applications, the contents will be beneficial to professors, researchers, and engineers. This book will provide support and aid to the researchers involved in designing the latest advancements in healthcare technologies. The conference "Emerging Technologies in Biomedical Sciences" encompasses all branches of biomedical signal processing and instrumentation, wearable sensors for healthcare monitoring, biomedical Imaging, and biomaterials, modeling, IoT, robotics, and simulation in medicine biomedical and health informatics. It presents the latest research being conducted on diverse topics in intelligence technologies with the goal of advancing knowledge and applications in this rapidly evolving field. Authors are invited to submit papers presenting novel technical studies as well as position and vision papers comprising hypothetical/speculative scenarios.

Keywords

   i. Biomedical signal processing and instrumentation
  ii. Wearable sensors for healthcare monitoring
 iii. Biomedical imaging micro/nano-bio-engineering and cellular/tissue
 iv. Engineering and biomaterials computational systems
  v. Modeling and simulation in medicine biomedical and health informatics.

Contents

Contents

Contents

About the Editors

Shruti Jain is an Associate Dean (Innovation) and Professor in the Department of Electronics and Communication Engineering at the Jaypee University of Information Technology, Waknaghat, Himachal Pradesh, India. She has received her Doctor of Science (D.Sc.) in Electronics and Communication Engineering. She has teaching experience at around 18 years. She has filed eight patents, of which two have been granted and six are published. She has published more than 25 book chapters, and 130 research papers in reputed indexed journals and at international conferences. She has also published 14 books. She has completed two government-sponsored projects. She has guided seven Ph.D. students and now has four registered students. She has also guided 11 M.Tech. scholars and more than 100 B.Tech. undergrads. She has organized 11 conferences of IEEE and Springer as Conference General Chair. Her research interests are Image and Signal Processing, Soft Computing, Internet-of-Things, Pattern Recognition, Bio-inspired Computing, and Computer-Aided Design of FPGA and VLSI circuits. She is a senior member of IEEE, an Executive member of the IEEE Delhi Section, a life member and Executive member of the Biomedical Engineering Society of India, and a member of IAENG. She is a member of the Editorial Board of many reputed journals. She is also a reviewer of many journals and a member of TPC of different conferences. She was awarded by Nation Builder Award in 2018–2019 and enlisted in 2% of scientists of world rankings of 2021 published by Elsevier, data compiled by Stanford University.

Nikhil Marriwala (B.Tech., M.Tech. and Ph.D. in Engineering and Technology) is working as Assistant Professor and Head of the Department Electronics and Communication Engineering Department, University Institute of Engineering and Technology, Kurukshetra University, Kurukshetra. He did his Ph.D.. at the National Institute of Technology (NIT), Kurukshetra in the department of Electronics and Communication Engineering. More than 33 students have completed their M.Tech. dissertation under his guidance. He has published more than five book chapters in different International books, has authored more than 10-books with Pearson, Wiley, etc., and has more than 40 publications to his credit in reputed International Journals (SCI, SCIE, ESCI, and Scopus) and 20 papers in International/National conferences.

He has been granted eight Patents with two Indian patents and six International Patents. He has been Chairman of Special Sessions at more than 22 International/ National Conferences and has delivered a keynote address at more than 7 International conferences. He has also acted as organizing secretary for more than 05 International conferences and 01 National Conference. He has delivered more than 70 Invited Talks/Guest Lectures in leading Universities/Colleges PAN India. He is having an additional charge of Training and Placement Officer at UIET, Kurukshetra University, Kurukshetra for more than 11 years now. He is the Single point of contact (SPOC) and head of the local chapter of SWAYAM NPTEL Local Chapter of UIET, KUK. He is the SPOC for the Infosys campus connect program for UIET, KUK. He is the editor of more than six book proceedings with Springer and guest editor for a special session in the Journal *Measurement and Sensors*, Elsevier. He has been awarded as the NPTEL ENTHUSIASTS for the year 2019–2020 by NPTEL IIT, Madras. He has also been awarded the "Career Guru of the Month" award by Aspiring Minds. His areas of interest are Software Defined Radios, Cognitive Radios, Soft Computing, Wireless Communications, Wireless Sensor Networks, Fuzzy system design, and Advanced Microprocessors.

C. C. Tripathi completed his Ph.D. in electronics from Kurukshetra University, Kurukshetra. Since 2016, he is working as a Director of the University Institute of Engineering Technology (an autonomous institute), Kurukshetra University, Kurukshetra. His research areas are microelectronics, RF MEMS for communication, and industrial consultancy. He has filed one patent and published over 80 papers in journals and conference proceedings. Professor Tripathi is an experienced professor with a demonstrated history of working in the higher education industry. He has been working extensively on graphene-based flexible electronic devices, sensors, etc. Presently he is working as Director, NITTTR, Bhopal.

Dinesh Kumar B.Tech. from IIT Madras, and Ph.D. from IIT Delhi is a Professor at RMIT University, Melbourne, Australia. He has published over 480 papers, authored eight books, and is on a range of Australian and international committees for Biomedical Engineering. His passion is for affordable diagnostics and making a difference for his students. His work has been cited over 7700 times and he has also had multiple successes with technology translation. He is a member of the Therapeutics Goods Administration (TGA), Ministry of Health (Australia) for medical devices. He is also on the editorial boards for *IEEE Transactions of Neural Systems and Rehabilitation Engineering* and *Biomedical Signals and Controls*. He has been the chair of a large number of conferences and has given over 50 keynote speeches.

Human Brain MRI Segmentation Approaches and Challenges: A Review

Puneet Bansal, Suraj Prakash Singh, and Krishan Gopal

Abstract In many clinical applications, brain MRI breakdown is a crucial step since it affects how the overall study turns out. This is so that various processing procedures, which depend on precise subdivision of structural sections, can function. Today, MRI subdivision is frequently considered for image guided involvements and medical planning, for assessing and imagining many brain constructions, for outlining lesions, for examining brain development, and more. Due to the variety of image processing applications, multiple segmentation approaches with varying degrees of accuracy have been developed. In this article, we cover the difficulties and the most widely used techniques for brain MRI subdivision.

Keywords Magnetic resonance imaging · Brain mining implement · White matter · Gray matter · Cerebrospinal fluid

1 Introduction

Image subdivision divides the image into collection of parts that are homogenous, non-overlapping, and have semantic significance. These sections can have comparable qualities like concentration, deepness, color, or smoothness. The end outcome of segmentation is either a group of outlines that show the boundaries of the homogenous regions or a copy of tags categorizing each one.

The distinguishing qualities of an object or segmental picture structure are represented by image features. Quantitative descriptors of shape and visual appearance which can be considered to distinguish between the structures of interest and their backdrop are examples of features which depend on mathematical measurements.

P. Bansal (✉) · S. P. Singh · K. Gopal
Department of Electronics and Communication Engineering, University Institute of Engineering and Technology, Kurukshetra University, Kurukshetra, Haryana 136119, India
e-mail: puneet4u@gmail.com

K. Gopal
e-mail: kgopal2015@kuk.ac.in

1

Right feature collection (selecting the best pertinent structures) and precise feature mining are crucial to the success of picture segmentation. Edges are one of the most often used features for picture segmentation. The borders of an object's surface when intensities abruptly shift are referred to as its edges [1]. Thresholding the principal and additional mandate 3D results of the concentrations is a common method for detecting such changes (the concentration grade and Laplacian). But, using this method to detect edges makes them susceptible to image noise [2] and frequently calls for image flattening as a preprocessing stage [3, 4]. The objective of this article is to present critical reviews of approaches already presented for the segmentation of human brain MRI images.

Section 2 presents various challenges faced during segmentation of human brain MRI images. Various types of segmentation approaches are discussed in Sect. 3. Comparative analysis of segmentation methods is shown in Sect. 4. Section 5 presents the conclusion and future scope of this article.

2 Challenges

As illustrated in Fig. 1, several processes are needed to properly make MR imageries for subdivision. The bias field alteration, image registering, and removal of non-brain tissue are the key procedures.

2.1 Bias Field Correction

Bias field known as the intensity inhomogeneity is a small frequency spectrally changing MRI object which causes an even signal concentration variation inside nerve with the same physical attributes. The magnetic field's three-dimensional inhomogeneity, changes receiving coil's sensitivity, and interactions with the body's magnetic field all contribute to the bias field's emergence [5, 6]. The bias field is affected by magnetic field's power. When MR images are looked over at 0.5 T, the bias field is practically undetectable and is disregarded.

Modern MR machines with compelling power of 1.5 and 3 T can obtain MR images; however, the bias field is present in an adequate amount to pose issues and significantly impact MRI investigation. In reality, skilled medicinal professionals can analyze photographic MRI data up to specific degrees of concentration inhomogeneity (15–35%) [6]. Involuntary MRI investigation and intensity-built separation techniques, however, perform much worse when the bias field is present.
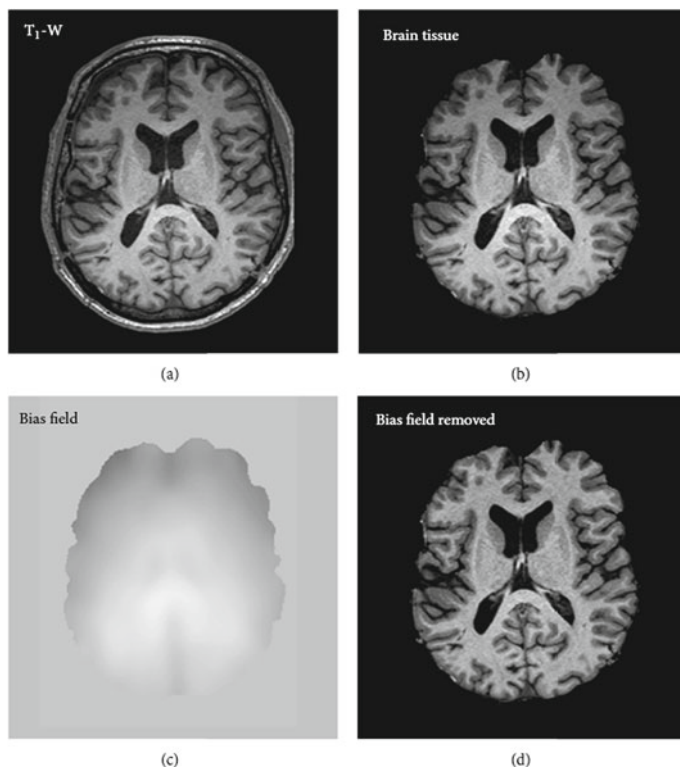
Fig. 1 **a** Original; **b** eliminating non-brain edifices; **c** the bias field; **d** bias field rectification

2.2 Image Registering

The practice of overlapping (spectrally bring into line) two or extra photographs with similar contents that were acquired at various intervals, at various angles, through various devices is known as image registering. Finding the transformation between images is necessary for spatial alignment of relevant image characteristics, which is what image registration entails. However, when brains have lesions or disorders, the inter-focus brain MRI registering will have issues since it is dreadful to tie the similar structures among fit and unhealthy brains.

2.3 Exclusion of Non-brain Muscle

The concentrations of non-brain muscles including fat, the skull, and the neck overlap with those of the brain tissues. In order to apply brain segmentation methods, the brain must first be retrieved. Voxels are categorized as brain or non-brain in this step. The

conclusion can be a twofold cover with an assessment of 1 for brain voxels and 0 for the residual muscles, or a fresh image with only brain voxels. The cerebral cortex's gray matter, white matter, and cerebrospinal fluid and subcortical constituents comparable to the brain stalk and cerebellum make up the majority of the brain's voxels. Non-brain voxels include the scalp, dura mater, chubby, membrane, sways, eyes, and skeletons. Utilizing prior knowledge of the brain's anatomy is the typical technique for brain extraction. Non-brain tissue can be eliminated by moving the brain cover from a deformable template that has been registered with an image [7]. The use of a probabilistic atlas for brain extraction, however, is typically not particularly accurate and can lead to misclassification at the brain boundary. The brain mining implement (BMI) [8, 9], that is a component of the openly accessible software suite, offers an alternate technique for removing the brain.

3 Segmentation Approaches

Segmentation techniques can be categorized as follows.

3.1 Manual Segmentation

The process of manually segmenting and labeling an image by a human operator is known as manual segmentation. When segmenting 3D volumetric imagery, this is often done "slice-by-slice." Because it is challenging to dependably and properly identify structures in medical imaging, manual technique is supposed to be the best precise. Atlas-based segmentation techniques involve manual subdivision of various brain areas, which is a critical stage in the creation of mind atlases [10–12].

3.2 Intensity-Centered Approaches

Specific pixels or voxels are classified using intensity-based segmentation techniques. Based on intensity, the three primary tissue classes in the brain MRI, white matter, gray matter, and cerebrospinal fluid, can be differentiated.

3.2.1 Thresholding

The modest picture subdivision technique is thresholding. The intensity histogram is used in a thresholding technique, which seeks to identify intensity levels, or thresholds, that divide the looked-for modules. The subdivision is completed by means of classifying every pixel concerning the edges. Thresholding is a quick and effective

computational technique, but it ignores the spatial properties of an image (locality statistics). As a result, noise and intensity in homogeneities might affect thresholding. It usually results in pixel clusters that are dispersed throughout low contrast photos. Thresholding can be used in iterative segmentation techniques like fuzzy C-means clustering to differentiate contextual voxels from mind muscle or to adjust the tissue modules. In [13, 21, 22], a review of thresholding approaches is given.

3.2.2 Region Growing

A method for mining a linked area of the image made up of clusters of pixels or voxels with comparable brightness is known as region expanding (also known as region merging) [14]. Region growth, in its most basic form, begins with a kernel fact (pixel/voxel) that is related to the thing of attention. An operator may choose the seed point by hand or a seed finding algorithm may automatically initialize the seed point. Then, if the intensities of all nearby pixels or voxels are sufficiently comparable (meeting a predetermined consistency or similarity condition), they are involved in the developing area. Up until there are no additional pixels or voxels that can be further to the area, this process is repeated. The subdivision of high-accuracy images, which are made up of enormous associated similar sections, is well suited for region growing.

3.2.3 Clustering Methods

Unsupervised segmentation techniques called clustering divide a picture into groups of pixels or voxels with comparable brightness without needing learning imageries. In reality, grouping techniques train on the copy figures that are already accessible. Through repeating among two phases—data grouping and approximating the physiognomies of all muscle category segmentation and training are carried out simultaneously. The K-means clustering [15], fuzzy C-means clustering [16, 17], and the expectation–maximization (EM) technique [18, 20, 23] are the three most used clustering methods.

4 Comparative Analysis of Segmentation Methods

Three threshold-raising methods, three clustering approaches, the Otsu threshold (OT), the adaptive threshold (AT), the seeded region growing (SRG), and utilizing the metrics MSE, PSNR, and accuracy, K-means (KM), fuzzy C-means (FCM), and expectation maximization (EM) are contrasted in Table 1 and Figs. 2, 3, and 4.

Table 1 Performance metrics [19]

Segmentation technique	MSE	PSNR (dB)	Accuracy
Otsu threshold (OT)	0.1335	56.2595	88.75
Adaptive threshold (AT)	0.0005	82.8575	99.75
Seeded region growing (SRG)	0.0545	62.7565	95.75
K-means (KM)	0.0345	65.5575	96.75
Fuzzy C-means (FCM),	0.0415	64.2575	96.95
Expectation maximization (EM)	0.1075	56.8505	84.75
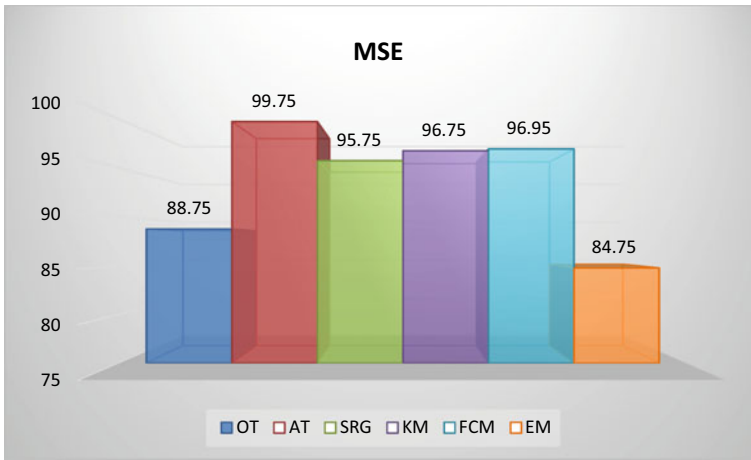
Fig. 2 MSE comparison of segmentation techniques
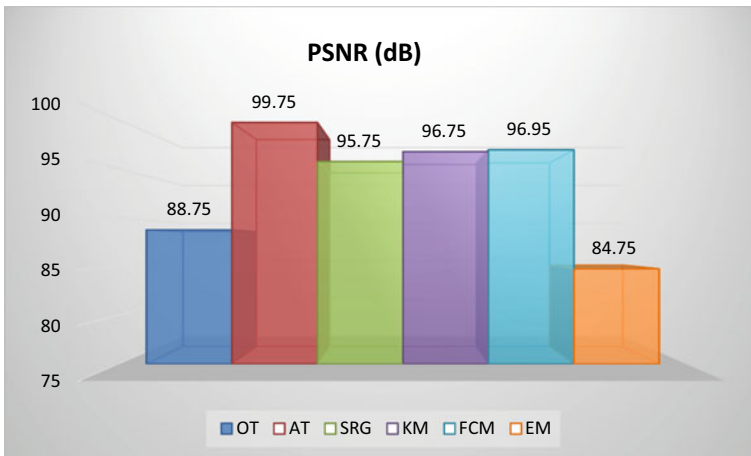
Fig. 3 PSNR comparison of segmentation techniques
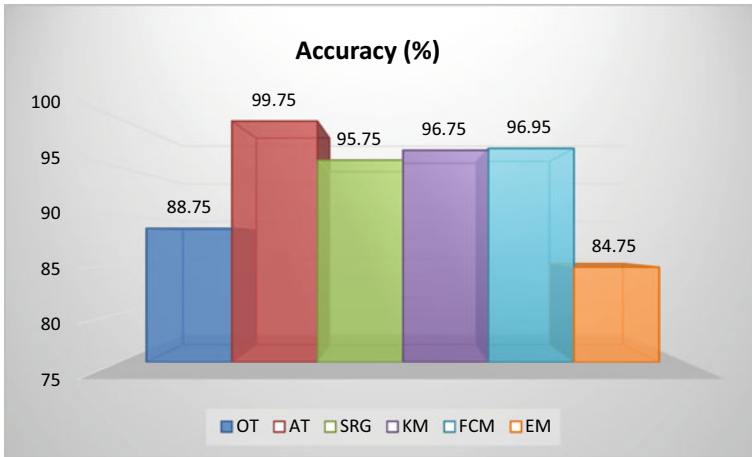
Fig. 4 Accuracy comparison of segmentation techniques

5 Conclusion and Future Scope

This study compares six distinct segmentation algorithms. These include expectation maximization, fuzzy C-means, K-means, Otsu, adaptive, seeded region growing, and seeded region growing. The six segmentation methods are tested on brain MRI images. MSE, PSNR, and accuracy are the measures used to evaluate the segmentation algorithms. According to the performance study, the adaptive segmentation algorithm efficiently segments the image. High precision and PSNR are present. Additionally, when compared to other segmentation methods, mean square error is extremely low. In future, the latest segmentation approaches may be used to minimize mean square error and improve PSNR and accuracy simultaneously.

References

1. Gonzalez RC, Woods RE (2008) Digital image processing. Pearson Education
2. Rogowska J (2000) Overview and fundamentals of medical image segmentation. In: Bankman I (ed) Handbook of medical image processing and analysis. Elsevier, Amsterdam, The Netherlands
3. Canny J (1986) A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 8(6):679–698
4. Marr D, Hildreth E (1980) Theory of edge detection. Proc Royal Soc London Biol Sci 207(1167):187–217
5. Collins CM, Liu W, Schreiber W, Yang QX, Smith MB (2005) Central brightening due to constructive interference with, without, and despite dielectric resonance. J Magn Reson Imaging 21(2):192–196
6. Sied JG, Zijdenbos AP, Evans AC (1998) A nonparametric method for automatic correction of intensity non uniformity in MRI data. IEEE Trans Med Imaging 17(1):87–97

7. Xue H, Srinivasan L, Jiang S et al (2007) Automatic segmentation and reconstruction of the cortex from neonatal MRI. Neuroimage 38(3):461–477

8. Smith SM (2002) Fast robust automated brain extraction. Hum Brain Mapp 17(3):143–155

9. Battaglini M, Smith SM, Brogi S, de Stefano N (2008) Enhanced brain extraction improves the accuracy of brain atrophy estimation. Neuro Image 40(2):583–589

10. Prastawa M (2007) An MRI segmentation framework for brains with anatomical deviations. Ph.D. thesis, University of North Carolina at Chapel Hill

11. Shi F, Shen D, Yap P-T et al (2011) CENTS: cortical enhanced neonatal tissue segmentation. Hum Brain Mapp 32(3):382–396

12. Murgasova M (2008) Segmentation of brain MRI during early childhood. Ph.D. thesis, Imperial College London

13. Sezgin M, Sankur B (2004) Survey over image thresholding techniques and quantitative performance evaluation. J Electron Imaging 13(1):146–168

14. Haralick RM, Shapiro LG (1985) Image segmentation techniques. Comput Vis Graph Image Process 29(1):100–132

15. Coleman GB, Andrews HC (1979) Image segmentation by clustering. Proc IEEE 67(5):773–785

16. Dunn JC (1973) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J Cybern Trans Am Soc Cybern 3(3):32–57

17. Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York, NY, USA

18. Pham DL, Xu C, Prince JL (2000) Current methods in medical image segmentation. Annu Rev Biomed Eng 2(2000):315–337

19. Farzana A, Sathik MM (2017) Review of brain MRI segmentation techniques. Int Res J Eng Technol (IRJET) 04(04). e-ISSN 2395-0056

20. Abdulraqeb ARA, Al-Haidari WA, Sushkova LT (2018) Anovel segmentation algorithm for MRI brain tumor images. In: 2018 Ural symposium on biomedical engineering, radio electronics and information technology (USBEREIT), Yekaterinburg, pp 1–4

21. Oliva D, Abd Elaziz M, Hinojosa S (2019) Multilevel thresholding for image segmentation based on meta-heuristic algorithms. In: Meta-heuristic algorithms for image segmentation: theory and applications. Springer, pp 59–69

22. Oliva D, Martins MS, Osuna-Enciso V, de Morais EF (2020) Combining information from thresholding techniques through an evolutionary Bayesian network algorithm. Appl Soft Comput 90:106147

23. Bercea CI, Wiestler B, Rueckert D, Albarqouni S (2021) FedDis: disentangled federated learning for unsupervised brain pathology segmentation. arXiv preprint arXiv:2103.03705

Hyperspectral Image and Deep Learning Methodology for Water Evaporation Prediction and Control System

M. Sivapriya and P. Mohamed Fathimal

Abstract Three types of surfaces influence the return of rain to the atmosphere. Vegetation, soil, and open water are the three. In vegetation, water evaporates through plant leaves, at or just below the soil-air interface, and through direct evaporation in open water. Because open water is a water consumer, precise open water evaporation has received a lot of attention. The goal of this research is to use drones to measure evaporation rates and the impact of algae blooms. A remotely operated vehicle (ROV) equipped with an aerial hyperspectral imaging camera and an airborne sensor will be launched above the lake. The sensors are used to assess suspended particles, coloured dissolved organic matter (CDOM), chlorophyll-a, and contaminants in the water, as well as predict water evaporation using deep regression. The SVM classifier is used to determine whether an algal bloom should be evicted or cultivated. With the help of executed bots, the proposed system's aggregated facts about water bodies can be shown automatically in the hands of respective officials.

Keywords Hyperspectral image · Water sensors · Mask RCNN · Linear regression · Deep regression · Support vector machine

1 Introduction

The elixir of life is water. Despite the fact that water covers 71% of the planet, just 1% of it can be utilized. It serves as a reminder that it is our responsibility to protect this priceless resource. Evaporation, condensation, and precipitation are the three fundamental processes involved in the global water cycle. Water evaporation is the first and most important step in the water cycle process. This crucial step must also be controlled in order to maintain the global water cycle's balance. Many scientists

M. Sivapriya (✉)
SRM Institute of Science and Technology, Ramapuram Campus, Chennai, India
e-mail: sivapris@srmist.edu.in

P. Mohamed Fathimal
SRM Institute of Science and Technology, Vadapalani Campus, Chennai, India

are attempting to control excessive water evaporation from bodies of water. Even though the water cycle is incomplete without evaporation, regardless, it is one of the leading causes of water loss in the majority of aquatic bodies. This must be kept under control in order to maintain a healthy global water balance.

The direct link between air and water surface is the primary source of water evaporation. Evaporation can be avoided by disrupting or liberating this link. Filling various chemical layers on the water surface to break this link is a common technique. In any case, the water quality parameters would be disrupted by this chemical layer. Nature provides us with natural biological coverings to help manage evaporation. We can limit the amount of water evaporation by using the biological cover efficiently and precisely. The advantages of the biological covers are controlling excess water evaporation, preventing erosion of embankments, improving aquatic life, and lower water surface temperatures. Biological covers are useful in many ways, but they also result in the loss of important minerals in the water. Also, one of the biggest challenges of this cover is its rapid growth. Therefore, a controlled biological cover is needed.

Polystyrene, wax, and foam were employed to float on the surface of the water body to decrease excess water evaporation. It pollutes the water with hazardous chemicals. Then, to control the direct contact between the water surface and the air, biological materials such as straws were designed to float on the surface. Therefore, the temperature can be lowered. Even when trying to do so, aquatic life in the water body suffers from straw rot. Then, solar panels were used on bodies of water. It is very efficient and powerful for generating electricity. But it is only suitable for small water bodies, the installation and maintenance costs are high, and the maintenance is complicated. Therefore, there remains a need for an efficient water evaporation control system that can operate in any weather condition.

2 Literature Review

Chen et al. [1] explained the long-term and short-term variation of water quality in Yuquia reservoir, China. They have taken specimens from three different stations and examined. They handled physical and chemical water quality parameters and did manual mathematical tests on those parameters. They used one way ANOVA to estimate spatial variations of the water quality.

Beck et al. [2] compared Satellite Reflectance Algorithms for Phycocyanin Values and Cyanobacteria Total Bio volume in a Temperate Reservoir and did estimation using coincident hyperspectral imagery and also they surveyed on the different metrics.

Pagán et al. [3] proved that photosynthetic active radiation can be used for normalization, and also he related the same with the ratio of transpiration to potential evaporation.

Vanella et al. [4] explained the estimation of evaporation process under water stress and irrigated conditions using local meteorology with the help of Sentinel 2 vegetation data.

Cheng and Kustas [5] used the two-source energy balance (TSEB) model to assess the component thermal in soil. Thermal observations have been taken from an aircraft at a few metres resolution.

Jason and David-related different measurements for measuring water evaporation in open water bodies in and around the area of small wetland [6]. They used one sided and two sided paired T test for comparing Priestley-Taylor (PT) values.

Daniel et al. [7] assessed the information acquired by hydrologist in the area of Brazilion savannah region for the purpose of agricultural development. They estimated evaporation and evaporation probability curves on a fortnightly period for minor reservoirs.

Yan et al. implemented convolutional neural network, a deep learning method for object detection in hyperspectral images. They trained the model with 400 high definition hyperspectral images and validated the framework. To avoid overfitting problem of the model, they reduced volume of features with the help of 3D convolutional neural network [8].

Lu Yan proposed two stage deep learning-based small object detection algorithm for identifying small objects in hyperspectral image [9]. This system works in two stages. In the first stage, pixel-wise information is taken for classification. After obtaining classification results, spatial information is taken for second stage. When along with pixel-wise classification results, spatial information is combined unlikely regions can be eliminated, and thus, better accuracy can be obtained.

3 Proposed System

3.1 Methodologies Involved

Remote sensing is a science of sensing object from distance and getting related information about sensed object. It generates radiation at distance and breed information from reflected signal. Remote sensors are sensors which help to measure environment parameters such as temperature and ozone concentrations water vapour.

The proposed idea is using such remote sensors to measure water quality parameters which is required to monitor level of evaporation ns. These remote sensors are mounted with remotely operated vehicle. Remotely operated vehicles are used widely in water body researches and can operate in all type of weather conditions. This vehicle can carry remote sensors and essential camera to investigate about the target.

Machine learning is recent fuzz word around the world. It is a subset of artificial intelligence which takes major part in decision-making problem. This pre-trained multi-output regression model to predict water evaporation level and water quality

parameters due to algae bloom. Machine learning simplifies and automates complex decision-making process with the help of trained model. Machine learning and remote sensing are a right combo for automating environmental-related decision-making. For the past decades, deep learning which is subsection of machine learning created a revolution in decision-making.

Regression is a machine learning technique to predict continuous variable with a help of labelled dataset. It comes in different flavour like simple linear, multiple linear, logistic, etc. Multi-output regression is a special deep learning model to solve such prediction problem for predicting multiple variables at a time.

LSTM stands for long short-term memory, a kind of deep learning architecture. This special child of recurrent neural network has the ability to carry or forget the information it earned.

Mask R-CNN works based on the function of faster-R-CNN. But faster R-CCNN gives two outputs which are class label and bounding box offset, whereas mask RCNN gives three which put which are class label, bounding offset, and mask for the object. This mask helps to increase the efficiency of segmentation.

Support vector machine (SVM) is a type of supervised machine learning algorithm. Even though it could be used for regression at most of the scenes, it is used for solving classification problem. Classifications used to be done using hyper and marginal plane in data space.

The proposed methodology uses two different inputs for an efficient water evaporation monitoring and controlling. The first input is hyperspectral image, given by hyperspectral camera which will be attached along with drone.

3.2 Hyperspectral Imaging

This imaging technique takes several images in different wavelengths for same target area. For a single scene, it gives continuous spectrum of light which also includes spectrums that could not be seen by human eyes. Due to this high calibre, it has potential information about target scene.

The hyperspectral image figured by hyperspectral camera will be input component for the LSTM model to measure the growth of floating biological matter on water body. To detect and localize any missing bodies or unwanted objects in water body, mask R-CNN is used. It will fix the location of target with the help of geo referenced hyperspectral image. This image is taken as input for mask-region-based convolution neural network architecture. This instance-based pre-trained network capable of classifying type of object.

The second input to the proposed algorithm is water quality parameters which are measured by water quality measuring sensors. The specifications for water sensor are given in Table 1.

In this proposed system, a remotely operated vehicle (ROV) equipped with an aerial hyperspectral imaging camera and an airborne sensor will be launched above the lake. The recommended drone's specification is as given Table 2.

Table 1 Specification for water sensor

S. No	Description	Specifications
1	Accuracy	± 0.02 pH
2	Body material	PEEK
3	Cable connection	Analogue
4	Calibration method	Two point automatic, one point automatic, two point manual, one point manual
5	Temperature accuracy	± 0.5 °C
6	Temperature sensor	NTC 300 Ω thermistor for automatic temperature compensation and analyzer temperature readout

Table 2 Specifications for drone

S. No	Description	Specifications
1	UAS type	Multi-rotor (with VTOL)
2	Mission	Surveying and mapping/professional grade
3	Landing and takeoff	Vertical takeoff and landing (VTOL)
4	Maximum takeoff weight (MTOW)	Up to the service provider
5	Flight height above ground level	Flight for aerial survey—60 m and altitude for video and images—40 m
6	GNSS grade	PPK enabled with GNSS base station (GNSS station of L1 and L2 frequency)
7	Nominal coverage at 120 m (400 ft) Forward overlap: 80% Side overlap: 60%	Minimum 0.7 km^2 with < 5 cm GSD
8	Endurance	Endurance must be minimum 30 min or more
9	Camera	Camera must be of minimum 20 MP or more
10	Operational wind speed	Minimum 8 m/s or higher
11	Average X, Y accuracy	< 10 cm (95% of total checked values (absolute accuracy) should be less than 10 cm)
12	Average Z accuracy	< 20 cm (95% of total checked values (absolute accuracy) should be less than 10 cm)

The captured hyperspectral images are processed using LSTM to get the details about algal bloom. The same will be processed by the mask R-CNN to detect missing bodies in water bodies.

The air borne sensors are used to (i) measure suspended sediments, coloured dissolved organic matter (CDOM), chlorophyll-a, and pollutants in the water, (ii) predict the degree of water evaporation which will be processed by deep regression.
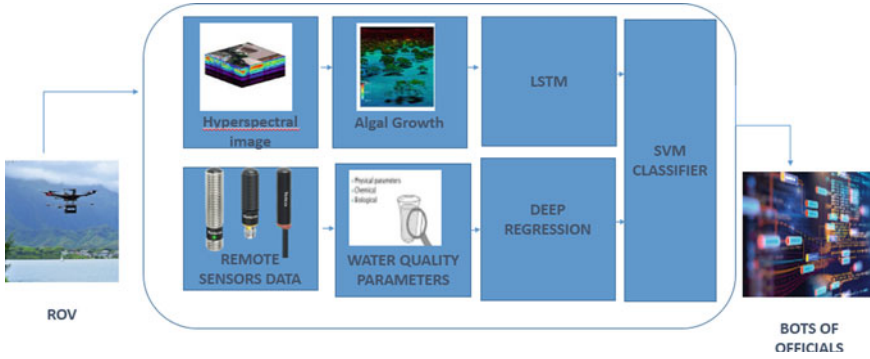
Fig. 1 Architecture diagram of proposed methodology

The support vector machine (SVM) classifier helps in decision of eviction or cultivation of algal bloom. Support vector machine catches the output of multi-output regression, and LSTM concludes whether to evict or cultivate the biological matter. Based on this decision, biological maters growth can be controlled.

3.3 Building a Chatbot

An intelligent bot is used for this entire machine learning process, and the consolidated details about water body can be displayed automatically in hands of respective officials with help of this executed bots. This bot will display the details in the officer's system and assists the officials to monitor the evaporation rate and take proper measures to control evaporation and to take decision about the cultivation or eviction of algal bloom. The architecture diagram of proposed methodology is given in Fig. 1.

3.4 Evaporation Measurement

Some of the empirical formulas for measuring evaporation are as follows:

1. *Fitzgerald's Equation*

$$E_v = (0.4 + 0.124w)(\text{sv} - \text{av}) \tag{1}$$

2. *Meyer's Equation*

$$E_v = C(1 + w/16)(\text{sv} - \text{av}) \tag{2}$$

3. *Rohwer's Equation*

$$E_v = 0.771(1.456 - 0.000732\,\text{Pa})(0.44 + 0.0733w)(\text{sv} - \text{av}) \qquad (3)$$

4. *Horton's Equation*

$$E = 0.4(2 - e - 0.124w)(\text{sv} - \text{av}) \qquad (4)$$

5. *Lake Mead Equation*

This equation helps to measure the quantity of evaporation rate of the day.

$$E_v = 0.0331\,w\left(v_p - v_s\right)[1 - 0.03(T_a - T_w)]24\,\text{h day}^{-1} \qquad (5)$$

where

E_v evaporation in mm/day
sv saturated vapour at the temperature of water surface in (mm)
av actual vapour pressure of air in (mm)
v_s saturated evaporation pressure
w the wind speed in the lowest 0.5 m above the ground (km/h)
v_p is evaporation pressure
T_a atmospheric temperature
T_w water temperature.

For the proposed system, Lake Mead equation has been taken since it consider several parameters like atmospheric temperature, evaporation pressure, saturated evaporation pressure, saturated evaporation pressure, whereas other formulas use only saturated vapour at the temperature of water surface and actual vapour pressure of air.

4 Results and Discussion

4.1 Dataset

The dataset used in the proposed paper is NOAA-GLERL data. NOAA stands for National Oceanic and Atmospheric Administration. This research centres are located in different parts of United States. Great Lakes Environmental Research Laboratory (GLERL) which is a part of NOAA intensively focusses on researches of environment and water quality. GLERL gathers environmental observations with the help of lot of techniques.

4.2 Experiment and Analysis

Data for the experiment is taken from GLERL Dashboard in which data was collected using daily simulations from NOAA-GLERL's one-dimensional Large Lake Thermodynamics Model. The observed data of three months (January, February, March 2021) which was taken from data set was compared with predicted data which was predicted by proposed algorithm. The prediction was done using deep regression. Figures 2, 3, and 4 show the analysis between observed and predicted data of three months January, February, March 2021.

Fig. 2 January month results
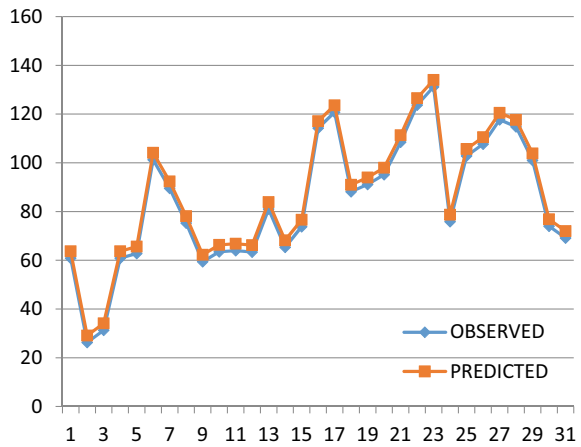
Fig. 3 February month results
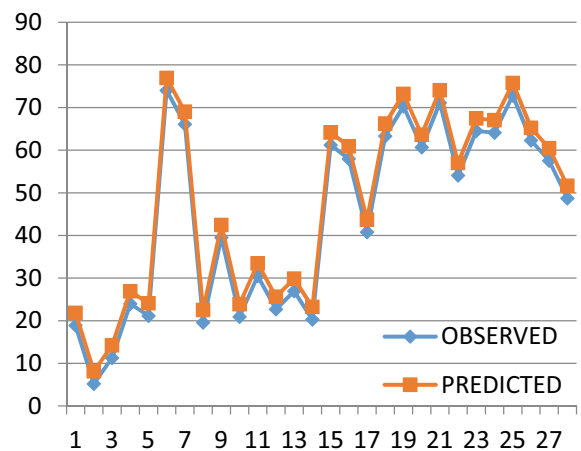
Fig. 4 March month results
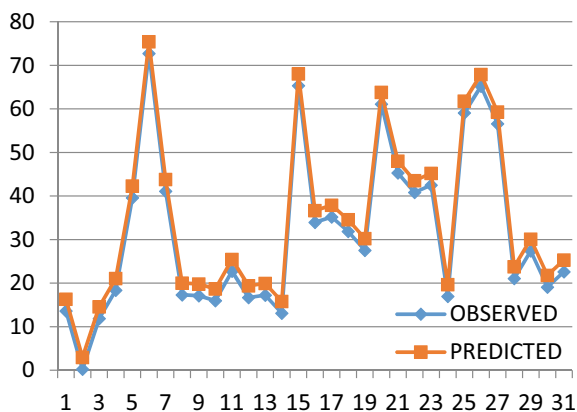
5 Conclusions

To date, several methods have been implemented to control excess water evaporation. Straws were used to floating the body of water. But when they rot, they are dangerous to aquatic organisms and also cause stinks. Then wax is spilled on the surface of the water, which is toxic to drinking water. Some models use solar panels above the body of water, which increases installation and maintenance costs. Therefore, a flotation material that is cheaper and must not disturb aquatic life at the same time is needed. The nature helps us in this regard with aquatic plants. These aquatic plants are an important part of the life cycle of aquatic organisms. These plants have the ability to act as a barrier to direct air and water contact. However, it is necessary to control the growth of this biological substance. This proposed automated intelligent evaporation monitoring system allows staff to help protect water and increase natural aquatic organisms with a small number of personnel without interrupting aquatic organisms and quality.

References

1. Chen YY, Zhang C, Gao XP, Wang LY (2012) Long-term variations of water quality in a reservoir in China. Water Sci Technol. https://doi.org/10.2166/wst.2012.034
2. Beck R, Xu M, Zhan S, Liu H (2017) Comparison of satellite reflectance algorithms for estimating phycocyanin values and cyanobacterial total biovolume in a temperate reservoir using coincident hyperspectral aircraft imagery and dense coincident surface observations. Rem Sens 9(538); 9(6). MDPI. https://doi.org/10.3390/rs9060538
3. Pagán BR, Maes WH, Gentine P, Martens B, Miralles DG (2019) Exploring the potential of satellite solar-induced fluorescence to constrain global transpiration estimates. Rem Sens 11(413):201. MDPI
4. Vanella D, Ramírez-Cuesta JM, Intrigliolo DS, Consoli S (2019) Combining electrical resistivity tomography and satellite images for improving evapotranspiration estimates of citrus orchards. Rem Sens 13(373). MDPI

5. Song L, Kustas WP, Liu S, Colaizzi PD, Nieto H (2016) Applications of a thermal-based two-source energy balance model using Priestley-Taylor approach for surface temperature partitioning under advective conditions. J Hydrol 574–587. Elsevier
6. Masoner JR, Stannard DI (2010) A comparison of methods for estimating open-water evaporation in small wetlands. Wetlands. https://doi.org/10.1007/s13157-010-0041-y
7. Althoff D, Rodrigues LN, da Silva DD (2019) Evaluating evaporation methods for estimating small reservoir water surface evaporation in the Brazilian Savannah. Water 11. MDPI, Springer. https://doi.org/10.3390/w11091942
8. Yan L, Zhao M, Wang X, Zhang Y, Chen J (2021) An algorithm and dataset for object-level hyperspectral image target detection. Sig Process Lett. IEEE, https://doi.org/10.1109/LSP.2021.3059204
9. Yan L, Yamaguchi M, Noro N, Takara Y, Ando F (2019) A novel two-stage deep learning-based small-object detection using hyperspectral images, vol 10. Optical Society off Japan, Springer. https://doi.org/10.1007/s10043-019-00528-0

Classification of Alzheimer's Disease Using Transfer Learning MobileNet Convolutional Neural Network

Monika Sethi, Saravjeet Singh, and Jatin Arora

Abstract Classification of Alzheimer's disease (AD) assists clinicians in formulating appropriate health care recommendations for their sufferers. Timely diagnosis could aid in adopting healthy choices and, when required, provide appropriate medical assistance. Deep learning (DL) is a viable strategy for reducing and understanding the manifestations or symptoms of AD. Moreover, DL allows the software developer to use a raw database as a source, permitting this to discover highly intricate features from the trained database without such involvement of users. Unfortunately, these techniques often generally confined either by requirement for a big collection of training images as well as optimal design customization. This research addresses such concerns via transfer training, wherein the state-of-the-art MobileNet has been utilized. One such model has previously been trained using a massive ImageNet benchmark dataset containing pre-trained weights. As a result, the model is improved through layer-specific tuning, that primarily applies to a selected group of layering for magnetic resonance imaging (MRI) neuro scans. The model was evaluated on performance metrics such as accuracy, loss, area under the curve (AUC), precision and F1-score. In comparison with existing transfer learning techniques, VGG and ResNet utilized for AD classification, and the presented MobileNet model on ADNI dataset attained an overall accuracy of 98.37%.

Keywords Alzheimer's disease · Deep learning · Transfer learning · MobileNet

M. Sethi (✉) · S. Singh · J. Arora
Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India
e-mail: sethi.monika4@gmail.com

S. Singh
e-mail: saravjeet.singh@chitkara.edu.in

J. Arora
e-mail: jatin.arora@chitkara.edu.in

1 Introduction

AD, one of the most common form of dementia in the elderly, is a degenerative brain illness that leads the destruction of neurons cells and tissue loss throughout the cerebral, severely affecting brain structure with times and impacting much of its capabilities [1–3]. Special report investigates primary care clinicians' usage for cognitive evaluations as a method toward enhancing early identification of dementia. The percentage of citizens struggling with AD in the United States might approach 13.8 million by 2050 [4]. Total fatality records revealed 121,404 losses from AD in 2017, putting as the 6th leading cause of death in the U.S and the 5th leading cause of death among Americans aged 65 and over. During 2000 and 2017, mortality from migraine, cardiovascular diseases, and malignancy dropped, although claimed losses from AD grew by 145% [5, 6].

A rapidly increasing range of machine learning (ML) approaches employing neuroimaging datasets attempt to build diagnostics techniques which helps in magnetic resonance imaging (MRI) brain classification and automated region segment, as well as to investigate the mechanics of pathologies, particularly neurological ones [7]. Numerous papers have recently showed up that are using deep learning (DL) for structural MRI with specific techniques varying from multidimensional learning to Deep Boltzmann Machines [8] to a mixture of auto-encoders and classification techniques [9]. Such techniques must not only automatically differentiate AD subjects from normal control (NC) subjects, and moreover foresee the chance of intermediate stage patients, i.e., mild cognitive impairment (MCI) continuing to evolve AD stage, so that MCI cases could be marked as MCI converters (cMCI) or MCI non-converters (ncMCI), based on the risk factors of development. As a basis, the initial prognosis of AD may be appropriately represented as a multi-classification issue. Earlier research findings restricted the issue to a binary classifier problem. Identification is now a critical step in the medical assistance about any medical issue. ML in the medical community might aid laboratory personnel by effectively identifying abnormal elements in bloodstream, fecal matter, and urinal, as well as skin cells. The ML model may recognize any suspicious or abnormalities in bloodstream, urinary, or cells autonomously. It is one major benefit of machine intelligence [10]. And from the other side, the utilization of computational tools in medical sectors is rising, and becoming the practice to collect patient information digitally, which was previously documented on paper-based documents. It leads to better access to a massive volume of electronic health records (EHRs). ML and mining approaches can be applied to such EHRs to improve the performance and profitability of healthcare services [11]. Also with growth of the DL techniques a subfield of ML, it has become possible to extract even the most complicated set of features directly from image data with hardly any user interaction. Several past researchers utilized images from a number of non-invasive brain imaging biomarkers, like as molecular, functional, and structural imaging to assist in the diagnosis of AD [12]. Structured imaging methods including computed tomography and structural MRI (sMRI) offer a great deal of information on structural alterations (orientation, size, and contour) within the nervous system. Functional imaging modalities including positron emission tomography (PET) and functional MRI are widely employed to exhibit the activity of neurons in various

brain regions by revealing both glucose and oxygen levels consumption. Molecular imaging, including single-photon emission computed tomography, is often applied to examine biochemical and molecular changes associated with certain diseases using radiotracers [13, 14].

In [15], authors applied 2D and 3D CNN as well as 3D CNN on the identical data using pre-trained SAE and showed that 3D CNN performed better than 2D CNN. The studies demonstrated 3D methodology could retain the localized 3D representations which contribute to ideal diagnosis. Additionally, 3D CNN is built around 3D CAE, which has also been pre-trained to recognize anatomical alterations there in MRI neuro imaging dataset. Next, 3D CNN dense layers were fine-tuned for maximize the classifier's accuracy on a particular destination datasets. In addition, to deal with restricted healthcare datasets, some other researcher's group in [16] adopted data augmentation for extension of dataset and transfer learning techniques to classify MCI-based neuro MRI scans. The researcher applied two distinct resources to pre-train the system: LIDC and OASIS. In HC versus MCI categorization, the team reported about 90% accuracy. Multimodalities MRI and PET have been defined as neuroimaging modalities by the researcher [17]. To start, various 3D deep CNN models were developed on the image patches to acquire discriminant features in both modalities. The image classifier was then merged using a 2D CNNs system that was cascaded upon top of all that. As a conclusion, model's classification performance for pMCI versus NC was 83 and 93% for AD versus NC.

However, the primary challenges with all these approaches include the demand for a significant amount of labeled training samples and subsequently the need for heavy model architectural tuning. Since transfer learning techniques minimize the needs of such a massive amount of dataset, nowadays researchers are more inclined to utilize these to overcome the difficulty of a smaller training dataset. In this paper, MobileNet a transfer learning model is proposed for AD classification on ADNI dataset. The remaining part of the paper is organized as follows: Sect. 2 comprises methods and materials including the details about dataset acquisition and the proposed methodology opted for AD classification, whereas, Sect. 3 explains the results obtained following the conclusion in Sect. 4.

2 Materials and Methods

With traditional ML approaches, it is possible to derive the eigen vectors from AD and cognitive normal (CN) MRI neuro scans; classify the scans into different classes. And moreover, classifier performance is influenced upon by those derived features. Manually, it is hard to extract the features. If such features extracted are not necessarily precise, the classifier's performance will be low. As a result, researchers utilize DL specifically MobileNet; a transfer learning network model, which can intelligently identify characteristics without human intervention and thus improving the classifier's accuracy. MobileNet is a CNN framework designed specifically for mobile

Fig. 1 MobileNet V1
baseline model
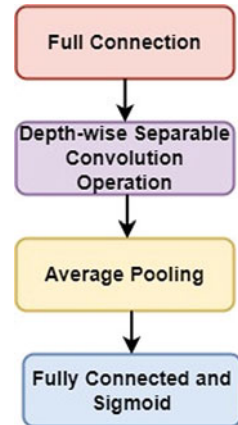
gadgets. As seen in Fig. 1, the primary architecture is based on depth-wise separated convolution operation, that are a form of quantized complexities which factorize a normal complexity together in depth-wise complexity, as well as a 1*1 complexity is referred to as point or element-wise complexity. Furthermore, the approach includes two basic global hyper-parameters to successfully optimize the latency and accuracy [18].

2.1 Dataset Acquisition and Preprocessing

In this research, we employed Alzheimer's disease neuroimaging initiative (ADNI) dataset for two different cohorts AD and CN. This dataset was collected from the official web page (https://adni.loni.usc.edu). ADNI is a never-ending investigation of clinical, imaging, and genetic dataset providers for the early diagnosis of AD. The project began in 2004. The ADNI research's main objective aimed to discover whether neuroimaging biomarkers including MRI and PET, together with all the other biomarkers (medical and neuropsychiatric measures), could have been used simultaneously for identify AD in the early stages of MCI.

The MRI brain images for this experiment have been downloaded after getting approval from the ADNI. The dataset collection utilized in this paper contains longitudinal-sectional MRI data of 20 subjects per class: AD and CN, for a total of 40 participants. Each of the subject includes around 3–4 MRI scans. The input was initially retrieved in NiFti version (nii extension). As a consequence, every NiFti file then transformed to a png files employing a NiBabel package, providing 18,432 and 17,408 for CN and AD, respectively. Only the middle 66 slices (135–200) per MRI scan were taken into consideration. Rest of the images from 1 to 134 and 201 to 256 were discarded as they do not contain any meaningful information [19]. Table 1 presents demographic details for the dataset used in this experiment.

Figure 2 shows the sample images extracted from MRI scan for AD and CN subjects along with the images that were discarded.

Table 1 Demographics dataset details

Subjects	Count	Gender (F/M)	Number of MRI scans (avg. 3–4 scan per subject)	No of extracted images	Images used
AD	20	11/9	68	17,408	4488
CN	20	13/7	72	18,432	4752

Fig. 2 Sample images of AD (top) CN (middle) and discarded images (bottom)

AD

CN

Discarded Images
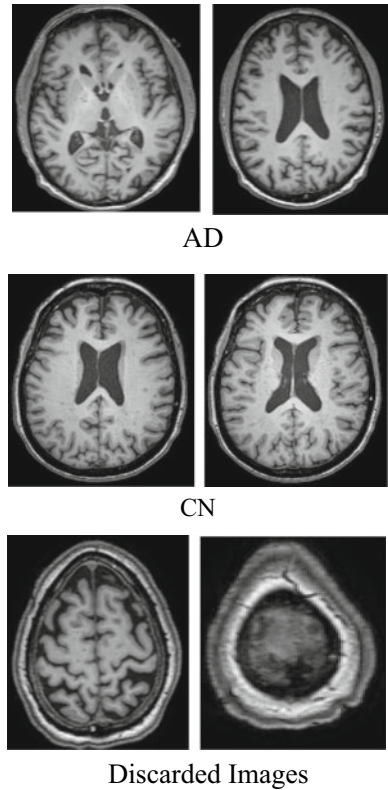
2.2 Proposed Methodology

A MobileNet pre-trained on ImageNet had demonstrated remarkable efficiency to produce image representations suitable for diagnosis. Researchers aim to utilize that power to featured generation to our AD brain MRI classification issue. In general, there have been two approaches of customizing an off-the-shelf model to a novel intent. The first one involves immediately fine-tune connections on the new dataset after altering its classification layers based on the new labels. The parameters of the pre-trained network are then used to initialize the model. It is simple to understand and implement. However, it isn't appropriate for brain image categorization
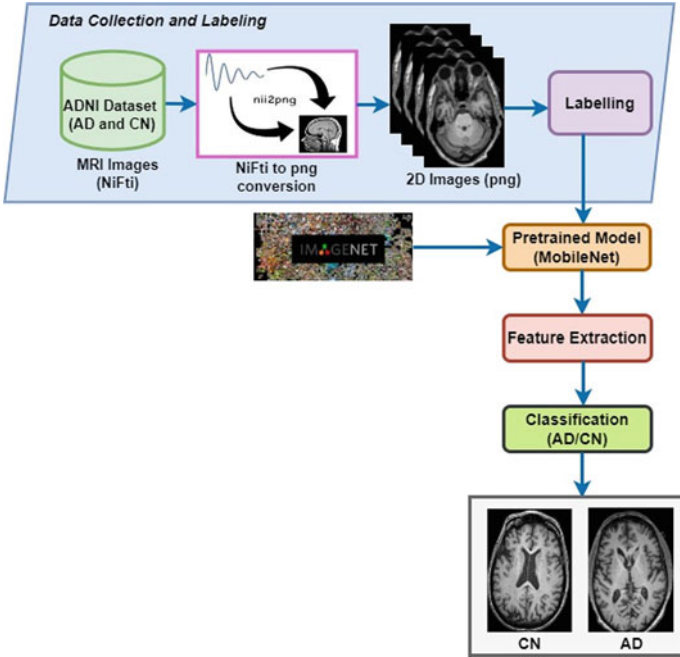
Fig. 3 Proposed approach MobileNet transfer learning model for AD classification

since brain image datasets often contain just hundreds of examples, and adjusting complex model for such a limited dataset could result in over fitting. So, in this research, MobileNet a CNN network model is applied to identify MRI brain scans of AD and NC on ADNI dataset. The MobileNet network model integrates conventional convolution with a deeper separable convolution, significantly eliminates the convolution kernel's redundancies and overhead [20]. Figure 3 illustrates the schematic layout of the proposed MobileNet AD classification approach. As explained in the preceding section, dataset gathered were all in NiFti form. After the translation from NiFti to png, every recovered image then labeled either in CN or AD and passed into MobileNet, originally pre-trained on the ImageNet database. The model was proved capable to identify images in two classes following training on the ADNI input data: AD and CN.

3 Results and Discussion

The proposed model is improved through layer-specific tuning, that primarily applies to a selected group of layering for MRI neuro scans. The model is evaluated on performance metrics such as accuracy, loss, AUC, precision, and F1-score. The graphs (as illustrated in Fig. 4) infer that the algorithms perform well on the untrained validation

set and has a better accuracy, area under curve, precision, and F1-score as compared to the training set.

Moreover, the algorithm undergoes the decrease loss with the increase in epochs. Therefore, it is best suited for unseen dataset in DL production pipeline. Furthermore, Fig. 5 shows that the accuracy of the training set attains a stabilized level at the start of the training and the same stabilization is retained for the validation set. However, algorithm is said to perform best for the validation sets. Similar relationship can be seen for the loss parameter where the loss reduces along with the every epoch on validation set.

Figure 6 shows the true positive and false negative scenarios for the AD prediction. The presented model that utilizes the ADNI dataset achieved an overall accuracy of 98.37%, which compares favorably to other transfer learning techniques in [21] for an accuracy of 85.07% for VGG16, 75.25% for ResNet50, 95.70% for AlexNet and 96.88% for ResNet 18 [22].
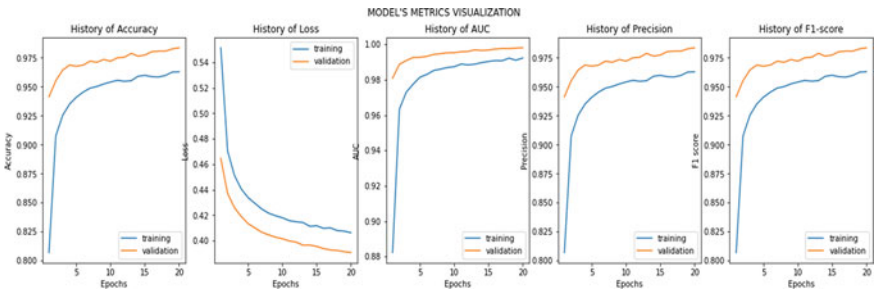
Fig. 4 MobileNet transfer learning model performance visualization
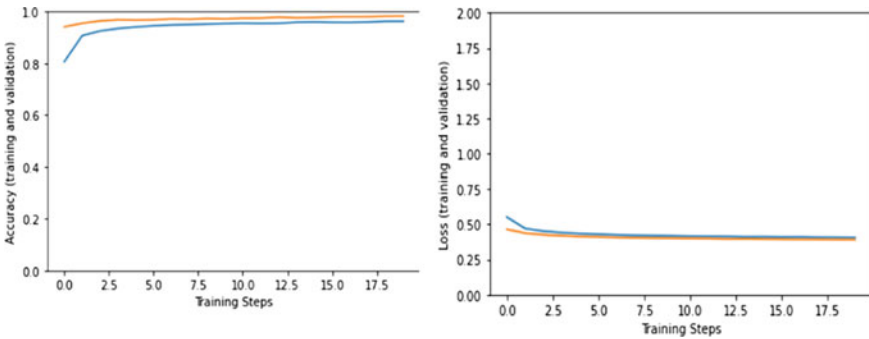
Fig. 5 Accuracy and loss for MobileNet AD classification model

Fig. 6 Predicted labels on
the test dataset
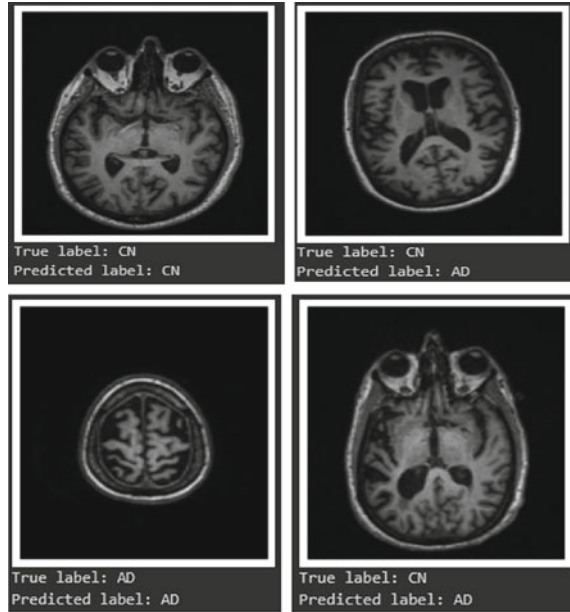
4   Conclusion

AD classification facilitates professionals in developing appropriate healthcare deci-
sions for their patients. A timely diagnosis might help people make healthier choices
and, if necessary, give appropriate medical care. With the rapidly expanding range
of ML and DL approaches employing neuroimaging datasets, attempting to build
diagnostics techniques to assist in MRI brain scans classification and automated
region segment has become a challenging task. The authors of this study made use
of a dataset acquired from the ADNI for their research on two distinct cohorts AD
and CN. On the ADNI dataset, the CNN network model MobileNet was utilized
to differentiate between MRI brain scans of AD and those of non-demented indi-
viduals. After being trained using ADNI input data pertaining to AD and CN, the
model demonstrated the ability to correctly identify images from both categories.
The performance of the model is assessed using metrics including accuracy, loss,
AUC, precision, and F1-score. Other researchers' transfer learning models achieved
an accuracy of 85.07% for VGG16, 75.25% for ResNet50, 95.70% for AlexNet and
96.88% for ResNet 18. In compared to these three models, the suggested MobileNet
attained an accuracy of 98.37%, which is a relative improvement overall.

References

1. Hosseini-Asl E, Gimel'farb G, El-Baz A (2016) Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional network. 502. [Online]. Available http://arxiv.org/abs/1607.00556
2. Punjabi A, Martersteck A, Wang Y, Parrish TB, Katsaggelos AK (2019) Neuroimaging modality fusion in Alzheimer's classification using convolutional neural networks. PLoS ONE 14(12):1–14. https://doi.org/10.1371/journal.pone.0225759
3. Sethi M, Ahuja S, Bawa P (2021) Classification of Alzheimer's disease using neuroimaging data by convolution neural network. In: 2021 6th international conference on signal processing, computing and control (ISPCC), Oct 2021, pp 402–406. https://doi.org/10.1109/ISPCC53510.2021.9609431
4. Liu X, Chen K, Wu T, Weidman D, Lure F, Li J (2018) Use of multimodality imaging and artificial intelligence for diagnosis and prognosis of early stages of Alzheimer's disease. Transl Res 194:56–67. https://doi.org/10.1016/j.trsl.2018.01.001
5. A. Association (2019) 2019 Alzheimer's disease facts and figures. Alzheimer's Dement 15(3):321–387. https://doi.org/10.1016/j.jalz.2019.01.010
6. Marzban EN, Eldeib AM, Yassine IA, Kadah YM (2020) Alzheimer's disease diagnosis from diffusion tensor images using convolutional neural networks. PLoS ONE 15(3):1–16. https://doi.org/10.1371/journal.pone.0230409
7. Korolev S, Safiullin A, Belyaev M, Dodonova Y (2017) Residual and plain convolutional neural networks for 3D brain MRI classification. In: Proceedings of international symposium on biomedical imaging, pp 835–838. https://doi.org/10.1109/ISBI.2017.7950647
8. Brosch T, Tam R (2013) Manifold learning of brain MRIs by deep learning. In: Lecture notes in computer science (including subseries Lecture notes in artificial intelligence and lecture notes in bioinformatics), vol. 8150 LNCS(PART 2), pp 633–640. https://doi.org/10.1007/978-3-642-40763-5_78
9. Liu S, Liu S, Cai W, Pujol S, Kikinis R, Feng D (2014) Early diagnosis of Alzheimer's disease with deep learning. In: 2014 IEEE 11th international symposium on biomedical imaging, ISBI 2014, pp 1015–1018. https://doi.org/10.1109/isbi.2014.6868045
10. Ahmad I, Pothuganti K (2020) Analysis of different convolution neural network models to diagnose Alzheimer's disease. Mater Today Proc (xxxx). https://doi.org/10.1016/j.matpr.2020.09.625
11. Shahbaz M, Ali S, Guergachi A, Niazi A, Umer A (2019) Classification of Alzheimer's disease using machine learning techniques. In: DATA 2019—proceedings of 8th international conference on intelligent data science technologies and applications (Data), pp 296–303. https://doi.org/10.5220/0007949902960303
12. Sethi M, Ahuja S, Rani S, Koundal D, Zaguia A, Enbeyle W (2022) An exploration: Alzheimer's disease classification based on convolutional neural network. Biomed Res Int 2022. https://doi.org/10.1155/2022/8739960
13. Wang SH, Phillips P, Sui Y, Liu B, Yang M, Cheng H (2018) Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling. J Med Syst 42(5):1–11. https://doi.org/10.1007/s10916-018-0932-7
14. Sethi M, Ahuja S, Rani S, Bawa P, Zaguia A (2021) Classification of Alzheimer's disease using Gaussian-based Bayesian parameter optimization for deep convolutional LSTM network. Comput Math Methods Med 2021. https://doi.org/10.1155/2021/4186666
15. Shang Q, Zhang Q, Liu X, Zhu L (2022) Prediction of early Alzheimer disease by hippocampal volume changes under machine learning algorithm. 2022. https://doi.org/10.1155/2022/3144035
16. Wang S, Shen Y, Chen W, Xiao T, Hu J (2017) Automatic recognition of mild cognitive impairment from MRI images using expedited convolutional neural networks. In: Lecture notes in computer science (including subseries Lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 10613 LNCS, pp 373–380. https://doi.org/10.1007/978-3-319-68600-4_43

17. Liu M, Cheng D, Wang K, Wang Y (2018) Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis. Neuroinformatics 16(3–4):295–308. https://doi.org/10.1007/s12021-018-9370-4

18. Bi C, Wang J, Duan Y, Fu B, Kang JR, Shi Y (2022) MobileNet based apple leaf diseases identification. Mob Netw Appl 27(1):172–180. https://doi.org/10.1007/s11036-020-01640-1

19. Jain R, Jain N, Aggarwal A, Hemanth DJ (2019) Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images. Cogn Syst Res 57:147–159. https://doi.org/10.1016/j.cogsys.2018.12.015

20. Lu X, Wu H, Zeng Y (2019) Classification of Alzheimer's disease in MobileNet. J Phys Conf Ser 1345(4). https://doi.org/10.1088/1742-6596/1345/4/042012

21. Acharya H, Mehta R, Singh DK (2021) Alzheimer disease classification using transfer learning. In: 2021 5th international conference on computing methodologies and communication (ICCMC). IEEE, pp 1503–1508

22. Ebrahimi A, Luo S, Chiong R (2020, November) Introducing transfer learning to 3D ResNet-18 for Alzheimer's disease detection on MRI images. In: 2020 35th international conference on image and vision computing New Zealand (IVCNZ). IEEE, pp 1–6

An Exhaustive Review on Emerging Healthcare Internet of Things Technology

Navdeep Prashar, Ashish Kumar, and Ashwani Sharma

Abstract Internet of things (IoT) being an emerging technology comprises of physical entity, hardware, and software that integrate together to frame an ecosystem. Currently, the healthcare system is confronted with various preeminent challenges such as availability of limited resources, lack of experienced professionals, and overpriced medical consultation, resulting in the essential inclusion of IoT-based mechanism to counter these deficiencies. IoT propound an impeccable platform enabling communication between various physical and virtual entities, including areas of personalized healthcare. To identify the primary application of IoT, its architecture, and characteristics in healthcare, a well-structured literature survey is conducted. The analytic study revealed that the home healthcare services are one of the key applications of IoT in healthcare. Furthermore, IoT is taxonomically classified into sensor-based, resource-based, communication-based, application-based, and security-based. Ultimately, discussing the merits and demerits of these classifications pertaining to open issues, future trends, and challenges prevailed during the execution of IoT technology in healthcare sector.

Keywords Healthcare Internet of things (HIoT) · Taxonomy · Architecture

1 Introduction

Presently, the entire world is facing severe health challenges related to chronic diseases brought on by viruses like COVID-19. Rapid rise in health issues accompanied by expensive medical consultations has catapulted mankind to switch toward remote health management system facilitated by computer-aided technology [1, 2].

N. Prashar (✉)
Department of Electronics and Communication Engineering, Chandigarh Engineering College Jhanjeri, Mohali, Punjab, India
e-mail: nav.prashar@gmail.com

A. Kumar · A. Sharma
Department of Applied Sciences, Chandigarh Engineering College Jhanjeri, Mohali, Punjab, India

IoT technology comprises of interconnected network devices that allow automation in different sectors, including remote and smart healthcare systems. This framework includes wireless body area networks (WBANs), wireless sensor networks (WSNs), and radio frequency identification (RFID) to transmit the procured data to the cloud. This data is then extracted and analyzed for prompt and effective decision-making [3, 4]. IoT is considered as an emerging tool in health management systems during global pandemic (COVID19), as there is a strong aspiration to build the healthcare more customized, intense, and affordable. Furthermore, healthcare environments are drastically transformed by the utilization of IoT prospects like the Internet of medical things (IoMT), integrating medical sensors, or customized medical equipment to offer a tailored healthcare report [5]. Terminology, IoMT, and HIoT are often used interchangeably in regards to healthcare informatics [6]. Integration of big data and IoT is a novel idea that enables smart handling of the healthcare operations [7].

This article outlines the advance research queries and trends future directions of HIoT. Exploration of distinct strategies for HIoT systems is essential due to the growing interest in IoT in healthcare. Figure 1 illustrates the structure of this paper. The background is discussed in Sect. 2, and the literature review of HIoT is discussed in Sect. 3. Adopted taxonomy is presented in Sect. 4, followed by current issues, future trends, and challenges which are highlighted in Sect. 5, and later, conclusion is elaborated in Sect. 6.
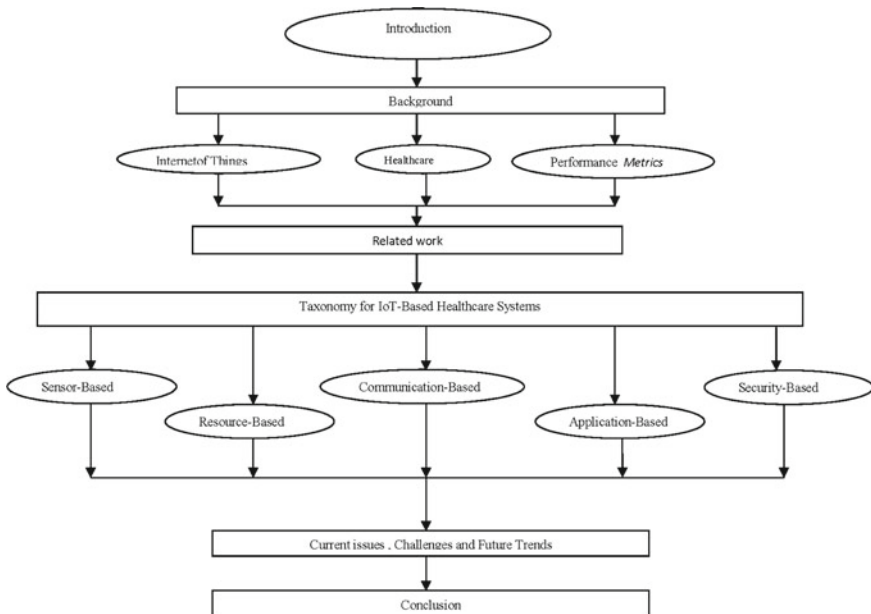
Fig. 1 Structural organization

2   Background

An overview of IoT and healthcare is described in this section. A brief introduction of IoT and healthcare is presented followed by broad description of layered HIoT architecture. Finally, the performance parameters that are employed in this domain are defined at the end.

2.1   Internet of Things

The terminology "IoT" was initially introduced in year 1999 for supply management domain. Further advancements in technology lead to the involvement of IoT in different segments that addressed the numerous concepts such as complex network of uniquely addressable and identifiable objects. Several components and devices such as distinct variety of sensors and actuators are collaborated that communicate with one another using Internet which constitutes the IoT hardware [8]. A typical IoT ecosystem is composed of sensors, sophisticated algorithms, communication interfaces, and cloud interfaces. Data collection from various devices is performed through sensors. Furthermore, network and communication infrastructures are provided by employing RFID technologies and WSN-based sophisticated algorithms for data processing and evaluation [9] followed by cloud domain that facilitate the exchange of client/server request, enabling users to simultaneously access multiple services [10, 11]. Major challenges like latency, reliability, and resources that limited the implementation of cloud computing are countered by fog computing. Fog computing delivers real-time analysis and efficient decision-making features to users [12]. IoT is still in its infancy despite all of its advancements, and there are several research topics in areas like standardization, heterogeneity of distinct devices, scalability, security, and privacy that seeks the attention of today's researchers. A specific challenge in this respect is the interoperability of smart devices to interact and link heterogeneous devices and various vendor systems offering convenient implementation and cost-effective alternative [13]. These challenges once overcome will facilitate convenient means for customers to deal with multiple vendors. Finally, the IoT supports the diverse infrastructures while minimizing the building costs and resolving organizational infrastructure complexity [14]. For instance, an IoT paradigm's introduction in the healthcare sector has the potential to revolutionize the sanitary systems that can be crucial in tale-monitoring at the hospital, particularly for elderly people with chronic illnesses residing at remote areas. By employing IoT technology, future healthcare systems will be benefit greatly in all aspects including early detection of abnormalities, higher-quality care, lower hospitalization expenses, and longer life expectancy [15, 16].

2.2 Healthcare

In the current era of pandemic, regular visits to hospitals are difficult due to restrictions and quarantine necessities [17]. In this regard, a realistic, comprehensive, and computer-aided technology is needed to both improve patients' quality of life and lessen the financial burden [18]. IoT has transformed the sanitary system by analyzing big data and linking multiple IoT devices to collect real-time patient physiological data such as blood glucose levels and monitoring of body temperature. The major objective is to provide people cutting-edge in medical services that includes ongoing illness monitoring and early disease identification. In fact, the IoT may support the healthcare sector in situations like remote clinical monitoring, supported care, illness management, and preventative care. Most prevailing sectors of HIoT are home healthcare, hospital administration, and mobile healthcare [19]. Prior to executing the HIoT framework, users need to understand its architecture and significance of each layer as displayed in Fig. 2.

This architecture composed of perception layer, networking layer, middleware layer, and application/business layer [20]. The perception layer is underneath layer in the architecture and recalled as hardware or physical layer. This layer handles the data collection and signaling before sending the processed data to the network layer. The network layer intends to link all smart devices and enable health data exchange and transmission of patient data securely to the base station via a distinct technologies like ZigBee, Bluetooth, infrared, and Wi-Fi. The middleware layer collects health information from the network layer and stores it in a database.

Finally, the application/business layer performed data evaluation that received from other respective layers. This layer handles all health services and displayed it
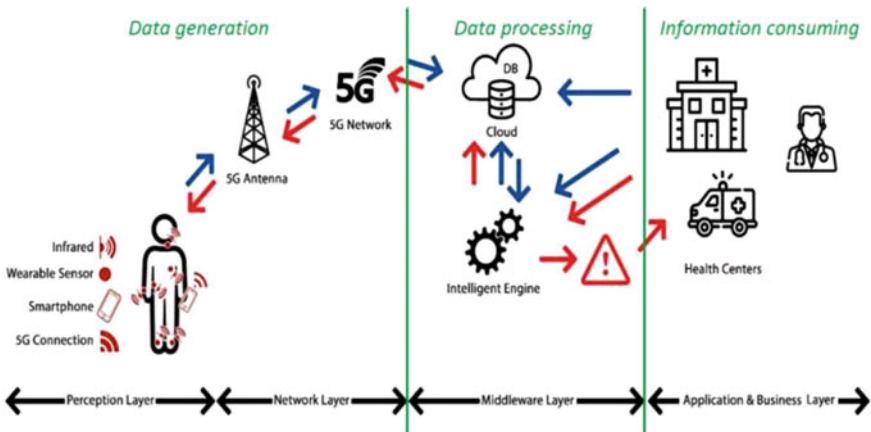
Fig. 2 IoT architecture in healthcare system [20]

to the customers using graphs, flowchart, and models [21]. To determine the effectiveness of HIoT mechanism, different performance parameters are required to address that are broadly described in next segment.

2.3 Performance Metrics

Distinct performance metrics that are used to assess the proposed HIoT techniques are briefly described in this section as follows:

Reliability: The capacity of a system to complete the task in specific circumstances and time. In HIoT, reliability aims to supply requested services to the patients in the majority of situations without any delay [22].

Security and privacy: In order to safeguard health data from attacks including side-channel attacks, physical attacks, malicious attacks, and to preserve privacy by preventing unauthorized access to the data, a smart healthcare systems need to perform some activities [23].

Accuracy: In HIoT, an accuracy is rely on system usage refereeing to the degree to which the fetched data accurately reflects the actual patient's condition [24].

Performance: This parameter is crucial for heath sector service providers to attain precision in data processing to offer treatments effectively. According to the reviewed articles, QoS factors including throughput, latency, delivery rate, mean time between failures, bandwidth use, and other parameters like efficiency, load balancing, resource utilization, overhead, and computing time [25].

Interoperability: The ability to appropriately comprehend data across system is known as data interoperability. Standardized communication and several other interoperability-supporting technologies, like Web services, clinical document architecture (CDA), and cloud computing, are essentially included in HIoT systems [26].

Scalability: Increasing demand and request from users required an extended HIoT mechanism by integrating smart devices, new operations as the service nodes for users and network infrastructures without compromising the quality factor and performance of healthcare services [27].

By evaluating the above distinct performance metrics, the performance of HIoT system can be easily accessed. A lot of research work related to implementation of IoT in healthcare has been already persist which is broadly described in next segment.

3   Related Work

Authors in [28] illustrate the evolution of IoT healthcare systems from clinic-based to patient-based. In year 2017, an advanced IoT mechanism is presented that enabled customized healthcare systems. An architecture comprises of sensor, network, data processing, and application layers which are presented along with the technologies which were thoroughly elucidated, but taxonomy is undefined in this paper [29]. In an IoT layer-based approach, authors in [30] demonstrated physical activity recognition and monitoring architecture and expanded their plan into four levels. For researchers, emerging trends were also described. However, the review was not carried out methodically. A survey that analyzed the most recent significant technologies and their applications in HIoT is presented in year 2019 particularly focused on WBAN and its security features. Additionally, a classification based on technology was presented along with challenges that can be addressed in future work [31]. Authors in year 2020 presented different topologies based on distinct applications for HIoT framework by investigating a wide range of technologies, including machine learning, block chains, big data, edge computing, and software-defined networks (SDNs) [32]. The Internet of medical things (IoMT) security challenges, various risk factors for security threats, and defense mechanisms are examined [32]. This study focused the need for providing device-level security while also demonstrating a moderate amount of communication-level security. Additionally, by leveraging IoT applications for remote diagnostics, this study helps to reduce recurring patient visits to the hospital [33].

4   Taxonomy for HIoT System

The taxonomy of HIoT system is further categorized as sensor-based, resource-based, communication-based, application-based, and security-based approaches as displayed in Fig. 3.

Each comprehensive approach is separately described in each subsection as follows.

4.1   Sensor-Based Approach

The sensor-based approach offered the inclusion of wearable sensors and environmental sensors. In wearable sensors category, a prototype of Android-based galvanic skin response system in smart e-healthcare applications is elaborated that delivers high performance but also offers low security and scalability [34]. Introducing advanced technologies alike artificial intelligence and machine learning empowering
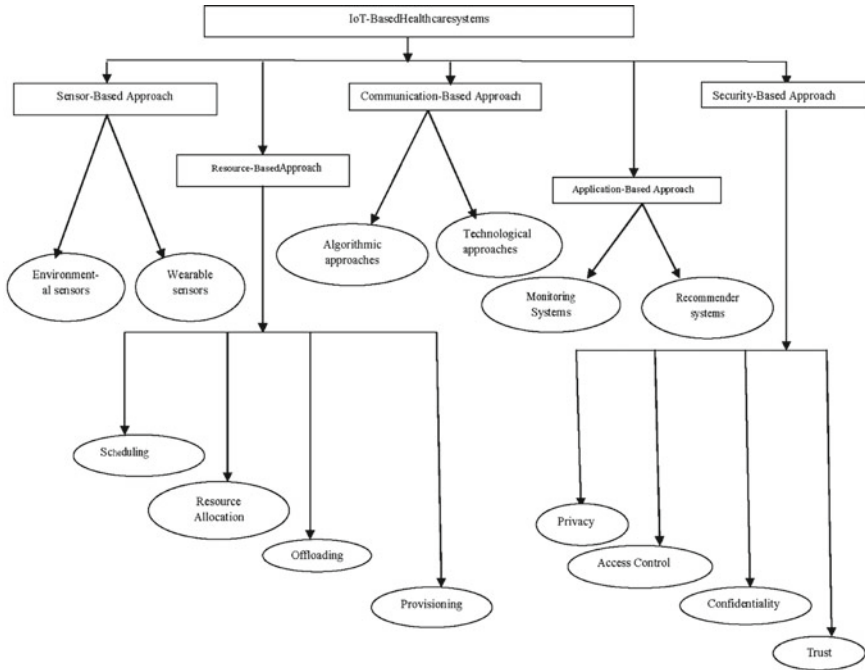
Fig. 3 HIoT taxonomy [20]

the wearable sensors to delivered better precision, high sensitivity, and greater speci-ficity [35]. Another category of sensor-based approach is environmental sensors that involved a high performance fog-assisted real-time system employed in hospitals to monitor the patient health. However, interoperability among heterogeneous devices is the major drawback inculcated in this research work [36].

4.2 Resource-Based Approach

Resource management problems are tedious because of the heterogeneity of resources, resource restrictions, dynamic nature, and HIoT environment unpre-dictability. Resource-based approach is sub-categorized as scheduling, allocating resources, offloading, load balancing, and provisioning.

Scheduling: Scheduling further includes three methods classified as dynamic, hybrid, and static methods. In context to dynamic methods, a delay-tolerant architecture is offered for dynamically allocating diverse tasks to resources and context-aware dew computing-based architecture for time-critical circumstances [16]. For hybrid methods, a schedule management strategy is introduced that minimizes the medical packet transmission delay in systems based on HIoT [37]. Similarly for static

methods, a hash-based polynomial approach is suggested that perform the patients' data calcification and then effectively schedules their processing according to patient health [38].

Resource Allocation: Many existing research papers have been focused on allocating the resources and their related challenges. In this context, a real-time heterogeneous HIoT architecture is defined that employ SDN and link adaptation. Simulation results signify that this architecture delivers high QoS, minimal packet loss, and low latency but also offers poor security and high level of complexity [39]. For healthcare systems, an architecture based on fog is introduced by utilizing elliptic curve cryptography for user identification and the virtual machine partitioning in fog nodes [40].

Offloading, load balancing, and provisioning: In context to offloading, privacy aware reinforcement learning-based technique is designed and simulated that delivers high privacy but offers low latency [41]. For latency reduction of retrieved data, an IoT, NDN, and edge cloud techniques are introduced [42]. Provisioning includes the artificial bee colony (ABC) to enhance the cloud performance in HIoT systems [43].

4.3   Communication-Based Approaches

Communication-based methodology includes technological and algorithmic approaches applied across each node and layers of HIoT system. Technological approaches comprise of smart gadgets, RFID, and WSN for patient monitoring in real-time [44]. For greater coverage, mobility, quick installation, and a long battery life, an advanced IoT-based analyzer and a Bluetooth-enabled electronic reader are introduced for personalized monitoring [45]. A clustered smart nodes network is introduced to lessen the load on RFID readers that further reduces the congestion across the channel [46]. Another category of communication-based methodology is algorithmic approach that covers distinct techniques such as route selection approach, synchronization protocols and convolutional neural networks to enhance data sensing and accurate data collection [47].

4.4   Application-Based Approach

Application-based approaches put the emphasis on offering a system that can perform multiple tasks. In other words, HIoT delivers the essential services for patients, caregivers, or users. The monitoring systems and recommender systems are two sub-categories that are included in this division of the application-based approach.

Monitoring Systems: A fuzzy-based neural classifier algorithm is proposed to diagnose the various diseases [48]. An embedded HIoT care monitoring system is prototyped in year 2019 that measures all three important parameters of human body

(blood pressure, heart rate, and temperature). Further, comparison of this proposed model with other conventional models revealed that this framework accurately and quickly identifies the chronic diseases like diabetes and kidney-related illness in elderly people by utilizing the ANN classification model [49].

Recommender systems: It determines the amount of medicine dosage as well as strict diet needs to follow by patient. In this respect, hybrid filtering techniques are also used to provide suitable recommendations with highly accurate and stable recommendation models. A monitoring system is developed that diagnose the few disorders and advised an preferred medical service according to patient condition [50].

4.5 Security-Based Approach

Security-based approaches addressed the different elements like privacy, access control, confidentiality, and trust of a secure HIoT system that offers the QoS.

Privacy: It deals with encryption algorithm that impart privacy, data integrity, and authentication [51]. For HIoT applications, a secure data collection method is proposed that safeguards user's privacy. Additionally, KATAN and secret cipher sharing algorithms are simulated to evaluate the cost, frequency rate, and computational time for method validation [52]. For effective clustering of data while protecting user's privacy and maintaining the secrecy of their private data, a data clustering approach in conjunction with hemimorphic encryption is introduced [53].

Access Control: It includes two categories authorization and authentication. Authorization process involved a mechanism to protect HIoT restricted resources from unwanted access using a policy-based access control [54]. An ownership transfer, energy-efficient, and secure authentication procedure for HIoT systems was introduced that maintained the privacy while providing access control [55]. Authentication process comprises of sign-on techniques, authentication protocols, and key agreement schemes for HIoT systems [56].

Confidentiality: It is related to encryption techniques to ensure data security and avoid unauthorized accesss. To overcome the issue of resource constraints and to meet the storage, security, and privacy needs, a hybrid IoT-fog-cloud infrastructure is introduced in which fog was dispersed between the IoT devices and cloud. A certificate-based signcryption and encryption approach for HIoT are presented that offers signature, encryption, unforgeability, confidentiality, and security while minimizing communication and computation costs [57].

Trust: To meet the trust objectives, a security-based methods are applied by distinct researchers. In this aspect, a block chain technology is introduced to provide a safe mechanism for storing health multimedia data, offering security chains of records, and generate hashes of each data [58]. A distributed trust framework employing block

chain is proposed which ensures the privacy, integrity, and confidentiality of patient data, enhances HIoT access control, and increases interoperability and security [59]. Last but not least, a malware detector using deep learning is developed to analyze the behavior of HIoT apps that identifies malware and increases trust [60].

Moreover, some of the merits and demerits of distinct classification approaches are also presented in Table 1.

Above exhaustive study of HIoT system identifies some of the crucial challenges and issues that need to be addressed in this article. A brief discussion related to current issues and challenges prevailed during the execution of HIoT system along with the future trend is displayed in the next section.

Table 1 Merits and demerits of distinct classification approaches

Type	Merits	Demerits
Sensor-based	Enhanced performance Greater trust ability High energy Low cost	Low security Low leveling Low mobility
Resource-based	Enhanced optimization Enhanced performance Better dependability High intermission High energy	Low security Low leveling
Communication-based	Better performance Better reliability Better latency Better energy	Low security Low leveling High cost
Application-based	Better performance Better flexibility Better accuracy Better latency	Low security Low leveling High cost Low energy Low interoperability
Security-based	Better privacy Better access control Better confidentiality Better trust Better performance Better accuracy Better latency	Low scalability Low leveling High cost Low interoperability

5   Current Issues, Challenges, and Future Trends

Scalability, interoperability, and mobility are the three key challenges that are encountered by HIoT systems. Another significant hurdle in the development of HIoT systems is deficiency of proper test bed environment for assessing their effectiveness in the real domain. On the other hand, some problems regarding the operational and technical aspects of the HIoT system, such as power and resource management issues relating to the confined storage and computing capacity of fog nodes, have remained as open issues that need to be addressed. These problems are relevant to fog computing environments that develop the HIoT system. Other unresolved concerns in HIoT systems include multi-objective optimization and imparting trust and privacy that rely on different intended purposes.

6   Conclusion

This study focused on the distinct features of IoT application in the healthcare sector. A broad description of IoT architecture and its applications in healthcare systems as well as the method of communication between their respective components is defined. Further, a complete investigation of taxonomy for the HIoT system is carried out, and the important characteristics of classification techniques are clearly examined. This article also addresses the most critical issues like as security, mobility, real-world test bed implementation, and interoperability in the healthcare system. The widespread use of IoT in healthcare has increased productivity, improved quality of life, and proportionally eliminating the health inequities.

References

1. Garg L, Chukwu E, Nasser N, Chakraborty C, Garg G (2020) Anonymity preserving IoT-based COVID-19 and other infectious disease contact tracing model. IEEE Access 8:159402–159414
2. Hosseinzadeh M et al (2020) A diagnostic prediction model for chronic kidney disease in internet of things platform. Multimedia Tools Appl 1–18
3. Mainetti L, Patrono L, Vilei A (2011) Evolution of wireless sensor networks towards the Internet of Things: a survey. In: SoftCOM, 19th international conference on software, telecommunications and computer networks, pp 1–6
4. AlFuqaha A, Guizani M, Mohammadi M, Aledhari M, Ayyash M (2015) Internet of things: a survey on enabling technologies, protocols, and applications. IEEE Commun Surveys Tutorials 17(4):2347–2376
5. Gatouillat A, Badr Y, Massot B, Sejdić E (2018) Internet of medical things: are view of recent contributions dealing with cyber-physical systems in medicine. IEEE Internet Things J 5(5):3810–3822
6. Habibzadeh H, Dinesh K, Shishvan OR, Boggio-Dandry A, Sharma G, Soyata T (2020) A survey of healthcare internet of things (HIoT): a clinical perspective. IEEE Internet Things J 7(1):53–71

7. Lu D, Liu T (2011) The application of IOT in medical system. In: 2011 IEEE international symposium on IT in medicine and education, vol 1. IEEE, pp 272–275
8. Nikravan M, Jameii SM, Kashani MH (2011) An intelligent energy efficient QoS-routing scheme for WSN. Int J Adv Eng Sci Technol 8(1):121–124
9. FossoWamba S, Anand A, Carter L (2013) A literature review of RFID-enabled healthcare applications and issues. Int J Inf Manag 33(5):875–891
10. Asghari P, Rahmani AM, Javadi HHS (2019a) Internet of Things applications: a systematic review. Comput Netw 148:241–261
11. Sultan N (2014) Making use of cloud computing for healthcare provision: opportunities and challenges. Int J Inf Manag 34(2):177–184
12. Farahani B, Firouzi F, Chang V, Badaroglu M, Constant N, Mankodiya K (2018) Towards fog-driven Io T eHealth: promises and challenges of IoT in medicine and healthcare. Future Gener Comput Syst 78:659–676
13. Jabbar S, Ullah F, Khalid S, Khan M, Han K (2017) Semantic interoperability in heterogeneous IoT infrastructure for healthcare. Wirel Commun Mob Comput 1–10, Artno. 9731806
14. Najafizadeh A, Kashani MH (2011) A novel intelligent mechanism for energy efficiency in hierarchical WSNs. Int J Adv Eng Sci Technol 10(1):139–144
15. Sakiz F, Sen S (2017) A survey of attacks and detection mechanisms on intelligent transportation systems: VANETs and IoV. AdHocNetw. 61:33–50
16. Ray PP, Dash D, De D (2019) Internet of things-based real-time model study one-healthcare: device, message service and dew computing. Comput Netw 149:226–239
17. Singh RP, Javaid M, Haleem A, Suman R (2020) Internet of things (IoT) applications to fight against COVID-19 pandemic. Diab Metab Syndr Clin Res Rev 14(4):521–524
18. Marengoni A et al (2011) Aging with multimorbidity: a systematic review of the literature. Ageing Res Rev 10(4):430–439
19. Ahmadi Z, Haghi Kashani M, Nikravan M, Mahdipour E (2021) Fog-based healthcare systems: a systematic review. Multimedia Tools Appl Inpress (2021)
20. Kashani MH, Madanipour M, Nikravan M, Asghari P, Mahdipour E (2021) A systematic review of IoT in health care: applications, techniques, and trends. J Netw Comput Appl 192:1–41
21. Vilela PH, Rodrigues JJ, Solic P, Saleem K, Furtado V (2019) Performance evaluation of a Fog-assisted IoT solution for e-Health applications. Future Gener Comput Syst 97:379–386
22. Bazzaz Abkenar S, Haghi Kashani M, Mahdipour E, Jameii SM (2020) Big data analytics meets social media: a systematic review of techniques, open issues, and future directions. Telematics Inf 57:101517–110555
23. Cutillo LA, Manulis M, Strufe T (2010) Security and privacy in online social networks. In: Furht B (ed) Handbook of social network technologies and applications. Springer US, Boston, MA, pp 497–522
24. Hogan WR, Wagner MM (1997) Accuracy of data in computer-based patient records. J Am Med Inf Assoc 4(5):342–355
25. Manishankar S, Srinithi CR, Joseph D (2017) Comprehensive study of wireless networks qos parameters and comparing their performance based on realtime scenario. In: International conference on innovations in information, embedded and communication systems, pp 1–6
26. Renner S (2001) A community of interest approach to data interoperability. In: Federal database colloquium, vol 1. CiteSeerX, SanDiego, CA, pp 1–7
27. Chen S, Xu H, Liu D, Hu B, Wang H (2014) A vision of IoT: applications, challenges, and opportunities with China perspective. IEEE Internet Things J 1(4):349–359
28. Farahani B, Firouzi F, Chang V, Badaroglu M, Constant N, Mankodiya K (2018) Towards fog-driven Io T eHealth: promises and challenges of IoT in medicine and healthcare. Future Gener Comput Syst 78:659–676
29. Qi J, Yang P, Waraich A, Deng Z, Zhao Y, Yang Y (2018) Examining sensor-based physical activity recognition and monitoring for healthcare using Internet of Things: a systematic review. J Biomed Inf 87:138–153
30. Qi J, Yang P, Waraich A, Deng Z, Zhao Y, Yang Y (2018) Examining sensor-based physical activity recognition and monitoring for healthcare using Internet of Things: a systematic review J. Biomed Inf 87:138–153

31. Dhanvijay MM, Patil SC (2019) Internet of Things: a survey of enabling technologies in healthcare and its applications. Comput Netw 153:113–131
32. Qadri YA, Nauman A, Zikria YB, Vasilakos AV, Kim SW (2020) The future of healthcare internet of things: a survey of emerging technologies. IEEE Commun Surv Tutorials 22(2):1121–1167
33. Kadhim KT, Alsahlany AM, Wadi SM, Kadhum HT (2020) An overview of patient's health status monitoring system based on internet of things (IoT). Wirel Pers Commun 114(3):2235–2262
34. Ray PP, Dash D, De D (2019) Analysis and monitoring of IoT-assisted human physiological galvanic skin response factor for smart e-healthcare. Sens Rev 39(4):525–541
35. Muthu B et al (2020) IOT based wearable sensor for diseases prediction and symptom analysis in healthcare sector. Peer-to-Peer Netw Appl 13(6):2123–2134
36. Vilela PH, Rodrigues JJ, Solic P, Saleem K, Furtado V (2019) Performance evaluation of a Fog-assisted IoT solution for e-Health applications. Future Gener Comput Syst 97:379–386
37. Awan KM et al (2019) A priority-based congestion-avoidance routing protocol using IoT-based heterogeneous medical sensors for energy efficiency in healthcare wireless body are a networks. Int J Distrib Sens Netw 15(6):1–16
38. Manikandan R, Patan R, Gandomi AH, Sivanesan P, Kalyanaraman H (2020) Hash polynomial two factor decision tree using IoT for smart healthcare scheduling. Exp Syst Appl 141:112924–112930
39. Asif-Ur-Rahman M et al (2019) Towards a heterogeneous mist, fog, and cloud based framework for the internet of healthcare things. IEEE Internet Things J 6(3):4049–4062
40. Awaisi KS, Hussain S, Ahmed M, Khan AA, Ahmed G (2020) Lever aging IoT and fog computing in healthcare systems. IEEE Internet Things Mag 2:52–56
41. Min M et al (2019) Learning-based privacy-aware off loading for healthcare IoT with energy harvesting. IEEE Internet Things J 6(3):4307–4316
42. Wang X, Cai S (2020) Secure healthcare monitoring framework integrating NDN-based IoT with edge cloud. Future Gener Comput Syst 112:320–329
43. Kumar P, Silambarasan K (2019) Enhancing the performance of healthcare service in IoT and cloud using optimized techniques. IETEJ Res 1–10
44. Catarinucci L et al (2015) An IoT-aware architecture for smart healthcare systems. IEEE Internet Things J 2(6):515–526
45. Catherwood PA, Steele D, Little M, Mccomb S, Mclaughlin J (2018) A community-based IoT personalized wireless healthcare solution trial. IEEE J Transl Eng Health Med 6:1–13
46. Abuelkhail A, Baroudi U, Raad M, Sheltami T (2020) Internet of things for healthcare monitoring applications based on RFID clustering scheme. Wirel Netw 27(1):747–763
47. Patan R, Pradeep Ghantasala GS, Sekaran R, Gupta D, Ramachandran M (2020) Smart healthcare and quality of service in IoT using grey filter convolutional based cyber physical system. Sustain Cities Soc 59:102141–102161
48. Kumar PM, Lokesh S, Varatharajan R, Chandra Babu G, Parthasarathy P (2018) Cloud and IoT based disease prediction and diagnosis system for healthcare using Fuzzy neural classifier. Future Gener Comput Syst 86:527–534
49. Tan E, Halim ZA (2019) Healthcare monitoring system and analytics based on internet of things framework. IETEJ Res 65(5):653–660
50. Asghari P, Rahmani AM, Haj Seyyed Javadi H (2019) A medical monitoring scheme and health-medical service composition model in cloud based IoT platform. Trans Emerg Telecomm Technol 30(6):3637–3652
51. Boussada,R.,Hamdane,B.,Elhdhili,M.E.,Saidane,L.A.,Privacy-preservingawaredatatransmissionforIoT-basede-healt.Comput.Network.162,106866–106890, (2019).
52. Tao H, Bhuiyan MZA, Abdalla AN, Hassan MM, Zain JM, Hayajneh T (2019) Secured data collection with hardware-based ciphers for IoT-based healthcare. IEEE Internet Things J 6(1):410–420
53. Guo X, Lin H, Wu Y, Peng M (2020) A new data clustering strategy for enhancing mutual privacy in healthcare IoT systems. Future Gener Comput Syst 113:407–417

54. Pal S, Hitchens M, Varadharajan V, Rabehaja T (2019) Policy-based access control for constrained healthcare resources in the context of the Internet of Things. J Netw Comput Appl 139:57–74
55. Aghili SF, Mala H, Shojafar M, Peris-Lopez P (2019) LACO: lightweight three-factor authentication, access control and ownership transfer scheme for e-health systems in IoT. Future Gener Comput Syst 96:410–424
56. Hou J-L, Yeh K-H (2015) Novel authentication schemes for IoT based healthcare systems. Int J Distrib Sens Netw 11(11):183659
57. Ullah I, Amin NU, Khan MA, Khattak H, Kumari S (2020) An efficient and provable secure certificate-based combined signature, encryption and sign cryption scheme for internet of things (IoT) in mobile health (M-Health) system. J Med Syst 45(1):1–14
58. Rathee G, Sharma A, Saini H, Kumar R, Iqbal R (2019) A hybrid framework for multimedia data processing in IoT-healthcare using blockchain technology. Multimed Tool Appl 79(11):9711–9733
59. Abou-Nassar EM, Iliyasu AM, El-Kafrawy PM, Song O, Bashir AK, El-Latif AAA (2020) DI Trust chain: towards blockchain-based trust models for sustainable healthcare IoT systems. IEEE Access 8:111223–111238
60. Amin M, Shehwar D, Ullah A, Guarda T, Tanveer TA, Anwar S (2020) A deep learning system for healthcare IoT and smartphone malware detection. Neural Comput Appl 32(21):1–12

Improved RPPG Method to Detect BPM from Human Facial Videos

Manpreet Kaur and Naveen Aggarwal

Abstract COVID-19 pandemic has led to a surge in many mobile applications where a medical physician interacts remotely with the patients through video conferencing mode. There is a need to measure the physical parameters of patients during the interaction. Remote photoplethysmography (rPPG) is one of the approaches that can be applied to measure heart rate in terms of beats per minute from the live facial video of patients. The PhysNet model is one of the earlier implementations of rPPG using a supervised learning approach. This model is augmented as a frequency contrast model using contrastive learning. Most of the implementations consider good quality videos and are based either on the publicly available dataset—UBFC-rPPG or on the private dataset. These models are not directly suited for video captured using different mobile phones. Each mobile phone has a different camera sensor and can induce different types of noise in the video. Further, depending upon the camera position, face resolution may vary. In this work, an improved rPPG method to detect beats per minute from human facial videos has been proposed. The model is validated on the UBFC dataset which is augmented by adding random noise varying from 10 to 30%. The dataset is also augmented by changing the resolution of images. The proposed experiments show the varied response of PhysNet and frequency contrast models at different noise levels with an increase in the loss at higher noise levels and lower resolution. Further, testing on the proposed dataset showed better results than the UBFC dataset. The experimentation has provided promising results with a 60% decrease in RMSE and MAE values for both the models.

Keywords Heart rate estimation · Remote photoplethysmography · Beats per minute · Heart rate variability

M. Kaur (✉) · N. Aggarwal
UIET, Panjab University Chandigarh, Chandigarh 160014, India
e-mail: kaur67931@gmail.com

1   Introduction

Heart rate is the key measure of a person's health and is widely used in identifying heart-related problems like uneven and very fast pulses that can result in coagulum. Variations in physiological signals like heart rate, rate of respiration, blood pressure, etc., can be observed using detectors that fall into two categories: invasive and non-invasive. Invasive technologies are based on measurements taken from sensors inserted inside the body. In non-invasive technologies, measurements of body parameters are taken without inserting any sensor inside the body. The non-invasive technologies are further subdivided into contact and non-contact categories.

Photoplethysmography (PPG) [1, 2] is a simple visual method that falls in the category of non-invasive technologies used for heart rate measurement. It is used to measure the reflection or transmission of skin light, which is used to detect heartbeats. The basic idea behind contact PPG is that the amount of plasma, movement of the walls of blood vessels, and orientation of red blood cells influence the quantity of light captured by the sensor and, as a result, affect the PPG signal. Hence, the standard PPG needed some light source for lightening the surface of the skin and a photodetector to monitor the intensity changes in light. So, to overcome the hectic problem of using instruments on the body of the patients which may cause inconvenience, the remote photoplethysmography method is devised. Remote PPG [3–5] (rPPG) is a type of non-contact pulse rate data collection. In rPPG, a video camera detects the subtle light inclusion changes persuaded by blood volume changes in the skin of the person. This is beneficial in situations where contact has to be prevented due to excessive sensitivity (e.g., infants), or where unobtrusiveness is required. The advantages of rPPG over contact PPG are that it only needs light (natural or artificial) and can record the changes in human skin color with a camera. As the camera can be mounted at a distance from the person, it becomes much more comfortable for patients to get diagnosed. That is why rPPG is widely used in realistic heart-related diagnostics.

During the COVID-19 pandemic, rPPG gained more popularity as it provided practitioners to measure vital parameters of patients interacting online. Even though many researchers [6–10] have used good quality cameras to capture video and apply rPPG techniques, recently some researchers [11, 12] have also explored the use of smartphone cameras. Smartphone cameras come in different form factors and have different resolution with AI sensors that smoothens the facial images. These characteristics of the smartphone camera to improve the visual fidelity of captured images become a hindrance to the application of rPPG techniques. They have different types of sensors that induce different types of noise in the videos. This can be due to several factors like heat and temperature, sensor illumination levels, etc. The various publicly available datasets used in the literature like UBFC, COHFACE, PURE, etc., and all these are collected using professional cameras. None of these focused on smartphones to capture the videos. Also, traditional rPPG methods did not estimate heart rate correctly in conditions where a facial video involved distortions or noise in them. This limits their efficiency since different facial emotions produce varying noise characteristics based on the region of the face. There are very few approaches

presented in the past that deal with noise effects on rPPG [13, 14]. Therefore, the main motivation of this work is to contribute to the development of techniques that focus on smartphones as a means of dataset collection and studying the effect of noise and resolution on rPPG algorithms and datasets.

This study investigates the impact of noise and resolution on the machine and deep learning models and different rPPG datasets. Various pre-trained models were studied, and, on the PhysNet and frequency contrast models as proposed by [10], multiple experiments by training on COHFACE [15] and adding noise to the publicly available dataset UBFC [16] at 10–30% levels were performed. Experimentation was also done with the distance and resolution of human faces by cropping the face area from the videos into a rectangular shape. For rigorous testing, a proposed dataset was collected by smartphone. The results obtained after experimentation were discussed in the results section.

2 Related Work

Over the past few years, the need for reasonably clean and steady input data was gradually eliminated as deep learning approaches have progressively demonstrated their ability to learn more effectively across the noise. The types of deep learning models proposed in the past years fall into two categories. These were either the end-to-end models or a combination of deep learning and machine learning models. In the conventional models, the methods fall into two categories—signal estimation and signal extraction. In signal estimation, the EVM-CNN model was given by Qiu et al. [17] in 2018, while in signal extraction, 3D-CNN, which was CNN-based and used deep spatiotemporal networks, was presented by Luguev et al. [18] in 2020. The SPAD model, which used the camera with deep learning methods, was proposed by Parachini et al. [19] in 2020. In the end-to-end deep learning methods, there were numerous models given. A VGG style CNN-based model named DeepPhys in which deep convolutional neural networks were employed, with separate prediction branches customized to motion and intensity was given by Chen and Mcduff [20] in 2018. It was an object recognition model with 19 layers.

After that, in 2019, Reiss et al. [21] proposed the model Deep-PPG, which addressed the shortcomings of the datasets and introduced a new dataset, PPG-DaLiA [22]. Both these models were VGG-style-based CNNs. The CNN-LSTM network named Meta-rPPG, which consisted of two components, one that extracts the features and the other that estimates rPPG, was proposed in 2020 by Lee et al. [23]. Also, in 2018, Spetlik et al. [24] presented a CNN model named HR CNN, which had two steps that employed a two-stage CNN with an estimator network and a per-frame feature extractor. It was a 3D-CNN network. Later, a two-stage end-to-end method that dealt with highly compressed videos was introduced by Yu et al. [25] that consisted of STVEN—spatiotemporal video enhancement network and rPPGNET for rPPG recovery. Their second work that augmented the concept of PhysNet [26] in 2019 explored various aspects of computational properties related

to time series and followed the 2DCNN and 3DCNN architectures. It used deep spatiotemporal networks to reconstruct precise signals of rPPG from the raw face videos. The Siamese-rPPG network was given in 2020 by Yun-Yun Tsou et al. [9], which proposed an rPPG network that simultaneously learn the diverse and comparable features of two facial regions and analyze the monotony of rPPG signals. Other works were by Perepelkina et al. [27] in 2020, a CNN-based model that used synthetic data training which not only included videos but PPG curves also; by Bousefsaf et al. [28] in 2019, which was a 3D CNN classifier; and by Lieu et al. in 2020, named Deep-rPPG, that was based on spatiotemporal convolution. Few works proposed in the past few years that covered the noise issue were discussed here. In 2022, Puneet and Lokendra [13] developed a denoising-rPPG network that improved the heart rate estimation by handling the temporal signals generated by facial expressions. It also eliminated the noise generated by the AUs based on the facial expressions.

After reviewing the literature, it is found that many researchers in the field of computer vision have already addressed the topic of rPPG [29, 30]. They have employed a wide range of approaches ranging from signal processing to machine learning and deep learning [31, 32]. Deep learning, which is popular for solving specific tasks successfully by identifying feature representations that are robust to many types of unwanted variations, has received a wide interest recently. But it is observed that most of the models focused on the datasets captured through professional cameras. These cameras require a fixed setup to capture the facial videos of the subjects and are not portable also. This causes a hectic stress on the person who captures the dataset to manage the setup. Also, not every researcher can afford these, so there is a need to work on the datasets that can be captured through smartphones. Furthermore, only a few previous works concentrated on the noise-robust models and noisy datasets. Due to all these reasons, it is found that there is a need to work on these gaps. We have proposed a dataset captured through smartphone and a technique that studies the impact of noise and resolution on deep learning models. This will provide flexibility to the researchers to create their own datasets from smartphones as they are not much expensive and are portable too.

3   Datasets

Various rPPG datasets are available in the literature, and some of them are publicly available while some are not. For evaluating the models, two publicly available rPPG datasets—COHFACE and UBFC available at [33, 34] —were used along with our newly created rPPG dataset [35] captured by a smartphone. The COHFACE dataset was used for training purposes only due to its greater size. UBFC and the proposed dataset were used for testing. As the ECG-Fitness dataset does not contain PPG as ground truth and MR-NIRP-Indoor is very basic and does not contain varying light and body motion conditions, so they were not used. The structure of the datasets is explained in Table 1, and the snaps of the datasets were shown in Fig. 1.

Table 1 rPPG datasets available in literature

Dataset name	No. of subjects	PPG (Y/N)	Freely available (Y/N)
COHFACE [15]	40	Y	Y
ECG-FITNESS [24]	17	N	Y
MAHNOB [36]	27	Y	N
MMSE-HR [37]	40	N	N
MR-NIRP-CAR [38]	19	Y	Y
MR-NIRP-INDOOR [39]	12	Y	Y
OBF [40]	106	Y	N
PURE [41]	10	Y	N
UBFC-RPPG [16]	42	Y	Y
VIPL-HR [42]	107	Y	N
VIPL-HR-V2 [43]	500	Y	N

Fig. 1 Images of some subjects from COHFACE, UBFC, and proposed dataset

3.1 COHFACE

The COHFACE dataset is made up of 160 one-minute-long videos by 40 subjects, with 12 women and 28 men recorded in two different light illuminations: studio and natural light. Recording was done using a Logitech HD Webcam C525 camera with a resolution of 640 × 480 and a 20 Hz frame rate. The videos of the subjects were synced with their recorded heart rate values and breathing rates. A contact PPG sensor was used to record the blood volume pulse (BVP) and respiratory rate.

3.2 UBFC-RPPG

The UBFC-rPPG dataset consists of 42 videos from 42 people in real situations captured using a Logitech C920 HD Pro webcam with the ground truth PPG and bpm values collected with a pulse oximeter. The subjects' HR were kept varied by asking them to solve a time-sensitive mathematics puzzle. Each subject's video is about 2 min long, and the data from the pulse oximeter was analyzed simultaneously using the CMS50E. To ensure that the results are correct, subjects 11, 18, 20, and 24 were dropped.

3.3 Proposed Dataset

The proposed dataset constitutes of three healthy individuals' physiological signals and nine videos. The dataset is collected from the subjects who are students at the UIET, Panjab University, Chandigarh. The data is acquired in new, more realistic conditions as compared to the UBFC and COHFACE datasets. The average subject age was 25 years old, with a standard deviation of 2.8 years.

Data Collection

The duration of each video in the dataset is around one minute, during which subjects were asked to sit relaxed and do some normal movements. The videos are captured in three different conditions: (i) subject's face without spectacles, (ii) subject's face with spectacles, and iii) subject at a greater distance than previous ones. Each video is recorded by a Samsung GW3 64MP sensor (S5KGW3) with a resolution of 1920 × 1080 pixels at a frame rate of 30 fps, and the ground truth bpm is recorded simultaneously during video recording with the help of a fingertip pulse oximeter, model AB-80.

Preprocessing

After collecting the dataset, it was preprocessed. Firstly, the videos were resized to a resolution of 640 × 480 pixels, and then, the ground truth bpm was distributed

at regular timestamps equal to the number of frames in the video. The file format for the videos and ground truth was selected as mp4 and .csv files. The purpose of preprocessing was to ensure a smooth testing process.

4 Proposed Model Implementation

In this study, the models PhysNet and frequency contrast are retrained and tested on the noisy datasets (UBFC dataset with added noise and the proposed dataset captured through smartphone). The implementation of the proposed model is shown in Fig. 2, and the details of the models along with their architectures are provided in the upcoming subsection.

4.1 PhysNet

PhysNet is an end-to-end deep learning network having spatiotemporal properties that merges the two steps of projecting RGB into a color subspace with higher representation capability and then re-projecting the color subspace to remove information that is not relevant. The overall framework and architecture of PhysNet were shown in Fig. 3a, b. The network takes T-frame facial images with color space as input.
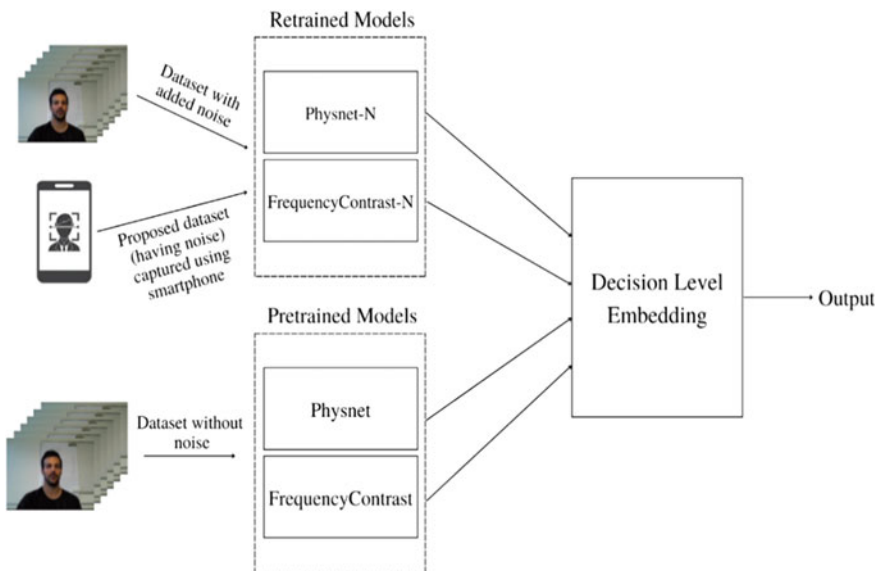
Fig. 2 Proposed model implementation

Fig. 3 **a** Framework of 3D CNN-based PhysNet, **b** network architecture of 3D CNN-based PhysNet

rPPG signal

Aggregation

3D CNN

Images 1 to T

time

(a)

1x1x1 Conv, 1

Spatial Global AvgPool

3x3x3 Conv, 64

3x3x3 Conv, 64

1x2x2 MaxPool

1x5x5 Conv, 32

x4

(b)
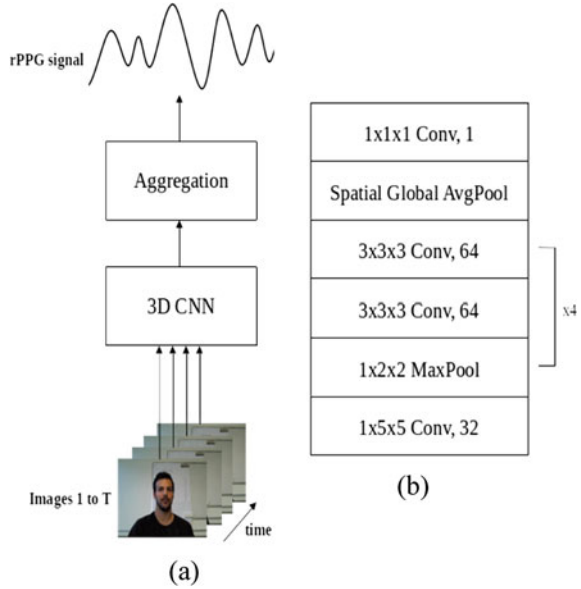
Multichannel vectors are generated after various convolution and pooling operations to represent the spatiotemporal features. Finally, the fundamental manifolds are forecasted into signal space using a convolution operation done channel-wise with a $1 \times 1 \times 1$ kernel to produce the estimated rPPG signal of length. The 3DCNN-based PhysNet shown in Fig. 3b was implemented as the spatiotemporal model, which employs $3 \times 3 \times 3$ convolutions to retrieve meaningful rPPG characteristics in both spatial and temporal dimensions simultaneously. As a result, it was possible to recover rPPG signals with less temporal volatility and learn more dependable context features.

4.2 Frequency Contrast

For PPG estimation, the modified version of PhysNet was also given by [10] as described in Fig. 4. This model reduced the aliasing problems that existed in the original PhysNet model by using interpolation and convolution in the decoder rather than transposed convolution.
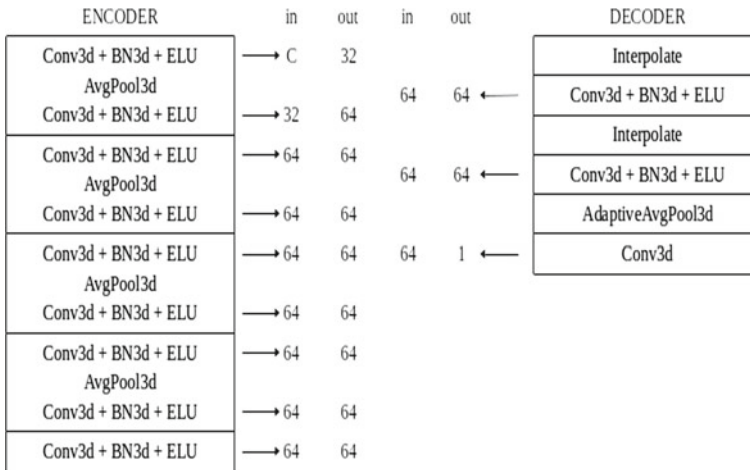
Fig. 4 Frequency contrast: modified PhysNet 3D CNN architecture (C → number of input channels)

5 Experiments and Results

Experimentation is performed on the above-mentioned models and datasets for evaluating their performance on noisy data by adding noise to the UBFC dataset and no noise to the proposed dataset. The noise was added with the help of FFmpeg [44] at 10–30% levels. Resolution of the facial videos was also changed from 640×480 to 480×320 and 320×160. Results of both the models and the datasets were then compared with each other.

5.1 Training on COHFACE Dataset

The two models were implemented using PyTorch 1.7.1 [45], and the training was done by using a single NVIDIA Geforce RTX 3090 GPU. Batch size of 4 for all the experiments, with w_s and w_t set to 1, and the AdamW optimizer was trained over 100 epochs with a learning rate of 10^{-5}. Before training, the facial videos and PPG signals were synchronized. The S3FD face detector [46] was used for each video clip to obtain the face area for the first frame and correct the region for the upcoming frames. S3FD stands for single shot scale-invariant face detector, which performs well on different scales of faces with a single deep neural network, notably for small faces. The face videos were then normalized to 64×64 to give input to the models.

5.2   Evaluation Metrics and Loss Functions

The evaluation metrics used in the experiments were RMSE, MAE, PC, and MCC which stand for root mean square error, mean absolute error, Pearson correlation coefficient and max cross-correlation, respectively. We are already familiar with these terms. MCC [8, 10] is based on the supposition that the HR is within a notable frequency band. MCC determines the correlation at the ideal offset between the signals under the assumption of PPG synchronization. If the heart rate is reasonably steady, this makes the loss more resistant to random temporal offsets present in the ground truth.

5.3   Results

The experimentation revealed the results of an ablation study performed on a UBFC dataset on the basis of noise levels and video resolution. Further details are available in the following sections.

Testing on the UBFC-rPPG dataset

Two different experiments on the UBFC-rPPG dataset of adding noise of 10–30% and changing the resolution of the facial videos to smaller scale were performed. The results obtained are discussed below.

Adding noise to the dataset

After adding noise to the UBFC-rPPG dataset, the results on both the PhysNet and frequency contrast models were evaluated using the discussed evaluation metrics. After performing testing on UBFC dataset variations, it was found that the RMSE and MAE errors for the basic UBFC dataset without any noise are 44.8 and 37.2, which were less than the variation that has 10% noise added to it. Also, when noise was increased to 15%, the value of RMSE and MAE increased. But after that, when the noise was increased more, it had the opposite effect on both RMSE and MAE, i.e., their values started decreasing and went on decreasing to 30% noise. The negative PC and MCC followed a different pattern, i.e., the Neg PC value for the UBFC dataset without any noise was − 0.026, and after adding 10% noise to it, it has a sudden increase in it. But as the noise increased further, it kept on decreasing up to 30% noise.

In the case of frequency contrast, the RMSE and MAE values were 37.8 and 27.8, respectively, which were less than that of UBFC with 10% noise. But as the noise was increased to 15%, the error values decreased. Then for 20% noise, the error values slightly increased. After that, for 30%, they again decreased. So, it is concluded that for the frequency contrast model, the variation of the RMSE and MAE was fluctuating between 10 and 30% in the range of (34–38) and had a wave-like behavior. In the case of negative PC, it followed a z-like pattern in the curve, i.e., up

to 15% Neg PC increases, and after that for 20% noise, it decreased and then again increased for 30% noise. The values of the testing metrics for PhysNet and frequency contrast are given in Tables 2 and 3, respectively.

Figure 5 illustrates the behaviors of RMSE, MAE, negative PC, and negative MC for the various noise levels of UBFC dataset.

Changing Resolution of videos

The UBFC dataset was augmented by changing its resolution of 640 × 480 to 480 × 320 and 320 × 160. It was found that for both the models PhysNet and frequency contrast, the value of loss functions RMSE, MAE, PC, and MCC increased as the resolution of the videos is decreased. It is clear from the results that 480 × 320 resolution has greater loss values than the 640 × 480 resolution and 320 × 160 has more loss than both 480 × 320 and 640 × 480 resolution videos. The results for the PhysNet and frequency contrast models are shown in Tables 4 and 5, respectively.

Figure 6 illustrates the behaviors of RMSE, MAE, negative PC, and negative MC for the three resolutions of UBFC dataset which are 640 × 480, 480 × 320, and 320 × 160.

In graph 640 × 480 resolution is shown by the scale value 640, 480 × 320 is shown by 480 and 320 × 160 by 320 for convenience.

Testing on the proposed dataset

On both the PhysNet and frequency contrast models, testing was done on the proposed dataset. The observed results are shown in Table 6.

We can see from the table that after training on the COHFACE dataset, the PhysNet model gave RMSE and MAE of 23.40 and 23.08, respectively, while the frequency

Table 2 Testing results of UBFC-rPPG dataset noise levels (PhysNet)

Dataset	RMSE	MAE	PC	MCC
UBFC-rPPG	44.8	37.2	− 0.026	− 0.274
UBFC-rPPG (noise 10%)	45.3	37.5	0.008	− 0.219
UBFC-rPPG (noise 15%)	47.6	40.9	− 0.001	− 0.259
UBFC-rPPG (noise 20%)	40.2	30.4	− 0.012	− 0.268
UBFC-rPPG (noise 30%)	39.5	29.7	− 0.037	− 0.262

Table 3 Testing results of UBFC-rPPG dataset noise levels (frequency contrast)

Dataset	RMSE	MAE	PC	MCC
UBFC-rPPG	37.8	27.8	− 0.006	− 0.368
UBFC-rPPG (noise 10%)	38.1	31.1	− 0.032	− 0.274
UBFC-rPPG (noise 15%)	36.6	31.0	− 0.007	− 0.256
UBFC-rPPG (noise 20%)	38.3	31.5	− 0.031	− 0.243
UBFC-rPPG (noise 30%)	34.0	28.7	− 0.038	− 0.280
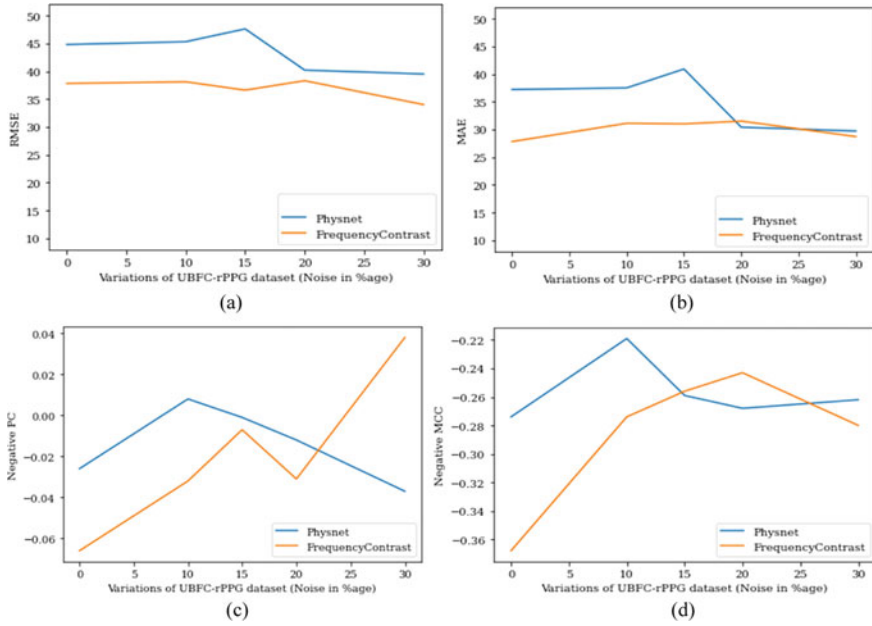
Fig. 5 Comparison of: **a** RMSE, **b** MAE, **c** Neg PC, **d** Neg MCC for UBFC dataset noise levels for models PhysNet and frequency contrast

Table 4 Testing results of UBFC-rPPG dataset resolution changes (PhysNet)

Dataset	RMSE	MAE	PC	MCC
UBFC-rPPG (640 × 480)	44.8	37.2	− 0.026	− 0.274
UBFC-rPPG (480 × 320)	47.1	39.3	0.012	− 0.209
UBFC-rPPG (320 × 160)	47.6	41.5	0.023	− 0.099

Table 5 Testing results of UBFC-rPPG dataset resolution changes (frequency contrast)

Dataset	RMSE	MAE	PC	MCC
UBFC-rPPG (640 × 480)	37.8	27.8	− 0.066	− 0.368
UBFC-rPPG (480 × 320)	39.6	33.9	0.074	− 0.244
UBFC-rPPG (320 × 160)	33.4	28.7	− 0.017	− 0.239

contrast model outputted 9.85 and 8.40 values, respectively. Also, the correlation between the predicted and ground truth values for PhysNet and frequency contrast models was 0.702 and − 0.229, which meant that there was a negative correlation in the case of the frequency contrast model. The standard deviation of predicted versus ground truth values for the models was 3.82 and 7.70, respectively, which concluded that the values were more scattered or dispersed in the case of the second model.
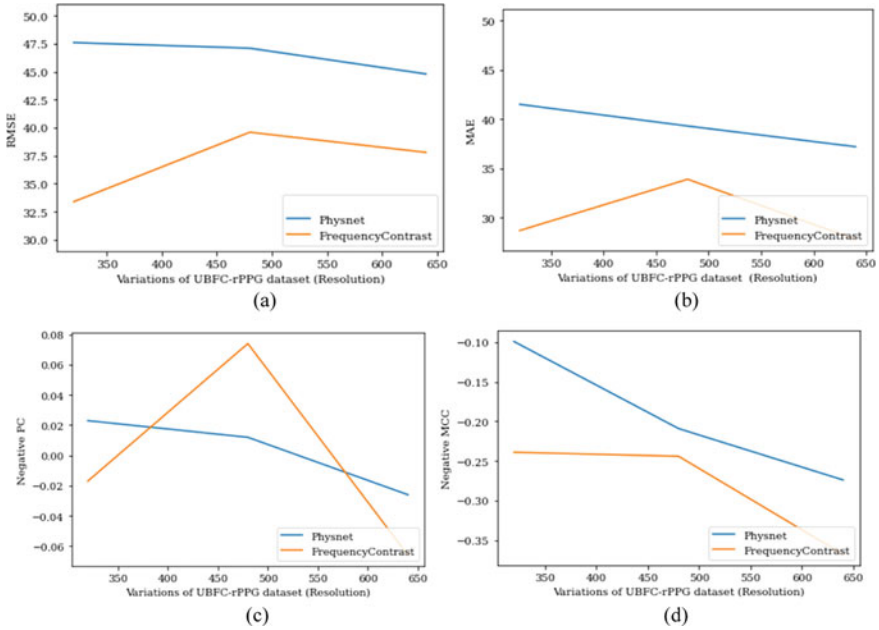
Fig. 6 Comparison of: **a** RMSE, **b** MAE, **c** Neg PC, **d** Neg MCC for UBFC dataset resolution levels for models PhysNet and frequency contrast

Table 6 Testing results of proposed dataset

Model	RMSE	MAE	Corr
PhysNet	23.40	23.08	0.702
Frequency contrast	9.85	8.40	− 0.229

Hence, after the experimentation, decision-level embedding has been used to obtain the output from the results of the different experiments.

6 Conclusion

In this work, two models PhysNet and frequency contrast were trained with the COHFACE dataset and tested with UBFC and the proposed dataset. The UBFC dataset was augmented by adding random noise varying from 10 to 30%. It was also augmented by changing the resolution of videos. Results obtained after experimentation showed that with an increase in noise, the loss of the models increased, and at a certain noise level, it became stable and did not increase. This implied that the models respond to noise and yield more loss in the case of noise. Further, the resolution change of the UBFC dataset discovered that with decrease in the resolution of

the videos, the loss values increased. So, there is a need to work on the noise present in the rPPG datasets and their resolution also. The second experiment showed that the proposed dataset had fewer RMSE and MAE values as compared to the UBFC dataset during testing, so it worked better than the UBFC dataset. Decision-level embedding after performing all the experiments and other above-mentioned findings pave the way for training the models on much larger, unlabeled, less resolution, and noisy datasets and developing new models that work efficiently on noisy datasets, such as those obtained from smartphones. In the future, a higher quality database could be created using smartphones with an increased subject number. This could improve the applicability of rPPG to further demanding situations and domains, like datasets with different environmental conditions.

References

1. Hertzman AB (1937) Photoelectric plethysmography of the fingers and toes in man. Proc Soc Exper Biol Med 37(3):529–534
2. Blažek V, Schultz-Ehrenburg U (1996) Quantitative photoplethysmography: basic facts and examination tests for evaluating peripheral vascular Funktions. Fortschritt-Berichte VDI. VDI-Verlag [online]. Available https://books.google.it/books?id=TfHaAgAACAAJ. Last accessed 1996
3. Wieringa FP, Mastik F, van der Steen AFW (2005) Contactless multiple wavelength photo-plethysmographic imaging: a first step toward "SpO2 camera" technology. Ann Biomed Eng 33(8):1034–1041
4. Humphreys K, Tomas W, Charles M (2007) Noncontact simultaneous dual wavelength photoplethysmography: a further step toward non-contact pulse oximetry. Rev Sci Instrum 78(4):044304
5. Verkruysse W, Svaasand LO, Stuart Nelson J (2008) Remote plethysmographic imaging using ambient light. Opt Express 16(26):21434–21445
6. Wang W, Stuijk S, Haan GD (2015) A novel algorithm for remote photoplethysmography: spatial subspace rotation. IEEE Trans Biomed Eng 63(9):1974–1984
7. Benedetto S, Christian C, Darren C, Greenwood NB, Virginia P, Paolo A (2019) Remote heart rate monitoring-assessment of the Facereader rPPg by Noldus. PloS One 14(11):e0225592
8. Lee E, Evan C, Chen-Yi L (2020) Meta-rppg: remote heart rate estimation using a transductive meta-learner. In: European conference on computer vision. Springer, Cham, pp 392–409
9. Tsou YY, Lee YA, Hsu CT, Hung Chang S (2020) Siamese-rPPG network: remote photo-plethysmography signal estimation from face videos. In: Proceedings of the 35th annual ACM symposium on applied computing, pp 2066–2073
10. Gideon J, Stent S (2021) The way to my heart is through contrastive learning: remote photo-plethysmography from unlabelled video. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 3995–4004
11. Qiao D, Zulkernine F, Masroor R, Rasool R, Jaffar N (2021) Measuring heart rate and heart rate variability with smartphone camera. In: 2021 22nd IEEE international conference on mobile data management (MDM). IEEE, pp 248–249
12. Shoushan MM, Reyes BA, Rodriguez AM, Chong JW (2021) Contactless heart rate variability (HRV) estimation using a smartphone during respiratory maneuvers and body movement. In: 2021 43rd annual international conference of the ieee engineering in medicine & biology society (EMBC). IEEE, pp 84–87
13. Lokendra B, Puneet G (2022) AND-rPPG: a novel denoising-rPPG network for improving remote heart rate estimation. Comput Biol Med 141:105146

14. Lu H, Han H, Zhou SK (2021) Dual-gan: Joint bvp and noise modeling for remote physiological measurement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 12404–12413
15. Guillaume H, Anjos A, Marcel S (2017) A reproducible study on remote heart rate measurement. arXiv:1709.00962
16. Bobbia S, Macwan R, Benezeth Y, Mansouri A, Dubois J (2019) Unsupervised skin tissue segmentation for remote photoplethysmography. Pattern Recogn Lett 124:82–90
17. Qiu Y, Liu Y, Arteaga-Falconi J, Dong H, Saddik AE (2018) EVM-CNN: real-time contactless heart rate estimation from facial video. IEEE Trans Multimedia 21(7):1778–1787
18. Luguev T, Seuß D, Garbas JU (2020) Deep learning based affective sensing with remote photoplethysmography. In: 2020 54th annual conference on information sciences and systems (CISS). IEEE, pp 1–4
19. Paracchini M, Marcon M, Villa F, Zappa F, Tubaro S (2020) Biometric signals estimation using single photon camera and deep learning. Sensors 20(21):6102
20. Chen W, McDuff D (2018) DeepPhys: video-based physiological measurement using convolutional attention networks. In: Proceedings of the European conference on computer vision (ECCV), pp 349–365
21. Reiss A, Indlekofer I, Schmidt P, Laerhoven KV (2019) Deep PPG: large-scale heart rate estimation with convolutional neural networks. Sensors 19(14):3079
22. Available online: https://archive.ics.uci.edu/ml/datasets/PPG-DaLiA
23. Lee E, Chen E, Lee CY (2020) Meta-rppg: remote heart rate estimation using a transductive meta-learner. In: European conference on computer vision. Springer, Cham, pp 392–409
24. Špetlík R, Franc V, Matas J (2018) Visual heart rate estimation with convolutional neural network. In: Proceedings of the British machine vision conference, Newcastle, UK, pp 3–6
25. Yu Z, Peng W., Li, X., Hong, X., Zhao, G.: Remote heart rate measurement from highly compressed facial videos: an end-to-end deep learning solution with video enhancement. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 151–160
26. Yu Z, Li X, Zhao G (2019) Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks. arXiv:1905.02419
27. Perepelkina O, Artemyev M, Churikova M, Grinenko M (2020) HeartTrack: convolutional neural network for remote video-based heart rate monitoring. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp 288–289
28. Bousefsaf F, Pruski A, Maaoui C (2019) 3D convolutional neural networks for remote pulse rate measurement and mapping from facial video. Appl Sci 9(20):4364
29. Sungjun K, Kim H, Suk Park K (2012) Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. In: International conference of the IEEE engineering in medicine and biology society, pp 2174–2177
30. Balakrishnan G, Durand F, Guttag J (2013) Detecting pulse from head motions in video. IEEE Comput Soc Conf Comput Vis Pattern Recogn: 3430–3437. https://doi.org/10.1109/CVPR.2013.440
31. Poh MZ, McDuff DJ, Picard RW (2010) Advancements in noncontact, multiparameter physiological measurements using a webcam. IEEE Trans Biomed Eng 58(1):7–11
32. Poh MZ, McDuff DJ, Picard RW (2010) Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Opt Express 18:10762–74. https://doi.org/10.1364/OE.18.010762
33. COHFACE dataset. https://www.idiap.ch/dataset/cohface
34. UBFC dataset. https://sites.google.com/view/ybenezeth/ubfcrppg
35. Proposed dataset. https://drive.google.com/drive/folders/1SEof0O7FjbM5is_1tU40VYBmHIk__mWj?usp=sharing
36. Soleymani M, Lichtenauer J, Pun T, Pantic M (2011) A multimodal database for affect recognition and implicit tagging. IEEE Trans Affect Comput 3(1):42–55
37. Tulyakov S, Alameda-Pineda X, Ricci E, Yin L, Cohn JF, Sebe N (2016) Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 2396–2404

38. Nowara EM, Marks TK, Mansour H, Veeraraghavan H (2020) Near-infrared imaging photoplethysmography during driving. IEEE Trans Intell Transp Syst
39. Nowara EM, Marks TK, Mansour H, Veeraraghavan A (2018) SparsePPG: towards driver monitoring using camera-based vital signs estimation in near-infrared. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 1272–1281
40. Li X, Alikhani I, Shi J, Seppanen T, Junttila J, Majamaa-Voltti K, Tulppo M, Zhao G (2018) The OBF database: a large face video database for remote physiological signal measurement and atrial fibrillation detection. In: 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018). IEEE, pp 242–249
41. Stricker R, Müller S, Gross HM (2014) Non-contact video-based pulse rate measurement on a mobile service robot. In: The 23rd IEEE international symposium on robot and human interactive communication. IEEE, pp 1056–1062
42. Niu X, Han H, Shan S, Chen X (2018) VIPL-HR: a multi-modal database for pulse estimation from less-constrained face video. In: Asian conference on computer vision. Springer, Cham, pp 562–576
43. Li X, Han H, Lu H, Niu X, Yu Z, Dantcheva A, Zhao G, Shan S (2020) The 1st challenge on remote physiological signal sensing (REPSS). In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp 314–315
44. FFmpeg Developers. ffmpeg tool (version be1d324) [Software]. Available from http://ffmpeg.org/ (2016)
45. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S (2019) PyTorch: an imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E, Garnett R (eds) NeurIPS, pp 8024–8035
46. Zhang S, Zhu X, Lei Z, Shi H, Wang X, Li SZ (2017) S3fd: Single shot scale-invariant face detector. In: Proceedings of the IEEE international conference on computer vision, pp 192–201

A Hybrid Classification Model for Prediction of Academic Performance of Students: An EDM Application

Amandeep Kaur and Deepti Gupta

Abstract A technology that aims at predicting the future possibilities based upon the past and present experiences is defined as prediction analysis. For performing prediction analysis, neural networks and classification techniques are implemented. Educational data mining (EDM) and learning analytics (LA) are the two communities grown together and associated with same research area, i.e., education. The main objective of the educational institution is providing the quality education to their students. The student performance prediction has several phases which involves preprocessing, feature extraction and classification. In this research work, novel model is proposed which is the combination of principal component analysis (PCA) and voting classifier for the student performance prediction. The voting classifier is the combination of multiple classifiers which are Gaussian Naïve Bayes, Random forest, k-NN and SVM. The approach used in this research work increases the accuracy for the prediction, obtained improved accuracy is 98.95%, and upgraded precision–recall and f-score values obtained is 99%. The novel approach is executed in Python using Anaconda framework, and outcome is examined with respect to the parameters, i.e., accuracy, precision, recall and f-score.

Keywords Data mining · Educational data mining · Gaussian Naïve Bayes · k-NN · Random forest · SVM · PCA · Voting classifier

A. Kaur (✉) · D. Gupta
Department of Computer Science and Engineering, UIET, Panjab University, Chandigarh 160014, India
e-mail: amandeepkaur.kaur06@gmail.com

D. Gupta
e-mail: deeptigupta@pu.ac.in

1   Introduction

The method of extracting the unseen predictive statistics from databases of considerable size is known as data mining (DM). It is an intense new technology with huge potential that assists organizations and focuses on the crucial information in the data warehouses. DM tools forecast upcoming trends and techniques, allowing work to be dynamic and knowledge-driven decisions. The methods used in DM are the aftereffect of an extensive procedure related to research and development. The advancement started when data related to trading and commerce begin to gathered on PCs, proceeded with improvements in data access, generated advances which permit clients to navigate through real time data. DM techniques yield the advantages of mechanization on former hardware and software applications that is applied on new systems as present applications are overhauled and innovative products are evolved, evaluating DM tools on multiprocessor system along with analysis of massive databases within minutes. Speedier processing implies clients can examine various models for understanding complicated data rapidly that makes analysis of large volume data easier for the clients. Therefore, huge database yields enhanced predictions [1]. The method used for depicting this performance in DM is called as modeling. Modeling demonstrates the evolving of model in one situation where the answer is known further, to seek in other situation whose answer is not known [2].

Data mining in education is an evolving interdisciplinary research field called educational data mining (EDM). Web-based educational system has been steadily rising which led to the storage of large volume of data. All this information delivers goldmine of learning data which can be explored and exploited to understand how student learn. Among the numerous problems in the learning process, student dropout is utmost complex and negative ones, both for the learner and for the institution as well. Large volume of significant data from different resources that comes from dissimilar formats and at different granularity levels can be stored in educational information system. Every specific educational problem has an explicit purpose with distinct characteristics that need different action of mining problem. The earlier data mining technique cannot be applied directly to these kinds of problems. As a consequence, the knowledge discovery process has been adapted, and few methods of DM are utilized [3].

1.1   EDM and LA

EDM and LA are the two major communities that grown together around the same research field related to education. EDM focusses on evolving, researching and implementing automated technique to identify patterns in huge collections of data; otherwise, it would be impractical to examine due to vast data in which it exists. LA is the assembling, measurement, surveying and reporting of information related to student, with the aim of interpreting along with optimizing learning as well as the

Fig. 1 Main areas related to EDM [4]
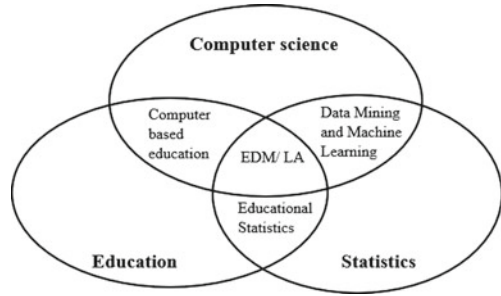
domains where it exists. Both the communities have common interest of data decision in scholastic research and having same objective of enhancing learning practices. EDM and LA are collaborative fields that are not limited to information retrieval (IR), recommender systems, visual data analysis, DM, social network analysis and so on. Moreover, both of them are considered to be the integration of three primary domains, namely computer science, education and statistics [4].

Figure 1 represents the mentioned three areas and their sub fields which are computer-based education (CBE), educational statistics, DM/machine learning (ML) and EDM/LA. EDM methods can be categorized with reference to its aim, i.e., clustering, classification and relationship mining.

1.2 Methods Used in EDM

Cluster analysis assembles data objects depending upon the information in the data that depicts the objects as well as relationships among them. The objective of clustering is that the objects within a cluster can be comparable to each other and not the similar as (or unrelated to) the objects in another cluster. Clustering is an unsupervised technique of classifying the samples into groups. Each one group known as cluster, containing objects that are similar between themselves and different to objects of other groups [5]. Clustering technique is viewed as a form of classification such that it labels the instance with class (cluster) labels. Whereas, classification is supervised learning technique in which new unlabeled instances are assigned to a class by utilizing a model developed from objects with well-known class labels. Hence, cluster analysis is now and then alluded to as unsupervised classification. When the classification utilized with no capability inside data mining, it regularly alludes to supervised classification [6]. The classification process involves two stages where the supervised learning is used initially for predefined class label to train data set. Next stage involves evaluating classification accuracy. Similarly, data prediction is also a two-stage process. DM algorithm uses the three distinct learning approaches: supervised, unsupervised or semi-supervised [7].

In classification task, every instance associated with a class is represented by the value of a unique target attribute or essentially the class attribute. The target attribute can go up against categorical values, each of them relating to a class [8]. The set of examples in the classification process is split into two sets, i.e., training and testing set [9]. Classification algorithms have the major goal of maximizing the predictive accuracy acquired by the model while classifying instances in the test set that remains unnoticeable during training process [10].

The paper is organized as: Sect. 2 describes the previous work done in EDM. Section 3 demonstrates the research methodology. Section 4 represents the results obtained, and comparative analysis has been done on the basis of evaluation metrics. Finally, Sect. 5 contains the conclusions and future work.

2 Related Work

Most of the researcher have done their research in data mining for educational purposes to get the prediction of students' achievement. Ragab et al. used ensemble technique for predicting the substitute model based on DM technique. For better results, several classifiers are combined together and then added to ensemble classifier with the help of voting method. Bagging method achieved an improved accuracy with decision tree (DT), with an increase in accuracy from 90.4 to 91.4% [11]. Prada et al. presented an online software tool that focuses on evaluating the student performance. This tool focuses on analyzing the performance of the students, with reference to the visible marks and accomplishment of their education. The advantage of this tool is that it does not require knowledge of data scientist. Clustering and visualization assist an analyst for finding patterns in the low-dimensional illustrating the data of scholars. Primary outcomes of classification and drop-outs are satisfactory as accuracy exceeded 90% in few scenarios [12]. Mengash H. A. focused on decision making in universities during the admissions by virtue of DM methods to forecast candidates' educational achievement at the university. This work used 2309 data related to students admitted in the College of Computer Science and Information of Saudi Public University in the span of 2016–2019 for validating the introduced method. The results indicate that initial admission criterion which is academic accomplishment entrance test marks yields highest accuracy in predicting the performance of students, and hence, this score should be given more importance in the admission process. Moreover, accuracy reaches above 79% using artificial neural network (ANN) [13].

Mihaescu et al. presented a concise, yet enlightening, explanation of the most commonly employed open-source data sources together with their respective EDM functions, algorithms used, test results and key discoveries [14]. Lien et al. aimed to overcome the gap by analyzing teachers' prediction quality where tutors foresee the scholar sequential actions called as expert prediction, and machine prediction system was developed and observed that there is a significant lower accuracy of expert prediction system compared to machine prediction algorithm [15] also summarized

in Table 1. Acharya et al. discussed that the expected graduate student always faced a problem in determining universities of their choice. Linear regression (LR), support vector machine (SVM) and decision tree were applied and concluded that linear regression performs well with low mean square error (0.0048) and high root mean square (0.724) [16].

Educational data mining focuses on studying and contributing with outcomes that seeks to find hidden problems and potential measures. Do Nascimento et al. involved the use of two non-parametric methods, i.e., quantile regression and support vector regression, for predicting the school dropout outcomes in the Brazil. The model is developed on the basis of Crisp-DM, and the performance is measured with respect to mean of the absolute error (MAE). The outcomes concluded excellent performance by support vector regression (SVR) classifier [17].

Rahman et al. used some data mining methods to measure the impact of two classes of features (behavior and absentee student). This work conducts classification using four classifier frameworks, namely Naïve Bayes (NB), ANN, DT and k-NN. In order to get more accurate results, this work also uses ensemble techniques such as bagging, random forest and AdaBoost. The technique used to discover the

Table 1 Related work done

Author	Year	Approach	Results
Ragab et al. [11]	2021	Several classifiers were combined together and then added to ensemble classifier using voting procedure	Bagging method achieved an improved accuracy with DT, with an increase in accuracy from 90.4 to 91.4%
Prada et al. [12]	2020	k-means algorithm was applied to form clusters, then PCA is used to reduce the dimensions, and for classification, SVM is used	Accuracy ranges in between 85 and 91%
Mengash [13]	2020	Linear regression and classification algorithms were applied (DT, SVM, ANN and NB)	Accuracy reaches above 79% using ANN technique
Lien et al. [15]	2020	Sequential pattern mining and machine prediction algorithm	Improved mean precision (MP) and mean reciprocal ranking (MRR) are 45.6% and 0.597, respectively
Acharya et al. [16]	2019	Computed and compared the error rate for different models (LR, SVM and DT)	LR perform well with low MSE of 0.0048 and high root mean square of 0.724
Rahman et al. [18]	2017	Measured the impact of two classes of features (behavior and absences) using DM techniques and also used the ensemble techniques such as bagging, random forest and AdaBoost	After applying ensemble approach, ANN gives the finest accuracy of 84.3%

misclassified examples that occur in the training set is ensemble filtering which optimizes the prediction accuracy. The outcomes describe that ensemble filtering and AdaBoost on ANN, respectively, yield 84.3% [18].

3   Research Methodology

In this study, scholar's academic performance is evaluated with the help of PCA and voting classifier as shown in Fig. 2. The methodology proposed to evaluate the performance of learner is given by following steps:
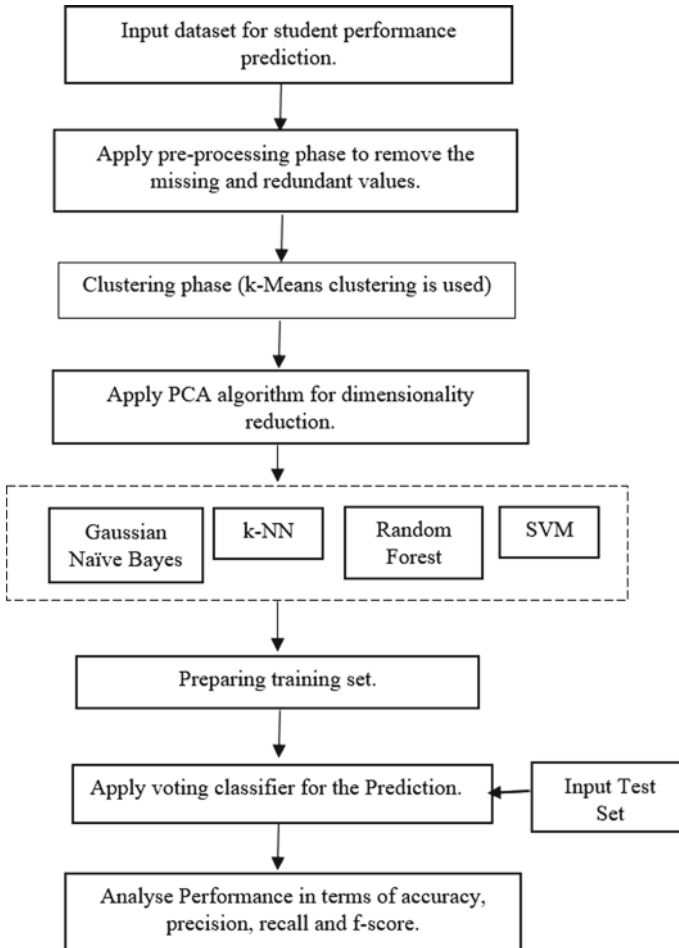
Fig. 2 Proposed model

Step 1: The dataset "student's academic performance dataset" is gathered from e-learning system known as KALBOARD 360 using learning management system. This dataset linked with students' performance as well as their behavioral consists of 16 attributes and 480 instances. The behavioral features that are used in this research work are participation in the discussion, announcement view, absences in the class and number of times visiting resources. The preprocessing of dataset is performed, and later no missing values are present.

Step 2: After data preprocessing, clustering is applied for dividing data points into the clusters such that similar data points belong to same cluster. For applying clustering, k-means algorithm is used, which is unsupervised and iterative technique that aim to find the local maxima at each point. It splits n data points into k-clusters.

Step 3: Next, the features are extracted where connection among attributes is established. For feature extraction, PCA technique will be applied for reducing the features [19]. PCA is generally used in the alteration of set of interconnected factors into a group of linearly independent subsets which further depends upon the variation resulting in distinct variables. It is also called orthogonal linear transformation that projects primary dataset into other projection approach aiming that the largest variance having projection of the first coordinate (first principal component), while the second largest variance having projection of the second coordinate given, and it is vertical to the first component. The target of PCA is to detect a linear transformation expressed as $z = W_k^T x$, where $x \in R^d$, and $r < d$, for improving the variance of the data in projected space. For a data matrix given $X = \{x_1, x_2, \ldots, x_i\}$, $x_i \in R^d$, $z \in R^r$ and $r < d$, the transformation can be depicted by a group of p-dimensional vectors of weights $W = \{w_1, w_2, \ldots, w_p\}$, $w_p \in R^k$, that matches every x_i vector of X.

$$t_{k(i)} = W_{|(i)|} T_{x_i} \tag{1}$$

In order to improve the variance, the condition that must be satisfied by an initial weight W_1 is given below:

$$W_i = \text{argargmax}_{|w|} = \left\{ \sum_i^n (x_i . W)^2 \right\} \tag{2}$$

An extension of the Eq. 2 is as follows:

$$W_i = \text{argargmax}_{\|w\|=1} \left\{ \|X.W\|^2 \right\} = \text{argargmax}_{\|w\|=1} \left\{ W^T X^T X W \right\} \tag{3}$$

A symmetric grid for, e.g., $X^T X$, can be evaluated by deriving the largest eigenvalue of the matrix, where the associated eigenvalue is represented by W. After W_1

isobtained, the primary principal component is derived by projecting the preliminary data matrix X onto the W_1 obtained from the transformation.

Step 4:   The voting classifier [20] is implemented for the prediction in the fourth phase. This method is the combination of Gaussian Naïve Bayes, k-NN, random forest and SVM. Naïve Bayes algorithm is based on Bayes theorem. These algorithms take into account that a value of a certain feature does not depend upon the other features, and the assumption that often considered while working with continuous data is the values associated with each class distributed based on the Gaussian distribution. The likelihood of features is calculated with the help of Eq. (4).

$$P(x_i|y) = \frac{1}{\sqrt{2\pi \sigma_y^2}} exp\left(\frac{-\left(x_{i-}\mu_y\right)^2}{2\sigma_y^2}\right) \tag{4}$$

Various trees are assembled in random forest (RF) which are integrated to acquire exact and suitable results. It is used for classification as well as regression. The hyperparameters used in this approach are analogous to decision tree and bagging classifier. Hence, random forest classifier comprises various decision tree on multiple subsets of given dataset which further takes the average to enhance the accuracy of the particular dataset [21]. Every single tree is established based on a training sample set and a random variable. Random variable belongs to n-represented as Θ_n, after n times operating, the sequence obtained by the classifier $\{h_1(x), h_2(x) \ldots h_n(x)\}$, where x denotes the input vector, applies this to at least one classification system, $H(x)$ denotes the combination of classification models, and Y represents the output variable; the concluding result of this system is drawn by usual majority vote, and the decision function is:

$$H(x) = \arg\left(\max \sum_{i=1}^{n} I(h_i(x) = Y)\right) \tag{5}$$

k-NN is simple classifier which is based on supervised machine learning classification technique [22, 23]. It considers the resemblance between the new information and the existing sample which further place the new sample into the class, alike to the given sample. k-NN just stores the data during the training phase, and when new sample arises, then it classifies data into a category that is most similar one. The sample is categorized on the basis of Euclidian distance, with this the distance of sample with nearby points is calculated, and nearest distance is chosen. The below equation represents the formula of calculating Euclidian distance.

$$ED = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{6}$$

Support vector machine (SVM) firstly maps the original data space to transform it into high dimensional feature space to further find the optimal linear classification surface. It is used to locate the best classification function to differentiate among individuals from the other categories in training data. The criterion for the idea of the "best" classification function can be acknowledged geometrically. In linearly separable dataset, a linear classification function correlates to a dissociated hyperplane $f(x)$ that passes over the mid of the two categories, splitting the classes. Since numerous linear hyperplanes are there, SVM guarantees that finest function to be found by amplifying the margin among the two classes. The margin is the separation between the two classes as defined by the hyperplane [24]. SVM aims to make either the optimal line or decision borderline which can separate n-dimensional space into classes so that new sample arise could be placed at correct category or class.

Step 4: The last step involves evaluation of performance with respect to the parameters—accuracy, precision, recall and f-score.

4 Results and Discussion

This research work is based on the learner's performance prediction. In this work, a novel model is designed which is the combination of multiple models like PCA and voting classifier. In Python 3.9.4, the proposed model is implemented with sklearn libraries, pandas, matplotlib and required library.

4.1 Evaluation Metrics

The evaluation of the proposed model is tested against the previously applied techniques in terms of accuracy, precision, recall and f-score metrics as discussed below:

- Accuracy—It denotes the ratio of number of correct predictions to the total number of input instances. Accuracy represents the percentage of rightly predicted outcomes, and it is calculated with the help of Eq. (7):

$$\text{Accuracy} = \frac{P_t + N_t}{P_t + N_t + P_f + N_f} \tag{7}$$

  where true positive (P_t) represents the number of samples that are exactly predicted as positive. False positive (P_f) indicates the samples that are improperly predicted as positive. True negative (N_t) specifies the samples that are rightly predicted as negative, and false negative (N_f) represents the samples that improperly predicted as negative.
- Precision—It is also positive predicted values; precision represents the ratio of actual positive among the total number of positively classified and is measured

with equation given below:

$$\text{Precision} = \frac{P_t}{P_t + P_f} \tag{8}$$

- Recall—It is also known as sensitivity and represents the positives that are rightly predicted as positive.

$$\text{Recall} = \frac{P_t}{P_t + N_f} \tag{9}$$

- *F*-Score—It represents the harmonic mean between sensitivity and precision.

$$F - \text{Score} = 2 * \frac{1}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} \tag{10}$$

4.2 Results

Table 2 represents the performance analysis on the basis evaluation metrics—accuracy, precision, recall and *f*-score values.

Table 3 demonstrates the percentage improvement in the performance metrics of the proposed model over the existing approaches.

As shown in Fig. 3, the accuracy of various models like decision tree, k-NN, random forest, SVM and proposed model is compared for the student performance

Table 2 Performance analysis

Model	Accuracy (%)	Precision (%)	Recall (%)	*F*-score (%)
Decision tree	80.20	80.6	80	80
Random forest	88.54	88	88	88.3
k-NN	81.25	80.3	82.6	81
SVM	82.29	84	80	81.6
Proposed model	98.95	99	99	99

Table 3 Percentage improvement of the proposed model over existing approaches

Percentage improvement with respect to the existing approaches				
Metrics	Decision tree (%)	Random forest (%)	k-NN	SVM (%)
Accuracy	18.75	10.41	17.7	16.66
Precision	18.4	11	18.7	15
Recall	19	11	16.4	19
F-score	19	10.7	18	17.4
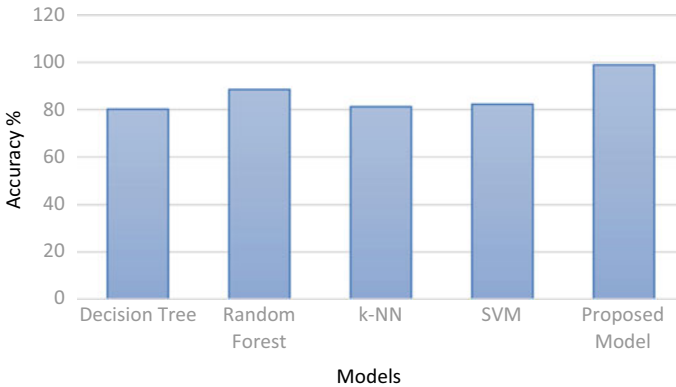
Fig. 3 Accuracy analysis

prediction. It is examined that the approach used in this research work achieved maximum accuracy which is 98.95%. The proposed work achieves $10.41 \approx 10\%$ higher accuracy as compared to previously applied or existing models among which random forest has maximum accuracy (88.54%).

Figure 4 shows the comparison of precision and recall values of proposed work with other previously applied techniques such as decision tree, k-NN, random forest and SVM. The proposed model achieves precision–recall values 99% which is 11% higher as compared to previously applied models among which random forest has the finest precision–recall value which is 88%.

Figure 5 represents the comparative analysis based on the f-score parameter, and it is clearly identified from the graph that the approach used in this research work has the finest f-score value, i.e., 99% as compared to the existing techniques. The f-score of the proposed approach increases by $10.7 \approx 11\%$ as compared to other existing approaches.
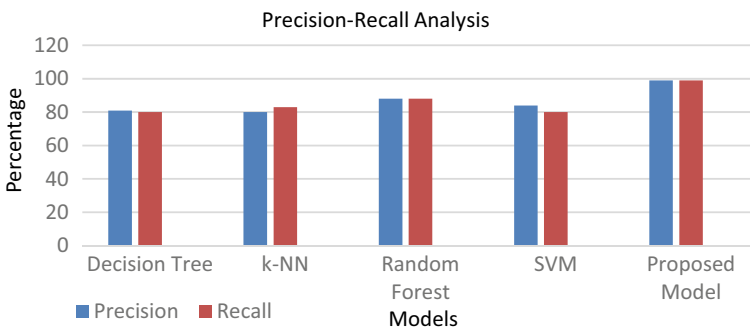
Fig. 4 Analysis of precision and recall
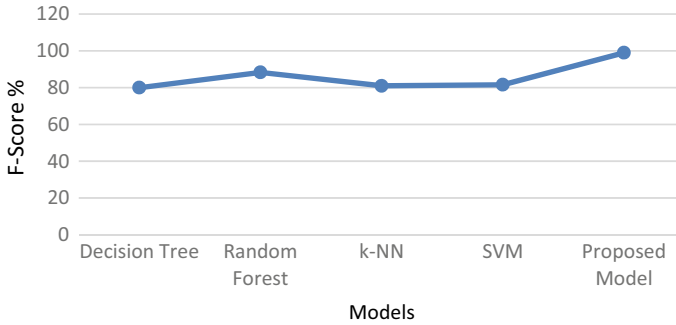
Fig. 5 *F*-score analysis

5 Conclusion

The prediction analysis approach predicts the future possibility with the help of the given current information, and it is practiced with the help of classification technique. The classification technique allocates data into certain target sets. The novel model which is developed in this research work is a hybrid model which is the combination of PCA and voting classifier for the scholar accomplishment prediction. The voting classifier is the combination of four classifier which are Gaussian Naïve Bayes, k-NN, random forest and SVM. The proposed model is implemented in Python, and outcomes are evaluated with respect to accuracy, precision, recall and *f*-score parameters. The upgraded accuracy achieved by using this approach is 98.95%, and precision–recall and *f*-score are 99%. The proposed model directly shows the 10% improvement in accuracy and 11% improvement in precision-recall and f-score as compared to existing models which are DT, random forest, k-NN and SVM. In future, deep learning technique can be used which may reduce the execution time.

References

1. Osmanbegović E, Suljić M (2012) Data mining approach for predicting student performance. Econ Rev J Econ Bus 10(1):3–12
2. Chaudhari KP, Sharma RA, Jha SS, Bari RJ (2017) Student performance prediction system using data mining approach. Int J Adv Res Comput Commun Eng 6(3):833–839
3. Baek C, Doleck T (2019) Educational data mining versus learning analytics: a review of publications from 2015 to 2019. Interact Learn Environ 28(8)
4. Romero C, Ventura S (2020) Educational data mining and learning analytics: an updated survey. WIREs Data Min Knowl Disc 10(3)
5. Khan A, Ghosh SK (2021) Student performance analysis and prediction in classroom learning: a review of educational data mining studies. Educ Inf Technol 26:205–240
6. Kumar AD, Selvam RP, Palanisamy V (2021) Hybrid classification algorithms for predicting student performance. In: 2021 IEEE international conference on artificial intelligence and smart systems (ICAIS). IEEE, Coimbatore, India, pp 1074–1079

7. Alloghani M, Al-Jumeily D, Mustafina J, Hussain A, Aljaaf AJ (2019) A systematic review on supervised and unsupervised machine learning algorithm for data science. In: Supervised and unsupervised learning for data science. Springer Conway, Arkansas, USA, pp 3–21

8. Patil R, Salunke S, Kalbhor M, Lomte R (2018) Prediction system for student performance using data mining classification. In: 2018, fourth international conference on computing communication control and automation (ICCUBEA). IEEE Pune, India, pp 1–4

9. Jain A, Solanki S (2019) An efficient approach for multiclass student performance prediction based upon machine learning. In: 2019, international conference on communication and electronics systems (ICCES). IEEE, Coimbatore, India, pp 1457–1462

10. Sarker IH (2021) Machine learning: algorithms, real-world applications and research directions. SN Comput Sci 2:160

11. Ragab M, Aal AMAK, Jifri AO, Omran NF (2021) Enhancement of predicting students performance model using ensemble approaches and educational data mining techniques. Wirel Commun Mobile Comput 2021

12. Prada MA et al (2020) Educational data mining for tutoring support in higher education: a web-based tool case study in engineering degrees. IEEE Access 8:212818–212836

13. Mengash HA (2020) Using data mining techniques to predict student performance to support decision making in university admission systems. IEEE Access 8:55462–55470

14. Mihaescu MC, Popescu PS (2021) Review on publicly available dataset for educational data mining. WIREs Data Min Knowl Disc 11(3)

15. Norm Lien YC, Wu WJ, Lu YL (2020) How well do teachers predict students' actions in solving an ill-defined problem in STEM education: a solution using sequential pattern mining. IEEE Access 8:134976–134986

16. Acharya MS, Armaan A, Antony AS (2019) A comparison of regression models for prediction of graduate admissions. In: 2019 international conference on computational intelligence in data science (ICCIDS). IEEE, Chennai, India, pp 1–5

17. do Nascimento RLS, das Neves Junior RB, de Almeida Neto MA, de Araújo Fagundes RA (2018) Educational data mining: an application of regressors in predicting school dropout. In: Perner P (ed) Machine learning and data mining in pattern recognition. MLDM 2018. Lecture notes in computer science, vol 10935. Springer, New York, USA

18. Rahman MH, Islam MR (2017) Predict student's academic performance and evaluate the impact of different attributes on the performance using data mining techniques. In: 2017 2nd international conference on electrical & electronic engineering (ICEEE). IEEE, Rajshahi, Bangladesh, pp 1–4

19. Verma C, Stoffova V, Illes Z, Tanwar S, Kumar N (2020) Machine learning-based student's native place identification for real time. IEEE Access 8:130840–213854

20. Ashraf M, Zaman M, Ahmed M (2020) An intelligent prediction system for educational data mining based on ensemble and filtering approach. Procedia Comput Sci 167:1471–1483

21. Batool S, Rashid J, Nisar MW, Kim J, Mahmood T, Hussain A (2021) A random forest students' performance prediction (RFSPP) model based on students' demographic features. In: 2021 Mohammad Ali Jinnah University international conference on computing (MAJICC). IEEE, Karachi, Pakistan, pp 1–4

22. Abdelkader HE, Gad AG, Abohany AA, Sorour SE (2022) An efficient data mining technique for assessing satisfaction level with online learning for higher education students during the covid-19. IEEE Access 10:6286–6303

23. Mohammad LJ, Islam MM, Usman SS, Ayon SI (2020) Predictive data mining models for novel coronavirus (COVID-19) infected patients' recovery. SN Comput Sci 1:206

24. Dai T, Dong Y (2020) Introduction of SVM related theory and its application research. In: 2020 3rd international conference on advanced electronics materials, computers and software engineering (AEMCSE). IEEE, Shenzhen, China, pp 230–233

Enhancement of Accuracy and Performance of Machine Learning System During Detection of Phishing Emails

Pallavi Sharma, Rohit Kumar, and Shelly Kalsi

Abstract Fraudulent attempts to get access to personal data like passwords or bank account details are known as phishing. Phishing scams are among the most common types of online fraud. Online payment, Webmail, banking institutions, file hosting, and cloud storage are just a few of the places where phishing attacks may take place. Among all industries, phishing was most prevalent in Webmail and online payment. Blacklist, heuristics, visual similarity, and machine learning are just a few of the anti-phishing solutions available. Since the blacklist strategy is straightforward to apply and does not detect new phishing attempts, it is the most widely employed. Detecting phishing is easier with the help of ML. It also eliminates one of the main drawbacks of the current method. We conducted a comprehensive review of the literature and developed a novel technique based on feature extraction and a machine learning algorithm for detecting phishing Websites.

Keywords Machine learning · Phishing emails · Security · Classification

1 Introduction

Phishing is seen as a serious threat in today's society, one that grows in importance year after year. It is a kind of fraud that utilizes social engineering and cutting-edge technology to steal sensitive information like usernames and passwords from unsuspecting victims. To put it another way, according to Lungu and Tabusca, the current economic crisis may be attributed to a rise in online attacks and abuses.

P. Sharma (✉) · R. Kumar
UIET, Panjab University, Chandigarh 160014, India
e-mail: pallusharma1999@gmail.com

R. Kumar
e-mail: rklachotra@gmail.com

S. Kalsi
GWC Prade, Jammu, J&K, India

According to the method of transmission, phishing strategies may be classed into malware, phishing emails, and phony Websites. Machine learning plays a crucial part in computer science research, and such research has already influenced real-world applications. Within the realm of information security, it is impossible to overestimate the potential applications of machine learning.

1.1 Machine Learning

It is possible to predict the result of a software application using a non-programming approach. Machine learning algorithms utilize prior data as input to predict future output values. Fraud detection, waste removal, virus detection, BPA, and predictive maintenance are just some of the uses of this technology (PM). Classical machine learning algorithms are often distinguished by how well they predict the future. Unsupervised and supervised learning are two of the most prevalent ways of learning. Algorithm selection may be predicted using data, according to scientists. Nowadays, machine learning may be found in a broad variety of settings and tasks.

Working on Supervised ML

With supervised learning, models are taught to recognize different categories of data by being exposed to labeled datasets. After the model has been trained, it is validated using test data and then used to make predictions about the output. Learning a function that maps an input to output by seeing examples of such mappings is the job of supervised learning (SL) in machine learning. This method infers a function from a collection of training samples that have been annotated with labels.

Working of Unsupervised ML

In unsupervised learning, AI algorithms are used for data that has not been labeled or classified in any way. The results are determined by the predetermined algorithms. Putting a system through its paces in the realm of unsupervised learning is a tried and true method of evaluating its efficacy. When there are no pre-defined categories available, unsupervised learning allows an AI system to organize data based on similarities and differences.

Working on Reinforcement Learning

The field of machine learning includes reinforcement learning. Here, agents learn to use reward and punishment systems on their own. The goal is to maximize benefits while minimizing costs based on what you learn about a given circumstance. It sends messages both for good and bad actions. A reinforcement learning agent can understand its surroundings, act on that understanding, and improve its performance via trial and error.
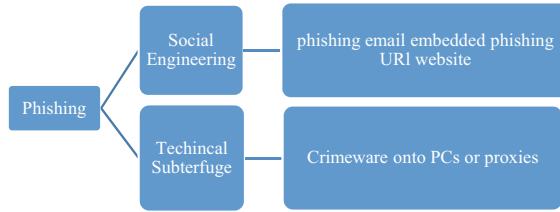
Fig. 1 Types of phishing emails

1.2 Phishing Emails

Spam emails include phishing emails. Scammers pose as reputable companies or banks and persuade users to click on a link included in the email. For private information like usernames, passwords, or credit card numbers, the link will take the user to an illegitimate Website that asks for such information.

There are two methods used by phishes to accomplish their objectives: misleading phishing and malware-based phishing (Fig. 1). By using social engineering techniques, such as sending phishing emails that seem like they came from a reputable organization or bank, the recipient is tricked into visiting a fake Website that asks for personal information, such as passwords and credit card details (Fig. 2).

The phishing technique's life cycle is shown in Fig. 1. The first step is to send emails to the recipients' inboxes, asking them to click on a link in the body of the message.

1.3 Detection of Phishing Techniques

Extracting data from the evaluated emails using a pre-defined set of criteria, a phishing detection system may classify an email as phishing or not. To categorize the data, a trained model uses feature vectors extracted from the data. Methods for detecting phishing attempts range from white-listing to blacklisting to content-based and URL-based approaches. Using whatever email addresses, they can get their hands on, and hackers will send these emails out. Your account has been compromised, and you must reply quickly by clicking on a link supplied in the email. Be sure your anti-spyware and firewall software is current and regularly updated to protect yourself against phishing scams.
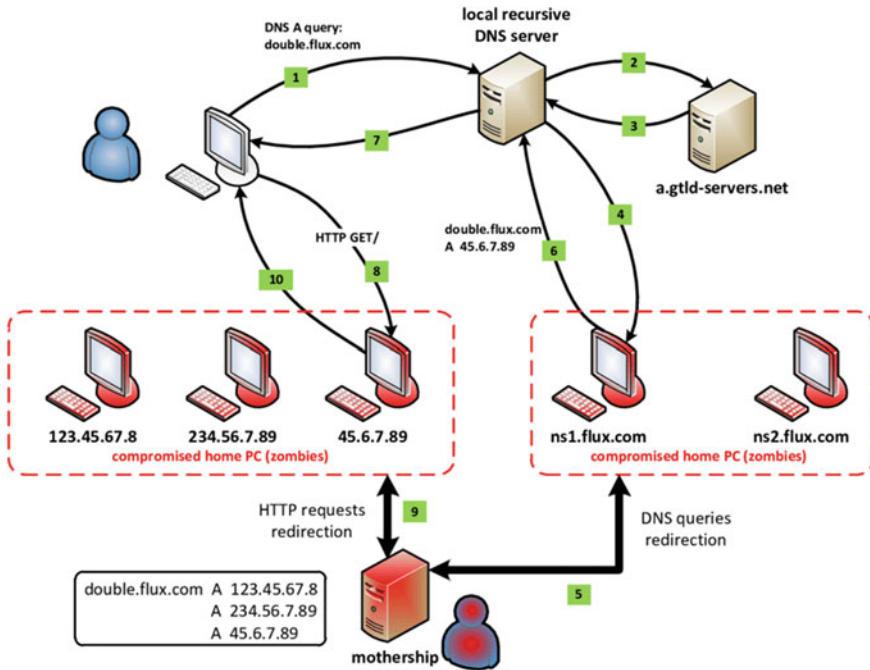
Fig. 2 Phishing Website hosting [1]

1.4 Machine Learning Used in Phishing

A machine learning algorithm uses markup visualization to identify phishing sites. Using machine learning models based on visual representations of Website code might improve the accuracy and speed with which phishing Websites can be identified. By applying machine learning models educated on visual representations of Website code, phishing Websites may be identified with greater precision and efficiency (Fig. 3).

1.5 Accurate Detection of Phishing

According to their experiments, the system has a 94% success rate in identifying phishing Websites. Due to its tiny neural network, it can operate on consumer devices and offer results in near-real time. They have tested the method using true phishing and legitimate sites, Stavros Shiaeles, co-author of the study, said to The Daily Swig. Using binary visualization and machine learning in cybersecurity is nothing new. ML and binary visualization were combined with Shiaeles' expertise in cybersecurity to create a new approach for malware detection that was published in 2019 and has
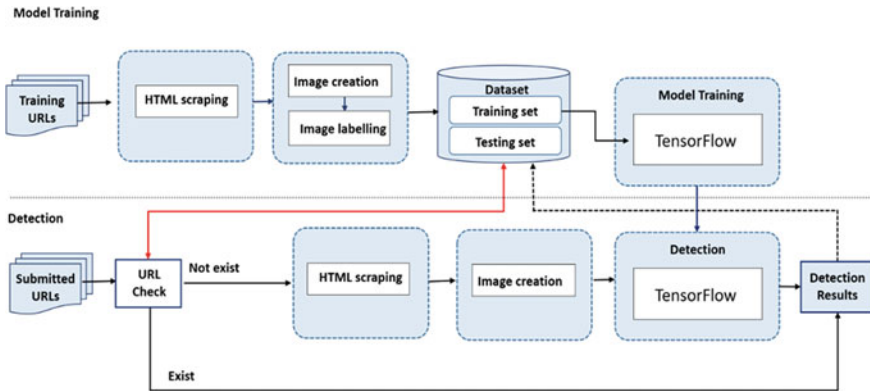
Fig. 3 Phishing interface

already shown promising results. Detection of phishing Websites is currently being completed by the team, who has completed testing of the system.

1.6 Classification Model Building

Our model's classification accuracy was evaluated against that of the J48, Naive Bayes, SVM, multi-layer perceptron, and random forest techniques after we extracted the set of features from the training dataset. Before we go into the classification findings for each method, let us take a quick look at how the algorithms work:

Naïve Bayes Classifier: Assuming all data characteristics are equally relevant and independent of one another, this classifier uses conditional probability and the Bayes rule to classify the data.

SVM: Algorithms for machine learning need our help. Even though it has a wide range of potential applications, vector machine is mostly employed for solving classification and regression issues.

Multi-layer Perceptron (MLP): MLPs are an example of a feed-forward ANN in which each neuron is treated as a processing unit after being triggered by an activation function.

Random Forest: In the training phase, it creates a collection of decision trees, each of which works on a preset number of randomly selected characteristics. This makes random forest an ideal classification approach for huge datasets.

1.7 Performance Metrics

We used a set of assessment measures for each algorithm to compare our proposed phishing email classification model to other classification approaches:

True Positive Rate (TP): From this information, the algorithm accurately detected the vast majority of phishing emails. In mathematical notation, $N P$ represents the total number of phishing emails in the dataset that were correctly labeled by the algorithm, as represented by P.

True Negative Rate (TN): The proportion of valid emails was accurately categorized as such by the algorithm. The total number of valid emails is equal to L minus the number of legitimate emails that were incorrectly identified by N_L as legitimate.

False Positive Rate (FP): Emails that were misclassified by the algorithm are included in this proportion. Since N_f is the total number of phishing-infected genuine emails, and L is how many were misclassified as phishing-infected by spam filters.

False Negative Rate (FN): Exactly how many malicious messages were falsely identified as safe by the system. The following formula may be derived if we assume that N pl is the total number of phishing emails in the dataset, and P is the number of phishing emails that the algorithm has assessed to be authentic.

Precision: Classifier precision is measured by how many emails classified as phishing are, in fact, phishing emails. This information is provided by:

Recall: The proportion of phishing emails that the classifier correctly identified as phishing is the metric used to determine the accuracy of the classifier's findings.

_F_-measure: Precision and recall are used to calculate the *F*-score, which is also known as the harmonic mean of these two metrics.

2 Literature Review

Ganguly et al. [2] offered a solid technique for offline development of the optimal SB classifier. Hierarchical clustering was utilized in conjunction with our patented labeling method for addressing the issue of disparate SB datasets in training data.

Rahul et al. [3] predicted high-risk companies early on using supervised machine learning models may be possible, according to our research. Unsupervised learning is also discussed in this section.

Peng et al. [4] introduced today that phishing is one of the most widespread and under-recognized problems in the world of online security. Natural language processing algorithms are used in this work to analyze text to identify phishing attacks.

SenthilMurugan et al. [5] were Twitter, the fast-evolving social network that millions of people use daily and grow hooked to. As a result, spam is playing a significant role in disengaging Twitter users and capturing their attention.

Hindy et al. [6] provide a model for monitoring SCADA-operated water systems for unexpected occurrences. The model was constructed and tested using six different machine learning methods.

Mani et al. [7] created that with the rise of e-commerce, the significance of customer evaluations has grown exponentially. Consumers have a true desire to learn what other people think about a product or service.

Martínez Torres et al. [8] reviewed machine learning techniques applied to cyber-security. Prediction, classification, data linkage, and data conceptualization are just a few of the many areas where machine learning methods may be used. The writers of this article provide a comprehensive overview of the uses of the machine.

Vinayakumar et al. [9] introduced preventing cyberattacks and creating safer computer networks and data systems. In the last ten years, machine learning has surpassed the traditional rule-based approach to tasks like natural language processing, image processing, and voice recognition.

Gaurav et al. [10] provided algorithms based on artificial intelligence for document labeling spam screening. Using a machine learning technique, we can automate the process of teaching a computer to classify data. There aren't many email spam datasets available; thus, the most likely problems are insufficient data and content written in an informal style, which prevents existing algorithms from performing well. Anticipations on the part of the classifier using the document labeling system, this research suggests a new approach to identifying spam emails use of forests (RF). This method is tested on three different datasets to see how well it performs. The experimental findings show that RF is more precise than other approaches.

Gangavarapu et al. [11] presented ML's potential in the realm of spam and phishing email filtering: a look back and some possible future directions. Our goal was to provide light on how to best extract features from email content and behavior, which features should be used to identify us, and how to choose the most discriminatory collection of features. The total accuracy of our suggested models for classifying UBEs was 99%.

Lakshmi et al. [12] introduced ADAM is a supervised deep learning classification and optimization method for detecting phishing in Web pages. In this research, we offer a unique approach to identifying phishing sites, which involves analyzing the HTML source code for potentially malicious connections. The proposed approach uses a 30-parameter feature vector to detect malicious domains.

Gupta et al. [13] reviewed phishing Website detection using machine learning. The easiest approach to getting private information from unsuspecting individuals was via a phishing attack. In an effort to steal personal or financial data, "phishers" create fake Websites that seem legitimate.

Ojewumi et al. [14] analyze and execute a rule-based method for phishing detection. Algorithms used in machine learning include KNN, RF, and SVM. The results showed that the random forest model performed the best out of the three algorithms tested.

Mohamed et al. [15] suggested a study centered on a three-stage phishing series assault. Uniform resource locators, Web traffic, and Web content based on phishing attack and non-attack features were the three input variables. Classification accuracy in phishing detection was calculated using three different classifiers, with results of 95.18% for NN, 85.45% for SVM, and 78.89% for RF. According to the findings, machine learning is the most effective method for identifying phishing attempts.

Mughaid et al. [16] proposed a detection model trained using ML methods, including test data used to verify the model's accuracy after training. They observed that the ones that made the most extensive use of attributes also performed the best in terms of precision and reliability. Accuracy for the top ML method boosted the decision tree and increased from 0.88 to 1.00 to 0.97 on the test datasets.

2.1 Limitation and Research Gap

There are several types of research in the area of machine learning and phishing emails but those researches considered traditional detection mechanisms. However, there are several types of research related to deep learning. But those research works failed to provide a real-life solution. Because of this, a new system that uses machine learning should be implemented to increase the security of emails.

2.2 Comparison of Features

See Table 1.

Table 1 Comparison of the feature of existing research

Citation	Machine learning	Phishing emails	Security	Classification
[17]	Yes	No	Yes	No
[18]	Yes	Yes	No	No
[19]	Yes	No	No	No
[20]	Yes	No	Yes	No
[21]	Yes	Yes	No	No
[22]	Yes	Yes	No	No
[23]	Yes	No	Yes	Yes
[24]	Yes	Yes	No	No
Proposed work	Yes	Yes	Yes	Yes

Literature review	Problem statement	Proposing novel approach	Evaluation
• Machine Learning • phishing email	• Accuracy • Performance • Reliability • Flexibility	• Hybridization of machine learning over phishing email • Advanced classifier	• Comparative analysis of accuracy and performance • Evaluating the reliability and flexibility

Fig. 4 Research methodology

3 Proposed Research Methodology

In the proposed research, machine learning has been used for detecting phishing emails. Advanced machine learning mechanism that should make use of detection techniques is capable to provide high-performance and security solutions (Fig. 4).

The proposed work is making using of compression technique to reduce the data content and apply security over contents. The proposed work is supposed to provide a more accurate and flexible solution for phishing emails.

4 Proposed Work

It has been observed that there have been several phishing email threats. There is a need to improve security by integrating a machine learning approach to classify phishing emails. The proposed work is focusing simulating the integration of compression and deep learning approach to detect and classify phishing emails to enhance the security of a system. Such a system is supposed to provide a more reliable and flexible approach by increasing classification accuracy. Several cyber assaults are identified and categorized in the proposed study using a machine learning model applied to a dataset including a variety of transaction types. The simulation took into account both filtered and unfiltered datasets. It is anticipated that the accuracy of detection and classification will be high once the dataset has been filtered using an optimizer (Fig. 5).

Multiple phishing email threats have been spotted, and it has been reported. There is a pressing need to tighten security by using a machine learning method to categorize phishing emails. The proposed study aims to improve system security by mimicking the combination of compression and a deep learning technique to identify and categorize phishing emails. The improved categorization accuracy promised by such a system would allow for a more trustworthy and adaptable method. In the proposed research, a machine learning model is used in a dataset consisting of different sorts of transactions to identify and classify several cyberattacks. Both filtered and unfiltered datasets were included throughout the simulation. Once an optimizer has been used
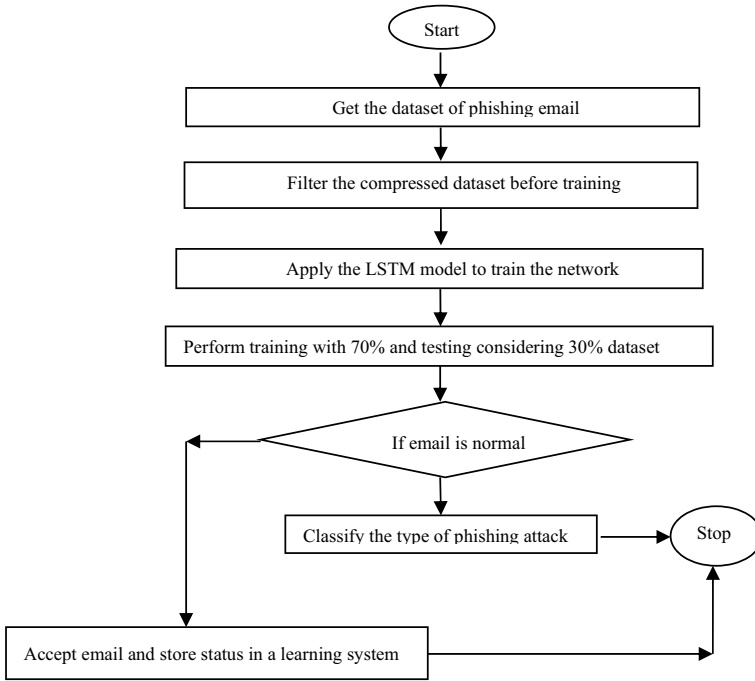
Fig. 5 Process flow of proposed work

to filter the dataset, it is expected that detection and classification accuracy would be high.

5 Result and Discussion

The proposed study simulates the combination of compression with a deep learning technique to identify and categorize phishing emails, to improve system security. The improved categorization accuracy promised by such a system would allow for a more trustworthy and adaptable method. In the proposed research, a machine learning model is used in a dataset that contains different kinds of transactions to identify and classify many phishing attacks. Both filtered and unfiltered datasets were included throughout the simulation. After an optimizer has been used to filter the dataset, it is expected that detection and classification accuracy would be high. A variety of phishing email schemes may be implemented using the suggested hybrid paradigm. For categorization purposes, LSTM-based ML has been used.

Table 2 Accuracy of confusion matrix of an unfiltered dataset

Class	n (truth)	n (classified)	Accuracy (%)	Precision	Recall	F1-score
1	1035	1000	91.18	0.84	0.81	0.83
2	971	1000	90.63	0.80	0.82	0.81
3	955	1000	91.18	0.80	0.84	0.82
4	1039	1000	89.63	0.81	0.78	0.80

Table 3 Accuracy of confusion matrix of filtered dataset

Class	n (truth)	n (classified)	Accuracy (%)	Precision	Recall	F1-score
1	1017	1000	95.53	0.92	0.90	0.91
2	985	1000	95.23	0.90	0.91	0.90
3	979	1000	95.58	0.90	0.92	0.91
4	1019	1000	94.78	0.91	0.89	0.90

5.1 Confusion Matrix of an Unfiltered Dataset

Table 2 has been processed using accuracy, precision, recall, and *F*1-score.

Results: TP: 3252 and overall accuracy: 81.3%.

5.2 Confusion Matrix of Filtered Dataset

Table 3 has been processed using accuracy, precision, recall, and *F*1-score.

Results: TP: 3622 and overall accuracy: 90.55%.

5.3 Comparison Analysis

Accuracy: The accuracy of completed work and recommended future work for each of the four classes has been shown in Table 4. Compared to the original, unfiltered data, the filtered version is more precise.

Taking into account the data in Table 4, we can now show the precision of the filtered dataset in comparison with the unfiltered one in Fig. 6.

Precision: Table 5 displays the results of comparing the accuracy of completed and suggested tasks in classes 1, 2, 3, and 4. The filtered dataset is found to be more precise than the unprocessed one.

Table 4 Comparison analysis of accuracy

Class	Unfiltered dataset (%)	Filtered dataset (%)
1	91.18	95.53
2	90.63	95.23
3	91.18	95.58
4	89.63	94.78

Fig. 6 Comparison analysis of accuracy
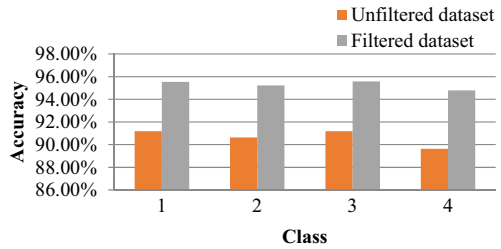
Table 5 Comparison analysis of precision

Class	Unfiltered dataset	Filtered dataset
1	0.84	0.92
2	0.80	0.90
3	0.80	0.90
4	0.81	0.91

Considering Table 5, Fig. 7 is drawn to visualize the precision of the filtered dataset to the unfiltered dataset.

Recall Value: Table 6 displays the recall values derived from past and projected work for classes 1, 2, 3, and 4. An increase in the recall value of the filtered dataset compared to the unfiltered one has been seen.

Considering Table 6, Fig. 8 is drawn to visualize the recall value of the filtered dataset to the unfiltered dataset.

Fig. 7 Comparison analysis of precision
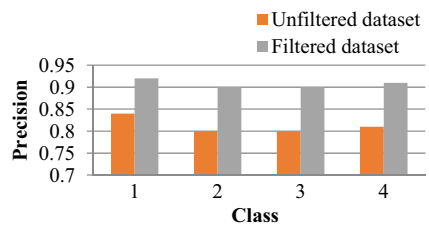
Table 6 Comparison analysis of recall value

Class	Unfiltered dataset	Filtered dataset
1	0.81	0.90
2	0.82	0.91
3	0.84	0.92
4	0.78	0.89

Fig. 8 Comparison analysis of recall value
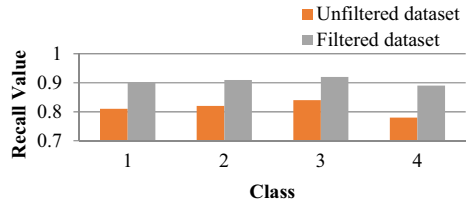
Table 7 Comparison analysis of F1-score

Class	Unfiltered dataset	Filtered dataset
1	0.83	0.91
2	0.81	0.90
3	0.82	0.91
4	0.80	0.90

F1-score: Table 7 displays the F1-scores for classes 1, 2, 3, and 4 based on the comparison of completed and planned tasks. The F1-score of the filtered dataset is higher than the unfiltered dataset.

Considering Table 7, Fig. 9 is drawn to visualize the F1-score of a filtered dataset to an unfiltered dataset.

Fig. 9 Comparison analyzes of F1-score
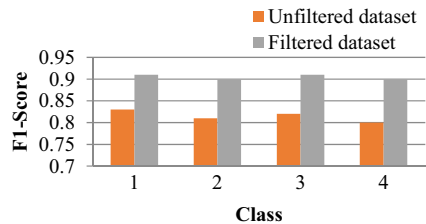
6 Conclusion

Phishing refers to fraudulent efforts to get access to personal information such as passwords or bank account numbers. Scams using phishing emails are rather widespread. Online payment systems, Webmail, banking institutions, file hosting services, and cloud storage services are just some of the many potential targets for phishing attacks. Among all sectors, Webmail and online payment were the most common targets of phishing attacks. Blacklisting, heuristics, visual resemblance, and machine learning are all examples of anti-phishing strategies. The blacklist method is the most popular since it is easy to implement and does not need constantly update a database with new phishing attempts. Phishing may now be easily identified with the use of machine learning.

1. Accuracy in the case of a filtered dataset is 94–96%, whereas it is 89–92% in the case of an unfiltered dataset.
2. Precision in the case of the filtered dataset is 0.9–0.92, whereas it is 0.8–0.84 in the case of an unfiltered dataset.
3. Recall in the case of a filtered dataset is 0.89–0.91, whereas it is 0.78–0.84 in the case of the unfiltered dataset.
4. $F1$-score in the case of a filtered dataset is 0.90–0.91, whereas it is 0.8–0.83 in the case of unfiltered dataset.

Furthermore, it resolves a significant issue with the current method. We conducted a comprehensive literature review and developed a novel approach to detecting phishing Websites using feature extraction and machine learning as part of our study.

7 Future Scope

Such types of research are supposed to play a significant role in the area of detection of phishing emails where data security is a must. Many real-time systems are capable to tackle the real-life problems. These systems might be extended by introducing machine learning mechanisms. The strategy places too much weight on the information presented on a Website, which is its main weakness. It is suggested that further research be conducted in this area. To acquire more precise findings, the article suggests additional data collection and the possible testing of other machine learning technologies.

References

1. https://www.researchgate.net/profile/Samuel-Marchal/publication/303458208/figure/fig5/AS:669487355404319@1536629632903/Phishing-web-site-hosting-using-a-double-flux-network.png
2. Ganguly S, Sadaoui S (2018) Online detection of shill bidding fraud based on machine learning techniques. LNAI, vol 10868. Springer
3. Rahul K, Seth N, Dinesh Kumar U (2018) Spotting earnings manipulation: using machine learning for financial fraud detection. LNAI, vol 11311. Springer
4. Peng T, Harris I, Sawa Y (2018) Detecting phishing attacks using natural language processing and machine learning. In: Proceedings—12th IEEE international conference on semantic computing (ICSC 2018), Jan 2018, pp 300–301. https://doi.org/10.1109/ICSC.2018.00056
5. SenthilMurugan N, Usha Devi G (2018) Detecting streaming of twitter spam using hybrid method. Wirel Pers Commun 103(2):1353–1374. https://doi.org/10.1007/s11277-018-5513-z
6. Hindy H, Brosset D, Bayne E, Seeam A, Bellekens X (2019) Improving SIEM for critical SCADA water infrastructures using machine learning. LNCS, vol 11387. Springer
7. Salihovic I, Serdarevic H, Kevin J (2019) The role of feature selection in machine learning for detection of spam and phishing attacks, vol 60. Springer
8. Martínez Torres J, Iglesias Comesaña C, García-Nieto PJ (2019) Review: machine learning techniques applied to cybersecurity. Int J Mach Learn Cybern 10(10):2823–2836. https://doi.org/10.1007/s13042-018-00906-1
9. Vinayakumar R, Soman KP, Poornachandran P, Akarsh S (2019) Cybersecurity and secure information systems. Springer
10. Gaurav D, Tiwari SM, Goyal A, Gandhi N, Abraham A (2020) Machine intelligence-based algorithms for spam filtering on document labeling. Soft Comput 24(13):9625–9638. https://doi.org/10.1007/s00500-019-04473-7
11. Gangavarapu T, Jaidhar CD, Chanduka B (2020) Applicability of machine learning in spam and phishing email filtering: review and approaches. Artif Intell Rev 53(7)
12. Lakshmi L, Reddy MP, Santhaiah C, Reddy UJ (2021) Smart phishing detection in web pages using supervised deep learning classification and optimization technique ADAM. Wirel Pers Commun 118(4):3549–3564. https://doi.org/10.1007/s11277-021-08196-7
13. Gupta P, Singh A (2021) Phishing website detection using machine learning. Lect Notes Netw Syst 154:183–192. https://doi.org/10.1007/978-981-15-8354-4_19
14. Ojewumi TO, Ogunleye GO, Oguntunde BO, Folorunsho O, Fashoto SG, Ogbu N (2022) Performance evaluation of machine learning tools for detection of phishing attacks on web pages. Sci Afr 16:e01165. ISSN: 2468-2276. https://doi.org/10.1016/j.sciaf.2022.e01165
15. Mohamed G, Visumathi J, Mandal M, Anand J, Elangovan M (2022) An effective and secure mechanism for phishing attacks using a machine learning approach. Processes 10:1356. https://doi.org/10.3390/pr10071356
16. Mughaid A, AlZu'bi S, Hanif A et al (2022) An intelligent cyber security phishing detection system using deep learning techniques. ClusterComput. https://doi.org/10.1007/s10586-022-03604-4
17. Jain AK, Gupta BB (2018) PHISH-SAFE: URL features-based phishing detection system using machine learning, vol 729. Springer, Singapore
18. Mani S, Kumari S, Jain A, Kumar P (2018) Spam review detection using ensemble machine learning. LNAI, vol 10935. Springer
19. Yuan F, Cao Y, Shang Y, Liu Y, Tan J, Fang B (2018) Insider threat detection with deep neural network. LNCS, vol 10860. Springer
20. Arif MH, Li J, Iqbal M, Liu K (2018) Sentiment analysis and spam detection in short informal text using learning classifier systems. Soft Comput 22(21):7281–7291. https://doi.org/10.1007/s00500-017-2729-x
21. Sartor RC, Noshay J, Springer NM, Briggs SP (2019) Identification of the expressome by machine learning on omics data. Proc Natl Acad Sci U S A 116(36):18119–18125. https://doi.org/10.1073/pnas.1813645116

22. Sarker IH, Kayes AS, Badsha S, Alqahtani H, Watters P, Ng A (2020) Cybersecurity data science: an overview from a machine learning perspective. J Big Data 7(1). https://doi.org/10.1186/s40537-020-00318-5
23. Stamp M (2018) A survey of machine learning algorithms and their application in information security. Springer
24. Kumar S, Faizan A, Viinikainen A, Hamalainen T (2018) MLSPD—a machine learning based spam and phishing detection. LNCS, vol 11280. Springer

Fog-Assisted Smart Healthcare Prediction System for Diabetics Patients

Subhranshu Sekhar Tripathy, Shashi Bhusan Panda, Abhilash Pati, Mamata Rath, Niva Tripathy, and Premananda Sahu

Abstract Accurately, forecasting human diseases continues to be a challenging issue in the search for better and more crucial studies. Diabetic multifunctional sickness is a potentially dangerous condition that afflicts people all around the entire globe. Among the important parts of the human body it affects are the heart, neuropathy, retinopathy, and nephropathy. A healthcare monitoring recommendation system accurately detects and also suggests diabetes complications using adaptive machine learning methods and data fusion techniques on medical management datasets. Recently, various machine learning approaches and algorithms for predicting diabetic illness have been created. Even so, a lot of systems are unable to handle the massive volume of multi-feature raw data on the diabetes mellitus condition. An intelligent diabetic health management strategy is created using deep

S. Sekhar Tripathy (✉)
School of Computer Engineering, KIIT (Deemed to be University), Bhubaneswar, Odisha, India
e-mail: subhranshu.008@gmail.com

S. Bhusan Panda
Department of Electronics and Telecommunication Engineering, Raajdhani Engineering College, Bhubaneswar, Odisha, India
e-mail: sashismile@gmail.com

A. Pati
Department of Computer Science and Engineering, Siksha 'O' Anusandhan University, Bhubaneswar, Odisha, India
e-mail: er.abhilash.pati@gmail.com

M. Rath · N. Tripathy
Department of Computer Science and Engineering, DRIEMS Autonomous College, Cuttack, Odisha, India
e-mail: mamata.rath@driems.ac.in

N. Tripathy
e-mail: nivatripathy@driems.ac.in

P. Sahu
Department of Computer Science and Engineering, SRMIST Delhi-NCR Campus, Ghaziabad, UP, India
e-mail: prema.uce@gmail.com

machine learning and data fusion technologies. By combining the data, we can reduce the unnecessary burden on organization computing power and increase the effectiveness of the suggested method, allowing us to predict and advise this life-threatening condition more accurately. Finally, the adaptive prediction system is assessed, and a machine learning model for diabetes diagnosis that is effective and up-to-date is trained. The suggested approach attained the maximum level of accuracy against standard advanced classification algorithms. The strategy we provide is therefore more effective in predicting and advising multimodal diabetic illness. Our proposed system's improved disease diagnosis performance warrants its inclusion in automated diabetes patient diagnostic and recommendation systems.

Keywords IoT · Machine learning · Cloud computing · Fog computing · Smart healthcare diabetes mellitus

1 Introduction

Recent advancements in modern biotechnology, as well as high-performance computing, are focusing on making e-healthcare survey methods and disease diagnosis more efficient and inexpensive. The accuracy of model development from e-healthcare large data is essential for efficiency and reliability. Diabetes is just one of several life-threatening illnesses [1, 2]. Diabetes affects around 422 million people globally, and the majority of whom live in low- and middle-income countries and diabetes is responsible for 1.5 million fatalities every year [3, 4].

Diabetes is a transdisciplinary illness that tends to cause abdominal organs in the human body, including the kidney, eyes, lungs, and heart. The sickness was diagnosed either manually by a healthcare professional or automatically by any automated equipment. Each of these diabetes illness assessment methods includes some upsides and downsides. Due to various concealed harmful impacts on the body, any qualified healthcare specialist cannot manually diagnose diabetes disease early. With the use of an intelligent prediction model and the employment of deep learning and AI-based technologies, this disease can be predicted at an initial stage with precision [5–7].

Recent findings result show certain healthcare advanced technologies for diagnosing and treating human ailments. Myocardial ischemia is a life-threatening condition that affects the heart's blood circulation. Deep CNN is used in this analysis to monitor and prevent heart attacks in humans. To massively reduce the remarkable calculation, a computer-aided diagnostic (CAD) system employs the transfer learning technique to provide accurate and rapid responses. In this analysis, a CAD system is designed to detect and prevent heart attacks in a precise and reliable manner. Blockchain is employed for the confidentiality of information in this study. In this study, data is captured through every stage using IoHT and then reinvented using blockchain to ensure seamless and consistent healthcare information for improved system consistency. In the current healthcare survey, many supervised

learning approaches have been suggested to boost performance [8, 9]. Hepatoma is a malignant malignancy illness in humans. In this paper, an intelligent prediction method for this dangerous and deadly condition is constructed using deep learning and healthcare data review with a stack learning method. Stack learning is a type of supervised method. The healthcare data on hepatocellular carcinoma disease is also observed using adaptive computational methods in this article. For superior outcomes, researchers employed a machine learning strategy for higher combinations in multistage edge cloud computing supported framework, 5G protected federated learning, and enhanced "coronavirus" illness diagnosis. An advanced approach with a machine learning algorithm is being used to diagnose COVID-19 disease, and it was confirmed to be completely accurate [10, 11]. The significance of an intelligent healthcare management system is developing more and more to give quality and more responsive projections. We need a suitable technique for detecting and successfully going to recommend life-threatening disorders like diabetes to mitigate the threat of existing common diseases. EHRs are critical components of smart health monitoring content-based filtering for anticipating life-threatening diseases, particularly trans-disciplinary with fatal diabetes mellitus disorders. The previously used data obtained from sensing elements and computerized recorded health textual data, on the other hand, is unstructured. Exporting important aspects and integrating them in a structured form is likewise a time-consuming and skill-intensive operation [12]. As a result, those segments are divided into two parts: a diabetes prediction system based on wearable sensors in addition to information retrieval from computerized recorded health textual data. Next, for improved outcomes and more accurate diabetes illness prediction with DML, data fusion is required. This analysis contributes considerably by strengthening the healthcare dataset for the most reliable estimation of multi-disciplinary diabetes conditions. Each patient's data was collected using wearable sensors and computerized recorded health textual data in the form of a textual record. The required information from bilateral sides is combined to improve the healthcare database after collecting patient records. In larger healthcare datasets, the integrated machine learning method performs more precisely as well as provides more accurate results [13, 14]. Finally, by gathering patient information and applying an integrated deep learning method to forecast that is correct and accurate with a suggestion based on multifunctional diabetes mellitus illness victims, we have produced a superior recommendation system. Figure 1 shows the fog computing attached between cloud computing and IoT.

The key contributions of this work can be listed as follows:

- It explores the latest recent innovations in the diagnosis and treatment of diabetes illness.
- For such treatment of diabetes, a smart healthcare recommendation system based on deep machine learning and data fusion viewpoints is described.
- Using a well-known diabetes dataset, the ensemble machine learning model is trained to forecast diabetes; its effectiveness is then examined to the most current advances in the literature.
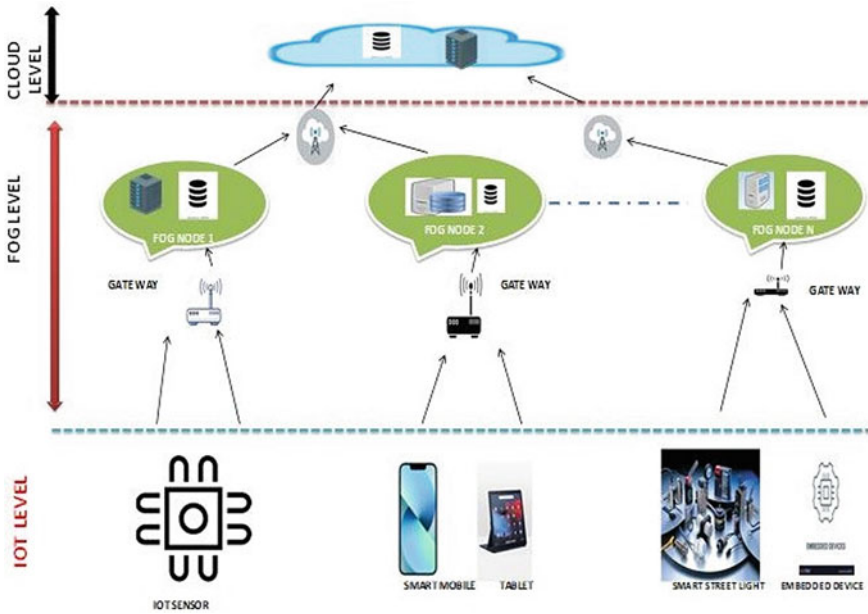
Fig. 1 Fog computing in cloud computing and IoT

The following is just how the paper proceeds: Sect. 2 discusses the most current advances in diabetic disease detection and literature recommendations, Sect. 3 covers the proposed method, Sect. 4 covers database choice, data preprocessing, data amalgamation, with discussion, Sect. 5 covers concluding remarks and subsequent scope with Segment 6 cover references.

2 Related Work

The diagnosis of diabetes syndrome was made possible, thanks to the efforts of several researchers. To forecast diabetic mellitus disorder, they implemented different predictive analytic methods with neural network techniques. We can simply acquire healthcare data through neural network techniques. It is possible readily forecast human diseases, including complex diabetes problems, after collecting massive data from the healthcare center. In [15], to diagnose diabetes disease early stage, the K-nearest neighbor (KNN) classifier model was used with the results compared to those of the support vector classifier, which generated 81.8% precision. The K-nearest neighbor algorithm is being applied to detect diabetic illness in this study, and the findings were contrasted with one more machine learning algorithm called support vector machine to validate the proposed model. As well as scientists employed a convolution neural network (CNN) to treat diabetes mellitus second type disorder [16] and assimilate

the findings of the time series forecast indicator and class of feed-forward artificial neural network. The semantic network is employed to detect diabetes mellitus second type condition in this study. Two types of machine learning algorithms were employed to compare the performance. In the area under the curve (AUC), 77.5% efficiency was accomplished. The SVM classifier was used to analyze earlier diagnoses of diabetes mellitus disorder using characteristics selection techniques [17], and the findings obtained were contrasted to random forest, Naive Bayes, decision tree, and K-nearest neighbor categorization models. SVM was obtained with the greatest precision of 77.73%. The feature selection technique was used in this study. We can minimize the processing capabilities of the system and improve accuracy by using the feature selection technique. For comparability and integrity of findings, many machine learning designs were used. Computational technologies are applied with quick and painless procedures for diabetic illness detection. The accuracy obtained with this method was 91.67%. The deep (DNN) approach was used to detect diabetic retinopathy [18]. CNN obtained a 74.4% overall accuracy. Artificial intelligence was used to detect multiclass retinal illness [19]. The CNN classifier was utilized, and it had a 92% prediction performance. With machine learning, a data-driven strategy is employed to forecast diabetes and cardiovascular illnesses [20]. The extreme gradient boost was used in this study, and it was compared to the logistic regression, support vector machine, random forest, and weighted average integrated method. In the field beneath, the relative operating characteristic curve, 95.7% accuracy was achieved. An ensemble classification algorithm was used to predict type 2 diabetes illness [21].

The AUC precision obtained by such a paradigm was 84.8%. A new method for detecting diabetes using a smartphone was presented [22]. Image data was explored in this work for diagnoses and future directions. Patients' blood glucose levels were measured using a microcontroller-based agent [23]. A diabetic patient's blood glucose was observed using a sensor-integrated medication [24] for diabetes conditions. A smart app for consciousness [25] was utilized and developed on recorded health information such as patients' daily physical activity and other essential diabetic indicators. For anticipating interdisciplinary disease, retrieving necessary details from transmission device information as well as the victim's electronic health data is equally difficult. To address this issue, various algorithms have been proposed for retrieving the most essential data from healthcare text documents [26] to generate a dataset for diabetes illness detection. As required by the smart healthcare proposed system, the text-based dataset gathered from EHRs was standardized and transformed into a useful pattern. A healthcare wireless sensor reading was also obtained using wireless devices. Data was standardized to eliminate unwanted wireless data after it was collected using a wireless communication device, resulting in a complete and consistent healthcare dataset. In this way, machine learning methods for forecasting multimodal diabetic disease may be easily applied.

3   Proposed Model

This segment explores the general framework of an intelligent healthcare management prediction model for individuals with diverse diabetic conditions. For an accurate interpretation of each layer's functionality, the suggested technique is split into multiple layers. Finally, the integrated decentralized machine learning organization is defined, which is then used during this proposed model for forecasting along with suggested diabetes mellitus disorder status. This must include a clear as well as a precise depiction of the analytical outcomes, as well as an interpretation and resolution to the experiments. Figure 1 depicts the suggested model's architecture, which lays out the basic framework for a multidisciplinary diabetes illness prediction system in a patient's smart healthcare setting.

First, the suggested model structure is explained. A proposed system's structure is divided into phases, specifically to allow for a precise examination of the behavior at each level. The integrated deep learning framework is then applied throughout this proposed model for predicting and advising diabetic mellitus disorder in the patient, and it is finally described. This proposed prediction model is shown in Fig. 1. The learning section and evaluation section are the two main sections, which are separated from each other. A proper multidisciplinary diabetic disease forecast requires other components. Seven layers make up the training phase: (I) tactile level, (II) electronic health record layer, (III) crude component level-I, (IV) crude component level-II, (V) merged unprocessed attribute level, (VI) preparation layer, and (vii) operation level. All of these levels are connected by a fog layer. The intake features in the tactile layer include lifetime, genealogy, carbohydrate, thick-skinned, hypertension, steroid, and queen let index. Through the Internet of things, the database, primitive layer-1, receives the input parameters from the tactile layer. The usage of wireless connectivity may cause data gathered and recorded from numerous feature nodes to be inconsistent. We called this category of data "feature raw data" as a result. The EHR layer is made up of lab results, inquiries, evaluations, and patient clinical data. The EHRs layer collects data in the form of documents, which must be converted into the appropriate format before further analysis. To give enhanced healthcare management data for multidisciplinary diabetes diseases, electronic health records are combined with the previously described sensitive information using the data integration approach. Such combined feature raw data is then saved at the level of combined features, where they will undergo additional processing to predict diabetic disease. The model's feasibility depends on the following preprocessing layer. The shortcomings discovered before through the tactile layer via the Wi-Fi network and the EHRs layer are included in this step-by-step process (Fig. 2).

Algorithm

Step 1: Input: Dataset D = (x1, y1), (x2, y2), …, (xm, ym);
Step 2: First-level learning algorithms L1, …, Lt
Step 3: Second-level learning algorithm L
Step 4: For t = 1, …, T;
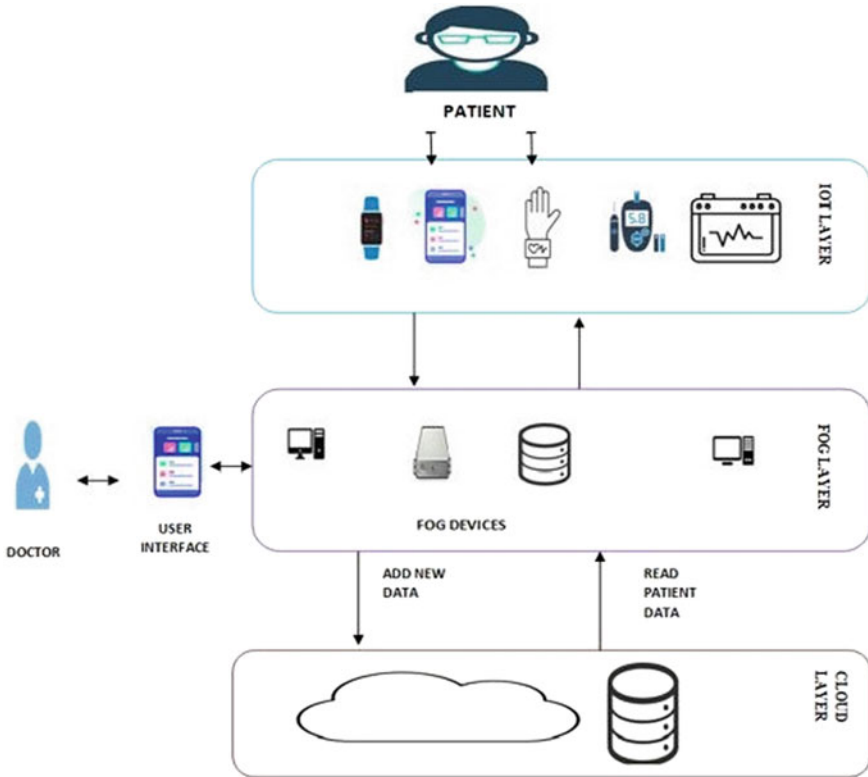
Fig. 2 Proposed model architecture

Step 5: Ht = Lt (D)
Step 6: End for
Step 7: D' = 0
Step 8: For i = 1, …, m;
Step 9: For t = 1, …, T;
Step 10: Zit = ht (Xi)
Step 11: End for
Step 12: D' = D'u ((Zi1, Zi2, …, Zit), yi)
Step 13: End for
Step 14: H' = L (D')
Step 15: Output: Binary (0/1)

Missing data is controlled using moving averages and standardization procedures to reduce noisy outcomes. The fused feature collection is then delivered to the application layer after filtering. The prediction layer is separated into two sub-layers: (i) the projection layer and (ii) the execution layer. An aggregate DML model is

used in the projection layer to determine multivariate diabetic syndrome. To optimize the classification performance of any predictive classification issue, ensemble deep learning integrates numerous independent models. The ensemble ML model's convergence is accomplished in three ways: (1) peak polling, (2) aggregating, and (3) balanced median for categorization. In contrast to various machine learning algorithms, the integrated machine learning proposed approach implemented an advanced encouraging mechanism for regularizing, reducing training error, and obtaining greater precision. In contrast to various machine learning architectures, this integrated machine learning algorithm's response rate is considerably faster. The training dataset is divided into several weak classifiers in this integrated machine learning model enhancement strategy, and this means that the fault rate of any one-week trainee is upgraded with the next weak trainee. As a result, the finishing strong trainee was discovered to possess the lowest prediction average fault rate. When compared to other ML algorithms, the ensemble ML model used an improved boosting mechanism for regularizing, reducing training error, and obtaining greater precision. In contrast with various machine learning algorithms, the integrated machine learning response rate is considerably faster. The training dataset is divided into several weak
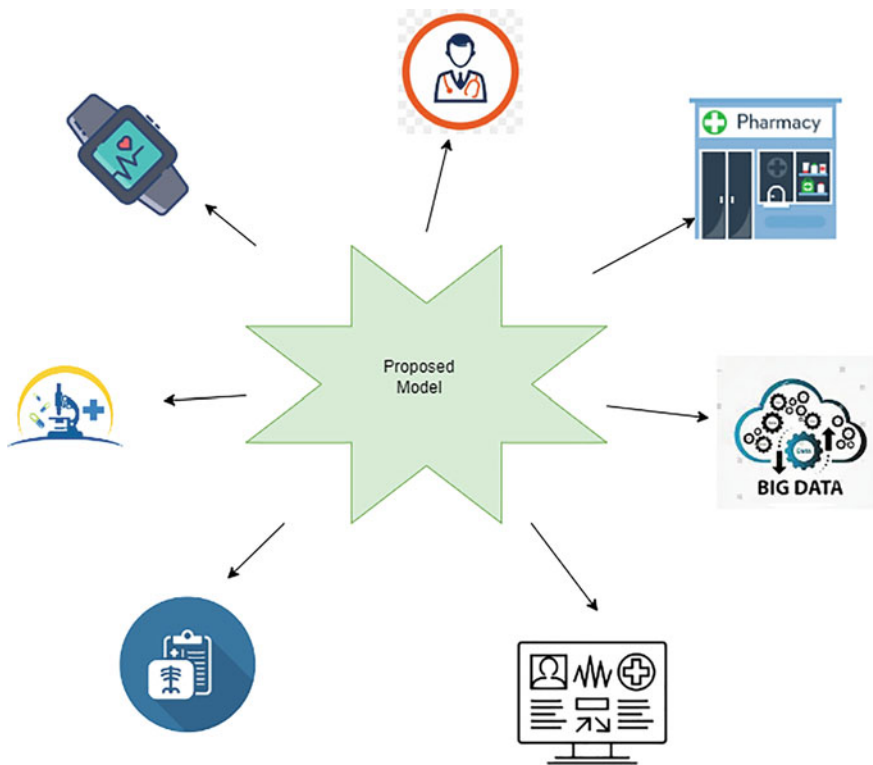
Fig. 3 Flowchart of the proposed model

classifiers in this integrated machine learning enhancement strategy, and this means that the fault rate of any one-week trainee is upgraded with the next weak trainee. Following that, the execution layer received the results. This correctness, exactness, recollect, average root means square deviation, and mean average fault rate attained through this SHRS-M3DP model are examined under this layer, and the execution layer performance is evaluated based on the accuracy, precision, recall, F1, root mean square error (RMSE), and mean average error (MAE) achieved by this intelligent health management prediction model. After reviewing the outcomes, "true" signifies that our suggested intelligent health management prediction model successfully detected diabetic disease, while "false" signifies that the suggested intelligent health management prediction model. The projection layer will be adjusted till the learning criteria requirements are fulfilled. Following the successful training of the suggested intelligent health management prediction model. This learned fused architecture is uploaded to the fog to again be integrated and projected diabetic illness. Figure 3 represents the flow of communication of the proposed work.

4   Results and Analysis

The type 2 diabetes dataset with nine important variables and 768 records has been utilized to run the suggested hybrid model [27]. Thirty percent of the data is used for testing, while seventy percent is for training. The experiment was run to demonstrate how well the suggested hybrid model would detect diabetics. The information used is gathered through sensors and then transferred to a database. Similarly, unstructured data from patients' lab results, observations, queries, and medical histories were turned into structured data for additional preprocessing. The processing module then examined the resultant fused, combined feature dataset in preparation for further processing.

    This part contains the findings of the aforementioned proposed model as well as a comparison with many other algorithms. When compared to all of the basic classifications, the recommended ensemble deep machine learning model surpassed them all [28, 29]. The accuracy of the suggested integrated intelligent healthcare management system in data file one for forecasting diabetes mellitus disorder by 779 instances is 72.7%. However, because of the tiny dataset, it is low. Ensemble ML outperformed all other classifiers in dataset 2 and reached 91% accuracy with 2000 examples, outperforming all others. The accuracy of diabetes mellitus disorder prediction with this suggested integrated intelligent healthcare management system is 95.6% in the concluded data integrated data file, with a small error rate. The suggested architecture by employing various assessment indicators compared with various DML classifiers. Only the Pima Indians diabetes dataset is used in this experiment, and no feature selection approach is used. Due to the tiny dataset and the unavailability of a feature set, the suggested architecture execution level is lower than various machine learning algorithms. When conducting the performance analysis of different data processing delays of the proposed method with other methods in [15, 16], we found

in Fig. 3 that the proposed approach exhibits a low delay in comparison with others. Figure 4 depicts the performance analysis based on processing delay.

Figure 5 presents a comparison of the accuracy percentage of prediction using different methods proposed by eminent researchers. It can be seen that in the proposed approach, the accuracy percentage is between [15, 16].

The proposed model evaluates various DML classifiers using a variety of assessment criteria. Because of the fused technique needed to interface a rich healthcare dataset, the suggested model's overall performance is much superior to the previous classifiers. The data fusion technique made our proposed model operate incredibly effectively and produce amazing outcomes. The suggested model outperformed existing DML classifiers in this trial by achieving 99.64% accuracy. In this experiment, the RMSE is 0%, and the MAE is just 0.6%. Other classifiers produce results that are just 85% accurate. This data combining method allowed our proposed framework to function effectively. Our suggested ensemble DML model has enhanced accuracy compared to the previous recent study. With the help of our extensive healthcare management dataset, we were able to forecast the multimodal diabetes state with greater accuracy than previous research, which had a 99.6 accuracy rate.
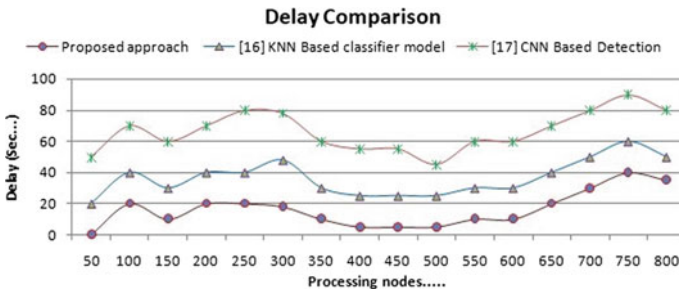
Fig. 4 Delay comparison between different approaches
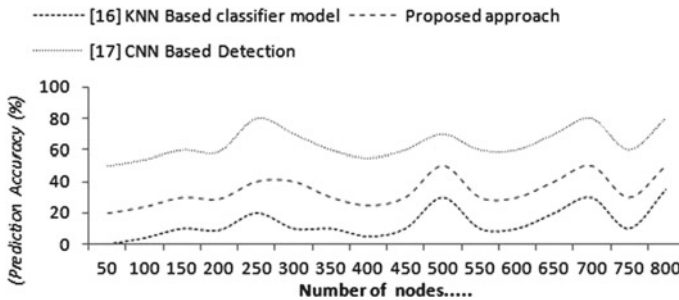
Fig. 5 Comparison of prediction accuracy

5 Conclusion and Future Plan

It is very difficult in detecting human diseases, particularly multimodal diabetes, and is difficult to provide effective and timelier treatment. In multimodal diabetes, there will be chance of affecting key human organs, as it considered as is very dangerous across the globe. In this research article, the author's model integrated an intelligent machine learning algorithm and data combination method adopted for swift reaction, a higher precision rate in this suggested architecture to predict with promoting interdisciplinary diabetic disease in patients quickly and easily. With the help of this suggested model, it accurately forecasts as well as regardless of whether or not this individual was dealing with multimodal diabetes disease. The suggested algorithm will detect these effects of living object components such as mono neuropathy, diabetic eye disease, and coronary thrombosis. Python programming language is used to design the suggested model's simulation. Finally, the findings of this investigation revealed that the suggested. The proposed algorithm's achievement is 95.6%, be excellent. In comparison with prior issued methodologies, this is exceptional.

References

1. Raman V, Then P, Sumari P (2016) Proposed retinal abnormality detection and classification approach: computer aided detection for diabetic retinopathy by machine learning approaches. In: 2016 8th IEEE ınternational conference on communication software and networks (ICCSN). IEEE, pp 636–641
2. Pati A, Parhi M, Pattanayak BK (2021) IDMS: an integrated decision making system for heart disease prediction. In: 2021 1st Odisha ınternational conference on electrical power engineering, communication and computing technology (ODICON). IEEE, pp 1–6
3. Tripathy SS, Mishra K, Barik RK, Roy DS (2022) A novel task offloading and resource allocation scheme for mist-assisted Cloud computing environment. In: Intelligent systems. Springer, Singapore, pp 103–111
4. Hammad M, Alkinani MH, Gupta BB, El-Latif A, Ahmed A (2022) Myocardial infarction detection based on deep neural network on imbalanced data. Multimedia Syst 28(4):1373–1385
5. Pati A, Parhi M, Pattanayak BK (2022) IADP: an ıntegrated approach for diabetes prediction using classification techniques. In: Advances in distributed computing and machine learning. Springer, Singapore, pp 287–298
6. Alghamdi M, Al-Mallah M, Keteyian S, Brawner C, Ehrman J, Sakr S (2017) Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: the Henry Ford ExercIse Testing (FIT) project. PLoS ONE 12(7):e0179805
7. Pati A, Parhi M, Pattanayak BK (2021) COVID-19 pandemic analysis and prediction using machine learning approaches in India. In: Advances in intelligent computing and communication. Springer, Singapore, pp 307–316
8. Tripathy SS, Barik RK, Roy DS (2022) Secure-M2FBalancer: a secure mist to fog computing-based distributed load balancing framework for smart city application. In: Advances in communication, devices and networking. Springer, Singapore, pp 277–285
9. Liu Y, Peng J, Kang J, Iliyasu AM, Niyato D, Abd El-Latif AA (2020) A secure federated learning framework for 5G networks. IEEE Wirel Commun 27(4):24–31
10. Sedik A, Hammad M, El-Samie A, Fathi E, Gupta BB, El-Latif A, Ahmed A (2022) Efficient deep learning approach for augmented detection of Coronavirus disease. Neural Comput Appl 34(14):11423–11440

11. Pati A, Parhi M, Pattanayak BK (2022) HeartFog: fog computing enabled ensemble deep learning framework for automatic heart disease diagnosis. In: Intelligent and cloud computing. Springer, Singapore, pp 39–53

12. Roul A, Pati A, Parhi M (2022) COVIHunt: an intelligent CNN-based COVID-19 detection using CXR imaging. In: Electronic systems and intelligent computing. Springer, Singapore, pp 313–327

13. Tripathy SS, Roy DS, Barik RK (2021) M2FBalancer: a mist-assisted fog computing-based load balancing strategy for smart cities. J Ambient Intell Smart Environ 13(3):219–233

14. Parhi M, Roul A, Ghosh B, Pati A (2022) IOATS: an intelligent online attendance tracking system based on facial recognition and edge computing. Int J Intell Syst Appl Eng 10(2):252–259

15. Mohebbi A, Aradottir TB, Johansen AR, Bengtsson H, Fraccaro M, Mørup M (2017) A deep learning approach to adherence detection for type 2 diabetics. In: 2017 39th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp 2896–2899

16. Sneha N, Gangil T (2019) Analysis of diabetes mellitus for early prediction using optimal features selection. J Big Data 6(1):1–19

17. Sah P, Sarma KK (2018) Bloodless technique to detect diabetes using soft computational tool. In: Ophthalmology: breakthroughs in research and practice. IGI Global, pp 34–52

18. Karthikeyan S, Sanjay KP, Madhusudan RJ, Sundaramoorthy SK, Namboori PK (2019) Detection of multi-class retinal diseases using artificial intelligence: an expeditious learning using deep CNN with minimal data. Biomed Pharmacol J 12(3):1577

19. Dinh A, Miertschin S, Young A, Mohanty SD (2019) A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Med Inform Decis Mak 19(1):1–15

20. Nguyen BP, Pham HN, Tran H, Nghiem N, Nguyen QH, Do TT, Tran CT, Simpson CR (2019) Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records. Comput Methods Programs Biomed 182:105055

21. Chen M, Yang J, Zhou J, Hao Y, Zhang J, Youn CH (2018) 5G-smart diabetes: toward personalized diabetes diagnosis with healthcare big data clouds. IEEE Commun Mag 56(4):16–23

22. Choudhary, P, De Portu S, Arrieta A, Castañeda J, Campbell FM (2019) Use of sensor-integrated pump therapy to reduce hypoglycaemia in people with Type 1 diabetes: a real-world study in the UK. Diabetic Med 36(9):1100–1108

23. Steinert A, Haesner M, Steinhagen-Thiessen E (2017) App-basiertes selb-stmonitoring bei Typ-2-diabetes. Zeitschrift für Gerontologie und Geriatrie 50(6):516–523

24. Davoodi R, Moradi MH (2018) Mortality prediction in intensive care units (ICUs) using a deep rule-based fuzzy classifier. J Biomed Inform 79:48–59

25. Fazlic LB, Hallawa A, Schmeink A, Peine A, Martin L, Dartmann G (2019) A novel NLP-fuzzy system prototype for information extraction from medical guidelines. In: 2019 42nd international convention on information and communication technology, electronics and microelectronics (MIPRO). IEEE, pp 1025–1030

26. Bernal EA, Yang X, Li Q, Kumar J, Madhvanath S, Ramesh P, Bala R (2017) Deep temporal multimodal fusion for medical procedure monitoring using wearable sensors. IEEE Trans Multimedia 20(1):107–118

27. Diabetes DataSet—UCI-ML Repository. https://archive.ics.uci.edu/ml/datasets/diabetes. Accessed 4 Oct 2021

28. Nayak DSK, Mahapatra S, Swarnkar T (2018) Gene selection and enrichment for microarray data—a comparative network based approach. In: Progress in advanced computing and intelligent engineering. Springer, Singapore, pp 417–427

29. Sahu B, Panigrahi A, Mohanty S, Sobhan S (2020) A hybrid cancer classification based on SVM optimized by PSO and reverse firefly algorithm. Int J Control Autom 13(4):506–517

Analyzing Land Cover Changes in Ganges Delta (Sundarbans) Using Remote Sensing and Machine Learning

Monika Sharma and Mukesh Kumar

Abstract The goal of this study was to look at how the Sundarbans mangrove and LU/LC environments have changed over the past 32 years. High-resolution multispectral Landsat satellite images from 1989 to 2021(1989–2000, 2000–2011, 2011–2021) were processed and analyzed. The main goal is to get useful information from satellite data. The maximum likelihood classification (supervised classification) and NDVI indices were used to analyze land cover changes in the region. The images were classified into four classes: water bodies, mangrove/dense forests, mangrove/spare forests, and wetland/mudflats. The NDVI values showed a gradual decrease in healthy vegetation and increased water level. The supervised classification results showed significant changes in forest density with 9.3% of dense forest areas, an increase in spare forest areas, and water levels are observed from 1989 to 2021. The classification uncovered that forest areas were shrinking by 0.82% (19.85 km^2). Significant changes can be visualized in Bhangaduni Island, where nearly 24 km^2 of the area covered with dense mangroves is now dissolved in water. For validation, a random sampling approach and the google earth pro platform was used for verifying the ground reality against classified results. The user, overall, producer, and Kappa statistics were used to assess categorization accuracy.

Keywords Remote sensing · LULC environment · Supervised classification · NDVI · Accuracy assessment

M. Sharma (✉) · M. Kumar
University Institute of Engineering and Technology, Chandigarh, India
e-mail: monikasharma68920@gmail.com

M. Kumar
e-mail: Mukesh_rai9@pu.ac.in

1   Introduction

The Sundarbans, also known as the Ganges Delta, is an organically grown mangroves area in the Bay of Bengal formed by the juncture of three rivers (Ganga, Meghna, and the Brahmaputra). From the Hooghly River to the Baleshwar River, it stretches from the West Bengal province of India to the Khulna Division in Bangladesh. It is surrounded by several streams and canals and features closed and open mangrove forests, farmland, mudflats, and uninhabited landscape. UNESCO has listed the Sundarbans Public Park, the West Sundarbans, the South Sundarbans, and the Eastern Sundarbans Wildlife Sanctuary as World Heritage Sites [1].

The flexible woods have been diminishing for a long time because of overexploitation and deforestation by people to make agrarian land, shrimp cultivating, and regular cataclysms. China already lost a major portion of its mangrove forest [2]. Coastal ecosystems are critical for preserving the diversity of species, genetic variation among them, storing and recycling nutrients, maintaining a pollution-free environment, and protecting shorelines from erosion and storms, all of which are impacted by climate change [3]. A maritime environment aids in environmental guidelines and serves as a vital source of oxygen and carbon sink. To counteract the effects of tides, a mangrove ecosystem helps to conserve the most important biodiversity in coastal and intertidal areas [4]. Storm hits are normal in the Ganges Delta and will turn out to be more continuous and exceptional in this time of environmental change. Sundarbans region has faced a number of cyclones strikes in the last two decades such as Sidr, Rashmi, Aila, Bulbul, and Amphan, with Bijli, Roan, Title, and Fani just missing out [5]. The testimony of dregs from the stream structure frames a baffling organization of waterway channels that cover the woods accompanying little islands and mudflats. The Sundarbans, the world's largest mangrove swamps, act as a carbon sink, give protection from severe storm damage, and destruction, safeguard from cyclones, and atmospheric instabilities, and provide a solid source of income as well [6].

Mangrove is an ordinary regular scene with extraordinary natural, social, and financial capacities and values and ought to be profoundly esteemed and secured. Mangroves have been severely harmed in recent decades, and there are numerous flaws in mangrove conservation [7]. As of now, the worldwide mangroves are disseminated to 124 nations [8]. The area of mangroves is falling slowly, dwindling at a pace of 1–2% each year all around the world [9].

The LULC alludes to the arrangement or characterization of human exercises and natural elements on the scene inside a particular period because of setting up logical and measurable strategies for the examination of fitting source materials. It has various methods of classification. There are many different sorts of land cover elements, such as metropolitan, agrarian land, and forestry land [10].

The remote detecting and geographic data framework has ended up being vital in surveying also dissecting land cover changes over a region. Remote sensing can give brief data on land cover in a specific general setting, which has altered the investigation of land cover change. Remote sensing is a useful technique for keeping

track of things, assessing, discovering, describing, and assessing natural resources on the planet [11].

The target of this study is to look at how the land cover has changed over time. Mangrove degradation occurs in the subpart of Sundarbans, classified as one of the core mangrove areas. This investigation aims to identify the major forcing parameter affecting the ecosystem and the number of species living in this deltaic region. According to the examination it is accepted that the Sundarbans have absorbed a major portion of co^2 [12, 13]. According to many experts, mangrove forests increased by 1.4% from 1975 to 1990 but fell by roughly 2.5% from 1990 to 2000 [13]. Various experts have found that between 2000 and 2012, forest loss in South Asia outpaced the rate of regeneration [14, 15]. So, it is crucial to study the LU/LC and change detection of the region to know the exact status and major factors behind the situation. We intend to analyze the changes in various features such as a change in net forest area, change in water level, and change in the condition of mangroves throughout the years. We intend to analyze the changes in various features such as a change in net forest area, change in water level, and change in the health status of mangroves over the years.

2 Materials and Methods

The Ganges Delta (otherwise called the Sundarbans) is a waterway delta in the Bay of Bengal, comprising Bangladesh and the Indian province of West Bengal. My study area is a subpart of Sundarbans which covers a significant area of Indian Sundarbans and a small area of the Bangladesh region; it lies between 21°30' and 21°50' North latitude and 88°43'-89°60' East longitude as shown in Fig. 1.

The Sundarbans cover an area of about 10,000 square kilometers. The Indian Sundarbans own 40% of these lands, while Bangladesh owns the rest. This delta is home to a diverse range of wildlife. In the Bay of Bengal, the three major rivers Ganga, Meghna, and Brahmaputra converge to form the Sundarbans delta. Reserved forest covers a large portion of this study region, covering approximately 2061 square kilometers.

The Sundarbans are made up of open and closed mangrove forests, horticultural and oceanic ranch lands, mudflats, and barren land, all of which are connected by flowing streams and channels. Natural disasters struck this region frequently during the monsoon season. As a result, most individuals are affected by natural disasters during the monsoon season. Cyclone Sidr wreaked havoc on the Sundarbans in 2007, destroying 40% of the region. Due to increasing sea levels and dwindling freshwater supplies, the woodland is suffering from increased salinity. Cyclone Aila wreaked havoc on the vast coastal region of the Sundarbans in May 2009, affecting more than 100,000 people. Amphan 2020 a super typhoon that made landfall in the Sundarbans in May 2020 caused $14.1 billion in damage. Amphan 2020 is the strongest storm to make landfall in the Ganges Delta since Sidr in 2007 and the first super typhoon to form in the Bay of Bengal since Cyclone Odisha in 1999.
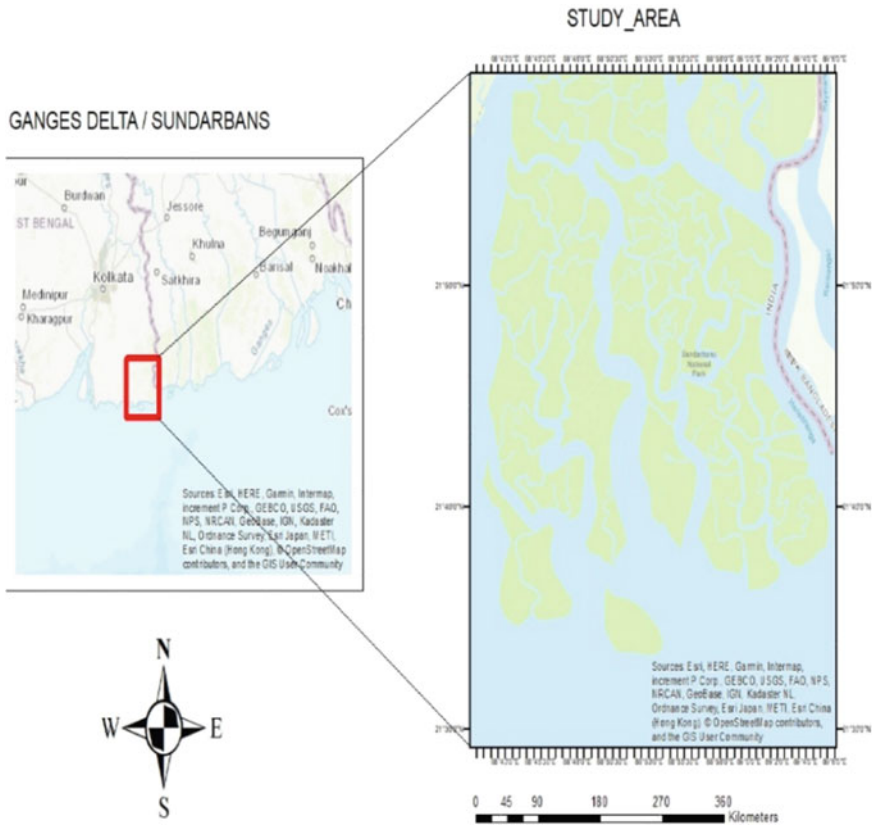
Fig. 1 Geographic locality of the study area

2.1 Data Source

The multispectral high-resolution Landsat data was sourced from the United States Geological Survey (USGS) website. Satellite data for the year 1989, 2000, 2011, and 2021 was used; data from the years 1989 and 2000 belongs to LANDSAT 5 TM, data from the year 2011 belongs to LANDSAT 7 ETM, and data from the year 2021 belongs to LANDSAT 8 OLI. The images are categorized as WGS_1984_UTM_ Zone projection coordinate system. The point-by-point portrayal is introduced in Table 1. Except for the 2021 images, all the others are clear and cloud-free.

Table 1 Satellite data used in this investigation

Year	Satellite data	Acquisition time	Spatial resolution	Cloud cover	UTM zone
1989	LANDSAT 5 TM	1989-03-16	30 m	0.0	45N
2000	LANDSAT 5 TM	2000-01-26	30 m	0.0	45N
2011	LANDSAT 7 ETM	2011-03-05	30 m	0.0	45N
2021	LANDSAT 8 OLI	2021-02-04	30 m	0.03	45N

2.2 Proposed Methodology

The purpose of this study is to explore and gather information from satellite images. The satellite data used in this study is multispectral Landsat data from various periods (1989, 2000, 2011, and 2021). In collecting desired data, the image preprocessing is the next proceeding step done using ArcMap 10.8 software. Figure 2 depicts the complete workflow or the strategy we followed to carry out this study.

2.3 Image Preprocessing

Geometric correction and radiometric calibration operations are used to reduce image distortion and radiometric inaccuracies. Geometric corrections are mathematical amendments that are made to address the irregularity between the area directions of the crude picture information, and the genuine area organized on the ground or base picture. Radiometric corrections are revisions that are made to address the errors in pixel values because of instrumentation, the wavelength reliance on sun-oriented radiation, the effect of the environment, and so forth. The RGB composite band was created using the composite band tool of the ArcMap system toolbox. For resampling, the nearest neighbor technique is utilized. The research area is defined and cropped using the manual digitization approach. As information had fundamentally less cloud cover, we did not have to play out any cloud masking. So, our clipped study area image is ready for classification.

2.4 Image Classification

For image classification, there are numerous approaches available such as pixel-level change detection, feature-level change detection, and object-class-based change detection techniques [16]. In this study, we have used two techniques, pixel-level change detection which is NDVI indices, and maximum likelihood supervised classification which is object-based classification.
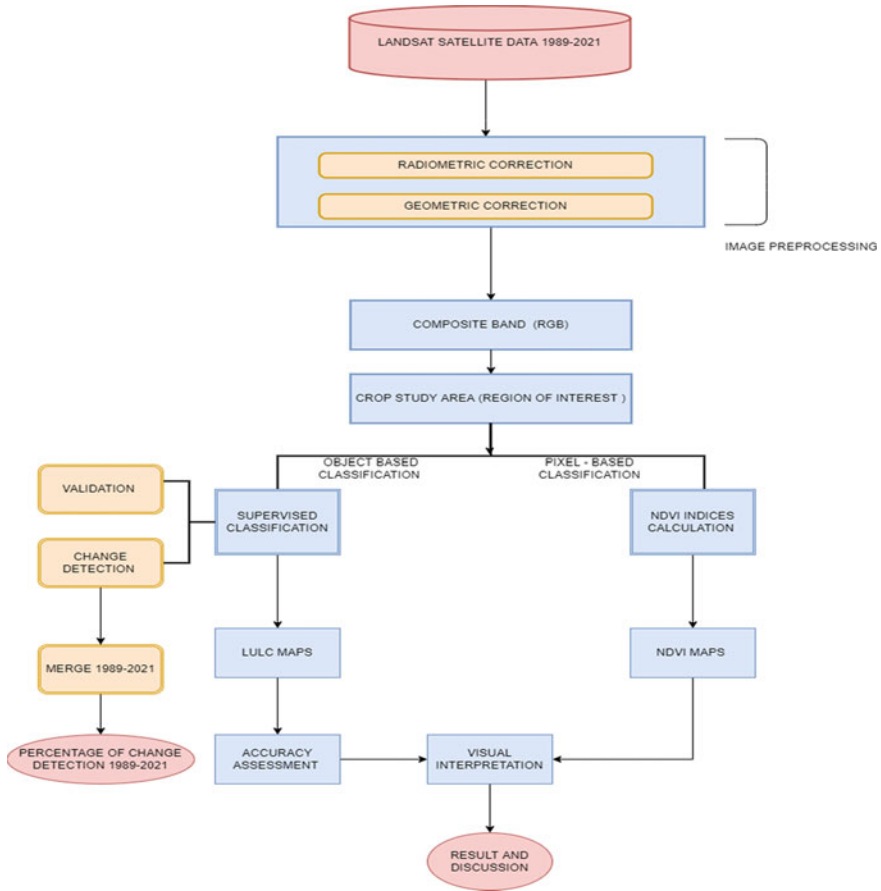
Fig. 2 Proposed methodology

2.4.1 NDVI Classification

This is a pixel-based vegetation mapping technique. Investigation of vegetation thickness and overhang design or fixes of greenness ashore and phantom vegetation ascribes are utilized. Standardized distinction vegetation record (NDVI) evaluates flora by estimating the contrast between close infrared and red light [17]. Vegetation lists permit us to depict the conveyance of vegetation and soil because of the trademark reflectance examples of green vegetation. The NDVI is a straightforward mathematical pointer that can be utilized to investigate the remote detecting estimations, from a far-off stage and survey whether the entity or thing being seen contains healthy organic vegetation or not.

NDVI calculated as:

Table 2 Various classes to categorize the objects

Classes	Satellite data
Water body	Area covered by water
Mangrove/dense forest	The area where forest cover is so dense > 50%
Mangrove/spare forest	Less dense forest area < 50%
Wetland/mudflats	Area covered by mudflats, flood plains, saltwater swamps, marshes, etc.

Table 3 Normalized difference vegetation index indices observation for 1989, 2000, 2011, and 2021

Year	Maximum	Minimum	Means	Standard deviation
1989	−0.181536	0.451597	0.038697	0.113719221
2000	−0.524659	0.517657	0.029767	0.129508304
2011	−0.159654	0.483587	−0.344467	0.135913289
2021	−0.137108	0.375781	0.1055248	0.231872071

$$\text{NDVI} = \text{Near Infrared (NIR)} - \text{RED}/\text{Near Infrared (NIR)} + \text{RED}$$

NDVI values lie between [-1, + 1] where extreme negative values are identified as water body, values near 0 are bare land and mudflats, and positive values show vegetation density [16].

NDVI Change Detection

In this work, Landsat imagery is used to create NDVI classification maps for 1989, 2000, 2011, and 2021, which are then classified depending on the reflectance value. Water body, mangrove dense forest, mangrove sparse forest, and wetland/mudflats are the four classes we designed to distinguish distinct items, as given in Table 2.

As per NDVI results given in Table 3, in the year 1989, the minimum which signifies a water body is −0.181536049 and the maximum is 0.451597303, which is dense vegetation. Similarly, −0.524659634 is the minimum and 0.517657459 is the maximum in 2000; −0.159654602 is the minimum and 0.483587682 is maximum in the year 2011; −0.137108698 is the minimum and 0.375781208 is maximum value that is observed in the year 2021.

In the present investigation, as given in Table 3, positive NDVI numbers were seen to diminish from 1989 to 2021, indicating that forest density has decreased over time. Furthermore, negative NDVI values grew within the same period, indicating that the number of water bodies increased over time (Fig. 3).
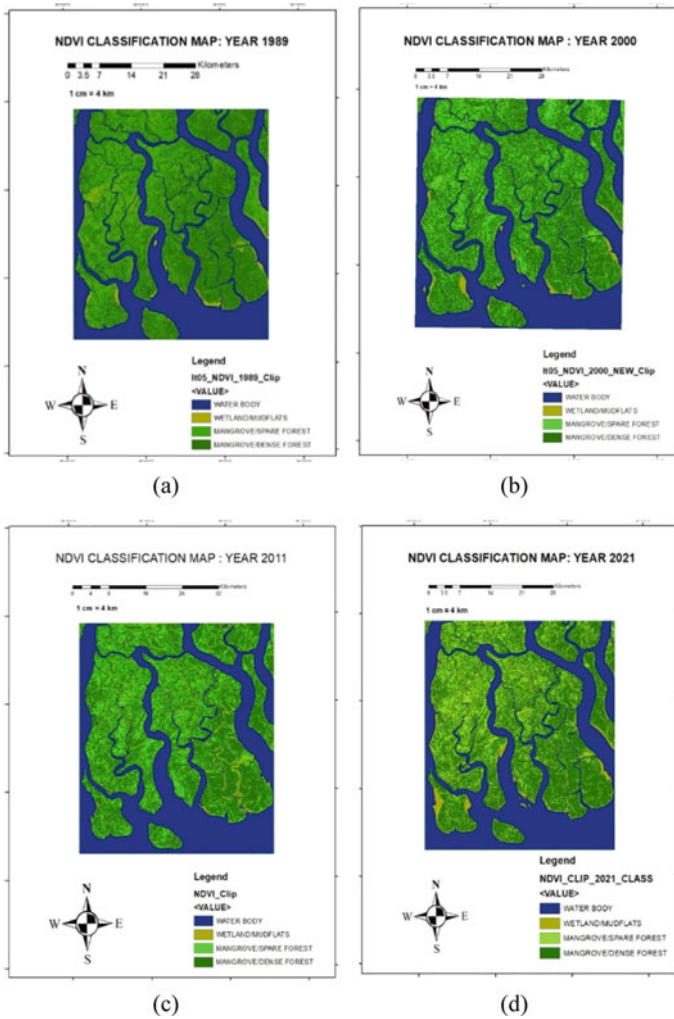
Fig. 3 NDVI classification of the Year **a** 1989, **b** 2000, **c** 2011, **d** 2021

2.4.2 Maximum Likelihood Classification (MLC)

Maximum likelihood classification is a regulated arrangement strategy that depends on the Bayes hypothesis. It is a class-based change recognition method to distinguish land cover changes. It utilizes a discriminant capacity to allocate pixels to the class with the most noteworthy probability. The class means vector and covariance network are the basic commitments and can be valued from the training sets of a class. Maximum likelihood (ML) is an administered arrangement technique got from the

Bayes hypothesis, which expresses that the deduced circulation, i.e., the likelihood that a pixel with an included feature vector belongs to a particular class.

Each picture element is either assigned to the class with the highest probability or left unclassified. The following are the general methods for most extreme probability grouping:

1. The number of land cover types included in the review is not set in stone.
2. Land cover data for the review zone is used to select the preparation pixels for each of the ideal classes.
3. The prepared pixels are then utilized to evaluate each class's mean vector and covariance network.
4. Finally, each point in the image is assigned to one of the ultimate land cover types or designated as obscure.

For image selection and categorization, we used ArcGIS 10.8 program. From the multispectral image, training samples were carefully collected. To identify various objects, we built four separate classes, as given in Table 2. Water body, mangrove dense forest, mangrove sparse forest, and wetland/mudflats are the four types. Information was acquired in the form of polygons during the preparation process. Due to sedimentation, the watercolor in the open ocean and coastal locations differs. As a result, we needed to gather data from numerous sources. Each class yielded an average of 100 polygons. The maximum likelihood classification (class-object-based change detection) technique is a promising classification algorithm that we applied with the help of ArcGIS 10.8 software. Two principles underpin this algorithm. One is that each class sample cell is spread in multidimensional space, and another is that decision making is based on the Bayes theorem.

Land Cover Change Detection

Change detection matrix illustrating the land cover changes in each period was generated from classified imagery of 1989 to 2000, 2000 to 2011, 2011 to 2021, and a change matrix was generated from 1989 to 2021 to measure the complete changes in total forest areas between 1989 and 2021. In evaluating these maps, Tables 4, 5, and 6 provides the total area that belongs to each class and the percentage of area change between two-time intervals, and Table 7 provides the change in the net forest area. It is noticed that the mangroves environment has been progressively diminishing at the pace of 19.25 km^2 each year and it got changed over into different features. Table 4 shows that in the year 1989, mangroves/dense forest and mangroves/spare forest area were 608.93 km^2 and 472.52 km^2, respectively; in the year 2000, mangroves/dense forest area decreased to 523.04 km^2 whereas mangroves/spare forest area increased to 635.47 km^2. In 2011, mangroves/spare forest area is 529.12 km^2 and the mangroves/dense forest area is 582.08 km^2 which shows the mixed status (loss and gain), which may be due to seasonal variations. Total/net area is decreasing, as seen in Table 7, with 2.41 percent net forest deterioration from 2000 to 2011, in continuity to that 0.47% forest area lost from 2011 to 2021. Seasonal variations could

Table 4 Area (sq. km) of LU/LC features over the years

Class	Year 1989	Year 2000	Year 2011	Year 2021
Water body	799	787.68	801.52	828.38
Mangrove/dense forest	608.93	523.04	582.08	417.31
Mangrove/spare forest	472.52	635.47	529.12	683.99
Wetland/mudflats	180.81	115.73	153.70	136.73

Table 5 Area (%) of LU/LC features over the years

Class	Year 1989 (%)	Year 2000 (%)	Year 2011 (%)	Year 2021 (%)
Water body	38.76	38.20	38.78	40.08
Mangrove/dense forest	29.54	25.36	28.16	20.19
Mangrove/spare forest	22.92	30.81	25.60	33.10
Wetland/mudflats	8.77	5.61	7.43	6.61

Table 6 Change detection matrix obtained post-classification

Class name	1989–2000 (Change in %)	2000–2011 (Change in %)	2011–2021 (Change in %)
Water body	−0.56	0.58	1.3
Mangrove/dense forest	−4.18	02.8	−7.97
Mangrove/spare forest	7.89	−5.21	7.5
Wetland/Mudflats	−3.16	1.82	−0.82

be to blame for this significant fluctuation (as the overall accuracy is 96.92% in the year 2011, as shown in the accuracy assessment section). Table 6 provides that while the area of mangrove/dense forest reduced, the area of mangrove/spare forest and wetland/mudflats rose. Mangrove/dense forest area was 608.93 km^2 in 1989, with mudflats of 180.81 km^2 and mangrove/spare forest cover of 472.52 km^2. However, in 2000, the mangrove/dense forest area was 523.04 km^2 and dramatically reduced due to an increase in mangrove/spare forest area of 635.47 km^2 (472.52 km^2 in 1989). Later in 2011, the mangrove/dense area was reduced to 582.08 km^2, with a significant increase in the wetland/mudflats area, which was 153.70 km^2, representing a 1.82% increase over the previous period. For the clipped study area of 2061 km^2, it is observed that 191.62 km^2 (9.3%) mangrove/dense forest cover area was decreased whereas 211.47 km^2 (10.18%) area of mangrove spare forest got increased (Fig. 4).

Table 7 Net forest change from the year 1989 to 2021

Category	1989–2000		2000–2011		2011–2021	
	Δ change in Area km^2	Change in %	Δ change in Area km^2	Change in %	Δ change in Area km^2	Change in %
Total/Net Forest area (dense-spare forest)	77.06	3.71	−47.31	−2.41	−9.9	−0.47

2.5 Accuracy Assessment

Accuracy assessment means a quantitative survey of how adequately the picture element was inspected into the correct class. In the present study, accuracy assessment is measured using a random sampling method for each interval, and a sample of size N is created for each period. A confusion matrix was utilized to test accuracy. Utilizing it we have tracked down the client exactness, maker precision, by and large, exactness, and Kappa coefficient by the accompanying recipes:

$$\text{Overall Accuracy} = \frac{\text{Correctly Classified Picture Elements/Points (Diagonal)}}{\text{Total No. of Reference Points}} \times 100$$

$$\text{Producer Accuracy} = \frac{\text{Correctly Classified Picture Elements/Points (Diagonal)}}{\text{Total No. of Reference Points (the column total)}} \times 100$$

$$\text{User Accuracy} = \frac{\text{Correctly Classified Picture Elements/Points (Diagonal)}}{\text{Total No. of Reference Points (the row total)}} \times 100$$

$$\text{Kappa Coefficient (T)} = \frac{\left((\text{TS} \times \text{TCC}) - \sum (\text{Column Total} \times \text{Row Total})\right)}{\left(\text{TS}^{\wedge}2 - \sum (\text{Column Total} \times \text{Row Total})\right)} \times 100$$

Here, TS = total sample; TCC = total correctly classified sample.

According to the accuracy evaluation statistics, overall accuracy gained from the random sample procedure in 1989 was 91.11%, 89.36% in 2000, 96.92% in 2011, and 96.07% in 2021. In this study, Kappa coefficient of range 0.85–0.96 was obtained which is rated as almost perfect agreement.

3 Result and Discussion

In the present research, four satellite images of the different years (1989, 2000, 2011, and 2021) have been processed in a GIS environment to formulate LU/LC maps. On evaluating these maps, it is observed that dense forest has been gradually decreasing and it got converted into other features such as spare forests, wetlands, or water bodies.
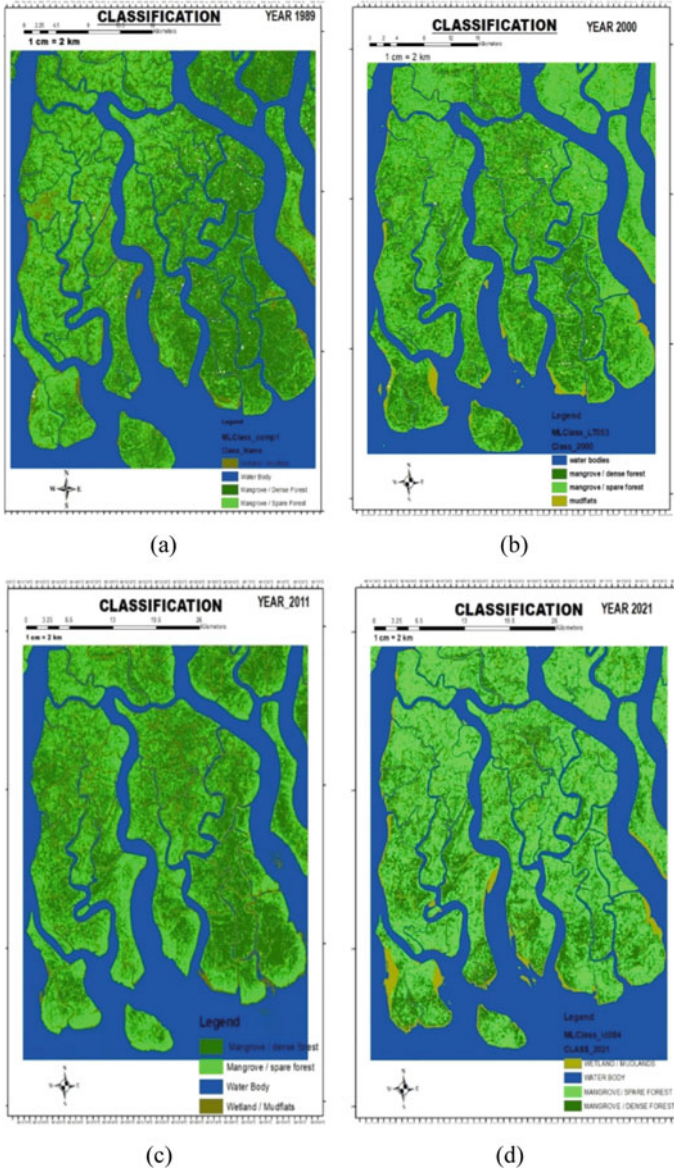
Fig. 4 LU/LC maps of the Year **a** 1989, **b** 2000, **c** 2011, **d** 2021

3.1 Visual Interpretation and Analysis

The main LULC changes that have happened in the Sundarbans over the review phase from 1989 to 2021 were initially broken down by visible translation. The visual translation can give an overall thought regarding the LULC changes throughout 32 years. Utilizing the Landsat satellite pictures, composite pictures were created. Regular and misleading variety composites were picked as an extremely simple, however exceptionally valuable strategy to outwardly extricate data from Landsat satellite pictures and to accomplish an overall depiction of LULC changes in the review region. For Landsat TM, ETM + and OLI information were utilized to make a typical variety composite. NDVI indices gave clarity about the vegetation and non-vegetation areas, and supervised classification helped in detecting temporal changes based on feature spectrums. The classification accuracy observed was better than other supervised classification methods and was very effective in improving the classification accuracy. The supervised classification using MLC classifier categorizes the pixels of composite image based on feature spectrum which has rarely been done using traditional LULC classification techniques.

3.2 Discussion

According to the accuracy evaluation figures, overall accuracy gained from the random sample procedure in 1989 was 91.11%, 89.36% in 2000, 96.92% in 2011, and 96.07% in 2021. In this study, Kappa coefficient of range, 0.85–0.96, was obtained which is rated as almost perfect agreement. The investigation resulted that there are 191.62 km^2(9.3%) of loss of dense mangrove over the selected region in the past 32 years which is one of the major concerns because of the area or subpart of the Ganges Delta (Sundarbans). The region selected for investigation is one of the core mangrove areas which is a reserved area, but results show a gradual decline in the vegetation health and an increase in spare forest cover by 10.18% which is showing poor vegetation health, the water level increased by 1.32%, and variation in wetland/ mudflats was observed showing some mixed result. As given in Table 8 in the year 1989–2000, the net forest area increased by 77.06 km^2 (3.71%). From the year 2000 to 2011, there is a drastic decrease in net forest area as well as dense mangrove area, and one of the reasons for this observed damage could be the Sundarbans assimilated the fundamental knock of typhoons [17]. From 2011 to 2021, there is again a severe decrease in dense mangroves and an increase in the water level; the biggest region for this observed degradation could be climate change, due to which cyclones are hitting the area very frequently and there is a rise in water level in the Bay of Bengal. In 2020, Amphan super cyclone struck the region and one of the authors investigated that the pH value also increased significantly in the study area due to the sudden intrusion of seawater in the estuarine system [18]. The studies recommend that the overflow in the Sundarbans' eastern streams has decreased in the last

32 years because of steadily increasing salinity and ocean water-sulfate fixations. Mangrove growth has been hampered by reduced freshwater runoff. Furthermore, farming is being harmed by undeniable degrees of saltiness in the soil because of rising tides, tornadoes, and storm floods, as well as water stagnation concerns, which have occurred in the past storm seasons [12, 17]. The pH of the estuarine water in eastern Indian Sundarbans is managed by the synergistic impact of the area of halophytes (prevalently known as blue carbon supply) and the magnitude of mixing of seawater from the Bay of Bengal [19]. Consequently, preservation of seaside vegetation and mangrove afforestation should be given need to forestall the fast speed of fermentation of the waterfront sea-going framework for the insurance of the whole Sundarbans (Ganges Delta) locale India and Bangladesh the two nations should make a joint arrangement and work on a singular level also. In India, there is even a lack of coordination between the central government body and the state government of Bengal. The central government manages the areas designated as protected forests for tiger conservation and coastal zone security, while the state government represents inhabited areas. Therefore, there is a breakdown in correspondence, a loss of time and assets, and task duplication between focal and state powers. With rising salinity and increasing ocean entrance, there is a pressing need to develop water across the board and supply more freshwater into the Sundarbans. However, due to a number of domestic pressures, there has been a lack of initiative in the Sundarbans to improve freshwater management. Entry into the forest's reserved area should indeed be strictly regulated by the government. Mangrove plantation projects and water management can also help to slow the rate of deforestation (Tables 9, 10, 11, 12, and 13).

Table 8 Confusion matrix 1989

Class name	Water Body	Mangrove/ dense forest	Mangrove/ Spare forest	Wetland/ mudflats	Total (user)	User accuracy
Water body	12	0	0	0	12	100%
Mangrove/ dense forest	0	9	1	0	10	90%
Mangrove/ spare forest	0	1	11	0	12	91.66%
Wetland/ mudflats	0	0	2	9	11	81.81%
Total (producer)	12	10	14	9	= 45	
Producer accuracy	100%	90%	78.5%	100%		OA = 41/45*100 = 91.11%

Table 9 Confusion matrix 2000

Class name	Water body	Mangrove/ dense forest	Mangrove/ Spare forest	Wetland / mudflats	Total (user)	User accuracy
Water body	11	0	0	0	11	100%
Mangrove/dense forest	0	11	1	0	12	91.66%
Mangrove/Spare forest	0	1	9	1	11	81.81%
Wetland/ mudflats	0	1	1	11	13	84.61%
Total (producer)	11	13	11	12	= 47	
Producer accuracy	100%	84.61%	81.81%	91.66%		OA = 89.36%

Table 10 Confusion matrix 2011

Class name	Water Body	Mangrove/ dense forest	Mangrove/ Spare Forest	Wetland / mudflats	Total (user)	User accuracy
Water body	15	0	0	0	15	100%
Mangrove/dense forest	0	19	0	0	19	100%
Mangrove/spare forest	0	0	16	0	16	100%
Wetland/mudflats	1	1	0	13	15	86.66%
Total (producer)	16	20	16	13	= 65	
Producer accuracy	93.75%	95.00%	100%	100%		OA = 96.92%

Table 11 Confusion matrix 2021

Class name	Water body	Mangrove/ dense forest	Mangrove/ Spare forest	Wetland / mudflats	Total (user)	User accuracy
Water body	17	0	0	0	17	100%
Mangrove/dense forest	0	11	0	0	11	100%
Mangrove/spare forest	0	1	9	0	10	90%
Wetland / mudflats	1	0	0	12	13	92.30%
Total (producer)	18	12	9	12	= 51	
Producer accuracy	94.44%	91.66%	100%	100%		OA = 96.07%

Table 12 Overall accuracy and Kappa efficiency in the present study

Year	Overall accuracy (OA)	Kappa efficiency
1989	91.11	0.88
2000	89.36	0.85
2011	96.92	0.96
2021	96.07	0.94

Table 13 User accuracy, producer accuracy, overall accuracy, and Kappa efficiency of the present study

Year	Landsat	Sample Size	Water body		Mangrove/ dense forest		Mangrove/ spare forest		Wetland / mudflats	
			UA%	PA%	UA%	PA%	UA%	PA%	UA%	PA%
1989	TM	45	100	100	90.00	90.00	91.66	78.57	81.81	100
2000	TM	47	100	100	91.66	84.61	81.81	81.81	84.61	91.66
2011	ETM +	65	100	93.7	100	95.00	100	100	86.66	100
2021	OLI	51	100	100	100	91.66	90.00	100	92.30	100

4 Future Work

The drawbacks of our investigation are that, because of depositions, the image contrast spectrum of the open sea and the coastal area are highly different. ArcGIS 10.8 software was unable to effectively separate them as a result, and mudflat's change detection score was quite poor. Also, this investigation does not contain any data regarding field tours; thus, the actual circumstance can potentially impact the precision of our calculations. Our strategy for additional project is to create a model employing data from the prior year can be used to forecast changes in the Sundarbans' land use. For future work, it would be surprisingly fascinating to utilize a blend of high-goal remote detecting symbolism to distinguish changes around here and for making the point-by-point land cover maps. Land cover planning of such an enormous region with medium-goal symbolism is exorbitant and frequently obliged by the absence of good preparation and approval information. The thought and proposal for additional work are to characterize the adjoining Landsat scenes utilizing the land cover grouping of the scene utilized in this concentrate as an underlying arrangement. Validate the results of such rapid classification to see whether the method is good enough to produce satisfactorily accurate LULC maps and to determine the trend and location of LULC changes for such a large area with similar spectral features.

5 Conclusion

This study expressed a gradual decline in the mangrove vegetation health in one of the core mangrove areas from 1989 to 2021 in 32 years. NDVI indices and NDVI maps show a decline in healthy vegetation. For change detection and land use and land cover classification, a supervised classification approach shows a drastic decrease in net forest area and an increase in water level by 1.32%. It is observed that dense forest areas converted to spare forests which show vegetation health is going down day by day. The study uncovered that mangrove are one of the weakest environments which are very inferable from constant anthropogenic anxieties in waterfront regions and climatic changeability. The present study revealed that mangroves are diminishing at the rate of 19.24 km^2. Even in the last 32 years, some regions of the Ganges Delta have already disappeared, and from classification maps, we can visualize Bhangaduni Island where 24 km^2 of the area covered with dense mangrove is now dissolved in water. In India, there is even a lack of coordination between the central government body and the state government of Bengal. The central government manages the areas designated as protected forests for tiger conservation and coastal zone security, while the state government represents inhabited areas. Therefore, there is a breakdown in correspondence, a loss of time and assets, and task duplication between focal and state powers. With rising salinity and increasing ocean entrance, there is a pressing need to develop water across the board and supply more freshwater into the Sundarbans. However, due to several domestic pressures, there has been a lack of initiative in the Sundarbans to improve freshwater management. Entry into the forest's reserved area should indeed be strictly regulated by the government. Mangrove plantation projects and water management can also help to slow the rate of deforestation.

The study recommended that some needed steps be taken into consideration by the law and policymakers for the preservation of Sundarbans such as water management and mangrove plantation on a large scale. Forest areas should be well delineated. A survey of forest resources on regular basis, making a long-term plan with state and center collaboration and a joint project with equal participation of India and Bangladesh, will enhance the capability to preserve the precious Ganges Delta.

References

1. Sievers M et al (2020) Indian Sundarbans mangrove forest considered endangered under red list of ecosystems, but there is cause for optimism. Biol Cons 251:108751. https://doi.org/10.1016/J.BIOCON.2020.108751
2. Liu K, Li X, Shi X, Wang S (2008) Monitoring mangrove forest changes using remote sensing and GIS data with decision-tree learning. 28(2):336–346, Jun 2008. https://doi.org/10.1672/06-91.1
3. Raha AK, Mishra A, Bhattacharya S, Ghatak S, Pramanick P (2022) (PDF) Sea level rise and submergence of Sundarban islands: a time series study of estuarine dynamics. Int J Ecol

Environ Sci. https://www.researchgate.net/publication/268813865_SEA_LEVEL_RISE_AND_SUBMERGENCE_OF_SUNDARBAN_ISLANDS_A_TIME_SERIES_STUDY_OF_ESTUARINE_DYNAMICS (Accessed 05 Feb 2022)

4. Thomas JV, Arunachalam A, Jaiswal R, Diwakar PG, Kiran B (2015) Dynamic land use and coastline changes in active estuarine regions—a study of Sundarban delta. Int Arch Photogrammetry Remote Sens Spat Inf Sci ISPRS Arch XL–8(1):133–139. https://doi.org/10.5194/ISPRSARCHIVES-XL-8-133-2014

5. Danda AA, Ghosh N, Bandyopadhyay J, Hazra S (2020) Strategic and managed retreat as adaptation: addressing climate vulnerability in the Sundarbans. ORF Issue Brief No 387

6. Mondal I, Bandyopadhyay J (2022) Coastal zone mapping through geospatial technology for resource management of Indian Sundarban, West Bengal, India. Int J Remote Sens Appl 4(2):103. Accessed: 05 Feb 2022. [Online]. Available: https://www.academia.edu/21760749/Coastal_Zone_Mapping_through_Geospatial_Technology_for_Resource_Management_of_Indian_Sundarban_West_Bengal_India

7. Fu XM et al. (2021) Resource status and protection strategies of mangroves in China. J Coast Conserv 25(4). https://doi.org/10.1007/S11852-021-00800-Z

8. Ross PM, Adam P (2013) Climate change and intertidal wetlands. Biology 2(1):445–480. https://doi.org/10.3390/BIOLOGY2010445

9. Liu M, Zhang H, Lin G, Lin H, Tang D (2018) Zonation and directional dynamics of mangrove forests derived from time-series satellite imagery in Mai Po, Hong Kong. Sustainability 10(6):1913, Jun 2018. https://doi.org/10.3390/SU10061913

10. Lunetta RS, Knight JF, Ediriwickrema J, Lyon JG, Worthy LD (2006) Land-cover change detection using multi-temporal MODIS NDVI data. Remote Sens Environ 105(2):142–154. https://doi.org/10.1016/J.RSE.2006.06.018

11. Schulz JJ, Cayuela L, Echeverria C, Salas J, Rey Benayas JM (1975–2008) Monitoring land cover change of the dryland forest landscape of Central Chile (1975–2008). Appl Geogr 30(3):436–447, Jul 2010. https://doi.org/10.1016/J.APGEOG.2009.12.003

12. Mahadevia K, Vikas M (2012) Climate change-impact on the sundarbans: a case study. Int Sci J Environ Sci 7–15. [Online]. Available: http://www.wildbengal.com/urls/en_habitat.htm

13. Raha A, Das S, Banerjee K, Mitra A (2012) Climate change impacts on Indian Sunderbans: a time series analysis (1924–2008). Biodivers Conserv 21(5):1289–1307. https://doi.org/10.1007/s10531-012-0260-z

14. Kumar R, Nandy S, Agarwal R, Kushwaha SPS (2014) Forest cover dynamics analysis and prediction modeling using logistic regression model. Ecol Ind 45:444–455. https://doi.org/10.1016/j.ecolind.2014.05.003

15. Cheng L, Hogarth PJ (2001) The biology of mangroves. Fla Entomol 84(3):459. https://doi.org/10.2307/3496515

16. Gandhi GM, Parthiban S, Thummalu N, Christy A (2015) Ndvi: vegetation change detection using remote sensing and Gis—A case study of Vellore District. Procedia Comput Sci 57:1199–1210. https://doi.org/10.1016/j.procs.2015.07.415

17. Khan Z, Zakir M, Khan H (2016) Disaster impact on Sundarbans-a case study on sidr affected area environmental management system of textile industry in Bangladesh: Constraint and remediation view project a case study on occupational health and safety of footwear manufacturing industry view project. Int J Res Appl Nat Soc Sci 5–12. [Online]. Available: https://www.researchgate.net/publication/310446619

18. Chakraborty S, Zaman S, Pramanick P (2022) Acidification of Sundarbans mangrove estuarine system. Discovery Nat 6. Accessed: 05 Feb 2022. [Online]. Available: http://www.discovery.org.in/dn.htm

19. Karim S (1970) Proposed REDD+ project for the Sundarbans: legal and institutional issues. Int J Rural Law Policy 1:1–7. https://doi.org/10.5130/IJRLP.I1.2013.3366

On Some New Similarity Measures for Picture Fuzzy Hypersoft Sets with Application in Medical Diagnosis

Himanshu Dhumras[ID] **and Rakesh Kumar Bajaj**[ID]

Abstract A variety of similarity measures for the extensions of fuzzy sets have been proposed and utilized in different fields such as data science, clustering techniques, medical diagnosis, econometrics and decision-making. In the present communication, two different types of similarity measures (arithmetic/trigonometric) for newly defined picture fuzzy hypersoft sets ($PFHSSs$) have been studied in detail. The inclusion of the picture fuzzy hypersoft set provides the decision-makers more flexibility to incorporate parameterizations of multi-sub-attributes and all the four components of picture fuzzy information. Necessary examples and results have been duly presented and established. Further, a novel methodology for determining the medical issue of the subject related to COVID-19 has been presented based on the presented similarity measures and the obtained results. On the basis of the technique discussed in the manuscript, a related illustrative example is solved. The provided similarity measures for $PFHSSs$ can be very effective and helpful for the decision-makers in establishing the connection between the sub-parametrization of attributes and available alternatives.

Keywords Picture fuzzy set · Hypersoft sets · Similarity measures · Decision-making · Medical diagnosis

1 Introduction

In view of the theory of multi-valent logic and consequently the vast literature on the fuzzy set theory [1], which have been developed in the due course of time by different researchers working in this field, this theory has brought an enormous archetypic change to enrich the multi-dimensional features of mathematics. In a real-

H. Dhumras · R. K. Bajaj (✉)
Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh 173 234, India
e-mail: rakesh.bajaj@juitsolan.in

H. Dhumras
e-mail: himanshudhumrash@gmail.com

life problem, choosing an ideal alternative for the sake of satisfying any practical purpose is an important task. Subsequently, the notion of multi-criteria decision-making problem gives a wider range and serves as one of the delightful options in the field of soft computing. It may be noted that the membership degree of the elements in a set (classical set theory) is investigated in binary form (0 or 1), while in case of fuzzy set theory, membership categorization is in the entire continuous range of 0 to 1. Atanassov [2] came up with the new concept of intuitionistic fuzzy set to handle with the indeterminacy/hesitation margin in the inexact information lying between membership degree and non-membership degree. Later, Yager [3] unveiled that the formal structure of fuzzy/intuitionistic fuzzy set is not able to portray the total flexibility and freedom of the decision-maker's opinion at its full. Further, the notion of Pythagorean Fuzzy Set ($PyFS$) [3] was introduced which was found to be capable enough for adequately intensifying the spell of information by presenting a new condition.

On the basis of this categorization and to address the additional component of refusal, Coung [4] introduced the notion of Picture Fuzzy Set (PFS) to deal with such situations which would be satisfactorily close to the adaptability of human nature, where all the four parameters are taken into account. It may further be noted that all the sets stated above are unable to deal with alternatives which have parametric values. In order to deal with such difficulties, Molodtsov [5] introduced the concept of soft set which has the capability to incorporate the parametric nature linked with the available alternatives/options in a decision science problem.

In spite of these extensions, Smarandache [6] observed that it is difficult to deal with some irreconcilable and inexplicit information where the involvement of attributes from a set of parameters containing further sub-attributes is present and subsequently introduced the concept of hypersoft set (HSS). Smarandache also introduced fuzzy hypersoft sets ($FHSS$) [6] to deal with the uncertainties involving sub-attribute family of parameters. Also, in order to include the indeterminacy component in the sub-attribute family of parameters, the concept of intuitionistic fuzzy hypersoft set ($IFHSS$) has further been investigated by Yolcu et al. [7]. The notion of hypersoft set has got many other important extensions with several aggregation operators [8–13].

The importance and use of similarity measurements in assessing the degree of resemblance between two or more things cannot be overstated. The idea of a similarity metric is significant in virtually every area of research and engineering. In the different areas of applications including pattern recognition, region extraction, coding theory, and image processing, similarity measures and their results are being widely utilized in literature. In assessing the degree of resemblance between two or more things, the importance and use of similarity measures cannot be overstated. The notion of similarity measures is significant in areas including "*pattern recognition, region extraction, coding theory, medical diagnosis etc.*". Various similarity measures between "*fuzzy sets, vague sets, soft sets, and fuzzy soft sets*" have been presented by many researchers. The similarity measures between two fuzzy sets have been proposed by Hyung et al.[15]; he also proposed the similarity measures between the elements of two fuzzy sets. Also, along the various extensions of fuzzy

sets (IFS, PFS, HFS, HSS) had been proposed in the literature [16]-[19]. There are various similarity and distance along the soft sets extensions, but along the hypersoft extensions similarity measures are not that much. Saqlain et al.[20] discussed the distance/similarity measures for neutrosophic hypersoft sets ($NHSSs$). Similarly, Jafar et al.[21] proposed the trigonometric similarity measures for the NHSSs, and Rahman et al. [22] proposed the similarity measures for the intuitionistic fuzzy hypersoft sets ($IFHSSs$).

In literature, the distance and similarity measures for picture fuzzy hypersoft sets and their applications are not available. The introduction of such notions of extended fuzzy sets and similarity measures would enhance the existing methodologies and provide more flexibility to the decision-maker with the inclusion of the degree of refusal and degree of abstain. The additional advantage would be to deal with any kind of sub-attribute family of parameters in decision-making problems. The objective of the present manuscript is to study the new notion of the picture fuzzy hypersoft set as an extended version so as to incorporate the sub-parameterization feature of the vague information and propose some new similarity measures with application in the field of medical diagnosis.

The contributions of the present manuscript have been sequenced as follows. In Sect. 2, the fundamental definitions and preliminaries in context with picture fuzzy hypersoft set have been presented. We propose two novel similarity measures (arithmetic and trigonometric) for the newly defined picture fuzzy hypersoft sets in Sect. 3 along with their proof, important properties and results. With the help of the obtained results, an algorithm to solve medical diagnosis problem has been outlined with a numerical illustration in Sect. 4. Finally, the paper has been concluded in Sect. 5 with some outline for future study.

2   Fundamental Notions

In this section, we are presenting the basic notions and definitions of various other fundamental sets which are available in the literature. These preliminaries would help to understand the proposed notions of picture fuzzy hypersoft set and increase the readability for the researchers.

Definition 1 *Picture Fuzzy Set(PFS) [4]. "A Picture Fuzzy Set R in V is given by $R = \{v, \rho_R(v), \tau_R(v), \omega_R(v)|v \in V\}$; where $\rho_R : V \to [0, 1]$ is the degree of positive membership of v in R , $\tau_R : V \to [0, 1]$ is the degree of neutral membership of v in R and $\omega_R : V \to [0, 1]$ is the degree of negative membership of v in R and ρ_R, τ_R, ω_R satisfies the constraint $0 \le \rho_R(v) + \tau_R(v) + \omega_R(v) \le 1$ ($\forall v \in V$)." And $\beth_R(v) = (1 - (\rho_R(v) + \tau_R(v) + \omega_R(v))$ is called the degree of refusal membership of v in V. We denote the set of all the PFSs over V by $PFS(V)$.*

Definition 2 *Soft Set(SS) [5]. "Let V be an initial universe and K be a set of parameters and $B \subseteq K$. A pair (R,B) is called a soft set over V, where R is a*

mapping from R: B→ P(V). In other words , the soft set is a parameterized family of subsets of the set V. Here, P(V) is the set of all subsets of the set V."

Definition 3 *Picture Fuzzy Soft Set(PFSS) [4]. "Let V be an initial universe and K be a set of parameters and B ⊆ K. A pair (R,B) is called a Picture Fuzzy Soft Set over V, where R is a mapping given R : B→ PFS(V) for every b ∈ B , R(b) is a picture fuzzy soft set of V and is called* **Picture Fuzzy Value** *for the set of parameter b. Here,* PFS(V) *is the set of all picture fuzzy subsets of V and R(b) = {v, ρ_{R(b)} (v) , τ_{R(b)} (v) , ω_{R(b)}(v)|v ϵ V} where ρ_{R(b)} (v) , τ_{R(b)} (v) , ω_{R(b)}(v) are the degrees of positive membership, neutral membership and negative membership respectively, with the constraint 0 ≤ ρ_{R(b)} (v) + τ_{R(b)} (v) + ω_{R(b)} (v) ≤ 1." The refusal degree is ⊐_{R(b)}(v) = (1 − (ρ_{R(b)} (v) + τ_{R(b)} (v) + ω_{R(b)} (v)) (∀ v ∈ V).*

Definition 4 *Hypersoft Set(HSS) [6]. "Let V be the universal set and P(V) be the power set of V. Consider k_1, k_2, ...k_n for n ≥ 1, be n well-defined attributes, whose corresponding attribute values are the sets K_1, K_2, . . . , K_n with K_i ∩ K_j = φ for i ≠ j and i, j ∈ {1, 2, . . . , n} . Let B_i be the non-empty subsets of K_i for each i = 1, 2, . . . , n. Then the pair (R, B_1×B_2× . . . .B_n) is said to be hypersoft set over V where R : B_1 × B_2 × . . . . × B_n → P(V). In other words, hypersoft set is a multi-parameterized family of subsets of the set V."*

Definition 5 *Similarity Measure Axioms [23]. "Let V be the universal set and F(V) be the set of all fuzzy sets over V. Consider a mapping S : F(V) × F(V) → [0, 1], for any A, B ∈ F(V), S(A, B) is called a similarity measure between fuzzy sets A and B if it satisfies the following conditions":*

  (i) 0 ≤ S(A, B) ≤ 1;
  (ii) S(A,A)=1;
  (iii) S(A, B) = S(B, A);
  (iv) if A ⊆ B ⊆ C, C ∈ F(V), S(A, C) ≤ S(A, B) and S(A, C) ≤ S(B, C).

3   Similarity Measures of Picture Fuzzy Hypersoft Sets

In this section, before defining the similarity measures on picture fuzzy hypersoft sets, we first proposing and presenting the definition of Picture Fuzzy Hypersoft Set (PFHSS) along with various important properties and fundamental operations. However, Chinnadurai and Robin [24] have very recently defined Picture Fuzzy Hypersoft Set(PFHSS) which restrict the decision-makers of being choice specific. In order to give the decision-makers more flexibility, we defined the PFHSS with a different approach. The following definition of PFHSS (the parametrization of multi-sub-attributes and all the four components of picture fuzzy information) is being proposed:

Definition 6 *(Picture Fuzzy Hypersoft Set).Let V = {v_1, v_2, . . . , v_n} be the universe of discourse and PFS(V) be the collection of all picture fuzzy subsets of V.*

Suppose $K_1, K_2, \ldots, K_m$ for $m \geq 1$ be m well-defined attributes, whose respective attribute values are the sets $K_1^a, K_2^b, \ldots, K_m^z$ with the relation $K_1^a \times K_2^b \times \cdots \times K_m^z$ where $a, b, c, \ldots, z = 1, 2, \ldots, n$.
The pair $\left(R, K_1^a \times K_2^b \times \cdots \times K_m^z\right)$ is called a picture fuzzy hypersoft set over V where $R : K_1^a \times K_2^b \times \cdots \times K_m^z \to \text{PFS}(V)$ defined by

$$R\left(K_1^a \times K_2^b \times \cdots \times K_m^z\right) = \big\{ < v, \rho_{R(\vartheta)}(v), \tau_{R(\vartheta)}(v), \omega_{R(\vartheta)}(v) > \mid v \in V,$$
$$\vartheta \in K_1^a \times K_2^b \times \cdots \times K_m^z \big\}.$$

Here, ρ, τ and ω represent the positive membership, neutral membership and negative membership degrees, respectively. Consider $PFHSS(V)$ as the collection of all $PFHSSs$ over V and $K_1^a \times K_2^b \times \cdots \times K_m^z$ by Λ.

Next, we define the similarity/weighted similarity measures between PFHSSs and study its fundamental operations.

Definition 7 *Let V be the domain of discourse and $PFHSS(V)$ be the collection of all picture fuzzy hypersoft sets over V. Consider a mapping $S : PFHSS(V) \times PFHSS(V) \to [0, 1]$, for any $\langle (R, \Lambda_1), (R', \Lambda_2) \rangle \in PFHSS(V)$, $S\langle (R, \Lambda_1), (R', \Lambda_2) \rangle$ is called a similarity measure between the picture fuzzy hypersoft sets (R, Λ_1) and (R', Λ_2) if it satisfies the following conditions:*

(i) $0 \leq S\langle (R, \Lambda_1), (R', \Lambda_2) \rangle \leq 1$;
(ii) $S\langle (R, \Lambda_1), (R', \Lambda_2) \rangle = 1 \Leftrightarrow R = R'$;
(iii) $S\langle (R, \Lambda_1), (R', \Lambda_2) \rangle = S\langle (R', \Lambda_2), (R, \Lambda_1) \rangle$;
(iv) Let (R'', Λ_3) be a picture fuzzy hypersoft set, if $(R, \Lambda_1) \subseteq (R', \Lambda_2)$ and $(R', \Lambda_2) \subseteq (R'', \Lambda_3)$, then $S\langle (R, \Lambda_1), (R'', \Lambda_3) \rangle \leq S\langle (R, \Lambda_1), (R', \Lambda_2) \rangle$ and $S\langle (R, \Lambda_1), (R'', \Lambda_3) \rangle \leq S\langle (R', \Lambda_2), (R'', \Lambda_3) \rangle$.

Note: Now, for any $PFHSS (R, \Lambda_1) \in PFHSS(V), \Lambda_1 \subseteq \Lambda$, we extend the $PFHSS (R, \Lambda_1)$ to the $PFHSS (R, \Lambda)$ where $R(\vartheta) = \phi, \forall \vartheta \notin \Lambda$, that is, $\rho_{R(\vartheta)}(v) = \tau_{R(\vartheta)}(v) = \omega_{R(\vartheta)}(v) = 0, \forall \vartheta \notin \Lambda$. So, now on we will take for any attribute subset of $PFHSS$ over V, same as the attribute set Λ.

Definition 8 *Let (R, Λ) and (R', Λ) be any two PFHSSs over the universe of discourse $V = \{v^1, v^2, ..., v^n\}$ with the attribute set values $K_1^a \times K_2^b \times \cdots \times K_m^z$. Then, a similarity measure between (R, Λ) and (R', Λ) can be defined as*

$$\mathbb{S}_{\text{PFHSS}}(R, R') = \tag{1}$$

$$\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} \frac{1-\frac{1}{2}\left[\begin{array}{c}\left(\min\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R'(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R'(\vartheta_j^s)}(v^i)|\}\right.\\ \left.+|\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R'(\vartheta_j^s)}(v^i)|\right)\end{array}\right]}{1+\frac{1}{2}\left[\begin{array}{c}\left(\max\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R'(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R'(\vartheta_j^s)}(v^i)|\}\right.\\ \left.+|\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R'(\vartheta_j^s)}(v^i)|\right)\end{array}\right]}$$

Table 1 Decision-maker's opinion for qualification

K_1^a (*Qualification*)	v^1	v^2	v^3	v^4	v^5
BS Hons.	(0.2, 0.3, 0.4)	(0.1, 0.3, 0.5)	(0.3, 0.5, 0.1)	(0.3, 0.2, 0.1)	(0.4, 0.1, 0.2)
MS	(0.1, 0.2, 0.4)	(0.4, 0.2, 0.3)	(0.3, 0.5, 0.1)	(0.1, 0.1, 0.5)	(0.1, 0.3, 0.4)
M.Phil.	(0.1, 0.3, 0.5)	(0.1, 0.3, 0.5)	(0.1, 0.3, 0.5)	(0.1, 0.3, 0.5)	(0.1, 0.3, 0.5)
Ph.D.	(0.3, 0.1, 0.5)	(0.4, 0.3, 0.1)	(0.2, 0.5, 0.1)	(0.2, 0.1, 0.5)	(0.3, 0.1, 0.5)

where, $j = 1, 2, ...m; i = 1, 2, .., n; s = a, b, ..., z; a, b, ..., z = 1, 2, ..., n$ *and* $\vartheta_j^s \in K_1^a \times K_2^b \times \cdots \times K_m^z$.

Definition 9 *Let* (R, Λ) *and* (R', Λ) *be any two PFHSSs over the universe of discourse* $V = \{v^1, v^2, ..., v^n\}$ *with the attribute set values* $K_1^a \times K_2^b \times \cdots \times K_m^z$. *Then, a tangent similarity measure between* (R, Λ) *and* (R', Λ) *can be defined as*

$$\mathbb{T}_{PFHSS}(R, R') = \qquad\qquad (2)$$

$$\frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \left\{ 1 - \tan \frac{\pi}{12} \left[\begin{pmatrix} |\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R'(\vartheta_j^s)}(v^i)| + |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R'(\vartheta_j^s)}(v^i)| \\ + |\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R'(\vartheta_j^s)}(v^i)| \end{pmatrix} \right] \right\}$$

where, $j = 1, 2, ...m; i = 1, 2, .., n; s = a, b, ..., z; a, b, ..., z = 1, 2, ..., n$ *and* $\vartheta_j^s \in K_1^a \times K_2^b \times \cdots \times K_m^z$.

In order to better understand the proposition, we take up an illustrative example of picture fuzzy hypersoft sets and show the computational part of the proposed similarity measures as follows:

Example 1 Suppose a need arises for a school to hire a mathematics teacher for tenth class. A total of five candidates have applied to fill up the void space. The human resource cell of the school appoints an expert/decision-maker for this selection process. Let $V = \{v^1, v^2, v^3, v^4, v^5\}$ be the set of all five candidates with their set of attributes as K_1 = Qualification, K_2 = Experience, K_3 = Age, K_4 = Gender. Further, their respective sub-attributes are
K_1 = Qualification = {BS Hons., MS, M.Phil., Ph.D.}
K_2 = Experience = {3yr, 5yr, 7yr, 10yr}
K_3 = Age = {Less than twenty five, Great than twenty five}
K_4 = Gender = {Male, Female}.

Let the function be $R : K_1^a \times K_2^b \times \cdots \times K_m^z \rightarrow \text{PFS}(V)$. Based on some empirical-hypothetical assumptions and the decision-maker's opinion, we present the following tables (Tables 1, 2, 3 and 4) with respect to each attribute and with their further sub-attributes.

Table 2 Decision-maker's opinion for experience

K_2^b (Experience)	v^1	v^2	v^3	v^4	v^5
3yr	(0.4, 0.3, 0.1)	(0.2, 0.3, 0.5)	(0.2, 0.2, 0.3)	(0.5, 0.1, 0.2)	(0.2, 0.3, 0.5)
5yr	(0.1, 0.3, 0.5)	(0.3, 0.3, 0.3)	(0.3, 0.3, 0.3)	(0.2, 0.4, 0.2)	(0.7, 0.1, 0.1)
7yr	(0.1, 0.7, 0.1)	(0.4, 0.3, 0.2)	(0.1, 0.3, 0.5)	(0.2, 0.2, 0.5)	(0.2, 0.5, 0.2)
10yr	(0.3, 0.5, 0.1)	(0.2, 0.4, 0.3)	(0.3, 0.3, 0.2)	(0.1, 0.3, 0.6)	(0.6, 0.3, 0.1)

Table 3 Decision-maker's opinion for age

K_3^c (Age)	v^1	v^2	v^3	v^4	v^5
< 25	(0.2, 0.1, 0.5)	(0.3, 0.3, 0.3)	(0.5, 0.2, 0.1)	(0.6, 0.2, 0.1)	(0.2, 0.3, 0.4)
> 25	(0.5, 0.3, 0.1)	(0.4, 0.3, 0.1)	(0.2, 0.4, 0.3)	(0.5, 0.2, 0.2)	(0.4, 0.3, 0.1)

Table 4 Decision-maker's opinion for gender

K_4^d (Gender)	v^1	v^2	v^3	v^4	v^5
Male	(0.2, 0.1, 0.2)	(0.3, 0.2, 0.3)	(0.1, 0.2, 0.6)	(0.4, 0.2, 0.3)	(0.5, 0.2, 0.1)
Female	(0.2, 0.1, 0.5)	(0.4, 0.3, 0.1)	(0.3, 0.5, 0.1)	(0.2, 0.2, 0.2)	(0.4, 0.3, 0.1)

Now, let us consider

$$R\left(K_1^a \times K_2^b \times K_3^c \times K_4^d\right) = R\,(\text{MS, 7yr, greater than twenty five, Male}) = \left(v^1, v^2, v^3, v^5\right).$$

For the above relational expression, the picture fuzzy hypersoft set can be expressed as

$$R\left(K_1^a \times K_2^b \times K_3^c \times K_4^d\right) =$$

$\{< v^1, (MS(0.1, 0.2, 0.4), 7yr(0.1, 0.7, 0.1), \text{Greater than twenty five}(0.5, 0.3, 0.1), Male(0.2, 0.1, 0.2)) >$
$< v^2, (MS(0.4, 0.2, 0.3), 7yr(0.4, 0.3, 0.2), \text{Greater than twenty five}(0.4, 0.3, 0.1), Male(0.3, 0.2, 0.3)) >$
$< v^3, (MS(0.3, 0.5, 0.1), 7yr(0.1, 0.3, 0.5), \text{Greater than twenty five}(0.2, 0.4, 0.3), Male(0.1, 0.2, 0.6)) >$
$< v^5, (MS(0.1, 0.3, 0.4), 7yr(0.2, 0.5, 0.2), \text{Greater than twenty five}(0.4, 0.3, 0.1), Male(0.5, 0.2, 0.1)) >\}$

Let

$$(R, \Lambda) = \{< v^1, (MS\,(0.1, 0.2, 0.4),\ 7yr\,(0.1, 0.7, 0.1), \text{Greater than twenty five}$$
$$(0.5, 0.3, 0.1), Male\,(0.2, 0.1, 0.2)) >\}$$

and

$$(R', \Lambda) = \{< v^3, (MS\,(0.3, 0.5, 0.1),\ 7yr\,(0.1, 0.3, 0.5), \text{Greater than twenty five}$$
$$(0.2, 0.4, 0.3), Male\,(0.1, 0.2, 0.6)) >\}$$

be two PFHSSs.

Then, $\mathbb{S}_{PFHSS}(R, R') = 0.8588$. Similarly, $\mathbb{T}_{PFHSS}(R, R') = 0.8588$.

Theorem 1 *Let (R, Λ), (R', Λ) and (R'', Λ) be three PFHSSs over the universal set V. Then $\mathbb{S}_{PFHSS}$ satisfies the four axioms of similarity measures as follows:*

   (i) $0 \leq S(R, R') \leq 1$;
  (ii) $S(R, R') = S(R', R)$
 (iii) $S(R, R') = 1 \Leftrightarrow R = R'$
 (iv) If $(R, \Lambda) \subseteq (R', \Lambda)$ and $(R', \Lambda) \subseteq (R'', \Lambda)$, then $\mathbb{S}_{\text{PFHSS}}(R, R'') \leq \mathbb{S}_{\text{PFHSS}}(R, R')$ and $\mathbb{S}_{PFHSS}(R, R'') \leq \mathbb{S}_{PFHSS}(R', R'')$.

Proof. Proof of (i) and (ii) can be easily done by making use of the definition of the proposed measure.

(iii) For the proof of this part, let us suppose $R = R'$

Then, $\rho_{R(\vartheta_j^s)}(v^i) = \rho_{R'(\vartheta_j^s)}(v^i)$, $\tau_{R(\vartheta_j^s)}(v^i) = \tau_{R'(\vartheta_j^s)}(v^i)$, $\omega_{R(\vartheta_j^s)}(v^i) = \omega_{R'(\vartheta_j^s)}(v^i)$.

$\Rightarrow S(R, R') = 1$.

Conversely, let $S(R, R') = 1$.

$$\Rightarrow \frac{1 - \frac{1}{2}\left[\begin{array}{c}\left(\min\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R'(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R'(\vartheta_j^s)}(v^i)|\}\right. \\ \left.+|\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R'(\vartheta_j^s)}(v^i)|\right)\end{array}\right]}{1 + \frac{1}{2}\left[\begin{array}{c}\left(\max\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R'(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R'(\vartheta_j^s)}(v^i)|\}\right. \\ \left.+|\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R'(\vartheta_j^s)}(v^i)|\right)\end{array}\right]} = 1$$

$\Rightarrow 1 - \frac{1}{2}[\min\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R'(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R'(\vartheta_j^s)}(v^i)|\} + |\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R'(\vartheta_j^s)}(v^i)|] = 1 + \frac{1}{2}[\max\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R'(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R'(\vartheta_j^s)}(v^i)|\} + |\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R'(\vartheta_j^s)}(v^i)|]$

$\Rightarrow \frac{1}{2}[\min\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R'(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R'(\vartheta_j^s)}(v^i)|\} + |\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R'(\vartheta_j^s)}(v^i)|] + \frac{1}{2}[\max\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R'(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R'(\vartheta_j^s)}(v^i)|\} + |\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R'(\vartheta_j^s)}(v^i)|] = 0$

$\Rightarrow |\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R'(\vartheta_j^s)}(v^i)| = 0$, $|\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R'(\vartheta_j^s)}(v^i)| = 0$ and $|\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R'(\vartheta_j^s)}(v^i)| = 0$

$\Rightarrow R = R'$.

(iv) $\langle R, \Lambda \rangle \subseteq \langle R', \Lambda \rangle \subseteq \langle R'', \Lambda \rangle$.

$\Rightarrow |\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R'(\vartheta_j^s)}(v^i)| \leq |\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R''(\vartheta_j^s)}(v^i)|$, $|\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R'(\vartheta_j^s)}(v^i)| \leq |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R''(\vartheta_j^s)}(v^i)|$ and $|\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R'(\vartheta_j^s)}(v^i)| \leq |\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R''(\vartheta_j^s)}(v^i)|$.

$\Rightarrow \min\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R'(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R'(\vartheta_j^s)}(v^i)|\} + |\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R'(\vartheta_j^s)}(v^i)| \leq \min\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R''(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R''(\vartheta_j^s)}(v^i)|\} + |\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R''(\vartheta_j^s)}(v^i)|$

and $\max\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R'(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R'(\vartheta_j^s)}(v^i)|\} + |\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R'(\vartheta_j^s)}(v^i)| \leq \max\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R''(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R''(\vartheta_j^s)}(v^i)|\} + |\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R''(\vartheta_j^s)}(v^i)|.$

$\Rightarrow 1 - \frac{1}{2}[\min\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R'(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R'(\vartheta_j^s)}(v^i)|\} + |\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R'(\vartheta_j^s)}(v^i)|] \geq 1 - \frac{1}{2}[\min\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R''(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R''(\vartheta_j^s)}(v^i)|\} + |\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R''(\vartheta_j^s)}(v^i)|]$

and
$1 + \frac{1}{2}[\max\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R'(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R'(\vartheta_j^s)}(v^i)|\} + |\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R'(\vartheta_j^s)}(v^i)|] \leq 1 + \frac{1}{2}[\max\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R''(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R''(\vartheta_j^s)}(v^i)|\} + |\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R''(\vartheta_j^s)}(v^i)|]$

$\Rightarrow \mathbb{S}_{PFHSS}(R, R'') \leq \mathbb{S}_{PFHSS}(R, R')$. Likewise, we can prove

$$\mathbb{S}_{PFHSS}(R, R'') \leq \mathbb{S}_{\text{PFHSS}}(R', R'').$$

Theorem 2 *Let* (R, Λ), (R', Λ) *and* (R'', Λ) *be three* $PFHSSs$ *over the universal set* V. *Then* $\mathbb{T}_{PFHSS}$ *also satisfies the four axioms of similarity measures.*

Proof. The proof can be done on the similar lines as above.

Remark: In order to address the importance of considered attribute set, we need to assign them with weights. Let w_j be the weights corresponding to ϑ_j^s, then the similarity measure for $PFHSSs$ (R, Λ) & (R', Λ) in the weighted form can be given by

$$\mathbb{W}_{PFHSS}(R, R') = \tag{3}$$

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} w_j \frac{1 - \frac{1}{2}\left[\begin{pmatrix} \min\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R'(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R'(\vartheta_j^s)}(v^i)|\} \\ + |\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R'(\vartheta_j^s)}(v^i)| \end{pmatrix} \right]}{1 + \frac{1}{2}\left[\begin{pmatrix} \max\{|\rho_{R(\vartheta_j^s)}(v^i) - \rho_{R'(\vartheta_j^s)}(v^i)|, |\tau_{R(\vartheta_j^s)}(v^i) - \tau_{R'(\vartheta_j^s)}(v^i)|\} \\ + |\omega_{R(\vartheta_j^s)}(v^i) - \omega_{R'(\vartheta_j^s)}(v^i)| \end{pmatrix} \right]}$$

where $w = (w_1, w_2, ..., w_m)^T$ is the weight vector of $\vartheta_j^s (j = 1, 2, ..., m; s = a, b, c, ...z; a, b, c, ..., z = 1, 2, ..., n)$ with $w_j \in [0, 1]$, $j = 1, 2, ..., m$; $\sum_{j=1}^{m} w_j = 1$. In particular, if $w = \left(\frac{1}{m}, \frac{1}{m}, ..., \frac{1}{m}\right)^T$, Eq. 3 reduces to Eq. 1, i.e., $\mathbb{S}_{PFHSS}(R, R') = \mathbb{W}_{PFHSS}(R, R')$. Similarly, we can define the weighted similarity measure for $\mathbb{T}_{PFHSS}$.

Note: We can also validate the weighted similarity measure by making use of similarity axioms as stated above.

Definition 10 *The two* $PFHSSs$ (R, Λ) *and* (R', Λ) *are said to be* $\approx^\alpha -similar$, *denoted by* $(R, \Lambda) \approx^\alpha (R', \Lambda) \Leftrightarrow \mathbb{S}_{PFHSS}(R, R') \geq \alpha$ *for* $\alpha \in (0, 1)$.

Definition 11 *The two* $PFHSSs$ (R, Λ) *and* (R', Λ) *are said to be significantly similar if* $\mathbb{S}_{PFHSS}(R, R') \geq 0.8$.

4   Application of Proposed PFHSSs Similarity Measures in Medical Diagnosis

Here, a methodology for the diagnosis of a medical problem on the basis of presented similarity measures of $PFHSSs$ has been proposed. The procedural steps of the methodology are outlined in Fig. 1.

Further, a numerical illustration has also been presented which involves the similarity measures of two $PFHSSs$ to detect whether a patient is suffering from a particular disease or not. Let us consider that there are two patients I_1, I_2 in a hospital having a symptoms of COVID-19. Suppose there are three stages of characterization of the symptoms as $severe(v^1)$, $mild(v^2)$ and $no(v^3)$, i.e., the universal set $V = \{v^1, v^2, v^3\}$.

Let

$$K = \{K^1 = \text{sense of taste}, \ K^2 = \text{temperature}, \ K^3 = \text{chest pain}, \ K^4 = \text{flu}\}$$

be the collection of symptoms which are classified into sub-attributes as: $K^1 =$ "*sense of taste*" $= \{$"*no taste*", "*Can taste*"$\}$ $K^2 =$ "*temperature*" $= \{$"$97.5 - 98.5$", "$98.6 - 99.5$", "$99.6 - 101.5$", "$101.6 - 102.5$"$\}$ $K^3 =$ "*chest pain*" $=$ $\{$"*shortness of breath*", "*no pain*", "*normal pain angina*"$\}$ $K^4 =$ "*flu*" $=$ $\{$"*sore throat*", "*cough*", "*strep throat*"$\}$. Now, let us define a relation $R : (K_1^a \times K_2^b \times K_3^c \times K_4^d) \longrightarrow P(V)$ defined as,
$R(K_1^a \times K_2^b \times K_3^c \times K_4^d) = \{\Im = \text{shortness of breath}, \ \wp = 101.3, \ \Re = \text{sore throat}, \ \mho = no \ \text{taste}\}$ is the most prominent sample of the patient for the confirmation of the COVID-19. Two patients are randomly selected based on the above sample. Let (R, Λ) be a $PFHSS$ over V for COVID-19 tabulated with the help of a medical expert as provided in Table 5.

Next, the $PFHSSs$ with respect to two patients which are under trial has been framed in Tables 6 and 7.
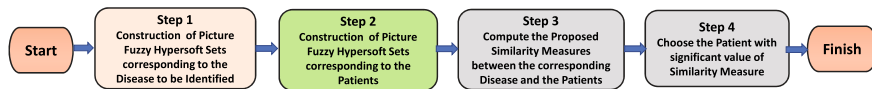
Fig. 1 Proposed methodology

Table 5 $PFHSS(R, \Lambda)$ for COVID-19

(R, Λ)	K_1^a	K_2^b	K_3^c	K_4^d
v^1	$(\Im(0.4, 0.1, 0.1))$	$(\wp(0.3, 0.0, 0.3))$	$(\Re(0.7, 0.0, 0.2))$	$(\mho(0.4, 0.2, 0.3))$
v^2	$(\Im(0.5, 0.1, 0.3))$	$(\wp(0.1, 0.3, 0.2))$	$(\Re(0.4, 0.3, 0.1))$	$(\mho(0.1, 0.2, 0.3))$
v^3	$(\Im(0.3, 0.5, 0.1))$	$(\wp(0.1, 0.3, 0.5))$	$(\Re(0.0, 0.4, 0.3))$	$(\mho(0.1, 0.2, 0.6))$

Table 6 $PFHSS(R, \Lambda)$ for the patient I_1

(I_1, Λ)	K_1^a	K_2^b	K_3^c	K_4^d
v^1	$(\Im(0.4, 0.2, 0.2))$	$(\wp(0.3, 0.1, 0.2))$	$(\Re(0.5, 0.1, 0.3))$	$(\mho(0.4, 0.1, 0.1))$
v^2	$(\Im(0.5, 0.2, 0.2))$	$(\wp(0.1, 0.2, 0.4))$	$(\Re(0.2, 0.3, 0.1))$	$(\mho(0.1, 0.5, 0.3))$
v^3	$(\Im(0.2, 0.5, 0.1))$	$(\wp(0.2, 0.3, 0.5))$	$(\Re(0.0, 0.3, 0.3))$	$(\mho(0.1, 0.2, 0.0))$

Table 7 $PFHSS(R, \Lambda)$ for the patient I_2

(I_2, Λ)	K_1^a	K_2^b	K_3^c	K_4^d
v^1	$(\Im(0.3, 0.2, 0.3))$	$(\wp(0.2, 0.3, 0.1))$	$(\Re(0.5, 0.1, 0.0))$	$(\mho(0.4, 0.0, 0.1))$
v^2	$(\Im(0.4, 0.2, 0.2))$	$(\wp(0.3, 0.2, 0.1))$	$(\Re(0.1, 0.4, 0.2))$	$(\mho(0.2, 0.4, 0.5))$
v^3	$(\Im(0.1, 0.5, 0.1))$	$(\wp(0.2, 0.3, 0.0))$	$(\Re(0.0, 0.5, 0.3))$	$(\mho(0.4, 0.2, 0.2))$

Now, by making use of the proposed similarity measure, we get $\mathbb{S}_{PFHSS}(R, I_1) = 0.8657 > 0.75$ and $\mathbb{S}_{PFHSS}(R, I_2) = 0.6892 < 0.75$.

Therefore, we conclude that the patient I_1 is suffering from COVID-19.

5 Concluding Remarks

The proposition of picture fuzzy hypersoft sets and their similarity measures can be a strong mathematical notions to deal with incomplete, inexact and vague information which may be the sub-parametrization feature. The propositions have been well validated with rigorous mathematical proof along with various important properties and results. The introduction of the arithmetic and trigonometric similarity measures has been successfully validated. These similarity measures of $PFHSSs$ provide the determining results for the diagnosis of a medical problem based on the systematic procedure. As an expansion of the work, the proposed methodology can be actively utilized in the wide range of decision-making issues where sub-parametrization becomes an unavoidable component. Also, the notion of entropy and directed divergence for picture fuzzy hypersoft sets may further be studied and applied in the different problems of machine learning decision-making processes.

References

1. Zadeh LA (1965) Inf Control. Fuzzy sets 8(3):338–353
2. Atanassov KT (1986) Intuitionistic fuzzy sets. Fuzzy Sets Syst 20(1):87–96
3. Yager RR (2013) Pythagorean fuzzy subsets. Joint IFSA world congress and NAFIPS annual meeting (IFSA/NAFIPS) 57–61

4. Coung B (2013) Picture fuzzy sets-first results Part 1. Seminar on Neuro-Fuzzy Systems with Applications, Institute of Mathematics
5. Molodtsov DA (1999) Soft set theory-first results. Comput Math Appl 37(4–5):19–31
6. Smarandache F (2018) Extension of soft set to hypersoft set, and then to plithogenic hypersoft set. Neutrosophic Sets Syst 22:168–170
7. Yolcu A, Smarandache F, Ozturk YT (2021) Intutionstc fuzzy hypersoft sets. Commun Fac Sci Univ Ank Ser A1 Math Stat 70(1):448–455
8. Rana S, Qayyum M, Saeed M, Smarandache F (2019) Pithogenic fuzzy whole hypersoft set: construction of operators their application in frequency matrix MADM technique. Neutrosophic Sets Syst 28:34–50
9. Zulqarnain RM, Xin XL, Saqlain M, Smarandache F (2020) Generalized aggregate operators on neutrosophic hypersoft set. Neutrosophic Sets Syst 36:271–281
10. Zulqarnain RM, Xin XL, Saeed M (2021) A development of pythagorean fuzzy hypersoft set with basic operations and decisionmaking approach based on the correlation coefficient. In: Theory and application of hypersoft set; Pons Publishing House Brussels, Brussels, Belgium, chapter5, pp 85–106
11. Zulqarnain RM, Siddique I, Jarad F, Ali R, Abdeljawad T (2021) Development of TOPSIS technique under pythagorean fuzzy hypersoft environment based on correlation coefficient and its application towards the selection of antivirus mask in COVID-19 pandemic. Complexity
12. Samad A, Zulqarnain RM, Sermutlu E, Ali R, Siddique I, Jarad F, Abdeljawad T (2021) Selection of an effective hand sanitizer to reduce COVID-19 effects and extension of TOPSIS. Complexity 1–23
13. Zulqarnain RM, Xin XL, Saqlain M, Saeed M, Smarandache F, Ahamad MI (2021) Some fundamental operations on interval valued neutrosophic hypersoft set with their properties. Neutrosophic Sets Syst 40:134–148
14. Zulqarnain RM, Xin XL, Ali B, Broumi S, Abdal S, Ahamad MI (2021) Decision-making approach based on correlation coefficient with its properties under interval-valued neutrosophic hypersoft set environment. Neutrosophic Sets Syst 40:12–28
15. Lee-Kwang H, Song YS, Lee KM (1994) Similarity measure between fuzzy sets and between elements. Fuzzy Sets Syst 62:291Ű-293
16. Zhang XL, Xu ZS (2014) Extension of TOPSIS to multiple criteria decision making with Pythagorean fuzzy sets. Int J Intell Syst 29:1061–1078
17. Hwang CM, Yang MS, Hung WL (2018) New similarity measures of intuitionistic fuzzy sets based on the Jaccard index with its application to clustering. Int J Intel Syst 33:1672–1688
18. Ulucay V, Deli I, Sahin M (2018) Similarity measures of bipolar neutrosophic sets and their application to multiple criteria decision making. Neural Comput Appl 29:739–748
19. Yang MS, Hussian Z (2019) Distance and similarity measures of hesitant fuzzy sets based on Hausdorff metric with applications to multi-criteria decision making and clustering. Soft Comput 23:5835–5848
20. Saqlain M, Riaz M, Saleem MA, Yang MS (2021) Distance and similarity measures for Neutrosophic HyperSoft set (NHSS) with construction of NHSS-TOPSIS and applications. IEEE Access 9:30803–30816
21. Jafar MN, Saeed M, Saqlain M, Yang MS (2021) Trigonometric similarity measures for Neutrosophic Hypersoft sets with application to renewable energy source selection. IEEE Access 9:129178–129187
22. Rahman AU, Saeed M, Khalifa HA, Afifi WA (2022) Decision making algorithmic techniques based on aggregation operations and similarity measures of possibility intuitionistic fuzzy hypersoft sets. AIMS Mathematics 7(3):3866–3895
23. Zwick, R., Carlstein, E., Budescu, D. V.: Measures of similarity among fuzzy concepts: A comparative analysis. Int J Approximate Reasoning 1:221Ű-242
24. Chinnadurai V, Robin V (2021) Picture fuzzy hypersoft TOPSIS method based on correlation coefficient. J Hyperstructes 10(2):86–107

Development of a Coconut De-Husking Machine

Oluwasina Lawan Rominiyi, Salim Mohammed, Ayodeji Olalekan Salau (ID)**, Shruti Jain** (ID)**, Johnson Felix Eiche, and Olarenwaju Thomas Oginni**

Abstract There are various techniques for de-husking coconut. These techniques are used to a great extent in taking off the exocarp of a coconut (husk), but these techniques have many concerns and restrictions. These issues reduce the production rate of de-husking coconut. These methods are more dangerous and harmful to the user. The paper aims to reduce human effort and accidents when de-husking using crude tools as well as maintaining a hygienic environment. This paper majorly focuses on the development of a coconut de-husking machine. The machine consists of a frame for support. It has two rollers with blades parallel to each other. The 5 h.p electric motor powers the machine by using a chain drive to transmit power to the shaft which rotates the roller blades. As the blade roller rotates, the husk is removed. The machine can de-husk a coconut within 30 s which is faster than the manual de-husking method which takes approximately 250 s. It can de-husk coconut of different types and sizes. The machine is also found to be approximately nine times faster than the manual de-husking method. This new coconut de-husking machine will help increase an economy that relies on coconuts for the renewable energy industry and coir product firm. This machine will be useful to the coconut processing factories.

O. L. Rominiyi · S. Mohammed
Department of Mechanical and Mechantronics Engineering, Afe Babalola University, Ado-Ekiti, Nigeria

A. O. Salau
Department of Electrical/Electronics and Computer Engineering, Afe Babalola University, Ado-Ekiti, Nigeria

S. Jain (✉)
Department of Electronics and Communications Engineering, Jaypee University of Information Technology, Solan, Himachal Pradesh, India
e-mail: shruti.jain@juit.ac.in

J. F. Eiche
Department of Mechanical Engineering, Olusegun Agagu University of Science and Technology, Okitipupa, Ondo State, Nigeria

O. T. Oginni
Department of Civil Engineering Bamidele, Olumilua University of Science Education and Technology, Ekiti State, Ikere-Ekiti, Nigeria

1 Introduction

Coconut (Cocosnucifera) is a crop that can last for more than one season and serves great importance in the world. The coconut fruit is made up of a thick fiber known as husk and the shell and can also be called the endocarp and exocarp. Coconut is the main crop grown in the coastal regions, many farmers deal with growing and selling coconuts as a cash crop and a food crop, and so many countries depend on it for their agricultural income. Coconut fruit comprises the meat which can be eaten, used to make ice cream, live-stock feed, or shredded. It can also be used in milk production and nutritious drinks and also for cooking as cooking oil. It is also used in the production of soaps, body cream, and medicines [1]. A dry coconut shell can be used for fuel purposes and acts as an alternate source of heat which has led to reduced usage of firewood. Fiber gotten from the husk of the coconut outer shell can be used in the manufacture of coir products such as coir carpets, coir geotextile, coir composite, coir pith, and coir mattings. The coir can be used for various things which are extracted from mesocarp tissue. Therefore, the process of de-husking is necessary [2].

De-husking of coconut is a very tiresome job and many people who don't want to partake in it hesitate because it can cause harm using the traditional or manual method. There are so many machines that have been made to ease and mechanize this operation but some of them are expensive and some are not in existence again [3].

The coconut also known as Cocosnucifera L. is a very utile fruit in the world today. The structure is shown in Fig. 1 [4]. So many countries depend on it for their agricultural economy as it provides food and shelter, it can serve as raw material for industries [1].

Fig. 1 Structure of the coconut fruit
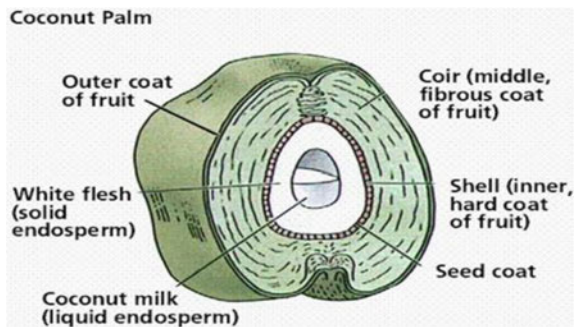
Coconut de-husking methods are divided into two categories, namely: manual de-husking method and automated de-husking method.

- **The manual de-husking process** is usually done with traditional tools such as ax, machetes, and knives. Aside from those traditional tools, there are existing manually operated de-husking tools that are upgraded [5]. Twin blade coconut husk removing tool, the coconut spanner, and the coconut husk removing tool are manual de-husking process. The twin blade coconut husk removing tool was developed as an altered version or an upgraded version of the smithy tong. The two lips are very sharp like knives when they are in a closed position. They are stroked on the coconut, and then disjoined to free the husk. The procedure was done repeatedly before removing the kernel [5]. *The coconut spanner tool* was developed by the Kerala's in India in the 1990s. It was an upgraded version of the twin blade tool which had was like a long sharp scissor. It is placed on the coconut to cut the husk out by using the sharp blade and then loosened to free the husk. Thus, the operator has to go down to change the position of the coconut before repeating the process [6]. *The coconut husk removing tool* was developed by Edward D. Hill. The tool comprised of a non-moveable blade and a movable blade, which was connected to the middle of a crossbar, the frame was formed from the ends of the crossbar that are placed on a rod. The coconut was set in a bowl in a vertical position and the blade connected to the crossbar can slip by in a vertical way which helps the blade to stake the coconut husk and free the non-moveable and movable blade resulting in the removal of the husk. This process has to done repeatedly to remove the husk completely. The issues with this tool were that the coconut doesn't always stay in the bowl when it's being fed.
- **Automatic De-husking Method**. In this method, coconut de-husking machines are electrically or mechanically operated. The mechanical coconut husking machine, continuous power-operated coconut husking machine, and the hydraulic coconut de-husking machine are the types of automatic de-husking method. The *mechanical coconut husking machine* consists of three components—inlet, de-husking component, and outlet. The coconut slides by holding it in a vertical position using hand close to the inlet where it is gripped between the two rollers. Then, a roller with sharp curved edges that look like knives impales the husk and the second roller presses the coconut to the other roller to remove the husk. *Continuous power-operated coconut husking machine* comprises of the inlet, de-husking component, separating component, and power component. The coconut is impaled at the inlet. As the coconut slides down, a blade punctures the husk in different parts of the nut rapidly. The coconuts that have been sheared completely enters into the separating component which is made up of two knurling rollers. Figure 2 shows the continuous power-operated coconut husking machine [5].

Fig. 2 Continuous
power-operated coconut
husking machine

The hydraulic coconut de-husking machine has a component used to support the coconut in an upright standing position. It consists of a hydraulic powered lifting component for raising the coconut. As the coconut is lifted, the arm is fed into the coconut which removes the husk. Figure 3 shows the hydraulic coconut de-husking machine [7].

The specific objective of the paper is to design and develop a coconut de-husking machine which can remove maximum coconut husk in less time. This will help overcome labor and accidents caused by using crude external devices.

The rest of this paper is structured as follows. Section 2 presents a review of related works, while Sect. 3 presents the proposed developmental process. The experimental results and discussion are presented in Sects. 4 and 5 concludes the paper.

Fig. 3 Hydraulic coconut
de-husking machine

2 Related Works

In [8], the authors carried out a study to develop a coconut de-husking machine consisting of two spiked rollers mounted on two shafts, powered by a 2 h.p electric gear motor with rated speed of 1500 rppm which was reduced to 21 rpm (ratio 70:1), using a belt and chain drive mechanism. The machine was developed to aid production in small scale. The objective of this study was to increase productivity by increasing its effectiveness. In methodology, it focused on the design concept generation, design calculation, the selection of pulley, and determination of speed and belt. The design generation tackles how Azmi et al. [8] can satisfy and meet customer needs through the use of 3d forms. Also, in the design calculation and selection of the pulley, authors in [8] determined the power needed to de-husk the coconut and in the selection of the pulley, it uses two v-belt mounted on the motor to manage the penetration of coconut. Overall, the study was successful for the small-scale farm owners and the machine was easy to operate and perform and also can save time, manpower, and labor.

In [9], the authors proposed a tool to de-husk a coconut, it needs human effort to de-husk and it is unsafe to operate and operate the KAU coconut husking tool, the pedestal with a base where on the top of it, it has a sharp object or known as stationary wedge/blade, the coconut will mount on the wedge, which needs to be pushed down on the hand lever to penetrate the coconut [10]. In methodology, it focused on the measurement of KAU coconut de-husking tool. Varghese and Jacob calculated the force amount to de-husk the coconut by using hand lever. Overall, the result of the study was successful and it was observed the de-husking of the coconut can be done manually with ease without damaging the coconut. Also, the machine was simple to construct that any steel fabricator can fabricate it.

Another study was conducted in 2016 by [6] "development of a new coconut de-husking and cutting machine". This study showed a new concept of de-husking and cutting a coconut on the same machine. In the methodology, they performed an experimental setup for the proposed machine comprising of one de-husking and one cutting unit. The coconut is de-husked using two spiked rollers while the cutting of the coconut is done using a strong sharp blade that looks like a knife positioned at an upright position. It needs human effort to penetrate the coconut to the knife. Overall, authors in [6] concluded that the de-husking unit study was more advanced and successful due to its performance, unlike to the cutting unit, which was unsafe to use and also it needs human effort to de-husk the coconut.

3 Methodology

The selection of material is based on some factors namely the accessibility of the materials, the quality of the materials for a specific working condition in service, and the value of the materials. The model consists of two rollers with blades parallel to each other. These rollers were linked to the gears for power transmission purposes. The blades thus will be utilized to remove the outermost portion of the nut. The rollers are linked to the shaft by using mounting them with the bearing and gear. As the rollers rotate in opposite directions, the coconut is being de-husked. A motor with 5h.p with a speed of 1440 rpm will be appropriate for performing this operation. The process is safer as only one operator is required to operate this machine. This serves as an added advantage to this machine. The frame is the structure that holds all the components together. It was constructed using 4 mm angle iron made of mild steel. The 4 mm angle iron was cut to size (572 × 600 × 774 mm).

With the aid of four pillows bearing the shaft was mounted on the frame, at the end of both shafts spur gears were mounted to transmit equal rotation of both shafts in the opposite directions which will help in the compression of the exocarp. A three-phase 5 h.p induction motor with a speed of 1440 rpm is used to move the rollers. The shaft diameter Ø was 25 mm and length of 530 mm. With the dimension of a coconut, the distance between each shaft was measured to be 70 mm. The blade is used to remove the husk on the coconut. The material used was mild steel with a dimension of (30 × 90 mm), a thickness of 10 mm for the blades. The blades were placed at an angle of 60°.

3.1 Design Calculations

i. **Determination of husking speed shaft**: The husking speed shaft is the ratio between the velocity of the driver and driven for an open belt [7, 11, 12]. To calculate the shaft speed Eq. (1) is used.

$$V = \frac{N2}{N1} = \frac{T2}{T1} \tag{1}$$

where,
$N1$ = speed of rotation of the smaller sprocket in r.p.m = 1440 rpm,
$N2$ = speed of rotation of the larger sprocket in r.p.m,
$T1$ = number of teeth on the smaller sprocket = 127 mm, and
$T2$ = number of teeth on the larger sprocket = 127 mm,
V = velocity.
Putting values in Eq. (1), authors evaluated $N2$ as

$$N2 = \frac{1440 \times 127}{127} = 350 \text{ rpm}$$

Therefore, the de-husking speed of larger sprocket is 350 rpm.

ii. **Torque required to de-husk a coconut (T)**: The torque to be used in de-husking a coconut manually is given by Eq. (2).

$$T = \text{the force required in shearing} \times \text{perpendicular distance}$$
$$= F \times L \tag{2}$$

The average force for de-husking mature and dry coconut to be 450N and distance is 80 nm. So the obtained torque $= 450N \times 80$ nm $= 36$N/m.

iii. **Determination of the required power**: The power was evaluated using Eq. (3).

$$P = \frac{T \times N}{9.554} \tag{3}$$

where $T =$ torque, $N =$ revolutions per minute. Putting values of T and N, the required power is

$$P = \frac{36 \times 1440}{9.554} = 5425.99 \text{ W}$$

iv. **Determination of length of the belt**: Driven pulley $D1$ was 300 mm as per standard available pulleys corresponding to the size of our proposed machine dimensions. Driver pulley $D2$ will be 100 mm and the center distance which is $x = 550$ mm. Using Eq. (4), the length of the belt has been evaluated.

$$L = \frac{\pi}{2}(d1 + d2) + 2x + \frac{(d1 - d2)2}{4x}$$
$$L = \frac{\pi}{2}(250 + 100) + 2(550) + \frac{(300 - 76)2}{4(550)}$$
$$L = 549.78 + 1100 + 0.13 = 1649.91 \text{ mm}$$
$$= 1.65 \text{ m} \tag{4}$$

will be required length of the belt.

v. **Selection of motor**: A5hp electric motor is used to transmit the power to the rotating shafts through sprocket and gears where the speed of the motor is 1440 rpm and the velocity ratio $= 1:1$.

vi. **The shaft and bearing diameter**: The mild steel material properties are yield strength, $\sigma_y = 378$ MPa, ultimate strength, $\sigma_u = 585$ MPa, shear modulus, $G = 81$ Gpa. The working/allowable normal stress is calculated using Eq. (5).

$$\sigma = \frac{\sigma_y}{\text{FOS}}$$
$$\sigma = \frac{378}{2} = 189 \text{ N/mm}^2 \tag{5}$$

The working/allowable shear stress is calculated using Eq. (6).

$$\tau = 0.5\sigma$$
$$= 0.5 \times 189 = 94.5 \text{ N/mm}^2 \tag{6}$$

Torsion torsional is evaluated using Eq. (7).

$$\frac{T}{J} = \frac{G\theta}{L} = \frac{\tau}{r} \tag{7}$$

where Torque $T = 198$ Nm $= 198 \times 10^3$ Nmm. The diameter of the shaft is calculated using Eq. (8).

$$= \frac{198 \times 10^3}{\pi d^4/32} = \frac{94.5}{d/2}$$
$$d^3 = \frac{198 \times 10^3 \times 32}{94.5 \times 2\pi} = 10670.96 \text{ mm} \tag{8}$$
$$d = 22.0158 \text{ mm}$$

Choosing the standard available diameter of the shaft from the ISO standard; shaft diameter is 25 mm. Select a bearing number 205 with bore of 25 mm and a shaft diameter is 25 mm.

4   Results and Discussion

The performance of the coconut de-husking machine is done. Both the coconut de-husking machine and the manual de-husking mechanism were compared by using cutlass. Work rate and efficiency were used as evaluation parameters. The work rate is determined using Eq. (9).

$$\text{Work rate} = \frac{N}{T} \tag{9}$$

where $N =$ number of coconuts de-husked and $T =$ time taken to de-husk coconut manually. The efficiency of the coconut de-husking machine was determined by comparing the rate of manual coconut de-husking by laborers and the coconut de-husking machine, therefore, it was calculated using Eq. (10).

Table 1 Results for performance evaluation of coconut de-husking machine

S/N	No of coconut de-husked (N)	TM (S)	Rate (N/m)
1	1	29	0.0344
2	1	22	0.0454
3	1	27	0.0370
4	1	26	0.0384
Mean values	1	26	0.0384

Table 2 Results for performance evaluation of manual coconut de-husking machine

S/N	No. coconut de-husked (N)	T (s)	Rate (N/m)
1	1	234	0.00427
2	1	212	0.00471
Mean values	1	223	0.00448

$$\text{Efficiency}(\eta) = \frac{\text{work rate of the coconut de - husking machine}}{\text{work rate of the laborer}} \quad (10)$$

Tables 1 and 2 show the results of the experiment. The results from the tables were used to calculate the work rate and efficiency of the machine.

Table 1 gives the time it took to de-husk four coconuts. Four trials were conducted using the machine and two trials were done using the manual de-husking method. Based on the results obtained above, the coconut de-husking machine is better than the manual de-husking process. The test results illustrate that the machine is more effective in de-husking coconuts compared with manual de-husking. For instance, it took approximately 212 s to de-husk one coconut manually while it took about 22 s to de-husk the same number of coconuts using the machine.

The efficiency of the coconut de-husking machine is evaluated using Eq. (10) and was obtained.

$$\text{Efficiency } (\eta) = \frac{0.0454}{0.00471} = 9.6$$

Its efficiency shows that it is 9.6 times faster than the manual de-husking method. The fabricated coconut de-husking machine is approximately nine times faster than the manual de-husking method.

5 Conclusion

The coconut de-husking machine was fabricated and tested in the welding workshop using materials locally sourced. From the tests carried out, it was found that the coconut de-husking machine was 9.6 times faster than the manual de-husking method. The machine is easier to operate since no special skills are required. The user only needs to set the work in position and power on the machine. The machine is environmentally friendly due to the lubrication and bearing system used; therefore, reducing the vibrations and noise caused by the moving parts. The design and fabrication of the machine helped us appreciate engineering principles and understand the role of engineering technology in the development of the world. The successful implementation of the work also requires effective interpersonal skills such as teamwork, discipline, and effective communication between the fabricators.

References

1. Nair RV, JO, O, Ikuenobe CE (2003) Coconut nursery manual. Institute for Oil Palm Research, pp 3–22
2. Nwankwojike BN (2012) Development of a coconut Dehusking machine for rural small scale farm holders
3. John E, Galvan PD, Shirly E, Agcaoili O, Marvin E, Adorio D, Ruth M, Cabutaje AB, George E, Curammeng M (2018) design and development of a motorized coconut Dehusking machine. Int J Res Appl Sci Eng Technol (IJRASET) 887. www.ijraset.com
4. Ketan M, Tonpe K, Vinod M, Sakhare P, Sakhale DCN (2010) JETIR1402010. J Emerg Technol Innov Res (JETIR) 1(2). ISSN-2349-5162. www.jetir.org
5. R RC, Scholar P (2017) Design and development of coconut Dehusking machine. © 2017 IJEDR | 5. www.ijedr.org
6. Raj SP, Cheriyan J, Chandran V (2016) Development of a new coconut Dehusking and cutting machine. Int J Sci Eng Res 7(4). http://www.ijser.org
7. Khurmi RJ, Gupta JK (2005) A text book of machine design. Eurasia Publishing House
8. Azmi H, Sanuddin AB, Zakimi MZ, Jamali MS, Radhwan H, Khalil ANM, Akmal ANA, Annuar AF (2015) Design and development of a coconut dehusking machine (machine component design). J Adv Res Design 4:9–19
9. Vargheser A, Jacob J (2014) A review of coconut husking machine. Int J Adv Res Eng Technol (IJARET) 5(3):68–78
10. Navaneethan R, Prasath NN, Pradeep G, Santhana Prabhu K (2020) Review on coconut Dehusking and cutting machine. Int J Eng Appl Sci Technol 4. http://www.ijeast.com
11. Salau AO, Adaramola BA, Rominiyi O, Agabalajobi K, Srivastava P, Kumar N (2021) Development of a smart multi-level car parking system prototype. Lecture Notes in Mechanical Engineering. Springer, Singapore. https://doi.org/10.1007/978-981-15-9956-9_64
12. Salau AO, Deshpande DS, Adaramola BA, Habeebullah A (2021) Design and construction of a multipurpose solar-powered water purifier. In: Smart innovation, systems and technologies, vol 196. Springer, Singapore, pp 377–387. https://doi.org/10.1007/978-981-15-7062-9_37

Health Sector: An Overview of Various Smartphone Apps

Neera Batra, Amandeep Kaur, and Sonali Goyal

Abstract By enhancing connectivity, performance, and service quality, mobile healthcare apps are revolutionizing the healthcare system. A person could make his life easier and healthy by utilizing number of beneficial mobile applications meant for health care. This paper analyzes numerous mobile applications for the efficient monitoring and treatment of diseases in patients in diversified domains used in the healthcare sector. This study divides mobile applications into two categories: general and unique and uses 21 criteria to compare them to show the areas in which more research is needed to help enhance the quality of health care and help link users to providers of health care. To conclude, several points that will help to strengthen the healthcare system are recommended.

Keywords Health care · Smart phones · Mobile apps · Treatment

1 Introduction

Health is a greatest importance factor as it affects the economic development and the well-being of a country in case of developing as well as developed countries [1]. Previously, patients go to the hospitals for their treatment which was not convenient and increases cost. To access, provide quality and affordability in health care is a big challenge all around the globe. The cost factor of health care is a big issue for millions of users [2]. To overcome with these types of issues, mobile technology helps by utilizing number of health apps, numerous types of sensors, medical devices, telemedicine, etc. In everyday lives, a number of mobile apps have become pervasive, and companies like Google, Apple, etc. have developed apps for differently able users by spending too much amount on the resources in order to update the operating system of mobile [3]. Due to this technology, care delivery can be facilitated, helps to lower

N. Batra · A. Kaur · S. Goyal (✉)
Department of CSE, MMEC, Maharishi Markandeshwar (Deemed to Be University), Mullana, Haryana, India
e-mail: Sonaliguglani21@gmail.com

the cost factor, users are easily connected with their care providers, time management is there, patient monitoring and so on. Now, users get services related to medical field with their cell phones rather than going to the hospitals. In a survey, it is concluded that there are approximately 5 billion cell phones in the world, out of which around 1 billion are smart phones [4]. There are number of mobile applications now which are used for patients who are suffering from dementia by using GPS function of cell phone for location tracking of patients, secondly there are number of social media websites who are also helping the patients, and thirdly wearable sensors are used for health care. Mobile application is evolving as a fast developing segment of the Information and Communication Technology in healthcare sector at global level. A vast amount of mobile apps have been developed during the past few years in an attempt to help patients to avail health facilities and doctors to monitor and support their patients at anytime from anywhere. This systematic review aims to shed light into studies that have developed and evaluated mobile apps for the monitoring, management, treatment, or follow-up of some general or specific diseases in patients.

- **Disease-specific apps**

Adolescents require a wide range of abilities to fulfill their aim of greater independence. Some teenagers struggle to make this adjustment. Their transition to independence can be stressful and upsetting for parents and families. Some aspects of this difficult transition are typical, and while stressful, parents should not be concerned. The MTNA [5] app is designed for an adolescent to deal with these issues. Based on four primary inputs in the app, an app for diabetes patients [6] was built for intervention and clinical decision-making for type 2 diabetics and was examined for practicality and acceptance. Both patients and general practitioners in the primary healthcare network approved of the development and use of a mobile application.

The Jakpros [7] application for pregnant women offered its patients "educational page" as a feature to get reproductive health information such as prenatal care, cervical cancer prevention, high-risk pregnancy, and contraception knowledge from the reliable obstetrician and gynecologist doctors. It is also directly connected the doctor and their patients with the help of question and answers' forum. In Jakpros, the patient could detect the doctor schedule and check the nearest hospital for them. Same way, the doctor could monitor their patient's health with the help of saved information such as estimated date of delivery and baby heart rate.

To cater to the need of spina bifida patients, five apps that comprised the iMHere [8] gallery were released to support preventive self-care for managing medications, neurogenic bladder and bowel, mood, and skin breakdown.

Manage my condition [9] created three smartphone apps to assist parents in managing their child's asthma and to increase oncology patients' adherence to prescribed drugs. The first app allowed parents to learn more about their child's condition. The second and third programs provided an organized schedule of reminders,

data logging, and active case management based on real-time data to the healthcare team.

- **General apps**

A mobile application [10] to make health care more convenient for the masses was proposed which provided general information about hospitals, cost, quality, facilities of hospitals, online appointment with the doctor, emergency call for an ambulance or healthcare service, and medicine alert system.

An integrated eldercare HER system (IEEHR) [11] that merged health data with sensors and telehealth (vital signs) measurements provided physicians' tools for chronic disease management, reduced nursing workload, and allowed the development of health context aware algorithms for predictive health assessment.

In view of patient data security, a secure system app [12] used public-key encryption allowing the secure sharing of medical resources. Most of the people lack first-aid knowledge which sometimes becomes life threatening. Keeping in view some remote locations and non-availability of internet services, an offline app [13] was programmed to give users' first-aid instructions based on the symptom/injury that they selected. It provided visuals of the symptom/injury and gave instructions in local language for easier understanding by the users in the community.

One more general Android application [1] had 11 features such as patient registration, login, insert health condition, view health condition chart, request appointment with doctor, view appointment schedule, view medical treatments, view health article, view message from doctor, do emergency calls, and logout feature.

2 Literature Review

Imteaj and Hossain [10] developed a mobile app in order to provide a healthcare system which is highly effective. This app is providing a number of benefits to the users, for example: ease of finding information of any hospital in city, cabin-related information, booking of cabin by using various modes of payment, suggestions for finding appropriate hospital, alert system for medications, etc. This app is very useful for users as they are able to find or select hospital, and they can book cabin when needed and so on. This paper describes the architecture and logic diagram for cabin booking system in online manner, and in future, this work can be extended for poor people by using AI concepts for the detection of diseases.

Amini et al. [14] proposed a technique for determining the sensor position on human body to ensure correctness and accuracy of the measurements in health monitoring. This paper uses the technique of accelerometers in a way to record motion data so that the location of device on human body is estimated. It uses both methods of time-series analysis: supervised and unsupervised. The proposed technique firstly uses the unsupervised way for time interval detection, in which a person is walking and then SVM is used for pattern analysis and estimation of device position on body.

This paper [11] presents an idea for US older adults that they prefer to live in an independent manner for as long as possible until some dementia condition occurs. So, a system is developed for their health monitoring by using sensor networks along with telehealth and electronic health record as there is no integrated system present till now. This paper describes an integrated system for elderly people with number of advantages. In paper [12], the concept of cloud computing is presented along with the usage of distributed networks and the combination of both technologies is described in a pervasive application. This paper mainly focuses on security requirements of unlink ability by using private as well as public clouds, and with the usage of these both clouds, the data of health records can be efficiently managed. The previous work related to this idea has not defined any security approach in medical record systems. The proposed idea of this work is to create a system which is highly secure by using elliptic curve encryption as well as public key encryption method.

In paper [1], a health index of Indonesia is compared with some other countries which shows a low value in Indonesia due to complex patterns of disease. So, to overcome this problem, smart health concept is used which monitors the health condition of persons from healthcare institutes in order to prevent diseases. This paper describes that the researchers just focus on the collection of sensor data, infect the requirement is to store, process and even synchronize between the systems, institutes and patients. That is why a proposal is given for the development of health monitor system in the form of an app which will be used by the patients later.

This paper [9] is describing an idea to create customized cell phone application for the management of medical conditions so that there will be number of improvements for patients as well as healthcare systems. The only barrier in this target is the paradigm of software design in medical field. In this paper, three mobile apps are developed: first is Manage My Asthma, second is Manage My Medications, and last one is Manage My Patients; from these three apps, a new app is created, i.e., Manage My Condition, an application for the development of medical conditions.

This paper [6] has shown the scenario of glycemic control by using effective mobile technology and diabetic's management in a proper manner. According to this paper, if tools are used for this scenario, it will require the commitment between the patient and physician for success. This paper [15] describes the strategy for fragile patients for their customized treatment at home. In this, mHealth apps can be utilized for home consultation management by following software resources for the result improvement. This paper described a mobile App related to health which is representing architecture for integration support and task automation with different phases of treatment so that it will help physicians in all types of home visits.

In paper [16], a model is presented named as electronic medical records which are highly sensitive due to shared information between peers for up-to-date patient history. In this paper, a challenge exits if there is a need to provide security, privacy, and availability to the sensitive data. There proposes an approach based on blockchain technology to provide security to medical records. For this approach, encrypted medical records are kept, and on the other hand, patient uses decryption key with healthcare professionals in which he/she trusts.

This paper [17] defines the concept of Internet of Things (IoT), that it is mainly used in healthcare industry by utilizing number of sensors and actuators, and using wireless technology, by use of cloud computing and analysis of data, these all are converting the health care from a paid service which is case-based to value-based service which is referred as IoT-embedded healthcare applications. This paper uses a Diffusion of Innovations (DOI) model with a survey questionnaire for data collection.

3 Methodology

The need for systematic performance analysis of mobile applications in health care is increased significantly as this platform is widely used in smart mobile devices. Performance of mobile applications from user point of view is usually measured in terms of ease of use, availability, friendly interface, emergency help, medication help, hospital search and link facility, doctor link and availability, health record maintenance, secure patient data, and maximum possible support in case of physical absence of doctor and medical facilities. Many authors have studied Android platform, graphical user interface, memory usage, etc., as a base for evaluating performance of mobile application. For more detailed and integrated performance analysis, 22 parameters are used to evaluate the usability of a mobile application as expected by a patient and shown in a tabular form.

- **Mobile Application**

A mobile application is a software application designed to run on a mobile device such as a phone, tablet, or watch. To accomplish a comprehensive analysis in the mobile applications, we considered parameters from different categories:

- Doctor availability: How to search a specialized doctor, providing link to the doctor, remote patient monitoring.
- Hospital access: Hospital search based on emergency, treatment availability, distance.
- Data access: Patient medical record storage, privacy and integrity of data, security of data.
- Remote data accessibility, SMS/notifications, first-aid info, etc.

4 Results and Analysis

- **Data analysis**

Results are summarized in Table 1. To have a better understanding on the impact of the use of discrete parameters in the performance of a mobile application, we created a pie chart for each parameter for all the eleven mobile applications chosen for analysis. Values in Table 1 show that the most of the mobile applications succeeded

in providing only the basic facilities to the patients in the categories, general as well as specific. Moreover, whereas the security and privacy of the patient data, hospital cabin booking, exercise management, online consultation fee are concerned, very little work has been done in this regard as shown in Table 1.

Values in Table 1 show that only few mobile applications succeeded in providing the much demanded facilities to the patients.

On an average, 45% mobile applications provided the feature of SMS/notifications/alerts (45%) as shown in Fig. 1.

Other features like doctor's link (72%), medicine alarm (63.6%), health analytics module (54.5%), first-aid information (54%), search options for a specialized doctor (45%) were also provided by a considerable number of applications as shown in Figs. 2a and b.

Table 1 Summary report

Parameters	1 BG Medical Practitioner	2 IEEHR Eldercare	3 secure	4 first aid	5 MOOBLE	6 MTNA (Teenagers)	7 Diabetes	8 Jackprus (Pregnant women)	9 Mcare	10 iMHere (spina bifida)	11. (Manage mycondition)
Emergency call	√				√	√		√	√		
Cabin Booking	√										
Doctor search	√		√		√			√	√		
Doctor Link	√		√		√	√		√	√	√	√
Medicine Alarm	√	√		√	√	√				√	√
Hospitals Link	√		√		√				√		
Hospital Search	√							√			
Payment Ease	√										
Room Booking	√										
First Aid Info	√			√		√		√		√	√
Hospital Map	√		√		√						
App	√		√	√	√	√	√	√	√	√	√
Sensors		√					√				
Patient A. Monitoring						√		√		√	
nursing visit module records	√	√	√		√	√	√	√	√	√	√
Mood Scale		√						√		√	
Secure Data			√	√							
Diet Management						√	√				
Health Analytics module	√					√	√	√	√	√	√
Exercise management							√				
SMS/Notifications						√	√	√	√	√	
Disease	General	Specific	General	General	General	Specific	Specific	Specific	General	Specific	Asthma
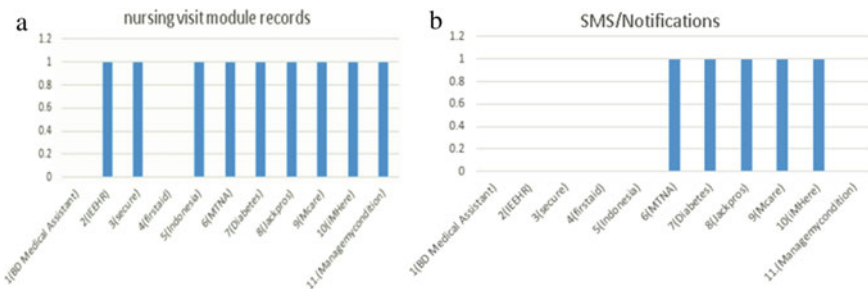
Fig. 1 **a** Nursing visit module feature in mobile applications, **b** SMS/notifications feature in mobile applications
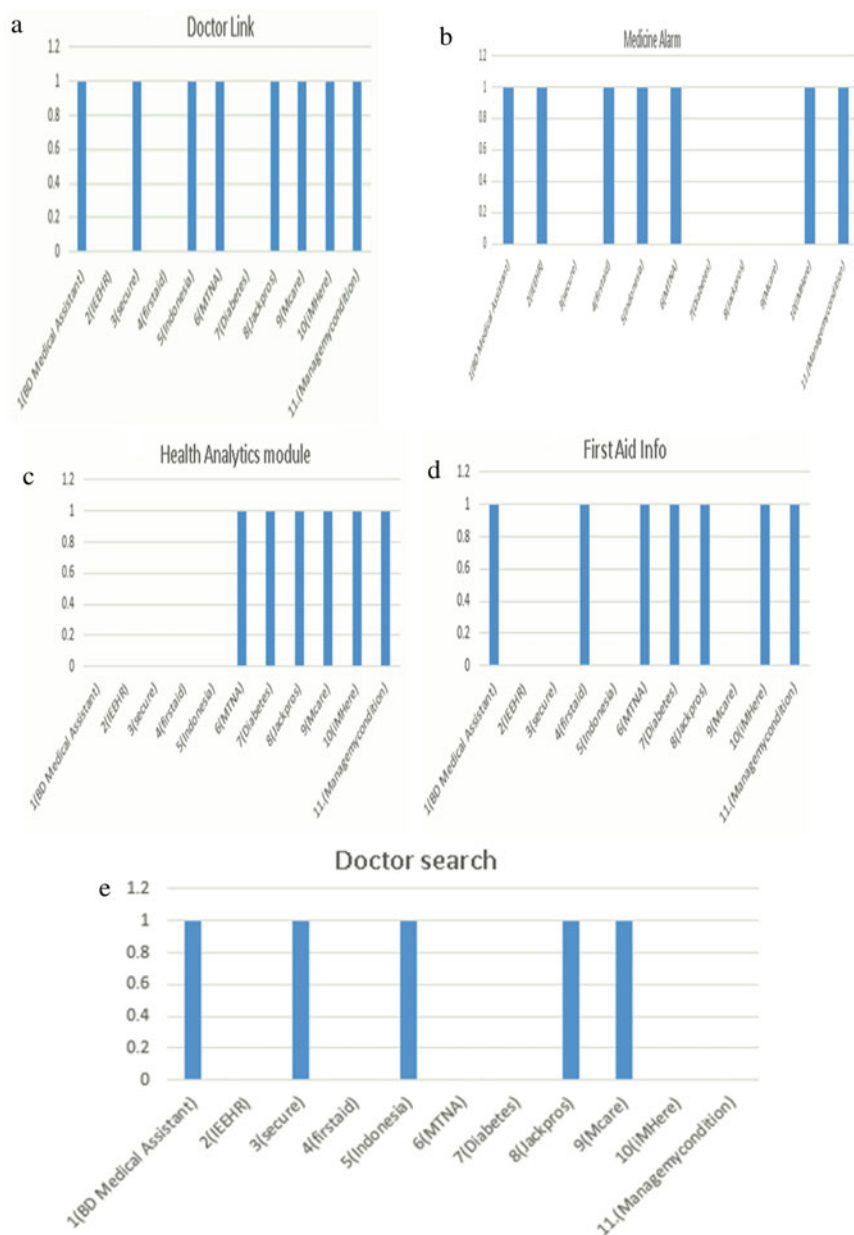
Fig. 2 **a** Doctor link feature in mobile applications, **b** medicine alarm feature in mobile applications, **c** health analytic feature in mobile applications, **d** first-aid info feature in mobile applications, **e** doctor search feature in mobile applications

But for most of the applications, there was a performance decay in terms of hospital link (36%), hospital map (27%), mood scale monitoring (27%), patient activity monitoring (27%), patient data security (18%), relevant hospital search (18%), diet management module (18%), sensor-based monitoring(18%), hospital room booking (9.0909%), exercise management module (only 9%), ease of payment (9%) that goes from slight (36%) mobile applications to very significant (only 9% mobile applications) as shown in Figs. 3a–k, respectively.

5   Conclusion and Future Work

The purpose of this study is to analyze mobile applications used by patients, doctors, and hospitals for health care in different domains aiming to provide remote health services in order to reduce the burden on hospitals and help patients and general public to maintain a healthy life. Based on the research that has been carried out, it has been concluded that:

- Most of the mobile apps under general category fulfill the basic requirements of patients and doctors, but scalability may become a bigger challenge when there is increase in number of patients.
- There is dearth of mobile applications which deal into specific diseases like heart failure, paralyses, skin problems, etc.
- Security is a major challenge in maintaining patient database. Very few mobile apps are dealing with issues of privacy and data integrity which is a basic necessity of the time.

In future, real-time analytics with its solution will be much valuable if considered [18–20].
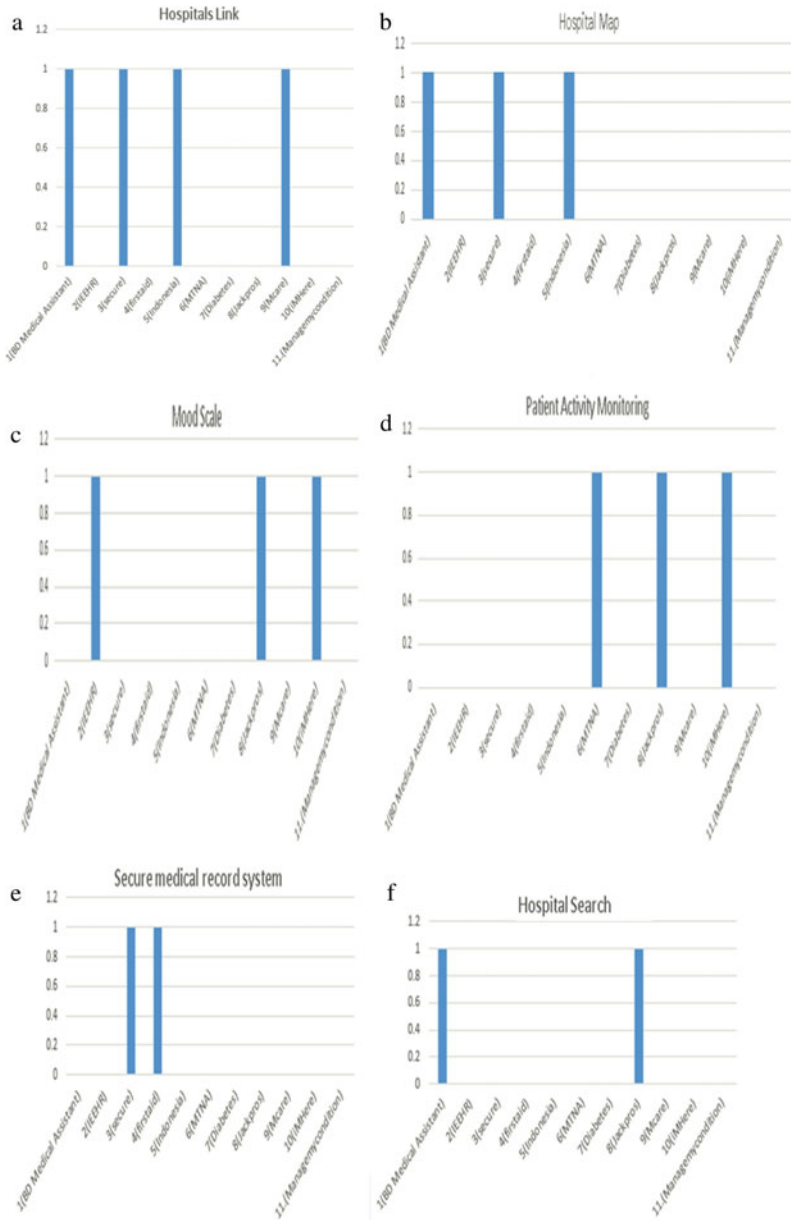
Fig. 3 **a** Hospital link feature in mobile applications, **b** hospital map feature in mobile applications, **c** mood scale feature in mobile applications, **d** patient activity monitoring feature in mobile applications, **e** secure medical record feature in mobile applications, **f** hospital search feature in mobile applications, **g** diet management feature in mobile applications, **h** sensor-based feature in mobile applications, **i** hospital room booking feature in mobile applications, **j** exercise management feature in mobile applications, **k** ease of payment feature in mobile applications
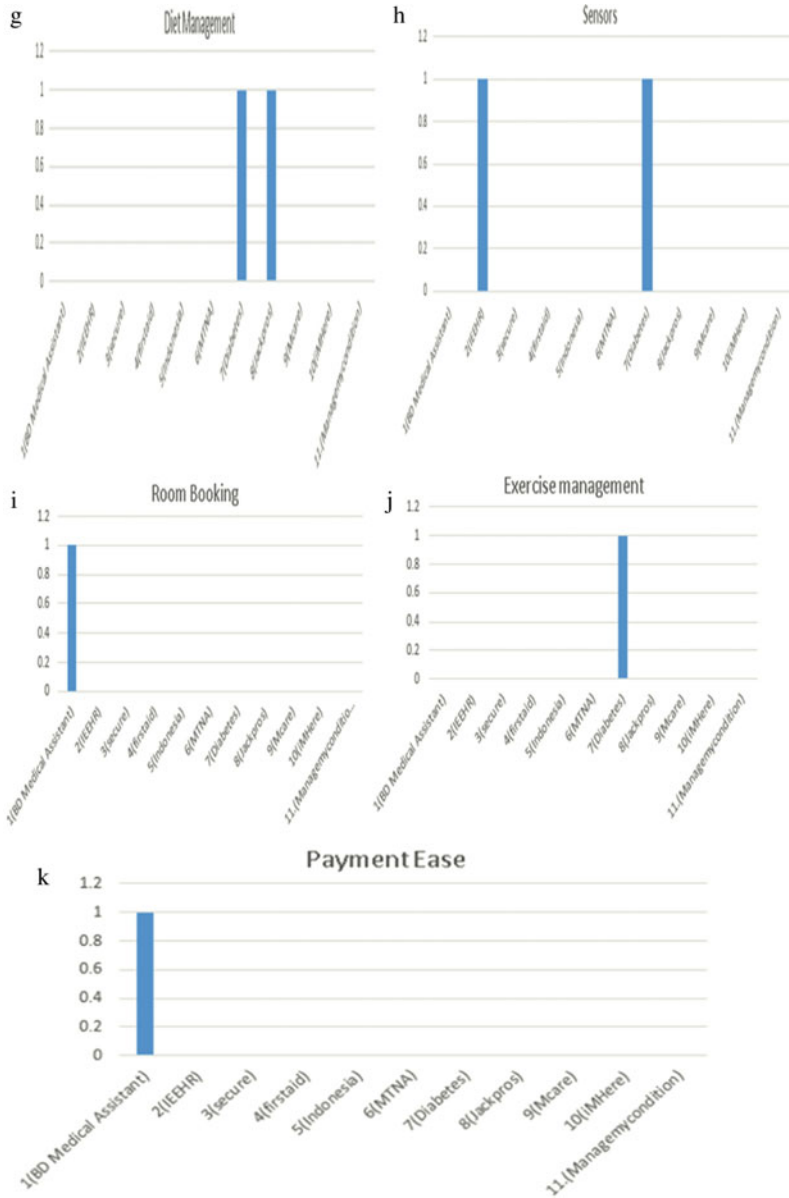
Fig. 3 (continued)

References

1. Zulkifi FY et al. (2018) Development of monitoring and health service information system to support smart health on android platform. IEEE
2. West DM (Oct. 2013) Improving healthcare through mobile medical devices and sensors
3. Kavitha R (2012) HealthCare industry in India. Int J Sci Res Publ 2(8)
4. Shankar M et al. (May 2015) Use of mobile technology in healthcare: case of Southern India
5. Dixon J et al. (2013) Designing, implementing and testing a mobile application to assist with pediatric to adult health care transition. Springer
6. Maryem B et al. (2020) Diabetic patients and physicians acceptability of a mobile health application for diabetes monitoring in Fez Region (Morocco). IEEE
7. Wiweko B et al. (2018) Jakpros: reproductive health education application for pregnant women. ICACSIS
8. Yu DX et al. (2014) Accessibility of iMHere smartphone apps for self-care. In: IEEE international conference on healthcare informatics
9. Tapparello C et al. (2016) Developing medical condition management applications using manage my condition. In: IEEE first conference on connected health: applications, systems and engineering technologies
10. Imteaj A, Hossain MK (2016) A smartphone based application to improve the healthcare system of Bangladesh. IEEE
11. Popescu M, Chronis G, Ohol R, Skubic M, Rantz M (2011) An eldercare electronic health record system for predictive health assessment. In: IEEE international conference on e-health networking, applications and services
12. Chen CL, Hunag PT et al. (2020) A secure electronic medical record authorizaztion system for smart device application in cloud computing environments. Springer
13. Garces AFC, Lojo JPO (2019) Developing an offline mobile application with health condition care first aid instruction for appropriateness of medical treatment. ISEC
14. Amini N, Sarrafzadeh M, Vahdatpour A, Xu W (2011) Accelerometer based on body sensor localization for health and medical monitoring applications. Pervasive Mob Comput
15. Araujo V et al. (2015) A health care mobile application and architecture to support and automate in–home consultation. In: IEEE international symposium on computer based medical systems
16. de Oliveira MT et al. (2019) Towards a Blockchain based secure electronic medical record for healthcare applications. IEEE
17. Yuan YS, Cheah TC (2020) A study of IOT enabled healthcare acceptance in Malaysia. J Critic Rev 7(3)
18. ArulanandaJothi S et al. (2019) A study on IOT and data analytics in healthcare systems. Int J Sci Technol Res 8(10)
19. HaqWani NUI et al (2013) Health system in India: opportunities and challenges for enhancements. IOSR J Bus Manag 9(2):74–82
20. Kiram M (2020) An analysis of healthcare infrastructure in India. Indian J Appl Res 10(2)

Survey on IoT Based Secure Health Care Framework Using Block Chain Technology

Mankiran Kaur, Puneet Kumar, and Anuj Kumar Gupta

Abstract The Internet of Things (IoT) is becoming increasingly popular in today's world, with applications in a wide range of disciplines, including healthcare. Internet of things-based healthcare solutions have benefited the healthcare industry by utilizing wearable and mobile devices. This results in the most accurate and timely diagnosis of healthcare data. Smart healthcare systems, on the other hand, are extremely susceptible to a variety of security vulnerabilities and malicious assaults such as data loss, tampering, and forging. Recent developments in the block chain technology in combination with IoT offer a viable defense against these flaws and challenges. Combining both the techniques provides better security, hence enhancing performance. This paper focuses on many challenges and the traditional security procedures used in the healthcare domain. Advantages of collaborating block chain with IoT have also been elaborated.

Keywords Block chain · Smart healthcare · Security · Privacy

1 Introduction

The Internet of Things (IoT) is a network of intelligent physical items known as "things" that connects humans. The Internet of Things (IoT) is the ability of any "thing" that connects and interacts, transforming the physical world into a massive information system. Many technologies are increasingly becoming an intrinsic aspect

M. Kaur (✉)
Chandigarh University/CGC Landran, Mohali, India
e-mail: mankiran.3321@cgc.edu.in

P. Kumar
CSEChandigarh University, Gharuan, India
e-mail: puneet.e11454@cumail.in

A. K. Gupta
CSE CGC Landran, Mohali, India
e-mail: anuj.coecse@cgc.edu.in

of IoT, from cloud computing and machine learning to data analysis and information modeling. The Internet of Things is allowing for the creation of new commercial techniques. The Internet of Things can do the following tasks: manage decentralized structures, use of divergent devices using IoT, keep track of different IoT records, and manage network diversity. All of these tasks lead to the issues such as IoT gadget heterogeneity, negative interoperability, IoT device resource restrictions, and privacy and security risks.

The layers that make up the internet of things are as follows:

- Perception Layer: Sensors, actuators, controllers, bar code readers, and other IoT devices are all available. These devices detect data in the surroundings and collect it.
- Communication Layer: IoT gateways and Wi-Fi access points can be used to connect a variety of wireless and wired devices, including sensors, RFIDs, actuators, controllers, and other tags, to create an industrial network (APs).
- Industrial Applications: Numerous industrial applications, including those in manufacturing, supply chain management, healthcare, and the food industry, can be supported by the Internet of Things (IoT).
- The Internet of Things (IoT) is essentially a network of physical things, or "things," that can be linked to the Internet and equipped with sensors that can collect data and communicate it to a centralized authority with little to no human involvement. We can say, The Internet of Things is a network of mechanical, digital, and networked computing objects, all of which have a unique identifier and the capacity to transmit data over a network without the need for direct human or computer contact [1]. Contacts with patients occur in medical settings [2]. The Internet of Healthcare Things and the Internet of Medical Things are two categories of IoT in healthcare [3]. The Internet of Healthcare Things device acts as a support link between patients receiving medical treatment far away and patients receiving healthcare information. The patient's condition is assessed through routine service monitoring, which also provides feedback. Interoperability, machine-to-machine connectivity, information exchange, and data transfer are all relevant to healthcare and are maintained in the Internet of Medical Things [4]. Data from IoT devices can help clinicians choose the best treatment technique for their patients and get the desired results [2]. Medical information, on the other hand, is vulnerable to data leakage and access issues [1] (Fig. 1).

2 Available Security Techniques

2.1 Confidentiality

Cryptographic techniques such as symmetric key algorithms like AES and DES ensure confidentiality. Hash functions preserve the integrity of the data (SHA-26, SHA-1, MD5, MD). Digital signatures and transfer of keys via asymmetric key

Fig. 1 Architecture of
layers of IOT
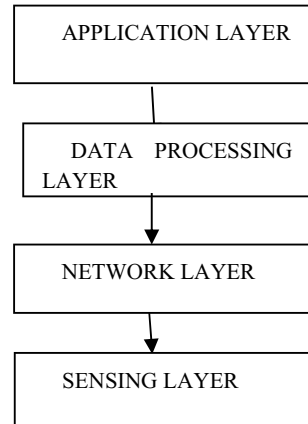
algorithms, insecure channels, Diffie-Hellman, elliptic curve cryptosystem, and the cryptography IC (elliptic curve cryptosystem) are attracting the greatest interest. Kakali et al. [5] suggested a Hellman Diffie curve authentication mechanism. This approach addresses, confidentiality, access control, and key management in wireless sensor networks. The authors compared their suggested scheme with various previously proposed ECDH schemes. In terms of attacks, two-way authentication security solution was presented by Kothmayr et al [6]. Datagram Transport Layer Security (DTLS) handshake is included in the scheme. It ensures that information is kept private and secure. A variable cypher security certificate was introduced by Wen et al. in [7]. This is applicable to the Internet of things sensor layer protected by IEEE 802.15.4 [8] using a common compliance architecture. Confidentiality of data, integrity of data, and a lightweight solution, and key management techniques all are provided by this framework [9]. Lightweight RFID mutual authentication protocol with cache in the Reader (LRMAPC) was proposed by Fan et al. for IoT in [10]. Lightweight protocol is used for authentication of large number of tags. IBE was proposed by Chen [11]. (Identity based encryption): It enables authentication in addition to quick and energy-efficient encryption. A mechanism called id-based multiple authentications (IMA) was proposed by Peng [12]. This technique is needed for protection of wireless sensor networks from node replication, data aggregation and enhances authentication determined to be unfit to handle dispersed systems such as IoT. After that, Cap BAC is presented. This protocol allows for rights delegation as well as more advanced access control measures. Oualha and others [13] suggested a cipher text policy-attribute based encryption with extended cipher text (CP-ABE). Because of their limited computation, IoT gadgets, e.g., sensors, actuators cannot be employed as CP-ABE enforcement points B.

2.2   Access Control

Access control is ensured through attribute based encryption (ABE) and role based Encryption (RBE). These two works are combined in the planned works. Gusmeroli and his colleagues. Scalability, manageability and efficiency of these schemes [14] are shown. As a result, they were determined to be unfit to handle dispersed systems such as IoT. After that, capability based access control (Cap BAC) is presented. This architecture allows for rights delegation as well as more advanced access control measures.

2.3   Privacy

Hosseinzadeh et al. [15] advocated that application interface (API) and the operating system utilized in IoT be obfuscated/diversified. It will prevent an attacker from exploiting the system for their advantage. However, this obfuscation/diversification increases computational costs and memory usage. The privacy-preserving techniques were used to a home automation test bed by Schurgot et al. in [16]. Emphasis was on maintaining anonymity through the use of information manipulation and cryptography. Ullah et al. [17] in order to safeguard privacy in the Internet of Things, the semantic obfuscation technique (SOT) was refined. The suggested approach, ESOT, guarantees location privacy. Antonio et al. [18] developed a personal security approach for a medical system in an Internet of Things. There are recommendations made about how to address the security and privacy issues. These solutions were used to develop the security framework.

2.4   RFID Security Protocols

RFIDs are the foundation of the Internet of Things (IoT) technologies. Access control are some of the protocols used to ensure security in an RFID system. Any powerful encryption algorithm is not supported by EMAP. It gives protection from replay attacks as well as compromising resistance. ASRAC [19] prevents cloning of data, attacks from the man in the middle, confidentiality of user data, forward secrecy, compromising resistance, and replay attack.

2.5   Secure Routing

Communication security refers to keeping unauthorized persons from intercepting and accessing telecommunications without interfering with the intended recipients'

communication. In the transport layer, the link is encrypted using the TLS/SSL protocols. On the other hand, IPSec helps in protecting the network layer. Each tier of the security protocols is built to protect the authenticity, secrecy, and integrity of communications. The InDRes system was proposed by Surendar et al. [20] (IDS for IoT with 6LoWPAN). In this system constrained-based on specifications to identify sinkhole attacks. In [21], Raza and colleagues developed and launched a novel intrusion detection system for Internet of Things.

3 Challenges of Internet of Things

There are various issues associated with the broad variety of IoT devices, including

(1) Heterogeneity: The property or state of including distinct or diverse elements.
(2) Network complexity: Factors that influence complexity are the number of nodes, different paths in a computer network, the range of communication mediums, protocols, communications equipment, hardware, and software platforms.
(3) Resource constraints for IoT devices: The word "resource constraint" refers to the limitations on the inputs that can be used to complete a task, such as people, time, equipment, and supplies. In this case, accepting no more work than you can handle every week will keep your time and resource restrictions in check.
(4) Privacy concerns: In the Internet of Things (IoT) environment privacy is the major concern as anyone has the ability to connect over the Internet, logical entity which is connecting to the Internet has a unique identifier, so privacy issues are needed to be addressed.

4 Smart Healthcare Security Requirements

Validation, automatic information compilation, and discrimination are all part of the IoT concept in the medical field. IoT enabled smart healthcare systems to manage most of the private information and past history of patients. So, data is extremely vulnerable to various attacks so effective security methods are needed [22]. The resources which are available for smart healthcare devices are limited and have low processing and storage capabilities. This results in need to implement efficient security methods to overcome security issues [22, 23]. These mobile gadgets may also need public network connections, like those found in homes, companies, and hospitals, which put them in danger. Designing dynamic and reliable security techniques is getting more and more challenging as connected IoT device numbers increase quickly [24] 25, e.g.,, the development record of a patient's health state is private, and therefore requires a safeguarding approach to prevent the information from being disseminated to an unapproved group. In this manner, no one can see or alter the data, and no one can pass a tampered-with patient health record. It also minimizes the likelihood of a doctor making an error when communicating with patients. If

Table 1 Smart healthcare security requirements

Requirement	Description
Integrity	The reliability of health information obtained or given to authorized parties that hasn't been tampered with. Inaccurate diagnoses and potentially harmful effects can result from data that has been altered or falsified [27]
Confidentiality	Ensures that health information is kept secure and that unauthorized persons cannot access it. The healthcare sector in the IoT domain includes a wide range of linked devices, applications, and parties, resulting in data being stifled by incorrect diagnosis [23]
Privacy	Only those who have been given authorization have access to the health data and information. Without patients approval, data and information about them is never shared among third party. Furthermore, data security in transit or storage is ensured by privacy
Availability	Healthcare data must be given as needed for timely diagnosis and treatment [28]
Authenticity	The health data can only be accessed or changed by the authentic party [29], which refers to the asking entity's honesty
Non-repudiation	Neither the user nor the patient will be able to deny the information presented. It could be dealt with using digital signatures and encryption [30]
Auditing	Ascertains a healthcare application's overall dependability. It refers to the process of recording and updating all transactions [31]
Data freshness	It ensures that the data is accurate and current. It is necessary to guarantee data availability for this attribute to be timely. [32] Delays have a negative, even fatal, impact on the diagnosis\and survival of a patient
Anonymity	The process of concealing a patient's identifying from the general public unauthorized parties. It ensures that the information is stored in a way that prevents patient identification [33]
Secure data transit	It assures that data is safe and secure while in transit, and that it is not tampered with or monitored in any manner. It prevents an attacker from viewing or modifying data in transit [23]

safety standards are not followed, medical practitioners may give patients the wrong medication or provide insufficient care [26]. Changes to a blood examination report, for example, could aggravate patient's condition if incompatible blood is transfused during a transfusion. Table 1 summarizes the critical security requirements in the IoT-enabled healthcare space. Table 1 gives smart healthcare security requirements.

5 Introduction to Block Chain Technology

Block chain technology has been used extensively nowadays and wide research is going on this topic. Satoshi Nakamoto coined the term "block chain" to describe the technology that underpins the digital currency bit coin. Block chain technology provides a framework for eight between unknown and untrustworthy things, as well as

validating the widely distributed attributes of mobile (smart health) devices, without the requirement for a central security and authentication authority, as in existing cloud computing infrastructures. BLOCK CHAIN is a consensus-based, secure decentralized database that saves data immutably across a peer-to-peer (P2P) network (impossible to forge/alter a block and very easy to detect tampering, if any). It is a decentralized database that keeps data in the form of transactions in its most basic form. The data is unchangeable, extremely secure and encrypted [34, 35]. Because of the implications of block chain technology, no trusted third party is required to record or access data in blocks [35]. In healthcare, the block chain is utilized to enable access to medical data for a set period of time. The block chain technology has an important part in healthcare by simplifying data sharing with remote patients. The use of this block chain improves the efficiency of monitoring and lowers service costs. The block chain is a distributed ledger platform for connecting patient records and sharing services [36]. The value of the patient-requested services is derived from the healthcare center's block chain-related data. The block chain therefore keeps global services in the IoT and obtains them through deployment requests. If the first block chain is unavailable, the work is completed by the second block chain by exchanging/swapping block chains, the failure rate of healthcare services is lowered. The management of healthcare services is done in a reliable manner here. As a result, healthcare data is a ledger that acts as a knowledge base for a variety of patient data [37].

6 The Benefits of Combining Block Chain and IoT

1. Improved interoperability of IoT systems: Block chains can effectively increase the interoperability of IoT systems by converting and storing IoT data in them. Throughout this process, a variety of IoT data types are translated, examined, retrieved, compressed, and then stored in block chains. As block chains are built on top of a peer-to-peer overlay network that enables worldwide Internet access, their ability to move through several disjoint networks further exemplifies interoperability.
2. Better IoT system security: Since IoT data is stored as block chain transactions that are encrypted and digitally signed using cryptographic keys, block chains can safeguard IoT data (e.g., elliptic curve digital signature algorithm [36]). Additionally, by automatically updating IoT device firmware to patch up security holes, integrating IoT systems with block chain technologies (like smart contracts) can help to improve IoT system security [38].
3. IoT data reliability and traceability: Anywhere and at any moment, block chain data may be identified and validated. In the meantime, all historical transactions contained in block chains can be traced, e.g., [39] built a block chain-based product traceability system that provides traceable services to suppliers and merchants. The quality and authenticity of the products can be checked and validated in this way.

Fig. 2 Represents flowchart of block chain

```
          ┌──────────────────────┐
          │ TRANSACTION          │
          │ REQUESTED            │
          └──────────────────────┘
                     ⇩
          ┌──────────────────────┐
          │ BLOCK        CHAIN   │
          │ REPRESENTINGTHE      │
          └──────────────────────┘
                     ⇩
          ┌──────────────────────┐
          │ NODE   VALIDATES     │
          │ TRANSACTION          │
          └──────────────────────┘
                     ⇩
          ┌──────────────────────┐
          │ BLOCK   IS   SENT  TO│
          │ EVERY  NODE  IN  THE │
          └──────────────────────┘
                     ⇩
          ┌──────────────────────┐      ┌──────────────────────────────┐
          │ NODE   WILL    GET  A│ ───▶ │ BLOCK  IS  ADDED  TO  THE     │
          │ REWARD FOR THEPROOF  │      │ BLOCKCHAIN                    │
          │ OF WORK              │      └──────────────────────────────┘
          └──────────────────────┘
```
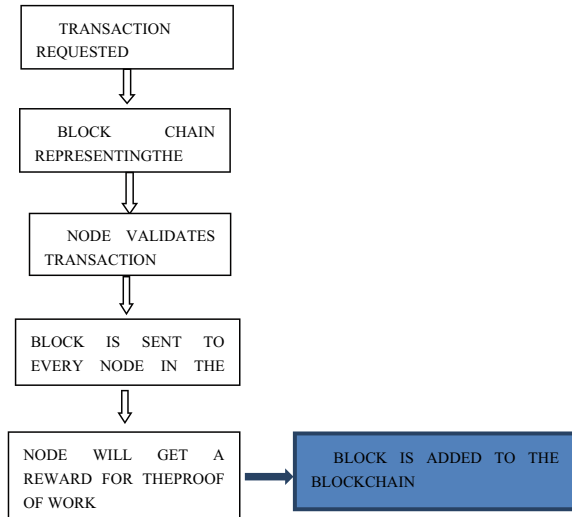
4. Automatic interactions of IoT systems: The block chain technique, for instance, supports the use of distributed autonomous corporations (DACs) to automate transactions that do not include conventional obligations such as governments or businesses. This allows IoT devices or subsystems to automatically link to one another. Smart contracts enable DACs to operate without the need for human intervention, resulting in cost savings (Fig. 2).

7 Literature Review

Several articles on the intersection between block chain and IoT have been published. This study, for example, [40] looks at how block chain technology can be used in the healthcare industry. Block chain-based smart contracts and an enterprise-distributed infrastructure can be used to monitor the patient's vital signs. Thanks to a comprehensive history log, patients can access their medical information at any time, from anywhere in the globe [41]. The focus of this study is on data security and privacy solutions that use block chain to avoid data abuse and corruption in huge datasets provided by the court, security, legislative, commercial code registries, and other sources [42]. This study shows how block chain technologies have taken into account multimedia data processing in IoT-healthcare systems, ensuring that patients' privacy and transparency are protected when each activity is collected by IoT devices inside the block chain [43]. This article provides an in-depth examination of BCoT and examines how block chain can be used in conjunction with IoT. The problems of the Internet of Things are briefly discussed. Brief overview of issues related to the protection of private information [44], such as bank transactions, voting systems, the Internet of Things, intelligent systems, and the sharing of personal information,

as well as a summary of technologies based on block chain for privacy protection has been discussed. To fulfill the consensus in the network method [45, 46], IoT applications must be lightweight. The authors devised a PoBT mechanism that requires significantly less compute for block validation and entry into the block chain network. Similarly, the authors have proposed lightweight scalable block chain (LSB) as a technique for IoT device scalability in a smart home utility. A block chain-based transaction validation method is suggested for a secure distributed IoT network in this study [47]. It contains a context-aware TX validation method that enables a miner to examine a TX in light of the importance of a service. In addition, we utilize the SDN-enabled gateway as a conduit between the IoT and the block chain network, which ensures a high level of network control and security. The core network was studied and compared with the proposed network model. The results show that prioritizing TX validation is more delay sensitive than the current network quality-of-service technique. Medical IoT is a collection of Internet-connected devices and software that connects to healthcare systems, according to this article [45]. Wearable devices and body sensors are used to track people's medical conditions. To be able to provide healthcare services, the collected data is processed, analyzed, and stored in cloud platforms. The data contains personal information such as the users' identify and location, as well as sensitive information such as mental health, drug addiction, sexual orientation, and genetics. As a result, maintaining a person's privacy is a huge concern for IoT service providers. The work's distinctiveness is significantly limited by the existing methodologies. The proposed system's keyless signature architecture for maintaining the secrecy of digital signatures also ensures features of authentication, as discussed in this paper [46]. Furthermore, the proposed block chain technology ensures data integrity. The performance of the proposed framework is evaluated by comparing features like average time, size, and cost of data storage and retrieval utilizing block chain technology to traditional data storage methods. According to the data, the proposed solution, which employs block chain technology, has a reaction time that is over 50% faster than traditional methods. They also claim that the system has a storage cost that is 20% less expensive. [48] The study discusses the role of a smart district model, which is an important step in the construction of a smart city using new technologies, as well as the role of an efficient energy management system integrated into a platform based on an IoT and BC approach for a repeatable and implementable model. Present a mechanism for assuring system security without the need for user intervention by continuously monitoring the user's genuine presence in a legal IoT-Zone. In an IoT context, every user interaction is saved in block chain as a transaction, and a user's IoT-trail is a collection of these transactions. For genuine user involvement, a one-of-a-kind digital crypto-token is necessary. This token is used as an identifier. Swarms of embedded devices introduce new privacy and security challenges are represented in [49]. According to us, permissioned block chains are an effective way to loosen and change those frameworks. When taking a long perspective of the block chain era, there are a number of requirements that are now unmet in existing block chain implementations. Permissioned block chains address the critical needs for scalability, agility, and incremental adoption as implemented by us [50]. Our permissioned block chain includes distributed identity management,

which allows for robust identity and attribute management for people and devices. [51] The goal of this research is to illustrate how to avoid these issues, the block chain era is being heralded as a game-changing technology that protects the control machine in real-time situations and provides anonymity. We used a block chain technique to extract data from IoT devices and store it in block chain to ensure transparency among several users in different regions. Furthermore, the suggested architecture has been tested against block chain's internal communication, which has allowed a number of intruders to obtain access to IoT devices. In person request time, falsification assault, black hollow assault, and probabilistic authentication situations, the findings were compared to the previous methodology and substantiated with stepped forward simulated consequences, which exhibited an 89% success rate. The paper suggests a smart district model [52], which is a necessary step in the creation of innovative products for smart cities, as well as the function of an effective energy management system integrated into a platform that is entirely based on IoT and BC method for a replicable and implementable version [53]. This study looks into the difficulties of combining IoT as well as block chain. An exploration of the potential benefits of block chain for the Internet of Things. Block chain IoT packages and platforms are available for the development of IoT applications. Possible topologies for that integration's topologies. The value of block chain nodes in IoT devices is assessed [54]. By implementing a variety of multimedia techniques, the healthcare system can employ multimedia to gather, process, and transmit patient and user-centered systems. However, managing massive volumes of data, such as reports and photos of every person requires more labor and poses a security risk. To address these issues, the Iot technology in healthcare improves patient care while cutting costs by allocating appropriate medical resources [55]. The proposed healthcare ecosystem incorporates a deep learning algorithm-based prediction model for diagnosing a patient's condition. The suggested framework's performance is evaluated in terms of security and prediction. The proposed framework's results reveal that it outperforms previous models in terms of security [56]. The study offers a secure and efficient health record transaction (SEHRTB) algorithm for transferring health record information between individuals, doctors, service providers, and institutions while maintaining privacy. Block chain technology is now being used in healthcare as a result of this endeavor. Patients may easily control and transfer their health records into cloud storage while keeping their privacy thanks to the efforts in healthcare [57]. The suggested method, as previously said, encrypts the user's health log data and context information records, and the side chain structure enables a significant amount of data to be kept in the knowledge base, assuring both tremendous expandability and security. It serves as a stand-in for the delayed and inefficient nature of block chain transactions [58]. The inhomogeneous Poisson point process, block header structure formulation, depth and height relationship, bloom filter mechanism, pay-to-public key hash (P2PKH) base stack process, and transaction management are a few of the crucial underlying arts models used in the study. Second, established models are implemented with the aid of certain crucial methods and techniques. The proposed scenario also analyzes and takes into account the block confirmation period, suitability for byzantine fault tolerance, smart contract rules, and SPV response. Finally,

the proposed scenario considers and examines the block confirmation time, byzantine fault tolerance suitability, smart contract policy, and SPV response [59]. The goal of this research is to provide an overview of block chain's application in healthcare systems. Not only can block chain be used to secure patient data in EHRs, but it can also be utilized to grant firms authorization when necessary. In this situation, hyper ledger can be used to control data access. To take use of the HER's decentralized structure, it can be dispersed among the network's many entities, such as doctors, medical practitioners, labs, and so on [36]. The goal of this research is to generate the most advanced deep learning-based secure block chain (ODLSB) enabled smart IoT and healthcare diagnosis model. Three critical processes are included in proposed approach: secure transaction, hash value encryption, and medical diagnosis. The orthogonal particle swarm optimization (OPSO) technique is used in the ODLSB strategy for secret sharing of medical pictures (Table 2).

Table 2 Depicts comparison of various techniques of block chain with IOT

S. no	Author	Method	Pros	Cons
1	Arul et al. [52]	Block chain-reliant adaptable	This of the aid	As the magnitude of the data grows, so does communication span
2	Rathee et al. [54]	A block chain-based hybrid framework for data processing in IoT-healthcare	The success rate of the product drop ratio, falsification attack, wormhole attack, probabilistic authentication scenarios higher	Biggest disadvantage of this method is the long transaction time and expense
3	Shanthapriya and Vaithianathan [57]	Block-health net (security)	With revocation communication	The most significant downside of this strategy is the lengthy transaction time and associated costs
4	Mubarakali et al. [58]	SEHR TB (secure and efficient health record transaction) the methodology ensures secure and efficient health dealings	Reduces latency, increases throughput	This model is not supported by an open-source platform
5	Kyung yong et al. [56]	For the management of health log data, a knowledge-based block chain network is used	High accuracy and repeatability are achieved	As the block formation shortens, performance

8 Conclusion

Security threats and issues confront IoT-enabled smart healthcare systems. It is critical to understand the security requirements of such systems in order to mitigate these dangers and difficulties [50]. Due to the scalability, greater cost, single point failure and resource constraints nature of IoT devices, standard security solutions are insufficient to address all of the security needs of an IoT-enabled smart healthcare system [60]. Block chain technology has recently ushered in a new era of privacy and security in healthcare. We examined at the security needs of IoT-enabled smart healthcare systems as well as the use of block chain-based security solutions in this study. We debated how block chain-based solutions can be used to solve a variety of safety problems in a cost effective, distributed, and scalable manner. We also debated infrastructural issues.

References

1. Hossain MS, Rahman M, Sarker MT, Haque ME, Jahid A (2019) A smart IoT based system for monitoring and controlling the sub-station equipment. Internet of Things 7:100085
2. Chethana C (2021) Determination of efficient and secured healthcare monitoring based on IoT along withmachine learning. New Visions Sci Technol 3:144–149
3. Ngan RT, Ali M, Fujita H, Abdel-Basset M, Giang NL, Manogaran G, Priyan MK (2019) A new representation of intuitionistic fuzzy systems and their applications in critical decision making. IEEE Intell Syst 35(1):6–17
4. Malik H, Alam MM, Le Moullec Y, Kuusik A (2018) NarrowBand-IoT performance analysis for healthcare applications. Proc Comput Sci 130:1077–1083
5. Chatterjee K, De A, Gupta D (Jan 2013) Mutual authentication protocol using hyperelliptic curve cryptosystem in constrained devices. Int J Netw Secur 15(1):9–15
6. Kothmayr T, Schmitt C, Hu W, Brunig M, Carle G (2012) A DTLS based end-to-end security architecture for the internet of things with two-way authentication. IEEE
7. Wen Q, Dong X, Zhang R (2012) Application of dynamic variable cipher security certificate in internet of things. In: Proceedings of IEEE CCIS
8. Sciancalepore S, Piro G, Vogli E, Boggia G, Grieco LA (2014) On securing IEEE 802.15. 4 networks through a standard compliant framework
9. Swetha MS, Pushpa SK, Muneshwara MS, Manjunath TN (2020) Blockchain enabled secure healthcare systems. In: Proceedings of IEEE international conference on machine learning and applied network technologies (ICMLANT), pp 1–6
10. Fan K, Gong Y, Du Z, Li H, Yang Y (2015) RFID secure application revocation for IoT in 5G. In: 2015 IEEE Trustcom/BigDataSE/ISPA, pp 175–181
11. Chen W (2012) An Ibe-based security scheme on internet of things. 2012 IEEE
12. Peng S (2012) An Id-based multiple authentication scheme against attacks in wireless sensor networks. In: Proceedings of IEEE CCIS 2012, pp 1042–1045
13. Suresh M, Neema M (2016) Hardware implementation of blowfish algorithm for the secure data transmission in Internet of Things. Global colloquium in recent advancement and effectual researches in engineering, science and technology (RAEREST 2016). Proc Technol 25:248–255
14. Gusmeroli S, Piccione S, Rotondin D (2013) A capability based security approach to manage access control in the Internet of Things. Math Comput Model 58:1189–1205
15. Hosseinzadeh S, Rauti R, Hyrynsalmi S, Leppänen V (2015) Security in the internet of things through obfuscation and diversification. 2015 IEEE

16. Schurgot MR, Shinberg DA, Greenwald LG (2015) Experiments with security and privacy in IoT Networks. IEEE, pp 1–6
17. Ullah I, Shah MA (2016) A novel model for preserving location privacy in internet of things. IEEE Conference Publications, pp 542–547
18. Antonio JJ, Zamora MA, Skarmeta AFG (2010) An architecture based on Internet of Things to support mobility and security in medical environments. In: IEEE CCNC 2010 proceedings
19. Butun I, Kantarci B, Erol-Kantarci M () Anomaly detection and privacy preservation in cloud-centric internet of things. In: IEEE ICC 2015—workshop on security and privacy for internet of things and cyber-physical systems, pp 2610–2615
20. Surendar M, Umamakeswari A In DReS: an intrusion detection and response system for internet of things with 6Lo WP AN. In: IEEE WiSPNET 2016 conference, pp 1903–1908
21. Raza S, Wallgren L, Voigt T (2013) SVELTE: Real-time intrusion detection in the Internet of Things. Ad Hoc Netw 11:2661–2674
22. Maamar Z, Kajan E, Asim M, Baker Shamsa T (2019) Open challenges in vetting the internet-of-things. Internet Technol Lett 2(5):e129
23. Tariq N, Asim M, Maamar Z, Farooqi MZ, Faci N Baker T (2019) A mobile code-driven trust mechanism for detecting internal attack sin sensor node-powered iot. J Parallel Distrib Comput 134:198–206
24. Abbas N, Asim M, Tariq N, Baker T, Abbas S (2019) A mechanism for securing IoT enabled applications at the fog layer. J Sens Actu Netw 8(1)
25. Tariq N, Asim M, Al-Obeidat F, Farooqi MZ, Baker T, Hammoudeh M, Ghafir I (2019) The security of bigdata in fog-enabled IoT applications including blockchain: a survey. Sensors 19(8):1788
26. Tariq N, Asim M, Khan FA (2019) Securing scada-based critical infrastructures: challenges and open issues. Proc Comput Sci 155:612–617
27. Khan FA, Haldar NAH, Ali A, Iftikhar M, Zia TA, Zomaya AY (2017) A continuous change detection mechanism to identify anomalies in ecg signals for wban-based healthcare environments. IEEE Access 5:13531–13544
28. Wang H, Li K, Ota K, Shen J (2016) Remote data integrity checking and sharing in cloud-based health internet of things, IEICE Trans Inf Syst 99(8):1966–1973
29. Strielkina A, Kharchenko V, Uzun D (2018) Availability models for healthcare iot systems: classification and research considering attacks on vulnerabilities. In: 9th International conference on dependable systems, services and technologies (DESSERT). IEEE, pp 58–62
30. Aghili SF, Mala H, Shojafar M, PerisLopez P (2019) Laco: Light weight three-factor authentication, access control and ownership transfer scheme for e-health systems in iot. Future Gener Comput Syst 96:410–424
31. Al-Issa Y, Ottom MA, Tamrawi A (2019) ehealth cloud security challenges: a survey. J Healthcare Eng
32. Talal M, Zaidan A, Zaidan B, Albahri A, Alamoodi A, Albahri O, Alsalem M, Lim C, Tan KL, Shir W et al (2019) Smart home-based iot for real-time and secure remote health monitoring of triage and priority system using body sensors: multi-driven systematic review. J Med Syst 43(3):42
33. Jaigirdar FT, Rudolph C, Bain C (2019) Cani trust the data i see? a physician's concern on medical data in iot health architectures. In: Proceedings of the Australasian computer science week multi conference, pp 1–10
34. Dinh TTA, Liu R, Zhang M, Chen G, Ooi BC, Wang J (2018) Untangling blockchain: a data processing view of blockchain systems. IEEE Trans Knowl Data Eng 30(7):1366–1385
35. Göbel J, Keeler HP, Krzesinski AE, Taylor PG (2016) Bitcoin blockchain dynamics: the selfish-mine strategy in the presence of propagation delay. Perform Eval 104:23–41
36. Ray PP, Kumar N, Dash D (2020) BLWN: Blockchain-based lightweight simplified payment verification in IoT- assisted e-healthcare. IEEE Syst J 15(1):134–145
37. Mohammed MN, Desyansah SF, Al-Zubaidi S, Yusuf E (2020) An internet of things-based smart homes and healthcare monitoring and management system. J Phys Conf Ser 1450(1):012079

38. Arul R, Alroobaea R, Tariq U, Almulihi AH, Alharithi FS, Shoaib U (2021) IoT-enabled healthcare systems using blockchain-dependent adaptable services. Pers Ubiquit Comput 1–15
39. Christidis K, Devetsikiotis M (2016) Blockchains and smart contracts for the internet of things. IEEE Access 4:2292–2303
40. Zhang Y, Wen J (2015) An IoT electric business model based on the protocol of bitcoin. In: Proceedings of 18th international conference on intelligence in next generation networks (ICIN), pp 184–191
41. Wang H (2020) IoT based clinical sensor data management and transfer using blockchain technology. J ISMAC 02(03):154–159
42. Suma V (2019) Security and privacy mechanism using blockchain. J Ubiquit Comput Commun Technol (UCCT) 01(01):45–54
43. Rathee G, Sharma A, Saini H, Kumar R, Iqbal R (2019) A hybrid framework for multimedia data processing in IoT-healthcare using blockchain technology
44. Dai HN, Zheng Z, Zhang Y (2019) Blockchain for internet of things: a survey. IEEE, pp 2327–4662
45. Biswas S, Sharif K, Li F, Maharjan S, Mohanty SP, Wang Y (2019) Pobt: a light weight consensus algorithm for scalable iot business blockchain. IEEE Internet Things J
46. Dorri A, Kanhere SS, Jurdak R, Gauravaram P (2019) Lsb: a lightweight scalable blockchain for iot security and anonymity. J Parallel Distrib Comput 134:180–197
47. Ji H, Xu H (2019) A review of applying blockchain technology for privacy protection. In: International conference on innovative mobile and internet services in ubiquitous computing. Springer, Cham, pp 664–674
48. Hosen ASMS et al (2020) Blockchain-based transaction validation protocol for a secure distributed IoT network. IEEE Access 8:117266–117277. https://doi.org/10.1109/ACCESS.2020.3004486
49. Alamri B, Javed IT, Margaria T (2020) Preserving patients' privacy in medical IoT using blockchain. In: Katangur A, Lin SC, Wei J, Yang S, Zhang LJ (eds) Edge computing—EDGE 2020. EDGE 2020. Lecture Notes in Computer Science, vol 12407. Springer, Cham. https://doi.org/10.1007/978-3-030-59824-2_9
50. Kaur A, Gupta AK (2018) A framework on healthcare monitoring system. Recent Trends Sens Res Technol 5(3):5–9 May 2019
51. agasubramanian G, Sakthivel RK, Patan R et al. (2020) Securing e-health records using keyless signature infrastructure block chain technology in the cloud. Neural Comput Appl 32:639–647 .https://doi.org/10.1007/s00521-018-3915-1
52. Lazaroiu C, Roscia M (2017) Smart district through IoT and Blockchain. In: 2017 IEEE 6th international conference on renewable energy research and applications (ICRERA), pp 454–461. https://doi.org/10.1109/ICRERA.2017.8191102
53. Agrawal R et al. (2018) Continuous security in IoT using block chain. In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 6423–6427. 0.1109/ICASSP.2018.8462513
54. Reyna A, Martín C, Chen J, Soler E, Díaz M (2018) On blockchain and its integration with IoT. Challenges and opportunities. Future Gener Comput Syst 88:173–190. ISSN 0167-739X
55. Rathee G, Sharma A, Kumar R, Iqbal R (2019) A secure communicating things network framework for industrial IoT using block chain technology. Ad Hoc Netw 94:101933. ISSN 1570-8705
56. Mubarakali A, Bose SC, Srinivasan K, Elsir A, Elsier O (2019) Design a secure and efficient health record transaction utilizing block chain (SEHRTB) algorithm for health record transaction in block chain. J Ambient Intell Humanized Comput 1–9
57. Rathee G, Sharma A, Saini H, Kumar R, Iqbal R (2020) A hybrid framework for multimedia data processing in IoT-healthcare using blockchain technology. Multimedia Tools Appl 79(15):9711–9733
58. Shanthapriya R, Vaithianathan V (2020) Block-healthnet: security based healthcare system using block-chain technology. Secur J 1–19

59. Chung K, Jung H (2019) Knowledge-based block chain networks for health log data management mobile service. Pers Ubiquit Comput 1–9
60. Kaur A, Gupta AK (October 2019) Supervised linear estimator modeling (SLEMH) for health monitoring. Int J Eng Adv Technol (IJEAT) 9(1) ISSN: 2249-8958
61. Mohammed MN, Desyansah SF, Al-Zubaidi S, Yusuf E An internet of things-based smart homes and healthcare monitoring and management system. J Phys Conf Ser 1450
62. Gupta S, Sinha S, Bhushan B (Apr 2020) Emergence of blockchain technology: fundamentals, working and its various implementations. In: Proceedings of the International conference on innovative computing and communications (ICICC) 2020
63. Khan R, Tao X, Anjum A, Kanwal T, Khan A, Maple C et al (2020) θ-sensitivek-anonymity: ananonymization model for iot based electronic health records. Electronics 9(5):716
64. Nakamoto S (2008) Bitcoin: a peer-to-peer electronic cash system, https://bitcoin.org/bitcoin.pdf. [Online; accessed June 15 2020]
65. Lu Q, Xu X (2017) Adaptable blockchain-based systems: a case study for product traceability. IEEE Softw 34(6):21–27
66. Kravitz DW, Cooper J (2017) Securing user identity and transactions symbiotically: IoT meets blockchain. In: 2017 global internet of things summit (GIoTS), pp 1–6. https://doi.org/10.1109/GIOTS.2017.8016280
67. Swapna Y, Sirisha G, Mary GA, Aparna G, Ruthvik K, Suneela B (2021) Covid-19-based intelligent healthcare monitoring system using deep learning and Iot principles for pulmonary breathing problem patients. Int J Fut Gener Commun Netw 14(1):1289–1297
68. Kaur A, Gupta AK, Kaur H (May 2021) Benefits of E-health systems during COVID-19 pandemic. In: Manocha AK, Shruti J, Singh M, Sudip P (eds) Computational intelligence in healthcare. Springer, pp 175–190

A Methodology for Increasing Precision for the Efficient Implementation of Sigmoid Function

S. Suseendiran, P. Nirmal Kumar, and Deepa Jose

Abstract In this brief, a high-precision sigmoid function is proposed and compared with the novel approximated method. The novel approximated function equation is modified and respective architecture is proposed. The high-precision sigmoid function is proposed to be used in artificial neural networks as an activation function. Since the activation function is the heart of the neural network, the function is made to give less error when compared with the original sigmoid function. So, increasing precision of sigmoid function will increase the precision of neural networks with high rate.

1   Introduction

In neural or deep neural networks, the final result depends on the activation function. There are many activation functions such as tanh, sigmoid, exponential. These are nonlinear functions, and implementation of this function is difficult due to its complexity of exponent, multiplication, division. So basically, implementation becomes difficult if we are having complex multiplication and division and implementation will be easy if we are having adders, shifters. Here in general, modules are designed to reduce area, power, and delay. In computation functional activities like activation function error also plays a major role. Error of the activation function should be minimum to reduce the errors at the output of neural networks. Activation function is the heart of the neural network. So reducing the area or power or delay or

S. Suseendiran · P. N. Kumar
Department of Electronics and Communication Engineering College of Engineering, Anna University, Guindy, Chennai, Tamilnadu, India

D. Jose (✉)
Department of ECE, KCG College of Technology, Chennai, Tamilnadu 600097, India
e-mail: deepa.ece@kcgcollege.com

error in activation function will reduce the total area power delay and error and not much reduction in other neural modules are needed. Many approximation methods already exist to reduce the parameters efficiently.

In [1], the look-up-table (LUT)-based method will store the sampled values using LUTs or ROMs, and this method directly uses LUTs to access data using input values and is a straight-forward approach to implement nonlinear functions [1, 2]. For higher precision, this LUT-based method needs more LUTs to store the sampled values which lead to significant increase in area. Another commonly used approach is piecewise linear (PWL) approximations [3]. In PWL approximations, the concept is dividing nonlinear function into several segments and a linear function is derived using the segments such that the segments will fit the curve accordingly. The hardware resources used in this method are multiplier, adder, and LUTs. These multipliers will be responsible for larger area and delay for this PWL method.

The hardware complexity is simplified using PLAN method [4, 5]. The PLAN method [4, 5] uses six special segments and equation is derived from these segments to approximate the sigmoid function. The multiplication used in the equation is replaced by the shift operations. This kind of simplification will be having many equations which will not fit correctly in the curve of the sigmoid function, and due to this, the precision will be too less for this approximation. Some other approximations are piecewise second-order approximation [6] and Taylor's theorem [7] for the efficient implementation of sigmoid activation function. More hardware resources are required in this method compared to PWL. In [8], the sigmoid function uses exponential function which requires complex division operation. The computation latency and hardware complexity are higher in these circuits. Bit-level mapping method is presented in [9]. The bit-level mapping method uses combinational circuit for implementing the sigmoid function and this method is applicable only for the small range of inputs. A Novel Approximation Method [10] which is the existing method overcomes hardware complexity and proposes two modules which efficiently reduces delay, area, and error. As said in the paper, the module 1 will reduce delay, error, and area, whereas the module 2 will reduce the error further by maintaining the same delay as of module 1 with increase in area with respect to module 1. Here in this module, we are going to propose a method to develop the novel approximation scheme [10–12]. The proposed concept is one of the main methods that will reduce area and reduces error.

2 Existing Approximation

A. Architecture for existing model

The existing method contains the approximated equations which are derived with the help of Taylor series expansion. In that approximation, hardware optimization is also done to increase its speed by reducing the multipliers to adders. The hardware optimization has been present in [a]. The detailed view of the existing approximation

has been present in [b]. The existing approximation has compared the precision with LUT, PLAN. Here, we are going to compare the proposed method with the existing method on the basis of precision without any change in the parameter, i.e., area, delay, and power. The existing approximation is compared with new approximation by increasing the area. So, the approximation which is to be proposed later will not increase area, power, and delay but reduces the error efficiently by using the concept of Horner's Rule of Precision. The hardware architecture for the existing approximation is given in Fig. 1.

B. **Simulation and synthesis for existing model**

The simulation result for the existing architecture and the is shown in Fig. 2. The inputs of 12 bits are given as an input, and the required 12-bit floating part result is obtained for both existing results and proposed architecture. The A[11:0] is the input and s2 is the corresponding output for existing architecture.
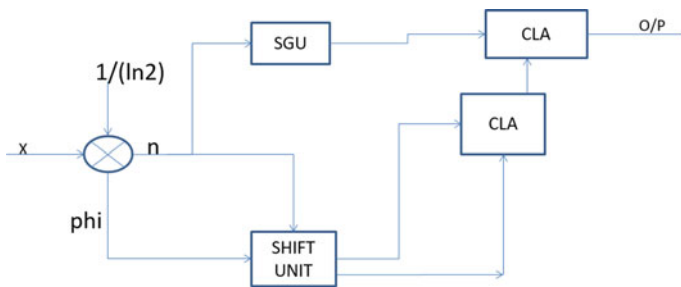
Fig. 1 Existing architecture
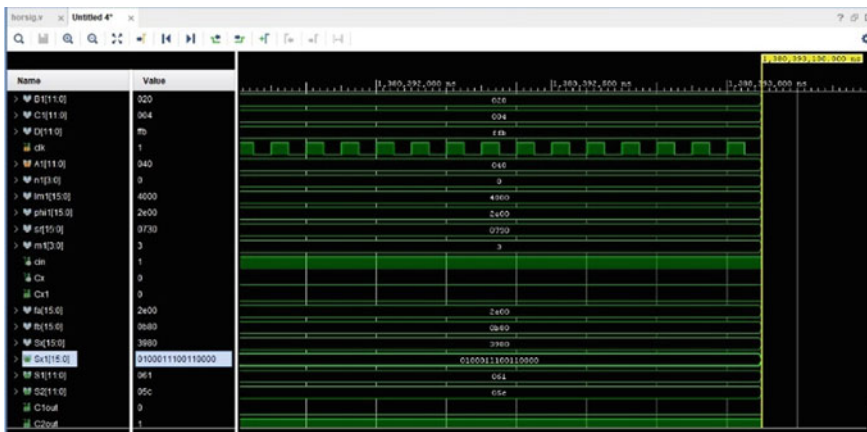
Fig. 2 Simulation result for existing model using Xilinx Vivado

3 Proposed Approximation

A. **Architecture for proposed model**

The equation of the existing approximation is given below:

$$f(\alpha) = \lambda + \phi \gg |m1| + \phi \gg |m2|. \tag{1}$$

Here in this equation, $m1$ and $m2$ are shift units and ϕ is shifted with respect to $m1$ and $m2$. This is the approximated sigmoid function equation. Here, we are going to optimize it further without increasing its adder and shift units. power of the sigmoid function. The concept can be explained with two examples given below:

Let us take 0.110011010101 as ϕ, $m1$ is -3 and $m2$ is -5. The λ for the given ϕ is 0.100000000000, so apply the above equation stated. The above equation binary addition for the $\phi \gg |m1| + \phi \gg |m2|$ part is as follows:

$$(0.110011010101 \gg 3) + (0.110011010101 \gg 5)$$
$$\geq (0.000110011010) + (0.000000110011)$$

The above one is the binary addition operation done on the approximated existing equation. Now, we can propose an approximated equation which we get from the existing equation. The existing equation is modified such that there will be no change in the number of adder and number of shifter. This will give no change in area, delay, and power, but the equation will give high precision by reducing the error when compared with the existing equation. The proposed equation is derived as follows:

$$f(\alpha) = \lambda + \phi \gg |m1| + \phi \gg |m2|, \tag{2}$$

$$= \lambda + (\phi + \phi \gg |m3|) \gg |m1|. \tag{3}$$

Consider the above final equation. From Table 1, we can see that $m1$ is smaller from $m2$. So taking $m1$ out and $m1$ with $m2$, we get $|m3|$. Further reduction in error will be explained by taking the same example. Here, $m1$ is -3 and $m3$ is -2 since $m3$ equals $-5 + 3$. Then, the procedure is as follows:

$$(0.110011010101 + 0.110011010101 \gg 2) \gg 3$$
$$\geq (1.000000001010) \gg 3$$
$$\geq 0.001000000001$$

The two results have difference in bits and not exactly the same. The proposed method will give the result nearly to the accurate value more than the result given by the existing method. This is because of the data loss in ϕ that is happening due to shifting of ϕ before addition. The original value of ϕ is shifted two times with value in the existing method. After shifting, the addition takes place. Since the shifting

Table 1 Error comparison between existing and proposed modules

Parameter	Existing model	Proposed model
Average error	0.0018	0.0011
Maximum error	0.0114	0.0106
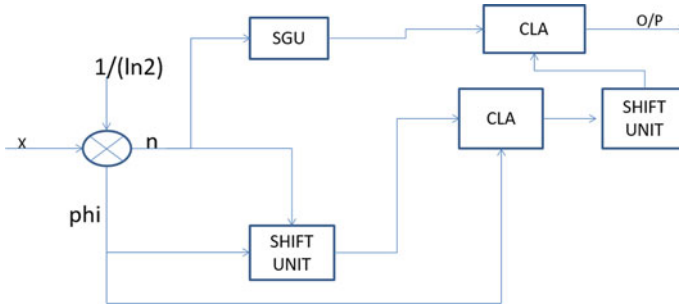
Fig. 3 Proposed architecture

process takes place first, the bit 1 might have been shifted out of the width, so in the addition process, the bits might be missing. But in the proposed method, only one time ϕ is shifted by less amount and added with the ϕ without shifting so that the out of width takes place only one time and added so that data loss in ϕ is reduced. This concept reduces the data loss so that error of proposed method will be lesser compared to that of existing method. In the proposed method, the adder used here was carry look ahead adder (CLA). This adder is the conventional fastest adder so that this adder has been proposed in the design. The following Fig. 3 is the schematic of the proposed design.

B. **Simulation result for proposed model**

The simulation result for the proposed architecture and the is shown in Fig. 4. The inputs of 12 bits are given as an input and the required 12-bit floating part result is obtained for both existing results and proposed architecture.

For the proposed architecture, the A1 is the input and sx1 is the corresponding output.

4 Experiment Results and Comparison

A. Metric of approximate precision

The precision of the modules is compared by calculating the average error and maximum error of the proposed and existing sigmoid modules. The average error and proposed error are given as follows.
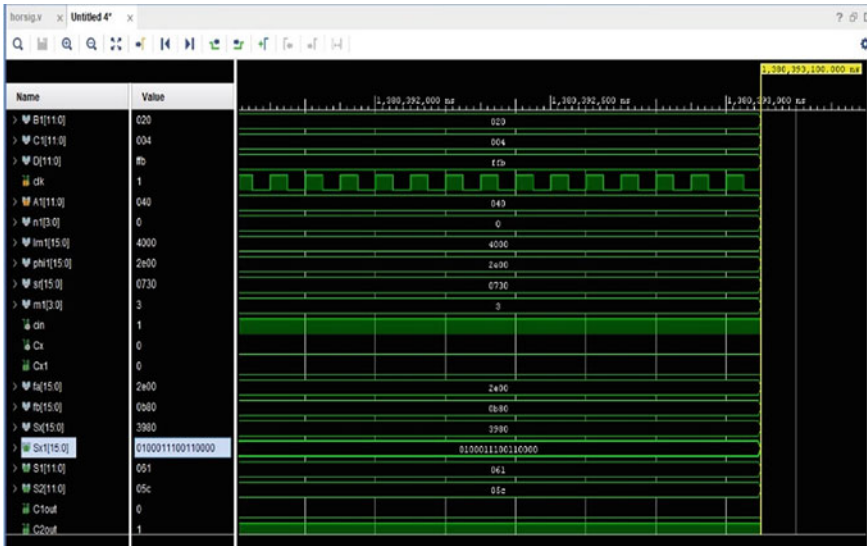
Fig. 4 Simulation result for proposed model

Here, in the above equation, g is the proposed function which is approximated from the original sigmoid function. In both g and f, equal samples at different instances are taken and subtracted to get the error at each instance. To find the average error, at each instance is added and divided by number of samples. To find the maximum error, take the maximum of the error samples.

B. Error analysis

The sampled error of the proposed and existing modules is plotted in MATLAB. The error is compared between the existing module and proposed module and the result is compared in Table 1.

The error analysis graph (Fig. 5) for the existing and proposed modules (Fig. 6) is plotted using MATLAB and the graph is shown below. The approximated errors are magnified by 50 × and plotted. From that plot, we have got the average error for the high-precision sigmoid function to be 0.0011 and maximum error to be 0.0106. For the existing novel approximation module, the average error from the graph equals 0.0018 and the maximum error equals 0.0114.

C. Comparison of synthesis results between different models

Table 2 shows the comparison between proposed and existing methods for plan method, existing method, and proposed method. The comparison shows that there is slight decrease in area when compared with both the methods. The delay remains same as existing method. The area reduction is due to the reduction in number of shift units. In proposed method, the number of shift units are x and y. Here, y can be

Fig. 5 Error graph for existing module

0.0114
0.0018
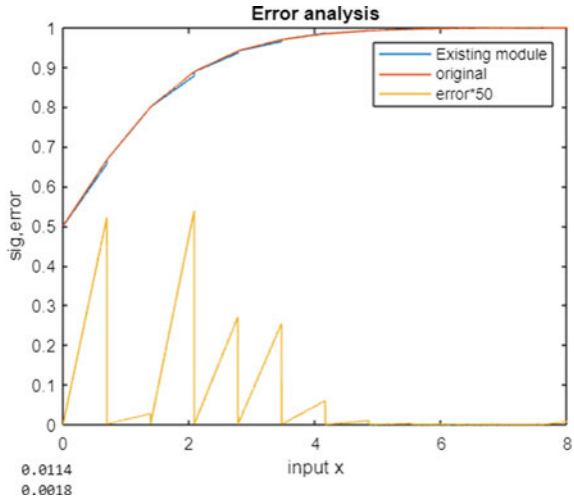
Fig. 6 Error graph for proposed module

0.0106
0.0011
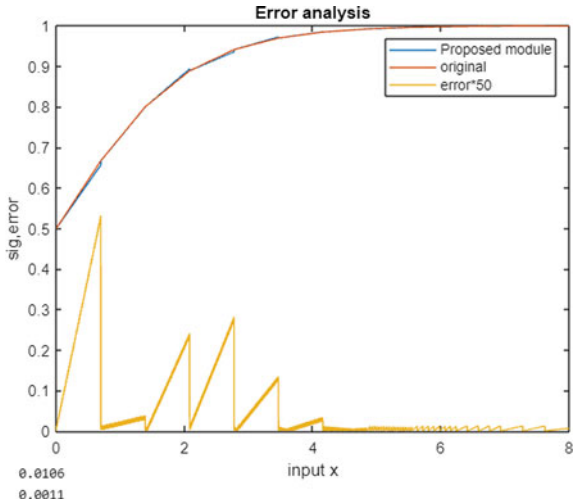
written as $x + a$, where a is lesser than x. If we take x out, the shift units are x and a only. This clearly tells that $x + a$ is reduced to a. Therefore, shifter area got reduced by x shifts.

Table 2 Area, delay, power comparison implemented on FPGA

Parameter	Existing model	Proposed model
Area (gate counts)	466	451
Delay (ns)	0.98	0.98
Power (uw)	519.52	519.52

D.   Analysis of results

The area and error of the proposed module are decreased compared with the existing module. The decrease in area is 3.21% with respect to the existing module, and the decrease in average error is 38.88% with respect to the existing module. This concludes that the proposed module is efficient than the existing module.

5   Conclusion

In this paper a high-precision sigmoid module and compared its precision with novel approximated sigmoid function is presented. The error graph for the high-precision sigmoid function and novel approximated sigmoid function is plotted and analyzed with respect to precision parameter. The error analysis results clearly show that the precision of the proposed module is higher than the precision of the existing module. The precision is improved by 38.88% with respect to the existing module. So, we can conclude that the proposed module has higher precision than existing module. Further, we can implement this module in neural networks to get the better results at the output.

References

1. Leboeuf K, Namin AH, Muscedere R, Wu H, Ahmadi M (Nov 2008) High speed vlsi implementation of the hyperbolic tangent sigmoid function. In: Proceedings of international symposium ICCIT, vol 1, pp 1070–1073
2. Piazza F, Uncini A, Zenobi M (Oct 1993) Neural networks with digital lut activation functions. In: Proceedings of international symposium IJCNN, vol 2, pp 1401–1404
3. Sun H, Luo Y, Ha Y, Shi Y, Gao Y, Shen Q, Pan H (2020) A universal method of linear approximation with controllable error for the efficient implementation of transcendental functions. IEEE Trans Circuits Syst I 67(1):177–188
4. Amin H, Curtis KM, Hayes-Gill BR (1997) Piecewise linear approximation applied to nonlinear function of a neural network. IEE Proc Circuits Devices Syst 144(6):313–317
5. Gaikwad NB, Tiwari V, Keskar A, Shivaprakash NC (2019) Efficient fpga implementation of multilayer perceptron for real-time human activity classification. IEEE Access 7:26 696–26 706
6. Zhang M, Vassiliadis S, Delgado-Frias JG (1996) Sigmoid generators for neural computing using piecewise approximations. IEEE Trans Comput 45(9):1045–1049
7. Campo ID, Finker R, Echanobe J, Basterretxea K (2013) Controlled accuracy approximation of sigmoid function for efficient fpga-based implementation of artificial neurons. Electron Lett 49(25):1598–1600
8. Gomar S, Mirhassani M, Ahmadi M (Nov 2016) Precise digital implementations of hyperbolic tanh and sigmoid function. In: Proceedings conference rec. asilomar conference signals systems and computers ACSSC, pp 1586–1589
9. Tommiska MT (2003) Efficient digital implementation of the sigmoid function for reprogrammable logic. IEE Proc Comput Digital Tech 150(6):403–411

10. Beiu V, Peperstraete J, Vandewalle J, Lauwereins R (1994) Closse approximations of sigmoid functions by sum of step for VLSI implementation of neural networks. Sci Ann Cuza Univ 3:5–34
11. Varalakshmi J, Jose D, Kumar PN (2019) FPGA implementation of hazeremoval technique based on dark channel prior. Adv Intell Syst Comput Book Ser AISC 1108. In: International conference on computational vision and bio inspired computing
12. Punitha L, Devi KN, Jose D, Sundararajan J (2019) Design of double edge-triggered flip-flop for low-power educational environment. Int J Electr Eng Educ. https://doi.org/10.1177/0020720919865836

Fog-Based Smart Healthcare Architecture in IoT Environment

Divya Gupta, Ankit Bansal, Shivani Wadhwa, and Kamal Deep Garg

Abstract The variety of services offered by IoT in different applications such as smart homes, smart mobility, smart wearables, and smart cities have made our lives easier. This transition from centralized approach to fully distributed environment proves applicability of IoT in healthcare sector. However, management of healthcare for ever increasing aging population with several diseases is a challenge due to insufficient availability of healthcare services and resources. The possible solutions could be shifting from hospital-centric treatment to home-centric treatment while connecting individual entities such as patients, hospitals, and services with each other for remote patient monitoring. This real time monitoring requires computation at fog layer to support minimal latency. This paper provides architecture of fog driven IoT e-healthcare along with its various services and applications. In addition, the challenges and solutions related to IoT healthcare in terms of data storage, scalability, and security concerns have been addressed.

Keywords IoT · Healthcare · Smart city · Challenges

D. Gupta
Department of CSE, Chandigarh University, Mohali 140413, India

A. Bansal (✉) · S. Wadhwa · K. D. Garg
Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India
e-mail: ankit.bansal@chitkara.edu.in

S. Wadhwa
e-mail: shivani.wadhwa@chitkara.edu.in

K. D. Garg
e-mail: kamaldeep.garg@chitkara.edu.in

1 Introduction

IoT integrates different softwares, hardwares, computing devices, humans, and objects on a network for computation, communication, storage, and exchange of data [1]. A traditional IoT system consists of various sensors, interfaces, algorithms, and cloud. In general, IoT architecture is divided among four layers (refer Fig. 1). In bottom up view, the bottom layer is for sensing the data provided through all the hardware devices present in the real world. Layer 2 offers networking support for transfer of data in network. Layer 3 creates and manages all the services required to fulfill user demands. Layer 4 provides interaction between user and application through different APIs [2].

Both technology and healthcare has a long history of connection. In addition, with increasing rise of IoT and wide acceptance of small wearable sensors have opened new possibilities for personalized healthcare [3]. Therefore, the IoT's four layer concept is directly applicable in healthcare sector. In this context, the different types of healthcare data is sensed through the wearable devices attached to patient's body by healthcare services. The collected data is transferred to cloud server for further processing through wireless network in support to networking layer. The processing of patient's health data in the form of testing blood sugar, heartbeat rate, pulse rate, and blood pressure are kind of services supported through service layer. Finally, the output generated from health data information in the form of patient's report is providing easy interface between user and application. Both doctors and patients can have access to these health reports and based on that doctor can suggest treatment or may fix an appointment. Using IoT in healthcare, the diagnosis and treatment of patient's health issues saves its time and money as well as offer effective utilization of hospital resources.

The rest of the paper is organized as follows: Sect. 2 provides knowledge on fog-based IoT healthcare architecture. Further, the challenges involved in implementing

Fig. 1 IoT in healthcare
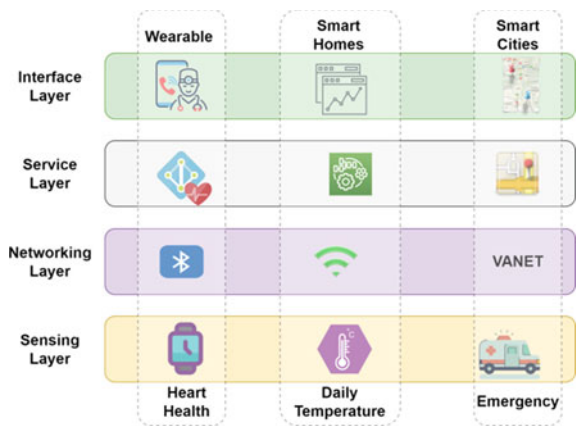
IoT in healthcare sector have been discussed in Sect. 3. Finally, Sect. 4 concludes the study.

2 Fog Driven IoT Healthcare Architecture

A traditional IoT system consists of various sensors, interfaces, algorithms, and cloud. Cloud itself supports various IoT services such as data computation, storage, and fast processing. However, high latency offered by cloud loses its suitability for real time applications [4]. In 2012, Cisco proposed a new paradigm called fog computing to overcome limitations of cloud computing. Therefore, fog computing can be embedded in to the applications that require real time response with minimal latency such as healthcare IoT system [5].

This fog driven multi-layer IoT health architecture has the potential to transform the way hospitals and caregivers deliver wellness to patients. This system is made up of three basic levels, as illustrated in Fig. 2: Device Layer, Fog Layer, and Cloud Layer. Although, various studies in past have already discussed IoT healthcare architecture using fog computing, still a comprehensive solution including everything from data collection and processing to cloud platforms and big data analytics [6, 7] is the need presented by this study.
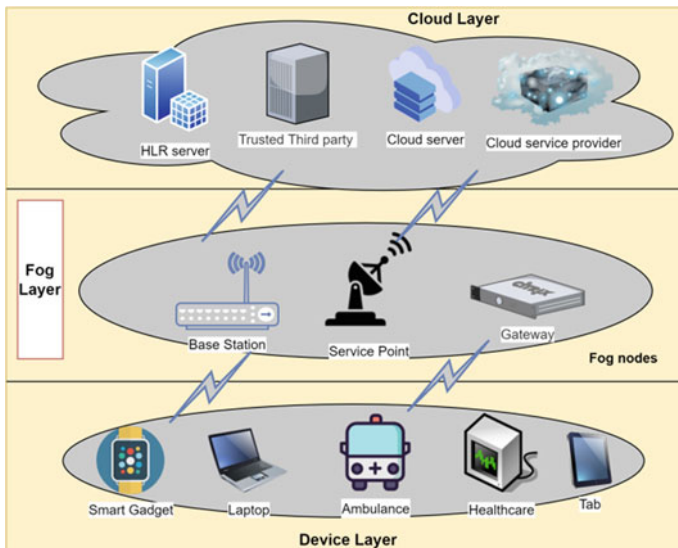
Fig. 2 Fog driven IoT healthcare architecture
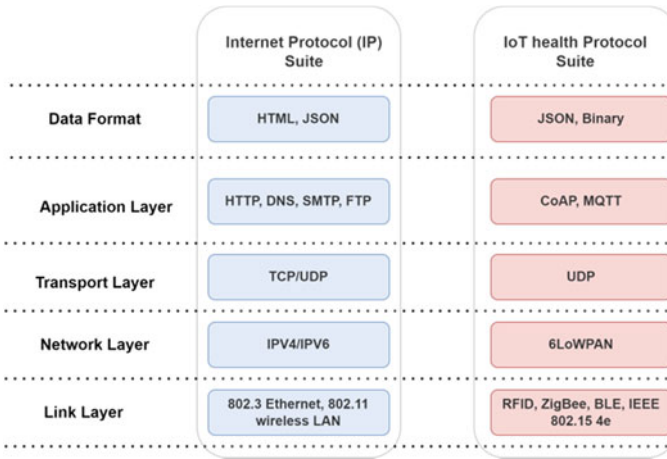
Fig. 3 IoT healthcare protocol stack

Device Layer

The presence of diverse variety of smart IoT healthcare devices have facilitated individual health data monitoring anytime, anywhere, and from any device along with real time synchronization at cloud. The only requirement for this real time monitoring is to establish a connection to a gateway or a fog node using a suitable communication protocol. The personal area network (PAN) and WSN protocols are widely accepted in this context. The IoT healthcare protocol stack is demonstrated in Fig. 3.

1. **Physical Sensor** In the healthcare sector, any medical equipment with a wired/wireless interface can be used to track patient's health physically or remotely. These devices include heartbeat monitor, body temperature monitor, smart shoes, smart gloves, smart watches, sleep monitor, elderly monitor, early warning system, and so on [8].
2. **Virtual Sensor** These sensors utilize mobile applications and different software services to sense the health data from the environment. This may include remote monitoring, remote diagnostic, and medical reference applications.

Fog Layer

Fog nodes exhibit computing and caching facility. In healthcare environment, the time sensitive decisions are taken on fog nodes for fast computation and processing. Such devices are always kept near to the devices producing health data. The rest non-sensitive data is passed to the cloud for further processing and main storage. Therefore, fog nodes are a viable option for processing, filtering, and compressing data sent between medical devices and the cloud.

Cloud Layer

Data from many sources is combined in the cloud. It collects data from a large number of fog nodes and saves it safely. As a result, it is always available to everyone involved in the patient's care. Non-sensor sources, like as data from EHRs, e-prescription platforms, and web sources, are also smoothly incorporated into this platform. Therefore, the data can be accessed by the patient, physicians, or anyone else on the patient's care team at any time and from any location. This improves collaboration across all specialties, resulting in a more efficient healthcare plan. Another benefit of the cloud platform is that it isolates the data and application layers while maintaining a single schema for capture and query transactions.

3   Challenges in IoT Healthcare

Although IoT aims to provide seamless connectivity across far away locations where all the health sector stakeholders such as patients, doctors, and health organizations could interact, communicate, co-operate, and orchestrate health related processes, still due to emergence of new technologies, there is a room for more development and some challenges which need to be addressed before it becomes a prevailing platform:

Data Storage and Management

The management of IoT health data is as challenging as in other IoT applications. Nevertheless, the healthcare data collected through sensors attached to human body is dynamic and changes rapidly. Moreover, due to continuous remote monitoring of patient health status, there is a huge amount of data sensed by sensors which needs processing at fog or cloud layer. The storage and management of such data is complex due to its variety, volume, and velocity. The variety of healthcare data is quite new as compared to previous existing challenges in past years. The format of healthcare data completely depends on the applications. For example: data collected for skin disease using camera need to be handle in image format while data from ECG has to be handle in XML format. On the other hand, the storage of high volume of data and managing this high velocity depends on capacity and capability of fog node hardware. As a result, fog administrators will be required to manage the data flow between the fog and cloud computing.

Security and Privacy

Each IoT device may pose a risk that might be abused to damage end-users or compromise their privacy. In case of compromised security at IoT device, any unauthorized user may have access to the private information through false authentication. Security and privacy is of great concern in IoT healthcare sector for complete span starting from information generation to its implementation, and from deployment to its result generation. However, maintaining security and privacy of complete IoT healthcare

ecosystem is a challenging task [9]. Therefore, challenges related to security in IoT healthcare needs to be addressed through multi-layer approach as discussed below:

- **Device Layer Security** The devices in IoT healthcare mainly comprises sensors, medical equipments, mobile devices, fog nodes, and gateways. These devices mainly capture, combine, process, and transfer data to cloud or fog for further computations. The most frequent threats on these devices are network spoofing, man-in-the-middle (MIM) attack, RF jamming, and direct connection. The only way to deal with MIM attack is to have assure guarantee from IoT device after complete verification and review of received certification. On the other hand, in direct connection, attackers use service discovery protocols to locate IoT devices.
- **Network Layer Security** The main responsibility of this layer is to establish a suitable connection among different network entities such as sensors, health devices, fog nodes, and cloud in support with appropriate network protocols set. The privacy of this layer is mainly compromised due to eavesdropping, sleep attack, and sybil attack. It is critical to use trustworthy routing mechanisms and encryption techniques to maintain integrity of information flow. To accomplish this, both symmetric and asymmetric algorithms based on cryptography approach can be implemented.
- **Cloud Layer Security** The data privacy and network security at cloud is the primary responsibility of any company delivering IoT/health services. The most common cloud vulnerabilities are Trojan houses, Denial of service (DoS), sniffing attack, and malicious file injection.

Scalability

The deployment of IoT in healthcare at small scale mainly requires sensors present on mobile devices for data collection and cloud for processing patient health records. The design must guarantee the access of medical data by patients and doctors via their portable devices. This service of IoT can be scaled up to cover entire hospital where patients can access all the required hospital services, book appointments, check health updates through their smart phones. Further, this service can be extended to cover entire city where all hospitals in the city can have access to the patient's data based on their terms and conditions. This would require sensors distributed in entire city, big data analytics algorithms, APIs for data processing and intelligent interfaces for analyzing user's requests in latency aware environment [10]. In this smart city environment, the collection of entire health related data and feedback based on analysis of data should all be done through smart phone applications. The scalability at entire city level improves system efficiency, better utilization of resources of both patient and hospital, saving time for doctors to consult more patients in a day, trust building between medical organizations and clients.

4 Conclusion

The transition from hospital-centric treatment to home-centric treatment is the need of the hour. In short, smart IoT healthcare has a bright future. Individual users may benefit from smart healthcare since it allows self-health management. Medical services can be availed anytime, anywhere based on the requirement, and treatment provided through these services will be more personalized, pervasive, participatory, and perpetual. In this study, we discussed fog driven IoT healthcare architecture to support real time health monitoring. Despite the various applications support provided by IoT health in different domains, there are some challenges for deployment of IoT in healthcare. In this context, various applications of IoT in healthcare and its challenges has been presented.

References

1. Gubbi J, Buyya R, Marusic S, Palaniswami M (2013) Internet of things (IoT): a vision, architectural elements, and future directions. Future Gener Comput Syst 29(7):1645–1660
2. Sobin CC (2020) A survey on architecture, protocols and challenges in IoT. Wirel Pers Commun 112(3):1383–1429
3. Kashani MH, Madanipour M, Nikravan M, Asghari P, Mahdipour E (2021) A systematic review of IoT in healthcare: applications, techniques, and trends. J Netw Comput Appl 192:103164
4. AbuKhousa E, Mohamed N, Al-Jaroodi J (2012) e-Health cloud: opportunities and challenges. Future Internet 4(3):621–645
5. Rahmani AM, Gia TN, Negash B, Anzanpour A, Azimi I, Jiang M, Liljeberg P (2018) Exploiting smart e-health gateways at the edge of healthcare internet-of-things: a fog computing approach. Future Gener Comput Syst 78:641–658
6. Mittal S, Bansal A, Gupta D, Juneja S, Turabieh H, Elarabawy MM, Bitsue ZK (2022) Using identity-based cryptography as a foundation for an effective and secure cloud model for e-health. Comput Intell Neurosci
7. Kaur K, Verma S, Bansal A (2021) IOT big data analytics in healthcare: benefits and challenges. In: 2021 6th international conference on signal processing, computing and control (ISPCC). IEEE, pp 176–181
8. Muthu B, Sivaparthipan CB, Manogaran G, Sundarasekar R, Kadry S, Shanthini A, Dasel A (2020) IOT based wearable sensor for diseases prediction and symptom analysis in healthcare sector. Peer-to-Peer Netw Appl 13(6):2123–2134
9. Manne R, Kantheti SC (2021) Application of artificial intelligence in healthcare: chances and challenges. Curr J Appl Sci Technol 40(6):78–89
10. Sengupta S, Bhunia SS (2020) Secure data management in cloudlet assisted IoT enabled e-health framework in smart city. IEEE Sens J 20(16):9581–9588

COVID-19 and SDG 3 in Bangladesh: A Statistical and Machine Learning Approach

Md. Mortuza Ahmmed, Md. Ashraful Babu, Jannatul Ferdosy, and Srikanta Kumar Mohapatra

Abstract Bangladesh is on the verge to confront severe prospective financial hurdle owing to the outcomes initiated by COVID-19. The objective of the study is to assess the status of the third sustainable development goal (SDG 3) in the nation preceding to COVID-19 advent as well as the evident effect of COVID-19 on SDG 3 in future. Information from national sources like IEDCR and BDHS have been employed. Findings show that the maternal mortality rate (MMR) fell by 6.3% while the total fertility rate (TFR) fell by 3.3% between 1993 and 2017. Different categories of early infantile mortality rates also dropped considerably during that period. Sizable progress happened in accompanying demographic factors during that period which led to enhanced level of maternal and child health. Lastly, how COVID-19 could impact maternal and child health through GDP has also been evaluated. Outcomes of the study would facilitate the policymakers to predict and ensure SDG 3 accomplishment accurately and take pertinent steps accordingly. Additional research is suggested to detect the causes for under deployment of optimal level healthcare facilities in the country.

Keywords COVID-19 · BDHS · FBProphet projection · Mortality · SDG 3

Md. Mortuza Ahmmed
Department of Mathematics, American International University—Bangladesh (AIUB),
Dhaka 1229, Bangladesh

Md. Ashraful Babu (✉)
Department of Quantitative Sciences, International University of Business Agriculture and
Technology, Dhaka 1230, Bangladesh
e-mail: ashraful388@gmail.com

J. Ferdosy
Department of Public Health, North South University, Dhaka 1229, Bangladesh

S. K. Mohapatra
Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura,
Punjab, India

1 Introduction

Bangladesh has attained considerable progress regarding health status of women and children in recent times and is on its way toward attaining sustainable development goal (SDG)-3 [1–3]. During 1993–2017, the under-5 death rate lowered by 33.33%, while the new born death rate lowered by 51.16%. The maternal death rate lowered from 57 to 19 in 1000 live births during those 24 years [4]. The health status of women and children has immense effect on the progress of a country. But the arrival of COVID-19 in 2020 impacted the whole economy of the country. The endless lockdown only worsened the situation [5]. The government flopped to evaluate the condition and take essential measures timely [6]. The garments sector lost half of their export profits due to the termination of order in 2020 [7]. The overseas remittance dropped by 13.98% in 2020 compared to 2019 [8]. Outcomes of several studies indicated that low death rate of women and children uplifts the GDP of a nation considerably [9–15]. Also, a successful immunization program substantially enhances the health status and consequently the GDP of a nation [16–18]. The immunization program in Bangladesh was on target preceding to COVID-19 influx in 2020 [19, 20].

To better the health condition of women and their infants has been the focal point of various public health plans in emerging nations like Bangladesh in recent times [21, 22]. The health authority of Bangladesh outlined the latest Health, Population, and Nutrition Sector Program to lower the maternal death rate to 12 in 10,000 live births by 2022 [23]. Newborn fatalities have been accounted for most infant fatalities across the globe [24, 25]. Newborn death rate plunged from 941 to 171 per 10,000 live births in Bangladesh between 1969 and 2018 [26]. Even though impressive progress was accomplished by Bangladesh about the Millennium Development Goals (MDGs), more improvement is needed to attain the SDGs [27–29]. It was established by the United Nations (UN) that financial downturn in emerging nations significantly influences the nutritional status of the mothers and their children [30, 31]. GDP growth rate showed strongly negative correlations with mortality rates in several studies [32–34]. Maternal and child health progress eventually consequences in improved GDP growth rate, guaranteeing additional improvement in people's living standard, nutrition level, government and non-government health care amenities, hygiene status, clean and safe water source, all of which eventually guarantee a better and more prolific forthcoming generation.

The objective of this study is to measure the improvement of SDG 3 goals against the COVID-19 effect of the country Bangladesh. For that, the progress of SDG 3 and associated factors before the COVID-19 appearance in the country has been demonstrated. The COVID-19 situation has also been fleetingly discussed. Besides, a forecast on COVID-19 has been made through FBProphet projection model. Finally, how COVID-19 could impact the SDG 3 fulfillment in Bangladesh has been discussed.

2 Materials and Methods

2.1 Source of the Data

Data regarding SDG 3 headway in Bangladesh have been unearthed from all the demographic and health surveys (BDHS) accomplished in the country so far. Besides, the COVID-19 related data is accumulated from the information circulated by Institute of Epidemiology, Disease Control and Research (IEDCR), Bangladesh.

2.2 Correlation and Regression Analyses

The correlation analysis operated to assess the association between two variables is defined as:

$$r = \frac{\text{SP}(x_1 x_2)}{\sqrt{\text{SS}(x_1)\text{SS}(x_2)}}$$

$$\text{SS}(x_1) = \sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n}$$

$$\text{SP}(x_1 x_2) = \sum x_1 x_2 - \frac{\sum x_1 \sum x_2}{n}$$

$$\text{SS}(x_2) = \sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n}$$

The correlation coefficient (r) ranges between -1 and 1, where $r = -1$ signals exact negative correlation, and $r = 1$ signals exact positive correlation. Besides, the regression model linking two variables is defined as:

$$X_2 = \vartheta + \lambda X_1 + \xi$$

ϑ is the intercept,
λ is the regression coefficient of X_2 on X_1,
ξ is the random error term.

The estimated model is defined as:

$$\widehat{X_2} = \psi + \eta X_1$$

$$\eta = \frac{\text{SP}(x_1 x_2)}{\text{SS}(x_1)} \text{ [estimate of } \lambda]$$

$$\psi = \overline{x_2} - \eta \overline{x_1} \text{ [estimate of } \vartheta]$$

The statistical significance of the associations among the selected variables under study has been established via *p*-values (*p*-value $\leq$ 0.05 signals significant association).

2.3 *FBProphet Projection Model*

FBProphet projection model developed by Facebook has been applied to project the COVID-19 trend in Bangladesh [35]. It employs the time series model encompassing three foremost facets—holidays, seasonality, and trend. Mathematically:

$$F(x) = N(x) + P(x) + H(x) + E(x)$$

$N(x)$ is the function of non-periodic changes,
$P(x)$ is the function of periodic changes,
$H(x)$ is the function of holiday effects,
$E(x)$ is the random error term.

2.4 *Analytical Tool*

All the required analyses are accomplished through Python together with Statistical Package for Social Sciences (SPSS).

3 Results and Discussion

3.1 *COVID-19 Situation in Bangladesh*

The first COVID-19 case in Bangladesh was spotted on March 7, 2020. There has been positive and significant correlation between the daily COVID-19 tests and identified cases ($r = 0.98$, *p*-value $= 0.00$) implying that more tests would result in more cases. But the test figures have not been sufficient to meet the expected level because of skilled work force limitations. Therefore, countless people were not identified on time, and they would have contaminated numerous other people by the time of their detection. Consequently, the nation is advancing toward a mild infection state in the long run (Fig. 1).

The summary statistics of the COVID-19 related information between March 8, 2020 and March 11, 2022 accumulated for the study have been encapsulated in Table 1. It can be observed that 18,597 tests were performed while 2655 individuals were infected each day on average. On average, 2614 individuals were recovered whereas 40 people perished every day.
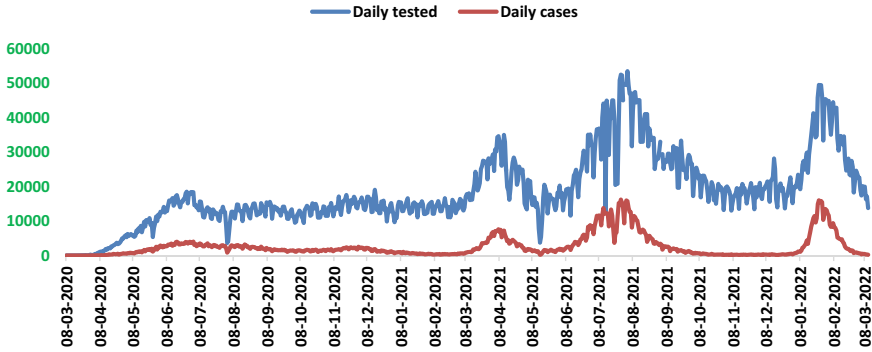
Fig. 1 Drifts of daily COVID-19 tests and identified cases in Bangladesh

Table 1 Summary statistics

	Overall	Lowest	Highest	Average	Variance
Tests per day	13,594,738	10	53,462	18,597	10,503
Infected per day	1,949,055	0	16,230	2655	3250
Deaths per day	29,105	0	264	40	52
Recovered per day	1,919,950	0	16,627	2614	3103

3.2 COVID-19 Projection for Bangladesh

FBProphet time series assessment has been performed to forecast the COVID-19 epidemic in the country as demonstrated in Fig. 2. It is appealing to witness in Fig. 2 that the frequency of infected people would be boosted approximately 5000 per day in next two months. The frequency of fatalities would be enhanced approximately 75 per day in next two months. The constant eruption in the model is due to the seasonality factor. The naturally developed immune system in human body would not allow COVID-19 to last too long. Also, measures like immunization, use of masks, physical isolation, and lockdown would counteract the dissemination of COVID-19.

3.3 SDG 3 Progress in Bangladesh

Figure 3 illuminates the progress of SDG 3 in Bangladesh between 1993 and 2017 regarding TFR, MMR, and different categories of early infantile mortality rates. The TFR was just about 3.3 between 1993 and 2000 which dropped down to 2.7 by 2007. By 2011, it dribbled further to 2.3 and got wedged there afterwards. The discernable fact is that as the TFR moved on lessening, MMR too began lessening. A highly significant positive association ($r = 0.91$, p-value $= 0.00$) among them vindicates tattoo. In case of different categories of early infantile mortality rates, the decline is

Forecasted Daily Infections for COVID-19 in Bangladesh
FBProphet-Machine Learning Model

Forecasted Daily Deaths for COVID-19 in Bangladesh
FBProphet-Machine Learning Model

Forecasted Cumulative Infections for COVID-19 in Bangladesh
FBProphet-Machine Learning Model

Forecasted Cumulative Deaths for COVID-19 in Bangladesh
FBProphet-Machine Learning Model

Fig. 2 FBProphet model for predictingCOVID-19 situation in Bangladesh
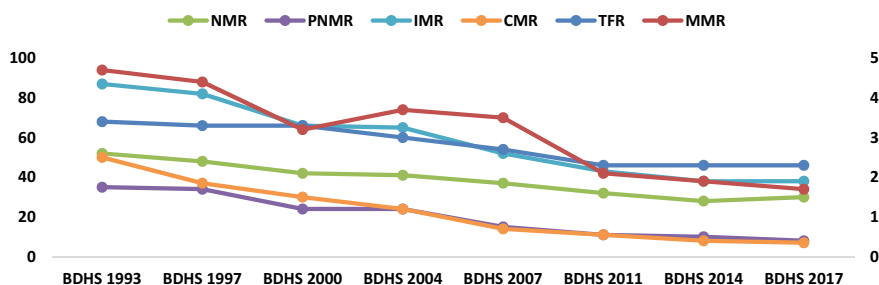
Fig. 3 TFR, MMR and early infantile death rates in Bangladesh between 1993 and 2017

quicker in elder age groups than the junior ones. The neonatal mortality rate (NMR) fell from 52 to 30 while the child mortality rate (CMR) fell from 50 to 7 between 1993 and 2017. NMR was responsible for almost 66% of total under-five mortality rate (UMR) according to BDHS 2017. Highly significant positive associations ($r = 0.96$, 0.95, 0.96, 0.94 correspondingly; p-value = 0.00, 0.00, 0.00, 0.00 correspondingly) have been observed among TFR and different categories of early infantile mortality rates too. These outcomes evidently suggest that fertility has strong association with MMR and different categories of early infantile mortality rates in Bangladesh.

Table 2 embodies the comparative distribution of various factors related with SDG 3 between 1993 and 2017 in Bangladesh. Fraction of population inhabiting in urban regions has been rising constantly between 1993 and 2017, while reverse trend can be witnessed for rural regions. Urban women remain in superior position than the rural ones concerning education, mass media exposure, dwelling standard, employment options, healthcare resources and so on, safeguarding the accomplishment of SDG 3 to some extent. Education authorizes a woman to develop self-reliance, permitting her to procure various crucial choices like marriage, contraception, gestation, childbirth place, etc. freely. An educated woman better understands various facets of maternal and child health. Rising trends is noticed in various classes of women education over the years in Table 2. Besides, the percentage of poor has waned incessantly while the ratio of middle class has been hiking over the years. The ratio of childbirths happening outside home has also been rising over the years. In fact, the ratio of childbirths taking place at home and outside is equal as per BDHS 2017. Childbirths at home is unsafe as qualified medical specialists are absent in most cases and are done by unqualified people like relatives and dais, raising the threat of insecure childbirth. Table 2 reveals lower status of mothers in decision-making contrasted to their husband about health matters.

Table 2 also reveals apparent progress in nutritional status of children between 1993 and 2017 in Bangladesh. Level of nutrition among the children is an important feature of their wellbeing. The proportion of mothers provided with antenatal care facilities has also enhanced between 1993 and 2017. A safe childbirth facilitated by a medical expert is an important aspect of SDG 3. Besides, CMR and vaccination coverage is observed to have highly significant negative association ($r = -0.92$,

Table 2 Distribution (%) of factors linked with SDG 3 during 1993 to 2017 in Bangladesh

		BDHS 1993	BDHS 1997	BDHS 2000	BDHS 2004	BDHS 2007	BDHS 2011	BDHS 2014	BDHS 2017
Safe childbirth	Yes	9	8	10	13	21	32	42	53
	No	91	92	90	87	79	68	58	47
Residing place	Urban	12	11	20	23	25	26	28	29
	Rural	88	89	80	77	75	74	72	71
Delivery place	Outside	4	5	3	9	18	28	35	50
	Home	96	95	97	91	82	72	65	50
Women education	Illiterate	58	55	47	40	34	28	25	17
	Primary	17	17	29	30	30	29	29	31
	Secondary	10	10	18	22	24	30	32	35
	Higher	15	18	5	7	12	12	14	17
Antenatal care	Yes	27	29	37	56	63	68	79	92
	No	73	71	63	44	37	32	21	8
Wealth index	Poor	NA	NA	NA	40	39	38	36	38
	Middle	NA	NA	NA	40	40	41	40	41
	Rich	NA	NA	NA	20	21	21	24	21
Vaccination coverage	Yes	46	54	66	68	76	83	78	86
	No	54	46	34	32	24	17	22	14
Decision on own health	Women	NA	NA	17	18	14	13	14	14
	Husband	NA	NA	40	48	32	31	29	29
Nutritional status of children	Stunting	NA	55	45	43	43	41	36	31
	Underweight	NA	56	48	48	41	36	33	22

(continued)

Table 2 (continued)

		BDHS 1993	BDHS 1997	BDHS 2000	BDHS 2004	BDHS 2007	BDHS 2011	BDHS 2014	BDHS 2017
	Wasting	NA	18	10	13	17	16	14	8
Decision on child health	Women	NA	NA	16	17	19	15	16	16
	Husband	NA	NA	29	32	16	20	28	28

p-value = 0.03) between them signifying that additional vaccination coverage will bring down the child fatalities and vice versa. Table 2 indicates improving trends of safe childbirth as well as vaccination coverage of children in Bangladesh between 1993 and 2017.

3.4 COVID-19 Effects on SDG 3 Through GDP in Bangladesh

The lockdown throughout the country made individuals to get involved in sexual contact more often than usual venturing unanticipated gestation. Moreover, a sizable section of skilled medical professionals stopped performing regular practices both in hospitals and their personal clinic sever since the influx of COVID-19 in the country which impacted the important health related services required for females, particularly, the expectant mothers. Subsequently, it heightened the hazard of maternal transience while giving birth (Table 3).

The regression model taking GDP and MMR into account can be written mathematically as:

$$MMR = 7.32 - 0.73(GDP)$$

The highly significant (*p*-value = 0.00) regression coefficient of the model suggests that 0.73 unit raise in MMR would result from one unit fall in GDP on average. Therefore, as the COVID-19 lengthens in the country, the GDP would incessantly drop, ensuing in higher MMR. The coefficient of determination score $\left(R^2 = 0.71\right)$ of the model suggests that GDP explains 71% variation of MMR through the model (Table 4).

Table 3 Impacts of GDP on maternal deaths

	Regression coefficient		*t*-statistic	*p*-values	95% confidence interval	
	β	Standard error (β)			Lower limit	Upper limit
GDP	− 0.73	0.16	− 4.70	0.00	− 1.05	− 0.41
Constant	7.32	0.94	7.80	0.00	5.37	9.26

Table 4 Impacts of GDP on child deaths

	Regression coefficient		*t*-statistic	*p*-values	95% confidence interval	
	β	Standard error (β)			Lower limit	Upper limit
GDP	− 0.83	0.21	− 3.90	0.00	− 1.27	− 0.39
Constant	8.33	1.29	6.48	0.00	5.67	11.00

In 2020, the government suspended its rubella and measles vaccination operation because of COVID-19. Deployment of essential health facilities for the children declined substantially due to COVID-19 spread across the country, subsequently, raising the hazard of child mortality. The regression model taking GDP and MMR into account can be written mathematically as:

$$CMR = 8.33 - 0.83(GDP)$$

The highly significant (p-value $= 0.00$) regression coefficient of the model suggests that 0.83 unit raise in CMR would result from one unit fall in GDP on average. Therefore, as the COVID-19 lengthens in the country, the GDP would incessantly drop, ensuing in higher CMR. The coefficient of determination score $\left(R^2 = 0.64\right)$ of the model suggests that GDP explains 64% variation of CMR through the model.

4   Conclusion

Findings of the study provide an understanding about the way COVID-19 impacted the progress of SDG 3 in Bangladesh. The status of maternal and child health was following an improving trend before the appearance of COVID-19 in the country. But the COVID-19 transmission across the nation stalled the development. The concerned authorities have already put forward their attention bout issues regarding SDG 3. Both the government and established NGOs should make sure additional efficacious strategies regarding SDG 3 targets fulfillment. The vaccination program required to be accomplished since life expectancy of a massive number of kids would be endangered if important vaccines were not provided to them on time. Additional research is required to discover other potential mediating causes through which the COVID-19 could impact the SDG 3 progress in Bangladesh. The outcomes of the study need to be brought into sensible attention by the relevant authorities to safeguard the welfare of mothers and their infants in Bangladesh during COVID-19 and comply with SDG 3.

References

1. Ahmmed M (2018) Maternal and child health in Bangladesh over the years: evidence from demographic and health surveys. Natl Univ J Sci 3–5:127–138
2. Ahmmed MM, Anee SA, Babu MA, Salim ZR, Shohel M, Babu JA, Darda MA (2020) Disparities in maternal and newborn health interventions in Bangladesh: evidence from the latest demographic and health survey. Ann Trop Med Public Health 23:231–639
3. Sack DA (2008) Achieving the millennium development goals in Bangladesh. J Health Popul Nutr 26(3):251–252

4. Ahmmed MM, Babu MA, Rahman MM (2021) A PLS-SEM approach to connect fertility, GDP, and childhood mortality with female life expectancy (FLE) in Bangladesh. AJSE 20(4):151–157
5. Amit S (2020) Long read: coronavirus and the Bangladesh economy: navigating the global COVID-19 shutdown. South Asia @ LSE, 01 Apr 2020. Blog Entry
6. Islam MD, Siddika A (2020) COVID-19 and Bangladesh: a study of the public perception on the measures taken by the government
7. Lalon RM (2020) COVID-19 vs Bangladesh: is it possible to recover the impending economic distress amid this pandemic?
8. Maliszewska M, Mattoo A, van der Mensbrugghe D (2020) The potential impact of COVID-19 on GDP and trade: a preliminary assessment. World Bank Policy Research Working Paper No 9211
9. Iacus SM, Natale F, Santamaria C, Spyratos S, Vespe M (2020) Estimating and projecting air passenger traffic during the COVID-19 coronavirus outbreak and its socio-economic impact. Saf Sci 129:104791
10. Saarela J, Finnäs F (2014) Infant mortality and ethnicity in an indigenous European population: novel evidence from the Finnish population register. Sci Rep 4:4214
11. Reidpath D, Allotey P (2003) Infant mortality rate as an indicator of population health. J Epidemiol Community Health 57(5):344–346
12. Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, Lawn E, Cousens S, Mathers C, Black E (2016) Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the sustainable development goals. Lancet 388(10063):3027–3035
13. Woldeamanuel T (2019) Socioeconomic, demographic, and environmental determinants of under-5 mortality in Ethiopia: evidence from Ethiopian demographic and health survey, 2016. Child Development Research
14. Mejía-Guevara I, Zuo W, Bendavid E, Li N, Tuljapurkar S (2019) Age distribution, trends, and forecasts of under-5 mortality in 31 sub-Saharan African countries: a modeling study. PLoS Med 16(3):e1002757
15. Hill K, You D, Inoue M, Oestergaard Z (2012) Child mortality estimation: accelerated progress in reducing global child mortality, 1990–2010. PLoS Med 9(8):e1001303
16. Jit M, Hutubessy R, Png E, Sundaram N, Audimulam J, Salim S, Yoong J (2015) The broader economic impact of vaccination: reviewing and appraising the strength of evidence. BMC Med 13(1):209
17. Ozawa S, Clark S, Portnoy A, Grewal S, Stack ML, Sinha A, Mirelman A, Franklin H, Friberg IK, Tam Y, Walker N, Clark A, Ferrari M, Suraratdecha C, Sweet S, Goldie SJ, Garske T, Li M, Hansen PM, Johnson HL, Walker D (2017) Estimated economic impact of vaccinations in 73 low- and middle-income countries, 2001–2020. Bull World Health Organ 95(9):629–638
18. Pathirana J, Munoz M, Abbing-Karahagopian V, Bhat N, Harris T, Kapoor A, Keene L, Mangili A, Padula A, Pande L, Pool (2016) Neonatal death: case definition & guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 34(49):6027–6037
19. Bangladesh EPI Coverage Evaluation Survey (2006) Directorate general of health services (DGHS), Dhaka, Bangladesh
20. Bangladesh EPI Coverage Evaluation Survey (2016) Directorate general of health services (DGHS), Dhaka, Bangladesh
21. Carrascal LM, Galva'n I, Gordo O (2009) Partial least squares regression as an alternative to current regression methods used in ecology. Oikos 118:681–690
22. Upadhyay KA, Srivastava S (2012) Association between economic growth and infant mortality: evidence from 132 demographic and health surveys from 36 developing countries. International Institute for Population Sciences Mumbai
23. BMMS (2010) Bangladesh maternal mortality and health care survey 2014. National Institute of Population Research and Training (NIPORT), Dhaka
24. (1991) Handbook of vital statistics systems and methods. Volume 1: Legal, organizational and technical aspects. United Nations studies in methods, glossary. Series F, No. 35. United Nations, New York

25. Hessol A, Fuentes-Afflick E (2005) Ethnic differences in neonatal and post-neonatal mortality. Pediatrics 115(1):e44–e51
26. Baqui A, Ahmed S, Arifeen S, Darmstadt G, Rosecrans A, Mannan I, Rahman S, Begum N, Mahmud A, Seraji H, Williams E, Winch P, Santosham M, Black R (2009) Effect of timing of first postnatal care home visit on neonatal mortality in Bangladesh: an observational cohort study. BMJ 339(7718):1–6
27. Ahmmed MM, Babu MA, Ferdosy J (2021) Direct and indirect effects of COVID-19 on maternal and child health in Bangladesh. J Stat Manag Syst 24(1):175–192
28. Ahmmed M (2017) Factors associated with safe delivery practice in Bangladesh. Int J Health Prefer Res (IJHPR) 2017(1):15–24
29. Darda MA, Ahmmed M (2019) Effect of maternal and child health outcomes on GDP growth rate in Bangladesh. Natl Univ J Sci 6(1):113–130
30. Haider M, Rahman M, Moinuddin M, Rahman A, Ahmed S, Khan M (2017) Impact of maternal and neonatal health initiatives on inequity in maternal health care utilization in Bangladesh. PLoS ONE 12(7):1–15
31. Ensor T, Cooper S, Davidson L, Fitzmaurice A, Graham J (2010) The impact of economic recession on maternal and infant mortality: lessons from history. BMC Public Health 10:727
32. Erdogan E, Ener M, Arıca F (2013) The strategic role of infant mortality in the process of economic growth: an application for high income OECD countries. Procedia Soc Behav Sci 99:19–25
33. Pervin MR, Parvin R, Babu MA, Ahmmed MM, Marzo RR (2021) The obstacles to combat against COVID-19 pandemic and the remedies: Bangladesh scenario. J Public Health Res 10(s2)
34. Kirigia M, Oluwole D, Mwabo M, Gatwiri D, Kainyu H (2006) Effects of maternal mortality on gross domestic product (GDP) in the WHO African region. Afr J Health Sci 13(1):86–95
35. Babu MA, Ahmmed MM, Ferdousi A, Mostafizur Rahman M, Saiduzzaman M, Bhatnagar V, Raja L, Poonia RC (2022) The mathematical and machine learning models to forecast the COVID-19 outbreaks in Bangladesh. J Interdisc Math 1–20

Analysis of Power Consumption and IoT Devices Using SPSS Tool

Anjali Nighoskar, Premananda Sahu, and Shweta Tiwari

Abstract India is the third largest producer of electricity in the world. For the sustainable energy transition and climate change mitigation, renewable energy integration and energy use optimization are important influencers. Technological innovations, such as the Internet of Things (IoT), have numerous applications in the energy sector, such as supply, generation and distribution, and requirements. The IoT can be useful for the enhancement of energy, increasing the use of renewable energy, and protecting the environment. In this paper, the analysis of energy consumption of IoT devices per home has been carried out and compared with the total energy consumption of India according to different states using the SPSS tool, which shows the linear regression, multi-correlation, and ANOVA test. The results are shown in graphs and figures.

Keyword Energy consumption · IoT devices · SPSS tool · Regression modeling

1 Introduction

Today's energy consumption is quite high. It is critical that the energy consumed is suitable for efficient and cost-effective use. The IoT energy management protocol is a new protocol for using smart devices in organizations and homes [12]. It should be noted that the IoT global energy market exceeded 6.8 billion (USD) in 2015 and can reach 26.5 billion (USD) by 2023, with a percentage per annum of 15.5% in 2023. According to the Central Electricity Authority (CEA), the total peak power

A. Nighoskar (✉)
School of Engineering and Technology, Jagran Lakecity University, Bhopal, India
e-mail: anjali.nighoskar@jlu.edu.in

P. Sahu
SRM Institute of Science and Technology, SRM IST, Ghaziabad, UP, India

S. Tiwari
Information Technology, Radharaman Institute of Technology and Science, Bhopal, India

consumption in India between April and August 2021 was 203,014 MW, up from 171,510 MW at the same time the previous year [5]. 5G is on the way, IP traffic is considerably higher than expected, and all cars, machines, robots, and artificial intelligence are being digitized, generating massive volumes of data that will be housed in data centers [2]. By 2025, the 'data tsunami' could consume one-fifth of global electricity. The astronomical data generated by the IoT industry will be critical in the design and implementation of new energy-saving solutions [14]. What is being missed, though, is the massive energy footprint that the IoT will leave behind. According to Cisco, the total number of IoT devices will reach 50 billion by next year, and the energy utilized by these devices will have a negative influence on global energy demands [18]. The Internet of Things (IoT) offers a comprehensive method of assessing and managing usage not only at the device level but also across the entire organizational system. IoT can also detect inefficient usage caused by outdated appliances, damaged appliances, or malfunctioning system components. This simplifies the process of monitoring and controlling energy while keeping costs low and precision high [15]. Here, the total consumption of India, including all the IoT devices and other industries, homes, and factories, is analyzed using the IBM SPSS tool. The paper is organized as follows: In Sect. 2, the discussion over power consumption and IoT devices, in Sect. 3, the literature review has been discussed, followed by Sect. 4, which shows the different challenges and suggestions. In Sect. 5, the analysis of the results with the SPSS tool is shown, and finally, in Sect. 6, the conclusion of the paper.

2 Power Consumption and IoT Devices

The 'linked home' is a collection of services and solutions that provide consumers with added value. It is controlled by the user through the IoT core, which connects to and communicates with the digital gadgets in the home. Several IoT applications can benefit from a constant power supply provided by energy harvesting systems, which are still in their early stages of development [10]. According to the statistics, the total number of smart homes in India is 13.15 million. And the forecast smart home penetration rate is 12.84% by 2025. India still lacks fully specified and comprehensive IoT standards that can be applied to all of today's smart gadgets and appliances. However, since the start of the IoT, we have come a long way. Figure 1 shows the impact of power consumption of IoT devices per smart home. Figure 2 shows the total power consumption in India for the fiscal year 2020. Analysis of power consumption between smart homes using IoT devices and those without IoT devices has been shown in Fig. 3.
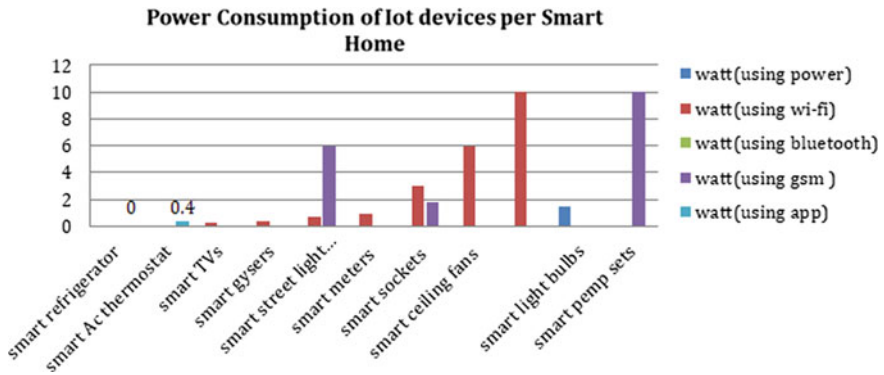
Fig. 1 Power consumption of IoT device per smart home
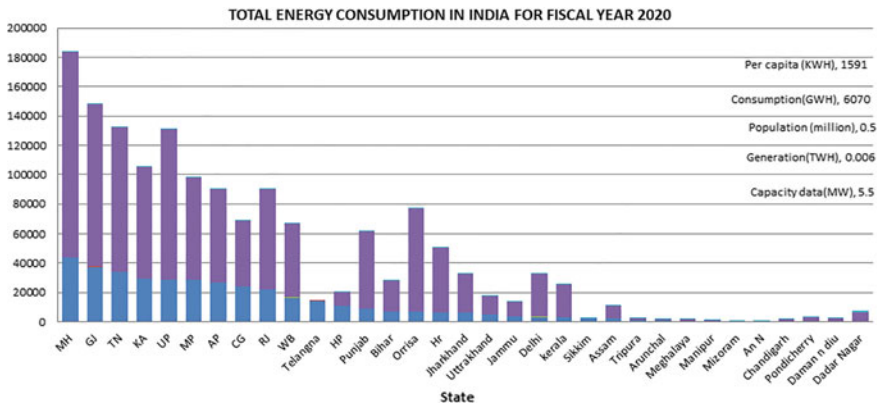
Fig. 2 Total power consumption in India for fiscal year 2020

Fig. 3 Power consumption analysis
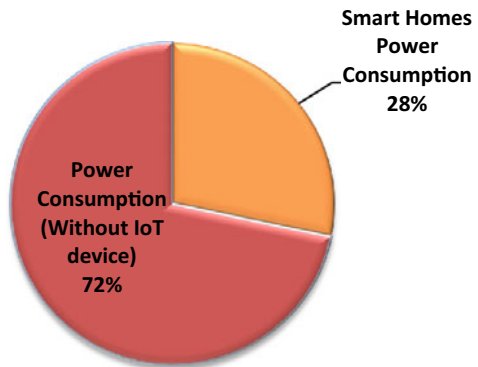
3 Literature Review

The critical review from different literature has been discussed:

Ahmed et al. [1] has study on different applications and analyze the usage of IoT in energy generation which is involved in a variety of renewable sources. Similarly authors Rana et al. [17] have offered a variety of aspects of IoT that lacked interoperability and energy efficiency and investigate the impact on IoT devices when various types of energy harvesters are linked with these devices. Iwendi et al. [13] have used the Whale Optimization Algorithm (WOA) with Simulated Annealing (SA) that has several performance measures like alive nodes, load, temperature, residual energy, and cost function to identify the best Cluster Head in IoT network clusters. Azari et al. [7] created a tractable model of NB-IoT connectivity. The method is to examine the influence of channel scheduling and coverage class interaction on the performance of IoT devices by deriving predicted latency and battery lifespan. Arshad et al. [3] have suggested an intrusion detection framework for energy-constrained IoT devices. The suggested method will detect the intrusion detection while minimizing energy consumption and communication overhead using Contiki OS. Azari et al. [6] have investigated low-complexity distributed learning algorithms. They provided a learning technique for adapting device communication parameters to maximize energy economy and data transmission dependability by using stochastic geometry. Hung et al. [1] have analysis of payload lengths and AES encryption types. It includes hardware-based AES-ECB, AES-CCM, and AES-CB. The implementation is done using Texas Instruments LAUNCHXL-CC1310 platform and showed the better results. Donatiello et al. [9] has done the probabilistic analysis of the energy consumed by mobile apps. The analysis was done on the occurrence of random events on a given time interval. Polynomial algorithm is used for calculation. Singh et al. [18] have analysis of lightweight cryptography algorithms. The analysis was done on the basis of key, block size, structures, and number of rounds. They also proposed a security service scheme to improve the resource constrained in IoT environment.

4 Challenges and Suggestions

The different challenges have been discussed and tried to provide suggestions to solve the issues:

1. Fault tolerance and discovery: This is a term that refers to network operations that are sensitive to events such as node failure or network resource shortages. Power reliability, node reliability, and route reliability are three crucial features that are difficult to achieve. IoT networks can be made effective if all of these aspects are handled together. Installing different check-points on devices will check the health status priorly [13].

2. Software interoperability and complexity: The selection of protocol gateways is the recent study on IoT device interoperability. However, the protocol appears to be exacerbating rather than resolving the interoperability issue. As a result, increasing server counts, reconfiguration attempts, analysis, and Internet speed become optimization issues. The protocol bridge and its interconnectivity, according to the industry, must be expandable, consistent, reliable, and protected [9].

3. Power supply and stability: A major challenge is met for self-contained IoT environments, especially facing the "constantly on" demands. Even if it has a battery, the device is only wireless. Engineers must exercise prudence when designing and developing an IoT device in order to minimize energy consumption and provide additional power backup [6].

4. Amounts of information and interpretation: Overloading of the data is the major challenge. This is especially difficult in the electricity sector, where massive amounts of data sets are collected from multiple devices, enterprises, and actuators by consumers as well as manufacturers. To reduce these issues, different methods should be used for producing energy, like green energy, solar energy, etc. [4].

5   Method to Calculate the Power Consumption

Method to calculate the power consumption of IoT devices and total power consumption of different states of India is by using IBM SPSS tool [8]. Using data set from https://www.niti.gov.in/edm/theanalysis has been done. In Fig. 4, the result for correlation b/w power consumption, per-capita and total population is calculated using SPSS tools which shows that the independent variable is having some relationship to the dependent variable. Here, the independent variable is population and dependent variables are per-capita and power consumption. The result shows the R value between total consumption and population is -.123 and -.2 respectively, that means they are strongly correlated to each other. In Fig. 5 the result for the correlation between power consumption, the per-capita, and the total population is calculated, which shows that the independent variable has some relationship with the dependent variable [16, 11]. Here, the independent variable is population, and the dependent variables are per-capita and power consumption. One-way "analysis of variance" (ANOVA) examines the means of two or more independent groups to see if there is statistical evidence that the related population means differ significantly. A parametric test is one-way ANOVA. Figure 6 shows the steps in SPSS tool to find the ANOVA test. F is the test statistic for a One-way ANOVA. The F statistic determines if the group means of an independent variable with k groups are substantially different. Because generating the F statistic is slightly more complicated than computing the paired or independent samples t test statistics, all of the F statistic components are frequently represented in a Fig. 5. Likewise, the coefficient value is also calculated and discussed in the result section [19].

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.226[a]	.051	-.015	293.54779

a. Predictors: (Constant), population, totalconsumption

b. Dependent Variable: percapitaconsumption

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	134092.985	2	67046.492	.778	.469[b]
	Residual	2498938.890	29	86170.307		
	Total	2633031.875	31			

a. Dependent Variable: percapitaconsumption

b. Predictors: (Constant), population, totalconsumption

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)	214.593	71.560		2.999	.006	68.236	360.950
	totalconsumption	.001	.002	.104	.369	.715	-.004	.005
	population	-1.714	1.640	-.296	-1.045	.305	-5.069	1.641

a. Dependent Variable: percapitaconsumption

Fig. 4 Results b/w total power consumption, per-capita and population

Descriptive Statistics

	Mean	Std. Deviation	N
percapitaconsumption	172.0625	291.43869	32
totalconsumption	35734.3750	38905.96400	32
population	41.1219	50.32363	32

Correlations

		percapitaconsumption	totalconsumption	population
Pearson Correlation	percapitaconsumption	1.000	-.123	-.216
	totalconsumption	-.123	1.000	.769
	population	-.216	.769	1.000
Sig. (1-tailed)	percapitaconsumption	.	.251	.118
	totalconsumption	.251	.	.000
	population	.118	.000	.
N	percapitaconsumption	32	32	32
	totalconsumption	32	32	32
	population	32	32	32

Fig. 5 Correlations b/w total power consumption, per-capita and population
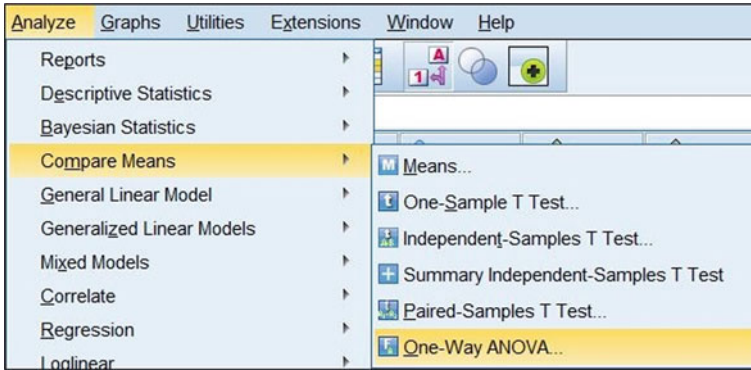
Fig. 6 Steps to calculate the ANOVA test

6 Result and Discussion

This article did an exploratory study on the energy consumption of IoT devices in smart homes in contrast to India's total and per-capita consumption. The inquiry and analysis reveal a relationship between IoT devices and energy usage in various states, as illustrated in Fig. 2. Similarly, Fig. 1 depicts the power consumption of IoT devices. These results clearly demonstrate that deploying IoT devices in smart homes reduces total power usage and saves energy. The pie graphic depicts the impact of both outcomes. Here, by using SPSS tool the result shows the R-value between total consumption and population is -0.123 and -0.216, respectively, which means they are strongly correlated with each other. Likewise, in Fig. 4 the regression analysis by ANOVA test says that if the value of $F < 4$ and the significant value is sig > 0.05 then the assumption is fulfilled. Hence, the hypothesis cannot be rejected. Thus, all group meanings are equal.

7 Conclusion

The power consumption of IoT devices in smart homes, as well as the total consumption of India with population and per-capita consumption, are shown in this study. Smart homes are very few in comparison to the total population, which affects energy consumption in India. The benefits of IoT-based energy management systems in enhancing energy efficiency and integrating renewable energy have been discussed and the results have been presented. We have also discussed the different challenges

faced by researchers. The results of correlation and regression analysis with an ANOVA test have been analyzed using the SPSS tool. The data is taken from Niti Aayog (Govt. of India) for the fiscal year 2020. Finally, it has been concluded that if we increase the number of IoT devices in our homes (or smart homes), then the energy consumption rate can be reduced in the coming years.

References

1. Ahmad T, Zhang D (2021) Using the internet of things in smart energy systems and networks. Sustain Cities Soc 68:102783
2. Akhavan-Hejazi H, Mohsenian-Rad H (2018) Power systems big data analytics: an assessment of paradigm shift barriers and prospects. Energy Rep 4:91–100
3. Arshad J, Azad MA, Abdeltaif MM, Salah K (2020) An intrusion detection frame-work for energy constrained IoT devices. Mech Syst Signal Process 136:106436
4. Arshad R, Zahoor S, Shah MA, Wahid A, Yu H (2017) Green IoT: an investigation on energy saving practices for 2020 and beyond. IEEE Access 5:15667–15681
5. Authority CE. CEA. https://cea.nic.in/old/
6. Azari A, Cavdar C (2018) Self-organized low-power IoT networks: a distributed learning approach. In: 2018 IEEE global communications conference (GLOBECOM). IEEE, pp 1–7
7. Azari A, Stefanović C, Popovski P, Cavdar C (2019) On the latency-energy performance of NB-IoT systems in providing wide-area IoT connectivity. IEEE Trans Green Commun Networking 4(1):57–68
8. Braun M, Altan H, Beck S (2013) Using IBM SPSS statistics to identify predictors of electricity consumption in a UK supermarket. Energy Sustain IV 176:119
9. Donatiello L, Marfia G (2018) Modeling the energy consumption of upload patterns on smartphones and IoT devices. IEEE Commun Lett 22(11):2258–2261
10. Foundation SSE (2018) A diagnostic study of the energy efficiency of IoT. https://shaktifoundation.in/report/diagnostic-study-energy-efficiency-iot/
11. Hossein Motlagh N, Mohammadrezaei M, Hunt J, Zakeri B (2020) Internet of things (IoT) and the energy sector. Energies 13(2):494
12. Hung CW, Hsu WT (2018) Power consumption and calculation requirement analysis of AES for WSN IoT. Sensors 18(6):1675
13. Iwendi C, Maddikunta PKR, Gadekallu TR, Lakshmanna K, Bashir AK, Piran MJ (2021) A metaheuristic optimization approach for energy efficiency in the IoT networks. Softw Pract Experience 51(12):2558–2571
14. Mandava S, Gudipalli A (2018) Analysis of home energy management system using IoT. Int J Pure Appl Math 118:3957–3968
15. Mocnej J, Miškuf M, Papcun P, Zolotov I (2018) Impact of edge computing paradigm on energy consumption in IoT. IFAC-PapersOnLine 51(6):162–167
16. NITI Aayog GoI. Electricity consumption. https://www.niti.gov.in/edm/

17. Rana B, Singh Y, Singh PK (2021) A systematic survey on internet of things: energy efficiency and interoperability perspective. Trans Emerg Telecommun Technol 32(8):e4166
18. Singh S, Sharma PK, Moon SY, Park JH (2017) Advanced lightweight encryption algorithms for IoT devices: survey, challenges and solutions. J Ambient Intell Humanized Comput 1–18
19. University KS (2022) SPSS tutorials: one-way ANOVA. https://libguides.library.kent.edu/spss/onewayanova

Analyzing Optimal Environment for the Text Classification in Deep Learning

Ochin Sharma

Abstract Deep learning is an area that emerged from the capabilities of machine learning and artificial intelligence. In the field of deep learning, various parameters are essential to building an efficient model. These parameters include activation function, loss function, number of layers, number of neurons, and many other essentials. Finding an optimal workable environment by setting all these variables is time-consuming and challenging due to time constraints or projects having sharp deadlines. In this paper, the work is done for the exploration of these parameters that make an environment in which deep learning models work. The experimental results will also be discussed to learn about the best environment that makes text classification optimal in deep learning. These results would help the professionals and researchers to extract a ready-to-work environment for the projects associated with the text classification.

Keywords Activation function · Relu · Accuracy · Loss function · Epochs · Weight · Machine learning

1 Introduction

Deep learning is a new field that has emerged in recent years and has proven its worth with many current technologies including text classification (TC). Deep learning comprises multiple levels of neurons. Such layers may vary from a few neurons to thousands of neurons. The input values are multiplied by the weight assigned to each input and the result is added. The activation function further scrutinizes this result. The deep learning model achieves greater precision in this way [1, 2]. In

O. Sharma (✉)

Chitkara University Institute of Engineering & Technology, Chitkara University, Rajpura, Punjab, India

e-mail: Ochin.sharma@chitkara.edu.in
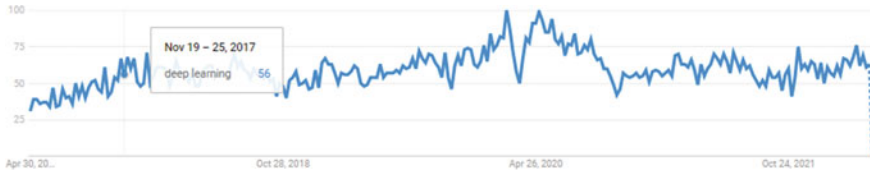
Fig. 1 Google trend for deep learning over the past five years

Fig. 1, Google trending shows that deep learning is a field of interest among professionals and researchers. The professions involved in deep learning have increased dramatically in the last five years.

Deep learning has applications like image detection, voice acknowledgment, computer visualization, NLP, bio-informatics, marketing, electronic commerce, digitalized marketing, and robotics [3]. The Google trend for deep learning over the past five years is shown in Fig. 1.

In natural language processing (NLP), TC, likewise identified as a standard problem of NLP, is to disperse textual information into sentences with labels or tags. TC has a range of applications, including answering questions, detecting spam, analysis of sentiment, categorization of news, classification of user intent, and moderating content.

Data from text comes through various sources, like Web information, e-mails, unstructured chats, social web media, user reviews, insurance claims, and questions and answers from to name a few, customer services. The text is an extremely rich source of data. But the extraction of insights from its non-structured nature is difficult and time-consuming.

There could be various types of text analysis like Sentiment Analysis (SA). This includes analyzing the views of people from their text-oriented information (e.g., various discussion forums, reviews related to any product or place or things like movies, etc. or tweets) and obtaining their perspective. It is possible to cast the mining task as either a problem of binary classification or multi-classification. The investigation of bi-sentiment classification put the texts into either positive or negative classes, in the case of a multi-classification problem, texts bind with several predefined classes.

News labeling is another important category of text analysis. Digital news information is among the vital news sources. This helps to obtain knowledge of interest on day-to-day basics, e.g., deep information resides in the news, critics analysis, or endorsing related news is based upon the users spending time on different topics. Movie recommendation, natural language inference, topic analysis, and language analysis are other fields related to text analysis [3–5].

Deep learning (DL) is useful to solve many of the existing problems with greater accuracy. DL has proven its worth to classify; images data, and analyze patients' disease history through various graphical health reports, numeral data, spatial data, and textual data. While numeral data is somewhat straightforward, this is not the case with graphical and textual data. In this paper, the focus is to address the challenges

related to the analysis of textual data and to find out some optimal arrangements that could be helpful to refer to in further text-related professional projects as well as research projects.

To achieve this, various experiments have been conducted in terms of analyzing results by simulating textual data (Reuters) with different activation functions, analyzing results by simulating the different number of layers, analyzing results by simulating different optimizers, and analyzing results by simulating different Loss Functions.

2 Review of Literature

Different researchers [6–9] have explored different activation functions and emphasis that although many new activation functions are being developed the competency of earlier activation functions is also not rule out.

Optimizers [10, 11] offer a vital role in terms of deep learning model accuracy, for this, a few optimizers are quite important like adam, SDG. So, these are included in this paper exclusively for experimentation.

Loss functions are used to minimize the errors so that maximum accuracy can be achieved in a deep learning model [12, 13]. A few loss functions that are used and discussed in [12, 13] are used are categorical_crossentropy, mean_squared_error, and mean_squared_logarithmic_error [14, 15]. So, in this paper, these loss functions are included to achieve optimal accuracy. The summary of other important reviews is placed in Table 1.

3 Research Gap

The following are the research gap that is being noted in the literature review. As there are many activation functions, optimizers, loss functions, and other attributes. It is essential to perform extensive experiments to find the desired attributes to apply for text classification.

3.1 **Several layers**: To decide the number of layers is an important factor to get the optimal results using deep learning. The more complex problems might take high levels of layers to solve the problem.

3.2 **Several neurons**: The next part concerns the number of neurons that should be possessed by each layer as the neuron is nothing but the elementary processing unit.

3.3 **Feature for activation**: After the processing is done by a neuron, does this calculated processing output has worth in the final output or it will give an overall contribution misleadingly and needs to be discarded. Hence to reach this information, the activation function plays a crucial role. However, there

Table 1 Review summary of some important studies

S. No.	Title	Author	Findings	Knowledge gap	References
1	A review of medical textual question answering systems based on deep learning approaches	E Mutabazi, J Ni, G Tang, W Cao (2021)	The focus is on the various architecture of deep learning	In an architecture, how changing different variables can lead to more accuracy, is missing	[16]
2	Deep learning models for bankruptcy prediction using textual disclosures	F Mai, S Tian, C Lee, L Ma (2019)	Relu activation function is used with two layers of 100 neurons to predict financial textual data	More accuracy of result could be achieved by using newly developed activation functions like a swish	[17]
3	Embracing textual data analytics in auditing with deep learning	T Sun, MA Vasarhelyi (2018)	Discussed the methodologies for auditing textual data	No significant experiments are conducted. Discussed approaches are quite theoretical	[18]
4	A fast deep learning model for textual relevance in biomedical information retrieval	S Mohan, N Fiorini, S Kim, Z Lu (2018)	Delta approach is developed for biomedical textual analysis by using DRMM	The focus is on Statistical comparisons rather than experimenting to find optimal results	[19]
5	A survey on deep learning for textual emotion analysis in social networks. *Digital Communications and Networks*	S Peng, L Cao, Y Zhou, Z Ouyang, A Yang, X Li (2021)	Discussed many attributes, expect an emphasis on different activation functions	Although discussed for sigmoid and tanh activation functions a significant discussion for the optimizer, loss functions, number of neurons, and optimal layers is required	[20]

exist several activation functions currently that without a proper exploration, the optimal result might be far away even after using all other parameters correctly.

3.4 **Function decay**: The decay function gives a numeric value to change the value of the weights that need to be multiplied by the static value of the input.

3.5 **Algorithms of optimization**: Optimization functions help to reduce the overall cost of the model by adjusting the weight. There exist many optimizers such as SGD, Adam, Adadelta, Nadam, and a few others.

3.6 **Several epochs**: The number of epochs: during processing and batch size: Epochs refer to the number of iterations/simulations needed upon the same data set with the same environment again and again. If more epochs are chosen, the cost of the training model will increase unnecessarily, and choosing a smaller number of epochs might result in inefficient output [12, 13, 16–19].

4   Methodology

The research work is being conducted in the following fashion (Fig. 2):

   i. First, the area of research had been decided.
  ii. A thorough literature survey is conducted.
 iii. Based upon the literature survey, out of the total available variables (activation functions, optimizers, loss functions), a few variables are selected for experimentation purposes.
 iv. Coded the model using python, TensorFlow, and Keras.
  v. Analyze the result based on the experimental results and concluded the best workable environment for text classification.

During the investigation, it is being found that the intense experimentations are being needed to solve the problem of optimization of the environment. These experimentations is a challenge as by freezing all the environments, only one parameter needs to be changes in terms of activation function, optimizer, loss functions, number of layers.
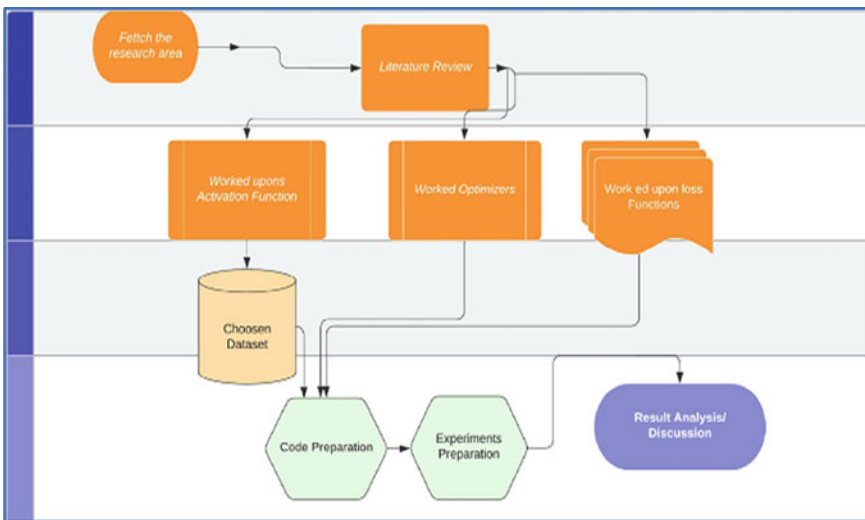
Fig. 2 Methodology

5 Analyzing Best Environment for Text Classification in Deep Learning

Reuters data set has been used to perform text classification using the concept of deep learning. The Reuters (21578) dataset is among the well-known and commonly used collections for TC [21]. the list of the activation function is large but in Table 1, a few important activation functions have been used to analyze the accuracy [22–24]. The experimentation is being conducted based upon various scenarios like analyzing results by simulating different activation functions, analyzing results by simulating different numbers of layers, analyzing results by simulating different optimizers and analyzing results by simulating different Loss Functions. In this analysis, categorical_crossentropy is represented by CC, Batch_size by BS, and Epochs by EP for making the table adequate and representative.

Table 2, shows the experiment result performed with different parameters and keeping these parameters constant and using different activation functions to find the best suitable activation functions upon text data.

Table 3 shows, how using a different number of layers can change the results and it is a myth that using many layers will give more accurate results and, in many cases, more layers only add the cost of training.

Table 4, displays the use of different optimizer functions and changes in the accuracy. It has been observed that for a text data set, adam optimizer should be a better choice.

Table 5 displays the analysis with various loss functions and changes noted in accuracy. For text data, it has been found that categorical_crossentropy is the desirable loss function to use. Further Fig. 3 shows the result of experimentation while using the Elu function.

The entire analysis depicts the accuracy depends upon several parameters and for text data swish activation function, categorical_crossentropy loss function, adam optimizer, and two layers are sufficient to obtain the optimal results.

Table 2 Analyzing results by simulating upon different activation functions

Loss function	CC	CC	CC	CC
Optimizer	Adam	Adam	Adam	Adam
BS	32	32	32	32
EP	2	2	2	2
Activation function (hidden layer)	RELU	SELU	ELU	SWISH
Activation function (outer layer)	Softmax	Softmax	Softmax	Softmax
No of layer	2	2	2	2
No of neurons	512	512	512	512
Results	93	94	94	94

Table 3 Analyzing results by simulating different numbers of layers

Loss function	CC	CC	CC	CC
Optimizer	Adam	Adam	Adam	Adam
BS	32	32	32	32
EP	2	2	2	2
Activation function (hidden layer)	Swish	Swish	Swish	Swish
Activation function (outer layer)	Softmax	Softmax	Softmax	Softmax
No of layer	1	2	3	4
No of neurons	512	512	512	512
Results	93	94	93	91

Table 4 Analyzing results by simulating different optimizers

Loss function	CC	CC	CC	CC
Optimizer	SGD	Adagard	Adam	Adadelta
BS	32	32	32	32
EP	2	2	2	2
Activation function (hidden layer)	Swish	Swish	Swish	Swish
Activation function (outer layer)	Softmax	Softmax	Softmax	Softmax
No of layer	2	2	2	2
No of neurons	512	512	512	512
Results	55	93	94	89

Table 5 Analyzing results by simulating different loss functions

Loss function	categorical_ crossentropy	mean_ squared_error	mean_squared_ logarithmic_error	hinge
Optimizer	Adam	Adam	Adam	Adam
BS	32	32	32	32
EP	2	2	2	2
Activation function (hidden layer)	SWISH	SWISH	SWISH	SWISH
Activation function (outer layer)	Softmax	Softmax	Softmax	Softmax
No of layer	2	2	2	2
No of neurons	512	512	512	512
Results	94	48	46	44

Fig. 3 Result of 'Elu' activation function with the accuracy of 94%

6   Conclusion

In this paper, analysis has been made on Reuters text data set with different environments of activation functions named: Swish, Selu, Relu, Elu, loss functions named: categorical_crossentropy, mean_squared_error, mean_squared_logarithmic_error, hinge, optimizers named: SGD, Adagard, Adam, Adadelta, and several layers. This analysis depicts the accuracy depends upon several parameters and for text data swish activation function, categorical_crossentropy loss function, adam optimizer, and two layers are sufficient to obtain the optimal results. However, for future expectations, this work can be simulated on other text datasets and more activation functions and loss functions can also be considered for experimentation.

References

1. Sharma A (2018) Understanding activation functions in neural networks. Retrieved from: https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0
2. Sharma.O (2019) Deep challenges associated with deep learning. In: Proceedings international conference machine learning big data cloud parallel computing (COMITCon), pp 72–75
3. Minaee S, Kalchbrenner N, Cambria E, Nikzad N, Chenaghlu M, Gao J (2020) Deep learning based text classification: a comprehensive review. arXiv preprint arXiv:2004.03705
4. Ranganathan H, Venkateswara H, Chakraborty S, Panchanathan S (2017) Deep active learning for image classification. In: 2017 IEEE international conference on image processing (ICIP). IEEE, pp 3934–3938
5. Janocha K, Czarnecki WM (2017) On loss functions for deep neural networks in classification. arXiv preprint arXiv:1702.05659
6. Rasamoelina AD, Adjailia F, Sinčák P (2020) A review of activation function for artificial neural network. In: 2020 IEEE 18th world symposium on applied machine intelligence and informatics (SAMI). IEEE, pp 281–286
7. Szandała T (2021) Review and comparison of commonly used activation functions for deep neural networks. In: Bio-inspired neurocomputing. Springer, Singapore, pp 203–224

8. Worked upon image data classification and shown to achieve greater results with swish activation function
9. Zhang X, Chang D, Qi W, Zhan Z (2021) A Study on different functionalities and performances among different activation functions across different ANNs for image classification. J Phys Conf Ser 1732(1):012026
10. Choi D, Shallue CJ, Nado Z, Lee J, Maddison CJ, Dahl GE (2019) On empirical comparisons of optimizers for deep learning. arXiv preprint arXiv:1910.05446
11. Zaman SM, Hasan MM, Sakline RI, Das D, Alam MA (2021) A comparative analysis of optimizers in recurrent neural networks for text classification. In: 2021 IEEE Asia-pacific conference on computer science and data engineering (CSDE). IEEE, pp 1–6
12. Xu J, Wang X, Feng B, Liu W (2020) Deep multi-metric learning for text-independent speaker verification. Neurocomputing 410:394–400
13. Gajowniczek K, Liang Y, Friedman T, Ząbkowski T, Van den Broeck G (2020) Semantic and generalized entropy loss functions for semi-supervised deep learning. Entropy 22(3):334
14. Sharma O (2019) A new activation function for deep neural network. In: 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon). IEEE, pp 84–86
15. Samek W, Binder A, Montavon G, Lapuschkin S, Müller KR (2016) Evaluating the visualization of what a deep neural network has learned. IEEE Trans Neural Netw Learn Syst 28(11):2660–2673
16. Mutabazi E, Ni J, Tang G, Cao W (2021) A review on medical textual question answering systems based on deep learning approaches. Appl Sci 11(12):5456
17. Mai F, Tian S, Lee C, Ma L (2019) Deep learning models for bankruptcy prediction using textual disclosures. Eur J Oper Res 274(2):743–758
18. Sun T, Vasarhelyi MA (2018) Embracing textual data analytics in auditing with deep learning. Int J Digit Account Res 18
19. Mohan S, Fiorini N, Kim S, Lu Z (2018) A fast deep learning model for textual relevance in biomedical information retrieval. In: Proceedings of the 2018 world wide web conference, pp 77–86
20. Peng S, Cao L, Zhou Y, Ouyang Z, Yang A, Li X, Yu S (2021) A survey on deep learning for textual emotion analysis in social networks. Digital Commun Netw
21. https://martin-thoma.com/nlp-reuters
22. https://www.tensorflow.org/
23. https://keras.io/activations/
24. https://www.datacamp.com/courses/deep-learning-in-python

Estimating Factors of Agile Software Development Using Fuzzy Logic: A Survey

Jahidul Hasan Antor, Sandhya Bansal, and Jamal

Abstract Agile software development is an iterative process that concentrates on producing a minimal viable product (MVP) quickly and then adjusting and adding features and capabilities in phases based on user feedback and behaviour. It is the most popular software development model nowadays. However, there exists various factors that plays a vital role for software development and are unknown, vague, and imprecision during the software development. Fuzzy logic is widely used for estimation of these factors in agile software development. In this paper, a survey on the use of fuzzy logic in agile software development for estimating various factors is presented. A comparative table of various research articles on different parameters is presented. Various related research questions are framed and finally answered. Overall, it is intended that the study will help advance knowledge creation and information accumulation in the area of agile software development using fuzzy logic by offering readers and researchers a road map to direct future research.

Keywords Fuzzy logic · Agile development · Software development · Parameter estimation

1 Introduction

Agile software development is the most versatile and widely utilized software development methodology. According to a 2005 poll of organizations in the United States and Europe, 14% of companies employed agile techniques, and 49% of those fully conscious of agile methods were eager to acquire them [1]. Even the best companies, including IBM, Cisco, Microsoft, and AT&T, adopt an agile initiative to enhance their operations because of its flexibility, predictability, and efficiency.

J. H. Antor (✉) · S. Bansal · Jamal
Department of Computer Science and Engineering, MMEC, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana, India
e-mail: antor4920@gmail.com

Fig. 1 Agile development
process

"Some organizations have their own custom software development procedures; however, the two most common ways are heavyweight and lightweight. Traditional software development is a heavyweight process, while agile modelling is a lightweight methodology" [2]. "The throughput of agile development to provide results rapidly and affordably on complex projects with inadequate needs is one of the crucial differences between agile and traditional development methodologies. Teams, functioning software, customer collaboration, and acclimating to change are prioritized in agile techniques, whereas contracts, plans, processes, paperwork, and tools are prioritized in traditional methods" [3]. The agile technique divides a project into smaller iterations or chunks to make it easier to manage and does not require long-term planning. It necessitates ongoing engagement with stakeholders and continual development at all stages. Once the work begins, teams go through the requirements analysis, planning, execution, quality assurance, and evaluation phase, as shown below in Fig. 1.

Agile methods are intended to facilitate software development vendors in rapidly developing and changing existing products and services, allowing them to react to competitive pressures [4]. The agile estimating method produces an expected output in agile development for delivering a project to a client within a specific time frame and set of requirements. This approach considers a few specific dominant factors (Fig. 2 shows some essential factors), which are uncertain and vague in nature. So, to estimate them, fuzzy logic is helping, e.g., the task size may be small, medium, or large; these small, medium, and large are fuzzy words. Similarly, task complexity, dynamic forces, and team experience have linguistic variables.

A fuzzy logic system is unusual, it is capable of simultaneously managing linguistic data and numerical data. Numbers are transformed into numbers using a nonlinear mapping from a data vector input to a scalar output [5]. Fuzzy logic is the specialized version of Boolean logic. True or false logic exists in Boolean sets. Though truth values for fuzzy logic, values may range from "0" to "1." Fuzzy logic comprises three major elements: the fuzzification process, inference from fuzzy rules, and the defuzzification process [6] (Fig. 3).
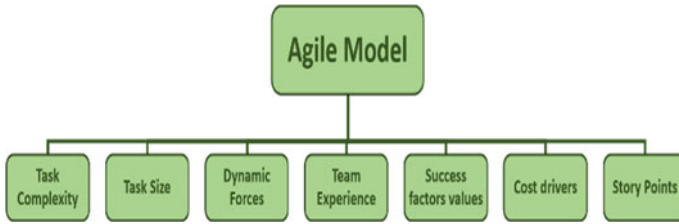
Fig. 2 Factors in agile model

Fig. 3 Fuzzy logic architecture
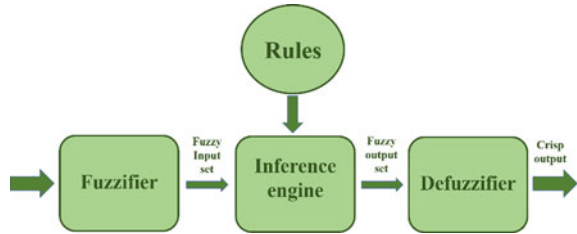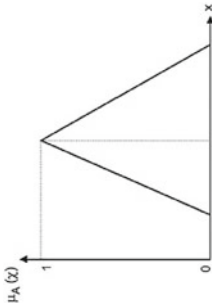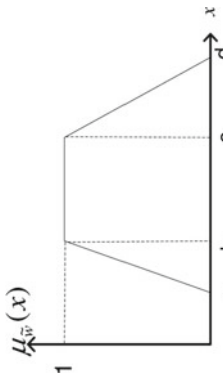
The fuzzification process turns practical or consistent values into membership grades and associated linguistic classes. Fuzzy rules are used to evaluate these fuzzified grades/classes for output grade classes. At last, in the defuzzification phase, these output grades are transformed back to real-world crisp output values using centroid computation [7]. The fuzzy logic system is based on fuzzy set theory and employs fuzzy factors to address ambiguity through the fuzzy inference system [8]. The core element of a fuzzy logic system that makes decisions is the fuzzy inference system. Mamdani and Takagi–Sugeno fuzzy model are two significant fuzzy inference system (FIS) approaches with various fuzzy rule consequences. "When calculating the centroid during defuzzification, the Mamdani technique takes the full fuzzy set into consideration. In contrast, the Sugeno method determines the weighted average of the singletons with the maximum membership value and the related x-point of the output gathered fuzzy set as the defuzzification output value"[7].

A membership function (MF) specifies how each data object is transformed to a membership value ranging from 0 to 1 [9].

Despite the great success and wide applicability of agile software development models, it is quite difficult to estimate the various parameters of the model which plays a crucial role during the software development. Fuzzy logic is widely used for the same purpose. In order to gain insight into the application of fuzzy logic in this context, this study seeks to present a systematic and thorough assessment of research articles published in the field of agile software development using fuzzy logic. The main goal of this work is to offer a thorough analysis of estimation of parameters of agile software development using fuzzy logic (Table 1).
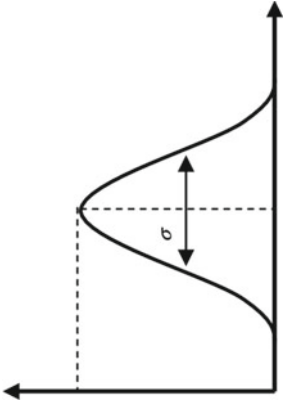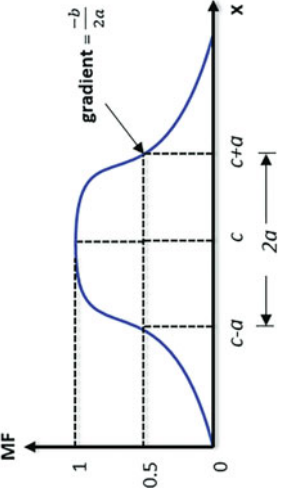
The rest of the paper is organized as follows: Sect. 2 presents the followed research methodology for selection of papers to be reviewed and research questions

Table 1 Available membership functions in graphical and equational form

Membership function (MF)	Equational representation	Graphical representation	Reference
Triangular MF	$\text{triangle}(x; a, b, c)$ $= \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ \frac{c-x}{c-b}, & b \le x \le c \\ 0, & c \le x \end{cases}$		[10]
Trapezoidal MF	$\text{trapezoid}(x; a, b, c, d)$ $= \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \frac{d-x}{d-c}, & c \le x \le d \\ 0, & d \le x \end{cases}$		[11]

(continued)

Table 1 (continued)

Membership function (MF)	Equational representation	Graphical representation	Reference
Gaussian MF	$\text{Gaussian}(x; \sigma, c)$ $= e^{-1/2\left(\frac{x-c}{\sigma}\right)^2}$		
Generalized bell MF	$\text{bell}(x; a, b, c)$ $= \dfrac{1}{1 + \left\| \frac{x-c}{a} \right\|^{2b}}$		[12]

Table 1 (continued)

Membership function (MF)	Equational representation	Graphical representation	Reference
Sigmoidal MF	Sigmoidal$(x; a, b)$ $$= \frac{1}{1 + e^{-\left[\frac{a}{x-c}\right]}}$$		

are framed. In Sect. 3, literature review along with comparative table of research articles is presented. Section 4 presents the results and discussion of the survey. Finally, Sect. 5 concludes the paper with future work.

2 Research Methodology

Because this is a review paper, the most relevant publications linked to fuzzy logic in agile software development must be chosen. For this goal, many search techniques are employed. The following review question has been the topic of this article:

- **RQ1**: Which fuzzy inference system (FIS) model and factors/parameters have been used by the researchers most in their application?
- **RQ2**: Which type of fuzzy membership function was maximum used in agile development?
- **RQ3**: What challenges are faced with using fuzzy logic in agile development?

The technique for selecting articles is depicted in Fig. 4. Firstly, keywords such as "fuzzy set theory and agile development," "fuzzy logic and agile development," "using fuzzy logic in agile development for effort estimation," and "uses of fuzzy logic in agile software development" were searched in the three most relevant databases: IEEE-Xplore, Google Scholar, and Springer. There was a total of 19,357 items retrieved. After that, a filter was done based on the removal of books, book chapters, and review articles, and a total of 15,237 items were chosen. Next, using a filtering method based on the quality check, appropriateness of topic, and relevance, this number is reduced to 3066. Then, applying a filtering method based on the title's keyword scan, the result becomes 1648. To facilitate more papers, we used other scanning methods based on title scans, articles, and research papers, and the number was reduced to 98. Finally, selection criteria were based on the removal of duplicates; as a result, the final 21 items for review were chosen.

3 Literature Review

Jasem et al. construct a fuzzy-based method that can enhance the Scrum framework's effort estimation. Story points are used to measure effort estimation, and fuzzifier, inference engine, and de-fuzzifier are the three components of the suggested model. The proposed model uses a trapezoidal fuzzy membership function to represent linguistic concepts, and this approach is the most efficient way to deal with complicated problems [8].

Assem et al. suggested a fuzzy-based paradigm for calculating agile software project success metrics. This article explains how to calculate the success metric value using the rates of success criteria and their significance values. The suggested

Fig. 4 Article selection procedure
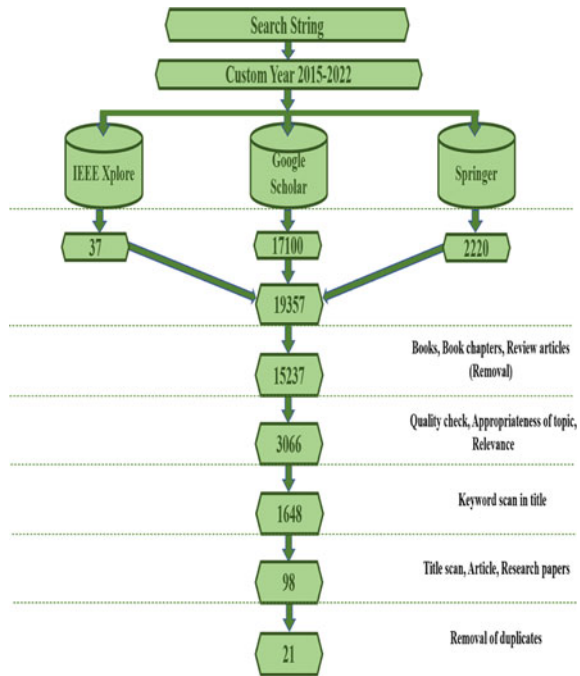
framework assists agile project stakeholders in representing the rates of success elements in a human-manner [13].

Raslan et al. developed an approach built on fuzzy logic that accepts fuzzy story points, application-level factors and dynamic forces as input parameters. This framework provides more accurate estimates than traditional effort estimating methodologies. In this approach, MATLAB is used to construct the suggested framework to prepare it for further trials with real datasets. This model aids in the calculation of appropriate effort estimates [14].

Rai et al. suggested a concept based on a fuzzy inference system to identify an optimal agile strategy for effective software development. In this framework, Mamdani FIS is utilized on a fuzzy rule basis, and the triangle membership function is used to represent linguistic words. The result shows that SCRUM and XP techniques are more prevalent in the industry. Still, FDD and DSDM methods are employed in certain situations, with DSDM being used for massive and safety–critical projects [15].

Tran demonstrated a practical fuzzy logic-based method for estimating productivity level inside the scaled agile model. A fuzzy logic-based system is used extensively in this unique strategy. During the manufacturers stage, agile software development teams replicate human thinking using an artificial intelligence technique. This special technique can assist software development teams in overcoming the challenge of simultaneously estimating various qualities [16].

Rola et al. provide a method for predicting the backlog products of scrum-based IT projects using ambiguous metrics and rigorous standards that take into consideration the human factor. This expert approach may estimate the time required to complete product backlog items [10].

Hamid et al. offer an expert strategy based on fuzzy logic for assessing the amount of effort in scrum projects. Fuzzy logic techniques are utilized to lessen the effects of various factors on work estimation for each sprint task. The suggested system's behaviour is tested using a MATLAB simulation. The results reveal that story point values can be simple or complicated based on the elements that influence task estimation [17]

Kumar Rai et al. describe the quality of design risk in agile software. Three risk indicators are established and utilized as input to a fuzzy logic MATLAB simulator-based model to evaluate design risk quality. The triangle membership function is used to represent linguistic terms. This article aims to focus on finding quality factors in agile software development and place more emphasis on quality metrics and evaluate them to deliver better results [18].

Dursun presents a multi-level hierarchical framework and fuzzy logic-based multi-criteria decision-making method for assessing agile providers. By allowing decision-makers to use linguistic concepts, the suggested decision-making approach lowers their mental burden throughout the evaluation process [19].

Dwivedi and Gupta show how an agile method engineering approach may create situation-specific methods for software companies based on the available agile methods ideas and practices. Fuzzy rules are employed to define the membership criteria of the best agile approaches, and the fuzzy logic controller is utilized to deal with the ambiguity of company goals. According to the study results, software organizations deconstruct methods and employ only a few of them to build projects [20].

Suresh and Patri examine the agility of universities clinic to propose a way for evaluating the agility of a health sector. They demonstrate how to evaluate a healthcare organization's agility, the characteristics that limit agile concert, and the proposed actions to increase its agility abilities. The case dispensary's agility has been measured using fuzzy logic, and the triangular relationship purpose is used to represent the verbal variables [21].

Atef et al. extend the agile principles to include the constructive cost model (COCOMO II), narrative points, and fuzzy logic models in an extended framework for agile project effort estimation. Initial iteration and constructive iteration are the two phases of the method that are being discussed. By providing genuine effort during the constructive iteration phase, the use of narrative points along with COCOMO factors may reduce the likelihood that a project would descend into chaos, and the application of fuzzy logic may improve the effort estimation. The results reveal that the suggested framework enhances prediction level (Pred) values to 80% [22].

Raslan et al. develop a model built on fuzzy logic to increase guesstimate correctness in agile development. The proposed model consists of three levels: input, processing, and output. The input layer of the fuzzy-SPEM contains the velocity and story size variables. The process layer transforms the velocity and tale size

variables into a fuzzy set, using a triangle MF. The output layer makes a decision using fuzzy inference and produces a crisp number for the projected velocity and story volume. Finally, the project work estimation is calculated based on the story's velocity and length [23].

Dragicevic et al. presented a Bayesian network model for estimating effort in agile initiatives. Their approach may be applied to any agile process, and they validated it using data from a single software business [24].

Sabour et al. proposes a more adaptable fuzzy set-based technique for human-like querying historical datasets based on the concepts of fuzzy sets and data-sensitive fuzzy sets. This method enables flexible time estimation and velocity based on scalable user profile elements in agile software development [25].

Saini et al. suggested a fuzzy logic approach for estimating expenses in agile software development. The fuzzy method is employed in this model to acquire raw data from actual figures, and the outcome is reported as the industry's estimated effort. When the data is unavailable for analysis using the conventional method, this alternative is preferred. The fundamental premise is derived from individuals whose acts and indecisions are neither apparent nor clear. This study will evaluate the effort outcomes using three input variables: user story, team experience, and difficulty. As a result, it isn't easy to define goals accurately, clearer, and more straightforwardly [26].

Semenkovich et al. represent a fuzzy-based extended scrum model. Their suggested methodology assesses user story points using a fuzzy inference system that incorporates team members' perspectives, such as the difficulty and size of work with various values. Overall, the results show that the strategy predicts more precisely for inexperienced scrum team members but is less accurate for seasoned scrum team members [27].

Sharma and Bawa developed different criteria decision-making approaches for selecting agile methodologies such as extreme programming (XP), etc. This technique combines fuzzy AHP and PROMETHEE to provide an experimental framework for evaluating agile systems based on the needs of a specific development. This research has several flaws, including a lack of criteria for assessing agile methodologies. This approach employs a modification of the fuzzy probe method [28].

Lin et al. employed fuzzy logic to overcome the uncertainty in agility evaluation. Identifying agility capacities, selecting linguistic elements to evaluate and comprehend the meanings of the linguistic phrases, fuzzy index labelling, and defuzzification are all part of the assessment procedures [29].

4 Results and Discussion

RQ1: Which fuzzy inference system (FIS) model and factors/parameters have been used by the researchers most in their application?

The following Table 2 represents an analytical view of the fuzzy inference system and factors/parameters.

RQ2: Which type of fuzzy membership function was maximum used in agile development? (Table 3).

**RQ3:** What challenges are faced with using fuzzy logic in agile development?

The decision-making process depends on fuzzy logic since the rules are pre-set. If the rules are faulty, the results may be unacceptably poor. One of the most challenging aspects of designing fuzzy systems is deciding on a membership function and fundamental limitations [30].

The use of fuzzy logic may be a challenge when validating system dependability and tuning membership functions. Fuzzy logic control does not scale well to big or complicated issues, as it only handles imprecision and vagueness, not uncertainty [30].

Significantly less paper has applied fuzzy on any real projects. [8].

5 Conclusion and Future Work

In this paper a survey on the use of fuzzy logic in agile software development model is presented. It shows the importance of parameter estimation in the field of agile software development. Various factors that affect the performance of software are estimated using fuzzy logic. A comparative table having parameters of estimation, fuzzy model used, highlighting the pros and cons of research article are presented. Various research questions are framed and answered through this table. Challenges using fuzzy logic in agile development is also presented. Future work includes a extending this review on using fuzzy logic in agile software development model for some real projects. Investigation into the scope of other soft computing techniques like neural network and evolutionary computation (swarm-based algorithms like particle swarm optimization (PSO) [31], bat swarm optimization, ant colony optimization (ACO) [32], bee colony optimization (BCO)) in improving the parameter estimation can also be studied.

Table 2 Uses of various factors and FIS model in agile development

NO	Purpose	Authors	Parameter	Model	Advantages	Disadvantages
1	More accurate, overtime, and gives a better effort estimation	Alostad et al. [8]	Team experience, task difficulty, amount of task, and estimation accurateness	Mamdani's model	It provides the most effective solution to complex issues	In the eXtreme programming (XP) technique, the suggested model cannot handle the effort estimating process
2	Improving the effort estimation accuracy	Raslan et al. [14]	Story points, application-level factor, friction factors, and dynamic forces	Mamdani's model	Help in producing an accurate estimation of effort	In this suggested framework, COCOMOII should be employed. The proposed framework can benefit from the use of COCOMOII
3	Estimate work effort by using a novel approach	Tran [16]	Complexity, amount of works, level of effort in story point	Mamdani's model	This suggested strategy has considerably stronger correlations with the typical human method of reasoning	This approach only can be used for some specific project
4	Risk assessment for the design quality of agile software	Rai et al. [18]	Correctness risk, maintainability risk, verifiability risk	Mamdani's model	The risk of agile software quality of design may be assessed by defining risk indicators	In this study, AI-based tools couldn't be applied to automate the agile software design risk assessment
5	Getting accurate effort estimation	Saini et al. [26]	User story points, team expertise, and complexity	Mamdani's model	This proposed model can estimate the effort by using a limited dataset	The neural network is not used in this suggested model. It is important for analysis on large amount of data

Table 2 (continued)

NO	Purpose	Authors	Parameter	Model	Advantages	Disadvantages
6	Improve the effort estimation accuracy	Raslan et al. [23]	Friction factors, dynamic forces, story points, implementation level factors, and complexity	Mamdani's model	This suggested methodology has the potential to increase the accuracy of effort estimation in agile projects	On real agile projects, fuzzy-SPEM is not used
7	Estimation of features based on linguistic terms	Semenkovich et al. [27]	The size and complexity of work	Mamdani's model	Convenient and more accessible in use than the "straight" story points estimation	This approach only can be used for scrum project
8	Increasing the value of the prediction level	Raslan et al. [22]	Scale factors, cost drivers, and KLOC; friction factors and dynamic force factors	Mamdani's model	It increases the value of the prediction level to 80%	Neuro-fuzzy approach and training algorithm should be used to improve the presented framework
9	Calculating success metrics for the agile development process	Mohammed et al. [13]	Success factors values, Importance value for each success factor value	TSK	It can evaluate the success factors values in human-like language	This framework isn't generic enough to handle all agile metrics
10	The evaluation of agile suppliers	Dursun et al. [19]	Delivery speed, flexibility, responsiveness	Mamdani's model	Enables the decision-makers to use linguistic terms	This proposed methodology doesn't employ subjective and objective weight assessments of the criteria and related sub-criteria

Table 3 Mostly used
membership function

Membership functions	References
Trapezoidal membership	[8, 10, 23, 24]
Triangular membership	[11–16, 18–20, 22, 25, 26]

References

1. Dybå T, Dingsøyr T (2005) About Agile software. IEEE Softw 26(5):6–9 (September–October 2009)
2. Islam AKZ, Ferworn DA (2020) A comparison between agile and traditional software development methodologies. Glob J Comput Sci Technol 7–42 (2020). https://doi.org/10.34257/gjcstcvol20is2pg7
3. Leau YB, Loo WK, Tham WY, & Tan SF (2012) Software development life cycle AGILE vs traditional approaches. In: International Conference on Information and Network Technology, vol 37, no. 1, IACSIT Press, Singapore, pp 162–167
4. Cao L, Mohan K, Xu P, Ramesh B (2009) A framework for adapting agile development methodologies. Eur J Inf Syst 18(4):332–343. https://doi.org/10.1057/ejis.2009.26
5. Mendel JM (1995) Fuzzy logic systems for engineering: a tutorial. Proc IEEE 83(3):345–377
6. Attarzadeh I, Ow SH (2009) Software development effort estimation based on a new fuzzy logic model. Int J Comput Theory Eng 473–476. https://doi.org/10.7763/ijcte.2009.v1.77
7. Jeganathan C (2003, March) Development of fuzzy logic architecture to assess sustainability of the forest management, ITC
8. Jasem M, Laila R, Sulaiman L (2017) A fuzzy based model for effort estimation in scrum projects. Int J Adv Comput Sci Appl 8(9):270–277. https://doi.org/10.14569/ijacsa.2017.080939
9. Bhatnagar R, Ghose MK, Bhattacharjee V (2011) Selection of defuzzification method for predicting the early stage software development effort using Mamdani FIS. Commun Comput Inf Sci 250 CCIS(4):375–381. https://doi.org/10.1007/978-3-642-25734-6_57
10. Rola P, Kuchta D (2019) Application of fuzzy sets to the expert estimation of Scrum-based projects. Symmetry 11(8). https://doi.org/10.3390/sym11081032
11. Kahraman C, Cebi S, Cevik Onar S, Oztaysi B, Tolga AC, Sari IU (eds) (2022) Intelligent and fuzzy techniques for emerging conditions and digital transformation, vol 307. https://doi.org/10.1007/978-3-030-85626-7
12. Chrysafis KA, Papadopoulos BK (2021) Decision making for project appraisal in uncertain environments: a fuzzy-possibilistic approach of the expanded NPV method. Symmetry 13(1):1–24. https://doi.org/10.3390/sym13010027
13. Assem H, Ramadan N (2016) A proposed fuzzy based framework for calculating success metrics of agile software projects. Int J Comput Appl 137(8):17–22. https://doi.org/10.5120/ijca2016908866
14. Tayh A, Nagy RD, Hefny HA (2015) Towards a fuzzy based framework for effort estimation in agile software development. IJCSIS Int J Comput Sci Inf Secur 13(1):37–45. http://sites.google.com/site/ijcsis/
15. Rai AK, Agarwal S, Kumar A (2018) A novel approach for agile software development methodology selection using fuzzy inference system. In: Proceedings of the international conference on smart systems and inventive technology, ICSSIT 2018, no. Icssit, pp 518–526. https://doi.org/10.1109/ICSSIT.2018.8748767
16. Tran HQ (2020) Software development effort estimation using a fuzzy logic-based system within the context of the scaled agile framework. IOSR J Comput Eng 22(1):10–19. https://doi.org/10.9790/0661-2201021019
17. Hamid M, Zeshan F, Ahmad A (2021) I. Conference, and undefined 2021. In: Fuzzy logic-based expert system for effort estimation in scrum projects. https://ieeexplore.ieee.org/abstract/document/9682239/. Accessed 23 Jul 2022

18. Rai AK (2021) Agile software quality of design risk assessment using fuzzy logic international journal of engineering research & management technology agile software quality of design risk assessment using fuzzy logic. December 2021
19. Dursun M (2017) A fuzzy MCDM framework based on fuzzy measure and fuzzy integral for agile supplier evaluation. AIP Conf Proc 1836. https://doi.org/10.1063/1.4982006
20. Dwivedi R, Gupta D (2017) The agile method engineering: applying fuzzy logic for evaluating and configuring agile methods in practice. Int. J. Comput. Aided Eng. Technol. 9(4):408–419. https://doi.org/10.1504/IJCAET.2017.086920
21. Suresh M, Patri R (2017) Agility assessment using fuzzy logic approach: a case of healthcare dispensary. BMC Health Serv Res 17(1):1–13. https://doi.org/10.1186/s12913-017-2332-y
22. Raslan AT, Darwish NR (2018) An enhanced framework for effort estimation of agile projects. Int J Intell Eng Syst 11(3):205–214. https://doi.org/10.22266/IJIES2018.0630.22
23. Raslan AT, Darwish NR, Hefny HA (2015) Effort Estimation in agile software projects using fuzzy logic and story points. December 2015. http://www.researchgate.net/publication/288839279
24. Dragicevic S, Celar S, Turic M (2017) Bayesian network model for task effort estimation in agile software development. J Syst Softw 127:109–119. https://doi.org/10.1016/j.jss.2017.01.027
25. Ramadan N, Sabour AA, Darwish NR (2018) Adaptive fuzzy query approach for measuring time estimation and velocity in agile software development information security view project agile software development view project adaptive fuzzy query approach for measuring time estimation and velocity in Ag. Researchgate.Net, no. February 2020. https://www.researchgate.net/publication/326000239
26. Saini A, Ahuja L, Khatri SK (2018) Effort estimation of agile development using fuzzy logic. In: 2018 7th international conference on reliability, Infocom technologies and optimization (trends and future directions) (ICRITO 2018), pp 779–783. https://doi.org/10.1109/ICRITO.2018.8748381
27. Semenkovich SA, Kolekonova OI, Degtiarev KY (2017) A modified scrum story points estimation method based on fuzzy logic approach. In: Proceedings of the institute for system programming of RAS, vol 29, no 5, pp 19–38. https://doi.org/10.15514/ispras-2017-29(5)-2
28. Sharma A, Bawa RK (2016) Modified fuzzy promethee approach for agile method selection using. I J C T a 9(41):641–649
29. Lin CT, Chiu H, Tseng YH (2006) Agility evaluation using fuzzy logic. Int J Prod Econ 101(2):353–368. https://doi.org/10.1016/j.ijpe.2005.01.011
30. Masoumi M, Hossani S, Dehghani F, Masoumi A (2020) The challenges and advantages of fuzzy systems application. Researchgate, May, pp 01–07. https://doi.org/10.13140/RG.2.2.22310.96328
31. Bansal S, Wadhawan S (2021) A hybrid of sine cosine and particle swarm optimization (HSPS) for solving heterogeneous fixed fleet vehicle routing problem. Int J Appl Metaheuristic Comput (IJAMC) 12(1):41–65
32. Bansal S, Goel R, Mohan C (2014) Use of ant colony system in solving vehicle routing problem with time window constraints. In: Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012). Springer, India, pp 39–50 December 28–30, 2012

Computational System Based on Machine Learning with Hybrid Security Technique to Classify Crime Offenses

Ankit Bansal, Vijay Anant Athavale, Kamal Saluja, Sunil Gupta, and Vinay Kukreja

Abstract Criminal investigation (CI) is a vital part of policing, with officers using a variety of traditional approaches to investigate crimes including robbery and assault. However, the strategies should be combined with the use of artificial intelligence to evaluate and determine different types of crimes in order to take real-time action. To combat cybercrime, new cyber security measures are needed. Conventional reputation frameworks perform poorly due to high management costs and also because of some other drawbacks such as false rate is high, more time taking process, and it is working on a limited number of data sources. This study proposes a hybrid approach based on intelligent dynamic malware analysis (IDMA) to address the issues raised above. The proposed approach uses advanced dynamic malware using decision tree classification machine learning (ML) algorithm. The proposed method emphases the risk scores (RS) as well as the confidence score (CS). At first, we compare ML algorithms to determine the good recall, precision and F-measure. As the next step, the comparison of complete reputation system with other reputation systems have been done. Using the confidential and risk scores, which are inversely proportional to each other, we can conclude that our findings are accurate. The risk score will automatically decrease as a result of an increase in confidence, reducing risk levels. The suggested architecture can not only be cross-checked with external sources, but it can also address security concerns that were previously overlooked by obsolete reputation engines.

A. Bansal · K. Saluja (✉) · S. Gupta · V. Kukreja
Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India
e-mail: kamal.saluja@chitkara.edu.in

A. Bansal
e-mail: ankit.bansal@chitkara.edu.in

V. Kukreja
e-mail: vinay.kukreja@chitkara.edu.in

V. A. Athavale
Walchand Institute of Technology, Solapur, Maharashtra, India

1 Introduction

IoT plays an important and effective role everywhere in the world. IoT collects a vast amount of data during a particular time period [1]. As a result, there are numerous challenges in data storage. A hybrid cyber security approach utilizing machine learning methods may be effective in combating IoT cyber threats.

Existing systems offer variety of options; however, due to a number of problems, reputation is being built for all types of objects [2]. Few criteria (such as risk and confidence) are investigated, demonstrating a thorough examination of existing techniques. Calculation of confidence is based on the pre-attack data. Those kinds of data are gathered from a variety of sources. A confidence score (CS) based on the information retrieved has been extracted. IP pre-attack data is gathered from a variety of sources, including net flow data and lots of similar sources are there for gathering IP's information. Based on this retrieved knowledge, a system provides a confidence score [3].

The fight between security experts and virus creators is never-ending. Malware behavior evolves at the same rate as technological advancements in the field of security. Advanced methods mainly focus the evolution process, which means malware patterns are changing as well. Machine learning (ML) approaches are used to detect and identify malware of this type. Security experts and specialists must constantly enhance their cyber defenses to keep alert of malware. Nowadays, machine learning is used extensively in cyber security to detect anomalies [4]. The severity is related to the previous performance. Furthermore, data sources that show an IP address's weighing factor contribute to the evaluation of measure of security [5].

The behavior of suspicious files is examined using machine learning techniques. Machine learning is used in reputation systems to identify patterns from data. ML algorithm is used for classification process. ML is also known as supervised learning. Decision tree (DT) is an effective classifier for classifying new malware, because it takes more probabilistic approach than support vector machine (SVM). SVM works well for datasets with limited features but tends to suffer when the number of key features is increased [6–8]. RS and CS must be assessed concurrently for reputation. Existing approaches do not take into account all of these factors at the same time. The Cuckoo sandbox evaluates malware associated with the identified IP addresses in real time [9]. The DT classifier is applied to malware samples to classify the behavior of zero-day malware. The risk of the threat is predicted by analyzing generated reports.

Malware families have been classified by the DT technique because of its comprehensive analysis nature and promising results. The DT method has a 95% success rate in predicting outcomes. We dynamically analyze the malware sample associated with an IP address to demonstrate the IP address's malicious behavior. Running a reputation calculation for every IP address gives the most current information.

The suggested framework is capable of detecting zero-day anomalies by analyzing malware behavior. It has been formed by considering the user having recent information with no additional maintenance cost. Upon reoccurrence in the system, calculation of reputation has been done and updated in real time and the risk score has been calculated. The proposed study attempts to increase confidence in the maliciousness of IP address by employing both internal and external sources.

2 Related Work

Existing systems provide firms with a variety of possible solutions; however, due to a number of flaws, reputation is being built. Risk score (RS) and confidence score (CS) has been computed, demonstrating a thorough examination of existing approaches. The CS for an IP address is calculated using pre-attack data [10]. To gather pre-attack IP information, net flow information, IP address location, and other sources are used. Based on the extracted knowledge, a module generates a CS. The RS, according to [11], is determined by the number of occurrences and previous incidents. Moreover, data sources that demonstrate an IP's weighing factor aid in the estimation of the RS for a range of IP addresses. A method for calculating an entity domain's approximate risk value has been proposed [12]. For the process of communication with network, endpoints were recorded to learn the behavior. Behavior-based approaches are preferable to signature-based techniques [13]. It is critical to detect domain names associated with command-and-control servers. Domain generation algorithms have fully checked areas used mostly by malicious software [14]. In general, risk computing measures blacklist hosts based on their association with specific malware for communicating with command-and-control servers [15–17].

Based on the user's activities, the demonstrated activities are classified and selected a RS. Within the entity, modifications with a variety of security products were discovered [18–20]. An attack is controlled automatically using security access points. A data mining is used for identifying website vulnerabilities and assigning them a numerical score [21–25]. Malicious code tracker is an application that detects malicious code hosted on suspicious websites.

3 Proposed Work

A hybrid method called IDMA utilizing machine learning in "IP Reputation for Forensics Data Analytics" for a well-known system has been proposed in this work. The proposed framework is depicted in Fig. 1. An IP address linked to several types of malwares can attack a machine multiple times. The IP address that assaults frequently is riskier than the one that hits infrequently.
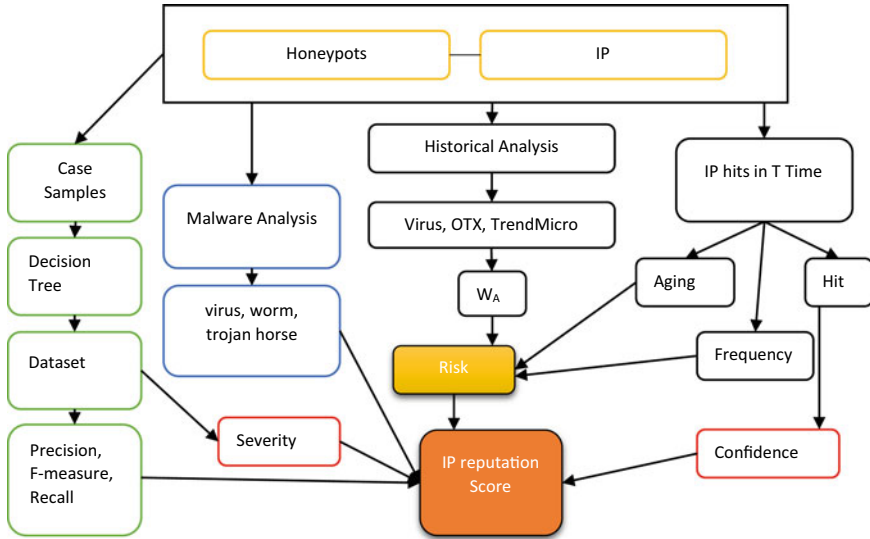
Fig. 1 Proposed IP reputation system

3.1 Preprocessing Phase

During this process, two honeypots have been used to detect malicious IPs and the attacks carried out by them. The IP address, the malicious binary file associated with it, the activation time, and the number of attempts is all extracted from an IP address's activation period. All these characteristics work together to help us determine the reputation of an IP address.

3.2 Malware Analysis

It is the process of determining the source and primary goal of a binary file (backdoor, worm or Trojan horse, virus, etc.). In general, there are two methods for assessing malware samples namely static and dynamic. In the present work, advance dynamic malware detection has been used.

3.3 Observed Behavior

The IP is observed by the following two behaviors.

Frequency. It is noted by taking the count of the attacks being carried out by an IP address within a particular period of time.

Aging. It denotes the time gap between two IP address occurrences. Both, internal and external sources are used.

3.4 Historical Analysis

For the purpose of calculating the risk and confidence value the history can be analyzed with the help of external sources as given in the Eq. (1).

$$\text{Ext.sources} = \left(\text{VT} * w_1'\right) + \left(\text{OTX} * w_2'\right) + \left(\text{MyIP} * w_3'\right) \tag{1}$$

where VT represents Virus Total which comes under the external source; w_1', w_2' and w_3' are the weights and the result is 1 or 0 (binary form).

3.5 Risk Score

Risk score is given as the combination of all the evaluated scores and it is represented by using the Eq. (2)

$$\text{RiskScore}(R.S) = (\propto (\text{Int.sources}) + \beta(\text{Ext.sources}))/((\propto +\beta)) \tag{2}$$

where,

$$\text{Int.sources} = \left[(\text{sev} * w_1) + (\text{freq} * w_2) + (\text{ag} * w_3)\right] \tag{3}$$

$$\text{Ext.sources} = \left[\left(\text{VT} * \left(w_1'\right)\right) + \left(\text{OTX} * \left(w_2'\right)\right) + \left(\text{MyIP} * \left(w_3'\right)\right)\right] \tag{4}$$

$$R.S. = \left(\alpha(S * w_1) + (F * w_2) + (A * w_3) + \beta\left(\text{VT} * w_1'\right) \\ + \left(\text{OTX} * w_2'\right) + \left(\text{MyI} * w_3'\right)\right)/(\alpha + \beta) \tag{5}$$

Here, sev represents the severity of the attack; freq represents frequency of the attacks; ag is aging, w_s represents weight. The values for parameters in the equations are evaluated as follows:

$\alpha = 86.5$;
$\beta = 11.4$;
MyIP $= [0, 1]$;
VT $= [0, 1]$;
OTX $= [0, 1]$;
$w_1 = 0.624$
$w_2 = 0.228$
$w_3 = 0.418$
$w_1' = 0.548$
$w_2' = 0.33$
$w_3' = 0.112$

Algorithm 1 Computation of risk score for discovering threat level from an IP address

1. **Input:**

   a. IP
   b. A malicious file MF

2. **Output:** Risk associated with an IP
3. **Procedure**
4. send IP to virusTotal
5. **if** IP exists **then**
6. VT $= 1$
7. **endif**
8. **else**
9. VT $= 0$
10. send IP to OpenThreatExchange
11. **if** IP exists **then**
12. OTX $= 1$
13. **Endif**
14. **else**
15. OTX $= 0$
16. send IP to MyIP
17. **if** IP exists **then**
18. MyI $= 1$
19. **endif**
20. **else**
21. Myl $= 0$
22. count Occurences
23. F $=$ Number of Occurrences/Unit Time
24. calculate Aging
25. A $=$ Aging (T_c, T_p)

26. calculate Severity
27. S = Severity(MF)
28. Risk = $\dfrac{\propto(S*w_1)+(F*w_2)+(A*w_3)+\beta(VT*w_1')+(OTX*w_2')+(MyIP*w_3')}{(\alpha+\beta)}$

A risk score indicates how dangerous an IP address is. An organization needs to determine where to communicate and what kind of privileges should be granted to the users based on the calculated risk score.

3.6 Confidence Level

We analyze the level of confidence in our reputation system after computing the risk score. To reflect assurance, a confidence level is calculated using data from other sources. We use the following formula to get the confidence level:

$$\text{Confidence level(CL)} = \left[\text{hit} + \text{hit} * (VT + OTX + MyIP)\right] \tag{6}$$

where,

Hit = 25%;
   All values other then Hit is same as the previous risk value.

Algorithm 2 Computation of confidence level for reflecting behavior of an IP address

1. **Input:** IP
2. **Output:** Confidence Score
3. **Procedure**
4. Send IP to virusTotal
5. **if** IP exists **then**
6. VT = 1
7. **endif**
8. **else**
9. VT = 0
10. send IP to OpenThreatExchange
11. **if** IP exists **then**
12. OTX = 1
13. **endif**
14. **else**
15. OTX = 0
16. send IP to MyIP
17. if IP exists then
18. MyI = 1
19. endif
20. else
21. MyI = 0

22. Confidence = (Hit * Internal) + VT * Hit_v + OTX *Hit_0 + MyIP * Hit_M
23. return Confidence

For illustration, if an IP address is found in the VT database, the value of VT is 1 or else 0. As same as this all other external source has processed. As shown in Algorithm 2, this is how we ensure the reputation of an IP address.

3.7 Performance Evaluation

In this work, we collected two datasets consisting of network and extended malware samples. These samples were collected from VirusShare [26]. The evaluation has been carried out by applying DT on dataset obtained from cuckoo sandbox that we have considered for this study. Precision, recall and F-measure have been considered to carry out the performance evaluation.

4 Experimental Results

Experimental results on network and extended dataset as given in Table 1.

The malwares depicted in the report with a name are already known to antivirus software, whereas those with unknown designations are needed to be considered. Because these unknown binaries are new and have no signature, antivirus software does not label them as malwares.

Dataset DT1 is the network dataset and it has contents about network activities. Dataset DT2 is the extended dataset and it contains malicious information.

4.1 Experimental Calculation of the Risk Score

The risk score that has been calculated for different aging is given in Table 2. For the calculation, internal and external sources are used as inputs.

Table 2 gives that the RS for IP addresses has been updated. Its risk score had previously been calculated, but because of the repeated attempt the weight of the RS is evaluated.

Table 1 Performance of DT in dataset 1 and dataset 2

Dataset	F-measure (%)	Precision (%)	Recall (%)
DT1	77	74	81
DT2	99.2	100	98.2

Table 2 Weighted risk score

Aging (mm/dd/yyyy)	Risk score frequency
09/09/2019	12.003
09/20/2019	5.069
09/28/2019	4.845
10/09/2019	0
10/19/2019	0
10/29/2019	0
11/16/2019	5.126
11/21/2019	4.85
11/23/2019	5.182
12/23/2019	4.843
12/14/2019	28.861
12/15/2019	40.779
01/11/2020	39.532
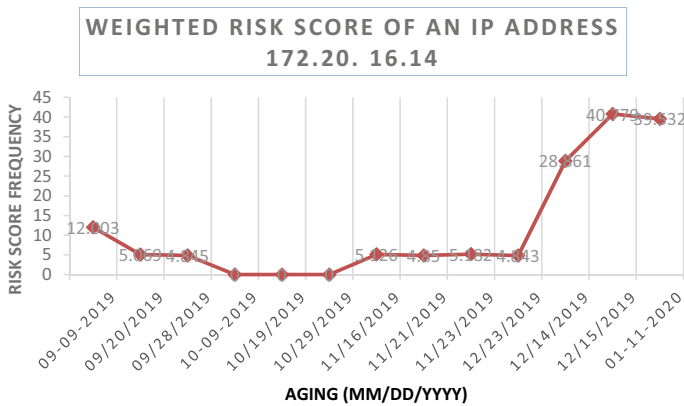
Fig. 2 Graphical representation of risk score

A weighted RS is considered as shown in Fig. 2. To save storage space, we update an IP's risk value rather than appending its RS to the database multiple times. We obtain efficiency by doing so since we reuse the same memory addresses when reactivating certain IP addresses.

4.2 Confidence Level Comparison

To find the confidence level we consider four IPs as given in Table 3 and the confidence level is compared as shown in Fig. 2.

Table 3 IP addresses with its confidence level

	IP 1	IP 2	IP 3	IP 4	IP5
IP num	213.186 33.3	61.142.1 73.4	202.103 218.88	172.201 6.4	93.178.3 229
Confidence level	100	25	75	25	75
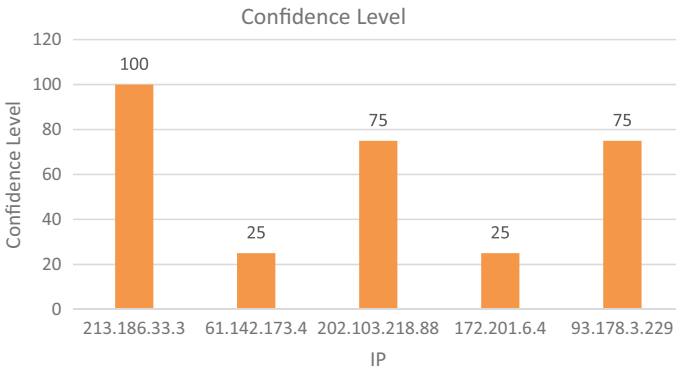
Fig. 3 Graphical representation of confidence level of IP addresses

To protect ourselves from cyber dangers, we determined the risk score and confidence level. Figure 3 shows the comparison of confidence level between different IP addresses. As per Fig. 3 the confidence level of IP 1 is better than all other four IP addresses.

When a reoccurrence occurred, the reputation is calculated and updated in real time, and the risk score is calculated. Using both internal and external sources, the present study aims to boost confidence in the maliciousness of IP address.

5 Conclusion and Future Work

By analyzing malware activity, the proposed system can identify zero-day anomalies. It has created with the idea of providing users with current information at no additional cost. On the dataset, the different ML approaches are examined, and we find that DT performs best when categorizing unknown malware samples. The prediction of performance has been done with the help of DT and the accuracy range of DT is 95%. A network dataset comprised of network activity yielded an accuracy rate of 77% with the help of DT1, while a malicious information-rich dataset yielded an accuracy rate of 99% with DT2. In this process, we have to deal with large data during the evaluation of risk and confidence level. The risk score and the confidence level have been calculated to protect ourselves from cyber threats. In future, the IP addresses

with a negative reputation will strengthen the system's security against any type of suspicious behavior and hence it is comfortably immersed in antivirus.

References

1. Usman N, Usman S, Khan F, Jan MA, Sajid A, Alazab M, Watters P (2021) Intelligent dynamic malware detection using machine learning in ip reputation for forensics data analytics. Futur Gener Comput Syst 118:124–141. https://doi.org/10.1016/j.future.2021.01.004
2. Kabanda PG (2021, May) Performance of machine learning and big data analytics paradigms in cybersecurity and cloud computing platforms. Glob J Comput Sci Technol 21(G2):1–25
3. Berman DS, Buczak AL, Chavis JS, Corbett CL (2019) Survey of deep learning methods for cyber security. Information 10:122. https://doi.org/10.3390/info10040122
4. Sarker IH, Kayes ASM, Badsha S, Alqahtani H, Watters P, Ng A (2020) Cybersecurity data science: an overview from machine learning perspective. J Big Data. https://doi.org/10.1186/s40537-020-00318-5
5. Smith R, Marck S, Assignee Inventors, Level 3 Communications LLC (2018) Identifying a potential DDOS attack using statistical analysis. United States patent US 9900344
6. Yanovsky B, Eikenberry S, Assignee Inventors, SonicWALL Inc (2019) Reputation based threat protection. United States patent US 10326779
7. Deepika C, Ashok K, Pradeep B, Vijay AA, Veeraiah D, Boppuru RP (2021) An effective face recognition system based on Cloud based IoT with a deep learning model. Microprocess Microsyst 81:103726. ISSN 0141-9331. https://doi.org/10.1016/j.micpro.2020.103726
8. Athavale VA, Gupta SC, Kumar D, Savita (2021) Human action recognition using CNN-SVM model. Adv Sci Technol 105:282–290. https://doi.org/10.4028/www.scientific.net/ast.105.282
9. Sihwail R, Omar K, Zainol A, Khairul A (2018) A survey on malware analysis techniques: static, dynamic, hybrid and memory analysis 8:1662. https://doi.org/10.18517/ijaseit.8.4-2.6827
10. Wang L, Wang B, Liu J, Miao Q, Zhang J (2019, March) Cuckoo-based malware dynamic analysis. Int J Perform Eng 15(3):772–781
11. Hillard DL, Munson A, Cayton L, Golder S, Inventors eSentire Inc Assignee (2019) System and method for determining network security threats. United States patent US 10412111
12. Ray KD, Reed SN, Harris MD, Watkiss NR, Thomas AJ, Cook RW, Samosseiko D, Assignee Inventors, Sophos Ltd (2018) Using indications of compro mise for reputation-based network security. United States patent US 9992228
13. Li Y, Liu F, Du F, Zhang D (2018) A simhash-based integrative features extraction algorithm for malware detection. Algorithms 11(8):124
14. Kruegel C, Bilge L, Kirda E, Balduzzi M (2019) Exposure: finding malicious domains using passive DNS analysis. In: Proceedings of 18th network and distributed system security symposium. NDSS'11, pp 214–231
15. Joo JU, Shin I, Kim M (2015) Efficient methods to trigger adversarial behaviors from malware during virtual execution in sandbox. Int J Secur Appl 9(1):369–376
16. Sharifnya R, Abadi M (2013) A novel reputation system to detect dga-based botnets. In: Computer and knowledge engineering (ICCKE), 3rd international e-conference. IEEE, pp 417–423
17. Coskun B (2017) (Un)wisdom of crowds: accurately spotting malicious IP clusters using not-so-accurate IP blacklists. IEEE Trans Inf Foren Secur 12(6):1406–1417
18. Mittal S et al (2022) Using identity-based cryptography as a foundation for an effective and secure cloud model for e-health. Comput Intell Neurosci 2022:7016554
19. Kaur K, Verma S, Bansal A (2021) IOT big data analytics in healthcare: benefits and challenges. In: 2021 6th international conference on signal processing, computing and control (ISPCC), pp 176–181. https://doi.org/10.1109/ISPCC53510.2021.9609501

20. Athavale VA, Bansal A, Nalajala S, Aurelia S (2020) Integration of block-chain and IoT for data storage and management. Mater Today Proc. https://doi.org/10.1016/j.matpr.2020.09.643
21. Bansal A, Athavale VA (2021) Big data and analytics in higher educational institutions. Lecture notes in networks and systems, vol 140
22. Mann KS, Bansal A (2014) HIS integration systems using modality worklist and DICOM. Proc Comput Sci 37:16–23
23. Kukreja V, Kumar D, Bansal A, Solanki V (2022) Recognizing wheat aphid disease using a novel parallel real-time technique based on mask scoring RCNN. In: 2022 2nd international conference on advance computing and innovative technologies in engineering (ICACITE), pp 1372–1377. https://doi.org/10.1109/ICACITE53722.2022.9823459
24. Saluja K, Bansal A, Vajpaye A, Gupta S, Anand A (2022) Efficient bag of deep visual words based features to classify CRC images for colorectal tumor diagnosis. In: 2022 2nd international conference on advance computing and innovative technologies in engineering (ICACITE), pp 1814–1818. https://doi.org/10.1109/ICACITE53722.2022.9823727
25. Katare A, Athavale VA (2011) Behaviour analysis of different decision tree algorithms. Int J Comput Technol Electron Eng 1:143–147
26. https://virusshare.comlast. Accessed 29 May 2022

Analysis of Robotically Controlled Percutaneous Needle Insertion into Ex Vivo Kidney Tissue for Minimally Invasive Percutaneous Nephrolithotomy (PCNL) Surgery

Ranjit Barua, Sumit Bhowmik, Arghya Dey, Surajit Das, and Sudipto Datta

Abstract Percutaneous nephrolithotomy or PCNL is a standard renal microsurgery method in modern urology. However, this is a challenging discipline for urologists throughout the surgical procedure. They should accomplish a manner with limited sensible and flexibility of invasive apparatuses. Presently in modern medical science especially in surgery, minimally invasive procedures achieved huge popularity where needle insertion plays an important part. In this experimental investigation, we have investigated the facilities and investigational implication of a bevel geometrical hypodermic needle piloting over studies in ex vivo duck kidneys, and also, evaluated the cutting forces, particularly the force developed in the axial direction during the insertion process. The fresh ex vivo duck kidney was used within 12 h of death as experimental material. The test data were saved by an adjusted measurement instrument (force). A set of insertion analyses with individual insertion speeds (15, 35, and 5 mm/s) in combination with different frequencies of vibration (450, 45, and 110 Hz) was deliberated for assessment to observe the impacts of these two factors on ex vivo perforation force. It is observed that the insertion force at 450 Hz gets up and around 19.3% is linked with the force at 45 Hz. In both the robotically and manual insertion methods, the force increases gradually, but most importantly, in the robotic insertion method, the nature of the insertion force is steady, and in the case of the manual insertion method, the nature of the insertion force fluctuates rapidly.

Keywords MIS · Ex vivo · Needle · PCNL · Surgery · Kidney

R. Barua (✉) · S. Bhowmik · A. Dey
OmDayal Group of Institutions, Howrah, West Bengal 711316, India
e-mail: ranjitjgec007@gmail.com

S. Das
R G Kar Medical College and Hospital, Calcutta 700004, India

S. Datta
Indian Institute of Engineering Science and Technology, Shibpur, Howrah 711103, India

1   Introduction

In the existing modern urology surgical area, minimally invasive surgery (MIS) is accepted due to the active features of expertise, negligible surgical wound and bleeding, and faster recovery rate [1]. In modern urology, PCNL (Fig. 1) is a popular method for treating renal stone surgical treatment [2–4]. At present, large numbers of researchers and urologist have examined the surgical hypodermic needle insert mechanism and intended a number of biomechanical prototypes and models constructed on the simultaneous trial records. The needle steer control described in the allied study was considered to accomplish particular inset of the urological needle into the particular spot of soft muscles. Clinical needles (specifically flexible) are inserted via the soft tissue to execute the collecting blood, drug delivery, perform local anesthetic managements, etc. [5–7]. The achievement of a needle addition in therapeutic events is enormously dependent on the precision of the perforation practice [8]. The core important aspects distressing needle errors are deflection and deformation of material specimen [9, 10]. These complications are unswervingly related to the cutting force and possessions on concentrated piercing force and bending of needle slant from the real route [11, 12]. To decrease the inaccuracies of needle supplement and material damage [13, 14], a cooperative process is to decline puncture force [15]. Decreasing damage force may possibly help with minutest pain and a lesser recovery interval [16]. The particular needle positioned into the kidney is very challenging and essential stage for operative "nephrolithotomy" preparation [17, 18]. Several approaches and functional measures have been elevated to efficiently acquire and develop the appropriate renal technique [19, 20]. It is defined significant in urology as it extents a huge range of renal stone or calculi surgery [21]. It will support tremendously in improving the puncture into the particular locus of the kidney. Right now, the excellence and exactness of clinical needle perforation method are affected by totally the proficiency or ability of the specialist as puncture or attachment into an essential organ has generally been done physically. An in vivo experimentation was arranged to consider the material belongings and outline the puncturing force analysis [22]. For investigation a cutting force, a sequence of insertion trials was completed [23], where the highest puncture force of different organs all over the addition into the viscoelastic material was defined. Another technique [24] was recognized to calculate the operational puncture forces with a smaller amount of cutting speed of needle and inspected the demonstrating and the boundary between the clinical hypodermic needle and biological tissue over the in vivo examinations. Additionally, researchers also proven a simulation technique for insertion experiment and distinguished the viscoelastic features of the in vivo research.

In our former paper [25], a PVA gel was used as a tissue mimic experimental material fora insertion test, where we calculated the frictional stress physical appearance of the gel, and needle bend incident, which is very major factor for invasive process. Here, a robotically and manually hypodermic urological needle insertion test was performed on the ex vivo kidney (duck), and furthermore matched the tentative actual result with the computational outcome. The purpose of this analysis is to
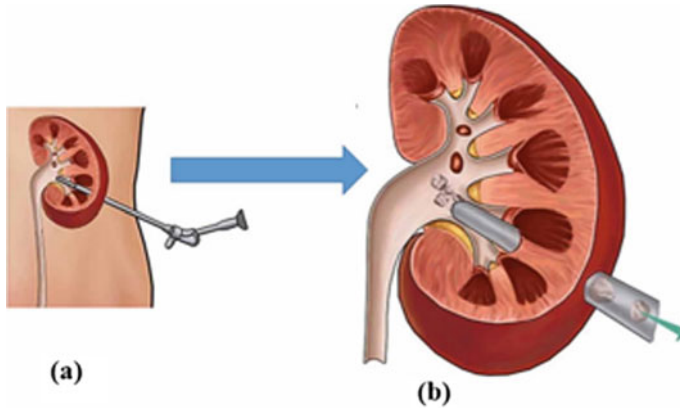
Fig. 1 **a** PCNL practice; **b** kidney stone elimination by PCNL process [3]

determine needle perforation forces during the course of the percutaneous practice and to explore the influences of speed and vibration rate on perforation force.

2   Mathematical Analysis of Percutaneous Force Model

The ex vivo needle insertion investigation was examined in this paper calculates the percutaneous force with inset distance. In numerous insertion techniques, it was observed that the cutting force alters promptly in an uneven manner as dissimilar anatomical structure of organs are being pierced or injected. Needle puncture force is the entirety aggregate of friction force, cutting force, and stiffness force [4].

$$F_I = f_s(Z) + f_F(Z) + f_C(Z) \tag{1}$$

[Z = Direction of the needle placing].

In the beginning, as the needle insets the outer surface of the ex vivo kidney during the experiment, a resistance force takes place which is also known as the stiffness force, and this force is depends on insert penetration length of the needle inside the ex vivo kidney (d_i).

So,

$$f_S(Z) = 2d_r \acute{r}_r d_i \tag{2}$$

[d_r = tissue deformation radius of during experiment.]

$\acute{r}_r$ is reduced modulus which arises when needle perforates the ex vivo kidney during the experiment procedure.

$$\acute{r}_r = \frac{1 - \upsilon_k^2}{\varepsilon_T} + \frac{1 - \upsilon_n^2}{\varepsilon_N} \tag{3}$$

$[\upsilon_k, \upsilon_n = \text{poissionratioofex vivokidneyandhypodermicneedle};$
$¥_k, ¥_n = \text{Young'smodulusofex vivokidneyandhypodermicneedle}]$

$$f_F = \mu F_R \tag{4}$$

$[\mu = \text{friction coefficient between the ex vivokidneyandhypodermicneedle, and} F_R$
is normal force.

Let, f_C is constant at stable conditions.
So,

$$F_I = \begin{cases} f_S & Z_I \leq Z \leq Z_{II} \\ f_F + f_C = \mu F_R + \text{Const.} & Z_{II} \leq Z \leq Z_{III} \\ f_F = \mu F_R & Z_{III} \leq Z \leq Z_{IV} \end{cases} \tag{5}$$

3 Materials and Procedures

In this ex vivo insertion study, a bevel hypodermic needle was used, and two altered trial methods were performed to check the ex vivo needle inset prototypical and evaluate the impacts of different speeds, and vibration frequency on puncture force; for example, the needle was horizontal straightly inserted into the ex vivo kidney at three different speeds (15, 35, and 5 mm/s), and the alternative test was performed at three dissimilar vibrations (450, 45, and 110 Hz) at 2.57 µm vibration amplitude (a const. speed of 15 mm/s). Sequences of insertion investigations (six times) with comparable constraints were done to save investigational records.

3.1 The Ex Vivo Experimental System

Figure 2a shows the ex vivo investigational arrangement. We have prepared a robotically controlled (2 DOF) insertion device attached with a 28G hypodermic needle (30° bevel tip) (Fig. 2b) which helps in the examination of the ex vivo needle insertion test. We have used a fresh duck kidney as the ex vivo investigational tissue material for this analysis (Fig. 2c), which was brought from a local market (meat shop), and during earlier the investigation, it was stored in the fridge for 14 h. At the time of the experiment, the ex vivo duck kidney was positioned in front of the investigational hypodermic needle. The ex vivo experimentations were set at the CHST, IIEST-Shibpur, in a biomedical engineering lab under the observation of a senior surgeon
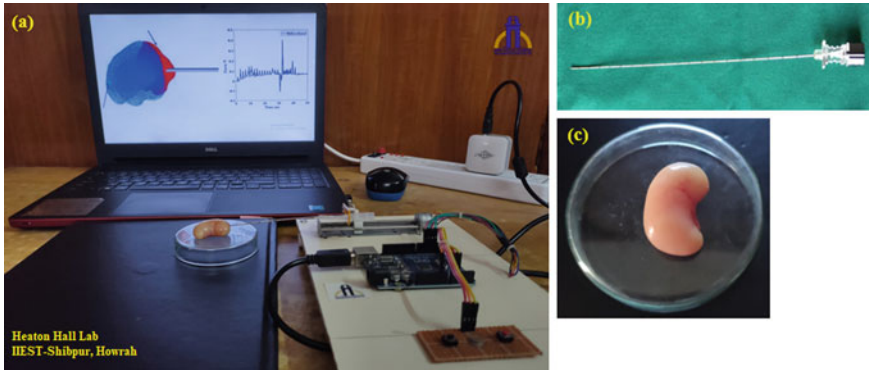
Fig. 2 **a** Ex vivo experimental arrangement of PCNL; **b** experimental hypodermic bevel needle; and **c** ex vivo experimental kidney

(urologist), a post-doctoral fellow, a biomechanical engineer, and an electrical engineer. Herein ex vivo investigation, even though the investigational needle perforated and passes from end to end in the investigational ex vivo tissue, an F/T sensor (6 DOF) [BHI160B, Bosch Sensortec's] was assisted to gather the experimental data of forces and torques. A data acquisition (NI 779205-01, DAQ, National Instrument Corp.) was employed to adapt the investigational data from analog data into digital data on the PC display. A piezo-actuator was used to generate the vibration, which was fastened with the exp. needle and the amplitude of the vibration alter from 0 to 130 μm.

4 Results and Discussion

Figure 3 displays that percutaneous force in ex vivo hypodermic surgical needles rises with different insertion speeds for every insertion experiment (specifically 15, 35, and 5 mm/s). Additionally, it is seen that in both precipitates, percutaneous force rises between 5 and 45 mm of the depth of insertion. In this study, nonlinear viscoelastic properties have been observed in the experimental ex vivo kidney tissue. During the insertion procedure, hypodermic needle insertion speed was determined and the ultimate cutting force was identified. With the speed and depth of the surgical needle insertion, the percutaneous force of the needle increases. Thus, the percutaneous force gets influenced by needle insertion speed at the time of minimal invasive PCNL surgery.

Figure 4 confirms the importance of variation in the frequencies of vibration on the percutaneous force of a hypodermic surgical needle at a fixed amplitude of 2.57 μm. The maximum cutting force was identified at the frequency of 450 Hz. As the result of ongoing insertion speed and amplitude of vibration, the slope of the hypodermic needle percutaneous force was the same.

Fig. 3 Ex vivo needle
percutaneous force with
different insertion speeds:
15, 35, and 5 mm/s
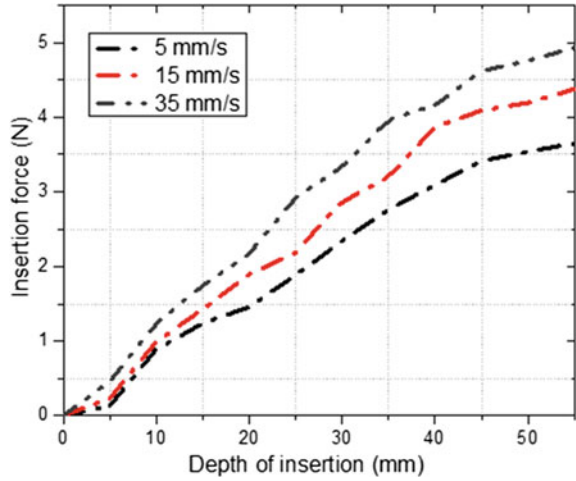
Fig. 4 Influences of
dissimilar vibration
frequencies on percutaneous
force
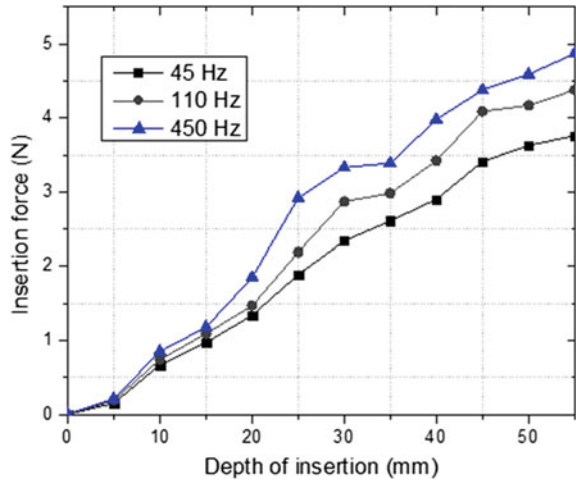
Figure 5 illustrates the outcome of the robotically controlled and manually inser-
tion. During the needle insertion practice, in both the insertion process the percuta-
neous force was equivalent and a steady rise was observed. Fluctuations in the force
of the needle insertion were noticed in manual insertion, as keeping constant speed
is not possible in this particular procedure. But due to constant speed, the insertion
force is maintained in robotically controlled insertion.

With the help of simulation software named AbaqusDassaultSystèmes®, the
simulation analysis of hypodermic needle insertion was experimented on in this
ex vivo investigation. As the soft biological tissue of the ex vivo kidney for the
experiment contains viscoelastic properties, we accept it to be nonlinear viscoelastic
properties. Figure 6 confirms that in both simulation and investigation, cutting force

Fig. 5 Cutting force profile of robotically and manually controlled experiment
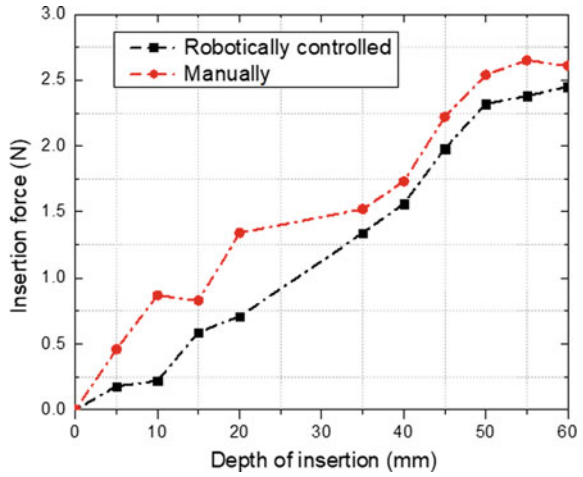
Fig. 6 Insertion force profile of exp. and simulation investigation
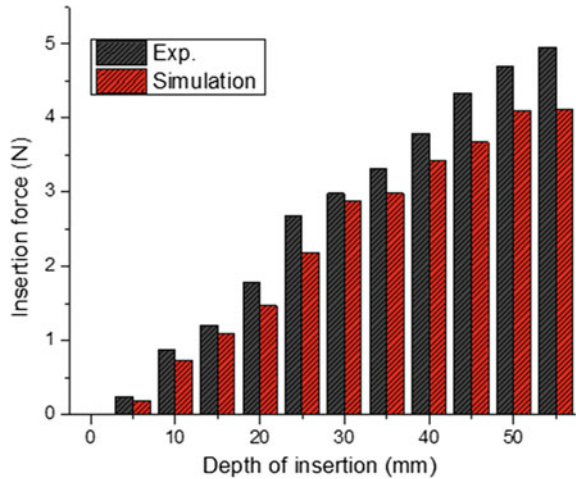
is rising simultaneously with the penetration of insertion, but in this practice of experimental procedure, the highest or ultimate percutaneous force was observed.

5 Conclusion

Here, we have accomplished examines to explain the influences of different factors of the needle insertion process like vibration frequency and speed on ex vivo insertion force, which is comparable to the cutting speed at the equivalent deepness of the

needle inset. To develop appropriate effects in this analysis, a sequence of experiments (six times) with altered speeds (15, 5, and 35 mm/s) and vibration frequencies (110, 450, and 45 Hz) were measured, and as a result, the ex vivo insertion practice can be explored truthfully. At 45 Hz vibration, 19.3% a smaller amount of the insertion force is noticed than the 450 Hz vibration. In the practices of PCNL, needle deflection is also a vital drawback. Upcoming work is required to model the stimulus of needle vibratory analysis on the needle deflection.

References

1. Misra S, Ramesh KT, Okamura AM (2008) Modeling of tool-tissue interactions for computer-based surgical simulation: a literature review. Presence 17(5):463. https://doi.org/10.1162/pres.17.5.463.PMID:20119508;PMCID:PMC2813063
2. Dogangil G, Davies BL, Rodriguez y Baena F (2010) A review of medical robotics for minimally invasive soft tissue surgery. ProcInstMechEng H 224(5):653–679. https://doi.org/10.1243/09544119JEIM591
3. Barua R, Das S, Datta P, RoyChowdhury A (2022) Computational FEM application on percutaneous nephrolithotomy (PCNL) minimum invasive surgery through needle insertion process. In: Advances in computational approaches in biomechanics. IGI Global, pp 210–222
4. Barua R, Giria H, Datta S, Roy Chowdhury A, Datta P (2020) Force modeling to develop a novel method for fabrication of hollow channels inside a gel structure. Proc Inst Mech Eng 234(2):223–231
5. Mei F, Yili F, Bo P, Xudong Z (2012) An improved surgical instrument without coupled motions that can be used in robotic-assisted minimally invasive surgery. ProcInstMechEng H 226(8):623–630. https://doi.org/10.1177/0954411912447145
6. Cepolina FE, Zoppi M (2012) Design of multi-degrees-of-freedom dexterous modular arm instruments for minimally invasive surgery. ProcInstMechEng H 226(11):827–837. https://doi.org/10.1177/0954411912453239
7. Saedi S, Mirbagheri A, Farahmand F (2011) Conceptual design of a miniaturized hybrid local actuator for minimally invasive robotic surgery (MIRS) instruments. AnnuIntConf IEEE Eng Med Biol Soc 2011:2140–2143. https://doi.org/10.1109/IEMBS.2011.6090400
8. Liu W, Yang Z, Li P, Zhang J, Jiang S (2019) Mechanics of tissue rupture during needle insertion in transverse isotropic soft tissue. Med BiolEngComput 57(6):1353–1366. https://doi.org/10.1007/s11517-019-01955-6
9. Jushiddi MG, Mani A, Silien C, Tofail SAM, Tiernan P, Mulvihill JJE (2021) A computational multilayer model to simulate hollow needle insertion into biological porcine liver tissue. Acta Biomater 136:389–401. https://doi.org/10.1016/j.actbio.2021.09.057
10. Barua R, Datta S, RoyChowdhury A, Datta P (2022) Study of the surgical needle and biological soft tissue interaction phenomenon during insertion process for medical application: a Survey. ProcInstMechEng H. 2022;9544119221122024. doi:https://doi.org/10.1177/09544119221122024. [published online ahead of print, 2022 Sep 14]
11. Jushiddi MG, Cahalane RM, Byrne M, Mani A, Silien C, Tofail SAM, Mulvihill JJE, Tiernan P (2020) Bevel angle study of flexible hollow needle insertion into biological mimetic soft-gel: simulation and experimental validation. J MechBehav Biomed Mater 111:103896. https://doi.org/10.1016/j.jmbbm.2020.103896

12. Li ADR, Plott J, Chen L, Montgomery JS, Shih A (2020) Needle deflection and tissue sampling length in needle biopsy. J MechBehav Biomed Mater 104:103632. https://doi.org/10.1016/j.jmbbm.2020.103632
13. Li ADR, Putra KB, Chen L, Montgomery JS, Shih A (2020) Mosquito proboscis-inspired needle insertion to reduce tissue deformation and organ displacement. Sci Rep 10(1):12248. https://doi.org/10.1038/s41598-020-68596-w
14. De Silva T, Fenster A, Bax J, Romagnoli C, Izawa J, Samarabandu J, Ward AD (2011) Quantification of prostate deformation due to needle insertion during TRUS-guided biopsy: comparison of hand-held and mechanically stabilized systems. Med Phys 38(3):1718–1731. https://doi.org/10.1118/1.3557883. PMID: 21520885
15. Reddy Gidde ST, Ciuciu A, Devaravar N, Doracio R, Kianzad K, Hutapea P (2020) Effect of vibration on insertion force and deflection of bioinspired needle in tissues. BioinspirBiomim. 15(5):054001. https://doi.org/10.1088/1748-3190/ab9341
16. Ravali G, Manivannan M (2017) Haptic feedback in needle insertion modeling and simulation. IEEE Rev Biomed Eng 10:63–77. https://doi.org/10.1109/RBME.2017.2706966
17. Liu Y, Zhu W, Zeng G (2021) Percutaneous nephrolithotomy with suction: is this the future? CurrOpin Urol. 31(2):95–101. https://doi.org/10.1097/MOU.0000000000000854
18. Zhao Z, Fan J, Liu Y, de la Rosette J, Zeng G (2018) Percutaneous nephrolithotomy: position, position, position! Urolithiasis 46(1):79–86. https://doi.org/10.1007/s00240-017-1019-5
19. Sugita N, Nakajima Y, Mitsuishi M, Kawata S, Fujiwara K, Abe N, Ozaki T, Suzuki M (2007) Cutting tool system to minimize soft tissue damage for robot-assisted minimally invasive orthopedic surgery. Med Image ComputComput Assist Interv 10(Pt 1):994–1001. https://doi.org/10.1007/978-3-540-75757-3_120
20. Alken P (2022) Percutaneous nephrolithotomy—the puncture. BJU Int 129(1):17–24. https://doi.org/10.1111/bju.15564
21. Kallidonis P, Liourdi D, Liatsikos E, Tsaturyan A (2021) Will mini percutaneous nephrolithotomy change the game? Eur Urol 79(1):122–123. https://doi.org/10.1016/j.eururo.2020.10.010
22. Podder TK, Sherman J, Fuller D, Messing EM, Rubens DJ, Strang JG, Brasacchio RA, Yu Y (2006) In-vivo measurement of surgical needle intervention parameters: a pilot study. ConfProc IEEE Eng Med Biol Soc 2006:3652–3655. https://doi.org/10.1109/IEMBS.2006.259917
23. Barua R, Das S, Datta S, Roy Chowdhury A, Datta P (2022) Experimental study of the robotically controlled surgical needle insertion for analysis of the minimum invasive process. In: Emergent converging technologies and biomedical systems. Springer, Singapore, pp 473–482
24. Barbé L, Bayle B, De Mathelin M, Gangi A (2006) Needle insertions modelling: identifiability and limitations. IFAC Proc Vol 39(18):129–134
25. Barua R, Das S, Datta S, Datta P, Chowdhury AR (2022) Analysis of surgical needle insertion modeling and viscoelastic tissue material interaction for minimally invasive surgery (MIS). Mater Today Proc 1(57):259–264

Location-Based Attendance Monitoring System with Facial Authentication

K. Jaspin, V. Pandiyaraju, M. Mohanraj, and Mohammed Basid Safwan

Abstract Modern technology in location-based services, data management, and pattern recognition has made it possible to replace the conventional process in various fields. Manual tracking of attendance is based on pen and paper system, which is time-consuming and inaccurate. By eliminating the manual tracking of attendance for reducing the administrative workload, an automated location-based attendance monitoring with facial authentication is introduced. In this work, we propose a smart location-based attendance tracking system with an anti-spoof mechanism. It is enabled by facial recognition based on pattern matching-based LBPH operation. It provides maximum of 97% accuracy, and system even performs better on less amount of training dataset which offered 96% hit ratio. The proposed system results are much higher compared as compared with other approaches such as Eigenfaces and Fisher faces. The proposed LBAM-FA system is implemented as an application, which uses highly secure event recording via blockchain and neglecting the chance of data manipulation. In this system, company's location can be determined by the organization coordinates, and GPS may be used on a smartphone to track the whereabouts of each employee. For location-based attendance monitoring, this location serves as the key. By utilizing the Haversine Formula, the location within the organization premises is determined once the location of the user has been verified. Based on proposed LBAM-FA system, attendance is recorded in a secure and trustable manner.

Keywords Image preprocessing · Face recognition · LBPH algorithm · Haversine distance · Location verification · Attendance

K. Jaspin (✉)
St.Joseph's Institite of Technology, Chennai, India
e-mail: jaspink@stjosephstechnology.ac.in

V. Pandiyaraju
School of Computer science and Engineering, Vellore Institute of Technology, VIT University, Chennai, India
e-mail: pandiyaraju.v@vit.ac.in

M. Mohanraj · M. B. Safwan
St.Joseph's Institite of Technology, Chennai, India

1 Introduction

Maintaining attendance records is one of the most tedious tasks within an organization. If the organization needs to verify other employees' attendance and timing, additional employees are required, that involves costs. The organizations reap many benefits from an automated time and attendance monitoring system. The hypothesis proposes to address the major problems and provide both the organization and its members with a software solution that is user-friendly and easy to use. The smart device in our proposal, in contrast to previous smart systems, would not require a peripheral other than a smart phone, which would cut down on computation time and expenses related to adding a second device. Anyone entering the border zone with a smartphone and the app open will be instantly tracked. Our proposed system offers an automatic and trustworthy attendance management system for a business, which reduces human process mistakes. It helps to save time for the employees in marking attendance common area of the organization and avoids crowding and physical contact by using common attendance marking system. For increasing security, there is in need of avoiding another person to present an employee while they are not by enabling facial authentication and logging the check-in and check-out time to the immutable storage network to ensure its integrity. As the existing systems are susceptible to spoofing, we integrate the facial recognition with the location tracking, which helps reduce computing power and increases the security on the user's device in order to mark attendance accurately and without errors in it. Logs will be stored on the blockchain prototype. To reduce searching time, an internal block will be created and temporary storage is allocated. Therefore, the system will be trusted to be authentic and it will be impossible to spoof an employer's name. For identifying the drawbacks in already related works, a survey is conducted. It is discussed in Sect. 2.

2 Related Works

The attendance of the users is currently marked using biometrics and basic applications. The algorithm compares the information with a database of known objects and contours to find a match and then marks the attendance. Various commercial applications have been found for it. Records are stored in a typical database in the form of tables making it volatile. The existing system lacks features to ensuring the user's presence in the organization premises. However, the major drawback of the existing system is the ability for other people to mark someone's presence and flaws in existing algorithms for face recognition such as Eigenfaces and Fisher faces. Fisher faces and Eigenface operates and needs larger storage for face data which prevents them from applying to the practical applications and comparatively requires more processing time in recognition. Sultana et al. [1] proposed that with the Android consisting of a location tracking method and a database consisting of a table to store records. In

a conventional system, a human support would be needed to monitor attendance, adding to the company's costs. Furthermore, barcode badges, magnetic stripe cards, electronic labels, barcode badges, biometric features (fingerprint, hand, or face), and touch displays are used in automated attendance systems.

Walunjkar et al. [2] proposed very simple and understandable system that was initially used a high-resolution camera, which must be placed in the classroom in a convenient location from which the entire class can be covered by a single camera image. After the camera captures the image, further enhancements were made using various methods. After enhancement, the image was given to recognize the faces of the students, using the face recognition algorithm. After recognizing the faces, the face of each student was cropped out from the image, and all these cropped out faces are equated with the faces in the database. In this database, all the information about the students was already stored with their image. By comparing the faces, one by one, the presence of the students was marked on the server. Boyko et al. [3] proposed the system that helps to finding faces to recognize people in photos or video. The system has the modules such as Face positioning, Determination of unique facial features, Identification of a person. If the data are similar, the system displayed the name of the person else, the system not yet recognized the person.

Fan et al. [4] implemented the system to identify the facial features using intellectual methods or statistically to create a face model and match it with the degree of correspondence between the recognized region and the face model. Based on this, the possible face region was determined. The AdaBoost face recognition algorithm includes the selection of Harr-like features and the computation of features with the whole image. Enhanced Harr-like features were selected to increase the recognition accuracy, which are divided into edge features, linear features, and environment features. Somasundaram et al. [5] developed the system using VB.NET and SQL Server which has two stages. In the first stage, a feasibility study was conducted, a requirement analysis was done to understand the need for the system, and the necessary modules such as the administration module, login module, student module, SMS, and Android module were developed. Chawhan et al. [6] proposed a system that helps for instructors to work less by cutting down on the time and computations required to manually update attendance records. The website will also allow students and their parents to examine information on attendance and curriculum. This work was based on J2ME technology. The professor enters attendance into the cell phone. After saving the absentees, the teacher can edit the attendance list in the cell phone. And these attendance details can be sent to a computer via GPRS. It lags badly in Windows, Android, and iPhone and is also platform, version specific. In order to decrease false attendance, GPS technology is not supported.

Syamala et al. [7] proposed an automated attendance management system which is less invasive, more affordable, and more efficient using facial recognition technology. By using convolutional neural network, the proposed system can predict the image of a student with higher accuracy. If the student is present in the trained group, the corresponding student presence is automatically stored in the database. The system will be more efficient and user-friendly because it can be run on devices that are available anywhere and can be used by anyone. Facial images can be taken even from

some distance, and no special authentication measures are mandatory. Because of all these special features, facial recognition technologies were used, not only for security claims, but also for natural user interfaces, image retrieval, and image indexing. Uddin et al. [8] proposed that this system consists of four hardware and software components. The smart phone is equipped with a GPS receiver that can receive radio signals from GPS satellites. Here, a personal position is determined using GPS readings and identified using Google Maps API (Program Programming Interface). The application may geolocate users to determine their present whereabouts. GPS may be used to identify the coordinates of an organization and an employee; if the coordinates match, the employee is present in the organization. The management software processes the data before storing it in the database. Vyshnavi et al. [9] proposed a system that a company can use an automated attendance management with real-time facial recognition. The camera captures the image, the faces are recognized, the faces are matched with the database, and finally the attendance is registered. In this work, we build a general deep neural network-based model for automatic person recognition, detection, and presence update of the identified faces in an attendance sheet. For this purpose, we have used Haar cascade classifier for face detection and LBPH face recognizer for face recognition. Smitha et al. [10] proposed a system that uses facial recognition technology to provide an efficient attendance system for classrooms. By using facial recognition, the system will be able to record attendance. Using a webcam, it will identify faces after which it will detect them. Following recognition, it will update the attendance list and record the acknowledged student's attendance. Each student in the class must sign up by providing the necessary details, after which their images are taken and saved in the file. Faces from the classroom's live streaming video were identified during each session. The images included in the dataset are matched with the faces that were detected. If a match is discovered, that student's attendance was recorded. A list of those who missed each session is forwarded to the academic member who was in charge of it at the conclusion of each one.

Porwal et al. [11] proposed a structure of the system. This structure needs a simple login procedure by the professor of the class through the host class to create a key-based QR code with some specific data. This technique can be carried out whenever the lecturer feels it necessary—during breaks in lectures, at the start, or after the lesson. The tutor uses a key to demonstrate the QR code to the students. The students then utilize the structure user class to scan the professor's QR code. From college's app market, this application may be obtained. Pandey et al. [12] developed an intelligent attendance system with automatic face recognition, automatic notification to parents of student absence, automatic sending of a report to staff via email depending on the number of students present and absent for live monitoring. In this, they developed an attendance system that observes the student's face and checks the appearance of the student's face data in the data archive and marks him as present or absent. This system uses Ultralight detector and Mobile FaceNet techniques to detect the face and compare it with the database. A SMS API was integrated into the system to automatically send text messages to students' guardians when their child has missed class or is absent. Third-party emailing was integrated to send emails to staff. Sasi et al. [13]

proposed the smart appearance design based on a high-megapixel camera that can capture the entire class. To start the system, initially student images were stored in the database. The images were cropped to a size of 24×24. Once the camera turns on, it will record the whole class for a few seconds and then stop automatically. The algorithm detects the faces and then crops them to 24×24 size. The clipped faces were then provided to the recognizer after this procedure. The database compares the faces, finds the attendance, and marks it.

Kishore et al. [14] proposed that a finite number of dataset classifiers embedded in a training network requires a huge set of possible data functions with all the edge and error cases, which have more runtime and take up more memory, compared to continuous steps of training data with coded objects. However, certainty can be a minimal observable factor when training data without training networks or recursive algorithms. Therefore, to exceed the certainty factor, a set of similar characteristic methods can be added to the steps of encoding objects that allow the process of classifier with a reasonable level of accuracy, and therefore, the certainty increases as more successive compilations and object encodings were performed. Lineament features can be encoded to perform different tests and basic recognition can be achieved using a library called Dlib. Ghalib et al. [15] proposed the system where to record the entire class, including the final bench, and a camera is positioned in the middle of the classroom at an appropriate height. After the students are seated, the camera snaps a photo and begins the face recognition process. The software then generates a folder with the learners, so they can be identified in the database. For image recognition, the database is utilized to access the already-posted images of each student. To determine if the student is in the class or not, the images were obtained and put in comparison with the database information.

The objectives of the proposed LBAM-FA work are to automatically record the attendance of employees with Face Authentication, which improves the security as compared with manual processes. The proposed framework for location-based attendance monitoring system is discussed in Sect. 3.

3 Framework for Location-Based Attendance Monitoring System

In this proposed LBAM-FA system, we integrate facial recognition with location verification to verify the person's identity and presence within the organization's premises, which helps in ensuring the physical presence of employees within the organization premises using Haversine Formula which suits best in case of dealing with the distance calculation on an ideal surface neglecting the influence of the terrain structure and reduces the time in distance calculation as stated theoretically, and the lightweight LBP approach for facial recognition consumes less computing power compared with other approaches using CNN, Eigenfaces, Fisher faces and increases the recognition chances using the facial features of a person as a dominant feature

to compare, whereas in other methodology, additional features like light influence the recognition result. LBP also performs robust on grayscale-converted images. User images are captured using the individual's device camera and transferred to the server in a base64-encrypted format for recognition and saved nowhere else. This ensures the security and privacy of user data. Attendance logs will be stored on the blockchain-based log storage system which ensures the trust, reduces the chance of data manipulation, and prevents data integrity. To reduce searching time from the blockchain, an internal block is created and temporary storage is allocated, and on finalizing the data for a day, the data get stored into the presence chain. Therefore, the system will be trusted to be authentic and avoids another person to spoof an employee's presence.

The system begins by collecting details from the user; the admin will also collect facial data. With the help of HOG and CNN classifiers, key features are extracted from the face and encoded into bits. After this, the organization's current location is gathered. To train the LBPH model, the encoded bits are compared with the trained facial data and the admin location is noted on the server. The first step in the recognition module is to capture the user input frames in real-time and process them to extract the facial features of the current frame. Afterward, the extracted facial feature is encoded in the same way as data in the training module, followed by a comparison with the existing datasets of a user in the encrypted and trained version. Once a match is found with any of the user's facial features in the dataset, a verification is implemented to ensure that the candidate is genuine. System architecture for the location-based attendance monitor with facial authentication model is shown in Fig. 1.

3.1 Training Module

In training, we encode the object into a pickle object and store the LBPH classifier model in YML format. This is to make computation lightweight and faster. Then, the data are stored in an encrypted format to ensure the security of the data file in byte format. The working of LBPH algorithm is stated in Algorithm 1. In this algorithm, NumPy Array of user image is used for training the user detail. For that purpose, each frame as a NumPy Array is processed individually and converted to gray frame. From the gray frame, the face end points $x1$, $y1$, $x2$, $y2$ are marked and each frame is cropped accordingly using Eq. (1).

$$\text{user\_face\_region} = \text{frame}[y1 - \text{offset} : y2 + \text{offset}, \\ x1 - \text{offset} : x2 + \text{offset}]. \tag{1}$$

Every region uses a local binary operator. A 3×3 window defines the Local Binary Patterns LBP operator. LBP is defined in Eq. (2) in which '(C_x, C_y)' is central pixel with intensity 'i_c'.
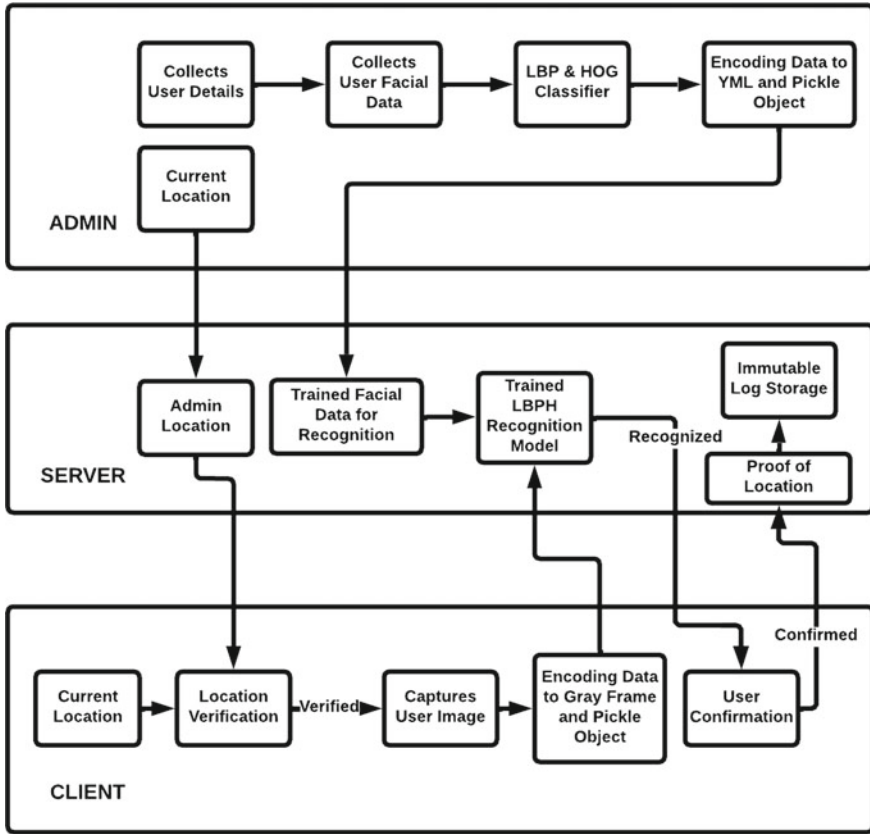
Fig. 1 Framework for location-based attendance monitoring system

$$\mathrm{LBP}(C_x - C_y) = \sum_{p=0}^{P-1} 2^p s(i_p - i_c).\tag{2}$$

Now, find the median pixel intensity value to set as threshold, it equivalences a pixel intensity value with its eight closest neighbor pixel values. If the intensity of neighbor is higher than or same as that of central pixel intensity value, it is set the value as 1, otherwise it is 0. As a result, eight binary values are derived from eight neighbors which are decoded to a decimal value. This is done to normalize the image and neglect illumination variation and the decimal number named the LBP value and the range is vary from 0 to 255. The various numbers of radius and neighbors are considered for improvisation as fixed neighborhood approach may not be good at all scenarios. The number of similar LBP values in each region is counted to construct the histogram for that region. The combined histogram that results from this step is identified as the feature vector which is used for processing. The histograms stored can use for recognition to find the match using Euclidean distance given in Eq. (3).

$$d(a, b) = \sqrt{\sum_{i=1}^{n} |a_i - b_i|^2}. \qquad (3)$$

In order to return the image object with the closest histogram, the test image's features are compared against database-stored features. Finally, the encoded user image object based on LBPH classifier is generated.

Algorithm 1. Working of LBPH Algorithm

Input: NumPy Array of user image for training with user detail
Output: Encoded user image object based on LBPH classifier

Read Face images
for 1:N
NumPy_Array ← Process each frame of Face images
Gray_frame ← Convert NumPy_Array to gray frame
end for
For each gray frame
    Mark face end points x1, y1, x2, y2
    Perform Cropping using frame$[y1 - \text{offset} : y2 + \text{offset}, x1 - \text{offset} : x2 + \text{offset}]$
    Return cropped_face
End for
Apply local binary operator in Eq. (1) of 3×3 window in to each region
Find the median pixel intensity value to set as threshold
If intensity_value $\geq$ central_pixel_intensity
    Return 1
Else Return 0
Merge all binary values
LBP_value ← Convert binary to decimal value
Create a histogram by counting the number of similar LBP values in each area
    Return the histogram and feature vector of the image
Use histograms for recognition by finding the match using Euclidean distance
    Return the image object with the closest histogram
Compare the test image features against database-stored features
If Matched
    Genuine candidate and start recording the attendance
Else Not a genuine candidate & reject recoding the data
End If

3.2 Recognition Module

In recognition, extracted facial features are encoded in the same manner as data in the training module proceeding with the comparison of the current encoded data with the existing datasets of users in the encrypted pickle object file and trained LBPH classifier which is stored as a YML file. The user's facial features in the dataset are compared, and if they match the user's features, then the genuineness of that candidate is verified.
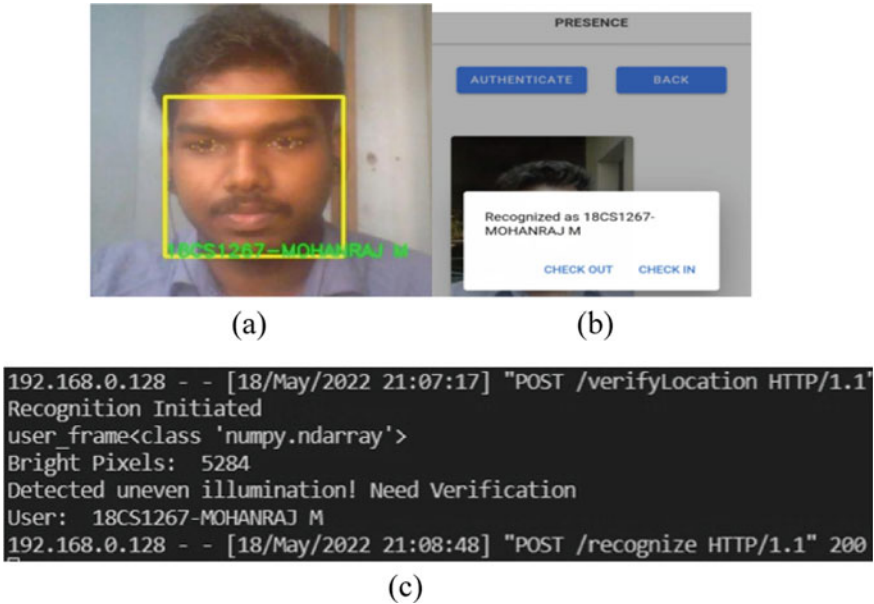
Fig. 2 Recognition process **a** recognition prompt, **b** user recognition, and **c** recognition response

The above Fig. 2a shows that the recognition prompt in the server side is displayed with the recognized result to be sent to the user for verification. Figure 2b shows that recognition requires capturing of user's facial frame and sends it to the server as a base64-encoded string and displaying the recognition result received from the same. Figure 2c shows that the recognition response is received from the server by finding the best match among the recognition result obtained from LBPH approach with the help of encoded user's facial data.

3.3 Location Verification Module

In the location verification, the latitude and longitude of the organization's location are stored on the server. With this solution, the user location coordinates obtained from the device are verified by using the Haversine formula to find the distance between the admin locations and storing the data in the log storage. It used user's

current coordinates and organization coordinates with distance threshold as the input. Then, the following steps are used for location verification.

Step 1: The coordinates of both user and organization are passed on to the Haversine formula to find the flat distance, i.e., greater circle distance between two points.

Step 1.1: Distance between the latitudes and longitudes is calculated and converted to radians.

$$D\_La = (La\_2 - La\_1) * m.pi/180.0$$

$$D\_Lo = (Lo\_2 - Lo\_1) * m.pi/180.0$$

$$La\_1 = (La\_1) * m.pi/180.0$$

$$La\_2 = (La\_2) * m.pi/180.0.$$

Step 1.2: Applying the Haversine formula for calculating the distance as follows:

$$A = pow(m.sin(D\_La/2), \ 2$$

$$B = pow(m.sin(D\_Lo/2), \ 2$$

$$C = m.cos(La\_1)$$

$$D = m.cos(La\_2)$$

$$E = m.asin(m.sqrt(A + B * C * D)).$$

Step 1.3: Multiply the E (unit considering earth points as 1 km) with the actual km 6371 and return the resultant distance in terms of kilometer.

Step 2: Checking the resultant distance is within the threshold distance and returning the binary result 1 for inside and 0 for outside to the system.

Finally, the binary result of user location with in threshold is recognized as an approved location for recording the data.

```
192.168.0.128 - - [18/May/2022 21:07:17] "OPTIONS /verifyLocation HTTP/1.1" 200 -
Location Verification Initialted
Organization Location(SJIT):
Latitude: 12.9099046
Longitude: 80.232425
Distance 0.005880646143205246
Radius 0.5
192.168.0.128 - - [18/May/2022 21:07:17] "POST /verifyLocation HTTP/1.1" 200 -
```

(a)
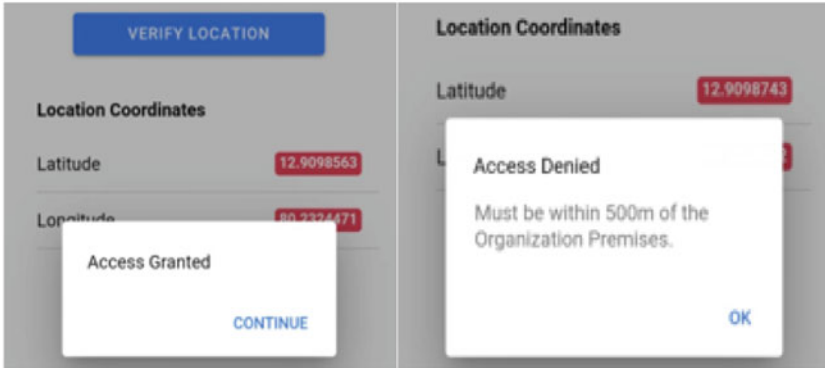
(b) (c)

Fig. 3 Verification process **a** location verification, **b** access-granted response and **c** access-denied response

3.4 Denied Access

The above Fig. 3a represents the location verification response received from the server on calculating the circular distance between the two coordinates. Figure 3a shows the location verification involving the collection of user's current coordinates and ensuring it within 500 m (variable) from the organization i.e., within the organization premises. Figure 3c shows the user's access denied as the current coordinates do not fall under the threshold radius fixed by the organization.

3.5 Log Storage Module

For the log storage section, you have the user's entry in the required location checked and marked in the log storage where there will be separate data for each user, as well as the time when the user checks out have to be marked and saved in the immutable log storage.

{"last_updated": "2022-06-22 20:59:55.888497", "presence_chain": [], "log_chain":
[{"index": 1, "timestamp": "2022-06-22 19:26:28.400375", "log": "18CS1267-
MOHANRAJ checked in at 2022-06-22 19:26:28.400375.", "prev_hash":
"7ba1f1f25ea31ac220483cbc4b99921eab3d2989b59e3aa6299baef0949b4c6f",
"hash":
"464c298d13542db1b1fde90ae64d58251c9d5653d0e749dade1606b8c213820f"},
{"index": 2, "timestamp": "2022-06-22 20:43:11.057828", "log": "18CS1267-
MOHANRAJ checked out at 2022-06-22 20:43:11.057828.", "prev_hash":
"464c298d13542db1b1fde90ae64d58251c9d5653d0e749dade1606b8c213820f",
"hash":
"e864d1ec2ead3794e071eca1a3ceafb404e066601711acf26f226de6b7a56d60"},
{"index": 3, "timestamp": "2022-06-22 20:43:34.766197", "log": "18CS1223-
BASID checked in at 2022-06-22 20:43:34.766197.", "prev_hash":
"e864d1ec2ead3794e071eca1a3ceafb404e066601711acf26f226de6b7a56d60",
"hash":
"9443fe64087dd084d2ede1a50961eefed0fb75af5bcae5301a06f0cf68fb591a"},
{"index": 4, "timestamp": "2022-06-22 20:43:43.612760", "log": "18CS1223-
BASID checked out at 2022-06-22 20:43:43.612760.", "prev_hash":
"9443fe64087dd084d2ede1a50961eefed0fb75af5bcae5301a06f0cf68fb591a",
"hash":
"9d1174099ac31bd66d9bd7569bd7d84517ed6fc1e9e98b7753ab5888da270238"},
{"index": 5, "timestamp": "2022-06-22 20:57:41.027527", "log": "18CS1267-
MOHANRAJ checked in at 2022-06-22 20:57:41.027527.", "prev_hash":
"9d1174099ac31bd66d9bd7569bd7d84517ed6fc1e9e98b7753ab5888da270238",
"hash":
"52db23f964959bb3c6b4561cdbf7b6c3d4b8345eab29619cdabe0754406b14c8"},
{"index": 6, "timestamp": "2022-06-22 20:58:45.076113", "log": "18CS1267-
MOHANRAJ checked out at 2022-06-22 20:58:45.076113.", "prev_hash":
"52db23f964959bb3c6b4561cdbf7b6c3d4b8345eab29619cdabe0754406b14c8",
"hash":
"2ed1f8e5876b5d325460d0c71928ba028fe7c17365daa7868f9cc10ce57fea6f"},
{"index": 7, "timestamp": "2022-06-22 20:59:55.885505", "log": "18CS1223-
BASID checked in at 2022-06-22 20:59:55.885505.", "prev_hash":
"2ed1f8e5876b5d325460d0c71928ba028fe7c17365daa7868f9cc10ce57fea6f",
"hash":
"453a4667181b2eaed2cac22caf023956337b74dc83ca9bb345f42a91a6a535f4"}]}

(a) (b)

Fig. 4 Log storage process **a** check-out prompt and **b** list of logs

Figure 4a shows check-out prompt to be displayed to the current user checked in to system facilitating the option to hit check-out request to the server. Figure 4b specifies the logs stored in the system temporarily in the log chain which is listed along with the hash value of each block, which will be then transferred to presence chain after finalizing the data.

4 Result Analysis

Existing methodologies tend to find the closest match and return it as a result in all cases. The proposed LBP and HOG models collectively ensure better result in recognizing the user registered in the system and also in finding the user not registered in the system which ensures maximum security in preventing the spoofing of a person.

The training images with the labeled data are given to the HOG detector and processed by the LBP operator and encoded for further recognition. After verifying the location, the user captures image for authentication which is passed on to the LBPH based recognition model which then classified as one with the maximum matches.

In the above Table 1, we have trained the LBP operator with ten PICS of a subject and followed with the recognition of 50 PICS of the subject and achieved the above hit ratio. The existing models prevent a person's features from becoming dominant, but they also regard illumination variations as a useful feature, completely deviating

Table 1 Hit ratio comparisons of proposed and existing algorithm-based models

Training models	Testing: 50 PICS		
	Correct	Error	Hit ratio (%)
Eigenfaces	46 PICS	4 PICS	92
Fisher faces	46 PICS	4 PICS	92
Proposed (LBPH)	48 PICS	2 PICS	96

the system from its intended use case since illumination variations are not suitable to extract because they are not a part of the actual face and produce a hit rate of 92%. With less training data, the proposed LBAM-FA method provides comparatively better results considering the 96% hit ratio.

$$\text{Accuracy} = \frac{\text{Correctly identified PICS}}{\text{Total number of PICS}} * 100, \tag{4}$$

$$\text{Incorrect identifications} = \frac{\text{incorrectly identified PICS}}{\text{Total number of PICS}} * 100. \tag{5}$$

The above formula (5) depicts the accuracy attained by the existing and proposed LBPH-based recognition model provided with the training data of 200 PICS of the subjects provides the above stated accuracy values by recognizing the similar subjects. It is observed that at different light and environmental conditions, the existing system fails to recognize the subjects as it considers the illumination as a major parameter even though it was not a part of the face and provides an accuracy of 93 and 89% in recognizing subjects with similar faces. Our proposed system performs better at different environment comparatively better providing a higher accuracy of 97%. Fig. 6 depicts the incorrect identifications of proposed and existing algorithms while testing 200 PICS. Our proposed system performs in providing less incorrect identification compared with two existing methods taken for analysis. From the above analysis, we can state the proposed LBPH-based recognition model which performs better in controlled recognition environment at different conditions and provides better results.

The Haversine methodology in calculating the shortest distance between the two points considering on an ideal surface ignoring the terrain structure suits well our needs and provides accurate result in measuring short distance with less computation power and time as it has been theoretically proven that 39,800 distances given latitude and longitude are calculated in 0.04 s.
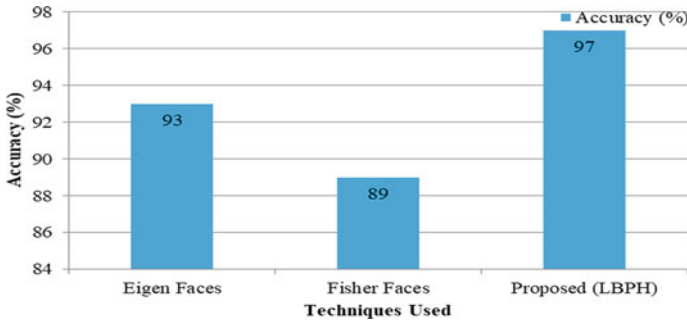
Fig. 5 Accuracy comparison of proposed and existing algorithms while testing 200 PICS
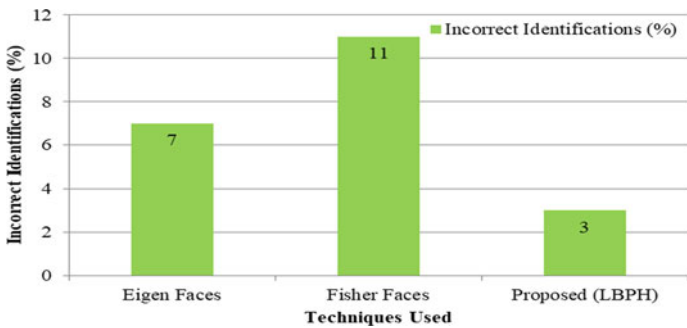
Fig. 6 Incorrect identifications of proposed and existing algorithms while testing 200 PICS

5 Conclusion and Future Works

This proposed work introduces a location-based attendance monitor with facial authentication using android application which uses location as the core component of attendance tracking using smart phone. Employee coordinates inside the area border illustrate that the employee is present in the organization as a threshold radius is set up for tracking using GPS. The main aim was to successfully capture the attendance log of the employees within the organization. The facial measures are taken when the person captures their face using phone camera. The fact that facial recognition is non-intrusive is one of the most major benefits of using it. The confirmed attendance was successfully entered into the log storage. Through facial authentication, we can avoid the spoofing of an employee and ensure the presence of the employee within the organization premises. The proposed LBP-based recognition methodology provides better results in terms of accuracy as 97% comparatively higher than other approaches such as Eigenfaces, Fisher faces and provides better hit ratio of 96% in case of less training dataset. This approach may be used to improve attendance and leave management, save time, and minimize the amount of physical effort required by the administration. Consequently, a location-based attendance

tracking system has been established with the anticipated outcomes, although there is still potential for improvement. The future scope of this proposed system is we can scale up to use in real time by deploying to a high-speed cloud service providers which ensures speed and accuracy of this model as it provides enough computational power. Instead of using image as bits, processing video stream can be done to ensure liveness by tracking the facial movements.

References

1. Sultana S, Enayet A, Mouri I (2015) A smart, location based time and attendance tracking system using android application. Int J Comput Sci Eng Inform Technol 5(1):1–05
2. Walunjkar PG (2019) Smart Attendance system using facial recognition. Int J Res Appl Sci Eng Technol 7(4):1064–1069
3. Boyko N, Basystiuk O, Shakhovska N (2018) Performance evaluation and comparison of software for face recognition, based on Dlib and OpenCV library. In: IEEE second international conference on data stream mining & processing (DSMP), pp 478–482
4. Fan X, Zhang F, Wang H, Lu X (2012) The system of face detection based on OpenCV. In: 24th Chinese control and decision conference (CCDC), pp 648–651
5. Somasundaram V, Kannan M, Sriram V (2016) Mobile based attendance management system. Indian J Sci Technol 9(35)
6. Chawhan S (2013) Mobile phone based attendance system. IOSR J Comput Eng 10(3):48–50
7. Syamala BRKNL (2020) Attendance management system using face recognition. Int J Res Appl Sci Eng Technol 8(7):1684–1688
8. Uddin MS, Allayear SM, Das NC, Talukder FA (2014) A location based time and attendance system. Int J Comput Theory Eng 36–38
9. Vyshnavi P (2021) Employee attendance management system using face recognition. Int J Res Appl Sci Eng Technol 9:5303–5308
10. Bin Hamed R, Fatnassi T (2022) A secure attendance system using raspberry pi face recognition. J Telecommun Digit Econ 10(2):62–75
11. Porwal Y, Rastogi D (2020) Smart attendance system. Int J Comput Sci Eng 7(6):13–16
12. Pandey P, John Arvindhar D (2020) Smart system for attendance. Int J Innov Res Appl Sci Eng 3(12):87
13. Sasi Chandra M, Radhika B (2019) The smart attendance monitoring system. Int J Innov Technol Explor Eng 8(11S):62–66
14. Kishore Shiva S, Mohanraj M, Vimalakeerthy D (2022) Computer vision lineament landmarks classifier and synchronic detection integration using OpenCV AndDlib archive. IJRAR Int J Res Anal Rev E-ISSN 9(1):431–434
15. Srinivas TS, Goutham T, Kumaran MS (2022) Face recognition based smart attendance system using IoT

Load Balancing in Cloud Computing Using Multi-agent-Based Algorithms

Shyama Barna Bhattacharjee

Abstract In distributed environments, cloud computing is widely used to manage user requests for resources and services. Resource scheduling is used to handle user requests for resources based on priorities within a given time frame. The resource scheduling concept of a virtual machine is used to manage servers. To manage load and perform resource scheduling in a smooth manner in this paper, an attempt has been made to propose a multi-agent-based load-balancing mechanism. In a multi-agent-based load-balancing mechanism, multiple agent nodes at different layers are used to manage load and requests and increase throughput.

Keywords Cloud computing · Agent · Multi Agent · Virtual machine and load balancing

1 Introduction

Cloud computing is one such computing model that is so popular among this generation for providing better services to its users. It highly depends upon virtualization technology [1] as well as the internet. In the cloud of Infrastructure-as-a-Service (IaaS), an application from the artificial resource pool is placed in a virtual machine (VM). VMs having common machines have the tendency to run several operating environments of an operating system at one time. It can also be configured to provide several services and software in accordance with several service requests. Additionally, VMs can be initiated and ceased on one machine. It can also be transmitted between different machines, which hugely expand the elasticity of the resource allocation. The issues of over provisioning [2] or under provisioning appear because of the unpredictability of the resource usage of applications by users. For maximising the infrastructure utilisation without breaching the Service Level Agreement (SLA)

S. B. Bhattacharjee (✉)
Department of CSE, UIET, Kurukshetra University, Kurukshetra, Haryana, India
e-mail: shyamabarna02@gmail.com

[3], the data centre workload can be easily balanced through the VM migration process.

State algorithms are such existing migration methods that choose suitable node for migration depending upon the recent execution and information of hardware workload [4]. One such traditional static VM migration method is the dual-threshold method [5] that keeps both a lower bound and an upper bound on host machines' workload and initiates the migration when the workload is below a lower bound or over an upper bound. Unfortunately, these strategies of threshold-based migration could not foresee the host machines' upcoming workload trend. Additionally, if the host machine's workload increases even for a second (for any reason), irrelevant and useless VM migration would be triggered. Different and useful methods finalise decisions depending upon the recent status as well as the past information of workload [6].

2 Load Balancing in Cloud Computing

To enhance the performance as well as the utilisation of resources in the system, dividing the employment among the accessible processors is all that is required. The aim of readjusting load is to dispense the load among various nodes in different cloud settings. Load adjustment is essentially an attempt to carefully balance the overall system of load in multiple nodes by transmitting employment between nodes [7]. This is done in order to carefully dispense the ready nodes to give satisfactory overall performance of the system [8].

2.1 Load-Balancing Taxonomy

 (i) *Static balancing of load*: It is an algorithm basically to ensure the system's typical behaviour. Mostly, static algorithms are not complex and are ineffective because they fail to reflect on the current and most recent state of the system [9].
(ii) *Dynamic balancing of load*: Dynamic algorithms dispense load between the nodes by utilising the most recent node state. Since dynamic load balancing has to gather as well as respond to the state of the system, it is more complex than static load balancing. Additionally, they have the tendency to behave in an extra efficient manner as compared to the state algorithm because of the dynamic load-balancing considerations [10].

2.2 Agent Based Load Balancing

(i) *Agent*: An agent is basically a process entity which usually acts for other objects by performing tasks and achieving goals. Agent systems are unique programmes of software systems which mainly adhere to the information domain [11]. They have the ability to react with a certain degree of independence, primarily for holding actions required for achieving their various goals [12].These agents are executed, especially when they are being operated in different environments.

(ii) *Agent-based load balancing*: An agent has the quality of performing particular tasks on its own, because of which it is known as an autonomous unit. The communication between agent and server usually does not take much time to process as compared to the centralised approach because of which the theme of agent-based load is readjusted and also minimises the servers' communication price [13]. Additionally, it enhances the pace of load balancing, which further increases the latency as well as the output of servers.

(iii) *Existing system*: The figure highlights the design of an existing system with several clients connected through the internet as a supplier of facilities that has revived many machines, systems of control, and the number of many mutual groups of resources that appear as clients [14]. One life cycle of an agent consists of the following steps:

   (i) *First step*: It acts as an preliminary server which gives commands to the final server for gathering data from all the servers that are actually needed for balancing load.

   (ii) *Second step*: Agent equals the ability of all the servers constructed to take the cloud's average load.

In the first step, an agent is started at an undefined server (mainly the initial server), which further finds several jobs in that server's line. After this, it calculates an average, depending on which, it will calculate the server's location and give a conclusion of whether it is underloaded or overloaded. The agent will then repeat this entire process till it reaches the last server. In the second step, the agent will reach from the initial server to the final server for the concerned load. On the particular server, it gives the solution for cross-checking the state. If the server is found to be overloaded, then the agent will transmit the jobs to those servers which are underloaded, and if the servers are underloaded, then it will get jobs from overloaded servers. The agents are known for doing this activity to the extent that the initial server will again adjust all the servers' load that contains the initial server. In this way, representatives are mainly utilised for balancing the load within the structure nodes.

3   Literature Review

Patel et al. [15] focused on distributing the remaining burden evenly across the compartment and shortening the make span. For load adjusting, various specialists have proposed various methodologies, such as genetic algorithm-based, PSO, ACO, Mn-Max, and so on. The authors of this paper has proposed a Grey-Wolf Optimization-based algorithm for adjusting load and shortening the make span. A GWO-based methodology was presented alongside GA and PSO-based algorithms.

Junaid et al. [16] proposed the ACOFTF half-breed algorithm, which takes into account significant QoS measurements such as SLA infringement, relocation time, throughput time, overhead time, and improvement time. Even in the presence of multiple datasets, the proposed model has demonstrated its ability to avoid untimely union, which was one of the goals of half-breed meta-heuristics. In the same way, a lack of variety encourages investigation, while a lot of variety often, but not always, leads to abuse.

Babou et al. [17] proposed the HEC Clustering Balance procedure, which fundamentally decreases the preparation time of client solicitations while productively utilising assets of the HEC node, MEC node, and the focal cloud on the three-level HEC engineering. The proposed strategy also allows for the use of HEC nodes that are not always selected. The proposed strategy uses a diagram to represent the cloud–MEC–HEC structure. Each MEC, known as a group head, is in charge of a sub-diagram, or bunch. There was a group head on each bunch (3-TER) that established the upper layer or 2-TER of the HEC design. The group heads are then linked together from one perspective and with the focal cloud from another.

Shahid et al. [18] focused on one of the CCMAN issues, such as load adjusting (LB). The goal of LB was to equilibrate the algorithm on the cloud nodes so that no host was under or overburdened. A few LB algorithms have been written in order to provide successful organisation and client fulfilment for fitting cloud nodes, to improve the general proficiency of cloud administrations, and to provide more fulfilment to the end client. A good LB algorithm improves productivity and resource use by dividing the remaining work among the framework's nodes in the best way possible.

Devraj et al. [19] proposed a load-adjusting algorithm based on the firefly and improved the multi-objective particle swarm optimization (MPSO) strategy, dubbed FMPSO. This method employs the firefly (FF) algorithm to narrow the search space, while the MPSO strategy was used to identify the improved response. The MPSO algorithm selects the global best (gbest) molecule with a small separation of highlights.

Saeedi et al. [20] present the asset skewness-mindful VMs' combination algorithm based on a more advanced thermodynamic recreated toughening strategy because asset skewness may prompt the algorithm to initiate extra workers. With the help of two heuristics and two meta-heuristics, the proposed SA-based algorithm has been tested in a wide range of situations with different amounts of skewness in the assets.

4 Proposed Work

4.1 Motivation

The existing system model agent-based dynamic load-balancing approach has a disadvantage, i.e. for every instance of a new job, the agent has to complete a life cycle (i.e. calculate the average and decide the status of the server). Hence, agents need to travel twice to allocate the job: once to find the average work load and a second one to balance the loads among nodes. To overcome this, we propose a multi-agent architecture for agent-based load balancing. A multi-agent architecture comprises Directory Agents and Mobile Agents. In this section, the proposed framework has been presented in detail. In the proposed work, we first provide a rank assignment framework to assign rank to the resources, and after that, an agent selection framework has been presented to select agent node. Finally, we present a resource allocation framework to allocate resources in an optimal manner to achieve high QoS and update the resource allocation table.

4.2 Rank Algorithm

Rc = 1 * initializing the resource counter*\
Rk = 1 * initializing resource rank*\
Cp = R[Rc]. Rpc * initializing the processing capacity of current resource to capacity of resource in the counter index*\
Cbw = R[Rc]. Rbw * initializing the bandwidth of current resource to bandwidth of resource in the counter index*\
Crt = R[Rc]. Rrt * initializing the response time of current resource to response time of resource in the counter index*\
Cl = R[Rc]. Rlc * initializing the latency of current resource to latency of resource in the counter index*\
For ∀ R[] **do** * repeat for all resources in the resource list*\
For i = 1:R[].size
If R[i]. Rpc ≥ Cp& R[i].Rbw ≥ Cbw&R[i]. Rrt ≥ Crt
&R[i]. Rlc ≥ Cl * compare the capacity of the current resource and the resource in the counter index*\
Cr = R[i] * set current resource to resource in the counter index if above condition is true*\
End for
Rank = Rk * rank the resource with the value in the rank variable*\
Update Cr Rank in At * update current resource in dynamic resource table*\
Remove(R[].Cc) *remove current resource from resource list*\
Rk = Rk + 1 * increment resource rank*\
Rc = Rc + 1 * increment resource counter*\

End for

In this algorithm, they assign rank to the resources. The rank of a resource is assigned based on parameters such as processing capacity, bandwidth, response time, and delay. The highest ranked resource capacity is updated for agent allocation in the agent information table.

After that, the host agent analyses the tasks and their requirements in the task list and chooses an appropriate layer for the completion of such tasks. Furthermore, once the connection is established, it is then up to the layer agent to perform the actual transmission.

4.3   Agent Selection Algorithm

Let Bmax, Pmax, Lcmax, Smax, and Rtmax represent bandwidth, processing capacity, latency, storage, and response time of the highest ranked resource in fog, high-capacity fog, and cloud layer.

Let Br, Pr, Lct, Sr, and Rtr represent task required bandwidth, processing capacity, task tolerable latency, task required storage, and response time.

The host agent selects an agent to allocate a resource for the execution of task by comparing the QoS requirement of the task and the capacity of the highest ranked resource in each host dynamic resource table.

For $\forall$ *t* ε R **do** * repeat for all task in a request*\\
If *td* $\leq$ *ac. Lcmax* + *ac.Rtmax&Sr* $\leq$ *ac.Smax** compare task deadline with sum of latency & response time of cloud *\\ Submit task to cloud agent * submit task to cloud agent if above condition is true*\\
Else
Return (Resource unavailability message) * return failure due to resource unavailability*\\
End if
End for

In this algorithm, the host agent selects an agent to allocate a resource for the execution of a task by comparing the QoS requirement of the task and the capacity of the highest ranked resource in each host dynamic resource table. An agent allocates a capable resource for a task received from a host agent by taking into account the deadline of the task, required processing capacity, bandwidth, response time, and tolerable latency. Any available resource that meets the requirements of the task is allocated for the execution of such a task.

4.4 Resource Allocation Algorithm

Let Bw, Pc, Lc, S, and Rt represent bandwidth, processing capacity, latency, storage, and response time of resource in fog, high-capacity fog, or cloud layer.

Let Br, Pr, Lct, Sr, Rtr, and td represent task required bandwidth, processing capacity, task tolerable latency, task required storage, response time, and task deadline.

The resource allocation algorithm selects the best available resource for execution of task that satisfies the QoS requirement of the task.

For $\forall$ t ε T[] * repeat for all task in task list*\
For $\forall$ r ε R[] * repeat for all resource in resource list*\
If td $\leq$ Lc + Rt& Pc $\geq$ Pr&r.status = free* compare task deadline with sum of latency & response time of resource *\
Allocate resource to task * allocate resource to task if above condition is true *\
r.status = allocated. * change resource status to allocated *\
t.status = allocated * change task status to allocated *\
End if
End for
If t.status = !Allocated * change task status not allocated *\
Return (Resource unavailability message) * return failure due to resource unavailability*\
End for

In this algorithm, the optimal resource for completing task is selected to fulfil the QoS of task.

4.5 Dynamic Resource Table Update Algorithm

Let R [], At, and Ct represent resource list, corresponding agent table, and current task executed by resource.

The dynamic resource table update algorithm updates the status of a resource after task execution is completed.

For $\forall$ r ε At. R[] * repeat for all resource in dynamic resource update table*\
tc = R[].Get (Ct) * get the current task executed by a resource*\
If tc = NULL * if task not found*\
r. status = free * change the status of the resource to free*\
End If
End for

Table 1 Simulation parameters

Parameters	Values
No of servers	100; 200; 300; 400
CTC (existing centralized scheme)	10 units
CTC (proposed)	1 unit
CTC (process of LB)	50 units
Platform	MATLAB 2013
Operating system	Windows 10

5 Results and Analysis

In this section, results and analysis have been presented. To implement the proposed mechanism, MATLAB is used. MATLAB is a general-purpose programming language developed by Mathworks. It provides a number of functions for different researchers to compute their simulations and plot results.

5.1 Performance Matrices

(i) *Throughput*: It indicates the total number of requests processed in perspective of total number of requests generated during data transmission.
(ii) *Latency*: It indicates the delay or time consumed by CPU during processing of requests by Agent.

5.2 Simulation Parameters

In this section, parameters used for simulation analysis have been discussed. To simulate a proposed mechanism, nodes are varied from 100 to 400. The simulation area is 100 × 100, and the initial energy of the node is 0.5Jule. Table 1 shows the complete parameters used for simulation.

5.3 Discussion

The simulation results of proposed and existing mechanisms are discussed in Figs. 1 and 2. In Fig. 1, a comparative analysis with the throughput parameter is presented, whereas Fig. 2 presents the latency of both proposed and existing techniques. The throughput is high in the proposed mechanism, whereas latency is less in the perspective of the existing single agent-based LB technique.
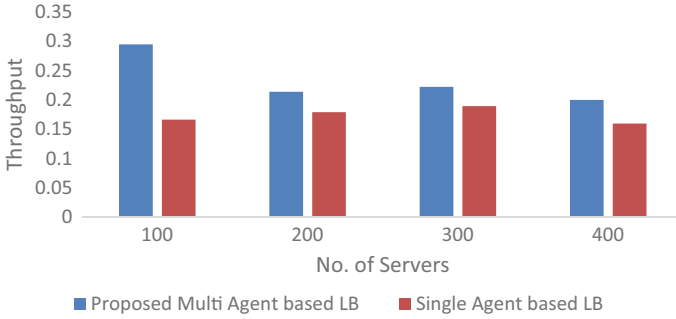
Fig. 1 Throughput of proposed Multi Agent_LB and Single Agent_LB
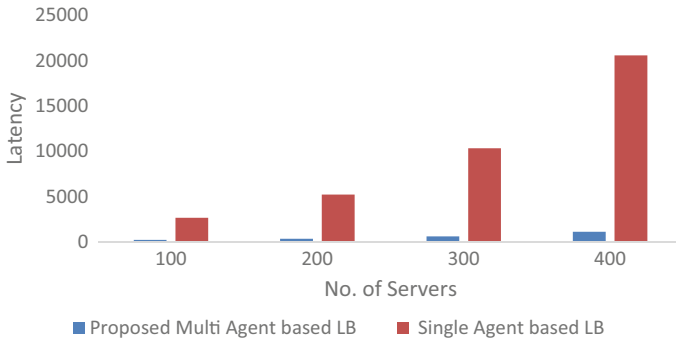
Fig. 2 Latency of proposed Multi Agent_LB and Single Agent_LB

6 Conclusion

The existing system model agent-based dynamic load-balancing approach has a disadvantage, i.e. for every instance of a new job, the agent has to complete a life cycle (i.e. calculate the average and decide the status of the server). Hence, agents need to travel twice to allocate the job: once to find the average work load and a second one to balance the loads among nodes. To overcome this, we propose a multi-agent architecture for agent-based load balancing. In this paper, an attempt has been made to propose a multi-agent-based load-balancing mechanism for cloud environments. A distributed load-balancing strategy based on multi-agent is proposed in this paper. The proposed mechanism has been simulated with the help of MATLAB. The simulation results show that throughput is high in the proposed mechanism, whereas latency is less in the perspective of the existing single agent-based LB technique. In the future, it is intended to continue working on it and apply the proposed mechanism in other environments, such as edge computing, which also improves throughput with less end-to-end delay.

References

1. Zhang J, Liu Q, Chen J (2016) A multi-agent based load balancing framework in cloud environment. In: 2016 9th international symposium on computational intelligence and design (ISCID), vol 1. IEEE, pp 278–281
2. George SS, Pramila RS (2021) A review of different techniques in cloud computing. Mater Today Proc 46:8002–8008
3. Shafiq DA, Jhanjhi NZ, Abdullah A, Alzain MA (2021) A load balancing algorithm for the data centres to optimize cloud computing applications. IEEE Access 9:41731–41744
4. Annie Poornima Princess G, Radhamani AS (2021). A hybrid meta-heuristic for optimal load balancing in cloud computing. J Grid Comput 19(2):1–22
5. Mohanty S, Patra PK, Ray M, Mohapatra S (2021) A novel meta-heuristic approach for load balancing in cloud computing. In: Research anthology on architectures, frameworks, and integration strategies for distributed and cloud computing. IGI Global, pp 504–526
6. Balaji K (2021) Load balancing in cloud computing: issues and challenges. Turk J Comput Math Educ 12(2):3077–3084
7. Kapila D, Dhir V (2021) Performance evaluation of new hybrid appraoch of load balancing in cloud computing. Des Eng 698–716
8. Jangra A, Mangla N, Jain A, Dewangan BK, Perumal T (2021) Classification of various scheduling approaches for resource management system in cloud computing. In: Autonomic computing in cloud resource management in industry 4.0. Springer, Cham, pp 149–157
9. Jangra A, Dubran H (2021) Simulation annealing based approach to enhanced load balancing in cloud computing. In: 2021 9th international conference on reliability. Infocom technologies and optimization (trends and future directions) (ICRITO). IEEE, pp 1–4
10. Mishra K, Majhi S (2020) A state-of-art on cloud load balancing algorithms. Int J Comput Digital Syst 9(2):201–220
11. Neelima P, Reddy A (2020) An efficient load balancing system using adaptive dragonfly algorithm in cloud computing. Clust Comput 23(4):2891–2899
12. Semmoud A, Hakem M, Benmammar B, Charr JC (2020) Load balancing in cloud computing environments based on adaptive starvation threshold. Concurr Comput Pract Exp 32(11):e5652
13. Jangra A, Mangla N (2021) Cloud LB using optimization techniques. In: Mobile radio communications and 5G networks. Springer, Singapore, pp 735–744
14. Singh N, Elamvazuthi I, Nallagownden P, Ramasamy G, Jangra A (2021) Assessment of microgrid communication network performance for medium-scale IEEE bus systems using multi-agent system. In: Mobile radio communications and 5G networks. Springer, Singapore, pp 377–387
15. Patel KD, Bhalodia TM (2019) An efficient dynamic load balancing algorithm for virtual machine in cloud computing. In: 2019 International conference on intelligent computing and control systems (ICCS). IEEE, pp 145–150
16. Junaid M, Sohail A, Rais RNB, Ahmed A, Khalid O, Khan IA, Huassin SS, Ejaz N (2020) Modeling an optimized approach for load balancing in cloud. IEEE Access 8:173208–173226
17. Babou CSM, Fall D, Kashihara S, Taenaka Y, Bhuyan MH, Niang I, Kadobayashi Y (2020) Hierarchical load balancing and clustering technique for home edge computing. IEEE Access 8:127593–127607
18. Shahid MA, Islam N, Alam MM, Su'ud MM, Musa S (2020) A comprehensive study of load balancing approaches in the cloud computing environment and a novel fault tolerance approach. IEEE Access 8:130500–130526
19. Devaraj AFS, Elhoseny M, Dhanasekaran S, Lydia EL, Shankar K (2020) Hybridization of firefly and improved multi-objective particle swarm optimization algorithm for energy efficient load balancing in cloud computing environments. J Parallel Distrib Comput 142:36–45
20. Saeedi P, Shirvani MH (2021) An improved thermodynamic simulated annealing-based approach for resource-skewness-aware and power-efficient virtual machine consolidation in cloud datacenters. Soft Comput 25(7):5233–5260

Viable Methods Adopted for Reducing SAR Value in Mobile Phone Antenna: A Review

M. Manikandan and S. Karthigai Lakshmi

Abstract This article presents a detailed review of different methods adopted from earlier to recent years to achieve low Specific Absorption Rate (SAR) value for mobile phone antennas. SAR value should be within the limit according to the available standards. The human fraternity should limit the use of mobile phone. In order to infer this, some proven biological effects caused due to heavy radiation exposure of mobile phones are listed. Due to this serious concern, the SAR reduction methods should be instantly reviewed. From the broad review, the SAR reduction methods by adding supplementary elements such as conducting materials, reflector, ferrite shielding, directional antennas and resistive cards are not suitable for recent mobile technology in terms of size and cost. Even though it gives better SAR reduction, it does not support multiband operation. The gain and efficiency of those methods are also low. The SAR reduction methods by modifying main antenna structure such as AMC structures, phased array, optimizing current distribution and EBG structures do not made good agreement with all parameters with SAR reduction. By analyzing all the methods, this article proposed suitable structure called split-ring resonator-based metamaterial that made good agreement with all parameters essential for viable product. This structure supports multiband of operation with the maximum efficiency of 80%, the maximum gain is above 3 dB, the return loss is less than -34 dB, and the SAR can be reduced up to 84%.

Keywords SAR · EBG · Metamaterials

M. Manikandan (✉) · S. K. Lakshmi
SSM Institute of Engineering and Technology, Dindigul, Tamilnadu, India
e-mail: manisece910@gmail.com

285

1   Introduction

The usage of mobile telephone has shoot up day by day that emits electromagnetic radiation which results in biological effects among living things particularly humans. The International Agency for Research on Cancer (IARC) at World Health Organization (WHO) has said that the radio frequency electromagnetic field (RF-EMF) from cellular phones is non-ionizing, as a Group 2B, i.e., a possible human carcinogen [1, 2]. Many articles have been published earlier and recently concluded that mobile phone is a suspicious device to produce many biological effects on human body such as thermal effects, blood–brain barrier effects [3], genetical effects [4], sleep and EEG effects [5, 6], non-thermal effects [7], cognitive effects [8], electromagnetic hypersensitivity [9], behavioral effects [10], infertility problems [11] and problem in salivary secretion [12]. The bottom line characteristics of mobile phone which assure the radiation exposure in humans to the safety level are Specific Absorption Rate (SAR). The guidelines given by International Commission on Non Ionizing Radiation Protection (ICNIRP) in 1998 [13], US Federal Communication Commission (FCC) in 1997 [14] and Institute of Electrical and Electronic Engineers (IEEE) in 2005 [15] form the primary regulation for ensuring safety levels with respect to human exposure to RF-EM waves. According to those guidelines, maximum SAR value should be 2 W/kg averaged over any 10 g volume of tissue and 1.6 W/kg averaged over any 1 g volume of tissue. Several methods have been proposed in the past years to reduce the SAR below ongoing levels of the international standards [16]. Earlier researches were done to reduce SAR value based on adding supplementary elements along with main antenna like conducting materials, reflectors, etc. Later, the researches were uplifted to modify the structure of main antenna like EBG, metamaterials, etc., and the recent methodologies follow the same. This article presents a wide review of past and present methods and gives a suitable antenna structure that makes a good agreement with all aspects especially in achieving multiband, efficiency, size and cost with reduced SAR value. In Sect. 2, SAR reduction methods by adding supplementary elements were analyzed; in Sect. 3, recent methodologies proposed to reduce SAR value by modifying the main antenna structure are presented and reviewed, followed by the summary and future directions.

2   SAR Reduction Methods by Adding Supplementary Elements

SAR can be reduced by adding supplementary elements like conducting materials, reflectors, directors, ferrite shielding, resistive cards, etc. Conducting materials play a crucial role in SAR reduction process, because it affords more electric field strength which results in increase in SAR value. The increase in electric field strength mainly depends upon the position of conducting materials [17]. Four different antennas are placed over the conducting surfaces at proper places, and increasing an antenna

distance from 0 to 20 mm leads to reduction in SAR value. The percentage of reduction for four different antennas is listed in Table 1. The constraints behind the technique are to place the conducting material in proper position. If an antenna structure is formed along with a reflector element, the radiation efficiency can be increased in conjunction with SAR. The percentage of reduction in SAR can be improved by using dual reflector when compared to single reflector [18]. A ferrite sheet will reduce induced current as it is a less conductive material that nullifies the magnetic field. The SAR mainly depends on the density of surface current, so ferrite sheet can be placed nearer to the high surface current density area to reduce SAR [19]. The main constraints behind adopting ferrite sheet are lagging in radiation efficiency, gain and bandwidth. About 99% of SAR reduction is possible using directional antennas at very high frequency [20]. Omnidirectional antennas are preferred for mobile phones because signals from all directions must be received, and using directional antennas will result in poor receptions in other directions; also it is not suitable for low frequencies. A resistive card with good EM properties has the ability to alter the impedance that reduces or blocks the EM waves in unwanted area [21]. The important constraints in this method are when the antenna is wrapped with resistive card, impedance mismatch occurs that affects radiated power gain. Table 1 shows SAR reduction by using supplementary elements with its frequency of operation which shows that supplementary elements do not support multiband operation. Even though percentage of SAR reduction is more from these methods, it will add product cost and size. From the operating mechanism of these structures, the gain and radiation efficiency are affected more.

Table 1 Percentage of SAR reduction in various methods by adding supplementary elements

Reference no	Methods	Frequency in GHZ	SAR reduction	
[17]	Conducting materials	2.2	λ/4 monopole	90.3%
			Helix	76.79%
			Patch	86.05%
			PIFA	80.62%
[18]	Reflector	0.85 to 2.2	Single reflector	83.01%
			Two reflectors	88.74%
[19]	Ferrite shielding	0.9	–	57.75%
[20]	Directional antennas	30	–	99%
[21]	Resistive cards	1.8	–	64.7%

3 SAR Reduction Methods by Modifying Main Antenna Structures

From the limitations of adding supplementary elements with main antenna structure, it is a challenging task for the antenna engineers to reduce SAR value in mobile phone antenna with increased gain and radiation efficiency. Low cost and size are also an essential requirement of a valuable product. By considering the above aspects, adopting changes within structure of main antenna will give satisfying results. Beam steering array, adopting AMC structures, EBG structure, optimizing surface current and metamaterials are the viable methods in recent years utilized for reducing SAR value in mobile phones. The operating methodology followed and the recent research on above methods with its convincing ideas and results are discussed further.

3.1 Adopting AMC Structure

The artificial magnetic conductors are periodic array structures implemented with other primary structures which reflect the incident waves with zero reflection phases. An AMC structure with 2×3 square patch array implemented with another square patch antenna [22] reduces SAR level up to 93.7% with improved gain of 8.69 dBi, and the radiation efficiency of this structure is more than 80%. A single rectangular split-ring antenna with AMC structure has been reported in [23] for wearable application. An AMC structure is backed with square metal patch covered by an outer square metal ring, and substrate is made up of leather with metallic ground plane. SAR value is measured for 1 g tissue of hand, chest and head. The AMC structure reduces SAR level to 89% in head tissue. It operates only at 5.8 GHZ with improved gain of 7.47 dB. A wearable 2×2 array of dumbbell-shaped AMC structure with monopole antenna is implemented for dual band of operation [24]. The gain is optimized up to 9.5 dB and 8.1 dB at 2.43 GHZ and 5.2 GHZ. AMC structure reduces SAR by more than 95% in hand model at both frequencies. In the above reported work, by including AMC structure, SAR reduction can be achieved above 80%, the gain value is above 7 dB, and this is due to the property of AMC structure. At resonant frequency, AMC will behave like high impedance surface, at that time the reflection phase will be zero for entire surface, and hence all the incident waves are reflected back which creates constructive interference with source wave. This leads to improved gain and reduced SAR value. Additional space is required for AMC structure, in above research works; it is not suitable for multiband of operation. Even though AMC structure is considered to be supplementary element, recently many works were reported based on this, so this method was included in this section.

3.2 Adopting Phased Array

The field strength can be improved in particular direction by using array antenna. For this, all the similar elements in an array should be oriented similarly. We can change the direction of main beam radiated by an array antenna by exiting each element in an array properly. A two-element λ/4 and two-element λ/2 arrays in [25] are exited properly which redirect the power opposite to head. The SAR reduction obtained for two-element λ/2 array is 91%.

Instead of forming array with elements, slots can be introduced to avoid additional spacing [26]. In this method, two subarrays are formed, one on top and other on the back surface of the mobile case, with 8 slots on each subarray. Subarray 1 is switched for talk mode; subarray 2 is switched for data mode. The beam can be steered from an array by giving progressive phase to each slot in an array. The detailed representation of beam steering array is shown (see Fig. 1). The excitation given to each slot in an antenna array is tilted by certain angle which leads to beam coverage in desired direction for both modes. The SAR value measured for subarray 1 with zero degree scan angle and 40 degree scan angle is 0.53 W/kg and 0.88 W/kg for 28GHZ. In these methods, maintaining phase for multiband is not possible. We can include additional elements for each band that increases design cost and size. It is suitable for high frequency only.
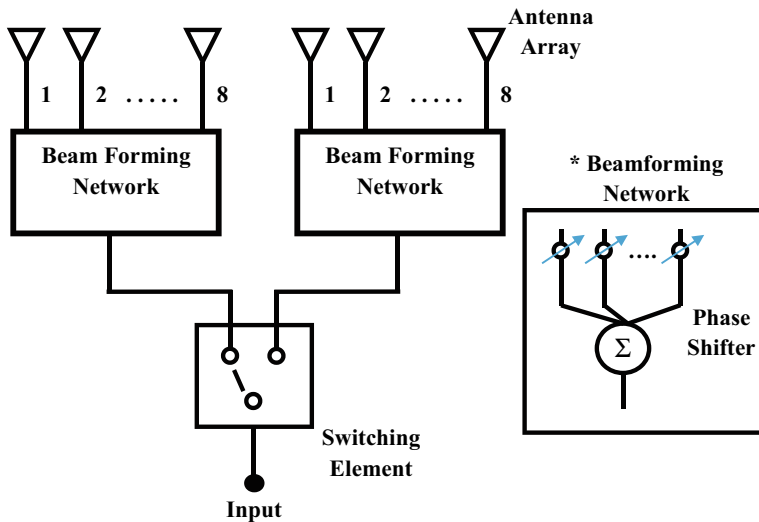
Fig. 1 Phased array architecture for beam steering array antenna [26]

3.3 Optimizing the Current Distribution

The surface current distribution can be optimized in two ways to reduce SAR value: one is by introducing reverse current and another way is to reduce surface current. SAR is directly proportional to the square of electric field strength (E) form Eq. (1). We can alter the distribution of electric (E) and magnetic field (H) strength by adjusting the surface current distribution (J) from Eqs. (2) and (3).

$$SAR = \frac{\sigma |E|^2}{\rho} (W/kg) \tag{1}$$

$$J = \sigma E \tag{2}$$

$$\nabla \times H = j\omega D + J \tag{3}$$

So by varying the surface current distribution, we can alter electric and magnetic field strength which in turn reduces SAR. Two metal-rimmed mobile phone antennas were designed in [27]. Antenna 1 is designed without parasitic branch, and Antenna 2 is designed with parasitic branch as shown in figure (see Fig. 2).

The two antennas are designed in such a way that to cancel the tangential component of electric field by introducing reverses current which will reduce SAR. From the boundary condition, tangential components of electric field radiating from mobile phone and the field penetrating to human body will be the same, and the normal component penetrating to human body will be very less. By introducing parasitic branch in Antenna 2, a pair of reverse current having equal magnitude was generated on metal rim and its neighboring ground. The tangential electric field generated by the reverse current partially canceled each other, resulting in very small amount of electric field induced, that leads to SAR reduction. As the electric field canceling effect in Antenna 1 is lesser than Antenna 2 due to parasitic branch, Antenna 2 will produce lesser SAR value than Antenna 1 with dual band of operation. Another way of optimizing current distribution is to reduce it by introducing rectangular slot at the place of high current density area [28]. But the structure supports only single band
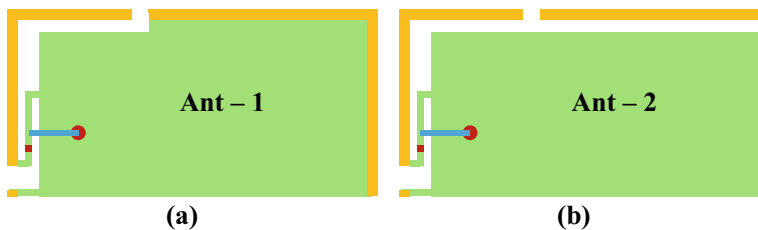
Fig. 2 **a** Ant 1 without parasitic branch, **b** Ant 2 with parasitic branch

of operation. Even though these methods avoid additional space for SAR reduction, it is not desirable for multiband of operation with higher SAR reduction rates.

3.4 Adopting Metamaterials (SRR)

Metamaterials are factitious materials which possess remarkable properties such as negative permittivity, negative permeability and negative refractive index. At desired frequency band of operation due to above-mentioned properties, this exhibits stop band characteristics. By using the property of metamaterial, the SAR value is reduced without downgrading radiation efficiency and radiated power. It also improves isolation between antenna elements by suppressing the surface waves [29] which in turn reduce SAR. Metamaterials are classified in to different categories based on permittivity and permeability, among those, split-ring resonator (SRR) which falls under negative permeability materials plays crucial role in multiband of operation. The split-ring shape may be square, spiral, hexagonal, rectangular, circular, pentagonal, etc. Some of double SRR in different shapes are shown in figure (see Fig. 3).

Increasing the split gap, spacing between adjacent rings and metal width of SRR leads to increase in resonant frequency, multiple split gaps in SRR shifts resonant frequency toward higher frequency [29]. A multiple split square ring that exhibits dual band of operation reduces SAR when exposed to human head [30]. Table 2 shows SAR reduction comparison for 1 g and 10 g of tissue volume for multisplit square ring.

Increasing the number of circular split ring increases number of resonant frequency [31]. From the parametric study of this research, each circular split-ring resonator contributes one resonant frequency. The effective negative permeability is obtained from scattering parameters to ensure the existence of metamaterial property. In evidence with the data of stop band and pass band characteristics undoubtedly, the SAR value will be very low for this circular SRR structure. Introducing an array of pentagonal split rings increases radiation efficiency, gain and reduces SAR value up to 84.5% [32].

The fabrication of epsilon negative (ENG) SRR with square SRR results in 93% of SAR reduction [33]. From the above literatures, we can observe that varying the
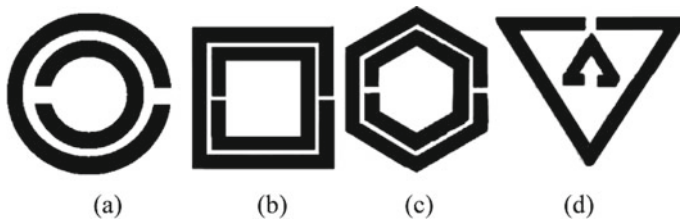
Fig. 3 **a** Circular SRR, **b** square SRR, **c** hexagonal SRR, **d** triangular SRR

Table 2 Comparison of SAR values of multiple split square ring metamaterial over 1 g and 10 g volume of tissue

Frequency in GHZ	1 g (W/Kg)		10 g (W/Kg)	
	Without MTM	With MTM	Without MTM	With MTM
		SAR reduction		SAR reduction
1.81	1.329	0.839	0.799	0.517
		36.86%		35.29%
3.79	1.728	0.941	0.732	0.378
		45.54%		48.36%

parameters of SRR not only reduces SAR but also improves other parameters such as radiation efficiency and gain and supports multiband of operation.

3.5 Adopting EBG Structures

EBG plays a major role in suppressing the surface waves and improving the gain of an antenna. The EBG exhibits stop band characteristics which behave like AMC that redirects the radiation opposite to human body in turn produces low SAR value [34]. A circular microstrip antenna with electromagnetic band gap structure was designed [35]. By adding M-shaped EBG unit cells around CMSA, the SAR value is reduced to 74.50% from 2.09 to 0.533 W/kg. On the other side, a metallic dual-feed annular ring antenna with circular defected ground structure (DGS) and EBG was developed [36]. SAR is evaluated for four different frequencies. When an array of mushroom-type EBG unit cells is arranged near dual feed lines, 83.64% of SAR reduction is obtained for 1.8 GHZ. The same author proposed hybrid EBG structures consists of an array of spiral EBG and M-shaped EBG placed near feed line of CMSA along with DGS [37]. The SAR value was analyzed for three different frequencies. Maximum SAR reduction is 87.05% achieved for 5.5 GHZ.

From the comparative study of all EBG structure reported above, it is made clear that the combination of EBG structure and defected ground structure (DGS) was effectively reducing SAR at greater level. In those combinations, when EBG structures are placed near feed lines, SAR will be reduced further. The design complexity is very high when compared to other structures like split-ring resonators. Also care must be taken to reduce the size of the overall antenna structure. Table 3 gives detailed parameters comparison of various methods discussed in Sect. 3 with SAR reduction (in head model).

Table 3 SAR reduction and parameters comparison of various recent methods (modifying main antenna structures)

Ref no., year	Methods	Frequency (GHZ)	Efficiency	SAR reduction (%)	Return loss	Gain (dB)
[23], 2021	AMC structures	5.8	64%	89	− 40 dB	7.47
[26], 2018	Beam steering array	28	–	90.75	Poor	14
[27], 2020	Reverse current technology	0.897	31–87%	23	− 12 dB	–
		1.745		46	− 12 dB	
		1.95		38	− 17 dB	
		2.535		39	− 28 dB	
[28], 2021	Adjusting magnetic field	2.45	60%	33.3	− 38 dB	–
[32], 2021	Metamaterials (SRR)	2.4	65–80%	84.5	− 15 dB	3
		5			− 34 dB	
		3.5			− 23 dB	
[37], 2022	EBG	2.4	–	82.39	Poor	2
		3.5		77.13		2.5
		5.5		87.05		2.5

4 Summary and Future Directions

The major requirement of a current mobile phone antennas is need of multiband frequency of operation with good efficiency, maximum gain, reduced return loss at operating frequency, simple design, reduced size and cost. However, an antenna structure made good agreement with the above parameters, and the primary requirement is SAR reduction. In this article, a comprehensive review was performed by analyzing all the viable methods so far proposed to reduce SAR value in mobile phone antenna. The methods discussed in Sect. 2 reduce SAR value by adding supplementary elements. The collective constraints noted from all those methods are increasing the size and cost of the equipment, the structures reduces antenna efficiency and gain also. Mostly, these methods are developed with external antenna structures, but nowadays for compact mobile phones, only internal antennas are preferred. The main antenna structures that are modified to overcome above limitations were reviewed in Sect. 3. Table 3 shows the SAR reduction and parameters comparison of various recent methods proposed from 2017 to 2022. From the detailed review, the structures like AMC [22–24], phased array [25, 26] and optimizing the current distribution [27, 28] were not compromising with all the parameters like gain, efficiency, size and cost with reduced SAR value. Even though EBG structures reported in [35–37] give better SAR reduction and support multiband of operation, the size is quite large and the design is more complex. The simple metamaterial-based SRR structure will be

the best option suitable for multiband of operation and reduces SAR and return loss at greater level with improved efficiency and gain [31, 32]. With the evidence of above research, a simple modification is further required to incorporate all the necessary parameters to build a perfect antenna by using metamaterial SRR structure. The limitation in this method is: size can be further reduced as only limited space is allotted for antennas inside the mobile phones. In the future, mobile phone antenna developers can focus on SRR-inspired metamaterial structure to reduce SAR value for multiband operation which satisfies the current demand of mobile phones.

References

1. Baan R, Grosse Y, Lauby-Secretan B, Ghissassi FEI, Bouvard V, Benbrahim-Talla L (2011) Carcinogenicity of radiofrequency electromagnetic fields. Lancet Oncol 12:624–626
2. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans (2013) Non ionizing radiation, part 2: radiofrequency electromagnetic fields, vol 102. IARC Press, Lyon
3. Salford L, Arne B, Eberhardt J, Malmgren L, Bertil R, Persson R (2003) Nerve cell damage in mammalian brain after exposure to microwaves from GSM mobile phones. Environ Health Perspect 111(7):881–883
4. VIjayalaxmi PT (2008) Genetic damage in mammalian somatic cells exposed to radiofrequency radiation: a meta-analysis of data from 63 publications (1990–2005). Radiation Res 169(5):561–574
5. Borbély A, Huber R, Graf T, Fuchs B, Gallmann E, Achermann P (199) Pulsed high frequency electromagnetic field affects human sleep and sleep electroencephalogram. Neurosci Lett 275(3):207–210
6. Andrzejak R, Poreba R, Poreba M, Derkacz A, Skalik R, Gac P, Beck AB, Steinmetz-Beck PW (2008) The influence of the call with a mobile phone on heart rate variability parameters in healthy volunteers. Indus Health 46(4):409–417
7. Blank M, Goodman R (2009) Electromagnetic fields stress living cells. Pathophysiology 16(2–3):71–78
8. Roy L, Ilan E, Ronen H, Menachem M, Nachshon M (2009) Cognitive effects of radiation emitted by cellular phones: the influence of exposure side and time. Bioelectromagnetics 30(3):198–204
9. Roosli M (2008) Radiofrequency electromagnetic field exposure and non-specific symptoms of ill health: a systematic review. Environ Res 107(2):277–287
10. Aldad T, Gan G, Gao X (2012) Fetal radiofrequency radiation exposure from 800–1900 MHz-rated cellular telephones affects neuro development and behavior in mice. Sci Rep 2:312
11. Agarwal A, Deepinder F, Sharma R, Ranga G, Li J (2008) Effect of cell phone usage on semen analysis in men attending infertility clinic: an observational study. Fertile Sterile 89(1):124–128
12. Kamyab N, Mordouei Z, Hosseini M, Sheikh Fathollahi M (2021) The effects of mobile phone waves on salivary secretion in dental students of Rafsanjan university of medical sciences. Int J Rad Res 19(1):81–87
13. International Commission on Non-Ionizing Radiation Protection (ICNIRP) (1998) Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (Up to 300 GHz). Health Phys 74:494–522
14. US Federal Communication Commission, Office of Engineering and Technology (1997) Evaluating compliance with FCC-specified guidelines for human exposure to radio radiofrequency radiation. OET Bulletin 65, Washington, DC
15. Institute of Electrical and Electronic Engineers (IEEE), IEEE C95.1-2005 (2005) Standards for safety levels with respect to human exposure to radio frequency electromagnetic fields. IEEE Press, New York

16. Pikale R, Sangani D, Chaturvedi P, Soni A (2018) A review: methods to lower specific absorption rate for mobile phones, international conference on advances in communication and computing technology (ICACCT). IEEE, 978-1-5386-0926-2/18/$31.00 ©2018
17. Faruque M, Islam M, Misran N (2011) Analysis of SAR levels in human head tissues for four types of antennas with portable telephones. Aust J Basic Appl Sci 5:96–107
18. Haridim M (2016) Use of rod reflectors for SAR reduction in human head. IEEE Trans Electromag Compatib 58(1):40–46
19. Islam M, Faruque M, Misran N (2009) Design analysis of ferrite sheet attachment for SAR reduction in human head. Prog Electromag Res 98:191–205
20. Kang W, Rahmat-Samii Y (1998) Handset antennas and humans at Ka-band: the importance of directional antennas. IEEE Trans Antennas Propag 46(6):949–950
21. Chou H, Hsu H, Chou H, Liu K, Kuo F (2009) Reduction of peak SAR in human head for handset applications with resistive sheets (R-cards). Prog Electromag Res 94:281–296
22. Mumin ARO, Alias R, Abdullah J, Dahlan SH, Ali J, Debnath SK (2020) Design a compact square ring patch antenna with AMC for SAR reduction in WBAN applications. Bull Elect Eng Inform 9(1):370–378
23. Saha P, Mitra D, Susanta Parui K (2021) Control of gain and SAR for wearable antenna using AMC structure. Radio Eng 30(1):81–88
24. Bidisha H, Banani B, Arnab N (2021) Design of dual-band conformal amc integrated antenna for SAR reduction in WBAN. Prog Electromag Res C 110:91–102
25. Mangoud M, Abd-Alhameed R, McEwan N, Excell P, Abdulmula E (1999) SAR reduction for handset with two-element phased array antenna computed using hybrid MoM/FDTD technique. Elect Lett 35(20):16931694
26. Jihoon B, Jaehoon C (2018) A SAR reduced MM-wave beam-steerable array antenna with dual-mode operation for fully metal-covered 5G cellular handsets. IEEE Antennas Wirel Propag Lett. https://doi.org/10.1109/LAWP.2018.2836196
27. Zhang HH, Yu GG, Liu Y, Fang YX, Shi G, Wang S (2021) Design of low-SAR mobile phone antenna: theory and applications. IEEE Trans Antennas Propag 69(2):698–707
28. Bao L, Bo P, Wei H, Wen J (2021) Low-SAR antenna design and implementation for mobile phone applications. IEEE Access 9:96444–96452
29. Praveen K, Tanweer A, Manohara Pai MM (2021) Electromagnetic metamaterials: a new paradigm of antenna design. IEEE Access 9:18722–18751
30. Ramachandran T, Faruque MRI, Ahamed E, Abdullah S (2019) Specific absorption rate reduction of multi split square ring metamaterial for L-and S-band application. Results Phys 15:1–10
31. ThamilSelvi N, Thiruvalar Selvan P, Babu SPK, Pandeeswari R (2020) Multiband metamaterial-inspired antenna using split ring resonator. Comput Elect Eng 84:1–13
32. Imaculate R (2021) A triple-band antenna with a metamaterial slab for gain enhancement and specific absorption rate (SAR) reduction. Prog Electromag Res C 109:275–287
33. Rosaline S, Raghavan S (2015) A compact dual band antenna with an ENG SRR cover for SAR reduction. Microw Opt Technol Lett 57:741–747
34. Priyanka D, Sanjeev KD (2019) Electromagnetic band gap structure applications in modern wireless perspective: a review. In: Proceedings of international conference on advances in computing & communication engineering (ICACCE-2019). IEEE, Sathyamangalam
35. Mahesh M, Anil N, Shankar D (2019) Low specific absorption rate antenna using electromagnetic band gap structure for long term evolution band 3 application. Prog Electromag Res M 80:23–34
36. Munde M, Nandgaonkar A, Deosarkar S (2021) Performance optimization of dual-feed UWB annular ring antenna with circular DGS and EBG for SAR reduction. Prog Electromag Res C 115:51–64
37. Munde MM, Nandgaonkar AB, Deosarkar SB (2022) Ultra-wideband circular microstrip antenna with hybrid EBG for reduced SAR. Adv Electromag (AEM) 11(1):51–57

Numerical Modelling of Cu$_2$O-Based Gas Sensor for Detection of Volatile Organic Compounds in Human Breath

Rahul Gupta, Anil Kumar, Mukesh Kumar, Pradeep Kumar, and Dinesh Kumar

Abstract The biomarkers presented in the human breath can lead to the detection of underlying diseases and alcohol detection. The semiconductor-based sensors are well suited for detection of volatile organic compounds (VOCs). The Wolkenstein-based numerical model is used to compute the sensor response of Cu$_2$O-based sensor towards VOCs. The adsorption–desorption equilibrium has been used to compute the sensor response for different concentrations of target gas analytes. The model has been simulated for various temperature range to obtain the optimized operating range of the sensor device. The plot between the normalized electrical conductance and the frequency provided the cut-off frequency (10^8–10^9 Hz). This cut-off frequency was further used to analyze the variations in sensor response. The presented design model of the gas sensor device exhibits improved sensor response towards 0.05 atm partial pressure in the operating temperature range of 360–380 K. Further, the operating temperature range of sensor device was found ~ 375 K.

Keywords Semiconductor sensor · Numerical modelling · Volatile organic compound (VOC)

R. Gupta (✉) · P. Kumar
J. C. Bose University of Science and Technology, YMCA, Faridabad, Haryana, India
e-mail: rgupta.iit@gmail.com

R. Gupta
University Institute of Engineering and Technology, Kurukshetra University, Kurukshetra, Haryana, India

A. Kumar · M. Kumar · D. Kumar
Department of Electronic Science, Kurukshetra University, Kurukshetra, Haryana, India

D. Kumar
Gurugram University, Gurugram, Haryana, India

1   Introduction

Human breath consists of various volatile organic gas compounds like ethanol, methanol, and gases like CO_2, CO, etc. These gases and VOCs are used as biomarkers for detection of various underlying diseased in the human body [1]. These compounds and gases can be detected and analyzed using semiconductor-based gas sensors. Gas sensors are becoming popular nowadays for daily life real-time applications like alcohol detection from breath, shell life prediction of eatables, industrial safety sensors, natural gas leakage detection [2], pollution analysis, and other biomedical sensor systems [1, 3–5]. Semiconductor-based sensors being small in size and low-cost and commercial availability are mostly used for all these applications. The semiconductor sensors based on Cu_2O, SnO_2, ZnOand Cr_2O_3, CuO_2, and CuO are few of the investigated materials by various researchers in the past decade [2, 6]. The sensing mechanism of such sensors includes oxidation and reduction processes involving the target analytes and the semiconductor surface. Adsorption causes generation of localized electrons states at the surface of the semiconductor metal oxide due to adsorption of oxygen from the environment. These electron states initiate a charge transfer between chemisorbed species and semiconductor by trapping of electrons or holes based on n-type or p-type semiconductor, respectively [7]. This charge transfer causes significant change in electrical resistance of the semiconductor device. Eventually, this relative change in electrical resistance can be measured and further analyzed to detect the presence of a particular gas at given temperature and pressure [8].

After the interaction of target gas/compound with the adsorbed oxygen atoms at the surface of the semiconductor, the electrical resistance of the surface will change depending upon the nature and concentration of the VOCs molecules as shown in Fig. 1. This relative resistance change is computed and reflected in terms of sensor response of the material.

The Wolkenstein proposed the gas sensor model based upon the chemisorption of gas analytes at sensor surface [9]. In the presented research work, this model has been improvised for the further analysis of sensor response in frequency domain yields the characteristics of presence of various VOCs and their concentration. This analysis is performed using noise spectroscopy technique mapped with basic model of chemisorption [10–13]. Additionally, the presented model has been further investigated and modified to study the sensor response for reducing gases like carbon monoxide, ammonia, etc., and some of the volatile organic compounds.

In Sect. 2 of the manuscript, the proposed numerical theory is described in detail. Moreover, the adsorption–desorption model is discussed to describe the sensing

Fig. 1   Basic steps involved in detection of VOCs using semiconductor-based gas sensors

mechanism of the semiconductor-based sensors. The simulated results of normalized electrical conductivity and sensor response at varying temperature and partial pressure of VOCs and related discussions are presented in Sect. 3.

2 Numerical Model

The presented work discusses the mathematical model to compute the sensor response to detect human breath containing VOCs using chemisorption at sensor surface. In this model, the first step is to initialize all the parameters like energy level of induced surface states due to chemisorption of environmental oxygen, the difference between conduction band at surface and Fermi level, doping concentration for Cu$_2$O, binding energy of neutral atoms, reaction rate constant of VOCs molecules, number of adsorption centre per unit surface area, oscillation frequency of charged and neutral chemisorbed species, molecular mass of oxygen, etc. Second step is modelling of chemisorption: this is done in further two steps of adsorption and desorption of oxygen molecules and the target VOCs molecules at the surface. The adsorption of environmental oxygen gas molecules in the dissociated form occurs at the Cu$_2$O surface. These gas molecules would be involved in the charge sharing with the sensing layer. The charge sharing at the sensor surface would result in the creation of a space charge region at the surface and hence the change in the film electrical resistance. Further, the desorption process of reversal of the adsorption occurs at surface of the sensor. The numerical model has been designed to calculate the adsorption–desorption equilibrium which will compute electrical conductance in the presence of oxygen and then in the presence of gas. Figure 2 explains the complete flow to compute the sensor response using the proposed model.
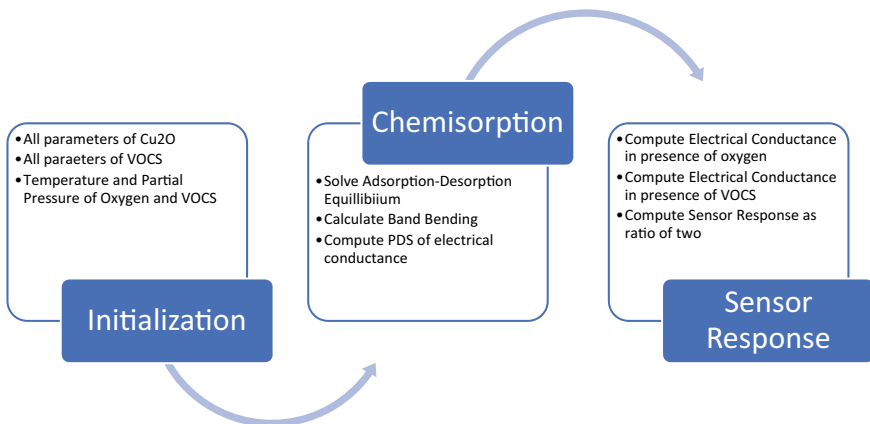
Fig. 2 Proposed flow of mathematical model for computation of sensor response

Furthermore, the sensor response (SR) has been computed as the ratio of electrical conductance before and after the presence of VOCs molecules. The equation used to calculate the sensor response is given below as

$$\text{Sensor response,} \quad SR = \frac{\text{Electrical conductance in presence of VOCs molecules}}{\text{Elecrtrical conductance in presence of } O_2}$$

(1)

The simulation flow of proposed model involves the computation of adsorption–desorption equilibrium of oxygen molecules at the sensor surface. The equilibrium state is called to be attained when the adsorption $a(N)$ and desorption $d(N)$ rate of oxygen ions become equal. The net effective adsorption rate is given by Eq. 2.

$$\frac{dN}{dt} = a(N) - d(N) = \alpha p \left(N^* - N \right)$$
$$- \left(\begin{array}{c} N^0 \vartheta^0 \exp\left(--\frac{q0}{kT}\right) + \\ N^- \vartheta^- \exp\left(-\frac{q0 + (E_c^b - E_{ss})}{kT}\right) \end{array} \right)$$

(2)

where N^* is the no. of adsorption centres/area,

$N = N^0 + N^-$ is total no. of adsorbed particles/area
N^0 and N^- are the no. of neutral adsorbate and chemisorbed adions/area, respectively
q^0 is binding potential
E_c^b is the energy level in the bulk of Cu_2O of conduction band
E_{ss} is the energy level of the chemisorbed surface state .
ϑ^0 and ϑ^- are the oscillation frequency of corresponding molecules

After calculation of equilibrium, when the sensor surface is exposed to gas molecules, the chemical reaction in Eq. 3 takes place, and the adsorption equilibrium shifts either to positive or negative side depending upon the gas nature.

$$VOC(gas) + O^- \rightarrow CO_2 + H_2O + e^-$$

(3)
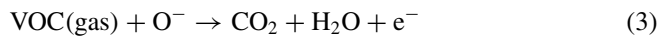
This shift in equilibrium is computed again using the same procedure as done for the adsorbed oxygen atoms/molecules. Hence, the change in electrical resistance after the exposure of VOCs has been calculated, and Eq. 1 is further solved for computing the sensor response toward particular gas analytes at given temperature and partial pressure.

3 Results and Discussion

The mathematical simulations were carried out using open-source platform Jupyter notebook using Python as the tool. The sensor parameters used for the presented model simulations have been extracted for n-type Cu$_2$O semiconductor material, energy band gap, E_{ss}, temperature, partial pressure of oxygen and target VOCs.

The normalized electrical conductivity as a function of gas concentration has been calculated and plotted in Fig. 3 in frequency domain. Initially, in the absence of target gas, i.e. VOCs = 0, the normalized electrical conductance is low at any given frequency. The conductance of the film increases after the gas concentration (partial pressure) is increased slightly. This shift in behaviour corresponds to the shift in adsorption–desorption equilibrium due to the interaction of sensor surface with the VOCs molecules. These molecules interact with already adsorbed oxygen ions and results into release of free electrons in to the conduction band of the semiconductor material as per Eq. 3. The shift in conductance is further dependent on the gas concentration as shown in Fig. 3. This graph of normalized electrical conductance with respect to frequency validates the adsorption–desorption theory and supports the proposed model to predict the sensing for VOCs molecules present in human breath. The cut-off frequency (10^8 to 10^9 Hz) signifies the specific adsorption frequency of adsorbed molecules into the consideration at which the maximum sensor response was observed. The cut-off frequency is a specific optimized parameter for a specific gas and needs to be optimized for different gases/VOCs.

Sensor response was computed and plotted for the computed cut-off frequency with respect to partial pressure of VOCs varied from 10^{-3} to 10^{-1} atm as shown in
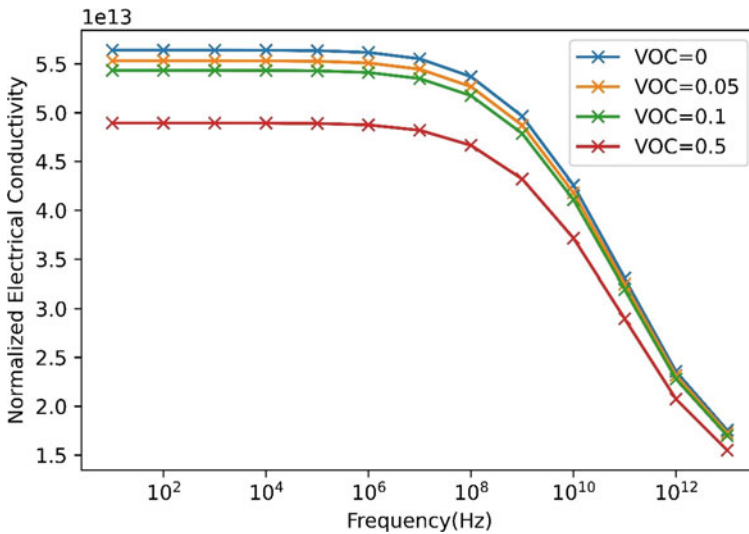
Fig. 3 Normalized electrical conductance as a function of frequency for different concentration
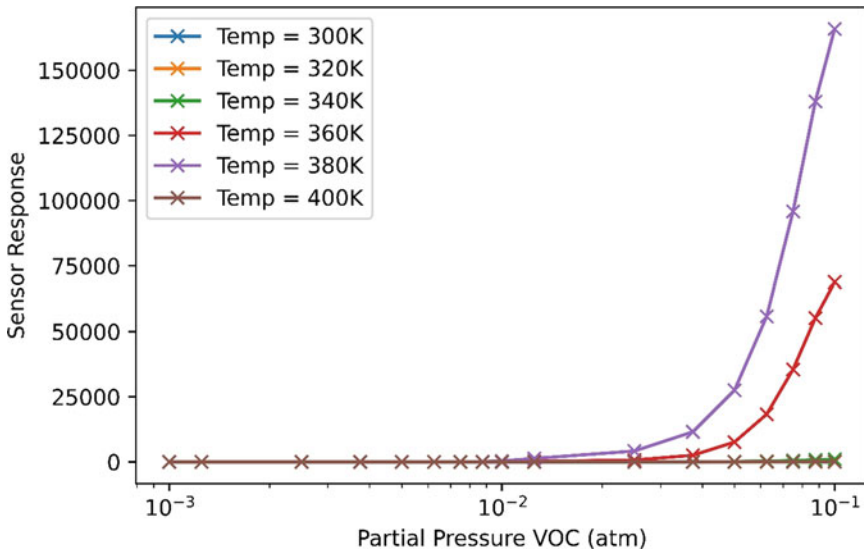
Fig. 4 Sensor response at various temperature as a function of gas concentration

Fig. 4. This computation was carried out for temperature range from 300 to 400 K. The results obtained from the study reveal that the sensor modelled with the parameters shows better response at 360 to 380 K and at partial pressure of 0.05 atm of gas molecules. The sensitivity of the simulated device has been recorded significantly at 360 K for 0.06 atm partial pressure of target VOCs.

Sensor response was also plotted as a function of temperature as shown in Fig. 5 to compute the optimum temperature for the modelled sensor. It has been observed that the sensor device demonstrates the similar behaviour towards various concentration of target VOCs. Moreover, there is a significant increase in the sensor response with the increasing concentration for the same operating temperature. Further, the sensor shows maximum response in the range of 375 to 385 K temperature, and it start reducing with the further increase in temperature. Hence, the operating temperature of the modelled sensor is 375 to 385 K for best sensor response towards exposed VOCs.

4 Conclusions

The presented model in this paper uses Wolkenstein chemisorption theory to compute the sensor response of Cu_2O-based sensor to the VOCs contained in human breath. The normalized electrical conductance has been plotted as a function of frequency which validates the proposed model. The presented model validates that the Cu_2O is a sensitive material towards VOCs, and it can further be optimized to improve
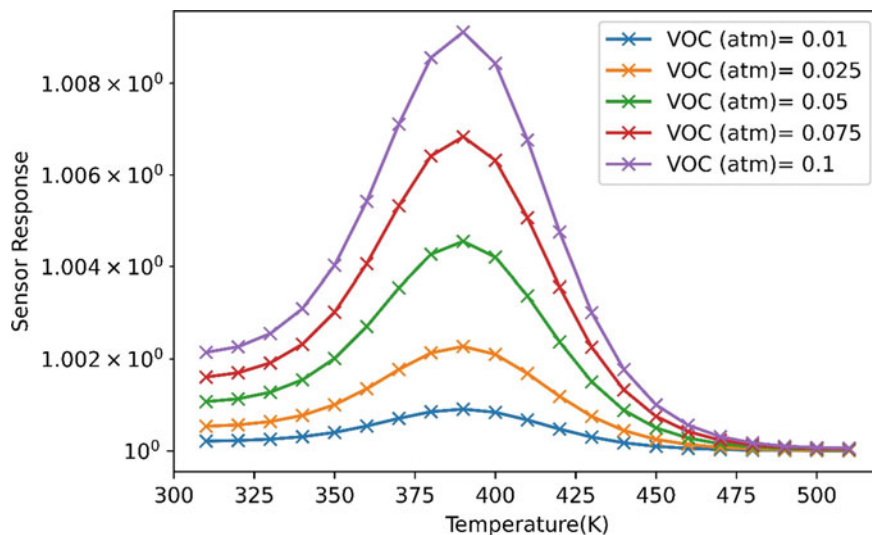
Fig. 5 Sensor response of Cu$_2$O for different concentrations of VOCs as a function of temperature

the same. The simulations also provided the cut-off frequency (10^8–10^9 Hz) after which the response start decreasing significantly. Further, the gas sensor response was calculated and plotted as a function of partial pressure of VOCs at varying temperature range. The results obtained from the study reveal that the sensor modelled with the parameters shows improved response in the temperature range of 360–380 K towards as low as 0.05 atm pressure of target VOCs. Further, this model can be modified to predict the sensor response of other gases present in the human breath by considering their parameter or using array of such sensors. Also, the effect of crystalline size and shape of sensing material can be studied to improve the material's sensitivity towards VOCs.

References

1. Bârsan N, Hübner M, Weimar U (2011) Conduction mechanisms in SnO$_2$ based polycrystalline thick film gas sensors exposed to CO and H$_2$ in different oxygen backgrounds. Sens Actuators B Chem 157:510–517. https://doi.org/10.1016/j.snb.2011.05.011
2. Korotcenkov G (2007) Metal oxides for solid-state gas sensors: what determines our choice? Mater Sci Eng B 139:1–23. https://doi.org/10.1016/j.mseb.2007.01.044
3. Bârsan N (2011) Transduction in semiconducting metal oxide based gas sensors—implications of the conduction mechanism. Proc Eng 25:100–103. https://doi.org/10.1016/j.proeng.2011.12.025
4. Barsan N, Koziej D, Weimar U (2007) Metal oxide-based gas sensor research: how to? Sens Actuators B Chem 121:18–35. https://doi.org/10.1016/j.snb.2006.09.047

5. Barsan N, Simion C, Heine T, Pokhrel S, Weimar U (2010) Modeling of sensing and trans-duction for p-type semiconducting metal oxide based gas sensors. J Electroceram 25:11–19. https://doi.org/10.1007/s10832-009-9583-x

6. Bejaoui A, Guerin J, Zapien JA, Aguir K (2014) Theoretical and experimental study of the response of CuO gas sensor under ozone. Sens Actuators B Chem 190:8–15. https://doi.org/10.1016/j.snb.2013.06.084

7. Brinzari V, Korotcenkov G, Schwank J, Boris Y (2002) Chemisorptional approach to kinetic analysis Of SnO_2: Pd-based thin film gas sensors. J Optoelectr Adv Mater 4(1):147–150

8. Shankar P, Rayappan JBB (2015) Gas sensing mechanism of metal oxides: The role of ambient atmosphere, type of semiconductor and gases-A review. Sci Lett J 4(4):126

9. Wolkenstein Th (1960) The electron theory of catalysis on semiconductors. In: Advances in catalysis. Elsevier, pp 189–264. https://doi.org/10.1016/S0360-0564(08)60603-3

10. Göpel S (1995) SnO_2 sensors current status and future prospects-annotated.pdf (n.d.)

11. Kumar A, Kumar M, Kumar R, Singh R, Prasad B, Kumar D (2019) Numerical model for the chemical adsorption of oxygen and reducing gas molecules in presence of humidity on the surface of semiconductor metal oxide for gas sensors applications. Mater Sci Semicond Process 90:236–244. https://doi.org/10.1016/j.mssp.2018.10.020

12. Bhowmick T, Ghosh A, Nag S, Majumder SB (2022) Sensitive and selective CO_2 gas sensor based on CuO/ZnO bilayer thin-film architecture. J Alloys Comp 903:163871. https://doi.org/10.1016/j.jallcom.2022.163871

13. Kumar A, Kumar R, Singh R, Prasad B, Kumar D, Kumar M (2021) Effect of surface state density on oxygen chemisorption, grain potential and carrier concentration for different grain sizes of nanocrystallite metal oxide semiconductors: a numerical modelling approach. Arab J Sci Eng 46(1):617–630. https://doi.org/10.1007/s13369-020-04630-3

Significance of Convolutional Neural Network in Fake Content Detection: A Systematic Survey

Pummy Dhiman and Amandeep Kaur

Abstract The digital era makes it possible to alter and manipulate digital content in ways that were not imaginable twenty years ago. In the future, digital media will be modified in ways that are currently unimaginable. As fake content is often produced to deceive users, it is quite thought-provoking to distinguish it. The classification of internet content—whether it is bogus or not—has significantly improved with the deep learning approach, convolutional neural network (CNN). This is the motivation for this study's analysis of CNN's significance in the identification of fake content.

Keywords Fake content detection · FCD · Convolutional neural network · CNN · Fake news · Affordable access · Technology

1 Introduction

As every coin has two faces, similarly the advancement of technology has proven both miracle and curse to humans. Due to mobile computing, now everyone has mobile phone. As the advancement in technology and low cost, people have smartphones with affordable internet access. This digital era has made content dissemination on social media. Most of people get information digitally. As the social platform is easy to use and free accessible, people use it frequently. And these social platforms are flooded with a lot of information (content). This scenario has its benefits as well as drawbacks, because with so much content available on internet, often confuses people about the credibility of content. The impact of fake content (news) has been shown during 2016 US President Election. Fake information has become a worldwide problem, and the researchers have been working to countermeasure this issue. There has always been fake information, even before the internet was invented. The

P. Dhiman (✉) · A. Kaur
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
e-mail: pummy.dhiman@chitkara.edu.in

A. Kaur
e-mail: amandeep@chitkara.edu.in

Fig. 1 Three elements of
fake information
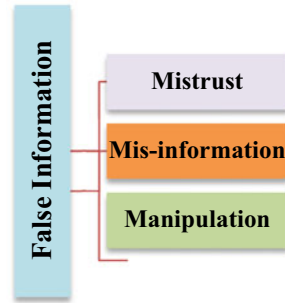
term "false information" covers a broad range of misinformation or disinformation including environment, health, etc. In this context, false information means news stories, hoaxes, or other materials meant to deceive or mislead readers. As fake content is often produced to deceive users, it is quite thought-provoking to distinguish it from real, and with the advancement in digital technology, the modification of digital media will be unimaginable in the future [1, 2]. According to Martina Chapman (Media Literacy Expert), there are three elements to fake information illustrated below in Fig. 1.

Since ancient times, fake information has been spread—in the Mahabharata, a piece of fake news stating "Ashwatthama is dead" led to Dhristadyumna beheading Dronacharya [3]. In the current era, fake news, misinformation, or disinformation is not a new concept, but it has become a frequent occurrence. We all have witnessed the effects of this fake news on mental health during COVID-19 times. There is so much fake news all over the internet regarding disease spread, home remedies, etc. WHO termed this the "infodemic". To combat the COVID-19 infodemic, WHO has collaborated with the UK and the study revealed that within the first 3 months of 2020 because of coronavirus misinformation, nearly 6000 people worldwide were hospitalized. Also during this time period, researchers also believed that it also caused the deaths of at least 800 people. Due to the low cost, ease of use, and quick information transmission, the internet has transformed how people interact and communicate [2]. Throughout the Russia-Ukraine war, the majority of social media feeds have been flooded with fake news because, as a part of propaganda, warring nations commonly use fake information to undermine the opponent's determination while boosting the morale of their own country [4]. The availability of online multimedia content generators makes it simple to produce content that offers benefits in communication. These advances also have a significant negative influence on the society in the context of the dissemination of false information [5]. Sometimes people not knowingly spread fake information, because if the content appeals some novelty or emotionally, they tend to forward it immediately without checking whether it is genuine or fake content. Echo chambers also played a significant role in dissemination of fake content. Human cannot detect fake information, as they can be biased. According to a survey of the American public, only 26% are confident in their ability to distinguish between true and fake news [5]. Various social platforms have taken steps to stop this fake content

spread, but still it is a challenging task. With rapid advancement in artificial intelligence, a lot of work has been done in this context using its various approaches. NLP, ML, and DL played a significant role in FCD. Various DL algorithms MLP, RNN, CNN, and bidirectional LSTM have been proved their importance in combating fake content. There has been outstanding performance by CNNs in text, image segmentation, classification, detection, and retrieval tasks, making them one of the best learning algorithms for understanding text and image contents. An evolution of deep CNN architectures from ConvNet is illustrated in Fig. 2.

This study seeks to provide the various CNN frameworks adopted by various researchers with dataset used by them and results achieved.
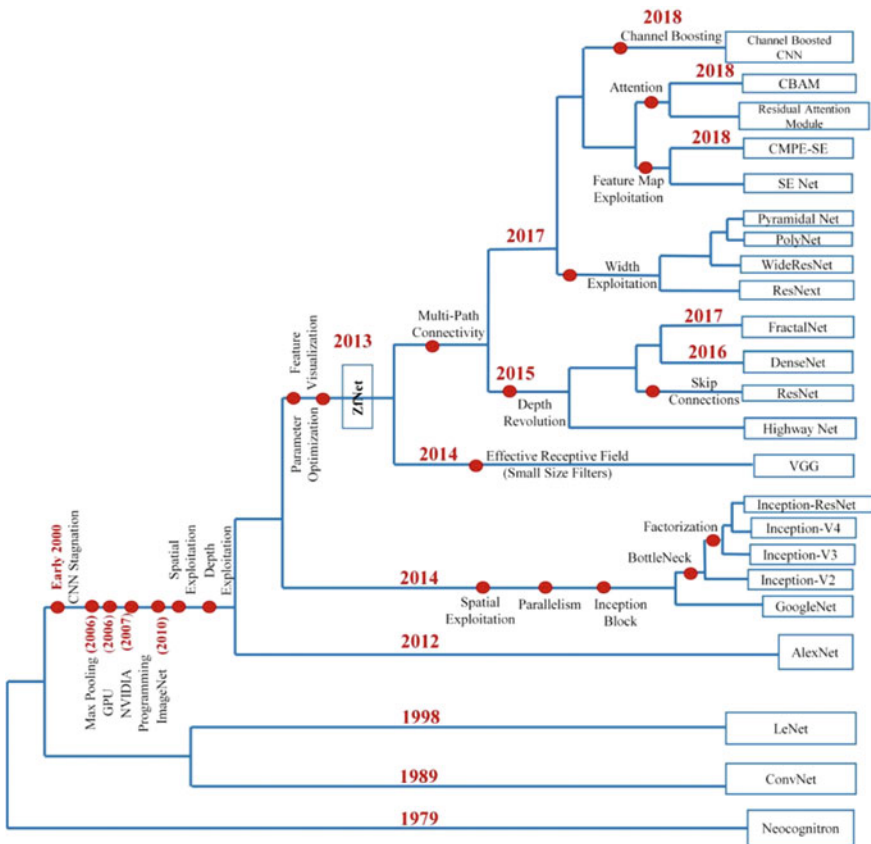
Fig. 2 CNN evolution [6]

2 Motivation

Whether it is journalism, economy, or democracy, fake content has an impact on all spheres. Various riots (e.g., the Delhi riots of 2019 against the Citizenship (Amendment) Act 2019) have occurred due to fake news. India's Election Commission of India (ECI) reported 154 instances of fake news or misinformation on social media platforms during the 2019 Lok Sabha Elections. During the 2020–2021 Indian farmers' protest, social media was flooded with fake information [7]. Violence has erupted across the country on the day of Ram Navami (April 10, 2022). After this, the hashtag #IndianMuslimsUnderAttack began trending on social media. A bogus story was seeded on social media on April 12, 2022. According to the story, Muslims in India are in danger. In today's tech era, digital media can now be edited and manipulated in ways that were unthinkable twenty years ago [1]. One of the emerging areas in artificial intelligence research that has drawn interest from researchers all over the world is the detection of fake news. Due to a lack of context-specific news data, fake news detection accuracy has not considerably improved despite garnering significant attention in the academic community. Deep learning has an advantage over the traditional feature-based (ML) model; in that it can identify the appropriate feature set on its own for a given problem for classification [8]. CNN was first introduced by Fukushima in 1998 [9], the special type of neural networks, which can automatically learn representations from the data due to the use of multiple feature extraction stages. It has demonstrated excellent performance in speech recognition, object detection, video processing, natural language processing, and image classification and segmentation [6]. This work seeks to provide the systematic review of literature on FCD using CNN with their dataset used, their resources requirement, and accuracy achieved.

3 Methodology

For this study, some guidelines have been followed, and this survey is divided into five steps.

1. Defining research questions
2. Formulation of search string
3. Data inclusion/exclusion criteria
4. Quality assessment rules
5. Data extraction.

Research Questions

RQ1. Which modalities have been used for fake content detection?

RQ2. What are the different CNN models that have been used for fake content detection?

RQ3. What are the various feature extraction methods used for FCD?

RQ4. What are the various types of datasets used for FCD?

RQ5. What is the software and hardware requirement for FCD?

RQ6. Which CNN model gives the highest accuracy?

The search string

The Boolean OR and AND rules are used to generate search string for finding paper from digital repositories Science Direct, ACM Digital Library, IEEE Xplore, and Google Scholar.

("Fake" OR "False") AND ("news" OR "information") AND ("detection") AND ("CNN" OR "Convolutional Neural Network")

Data inclusion/exclusion criteria

To answer the framed research questions, inclusion and exclusion criteria have been followed to select relevant papers. This study aims to survey fake information detection using CNN, so to fulfill this purpose, the articles related to both FCD and CNN are included. Fake content detection studies that do not incorporate the use of CNN are excluded.

Quality Assessment Rules

After inclusion and exclusion criteria, several research articles have been selected. Quality assessment rules are used to identify the papers to consider for this survey. Here, if we get answers to at least half of the research questions, then we consider that study.step is followed to ensure the high quality of selected studies and to get the answer of at least some of the research questions as shown in Table 1.

Data Extraction

To analyze the finally obtained research papers, data extraction is applied in order to get the answers of the research questions. First, general details like publication year, journal name, publisher, impact factor, citations, etc. are extracted from the finally selected papers; after that, more specific information such as framed in research questions as types of fake information, CNN models, feature extraction methods, dataset details, and the result obtained is fetched.

4 Result

Based on the information retrieved from the selected studies in Table 1, this section represents the answers to the framed research questions.

RQ1. Which modalities have been used for fake content detection?

Table 1 Framed research question answers

References	Year	RQ1	RQ2	RQ3	RQ4	RQ5
[11]	2022	✓	✓	✓	✓	×
[5]	2021	✓	✓	✓	✓	✓
[12]	2018	✓	✓	✓	✓	×
[13]	2019	✓	✓	✓	✓	×
[14]	2021	✓	✓	✓	✓	✓
[15]	2020	✓	✓	✓	✓	✓
[16]	2021	✓	✓	✓	✓	×
[8]	2020	✓	✓	✓	✓	×
[17]	2021	✓	✓	✓	✓	✓
[18]	2019	✓	✓	✓	✓	×
[19]	2018	✓	✓	✓	✓	✓
[20]	2021	✓	✓	✓	✓	×
[21]	2020	✓	✓	✓	✓	×
[22]	2019	✓	✓	✓	✓	×
[23]	2021	✓	✓	✓	✓	✓
[24]	2020	✓	✓	✓	✓	×
[25]	2022	✓	✓	✓	✓	×
[26]	2021	✓	✓	✓	✓	×
[27]	2020	✓	✓	✓	✓	×

Based on the findings of the selected studies, it has been seen that most of the study followed unimodal approach, by unimodal means using only textual data [10]. But fake content is not limited to text only, it includes visual content also. So the counter-measures this global problem of fake information dissemination, researchers is not limit themselves to textual content. Researchers have also worked in this area as depicted in Fig. 3. They have worked on images only, videos only as well as followed multimodal approach having the combination of unimodal and performed binary fake content classification (fake or not). Figure 4 is an illustration of the languages used in selected FCD studies. The English language received 90% of the attention.

Fig. 3 Modalities in FCD
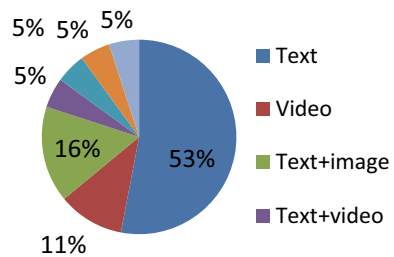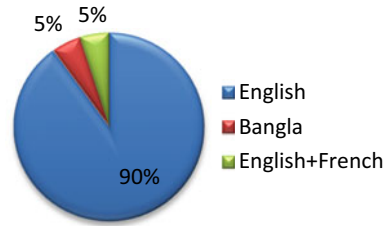
RQ2. What are the different CNN models that have been used for fake content detection?

In literature, CNN has proved to be the most powerful deep learning technique in various aspects of classification, detection task whether in terms of text, image, and video. The popularity of CNN is primarily due to its hierarchical feature extraction ability [6]. Research is still going on to get more advancement. The representational capacity of a CNN usually depends on its depth [6]. Selected literature showed that CNN has been adopted by the researchers to detect fake text content as well as images and videos. Various CNN models have been used as shown in Table 2.

Table 2 Type of content used by various CNN models

References	Year	Modality	Language	Type
[11]	2022	Video	English	News
[5]	2021	Text	English	News
[12]	2018	Text + Image	English	News
[13]	2019	Text (email, spam)	English	Fraud spam
[14]	2021	Text	Bangla	News
[15]	2020	Text	English	Rumor
[16]	2021	Text	English	Financial
[8]	2020	Text	English	Political
[17]	2021	Text	English	Political
[18]	2019	Text, video	English, French	Satire (vaccination)
[19]	2018	Image		Forgery
[20]	2021	Text + Image + Sentiment	English	Rumor
[21]	2020	Text + Image	English	News
[22]	2019	Video	English	News
[23]	2021	Text	English	News
[24]	2020	Text	English	Political
[25]	2022	Text, Image	English	News, account, network
[26]	2021	Text	English	News
[27]	2020	Text	English	News
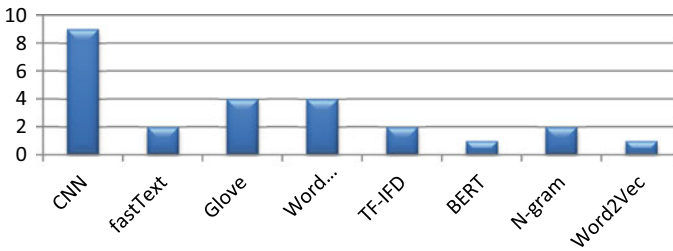
Fig. 5 Feature extraction

Some studies used standalone CNN architecture, while others used 3D CNN and GCNN. VGG, ResNet, and attention-based CNN also played a significant role in FCD. Hybrid approach containing CNN with RNN also proved significance results as CNN is used to extract features automatically and then those features are fed to RNN to get better results. Researchers and developers use ImageNet, a large-scale, organized visual image database, to train their models.

They now hold the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), an annual contest for large-scale object recognition and image categorization, and the competition's top performers are able to establish a standard for object classification [28].

RQ3. What are the various feature extraction methods used for FCD?

The main advantage of DL over ML is its automatic feature detection, as feature extraction is one of the necessary steps before actual working of FCD. CNN does this work itself due to the utilization of many feature extraction stages that can learn representations from data automatically [29]. Along CNN, various feature extraction methods like fastText, Glove, TF-IFD, word embedding, and BERT have been employed in study by researchers as shown in Fig. 5.

RQ4. What are the various types of datasets used for FCD?

Studies with various benchmark datasets have been listed in Table 3, which has been used in fake content detection. From this table, it has been concluded that most of the work has been conducted on text data and focused on the English language. So there is a wide scope of doing work in multimodality, and other languages apart from the resource-rich language of English would be explored. Inadequate linguistic resources such as stemmers and annotated corpora make the research more challenging and inspiring [30].

RQ5. What is the software and hardware (resources) requirement for FCD?

As the information containing human language and we use technology to detect whether the given content is fake or true, this human (natural) language is not understandable by machines. To fill up this gap, natural language processing (NLP) comes into the picture. It is a technology that makes natural language understandable by machines. From the literature study, it is clear that most of the studies used Python

Table 3 Dataset used by various studies

Study	Dataset
[11]	FaceForensics++
[5]	Fake News Dataset, Fake or Real News Dataset, Fake News Detection Dataset, ISOT Fake News Dataset
[12]	fake.csv from Kaggle
[13]	Enron email dataset and the 419 spam fraud corpus
[14]	BanFakeNews
[15]	Yelp Polarity (YELP-2), Five Breaking News (FBN)
[16]	Self-collected data
[8]	Fake news dataset
[17]	LIAR
[18]	Vaccination fake news dataset provided by the Storyzy company
[19]	CELEBAHQ, fake face image database
[20]	FakeNewsNet, DAT@Z20
[21]	PolitiFact, GossipCop
[23]	Dataset from Kaggle, FakeNewsNet, FA-KES5,ISOT
[24]	PolitiFact, Twitter
[26]	George M., Kaggle, GossipCop, F
[27]	Kaggle fake news dataset

and Natural Language Toolkit (NLTK) which has been called "an amazing library to play with natural language" [31]. Google Colab is a cloud-based platform that provides free access to GPU and other essentials for deep learning tasks and various deep learning libraries like TenserFlow, PyTorch, and Keras, which provide built-in classes of NNs, fast numerical computation, and automated gradient estimation for both CPU and GPU [6]. Deep learning accelerators from NVIDIA, which are both energy and computation efficient, have also been used to detect fake news.

RQ6. Which CNN model gives the highest accuracy?

According to the proposed work, only the studies which are using CNN for FCD are considered, and it has been recorded that selected studies showed good results in binary classification of fake content. Figure 6 depicts the performance graph of each study using either standalone CNN approach or hybrid approach with CNN.

5 Issues and Challenges

The observed challenges from this study are that most of the work has been done in the resource-rich language of English. Other languages should be explored as this fake content dissemination is not limited to one language. The news can sometimes
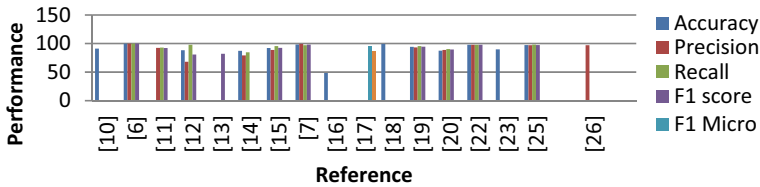
Fig. 6 Performance analysis

be partially true, textually true but visually faked, or both textually and visually faked. As a result, multiclass classification must be investigated. Most FCD studies have relied solely on textual data. However, multimodality, including videos, audio, and images, should be prioritized. A large and balanced dataset will play a great role in achieving high accuracy.

6 Conclusion

This study defines the systematic review of fake content detection through deep learning. By conducting this survey, some observations have come across that will help the upcoming researchers to help in their research. As the CNN is the most popular technique in deep learning, so for this piece of work, the studies related to FCD and CNN have been selected. It is observed that most of the research has been emphasized on text content detection only. Adding image or video along text attracts more attention as compared to text only news; therefore, a lot of work would be done in multimodality. Apart from this, other languages should be explored. Hybrid techniques have proved their potential in FCD task.

References

1. Farid H (2006) Digital doctoring: how to tell the real from the fake. Significance 3(4):162–166. https://doi.org/10.1111/j.1740-9713.2006.00197.x
2. Mittal N, Rajpurohit U, Sarangi PK, Goel N (2022) Implementation and analysis of fake news using machine learning technique. In: 2022 IEEE International Conference on Current Development in Engineering and Technology (CCET), Bhopal, India, pp 1–6. https://doi.org/10.1109/CCET56606.2022.10080126
3. Ansar W, Goswami S (2021) Combating the menace: a survey on characterization and detection of fake news from a data science perspective. Int J Inf Manag Data Insights 1(2):100052. https://doi.org/10.1016/j.jjimei.2021.100052
4. Vikram M (2022) Misinformation is playing a significant role in Russia-Ukraine War. Forbes. https://www.forbes.com/sites/vikrammittal/2022/03/07/misinformation-will-play-a-significant-role-in-russia-ukraine-war/. Accessed 25 May 2022
5. Sastrawan IK, Bayupati IPA, Arsa DMS (2021) Detection of fake news using deep learning CNN–RNN based methods. ICT Exp 8:396–408. https://doi.org/10.1016/j.icte.2021.10.003

6. Saini A, Guleria K, Sharma S (2023) An automatic fake news identification system using machine learning techniques. In: 2023 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT), Karaikal, India, pp 1–5. https://doi.org/10.1109/IConSCEPT57958.2023.10170307

7. Ahmed AAA et al (2021) Detecting fake news using machine learning: a systematic literature review. Psychol Educ J 58(1):1932–1939. https://doi.org/10.17762/pae.v58i1.1046

8. Kaliyar RK, Goswami A, Narang P, Sinha S (2020) FNDNet—a deep convolutional neural network for fake news detection. Cogn Syst Res 61:32–44. https://doi.org/10.1016/j.cogsys.2019.12.005

9. Dhillon A, Verma GK (2020) Convolutional neural network: a review of models, methodologies and applications to object detection. Prog Artif Intell 9(2):85–112. https://doi.org/10.1007/s13748-019-00203-0

10. Dabas K (2018) Multimodal fake news detection on online social media. B. Tech, Indraprastha Institute of Information Technology New Delhi

11. Roy R, Joshi I, Das A, Dantcheva A (2022) 3D CNN architectures and attention mechanisms for deepfake detection. Adv Comput Vis Pattern Recogn 213–234. https://doi.org/10.1007/978-3-030-87664-7_10

12. Yang Y, Zheng L, Zhang J, Cui Q, Li Z, Yu PS (2018) TI-CNN: convolutional neural networks for fake news detection. http://arxiv.org/abs/1806.00749

13. Dhamani N et al (2019) Using deep networks and transfer learning to address disinformation. http://arxiv.org/abs/1905.10412

14. Adib QAR, Mehedi MHK, Sakib MS, Patwary KK, Hossain MS, Rasel AA (2021) A deep hybrid learning approach to detect Bangla fake news. In: ISMSIT 2021— proceedings of 5th international symposium on multidisciplinary studies and innovative technologies, pp 442–447. https://doi.org/10.1109/ISMSIT52890.2021.9604712

15. Guo M, Xu Z, Liu L, Guo M, Zhang Y, Kotsiantis SB (2020) An adaptive deep transfer learning model for rumor detection without sufficient identified rumors. Math Probl Eng 2020:1–12. https://doi.org/10.1155/2020/7562567

16. Zhi X et al (2021) Financial fake news detection with multi fact CNN-LSTM model. In: 2021 IEEE 4th international conference on electronics technology. ICET, pp 1338–1341. https://doi.org/10.1109/ICET51757.2021.9450924

17. Xing J, Wang S, Zhang X, Ding Y (2021) HMBI: a new hybrid deep model based on behavior information for fake news detection. Wirel Commun Mob Comput 2021:1–7. https://doi.org/10.1155/2021/9076211

18. Guibon G et al (2019) Multilingual fake news detection with satire to cite this version. HAL Id: Halshs-02391141

19. Mo H, Chen B, Luo W (2018) Fake faces identification via convolutional neural network. IH MMSec 2018— Proceedings of the 6th ACM workshop on information hiding and multimedia security, pp 43–47. https://doi.org/10.1145/3206004.3206009

20. Azri A, Favre C, Harbi N, Darmont J, Noûs C (2021) Calling to CNN-LSTM for rumor detection: a deep multi-channel model for message veracity classification in microblogs. Lecturer notes on computer science (including subseries lecturer notes on artificial intelligence and lecturer notes on bioinformatics), vol 12979 LNAI, pp 497–513. https://doi.org/10.1007/978-3-030-86517-7_31

21. Wu XZBJ, Zafarani R (2020) SAFE : Similarity-aware multi-modal fake. Springer International Publishing

22. Yao G, Lei T, Zhong J (2019) A review of convolutional-neural-network-based action recognition. Pattern Recogn Lett 118:14–22. https://doi.org/10.1016/j.patrec.2018.05.018

23. Saleh H, Alharbi A, Alsamhi SH (2021) OPCNN-FAKE: optimized convolutional neural network for fake news detection. IEEE Access 9:129471–129489. https://doi.org/10.1109/ACCESS.2021.3112806

24. Sansonetti G, Gasparetti F, D'Aniello G, Micarelli A (2020) Unreliable users detection in social media: deep learning techniques for automatic detection. IEEE Access 8:213154–213167. https://doi.org/10.1109/ACCESS.2020.3040604

25. Varlamis I, Michail D, Glykou F, Tsantilas P (2022) A survey on the use of graph convolutional networks for combating fake news. Futur. Internet 14(3):1–19. https://doi.org/10.3390/fi14030070

26. Choudhary M, Chouhan SS, Pilli ES, Vipparthi SK (2021) BerConvoNet: a deep learning framework for fake news classification. Appl Soft Comput 110:107614. https://doi.org/10.1016/j.asoc.2021.107614

27. Agarwal A, Mittal M, Pathak A, Goyal LM (2020) Fake news detection using a blend of neural networks: an application of deep learning. SN Comput. Sci. 1(3):143. https://doi.org/10.1007/s42979-020-00165-4

28. Chatterjee HS (2022) Various types of convolutional neural network. In: Towards data science. https://towardsdatascience.com/various-types-of-convolutional-neural-network-8b00c9a08a1b. Accessed 21 June 2022

29. Different types of CNN models (2022) https://iq.opengenus.org/different-types-of-cnn-models/. Accessed 21 June 2022

30. Akhter MP, Zheng J, Afzal F, Lin H, Riaz S, Mehmood A (2021) Supervised ensemble learning methods towards automatically filtering Urdu fake news within social media. PeerJ Comput Sci 7:1–24. https://doi.org/10.7717/peerj-cs.425

31. NLTK: Natural Language Toolkit (2022) https://www.nltk.org/. Accessed 17 June 2022

Wavelet and Savitzky–Golay Filter-Based Denoising of Electrocardiogram Signal: An Improved Approach

Nisha Raheja and Amit Kumar Manoacha

Abstract Today, heart disease has reduced the chance of a human living. One of the main causes of morality is cardiovascular disease (CVD). The determination of the electrical activity of the signal is called an electrocardiogram. Different noises are present in the signal as it is being recorded. So, noise needs to be eliminated before ECG signal analysis. Various noises, including baseline wander, power line interference, and electromyogram, can be found in the ECG signal. In this article, develop a one-stage median filter to eliminate baseline wanders, Savitzky–Golay filtering (SG) and wavelet transform is employed for the elimination of artifacts from the electrocardiogram waveform is presented and the SNR (signal to noise ratio) measurements results have shown been calculated using MIT-BIH database for various records and contrasted these findings with earlier works. These suggested results outperform those found with the state of artworks.

Keywords ECG · Base line wander · Power line interference · Electromyogram · MATLAB

1 Introduction

Electrical activity is represented via an ECG signal. The main reason that ECGs are utilized is because they are inexpensive and provide an abundance of clinical information. Additionally, it serves as a digital recorder. It experiences a variety of noises when recording, including baseline wander, power line interference, and electromyogram. The two steps in the analysis of the electrocardiogram signal are preprocessing and feature extraction. The electrocardiogram signal's noise is removed during the preprocessing stage, and ECG-related data is extracted during feature extraction [1]. Different waveforms, including the P, QRS, and T waveforms, make up the ECG

N. Raheja · A. K. Manoacha (✉)
Department of Electrical Engineering, Maharaja Ranjit Singh Punjab Technical University, Bathinda, India
e-mail: manocha82@gmail.com

signal. Figure 1 depicts a typical ECG waveform [2]. For the reduction of baseline
wander noise, an approach based on an adaptive filter is proposed in [3]. The [4]
proposes an analog electrical device for a portable ECG system that depends on a
telecardiology platform. The purpose of removing artifact and an adaptive filter tech-
nique for enhancing the electrocardiogram signal for real-time monitoring systems
is reported in [5]. The [6] suggests a bias-compensated least mean square approach
for the input signal. An LMS method is provided in [7] for the elimination of power
line interference (PLI). The elimination of irregularities from the electrocardiogram
signal is suggested in [8] using an S-median thresholding algorithm. The elimination
of additive white Gaussian noise in ECG data is suggested in [9] using a non-local
wavelet approach. The [10] introduces a brand-new biotelemetry-based technique
for continually monitoring heart activity. A telecardiology-based method for the
real-time detection of cardiovascular diseases in rural areas is suggested in [11].
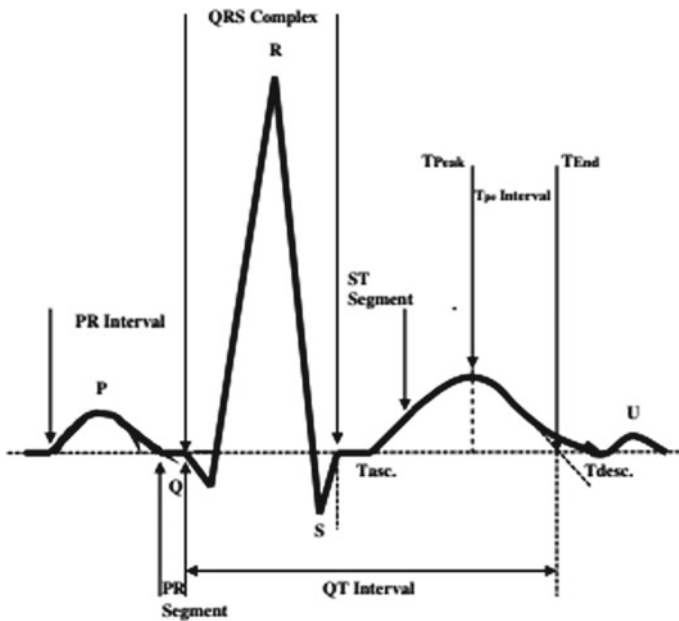
Fig. 1 Components of ECG waveform

2 Artifacts in ECG Signal

2.1 Baseline Wanders

The movement of the patient's body, breathing, and improper electrode placement on the human chest are the main causes of this noise. The frequency band of this artifact is 0.5–0.6 Hz. The baseline wander noise is challenging to identify since it has a very low-frequency component [12].

2.2 Power Line Interference

PLI appears in electrocardiogram when machine is configured incorrectly, making it more challenging to evaluate the ECG signal. PLI takes place between 50 and 60 Hz. Therefore, it is crucial to eliminate this noise. If not, an ECG waveform is superimposed over the P and T waves [12].

2.3 Electromyogram Noise

The movement of muscles is what is making this noise. The ECG signal becomes problematic while being recorded as a result. The sub-harmonic waveform is totally suppressed. The frequency of this form of noise is 10 kHz. This makes getting rid of his kind of noise very challenging. The PQRST complex is overlapping if it is not removed completely [12].

3 Materials and Methods

3.1 ECG Database

There are numerous types of databases available in PhysioBank that contain recordings of various signals that are used by the biomedical research groups. This article makes use of the MIT-BIH dataset. This kind of database contains 48 signals, each of which was sampled at a frequency of 360 Hz and has a period of 30 min. According to [13], this database contains 110,007 beats.

3.2 Method

Biomedical data can be examined separately in the time–frequency domain and the frequency and time domains. There are two transforms available for data analysis [14]:

(I) **Fourier transform**: This potent instrument analyzes stationary signals in the frequency range. It is not appropriate for non-stationary data analysis.

(II) **Wavelet transform**: This technique analyzes signals in the temporal and attributes. It is a technology used to reduce noise and analyzes a non-stationary signal's constituent parts. The definition of the wavelet function is:

$$Wf(u,s) = \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{s}}\Psi^*\left(\frac{t-u}{s}\right)dt = f * \Psi_s(u) \tag{1}$$

where s is a scaling factor and Ψ is pre-dilation of basic wavelet (x) with scaling factor s.

3.3 Discrete Wavelet Transform (DWT)

The signal is split into two signals in this section using discrete wavelet transform (DWT). The sequence of approximation coefficient comes first, followed by the detailed coefficient. The information needed to reconstruct the signal is contained in the approximation and detail coefficient at each stage. The decomposition tree is the name of this process. As seen in Fig. 2, this is also referred to as Mallet tree decomposition (Fig. 3).
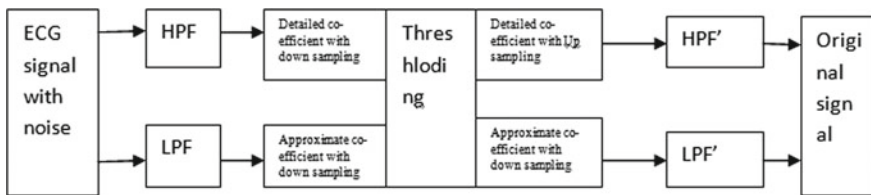
Fig.2 Process of wavelet decomposition and reconstruction

Fig. 3 Workflow diagram
for eliminating ECG artifacts

```
                            ┌─────────────┐
                            │    Start    │
                            └──────┬──────┘
                                   │
                                   ▼
        ┌───────────────────────────────────────────────────┐
        │ Load the ECG signal from MIT-BIH database          │
        └───────────────────────┬───────────────────────────┘
                                   │
                                   ▼
        ┌───────────────────────────────────────────────────┐
        │ Develop a one-stage median filter to eliminate     │
        │ baseline wanders                                   │
        └───────────────────────┬───────────────────────────┘
                                   │
                                   ▼
        ┌───────────────────────────────────────────────────┐
        │ Use a Savitzky-Golay (SG) filter to smooth the     │
        │ median filter's outcome                            │
        └───────────────────────┬───────────────────────────┘
                                   │
                                   ▼
        ┌───────────────────────────────────────────────────┐
        │ Decompose the ECG data and eliminate the artifacts │
        │ presented using the discrete wavelet toolbox [14]  │
        └───────────────────────┬───────────────────────────┘
                                   │
                                   ▼
        ┌───────────────────────────────────────────────────┐
        │ Use the MIT-BIH database to calculate signal to noise│
        │ ratio and then compare the findings with ref [16]. │
        └───────────────────────┬───────────────────────────┘
                                   │
                                   ▼
                            ┌─────────────┐
                            │     End     │
                            └─────────────┘
```

4 Results and Discussion

4.1 Elimination of Baseline Wanders Artifact

The movement of the patient's body, breathing, and improper electrode placement on the human chest are the main causes of this noise. This noise has a frequency spectrum of 0.5–0.6 Hz. The baseline wander noise is difficult to identify since it has an extremely lower frequency. Figure 4 shows that baseline signal, which moves the DC level up and down, is a low-frequency signal. The DC component can be removed by taking the median of the ECG signal and subtracted it from the given input, and

the result was then run through a Savitzky–Golay filter to flatten the waveform. Thus, the name baseline-free ECG signal for this signal. Figure 4 shows the elimination of baseline wanders of MIT-BIH database for records 109, 111 [16].

4.2   Elimination of High-Frequency Noise

This sort of noise happens when an ECG machine is configured incorrectly, making it more challenging to evaluate the ECG signal. Therefore, it is crucial to eliminate this noise. If not, an ECG waveform is overlaid over the P and T waves. Figure 4 shows that the baseline signal, which alters the DC level up and down, is a low-frequency signal. The median of the electrocardiogram signal was extracted and subtracted from the input signal to remove DC component, and the result was then run through a Savitzky–Golay filter to flatten the waveform. Hence, this signal is known as the baseline-free ECG waveform. A wavelet transform is then utilized for partition of the electrocardiogram signal up to the tenth level (Cd1, Cd2, … Cd10) and (Ca1, Ca2, … Ca10). There is a high-frequency component in Cd1, Cd2. Thus, the high-frequency noise can be eliminated by employing wavelet to filter out the coefficient Cd1, Cd2 having frequencies similar to high frequency. Therefore, as illustrated in Fig. 5 [16], this proposed method removes this high-frequency noise using records 109 and 111 from the MIT-BIH dataset.

5   Comparison with State of the Artworks

In the literature, the numerous methods for removing baseline wander, powerline interference, and electromyogram have been described. In [15], the signal-to-noise ratio (SNR) is computed after the noise is removed using the wavelet transform method. The correlation of signal-to-noise ratio (SNR) for the records in the MIT-BIH database, specifically records 109, 111, 112, 114, 115, and 116, with reference [15] yields better findings, as shown in Table 1. After evaluation, the suggested technique uses the db4 wavelet function and has a high SNR value.

6   Conclusion

In order to analyze the electrocardiogram signal and eliminate the various distortions, the median filter, SG filter, and wavelet transform are utilized in this paper. The method was implemented in MATLAB using threshold method, and the wavelet transform provides data in the time and frequency domains. Using the MIT-BIH database, the deliberate outcomes, or signal-to-noise ratio (SNR), were determined.
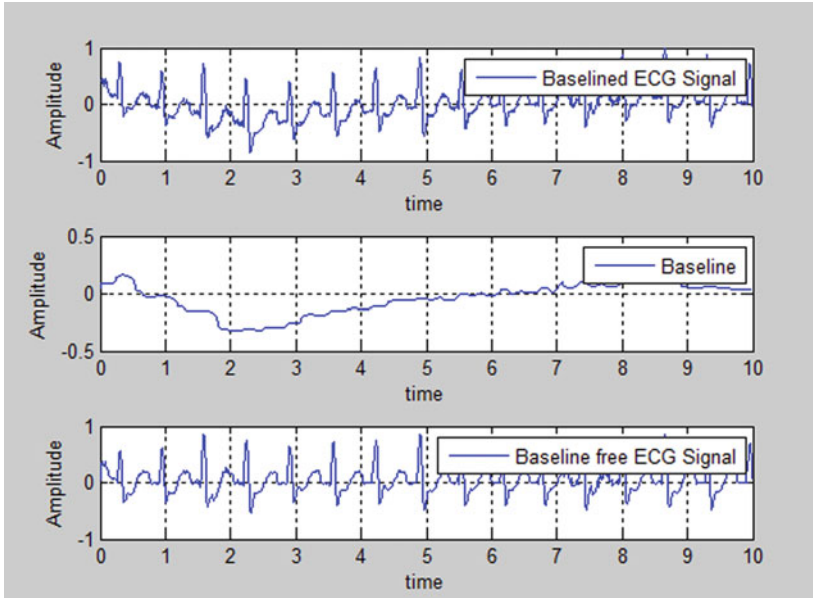
(a) Elimination of baseline wander noise for 109 record of MIT-BIH database
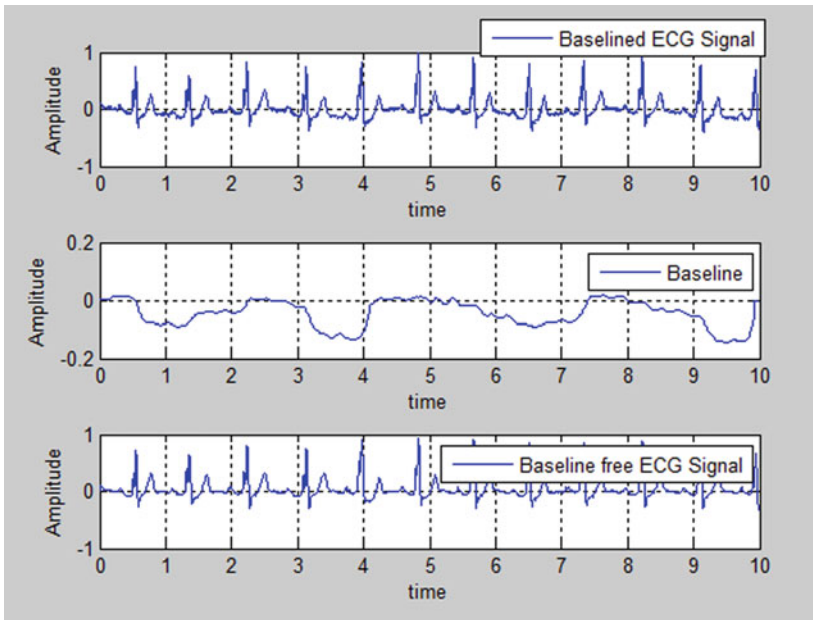
(b) Elimination of baseline wander noise for 111 record of MIT-BIH database

Fig. 4 Elimination of baseline wanders noise of MIT-BIH dataset for records **a** 109 and **b** 111
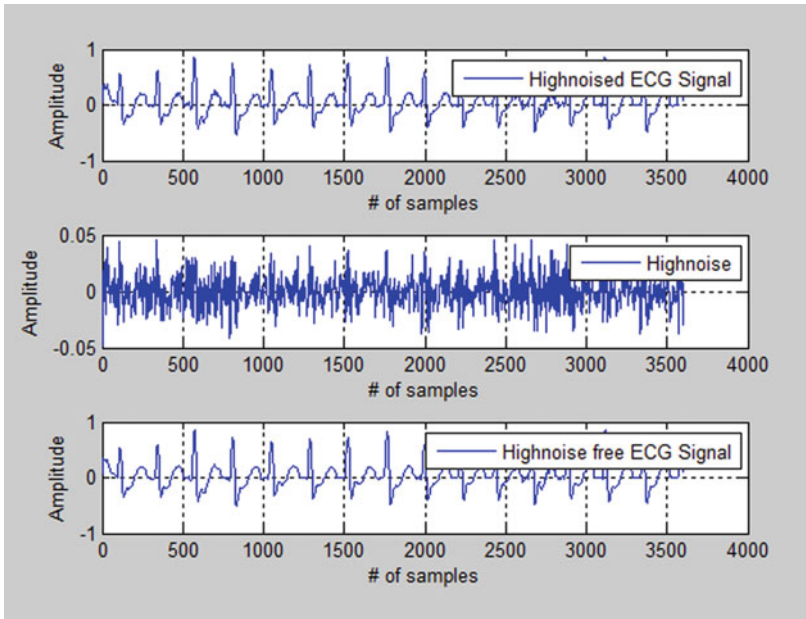
(a) Elimination of high frequency noise for 109 record of MIT-BIH database

(b) Elimination of high frequency noise for 111 record of MIT-BIH database

Fig. 5 Outcomes of removing high noise from the MIT-BIH dataset **a** 109 and **b** 111

Table 1 Analysis of signal-to-noise ratio (SNR)

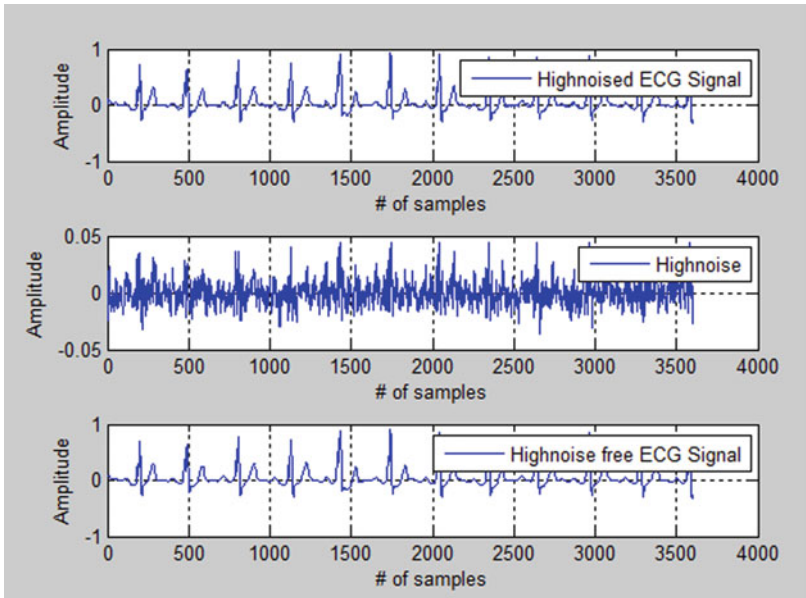| Sr. no.	Analysis of SNR and comparison with the state of artworks		
	MIT-BIH database	For the MIT-BIH database, wavelet transform with IIR filter [16]	Wavelet transform and SG filter for MIT-BIH database (proposed)
1	109	− 1.8472	25.5315
2	111	− 2.7015	24.1984
3	112	− 0.1926	23.6709
4	114	− 0.6912	19.1777
5	115	− 10.997	28.5704
6	116	− 5.7531	28.9598

Additionally, the outcomes of the proposed work contrasted with those of the existing state of artworks and were found to be superior.

References

1. Priya MS (2015) MATLAB based ECG signal noise removal and its analysis. Int Conf Recent Adv Eng Comput Sci IEEE
2. Pater C (2005) Methodological considerations in the design of trials for safety assessment of new drugs and chemical entities. Trials 6(1):1–13
3. Jane R, Laguna P et al (1993) Adaptive baseline wander removal in the ECG: comparative analysis with cubic spline technique. In: Proceedings of computers in cardiology conference, USA, pp 143–146
4. Tsai T-H, Hong J-H, Wang L-H, Lee S-Y (2012) Low-power analog integrated circuits for wireless ECG acquisition systems. IEEE Trans Inf Technol Biomed 16(5):907–912
5. Rahman MZU, Karthik GVS, Fathima SY, Lay-Ekuakille A (2013) An efficient cardiac signal enhancement using time–frequency realization of leaky adaptive noise cancelers for remote health monitoring systems. Measurement 46(10):3815–3835
6. Kang B, Yoo J, Park P (2013) Bias-compensated normalised LMS algorithm with noisy input. Electron Lett 49(8):538–539
7. Taralunga DD, Gussi I, Strungaru R (2015) Fetal ECG enhancement: adaptive power line interference cancellation based on Hilbert Huang transform. Biomed Signal Process Control 19:77–84
8. Awal MA, Mostafa SS, Ahmad M, Rashid MA (2014) An adaptive level dependent wavelet thresholding for ECG denoising. Biocybern Biomed Eng 34(4):238–249
9. Yadav SK, Sinha R, Bora PK (2015) Electrocardiogram signal denoising using non-local wavelet transform domain filtering. IET Signal Process 9(1):88–96
10. Wang L-H, Chen T-Y, Lin K-H, Fang Q, Lee S-Y (2015) Implementation of a wireless ECG acquisition. IEEE J Biomed Health Informan 19(1):247–255
11. Raheja N, Manoacha AK (2020) A study of telecardiology-based methods for detection of cardiovascular diseases. Adv Intell Syst Comput Book Ser 1124
12. Rahul K (2019) Signal processing techniques for removing noise from ECG signals. J Biomed Eng Res 3:101
13. Kumar A, Singh M (2015) Optimal selection of wavelet function and decomposition level for removal of ECG signal artifacts. J Med Imaging Health Inform 5:138–146
14. Kumar A, Sing M (2016) Robust multiresolution wavelet analysis and window search based approach for electrocardiogram features delineation. J Med Imaging Health Inform 6:146–156

15. Arun KS, Maheswari A, Bian G-B (2019) An intelligent learning approach for improving ECG signal classification and arrhythmia analysis. Artif Intell Med 103:101788
16. Amit K, Mandeep S (2016) Statistical analysis of ST segments in ECG signals for detection of ischemic episodes. In: Transactions of the institute of measurement and control, pp 1–12

Heart Disease Detection Using Phonocardiogram (PCG) Signals

Aarti Kashyap and **Babita Majhi**

Abstract Heart disease detection plays an important role in the early diagnosis of cardiovascular disease, especially for small medical services centers. In the past few years, there is progress in heart disease detection by using a conventional methods which are insufficient in recognizing heart disease, and the performance degrades while the cardiac environment changes. Analysis of PCG signal is a typical method for assessing the state of the heart and identifying the potential problems like cardiac disease. This study concentrated on the PASCAL challenge competition dataset to develop various machine learning (ML) models for heart disease detection using PCG signals, which can detect whether the person is healthy or unhealthy. Firstly, the PCG signals are filtered, removed noise and then decomposed to 10 levels using DWT. The classification is done using five ML models, i.e., support vector machine (SVM), k-nearest neighbor (k-NN), decision tree (DT), random forest (RF) and extreme gradient boosting (XGBoost). It is demonstrated from the experiments that RF achieves the best AUC of 98.33% (holdout method, 70:30 ratio), 96.42% (holdout method, 80:20 ratio) and 83.35% (fivefold CV) in all three approaches in comparison to others.

Keywords Heart disease · Phonocardiogram · Signals · Discrete wavelet transform · Support vector machine

1 Introduction

The phonocardiogram (PCG) signal represents the heart sound plot created by the mechanical movement of cardiovascular parts. The heart is one of the principal organs of the human body that conveys blood from one position to other parts of the body. It works same as a pump, and heart beats 72 times per minute [1]. The cardiac cycle contains two phases: diastole and systole. As the heart beats, they take

A. Kashyap (✉) · B. Majhi
Guru Ghasidas Vishwavidyalaya, Central University, Bilaspur, Chhattisgarh, India
e-mail: aarti.kas2009@gmail.com

place as blood is pumped via a network of blood vessels to reach every region of the body. Diastole happens when the heart relaxes after contracting during systole, which happens when the heart contracts to push blood out. It produces the lub and dub sound. The sequence of the heart beats of a normal person is lub dub lub dub (s1s2s1s2….) and so on. This arrangement will change if the person's heartbeat is abnormal, known as a murmur containing some noisy sound. The heart beat sequence is monitored by phonocardiogram (PCG), electrocardiogram (ECG) and plethysmography (PPG) [2] which are used to monitor the electrical and mechanical characteristics of the heartbeat.

Heart disease patients are dyeing due to mental pressure, work over-burden and numerous different reasons [3]. World Health Organization (WHO) reported that the cardiovascular disease is still one of the reasons for death. Approximately 17.9 million individuals overall pass on from CVD-related problems yearly, addressing 31% of deaths worldwide [4].

Conventionally, a medical specialist uses an auscultation process to detect the heart sound. Auscultation is the process of identifying whether heart sound is normal or abnormal. It is painless, financially savvy and requires negligible hardware, making it appropriate for heart assessment, particularly in small essential medical care clinics. Heart sound auscultation depends vigorously on a doctor's clinical experience and assessment abilities [5].

There is a need to develop computer-based heart sound detection techniques for the analysis, classification and detection of PCG signals. The computer-based heart detection techniques involve several steps such as (1) preprocessing of the PCG signal using filters, (2) extracting features and (3) designing of classifiers [6]. Various machine learning models have been developed for heart disease detection in the past years. To overcome the problems faced by the physicians and patients, five machine learning models—support vector machine (SVM), k-nearest neighbor (k-NN), decision tree (DT), random forest (RF) and extreme gradient boosting (XGBoost)—have been developed in this study.

The organization of the paper is as follows: The second section describes the related review literature year wise, and the third section clarifies the preprocessing methods such as bandpass filtering, zero-phase filtering and denoising using DWT and IDWT. The simulation study is done in the fourth section. In the fifth section, results and discussion are presented. In the last section, conclusions and future works are discussed.

2 Related Review Literature

This section manages the literature survey connected with heart disease identification to have a superior comprehension and to know the latest things. Many authors have developed various machine learning and deep learning models for heart disease classification. A systematic review of related studies is described in Tables 1 and 2.

Table 1 Systematic review of related literature year wise

Study	Year	Dataset	Feature extraction	Classifiers used
Rath et al. [7]	2022	Physionet Pascal CHSE	Discrete wavelet transform (DWT) Mel frequency cepstral coefficient (MFCC), DWT + MFCC	Random forest (RF), k-nearest neighbor (k-NN), extreme gradient boosting (XGB), **RF + moth fly optimization (MFO) + XGB + ensemble learning (EL)**
Rath et al. [8]	2022	PTB-ECG MIT-BIH	–	Autoencoder (AE), radial basis function network (RBFN), self-organizing map (SOM), restricted Boltzmann machine (RBM) **SOM + AE**
Rath et al. [9]	2021	MIT-BIH PTB-ECG	–	Support vector machine (SVM), Naive Bayes (NB), multilayer perceptron (MLP), long short-term memory (LSTM), generative adversarial network (GAN), **GAN-LSTM**
Abduh et al. [4]	2020	Physionet	MFSC, fractional Fourier transform (FrFT), FrFT-MFSC, principal component analysis (PCA)	**SVM**, ensemble classifier, k-NN
Soares et al. [10]	2020	Physionet	Dominant frequency, spectrum entropy, MFCC	**Autonomous learning multiple-model (ALMMo*), ALMMo-0**, **AdaBoost& CNN**, ensemble of SVMs, regularized neural network, MFCC + Wavelet + Tensors + k-NN, RF + LogitBoost, ensemble of NN, deep structured features
Demir et al. [11]	2019	PASCAL CHSE	Feature extraction using CNN	AlexNet, VGG16,19, AlexNet-VGG16, VGG16-19, **AlexNet-VGG16-19**
Malik et al. [12]	2018	PASCAL CHSE	Fast Fourier transform	**Robust adaptive algorithm**
Zhang et al. [13]	2017	PASCAL CHSE	Scaled spectrogram and tensor decomposition	J48, MLP, CS UCL, **SS-Method**, SS-PLSR, 2D-PCA, SVM-DM, **SS-TD**
Zhang et al. [14]	2017	PASCAL CHSE	DWT	J48, MLP, **CS UCL**, **SS-Method**, SS-PLSR

Table 2 Performances of various studies

Study	Subjects	Train/test scheme	Performance measures					
			Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)	AUC (%)
[7]	HD-816, N-2725 HD-100, N-231	Holdout (70:30) (80:20)	89.80 89.53	–	–	–	90.0 90.0	96.0 95.0
			86.25 87.89	–	–	–	86.0 88.0	91.0 93.0
[8]	HD-54, N-146 HD-104, N-164	70:30	99.20	–	–	–	98.6	99.5
			98.40	–	–	–	97.1	99.7
[9]	HD-146, N-214 HD-346, N-54	80:20	99.20	–	–	–	98.7	98.4
			99.40	–	–	–	99.3	99.5
[4]	HD-619, N-2249	CV	94.39	86.57	96.34	–	–	–
		80:20	87.35	96.66	94.69	–	–	–
[10]	HD-9857, N-3158	70:30	93.04	94.24	95.26	–	–	–
[11]	HD-243, N-212	Dataset A, B (80:20)	–	100	69.00	100	–	–
			–	54.00	91.00	81.0	–	–
[12]	HD-153, N-481	80:20	80.00	97.00	88.00	–	–	–
[13]	HD-818, N-3056	Dataset A, B, C (80:20)	–	–	–	67.00	–	–
			–	–	–	83.00	–	–
			–	–	–	96.0	–	–
[14]	HD-153, N-481	Dataset A, B (80:20)	–	–	–	67.00	–	–
			–	–	–	77.00	–	–

3   Materials and Methods

3.1   Proposed Methodology

The five proposed machine learning-based models, SVM [15], k-NN [15], RF [16], DT [17] and XGBoost [18], are developed in several steps, as shown in Fig. 1.

Step 1: PASCAL challenge competition [19] dataset is gathered from the website, which contains PCG signal in the form of .wav format.

Fig. 1 Proposed
methodology for the
classification of PCG signals
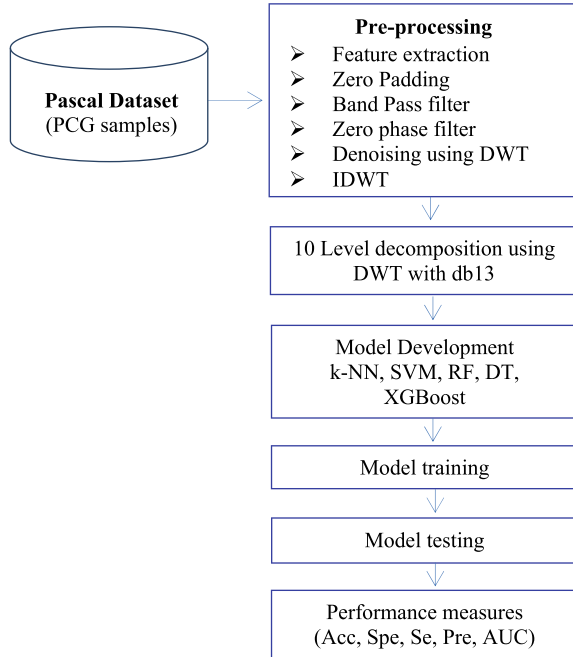
Step 2: The PCG signals are preprocessed using various preprocessing techniques such as bandpass filtering [7], zero-phase filtering [7], DWT [7], IDWT [7] and 10-level decomposition using DWT.

Step 3: Five machine learning-based models, SVM, k-NN, RF, DT and XGBoost, are developed for classification.

Step 4: Training and testing of the models with 70:30, 80:20 ratio and with fivefold cross-validation.

Step 5: Finally, the performance measures are evaluated (Accuracy, Sensitivity, Specificity, Precision, F1-score and AUC).

The proposed system will identify and forecast the patient's heart health condition by employing machine learning models. Five different machine learning methods are used for the proposed, and out of it, RF-based model outperforms the other four models on the basis of all performance measures.

3.2   Data Description

In this study, PCG signals are gathered from the standard dataset, i.e., PASCAL challenge competition [19] available on the website, which is provided in two sets, i.e., set A and B. Dataset A consists of total 124 recordings, and dataset B consists

Table 3 Dataset used in the experiment

Dataset	Set	Total no. of recordings	Total no. of recordings used	Normal PCG signal	Abnormal PCG signal
PASCAL challenge competition	Set A	124	65	31	34
	Set B	312	266	200	66
Total	A + B	436	331	231	100

of 312 recordings. Out of these, 65 recordings of set A and 266 of set B have been taken for the experiment to classify PCG signal as abnormal or normal, as presented in Table 3. The varying length of each audio file is in between 1 and 30 s long.

3.3 Preprocessing of PCG Signal

After gathering data from the websites, the signals need to preprocess using various techniques. This section presents feature extraction, filtering, denoising and bringing the signals back to the original form using inverse DWT.

Firstly, features are extracted from the PCG signals; then in the end of each signal, zero is padded using the zero padding technique. After that, the signals are passed through the band pass filter with the range of 50–600 Hz to eliminate noisy waves from the PCG signals. Still some contaminated noise is present in the PCG signals; zero-phase filtering is used to eliminate these. After that, the third-order Butterworth filter is applied to eliminate the contaminated noise effectively. Discrete wavelet transform (DWT) is deployed to the PCG signal, and the output is set at the suitable threshold. Finally, to obtain the denoised PCG signal, the signal is converted back to its original time domain form using inverse DWT.

In this study, the DWT is used to extract relevant features from the signals. The 'Daubechies 13', family of mother wavelet, is deployed up to ten decomposition levels. After decomposition, DWT extracts 545 features, which is passed to the proposed models as input. Figure 2a–c show the original, filtered and denoised PCG signal representation. The equation used for DWT is defined as [15]:

$$\theta(a \cdot b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} z(t)\theta' \left(\frac{t-b}{a} \right) dt \tag{1}$$

where

$\theta(a \cdot b)$ is the discretized wavelet,
a is the after translation of the wavelet $\theta(t)$,
b is the expansion of wavelet $\theta(t)$.

(a) Original PCG signal

(b) Filtered PCG signal

(c) Denoised PCG signal
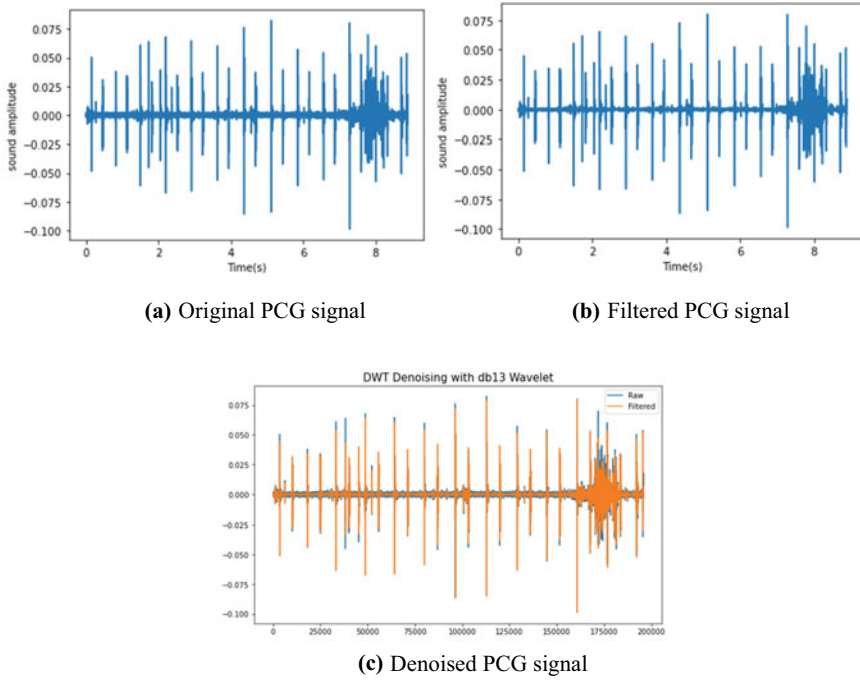
Fig. 2 **a** Original PCG signal. **b** Filtered PCG signal. **c** Denoised PCG signal

3.4 Generalized Models for Heart Disease Detection

Five generalized models, SVM, k-NN, RF, DT and XGB, are proposed in this study as classification. The proposed methodology is shown in Fig. 1. The PCG signal is preprocessed using filters, DWT and IDWT. The extracted features are then passed as input to the proposed models for detecting heart disease using PCG signals. Proposed classification models are described below:

3.4.1 Support Vector Machine (SVM)

It is a supervised machine learning algorithm, utilized for classification as well as for regression problems. It aims to create a hyperplane in 'm'-dimensional space to classify data points. Dimension depends upon the quantity of features. If the quantity of features is two, it creates a single-line hyperplane; if three, the hyperplane is a two-dimensional plane. SVM kernel is one function used to covert low-dimensional space into high-dimensional space. Its advantage is it is more effective in high-dimensional instances. Due to its good performance, most researchers used the SVM algorithm to analyze binary classification problems [15].

3.4.2 k-Nearest Neighbor (k-NN)

A supervised machine learning algorithm, k-NN [15], is used for classification as well as for regression problems. K neighbors are selected based on the similarity measures. Firstly, the k neighbors are chosen, and then Euclidean distance is calculated for all the k neighbors, select the k neighbors according to the minimum Euclidean distance, then count the data points of every class, and finally, a new data point is assigned to a class for which the quantity of neighbors is highest. It is a lazy learner because it does not have its training phase.

3.4.3 Random Forest (RF)

An ensemble learning algorithm, RF, is one of the powerful and popular machine learning techniques for classifying high-dimensional and skewed problems. It comprises multiple decision trees. Based on the majority voting strategy from the predicted class, it decides the final class. From the overall dataset, multiple decision trees use the subsets. Every node of multiple decision trees is splitted using Gini impurity which measures the Gini quality [16].

3.4.4 Decision Tree (DT)

The DT [17] is a tree-like structure, where every node represents the internal node which conducts the test on features, every branch of the tree addresses a result of the test and every terminal node (leaf node) represents a class label. A decision tree can learn by splitting the dataset into several subsets based on the feature test. The process is repeated for all the branches recursively. The recursion is finished when the subset at a node all has a similar value of the objective (target) variable. The advantages of DT are it takes less computation to perform classification, and it can deal with continuous as well as categorical variables.

3.4.5 Extreme Gradient Boosting (XGBoost)

XGBoost [18] is a boosting algorithm which is a successor of gradient boosted decision tree. Here, decision trees are sequentially created. Weight plays a significant role in the XGB and is assigned to all the input features and then fed to the decision tree, which produces the results. When the decision tree wrongly predicts that the weight is increased, then the features fed to the next decision tree. Finally, all the outputs from the trees are combined by using the prediction function to obtain the final class. Its advantages are that it is faster than other model gradient boosting, parallel processing, in-build features to handle missing values and updating weights using regularization.

4 Simulation Study

In this study, five machine learning models, SVM, k-NN, DT, RF and XGB, have been developed for heart disease classification using PCG signals. The implementation uses Python 3.7 version with various packages such as scikit-learn [20], NumPy and ensemble packages. The PCG signals are available in the .wav format audio files, which are publicly available on the PASCAL challenge competition website. The input given to all the five models is the preprocessed signals obtained from the original PCG signals. The DWT is applied to the models to extract features. Training and testing are done using two methods, i.e., holdout (70:30, 80:20) and cross-validation.

Firstly, models have been developed by splitting the datasets using 70:30 and 80:20 ratios, as shown in Table 4. 70%, 80% samples are used for training the models, and 30%, 20% samples are used for testing the models. In the training stage, the model learns from the 70 and 80% data. A total of 545 inputs are passed to the proposed models, and the hyperparameters are tuned for each model whose details are shown in Table 5. For the k-NN, the number of neighbors is set to 5. In SVM, linear kernel is chosen, with gamma = 'auto', class_weight = 'balanced', probability = 'true'. For DT, tree's maximum depth is set to 20, criterion = 'gini', which measures the number of splits of gini impurity. Similarly, for RF, the number of estimators is 100, and the maximum depth is 20. Finally, for XGB, estimators are set to 100 with a learning rate = 0.01, a maximum number of depth of the tree = 20, booster of the tree = 'gbtree' and evaluation metrics set to 'logloss'.

Cross-validation is one of the resampling techniques in which the data are divided into the k folds for training and testing. Fivefold cross-validation is used in this study to overcome the overfitting problem. 4/5th part is for the training, and the remaining

Table 4 Training and testing data

Ratio (%)	Training data	Testing data
70:30	231	100
80:20	264	67

Table 5 Hyperparameters used in the proposed models

Models	Hyperparameters used in the experiment
k-NN	n_neighbors = 5
SVM	Kernel = 'linear', gamma = 'auto', verbose = True, class_weight = 'balanced', probability = True
DT	max_depth = 20, random_state = 42, criterion = 'gini'
RF	n_estimators = 100, max_depth = 20
XGBoost	n_estimators = 100, use_label_encoder = False, learning_rate = 0.01, max_depth = 20, scale_pos_weight = 1.5, booster = 'gbtree', eval_metric = 'logloss'

1/5th part is used for the testing of the models on a number of iterations. The mean of each fold is the final result. Also, the same hyperparameters are tuned for all the models while using cross-validation.

5 Results and Discussions

5.1 Various Performance Measures

The results are obtained using various performance measures such as accuracy, sensitivity, specificity, precision, F1-score and AUC. The results of all the performance measures are based on the confusion matrix, a two-dimensional matrix representing true positive, true negative, false positive and false negative. All the performance measures are defined as [21]:

$$\text{Accuracy (Acc)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{2}$$

$$\text{Sensitivity (Se)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{Specificity (Spe)} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{4}$$

$$\text{Precision (P)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

$$\text{F1 - score} = 2 \times \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right) \tag{6}$$

where TP is the quantity of true-positive outcome, TN is the quantity of true negative outcomes, FP is the quantity of false-positive outcomes and FN is the quantity of false negative outcomes. The receiver operator characteristic (ROC) curve is also plotted between the TPR and FPR. Area under curve (AUC) score is also obtained.

5.2 Results of ML Models in Heart Disease Classification

Following are the results obtained from the five ML models using PCG signals.

It is observed from the above Tables 6, 7 and 8, RF outperforms the other four models and achieved AUC score of 98.33% using the holdout method (70%:30%). DT and RF achieved the same AUC of 96.42% by using the holdout method (80%:20%). Again RF outperformed and achieved AUC of 83.35% by using fivefold

CV. Figures 3, 4 and 5, show the graphical representation of various performance measures. Figures 6 and 7 represent the ROC plots for all the proposed models.

Table 6 Performance measures obtained by the various models using (70:30) ratio

Performance measures	70:30 (holdout method)				
	k-NN (%)	SVM (%)	DT (%)	RF (%)	XGBoost (%)
Accuracy	97	93	98	99	98
Sensitivity	93.33	96.66	96.66	96.66	93.33
Specificity	98.57	91.42	98.57	100	100
Precision	96.55	82.85	96.66	100	100
F1-score	94.91	89.23	96.66	98.30	96.55
AUC	95.95	94.04	97.61	**98.33**	96.66

Table 7 Performance measures obtained by the various models using (80:20) ratio

Performance measures	80:20 (holdout method)				
	k-NN (%)	SVM (%)	DT (%)	RF (%)	XGBoost (%)
Accuracy	95.52	94.02	98.50	98.50	97.01
Sensitivity	85.71	92.85	92.85	92.85	85.71
Specificity	98.11	94.33	100	100	100
Precision	92.30	81.25	100	100	100
F1-score	88.88	86.66	96.29	96.29	92.30
AUC	91.91	93.59	96.42	**96.42**	92.85

Table 8 Performance measures obtained by the various models using fivefold CV

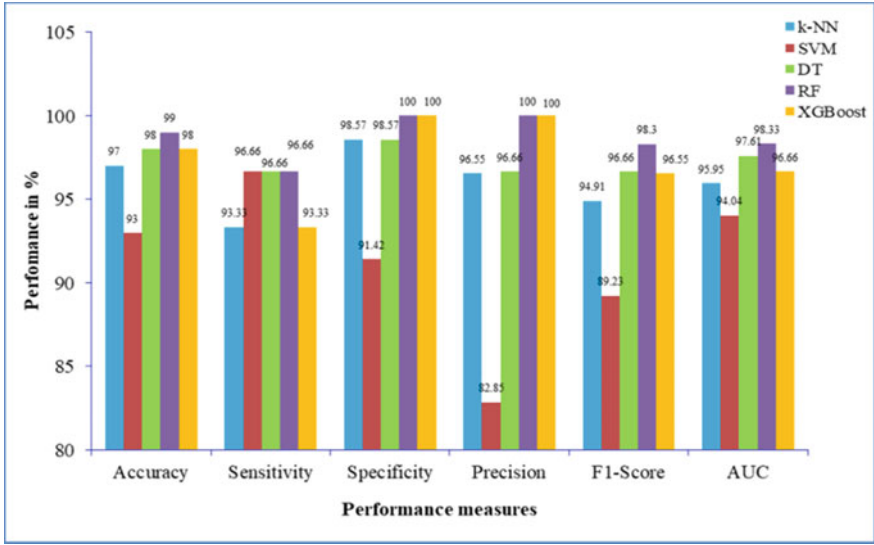| Performance measures	Fivefold cross-validation (CV)				
	k-NN (%)	SVM (%)	DT (%)	RF (%)	XGBoost (%)
Accuracy	79.39	81.47	75.75	**83.09**	77.59
Sensitivity	76.09	84.87	80.49	86.34	79.02
Specificity	82.47	78.60	71.72	80.36	76.40
Precision	85.90	85.56	82.76	87.33	82.43
F1-score	77.92	78.82	74.21	81.61	74.91
AUC	79.28	81.73	76.10	**83.35**	77.71

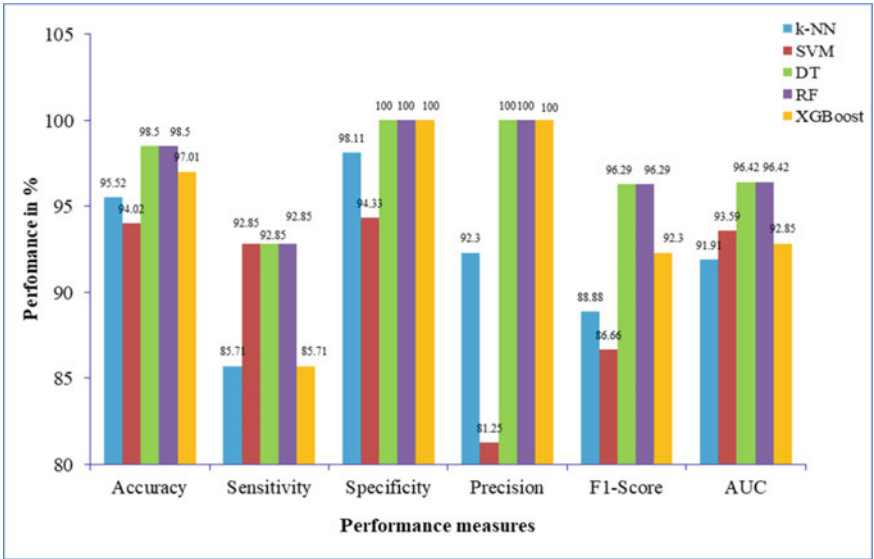**Fig. 3** Graphical representation of various performance measures using 70:30 ratio

Fig. 4 Graphical representation of various performance measures using 80:20 ratio
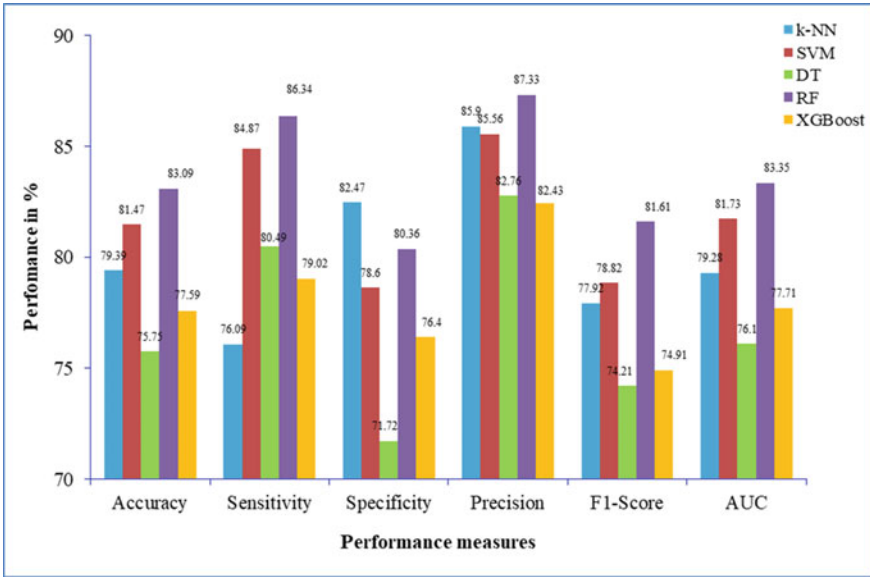
Fig. 5 Graphical representation of various performance measures using fivefold CV ratio
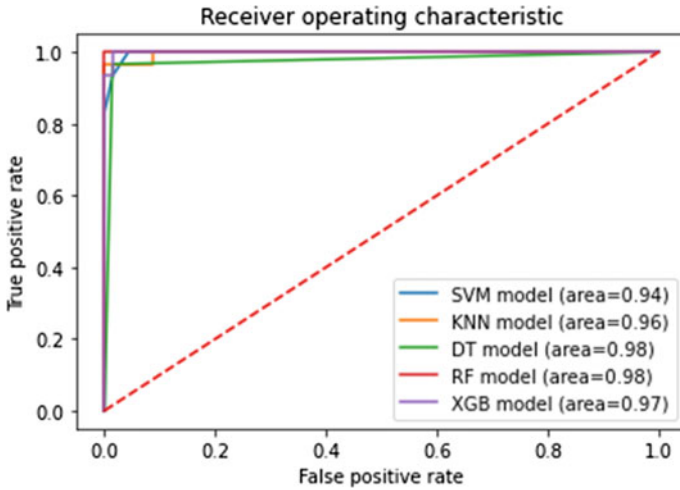
Fig. 6 ROC plot of all five models using 70:30 ratio
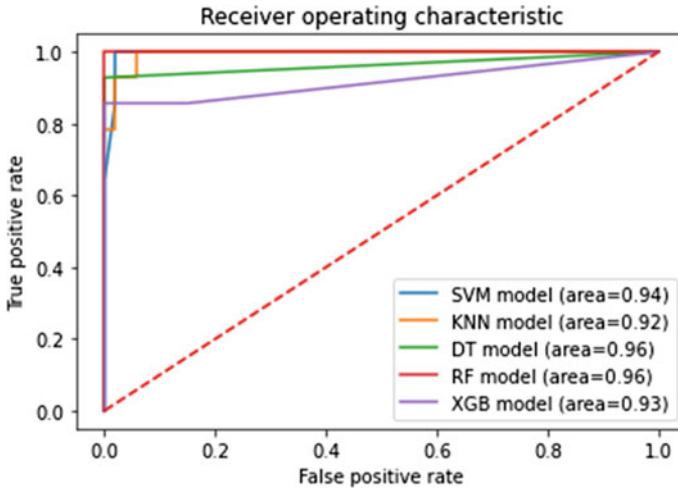
Fig. 7 ROC plot of all five models using 80:20 ratio

5.3 Comparison and Analysis of Proposed Model with Baseline Models

The comparison of proposed models with the existing baseline models is shown in the above Table 9.

In this study, three training approaches 70:30, 80:20 ratio and fivefold cross-validation are used. By applying the 70:30 and 80:20 splitting, RF model achieves better performances nearer to the existing baseline models [8, 9] (Acc = 99, Sen = 96.66, F1_scr = 98.30, AUC = 98.33) and (Accuracy = 98.50, Sensitivity = 92.85, F1_score = 96.29 and AUC = 96.42), where specificity and precision achieve better results (Specificity = Precision = 100) as per their results.

5.4 Challenges Faced During Implementation

The challenges faced while implementing the proposed models are as follows:

1. Conversion of the raw data from the available PCG signals recorded for 1–30 s of different patients to comma separated value format.
2. Since the dimension of the data is very high, it increases the computational complexity and requires more memory space which is more than 8 GB RAM particularly in our case.
3. Careful preprocessing task is to be performed to get accurate results.

Table 9 Comparison of proposed model with baseline models

Study	Classifiers	Performance measures					
		Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)	AUC (%)
[7]	RF + moth fly optimization (MFO) + XGB + ensemble learning (EL)	89.80 89.53	–	–	–	90.0 90.0	96.0 95.0
		86.25 87.89	–	–	–	86.0 88.0	91.0 93.0
[8]	SOM + AE	99.20	–	–	–	98.6	99.5
		98.40	–	–	–	97.1	99.7
[9]	GAN-LSTM	99.20	–	–	–	98.7	98.4
		99.40	–	–	–	99.3	99.5
[4]	SVM	94.39	86.57	96.34	–	–	–
		87.35	96.66	94.69	–	–	–
[10]	(ALMMo*), ALMMo-0, AdaBoost & CNN	93.04	94.24	95.26	–	–	–
[11]	AlexNet-VGG16-19	–	100	69.00	100	–	–
		–	54.00	91.00	81.0	–	–
[12]	Robust adaptive algorithm	80.00	97.00	88.00	–	–	–
[13]	SS-Method, SS-TD	–	–	–	67.00	–	–
		–	–	–	83.00	–	–
		–	–	–	96.0	–	–
[14]	SS-Method, CS UCL	–	–	–	67.00	–	–
		–	–	–	77.00	–	–
Proposed methodology	k-NN, SVM, **RF**, DT, XGB	**99.00**	**96.66**	**100**	**100**	**98.30**	**98.33**
		98.50	92.85	**100**	**100**	96.29	96.42
		83.09	86.34	80.36	87.33	81.61	83.35

6 Conclusion and Future Work

This study addresses the PCG recording problems on the PASCAL challenge competition data. The work lies in the detection of heart disease from the PCG signals and involves various preprocessing techniques such as filtering the signals using a bandpass filter, removing noise present in the signals using zero-phase filters, denoising a signal using DWT and bring signals back to the time domain format using IDWT and features extracted using DWT. Five machine learning models have been developed in this study using SVM, k-NN, RF, DT and XGBoost. It can be seen that from all the five proposed models, RF performs better and achieves the best AUC of 98.33% in 70:30 ratio, 96.43% in 80:20 ratio and 83.35 in fivefold cross-validation. It is still a challenging problem for heart disease detection using PCG signals. The research is continuous, and many more authors are trying to develop machine learning and deep

learning models from which models can work more accurately. It will be beneficial for the doctors and cardiac patients to early diagnose the disease. In the future, authors will try to develop deep learning models with various preprocessing techniques and different datasets.

References

1. Mayr FB, Spiel A, Leitner J, Marsik C, Germann P, Ullrich R, Wagner O (2005) Effects of carbon monoxide inhalation during experimental endotoxemia in humans. Am J Respir Crit Care Med 171(4):354–360
2. de Carvalho P, Paiva RP, Couceiro R (2010) Comparison of systolic time interval measurement modalities for devices. In: 2010 Annual international conference of the IEEE engineering in medicine and biology. IEEE, pp 606–609
3. Dwivedi AK (2018) Performance evaluation of different machine learning techniques for prediction of heart disease. Neural Comput Appl 29:685–693. https://doi.org/10.1007/s00521-016-2604-1
4. Abduh Z, Nehary EA, Wahed MA, Kadah YM (2020) Classification of heart sounds using fractional Fourier transform based Mel-frequency spectral coefficients and traditional classifiers. Biomed Signal Process Control 57:101788. https://doi.org/10.1016/j.bspc.2019.101788
5. Deng M, Meng T, Cao J, Wang S, Zhang J, Fan H (2020) Heart sound classification based on improved MFCC features and convolutional recurrent neural networks. Neural Netw 130:22–32. https://doi.org/10.1016/j.neunet.2020.06.015
6. Yang X, Yang F, Like G (2016) A multi-modal classifier for heart sound recordings. In: 2016 Computing in cardiology conference (CinC), pp 1165–1168
7. Rath A, Mishra D, Panda G, Pal M (2022) Development and assessment of machine learning based heart disease detection using imbalanced heart sound signal. Biomed Signal Process Control 76:103730. https://doi.org/10.1016/j.bspc.2022.103730
8. Rath A, Mishra D, Panda G, Satapathy SC, Xia K (2022) Improved heart disease detection from ECG signal using deep learning based ensemble model. Sustain Comput Inform Syst 35:100732. https://doi.org/10.1016/j.suscom.2022.100732
9. Rath A, Mishra D, Panda G, Satapathy SC (2021) Heart disease detection using deep learning methods from imbalanced ECG samples. Biomed Signal Process Control 68:102820. https://doi.org/10.1016/j.bspc.2021.102820
10. Soares E, Angelov P, Gu X (2020) Autonomous learning multiple-model zero-order classifier for heart sound classification. Appl Soft Comput 94:106449. https://doi.org/10.1016/j.asoc.2020.106449
11. Demir F, Sengur A, Bajaj V, Polat K (2019) Towards the classification of heart sounds based on convolutional deep neural network. Health InfSci Syst. 7(1):16. https://doi.org/10.1007/s13755-019-0078-0
12. Malik SI, Akram MU, Sissiqi I (2019) Localization and classification of heartbeats using robust adaptive algorithm. Biomed Signal Process Control 49:57–77. https://doi.org/10.1016/j.bspc.2018.11.003
13. Zhang W, Han J, Deng S (2017) Heart sound classification based on scaled spectrogram and tensor decomposition. Expert Syst Appl 84:220–231. https://doi.org/10.1016/j.eswa.2017.05.014
14. Zhang W, Han J, Deng S (2017) Heart sound classification based on scaled spectrogram and partial least squares regression. Biomed Signal Process Control 32:20–28. https://doi.org/10.1016/j.bspc.2016.10.004
15. Singh SA, Devi ND, Majumde S (2021) Heart abnormality classification using PCG and ECG recordings. Computación y Sistemas 25(2):381–391. https://doi.org/10.13053/CyS-25-2-3447

16. Breiman L (2001) Random forests. Mach Learn 45:5–32
17. Langley P, Murray A (2017) Heart sound classification from unsegmented phonocadiogram. Physiol Meas 38:1658–1670. https://doi.org/10.1088/1361-6579/aa724c
18. Ahmad TA, Hanaa IE, Aboul EH, Abeer ME (2014) A random forest classifier for lymph diseases. Comput Methods Prog Biomed 113(2):465–473. https://doi.org/10.1016/j.cmpb.2013.11.004
19. Bentley P, Nordehn G, Coimbra M, Mannor S (2011) Classifying heart sounds challenge. http://www.peterjbentley.com/heartchallenge
20. Pedregosa F, Weiss R, Brucher M (2011) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830
21. Han J, Pei J, Kamber M (2011) Data mining: concepts and techniques, 3rd edn. Morgan Kaufmann Publishers, Elsevier

Enigmas of Various Techniques to Implementing Authentication and Integrity in Blockchain-Based Wireless Sensor Networks

Tejbir Singh and Rohit Vaid

Abstract Wireless sensor network (WSN) is a part of wireless network containing large number of sensor nodes with limited computation capability to handle sensing data. A major challenge in WSN is security and privacy of user data, as this data is transferred through sensor nodes on wireless medium at the command station. This paper discusses enigmas of security to prevent the data from unauthentic and unauthorized nodes by using communication rule-based, association rule-based and block validity-based techniques. The purpose of using these three techniques is for proper communication and secure transmission in WSN using blockchain-based technology where communication rule-based technique is used for proper communication, whereas association rule-based techniques is used to maintain the links for proper data transfer and block validity-based technique is best for proper data transmission in WSN.

Keywords WSN · Security · Blockchain · BWSN · BAI · Privacy

Abbreviations

AN	Authentic node
BS	Base station
BWSN	Blockchain-based WSN
BAI	Blockchain-based authentication and integrity
CH	Cluster head
DN	Destination node
IN	Intermediate node
NS2	Network simulator 2
SN	Source node

T. Singh (✉) · R. Vaid
Department of Computer Science and Engineering, Maharishi Markandeshwar (Deemed to be University), Mullana, Haryana, India
e-mail: Tejbirrana662@gmail.com

UN   Unknown node
WSN   Wireless sensor network

1 Introduction

A present problem in wireless sensor networks is security, with different limitations, small resources with hardware functionalities and energy utilization. This work is focused on security on WSN, i.e., malicious nodes occur when the data transfer in wireless sensor network from one source node to other destination node. Together have been dynamically search the last ten years, with a number of solutions discuss on some literature survey in this paper.

The motivation of this research is securities is biggest problem of wireless sensor network, and when discuss the blockchain on behalf of wireless sensor network, first focus on security of data and transactions. So, according to previous research, focus on security for blockchain-based WSN with the techniques of authentication and integrity.

This paper introduces blockchain technology with a relate algorithm to make certain the authority of the blockchain in the wireless sensor network (WSN) known as blockchain-based WSN (BWSN). First part of this paper explains briefly the WSN with its characteristics and also explains blockchain data pattern for security in WSN. Section two present concerns the techniques of authentication and integrity based on blockchain for WSN. The last part of this paper illustrates there presentation of a blockchain-based algorithms on behalf of authentication and integrity which offer methods to handle the user data in WSN.

When consider the security for WSN-based blockchain in future, first focus on secure data transfer from end-to-end node, second focus on transfer the secure information between intermediate nodes and third focus on secure data storage in system.

If focused on security under leveraging for personal data running platform in WSN, then use the blockchain. To propose a structure based on blockchain for decentralized authentication and data integrity. Following Fig. 1 shows the deployment of sensor node with authentic and unknown nodes which connected to base station in WSN as:
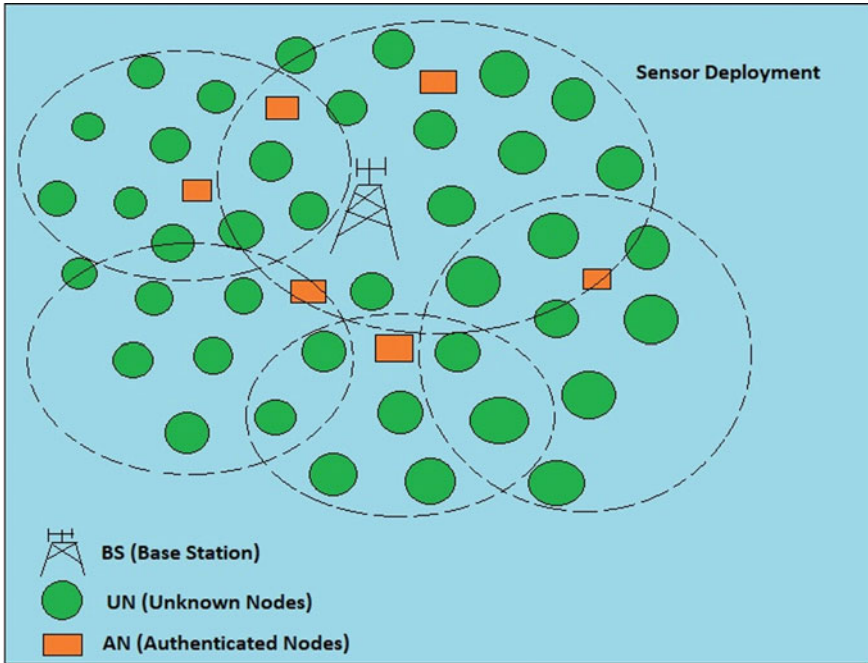
Fig. 1 Sensor deployment on BWSN

2 Literature Review

Following Table 1 shows the review of literature for the work on next research directions as:

3 Theoretical Background

A wireless sensor network (WSN) is built by a large number of sensor nodes with low cost that have restricted sensing communication and computation capabilities [19, 20]. It is main the amount of data broadcast is minimized due to resource limited sensor nodes so that the normal sensor lifetime as well as the generally bandwidth consumption are enhanced [21]. It is a difficult work to keep perceptive information sends out through wireless sensor networks due to unreceptive environments with exclusive properties of WSN [20]. As well, WSN has protection difficulties that conventional networks do not face. Therefore, security is a main problem for WSN. Following Fig. 2 shows the conceptual structure of WSN as:

If concern the blockchain for a private data management proposal [22], considered on privacy [23]. They summarized how the blockchain assists control the leveraging more than data in

Table 1 Literature review

References	Category	Research contribution	Research gap
Alladi et al. [1]	Security of vehicular network	In this review, analysis the security on blockchain framework with three important perspectives such as security, application and blockchain in vehicular network security [1]	Focus on requirements of security under low computation, low latency and data storage
Amjad S. et al. [2]	Blockchain-based authentication model under Internet of sensor Things	Proposed the work on IosT-based network with malicious activity and less computation cost by node under blockchain [2]	Focus on efficient ML techniques for malicious node detection
Awan S. et al. [3]	Trust management and secure routing based on blockchain under WSN	The author's proposed, trust evaluation model with secure routing for authentication and trust management under WSN [3]	Focus on real-world network for trust model implementations
Ramasamy et al. [4]	Malicious node detection by BWSN	Proposed the survey of blockchain-based WSN (BWSN) under data sharing, data security, data storage and malicious node detection [4]	Focus on improvement of scalability, energy expenditure, complexity and privacy
Gautam and Kumar [5]	WSN with techniques of trust management, authentication and key management	Proposed the paper of WSN with security issues under unknown environment and point to authentication, trust and key management under WSN [5]	Focus on stability, scalability and storage
Hsiao and Sung [6]	Security of WSN by blockchain technology	Proposed the paper for data security in case of transmission delay and data amount increase by block chain in WSN, when data transmission [6]	Focus on the difficulties of blockchain under innovative methods
El-Moghith and Darwish [7]	WSN with trust routing scheme as deep blockchain approach	In this paper proposed the trust routing approach for improve the performance of routing network under WSN using deep blockchain approach [7]	Focus on the route scheduling techniques for portability and efficacy

<div align="right">(continued)</div>

Table 1 (continued)

References	Category	Research contribution	Research gap
Awan et al. [8]	Secure routing in WSN based on blockchain for authentication and trust evaluation	In this paper proposed the security to WSN with respect to authentication and trust management using blockchain for smart connectivity with secure routing [8]	Focus on multi-WSN for inter-network performance
Hsiao and Sung [9]	Sensor data transmission under secure WSN using blockchain utilization	In this paper, propose the method for improve the security by blockchain utilization of WSN with handle the large amounts of data [9]	Focus on multi-sensor data with public record database in WSN
Hidar et al. [10]	Authentication-based solution of remote surgery using blockchain under tactile internet	The author's proposed to security of tactile internet environment for user authentication by using blockchain on remote surgery [10]	Focus on latency under fog computing with high level of QoE (quality of experience)
Babu and Jayalakshmi [11]	Trust-based key management in WSN by blockchain	In this implementation, proposed the trust-based key management with focus the issues of security like data confidentiality and integrity on behalf of data aggregation in WSN through blockchain [11]	Focus on energy efficiency under high-level cryptosystem
Miraz and Ali [12]	Security perspective of IoT with blockchain	In this research proposed blockchain-based security in IoT system with emerging technology, issues of security and security improvement by blockchain with application [12]	Focus the data and network redundancy when distribution in decentralized application
Tariq et al. [13]	Blockchain-based VANET, WSN, healthcare and with IoT	In this review paper, discuss the security of WSN, VANET, healthcare devices and IoT using blockchain [13]	Focus on the issues of security, privacy, data storage and computational efficiency with scalability by consensus algorithm

Table 1 (continued)

References	Category	Research contribution	Research gap
Guerrero-Sanchez et al. [14]	Blockchain-based WSN with symmetric encryption	In this paper, analysis the humidity and temperature data with respect to data accuracy and transfer to user with security by symmetric encryption method as AES in IoT under WSN [14]	Focus on next issues like linking and DDOS attack in WSN
Zhang et al. [15]	Applications of blockchain-based systems	The author's proposed the review of blockchain-based systems in sensor network with its applications in different fields [15]	Focus on challenges for the development of industry
Ferrag et al. [16]	Green IoT-based agriculture with privacy and security with blockchain	The author's proposed the survey of blockchain-based security in green IoT-based agriculture with security issues, challenges and solutions [16]	Focus on the challenges of blockchain-based solutions for scalability and intrusion detection with the techniques of machine learning
Torky and Hassanein [17]	Blockchain on precision agriculture discipline	In this review proposed the analysis, challenges and opportunities of integrated blockchain on behalf of precision agriculture-based discipline with IoT under WSN environment [17]	Focus on the connectivity between IoT and blockchain
Alsaedy et al. [18]	Blockchain-based WSN	The authors proposed the review of blockchain-based WSN on behalf of energy efficiency, security, data reliability, node recovery and data storage with combination of multiple technologies [18]	Focus on blockchain technology against hacking

the perspective of social networks as well as big data [24]. In information-centric networking-based architectures use distributed related privacy structure by blockchains for safe data access [25]. These schemes verify with working of the blockchain with safe distributed data structure on behalf of novel applications [26], other than not any exploit to give node authentication with integrity in WSNs.

There are some important functions used in blockchain-based authentication and integrity (BAI) under BWSN as:

ObjIdExists (Integer objId, Blockchain BC): This function is used for the object identifier under blockchain. When finding a particular selective node or selective
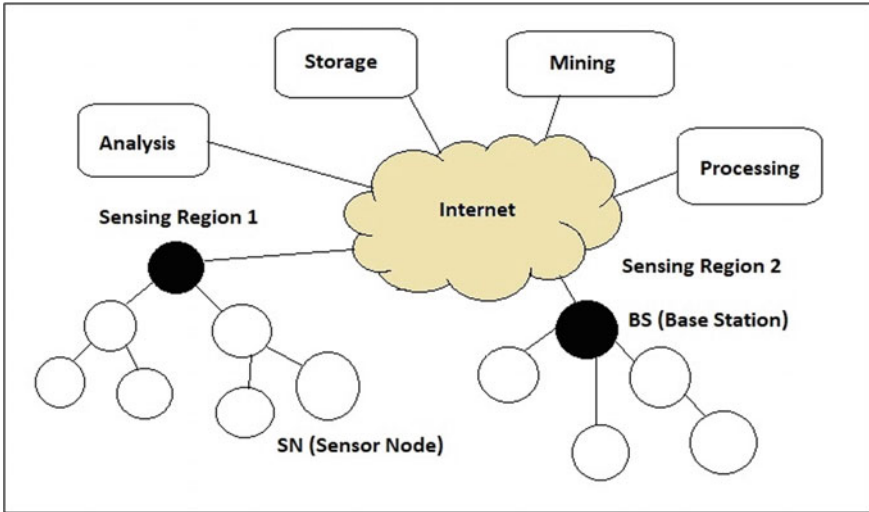
Fig. 2 WSN with sensing region

object for discover a path on WSN and finding the node for observe the security under blockchain, then use this function. In this research, this function is used for select a particular node under large number of nodes in WSN.

GrpIdExists (Integer grpId, Blockchain BC): This function is used for the group identifier in blockchain. When finding the group of nodes with different id of nodes, each node has its cluster head in sensor network and then uses this function. In this research work, this function is used for selecting a group of nodes for data transmission with security in WSN.

AddrExists (Integer objAddr, Blockchain BC): This function is used for find the address of object. When finding the address of particular object node under many object nodes on WSN for secure data transmission with discover a particular path, then use this function. In this research, this function is use for discover the address of particular node for transfer the data from one source node to other destination node with security.

Error (): This is used for return error in blockchain algorithms. When source node sends the data with message to end node and destination node not accepting the data properly with message due to error occurs in network, then use this function. In this research, when message delivery is failed due to data transmission is break, then use this function. The conceptual structure of blockchain algorithms is shown in Fig. 3 as:
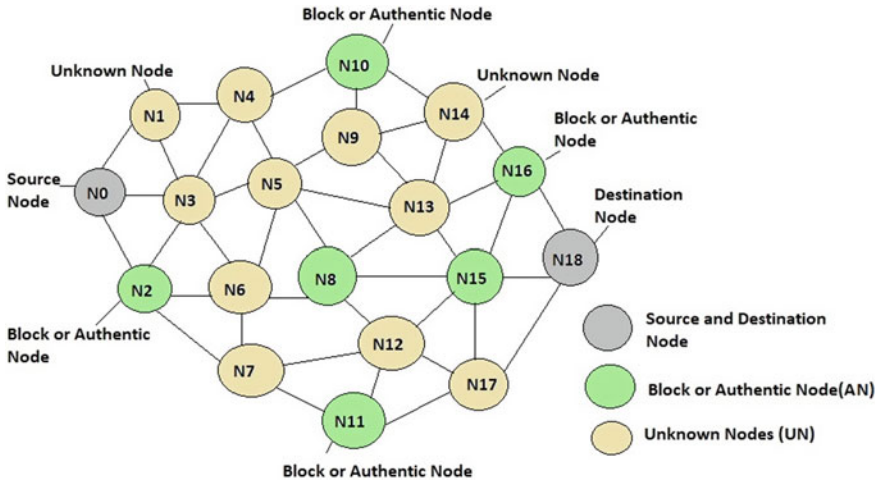
Fig. 3 Blockchain under WSN

4 Proposed Work

In this section discussed the proposed work for this research. Firstly study the different research papers regarding to this research under secondary data and then find the limitations or research gaps. According to research gaps, the different security techniques are required for blockchain-based WSN under different routing attacks in network. On behalf of research gaps, find the new work like this research. Following points shows the propose work as:

4.1 Theoretical study of background for blockchain-based WSN.
4.2 Collection of literature survey on behalf of blockchain-based WSN with latest works.
4.3 Design of blockchain structure on behalf of WSN.
4.4 Design of structure and algorithm of techniques for authentication and integrity of blockchain-based WSN.

5 Techniques of Blockchain-Based Authentication and Integrity (BAI) in WSN

There are three main techniques of blockchain-based authentication and integrity (BAI) in BWSN for security as:

5.1 Communication rule-based BAI
5.2 Association rule-based BAI
5.3 Block validity-based BAI.

5.1 Communication Rule-Based BAI

Communication rule-based BAI is a technique, for proper communication between source node and destination node using blockchain in WSN [26]. In this technique, source node sends the data to destination node through unknown activate node in large sensor network, and this network is based on multi-path. When using the communication-based technique, every nodes connected properly in sensor networks, shown in Fig. 4 as:

The algorithm of BAI is based on communication rule as:

Algorithm (identify source node 'S' and destination node 'D' with blockchain 'BC' and object 'OBJ')
Step 1. Start
IF (ObjIdExists (source.id, BC) = FALSE
{
(ObjIdExists (destination.id, BC) = FALSE)
{
Then Return Error ();
}}
Step 2. IF (source.grpId = destination.grpId)
{
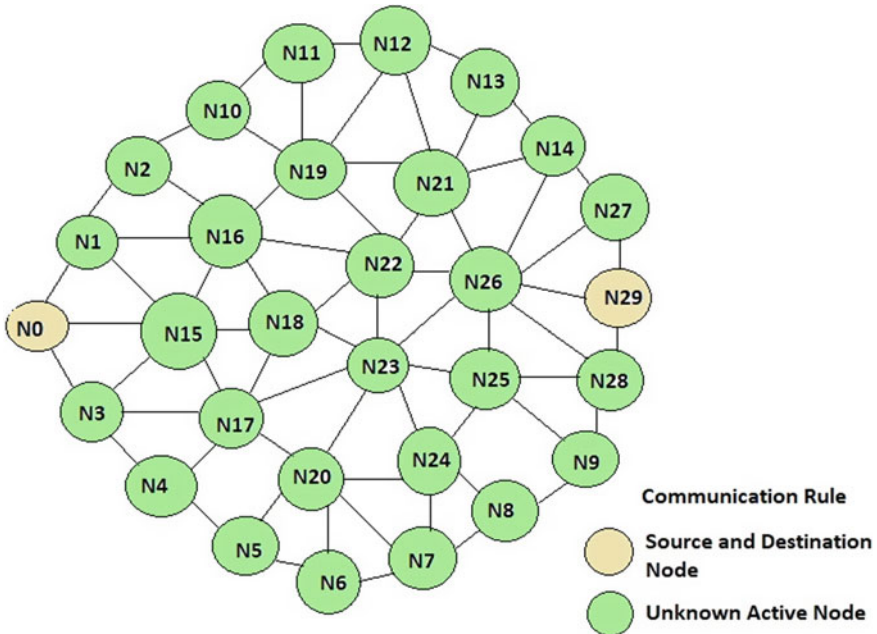
Fig. 4 BAI with communication rule

Then
{
Return Error ();
}}
Step 3: IF (bc.SignVerif(source.msg)) = Failed
{
Then
{
Return Error ();
}}
// protected data exchange completed as well as success

5.2 Association Rule-Based BAI

When intermediate node is out of control and difficult to handle the large datasets in sensor network, then use association rule [26]-based BAI technique. In this technique, secure node finds the interesting links for relationships between large numbers of nodes with handling different size of data in networks for proper data transfer, shown in Fig. 5 as:

The algorithm of this technique is as:

Algorithm (identify source node 'S' and destination 'D' with master node 'M' and follower node 'F' with BlockChain 'BC')
Step1: IF (ObjIdExists (obj.id, BC) = TRUE) then
{Return Error();
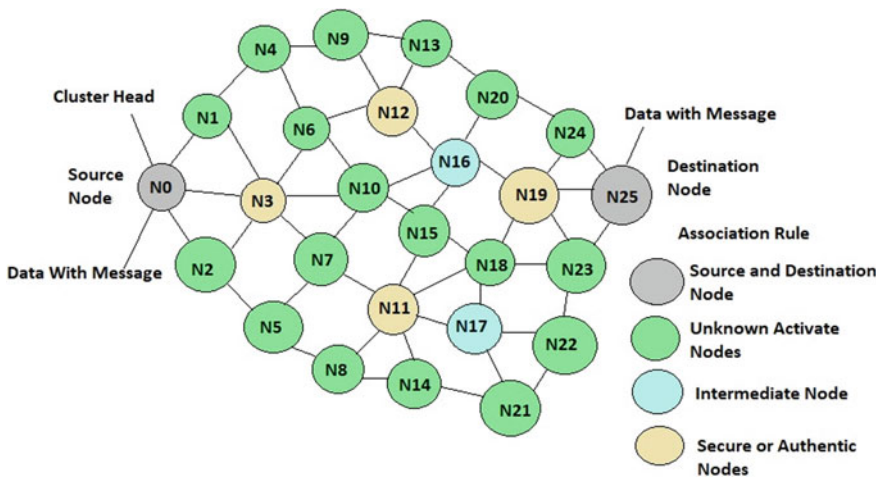
Fig. 5 BAI with association rule

```
}
Step2: IF (ObjIdExists (obj.id, BC) = TRUE) then
{
Return Error();
}
Step 3: IF (obj.type = M) then
{
IF GrpIdExists(obj.grpId, BC) = True then
{
Return Error ();
}
}
Step 4: ELSE IF (obj.type = F) then
{
IF GrpIdExists(obj.grpId, BC) = FALSE then.
{
Return Error ();
}
}
Step 5: IF (BC.TicketVerif(obj.ticket) = FAILED) then.
{
Return Error ();
}
ELSE
{
Return Error ();
}
}
// Association completed and Success
```

5.3 Block Validity-Based BAI

In this method, combine the communication rule and association rule-based BAI with block mining and mutual influence [27]. This technique is used for handling to multiple large number of nodes with security and large number of data sets. When using this technique in sensor network, every nodes is secure in network which passes the data from source to destination. In this technique, each secure node has blocks and activates with authentication and integrity, shown in Fig. 6 as:

The algorithm for block validity-based BAI is as:

Algorithm (Identify Present Block 'C' and Previous Block 'P', Hash Current Block 'H' with block validity 'V')

Fig. 6 BAI with block validity methods

```
Step 1: IF NOT (solve problem of H) then
    {
    Return FALSE;
}
Step 2: End IF
{
    IF NOT (Miner support payload VALID) THEN
    {
        Return FALSE;
    }
}
Step 3: End IF
{
    IF C has Event Payload on behalf of miner Then
    {
        Return FALSE;
    }
}
Step 4: End IF
{
    IF NOT (Every one payloads in V) Then
    {
        Return FALSE;
    }
}
```
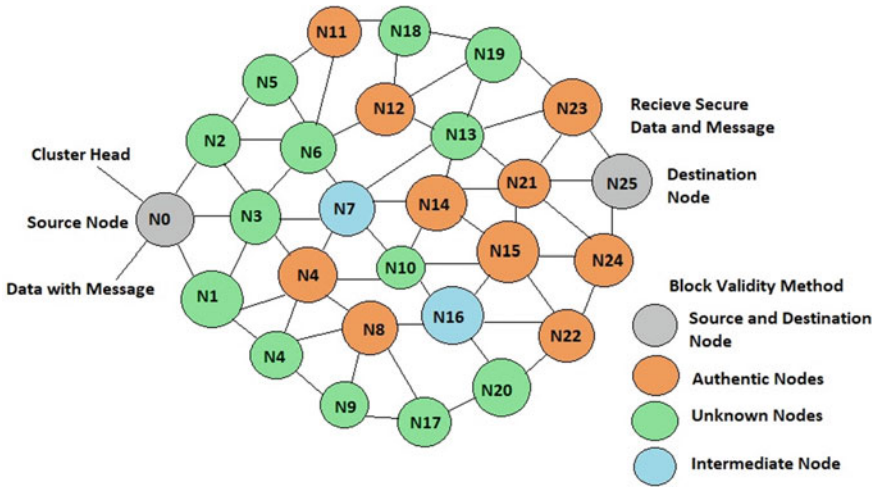
Step 5: End IF
{
        Return TRUE;
}
// block validity successfully check receive

6   Simulation

When consider the implementation of algorithm for authentication and integrity under blockchain-based WSN, use network-based simulator for analysis using different performance parameters which indicates to security. In this research work, we use NS2 simulator for implementation of authentication and integrity-based algorithms in blockchain-based WSN.

7   Conclusion and Future Scope

Security and privacy is a key part of WSN. Before, starting to this review paper, study to different article, review and implementation-based papers on behalf of blockchain-based security and WSN-based security. If completed the study, then find the limitations and then start new work on behalf of research gaps. According to research gaps, Authentication and integrity is an important limitation of blockchain-based security. In this research study paper, discuss the authentication and integrity-based techniques of blockchain-based WSN. In this study, describe the WSN with security using blockchain technology. WSN is a sensor-based network which has dynamical nature, so face the challenges of security. In this article, focus the security via blockchain technology in expressions of authentication and integrity. Blockchain is a protected data structure which holds the records of data and transactions with security. In this paper, discuss on blockchain-based WSN (BWSN) for security on behalf of authentication and integrity as blockchain-based authentication and integrity (BAI). The three algorithm used in this work as communication rule-based BAI, association rule-based BAI, and block validity-based BAI. The outcome of this study is blockchain is best technology for security in wireless sensor network (WSN). The future work of this paper is practical implementation of this study in blockchain-based WSN (BWSN) network on behalf of blockchain-based authentication and integrity (BAI) techniques. The future scope of this study paper is focus on data security, secure transaction and secure data storage with information in BWSN when handling the multi-sensor data.

References

1. Alladi T, Chamola V, Sahu N, Venkatesh V, Goyal A, Guizani M (2022) A comprehensive survey on the applications of blockchain for securing vehicular networks. arXiv preprint arXiv:2201.04803

2. Amjad S, Abbas S, Abubaker Z, Alsharif MH, Jahid A, Javaid N (2022) Blockchain based authentication and cluster head selection using DDR-LEACH in internet of sensor things. Sensors 22(5):1972

3. Awan S, Javaid N, Ullah S, Khan AU, Qamar AM, Choi J-G (2022) Blockchain based secure routing and trust management in wireless sensor networks. Sensors 22(2):411

4. Ramasamy LK, Firoz Khan KP, Imoize AL, Ogbebor JO, Seifedine K, Seungmin R (2021) Blockchain-based wireless sensor networks for malicious node detection: a survey. IEEE Access 9:128765–128785

5. Gautam AK, Kumar R (2021) A comprehensive study on key management, authentication and trust management techniques in wireless sensor networks. SN Appl Sci 3(1):1–27

6. Hsiao S-J, Sung W-T (2021) Employing blockchain technology to strengthen security of wireless sensor networks. IEEE Access 9:72326–72341

7. Abd El-Moghith IA, Darwish SM (2021) Towards designing a trusted routing scheme in wireless sensor networks: a new deep blockchain approach. IEEE Access 9:103822–103834

8. Awan S, Sajid MBE, Amjad S, Aziz U, Gurmani U, Javaid N (2021) Blockchain based authentication and trust evaluation mechanism for secure routing in wireless sensor networks. In: International conference on innovative mobile and internet services in ubiquitous computing. Springer, Cham, pp 96–107

9. Hsiao S-J, Sung W-T (2021) Utilizing blockchain technology to improve WSN security for sensor data transmission. CMC Comput Mater Continua 68(2):1899–1918

10. Hidar T, El Kalam AA, Benhadou S, Mounnan O (2021) Using blockchain based authentication solution for the remote surgery in tactile internet. Int J Adv Comput Sci Appl (IJACSA) 12(2):277–281

11. Babu TG, Jayalakshmi V (2021) Trust-based key management conglomerate ELGamal encryption for data aggregation framework in WSN using blockchain technology. Int J Adv Comput Sci Appl (IJACSA) 12(11):228–236

12. Miraz MH, Ali M (2020) Integration of blockchain and IoT: an enhanced security perspective. Ann Emerg Technol Comput (AETiC) 4(4). arXiv preprint arXiv:2011.09121

13. Tariq F, Anwar M, Janjua AR, Khan MH, Khan AU, Javaid N (2020) Blockchain in WSNs, VANets, IoTs and healthcare: a survey. In: Workshops of the international conference on advanced information networking and applications. Springer, Cham, pp 267–279

14. Guerrero-Sanchez AE, Rivas-Araiza EA, Gonzalez-Cordoba JL, Toledano-Ayala M, Takacs A (2020) Blockchain mechanism and symmetric encryption in a wireless sensor network. Sensors 20(10):2798

15. Zhang J, Zhong S, Wang T, Chao H-C, Wang J (2020) Blockchain-based systems and applications: a survey. J Int Technol 21(1):1–14

16. Ferrag MA, Shu L, Yang X, Derhab A, Maglaras L (2020) Security and privacy for green IoT-based agriculture: review, blockchain solutions, and challenges. IEEE Access 8:32031–32053

17. Torky M, Hassanein AE (2020) Integrating blockchain and the internet of things in precision agriculture: analysis, opportunities, and challenges. Comput Electron Agric 178:105476

18. Alsaedy S, Alraddadi S, Owais A (2020) A review on using blockchain in wireless sensor networks. J Theor Appl Inform Technol 98(23):3879–3887. ISSN: 1992-8645, E-ISSN: 1817-3195

19. Yang J, He S, Xu Y, Chen L, Ren J (2019) A trusted routing scheme using blockchain and reinforcement learning for wireless sensor networks. Sensors 19(4):970

20. Ozdemir S, Xiao Y (2009) Secure data aggregation in wireless sensor networks: a comprehensive overview. Comput Netw 53(12):2022–2037

21. Raj Anand S, Tanguturi RC, Soundarrajan DS (2019) Blockchain based packet delivery mechanism for WSN. Int J Recent Technol Eng (IJRTE) 8(2):2277–3878

22. Uddin MA, Stranieri A, Gondal I, Balasubramanian V (2018) A patient agent to manage blockchains for remote patient monitoring. Stud Health Technol Inform 254:105–115
23. Kim T-H, Goyat R, Rai MK, Kumar G, Buchanan WJ, Saha R, Thomas R (2019) A novel trust evaluation process for secure localization using a decentralized blockchain in wireless sensor networks. IEEE Access 7:184133–184144
24. Uddin MA, Stranieri A, Gondal I, Balasubramanian V (2019) Blockchain leveraged task migration in body area sensor networks. In: 2019 25th Asia-Pacific conference on communications (APCC). IEEE, pp 177–184
25. Ren Y, Liu Y, Ji S, Sangaiah AK, Wang J (2018) Incentive mechanism of data storage based on blockchain for wireless sensor networks. Mobile Inform Syst 2018:1–10
26. Hammi MT, Hammi B, Bellot P, Serhrouchni A (2018) Bubbles of trust: a decentralized blockchain-based authentication system for IoT. Comput Sec 78:126–142
27. Moinet A, Darties B, Baril J-L (2017) Blockchain based trust & authentication for decentralized sensor networks. arXiv preprint arXiv:1706.01730

A Systematic Review on Blockchain-Based e-Healthcare for Collaborative Environment

Deepak Singla and Sanjeev Kumar Rana

Abstract The utilization of recent data management support can be utilized to continuously carry out a variety of analytical activities, speeding up interactions between parties. Conventional healthcare service systems can help build a secure network. The existing EHR software's security and privacy concerns restrict how much information about specific patients can be gathered from the multiple health providers' records. Conventional e-healthcare solution technologies are built on centralized access regulation and are not adequately suitable for cooperative groups. The necessity of a decentralized strategy in the e-healthcare system stems from this centralized approach's difficulties. In this work, an integrated blockchain technology conceptual model of an e-hospital emergency management platform that correlates the Internet of things (IoT) is proposed. In Sect. 3, a critical examination of 50+ papers on blockchain-based health systems that have been published during 2015 and 2020 is presented. Thirty-six of which were articles, seven are all from conferences, four were from different symposiums, and three were from seminars, while one was a section from a book. Current e-healthcare system is trust less in collaborative environment. Three key questions will be covered after literature survey. Firstly, what are challenges that occur with traditional e-healthcare system in collaborative environment? Secondly, need of decentralized access control for digital e-healthcare system. Third, see challenges with established methods in the e-healthcare system and whether a blockchain-based model can be a fine replacement.

Keywords e-healthcare · Digital content · Interactive environment · Blockchain

D. Singla (✉) · S. K. Rana
Department of Computer Science and Engineering, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana, India
e-mail: singladeepak211@gmail.com

S. K. Rana
e-mail: dr.sanjeevrana@mmumullana.org

1 Introduction

E-health is a field of technology that is gaining importance throughout time, from individual real-time data interchange to faraway access to medical information like electronic health records (EHR) or electronic medical records (EMR) [1]. With the development of IoT and linked products, e-capacity health's to give patients access to their clinical data and real-time health monitoring throughout the globe is crucial. Significantly larger availability of healthcare, performance in remedies and health management, and less strain on community health budgets are all benefits of improved patient-health effective communication [2]. EHR is a consistent data that enables inclusion among various healthcare providers, and this integration is thought to be its main benefit [3]. EHR offers a number of advantages, including assisting with prescriptions, enhancing illness management, and reducing serious pharmaceutical errors. However, the accessibility of EHR, the confidentiality of data sent across healthcare organizations, and the exclusion of information on patient well-being are all constraints [4]. As a result, blockchain transforms the way clinical data is kept and communicated in the context of e-health by acting as a secure data and decentralized Internet platform of numerous computers referred to as nodes. It streamlines operations, keeps an eye on data security and correctness, and lowers maintenance costs.

As the blockchain technology is growing in the area of e-hospital rapidly, Table 1 tells the publications trend in e-hospital research since 2017–2021. Scope of blockchain is increasing day by day in healthcare domain to provide high security and trust in medical sector. Blockchain presents a chance to ensure the credibility of the data, the seamless transfer of data between the stakeholders, and the confidentiality of the customer. For the application of blockchain technology, trust issues like adoption, legislation, and morality should be taken into account. Medical care is necessary as much as there is modern society.

The production of health information, or patient history in general, has evolved from the initial paper medical records (PMRs) to the present electronic medical records as a result of the ongoing improvement of technology [5] (EMRs). Without a doubt, e-health systems make it simpler to save, exchange, collect, and trace EMRs [6]. Blockchain-based e-healthcare system is working as shown in Fig. 1.

The Internet of things (IoT) gadget [7] provides clinical information to patients and creates a supportive connection between faraway clients and medical service.

Table 1 Number of publication count	Publications year	Count
	2017	2
	2018	8
	2019	10
	2020	18
	2021	7
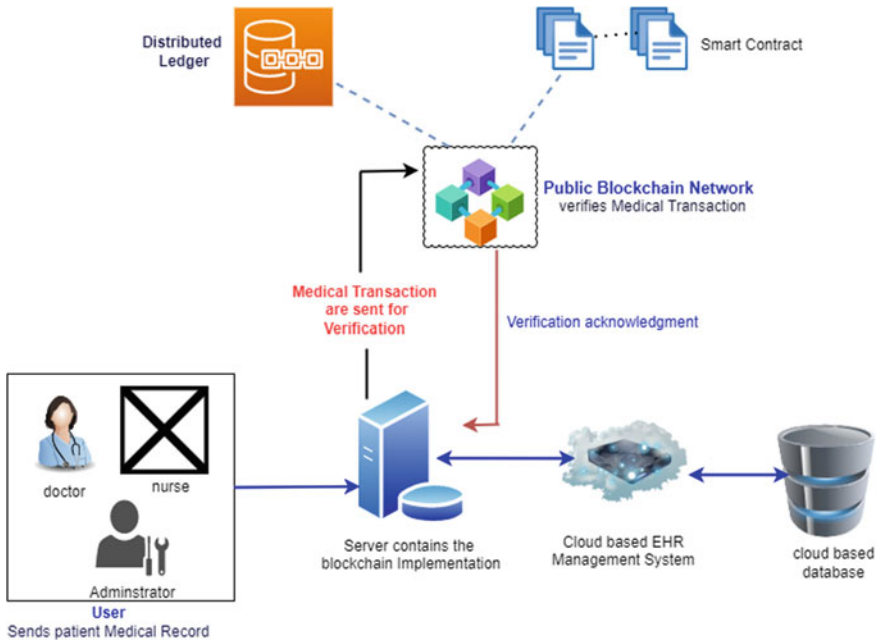
Fig. 1 Blockchain-based e-healthcare system

In this way, the patient's state is determined, and feedback is given via the routine care tracking. Interchange, machine-to-machine connection, data exchange, and data transfer are all related to healthcare in the IoT. Depending on social, economical, and health policy factors, availability of medical treatment may differ between countries, groups, and people. In order to achieve the optimum results, delivering healthcare services entails "the timely utilization of personal health services overall. The availability of medical records for the predetermined amount of time is made possible by the blockchain in healthcare [8]. By executing the information transfer to the distant patient, the blockchain technology plays a crucial role in healthcare. The blockchain used in this case is a distributed ledger system that integrates medical charts and facilitates communication to share services, improving efficiency monitoring and lowering service costs [9]. In this way, the blockchain maintains the global services in IoT and obtains the services by deploying the request [10]. The patient-requested solutions are retrieved from the blockchain-related information from the healthcare center and deliver the consequent value.

2   Literature Survey

Zheng et al. [1] analyze the various consensus algorithm used in blockchain and then discuss the various IoT-based solution in e-healthcare system. Then advantage and limitations of old and latest mechanism to provide security to IoT enable data using blockchain are explained.

Fan et al. [2] in this paper discussed the current condition of blockchain in healthcare system. In this analysis of poor scalable technique to improved scalable system using blockchain. The limitation of this paper is more focused on survey paper not the implementation of corresponding work and in future perform in depth using various blockchain-based metric.

Kamau et al. [3] have discussed about architecture of blockchain technology. Agreement procedure is introduced and also reviewed various applications. It also elaborated technical challenges faced in this technology.

Dubovitskaya et al. [4] contributed to support the protection of big data sources with the help of access control based on the concept of the blockchain. It also outlines the mechanism for access control framework using blockchain.

Omar et al. [5] proposed a framework to achieve supply chain management in blockchain. This paper designed about how the supply chain management work using blockchain in various collaborative environment. This paper faces issue about the implementation of supply chain using blockchain.

Omar et al. [6] in this paper told about the various issues in blockchain-based healthcare system like data breaches, data theft, and data direction. The solution of these problems is given in terms of private.

Premarathne et al. [7] present a framework to apply access control policies for safe shared authentication using blockchain to maintain privacy and security. Researcher can use blockchain and access policies together to answer the existing problems.

Uddin et al. [8] the healthcare procedure issues that can be obtained from various blockchain-based approaches were specifically covered in this research and analyzed accordingly.

Radanović and Likić [9] explained the public blockchain that uses sharing known as Rapidchain which can protect from complex faults. It divided the data and stored them different nodes of the collaborative environment. It can also prevent from gossiping transactions by some efficient authentication method.

Zhuang et al. [10] proposed an access control method that uses decentralized capacity to safeguard the resources like tools, services, and data in system of Internet of things. It incorporates the basic concept of blockchain.

Yánez et al. [11] proposed multi-organizational verification system that used blockchain as its base. Common ledger and unbreakable properties of blockchain are incorporated in the above proposed system. Hash chain concept is employed to maintain the transparency in the proposed system that uses one time password.

Li et al. [12] investigated Internet of things and find out an issue related to access control. It also proposed a structure that uses smart contract. Smart contracts contains

many access control contracts like judge contract, register contract to obtain trust in collaborative environment in Internet of things.

Dziembowski et al. [13] presented an application that restricts the access to physical locations. This application is based on blockchain technology and hyperledger concept. Components of this application are elaborated. In the last, the efficiency is measured using performance metrices and assets consumed.

Bentov et al. [14] presented an approach that gives the answer to big data protection. The privacy challenges are managed by hyperledger. Users are not allowed to share their information other than predefined roles. Finally, in this approach partial transparency is selected that influence the trustworthiness of the system.

Mertz [15] demonstrated the access control policies categorized into general policies and special policies. The public database can be used to store general policies and private blockchain can be used to store special policies. These policies are evaluated by smart contract and results in recognition. This approach can be used in smart traffic problem.

Hölbl et al. [16] introduced a technique to verify the identity of user by using blockchain computers with a decentralized access control and attach aspect of time to file sharing. A validity time was also introduced to access. They introduced a validity time to the access permission without extra revocation expenditure. Implementation for the same was also discussed.

Bach et al. [17] discussed a model that will allow data sharing among cloud with the help of blockchain. Smart agreements' advantages with access control mechanisms are used to monitor information exchange and revoke permission in cases of rule violations. This model was developed for the environment where trust between the communicating parties is less.

Hussien et al. [18] developed the Fairaccess network methodology. In this approach, the FairAccess access control architecture, which makes utilization of the blockchain's stability, is used for the first time. This model provides the control to resource possessor. A scripting language is used to evaluate the guidelines.

Kombe et al. [19] proposed a framework to maintain the privacy of user summary by using blockchain technology. This model was implemented by a blockchain network that uses multichain that allows managing the privacy of user profile. Only trusted parties can share the data with this model. But user is not able to know who see their information in the organization.

Zhang and Poslad [20] proposed a model where IoT devices are used to collect the data from the remote areas because it is very difficult to collect.

Benchoufi and Ravaud [21] developed an integrated healthcare system based on blockchain in which clients' health data (HD) are gathered by an unmanned aerial vehicle (UAV) and kept on the closest server. The viability of the suggested protected framework is assessed through a security analysis. On the UAV, simulations are run to examine the effects of the suggested authorization system.

Siyal et al. [22], this paper introduced a cloud computing based technology to push medical information on to cloud.

Zhou et al. [23], The existing EHR system's security and confidentiality concerns restrict the information that may be provided on specific patients, as discussed in this

paper. They recommended a new platform based on blockchain termed MedBlock to govern the client database held in EHRs in order to solve the above problem.

3 Review Design

The confidentiality of Internet of things (IoT) devices, which is generally viewed to encompass mobile devices as well, and the supporting health records network, is one of the most significant concerns in the digitized and ubiquity healthcare. A sizable volume of medical data is gathered and exchanged hourly, daily, weekly, and so forth [11] because to the development of IoT and the prevalence of healthcare devices and applications (apps). Interoperability, lengthy procedures, delays in diagnosis and treatment, information-sharing difficulties, excessive operating expenses, lengthy insurance processing times, and governance, transparency, and security issues are just a few of the difficulties facing current healthcare systems. The performance for the healthcare sector to adapt to the blockchain revolution has been built. It is a leader in its field [24]. The medical environment places a high priority on patients, and one of their main objectives is to increase the security, privacy, and interchange of health data. The health sector generates enormous amounts of data, such as client clinical notes, information about clinical trials, payment, medical research, and many others. Adoption of IoT and mobilization are changing over time. The access administration of health records is another key issue in a centralized structure, making blockchain healthcare technology a major challenge for all industries and researchers [12]. Blockchain technology can be utilized to get around third parties hiding patient personal information and medical data [13]. Blockchain-based medical practices significantly increase the volume of data available to track patients' medical records in order to assure their safety, privacy, and dependability [14]. With these techniques, it can also be useful for medical record to be interoperable between different organizations. Electronic health files should not contain any additional potential dangers. Patient lives and jobs may be affected by the release of personal health information.

4 Limitations of Traditional Authorization Models in Collaborative Environment

The persons participating inside a collaborative setting may not have known one another before the interaction. In light of numerous technical issues, this section presents ideas put forth in earlier studies on e-healthcare in collaborative settings. Goals for non-repudiation, traceability, and verifiability are only a few of the concerns that should be addressed by the present set of healthcare system security objectives. Other objectives include security, reliability, accessibility, privacy, and goals for legitimacy and credibility. When trying to address the security procedures of the

healthcare information system, some studies have tried to address all of these necessities, while others have concentrated on a single problem. If we utilize conventional access control frameworks for the identical use case, those certain frameworks suffer from various issues [15], as listed below:

1. **Problem of medical data security**: The confidentiality of patient records contained in EHR for patients or health insurers is the first sort of technical issue that has been raised in a number of researches. According to studies, the current EHR storage system in e-healthcare is contingent on reliable external parties (like cloud servers). Secondly, IoT-RPM has generated a number of privacy and security concerns due to the inability of the IoT server architecture to completely disrupt the network and the devices' inherent susceptibility to DDoS attack, information theft, trying to hack, and virtual hijacking [16].

2. **Problem of medical data privacy**: These issues deal with the worries about the preservation of clinical data security for patients and healthcare professionals in EHR. Issues with cyberattacks compromising the privacy of medical data kept in the cloud, including such lack of responsibility and pseudonymity [17].

3. **Problem of medical data integrity**: The sort of issues pertain to data security and the accessibility of clinical information that has been saved. Making a connectivity control plan and safeguarding medical information with the patient's key are the two techniques for achieving integrity in the current system [18]. The drawback of the first approach is the vulnerability to internal database data modification and deletion. The issue with the second approach is that if the patient passes away during the course of the diagnostic or therapy, it is impossible to share the key. As a result, there are two issues that affect data availability.

4. **Problem of access control in medical system**: Since unapproved links can be made when customer data is transferred from one vendor to another, this type of issue can arise [19].

5. **Problem of medical data interoperability**: This issue arises from the gathering and processing of clinical information, which results in a lack of accessibility in e-healthcare [20]. The demand for high-speed data sharing poses a difficulty to interchange among different providers since it causes record data to become fragmented rather than unified. Client privacy is at stake due to the possibility of hostile acts that could significantly harm status and revenue.

6. **Problem of managing the massive volume of patient data**: Due to their size and complexity, patient records can have concerns with their quality, including difficult analysis, diagnosis, and forecasting, but also their security due to the steadily rising incidence of cybercrimes [21].

5   Need of Decentralized Authorization Model
for e-Healthcare System

All of the centralized classical permission approaches that were covered above. Using centralized methods is usually plagued by problems like single points of failure and lack of confidence. Centralized approaches are not appropriate in collaborative settings since trust issues would always arise. Consequently, a decentralized strategy is needed to engage in a cooperative atmosphere. This decentralized strategy would eliminate the problem of trust and eliminates the risk by removing the necessity of middleman. Let's take the pharmaceutical medicine supply chain as an example [22]. Every nation is presently addressing the threat posed by the sale of counterfeit medications, which is a worldwide problem. It is a growing global issue that has a significant influence on low low- and middle-income countries. According to current estimates from the WHO, one in ten of the medications circulating in low- and middle-income nations are either subpar or fake. According to the National Drug Survey 2014–2016, carried out by the National Institute of Biologics, Ministry of Health and Family Welfare [23], counterfeit or substandard drugs make up about 3% of all pharmaceuticals sold in India. Most of the time, it is discovered that medications obtained straight from the production facility are reliable, whereas the chance of receiving substandard medications increases as soon as the items are transferred between the multiple phases and layers of the intricate supply chain (i.e., wholesalers, distributors, or sub-distributors). Drugs are susceptible to theft, adulteration, and replacement at every point of transit from the factory to the patient. Such misconduct results in financial loss for the medication manufacturers and, more crucially, poses a serious danger to consumer safety [25]. Internal systems automatically placed each activity into the ledger as the pharmaceutical drug [26] travelled through the supply chain. This was done to assure the manufacturer's confidentiality and safety. A significant amount of linked data can be presented to the consumers without endangering the information integrity thanks to decentralization, encryption techniques, and irreversible maintaining records. Even the processing inputs, such as adjuvants and pharmaceutically active components, were recorded and connected to the finished pharmaceutical goods. Further, the blockchain recorded crucial information from IoT sensors attached to the parcels, such as latitude and temperatures, enabling the journey available to all partners and reducing the risk of document forgery [27]. The ideal case of blockchain implementation is shown in Fig. 2.

6   Blockchain: The Imminent Approach in e-Healthcare

A blockchain is a decentralized ledger that is kept over a collection of blocks that are linked to one another using encrypted hashes [28]. These blocks are used to securely and openly hold all ledger transactions carried out across a peer-to-peer network. The ledger contains all of the details of an action, from its inception to its current state.
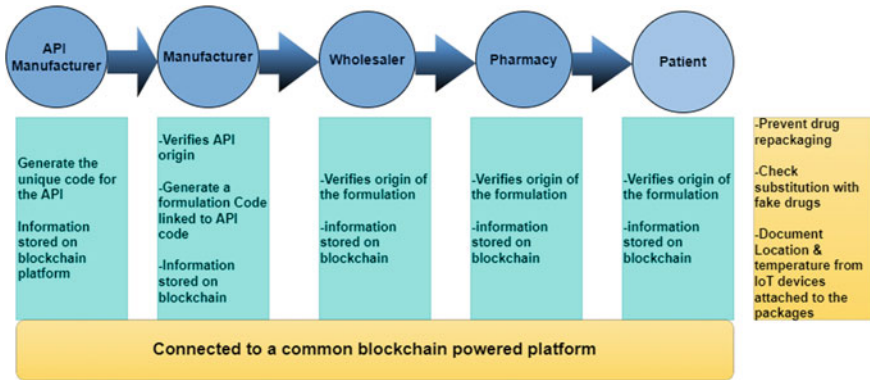
Fig. 2 Ideal case blockchain implementation

Every node in the network can see and affect every transaction, which is irreversible. Public (permission-less) or private (permissioned) ledgers are two different types of blockchain designs. Anyone can participate in a public blockchain without needing permission, and users can remain anonymous to one another. Before linking to a private blockchain, a company must obtain some sort of authorization.

6.1 Blockchain Features for e-Healthcare

1. **Decentralization**: Distributed innovation allows you the ability to preserve your resources (such as contracts and documents), making it possible for you to retrieve them via the Web. Here, the owner has entire authority over his account, which gives him the authority to transmit his resources to anybody he chooses [29].
2. **Transparency**: A blockchain's openness originates from the fact that every public address's holdings and transactions are visible to everyone.
3. **Immutability**: The information cannot be changed once it has been entered into the ledger. Regardless of who you are, you cannot alter the situation [30]. In order to correct an error, a payment update must be generated.

6.2 Types of Blockchain

The connectivity of nodes that operate on a system for transactional or authentication purposes is called to as blockchain architecture. When a framework is accessible to the public, any entity node can be an associate of the network; therefore, this blockchain is found in public, such as Ethereum and bitcoin [31]. If the participants of the participating nodes in blockchain have become identified to the network, therefore such a blockchain is regarded to as order to make this research, such as

hyperledger fabric and ripple. Users are enabled to instantly perform and verify the transfer of funds without such a centralized government thanks to the database systems of blockchain. This enables the construction and exchange of a digital ledger for distributed transactions between P2P networks of nodes. Due to the fact that this process should only be carried out once in a central place, this decentralized approach considerably lowers the cost of system configuration, maintenance, change, and adjudication in communication. Despite having tremendous performance in many circumstances, this kind of technology would only experience a single failure or a very small number of failures, and it would have scalability issues.

6.2.1 Permission-less or Public Blockchains

When nodes that belong to the network are reachable by anybody online, the ledger is referred to as a public blockchain (such as Bitcoin or Ethereum). Consequently, any activity uses can verify a transaction and contribute in the authorization through the consensus mechanism, including proof of work [32] or stake evidence. A blockchain's main purpose is to safely reduce centralized control during the exchange of digital assets. P2P transactions create a block of chains to guarantee decentralization. Before being added to the system's irreversible database, each transaction is connected to the preceding it using the cryptographic hash Merkle tree as a piece of the chains. The blockchain-based ledger is thus coordinated and compatible with each access point. A comprehensive blockchain ledger will be made available to anyone who has computer and a connection to the Internet who registers to become a station. The technology is entirely safe due to the duplication of synchronized public blockchains among each station in the network [33]. However, the operation of validating activities for this kind of blockchain has been slow and ineffective. The amount of electrical power required to validate each transaction is enormous, and as more nodes are added to the network, this power should grow dramatically.

6.2.2 Permissioned or Private Blockchains

This particular class of constrained blockchains enables a facilitator to be fairly compensated. Connectivity authorization is strictly controlled in private blockchains [34]. Without authorization, stations in the P2P network cannot take part in checking and authenticating transactions. Instead, the broadcaster's transactions can only be verified and validated by businesses or organizations.

6.3 IoT-Based Blockchain for e-Healthcare

The Internet of things (IoT) is the expansion of network connectivity into real-world equipment and ordinary items. These gadgets can communicate and collaborate with

others via the network and can be tracked and managed from a distance since they are integrated with transistors, Internet connectivity, and other types of technology [35]. Let us talk about the flaws in current approaches and the potential blockchain has for the IoT-based e-healthcare paradigm.

The majority of distributed systems now in use are focused on the server-client model, in which all equipment are recognized, verified, and associated via cloud servers, necessitating a huge amount of computing resources and storage space. Additionally, even if the gadgets are close to one another, all conversations between them must occur over the Internet. Despite being useful for small IoT networks, this strategy does not scale effectively. For substantial IoT networks, it is also expensive to set up a large number of communication links, operate centralized clouds, and connect all of the equipment. The majority of the IoT networks in use today are built on the server-client model. In addition to the expenditures, the architecture is vulnerable to a singular failure point due to its dependence on cloud servers. Additionally, IoT devices need to be resistant to physical and informational hacking. Despite the fact that some of the currently used techniques help to encrypt IoT devices, they are complicated and unsuitable for devices with low resources and processing power.

By developing a peer-to-peer network [36], blockchain reduces the price of setting up and maintaining centralized clouds, data centers, and networking hardware by dispersing the computing and storage demands among all the platform's devices. The single-point-of-failure issue is fixed by this communication network. By utilizing cryptographic techniques, blockchain addresses the privacy issues for IoT networks. Additionally, it uses tamper-resistant ledgers to address the reliability problems in IoT networks. IoT devices are work with blockchain is mention in Fig. 3.
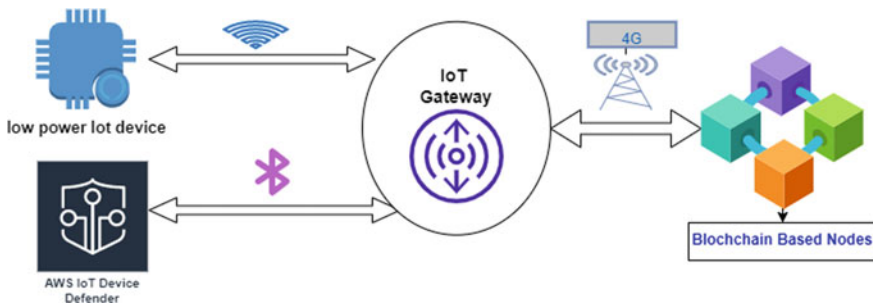

Fig. 3 IoT connectivity for blockchain e-healthcare

7   Blockchain Consensus Algorithm in e-Healthcare

1. **Proof of work (PoW)**: According to information published in a 1993 journal paper, the idea was created by C. Dwork and M. Naor. There is a moderate likelihood that PoW will be included into e-healthcare services because PoW is widely employed across numerous platforms [37].

2. **Proof of stake (PoS)**: Nodes are selected by lottery to extract the following block and solve it. It is incredibly democratic. There is no mining incentive or coin production; instead, rewards consist entirely of transaction fees [38]. A node may receive a reward in the form of a transaction fee as a result of the nothing at stake dilemma. We assume that using it as a viable solution for e-healthcare application submissions.

3. **Delegated proof of stake (DPoS)**: It is democratically representative. Although it speeds up transactions, centralization costs more [39]. There is a mechanism in place for identifying and voting out rogue delegates. As a result, it has a high likelihood of application in e-healthcare contexts.

4. **Leased proof of stake (LPoS)**: It resolves the PoS dominance issue, supports low-balance networks and the leasing contract, and distributes the incentive according to wealth [40].

5. **Proof of importance (PoI)**: Over PoS, it is an upgrade. It takes into account the reputation and balance of nodes [41]. It is a network that is more effective. We advise using it for online healthcare services since patients may use a doctor's reputation to help them make a decision.

6. **Practical byzantine fault tolerance (PBFT)**: This concept is comparable to PoW and PoS and better suited for private blockchains [41]. It does not tolerate infection very well. We would advise it to increase the utilization of e-health services.

Because they meet all the criteria for a digital asset ledger, such as integrity, attribution, and visibility blockchain-based ledgers can be utilized as a substitute for centralized digital asset management [42]. Table 2 shows a comparative analysis of traditional authorization models with blockchain-based models.

Some major challenges occur when blockchain is working for healthcare system. Figure 4 shows the some of major challenges that comes out after studies various research paper between the year 2017 and 2021.The study shows that the total 19% of study is related to challenge in data confidentiality of healthcare system that uses blockchain. Data integrity and data availability is the next major concern in healthcare domain using blockchain.

Table 2 Comparison of blockchain-based models vs conventional models

Features/ approach	Traditional models	Blockchain-based model
Confidentiality	✓	✗
Integrity	✓	✓
Immutability	✗	✓
Availability	✓	✓
Transparency	✗	✓
Trust	✗	✓
Single point of failure	✓	✗
Consensus	✗	✓
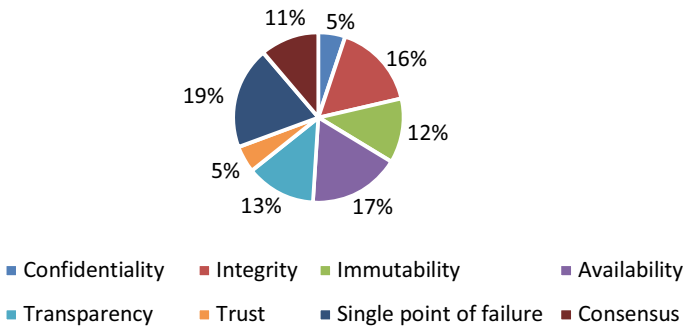
Fig. 4 Challenges occurs with blockchain in healthcare

8 Conclusion and Future Scope

In this work, problems with conventional permission models for usage in collaborative settings in e-healthcare systems are highlighted. The necessity for a decentralized strategy for a community setting was then investigated. The blockchain can address a number of complex problems in e-healthcare thanks to its immutable processes, cryptography methods, and decentralization mechanism. Peer-to-peer transactions using the blockchain system are significantly less expensive. Protecting the privacy of all Internet-enabled IoT devices is the biggest challenge facing the medical industry. The adoption of new technology by the health industry to build technological infrastructure is a key factor in determining blockchain's potential in the field of healthcare. Utilizing this innovation provides the ability to link disparate systems and generate ideas while improving the value of care. Long-term efficiency gains and enhanced outcomes for patients are possible with a national blockchain EMR network. The

improvement of insurance claim or other administrative processes inside the pharmaceutical systems, as well as the availability of survey data for scientists and clinicians, are other facets of medical care that the distributed ledger technology may assist with. A variety of industrial fields are implementing blockchain to expedite procedures, boost security, and improve efficiency. Because of this technology's shown value, businesses in the financial, energy, supply chain, healthcare, and manufacturing sectors are implementing it. As a result, a lot of new feature has been created to increase the blockchain technology's usefulness in various fields. This technology is only now being used by researchers to tackle various e-healthcare concerns in a group setting. With the use of this technology, numerous issues related to collaborative environments can be solved. Therefore, our main focus will be on using blockchain-based business models to solve challenges that cannot be handled via conventional methods.

Conflict of Interest The author declares that there is no conflict of interest.

References

1. Zheng Z, Xie S, Dai H-N, Chen X, Wang H (2018) Blockchain challenges and opportunities: a survey. Int J Web Grid Serv 14(4):352–375
2. Fan K, Wang S, Ren Y, Li H, Yang Y (2018) Medblock: efficient and secure medical data sharing via blockchain. J Med Syst 42:1–11
3. Kamau G, Boore C, Maina E, Njenga S (2018) Blockchain technology: is this the solution to EMR interoperability and security issues in developing countries? In: 2018 IST-Africa week conference, pp 1–8
4. Dubovitskaya A, Xu Z, Ryu S, Schumacher M, Wang F (2017) Secure and trustable electronic medical records sharing using blockchain. In: Proceedings of AMIA annual symposium, p 650
5. Omar AA, Rahman MS, Basu A, Kiyomoto S (2017) Medibchain: a blockchain based privacy pre-serving platform for healthcare data. In: International conference on security, privacy and anonymity in computation, communication and storage. Springer, 2017, pp 534–543
6. Omar AA, Rahman MS, Basu A, Kiyomoto S (2019) Privacy-friendly platform for healthcare data in cloud based on blockchain environment. Future Gener Comput Syst 95:511–521
7. Premarathne U, Abuadbba A, Alabdulatif A, Khalil I, Tari Z, Zomaya A, Buyya R (2016) Hybrid crypto graphic access control for cloud-based HER systems. IEEE Cloud Comput 3(4):58–64
8. Uddin MA, Stranieri A, Gondal I, Balasubramanian V (2020) Blockchain leveraged decentralized IoT eHealth framework. Internet Things 9:100159
9. Radanović I, Likić R (2018) Opportunities for use of blockchain technology in medicine. Appl Health Econ Health Policy 16(5):583–590
10. Zhuang Y, Sheets LR, Chen YW, Shae ZY, Tsai JJ, Shyu CR (2020) A patient-centric health information exchange framework using blockchain technology. IEEE J Biomed Health Inform 24(8):2169–2176
11. Yánez W, Mahmud R, Bahsoon R, Zhang Y, Buyya R (2020) Data allocation mechanism for Internet-of-Things systems with blockchain. IEEE Internet Things J 7(4):3509–3522

12. Li W, Andreina S, Bohli JM, Karame G (2017) Securing proof-of-stake blockchain protocols. In: Data privacy management, cryptocurrencies and blockchain technology. Springer, New York, NY, USA, pp 297–315

13. Dziembowski S, Faust S, Kolmogorov V, Pietrzak K (2015) Proofs of space. In: Proceedings of annual cryptology conference, pp 585–605

14. Bentov I, Lee C, Mizrahi A, Rosenfeld M (2014) Proof of activity: extending bitcoin's proof of work via proof of stake. ACMSIGMETRICS Perform Eval Rev 42(3):34–37

15. Mertz L (2018) (Block) chain reaction: a blockchain revolution sweeps into health care, offering the possibility for a much-needed data solution. IEEE Pulse 9(3):4–7

16. Hölbl M, Kompara M, Kamišalić A, Nemec Zlatolas L (2018) A systematic review of the use of blockchain in healthcare. Symmetry 10(10):470

17. Bach LM, Mihaljevic B, Zagar M (2018) Comparative analysis of blockchain consensus algorithms. In: Proceedings of 41st international convention on information and communication technology, electronics and microelectronics (MIPRO), May 2018, pp 1545–1550

18. Hussien HM, Yasin SM, Udzir SNI, Zaidan AA, Zaidan BB (2019) A systematic review for enabling of develop a blockchain technology in healthcare application: taxonomy, substantially analysis, motivations, challenges, recommendations and future direction. J Med Syst 43(10):320

19. Kombe C, Ally M, Sam A (2018) A review on healthcare information systems and consensus protocols in blockchain technology. Int J Adv Technol Eng Explor 5(49):473–483

20. Zhang X, Poslad S (2018) Blockchain support for flexible queries with granular access control to electronic medical records (EMR). In: Proceedings of IEEE International conference on communications (ICC), May 2018, pp.1–6.

21. Benchoufi M, Ravaud P (2017) Blockchain technology for improving clinical research quality. Trials 18(1):1–5

22. Siyal AA, Junejo AZ, Zawish M, Ahmed K, Khalil A, Soursou G (2019) Applications of blockchain technology in medicine and healthcare: challenges and future perspectives. Cryptography 3(1):3

23. Zhou L, Wang L, Sun Y (2018) MIStore: a blockchain-based medical insurance storage system. J Med Syst 42(8):149

24. Grossman R, Qin X, Lifka D (1993) A proof-of-concept implementation interfacing an object manager with a hierarchical storage system. In: Proceedings of 12th IEEE symposium on mass storage systems, pp 209–213

25. Zhang P, Schmidt DC, White J, Lenz G (2018) Blockchain technology use cases in healthcare. In: Advances in computers, vol 111. Elsevier, Amsterdam, The Netherlands, pp 1–41

26. Angeletti F, Chatzigiannakis I, Vitaletti A (2017) Privacy preserving data management in recruiting participants for digital clinical trials. In: Proceedings of the first international workshop on human-centered sensing, networking, and systems, pp 7–12

27. Clim A, Zota RD, Constantinescu R (2019) Data exchanges based on blockchain in m-Health applications. Procedia Comput Sci 160:281–288

28. Ali F, Islam SR, Kwak D, Khan P, Ullah N, Yoo SJ, Kwak KS (2018) Type-2 fuzzy ontology–aided recommendation systems for IoT–based healthcare. Comput Commun 119:138–155

29. Aktas F, Ceken C, Erdemli YE (2018) IoT-based healthcare frame-work for biomedical applications. J Med Biol Eng 38(6):966–979

30. Aladwani T (2019) Scheduling IoT healthcare tasks in fog computing based on their importance. Procedia Comput Sci 163:560–569

31. Prikhodko P, Izumchenko E, Aliper A, Romantsov K, Zhebrak A (2018) Converging blockchain and next-generation artificial intelligence technologies to decentralize and accelerate biomedical research and health-care. Oncotarget 9(5):5665

32. Kwak K-S (2021) An intelligent healthcare monitoring framework using wearable sensors and social networking data. Future Gener Comput Syst 114:23–43

33. Zhang P, White J, Schmidt DC, Lenz G (2017) Design of blockchain-based apps using familiar software patterns to address interoperability challenges in healthcare. In: Proceedings of 24th pattern languages of programs, Ottawa, ON, Canada

34. Roehrs A, da Costa CA, da Rosa Righi R (2017) OmniPHR: a distributed architecture model to integrate personal health records. J Biomed Inform 71:70–81

35. Genestier P et al (2017) Blockchain for consent management in the ehealth environment: a nugget for privacy and security challenges. J Int Soc Telemed eHealth 5:1–4

36. Rana SK, Rana SK, Nisar K, Ag Ibrahim AA, Rana AK, Goyal N, Chawla P (2022) Blockchain technology and artificial intelligence based decentralized access control model to enable secure interoperability for healthcare. Sustainability 14:9471. https://doi.org/10.3390/su14159471

37. Kim JY (2018) A comparative study of block chain: bitcoin namecoin medibloc. J Sci Technol Stud 18(3):217–255

38. Rana SK, Rana SK (2020) Blockchain based business model for digital assets management in trust less collaborative environment. J Crit Rev 7(19):738–750

39. Xia Q, Sifah EB, Asamoah KO, Gao J, Du X, Guizani M (2017) MeD- share: trust-less medical data sharing among cloud service providers via blockchain. IEEE Access 5:14757–14767

40. Aileni RM, Suciu G (2020) IoMT: a blockchain perspective. In: Decentralised internet of things. Springer, Berlin, Germany, pp 199–215

41. Rana SK, Kim H-C, Pani SK, Rana SK, Joo M-I, Rana AK, Aich S (2021) Blockchain-based model to improve the performance of the next-generation digital supply chain. Sustainability 13:10008. https://doi.org/10.3390/su131810008

42. Rana SK, Rana SK (2021) Intelligent amalgamation of blockchain technology with industry 4.0 to improve security. In: internet of things. CRC Press, 1st ed, ISBN: 9781003140443

Recognition of Hypertension Through Retinal Fundus Images by Segmentation of Blood Vasculature

Jaskirat Kaur, Deepti Mittal, Parveen Singla, Ramanpreet Kaur, and Sharad Kumar Tiwari

Abstract Condition of high blood pressure is termed as hypertension. Hypertension also causes variations in the blood vessels of retina which forms as a base to identify hypertension at a very early stage. These variations occur as narrowing of arteries, bifurcation variations and may lead to torturous blood vessels. The present paper introduces an automated method of vessel segmentation based on top-hat filtering and slicing of bit-planes on which measurements like arteriovenular ratio (AVR), bifurcation points and tortuosity index are calculated. The performance of the designed retinal vasculature identification and detection of bifurcations are evaluated on two open-source datasets, namely: DRIVE and STARE. DRIVE dataset achieved the mean accuracy of 94.89% and STARE dataset achieved the mean accuracy of 94.56%. Arteriovenous ratio of the retinal vessels is computed on a standard dataset INSPIRE-AVR with the aid of two medical specialists. Specialist 1 reported the mean accuracy of 98.43 and specialist 2 reported the mean accuracy of 98.43%.

J. Kaur (✉)
Department of Electronics and Communication Engineering, Punjab Engineering College (Deemed to be University), Sector 12, Chandigarh 160012, India
e-mail: jaskiratkaur17@gmail.com; jaskiratkaur@pec.edu.in

D. Mittal
Electrical and Instrumentation Engineering Department, Thapar Institute of Engineering and Technology, Patiala 147004, India

P. Singla
Electronics and Communication Engineering Department, Department of Research and Development, Chandigarh Engineering College, Mohali 140307, India
e-mail: parveen.rise@cgc.edu.in

R. Kaur
Electronics and Communication Engineering Department, Chandigarh Engineering College, Mohali 140307, India

S. K. Tiwari
Electronics and Communication Engineering Department, Vignan Foundation for Science and Technology, Vadlamudi, India
e-mail: drskt_ece@vignan.ac.in

Also, retinal vessels curve and tortuosity-based computations are made using the standard dataset, namely RETTORT dataset.

Keywords Arteriovenular ratio · Bifurcations · Hypertensive retinopathy · Region of interest · Tortuosity

1 Introduction

Hypertension is a chronic abnormality due to high blood pressure. High blood pressure results in abnormal variations in vasculature of retina. Retinal abnormality due to hypertension is the foremost cause of fatalities around the world [1, 2]. Retinal abnormities due to hypertension are not recognized until it reaches an advanced stage and it starts attacking the cardiac system. Thus, initial identification of retinal abnormality due to hypertension is necessary to avoid reaching it at chronic stage where it starts affecting other organs of the body. High blood pressure (systolic more than 140 mmHg and diastolic more than 90 mmHg) results in broadening of the vasculature boundaries and thinning of vessels [3]. This limits the flow of blood inside retina resulting in high pressure. This prolonged high blood pressure causes various vascular changes in retina which are termed as hypertensive retinopathy (HR). The several indications of retinal abnormality due to hypertension are narrowing of arteries, inflammation in the retinal optic disk, hemorrhages of varying shapes and sizes, microaneurysms, bright retinal lesions [4]. Figure 1 shows a normal retina and a pathological retina depicting indications of HR.

Numerous techniques were designed for the extraction of blood vasculature of retina. These techniques were primarily designed for non-specialists for the successful extraction of retinal blood vasculature. The detection and extraction
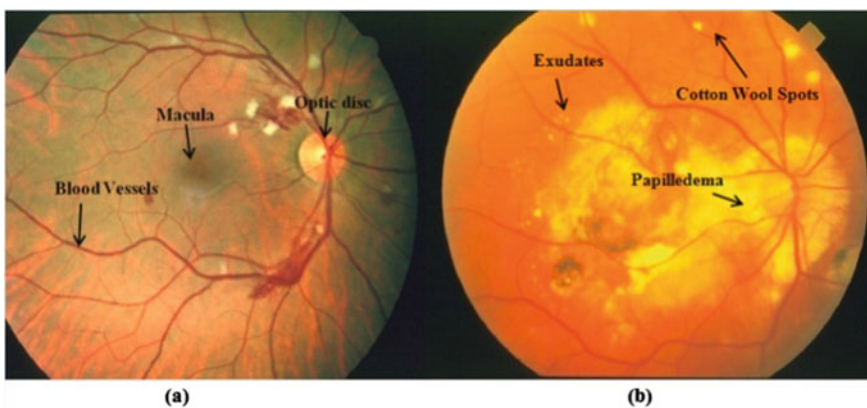

Fig. 1 **a** Image showing landmarks, **b** image depicting retinal abnormality related to hypertension

of vasculature of retina is important for evaluating the measurements of different attributes like retinal vasculature diameter, tortuosity-based indices and vessels junction points. The effectiveness of preventive protocols for HR can be administered only in the early stages of the retinal diseases. Therefore, these parameters are exploited for the initial detection of HR.

The retinal vasculature diameter is one of the prime attributes for the identification and leveling of retinal abnormality related to hypertension. It is usually expressed in terms of artery-to-vein ratio (AVR). The medical standards proposed by Keith, Wagener and Barker to classify and grade the retinal abnormality related to hypertension [5]. The quantitative measurement of AVR aid to ophthalmologist in the early detection of retinal abnormality related to hypertension. The grading of the retinal abnormality related to hypertension in terms of AVR is defined in four stages, namely normal, mild, moderate and severe. The range of AVR for various stages is detailed in [3]. The measurement of retinal geometrical structure is very important to make the quantitative assessment of retinal abnormality related to hypertension [6]. Therefore, the identification of vascular bifurcations, cross and end points are the initial steps in the detection of retinal abnormality related to hypertension [7]. The geometrical attributes of all linear segments in the concerned bifurcation points are calculated. These geometrical properties are necessary to form a strong correlation between the bifurcation measurements and hypertensive retinopathy. Figure 2 shows the image of normal retina with no instance of abnormality [8].

Tortuosity present in the retinal blood vasculature is another quantitative measurement for the detection of retinal abnormality related to hypertension [9]. Tortuosity arises due to abnormal elongation of the blood vasculature, and since the corners of the blood vasculature are static, thus increase in the length of the vessel results in twists and turns [10]. Figure 2b depicts a tortuous retinal image. The vessel tortuosity is measured largely by employing two methods. In the first method, tortuosity of the vessel segment was computed as the variations in the original path length and the straight-line length with respect to the straight-line length. Since this method is not sensitive toward curvature of the retinal vessel; therefore; other methods are

Fig. 2 **a** A sub-image from DRIVE dataset showing number of bifurcations, **b** original image from the RETTORT database showing tortuous vessels
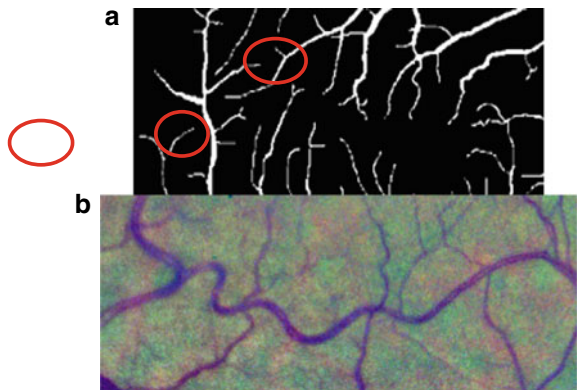
employed to measure tortuosity based on blood vasculature curvature [11, 12]. As described in Table 1, several algorithms are designed previously for the detection and extraction of vessels. These techniques are able to provide acceptable accuracy in few cases, but there requires an improvement in the proper extraction of blood vessels [13–45].

The major shortcoming of the current methods on vasculature extraction is the evaluation of techniques on few numbers of datasets and resulting in a not up to the mark performance in terms of sensitivity, specificity and accuracy. Therefore, the present work focuses on the designing of segmentation technique for obtaining the retinal vasculature on which various geometrical measurements are carried out for the detection of retinal abnormality related to hypertension. This method allows (a) automated segmentation of the retinal vessels; (b) computation of arteriovenous ratio for the identification and severity detection of hypertensive retinopathy; (c) identification of junction points and (d) tortuosity measurements. The proposed technique is robust in the sense that it has been tested on huge number of images as compared to other methods.

2 Proposed Method

The proposed method is designed in three steps, namely (i) preliminary processing using morphology-based filtering and slicing of bit-planes, (ii) identification of blood vasculature and optic disk, (iii) discrimination of arteries and veins and (iv) postprocessing. A new method for the identification of vasculature along with the measurements of various geometrical parameters for the early detection of retinal abnormality related to hypertension. Figure 3 details the steps involved in the proposed technique.

2.1 Preliminary Processing

In order to remove the noise from the background of retinal fundus images and to provide the clear visibility of retinal blood vasculature green channel with the highest contrast is extracted for further processing [46, 47]. This also aids in proper visibility of landmark structures of retina. In the next step, artifacts in retinal blood vasculature are eliminated [48, 49]. This is carried out by linear morphological transform. The transform is used in such a way that the artifacts present in retinal independent blood vessels are filled resulting in better contrast of the image with respect to the background. The resulting image is normalized between the fixed intensity variations for further processing [50]. Lastly, the vessels are enhanced by applying morphological top-hat transform in different bit planes.

Table 1 Literature of vasculature segmentation

Author	Year	Vessel segmentation method	Performance	Measures	Pros	Cons
Drive stare						
[13]	1989	2D Gaussian matched filter	Acc = 0.8773	–	Use of simple enhancement, thresholding operators	Large computational time
[14]	1999	Region growing algorithm	SN = 0.785 SP = 0.949 Acc = 0.932	–	Overcomes the problem of contrast inherent variations	Low sensitivity
[15]	2000	Matched filtering, threshold probing	Acc = 0.9267	SN = 0.6751 SP = 0.9567	Proposed new method	Used database of five images
[16]	2001	Top-hat transform, cross-curvature evaluation	SN = 0.6971 Acc = 0.9377	–	Can be used for image registration	Background structure remains noisy
[17]	2003	Adaptive local thresholding	SN = 0.659 SP = 0.96 Acc = 0.928	–	It can be applied to other lesion-based images	High time complexity
[18]	2004	Laplacian profile model	–	SN = 0.924 SP = 0.921	Aids in segmentation of vessels that split due to specular refection	Low specificity
[19]	2004	Morphological bit plane slicing	–	Acc = 0.9442 TPR = 0.7194 FPR = 0.0227	Used color features for segmentation	Poor results with vessel having light reflex
[20]	2006	DOG filter, Multiscale morphological reconstruction	SN = 0.7344 SP = 0.9764 Acc = 0.9452	SN = 0.6996 SP = 0.9730 Acc = 0.9440	Less time complexity	Unable to remove background noise
[21]	2007	Improved Gaussian matched filter	Acc = 0.9535	–	Improved responses for matched filter	Limited to find filter parameters

Table 1 (continued)

Author	Year	Vessel segmentation method	Performance	Measures	Pros	Cons
[22]	2007	Maximum principal curvature, gradient magnitude, region growing	SN = 0.7246 SP = 0.9655 Acc = 0.9344	SN = 0.6996 SP = 0.9569 Acc = 0.9410	Automated technique	Unable to eliminate abnormalities
[23]	2007	ITK serial implementation	SN = 0.660 SP = 0.9612 Acc = 0.9220	SN = 0.779 SP = 0.9409 Acc = 0.9240	Improved computational efficiency	Unable to handle high -esolution images
[24]	2007	Multiresolution Hermite model	SN = 0.780 SP = 0.978	SN = 0.752 SP = 0.980	New algorithm is proposed	No measurements performed
[25]	2007	Snake algorithm, topological properties	SN = 0.6634 SP = 0.9682 Acc = 0.9316	–	Good for localization of optic disk	Low sensitivity due to unsegmented thin vessels
[26]	2008	Top-hat transform, fuzzy clustering	–	Acc = 0.952	High accuracy and low misclassification	Low connectivity
[27]	2008	Scale space analysis parameter search	Acc = 0.9419	–	High quality and speed	Poor results for pathological images
[28]	2008	Multiscale line operator, region growing	–	Acc = 0.940	Found good estimate of vessel width	No fine vessels detected
[29]	2008	Divergence of vector fields	–	Acc = 0.9474	High accuracy	Did not discussed dark abnormality
[30]	2008	Snakes algorithm, morphological processing	SN = 0.7436 SP = 0.9615 Acc = 0.9352	–	Good results in arteriovenous structures	Low sensitivity
[31]	2009	Multiscale line tracking	SN = 0.747 SP = 0.955 Acc = 0.929	–	Fast tracking process	Misclassification of optic disk

Table 1 (continued)

Author	Year	Vessel segmentation method	Performance	Measures	Pros	Cons
[32]	2009	Ribbon of twin active contour model	SN = 0.7282 SP = 0.9551	SN = 0.7521 SP = 0.9681	Location of vessel segments is accurate	Low specificity
[33]	2010	Top-hat transformation, double-ring filter	SN = 0.793	–	Automated segmentation and classification	Segmentation method can be improved
[34]	2010	Maximum likelihood estimation of scale space parameters	–	SN = 0.7000 SP = 0.9530	Estimates vessel parameters directly from image rather than binarized image	Low true positive value
[35]	2010	Phase concurrency, log-Gabor filter	Acc = 0.92	–	Very low computational time of 10 s	Low accuracy
[36]	2010	Vessel tracing algorithm	SN = 0.744 SP = 0.966 Acc = 0.943	–	High accuracy for vessel segmentation and diameter estimation	Low sensitivity
[37]	2011	Curvelet transform, multistructure elements morphology by reconstruction	SN = 0.7352 SP = 0.9795 Acc = 0.9458	–	High sensitivity of multistructure elements provides better segmentation	Improper thresholding
[38]	2011	Top-hat transform, automatic thresholding	SP = 0.9788 SP = 0.7031 Acc = 0.9433	–	Automated method results in high specificity	Unable to detect narrow vessel extremities
[39]	2012	Gabor filter, morphological transformations	SN = 0.7224 SP = 0.9711 Acc = 0.9469	–	Used orientation analysis of gradient vector field and obtained good results	Poor segmentation due to bright central reflex
[40]	2013	Top-hat transform, shape-based morphology	TP = 0.6924 TN = 0.9779 Acc = 0.9413	TP = 0.7149 TN = 0.9749 Acc = 0.9471	Simple filtering and segmentation	Constrained vessel connectivity

(continued)

Table 1 (continued)

Author	Year	Vessel segmentation method	Performance	Measures	Pros	Cons
[41]	2013	Multiscale hierarchical, decomposition adaptive thresholding	Acc = 0.9461	Acc = 0.9501	Generates competitive results in terms of accuracy	Non-classification of arteries and veins
[42]	2014	Graph cut technique	TPR = 0.7512 FPR = 0.0316 Acc = 0.9412	TPR = 0.7887 FPR = 0.0367 Acc = 0.9441	Good results for vessel segmentation and optic disk detection	Suffers from overlapping tissue segmentation
[43]	2014	Self-adaptive matched filter	SP = 0.9579 SN = 0.7205 Acc = 0.9370	SP = 0.9586 SN = 0.6786 Acc = 0.9379	High accuracy	Low sensitivity
[44]	2015	Gaussian filtering, simple linear iterative clustering	SN = 0.8173 SP = 0.9733 Acc = 0.9767	SN = 0.8104 SP = 0.9791 Acc = 0.9813	Automated method	High computational time of 120 min
[45]	2016	Gabor wavelet filter, Firangi's filter	Acc = 0.9332	Acc = 0.8600 s	Better responses compared to single filter	False positive in areas close to border

2.2 Identification of Blood Vasculature and Optic Disk

Accurate and reliable segmentation is an essential step in determining the valuable information such as size and shape, which may help ophthalmologists in the diagnoses of retinal abnormality related to hypertension. Retinal blood vasculature is identified by applying repetitive thresholding technique [51].

This method is based on the Taylor series expansion where threshold values keep on varying resulting in the extraction of blood vessels. Figure 4 depicts the resulting retinal blood vasculature of the retinal image from the standard dataset. Precise identification of optic disk is important for the computation of various geometrical parameters of retinal abnormality related to hypertension. It appears as the brightest blob, circular in shape and largest anatomical structure with maximum intensity. In order to identify the optic disk location and boundary precisely, circular Hough transform is used [52]. Once optic disk is located, AVR is computed [53]. It is computed within the optic region of 1–1.5 disk diameter as most prominent vessels
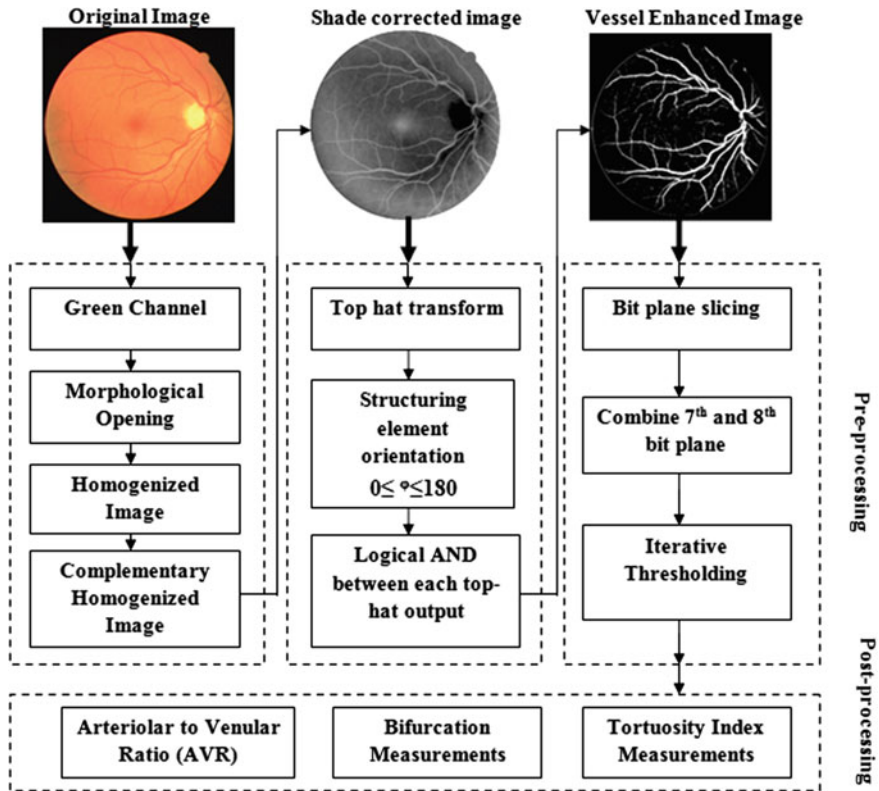
Fig. 3 Flowchart of the proposed methodology

are present in this region of the optic disk as [54, 55]. Afterward junction points of the retinal blood vasculatures were calculated on the standard open benchmark databases.

2.3 Discrimination of Arteries and Veins

In order to computer Arteriovenous ratio (AVR), it is important to precisely discriminate arteries and veins in retinal blood vasculature. Retinal blood vasculature can be discriminated as arteries and veins by visualizing its geometrical attributes. Generally, veins in the retina are dark red in color and thick as compared to arteries and arteries are thin [56]. Figure 5 depicts retinal vasculature arteries in blue and veins in blue separately.
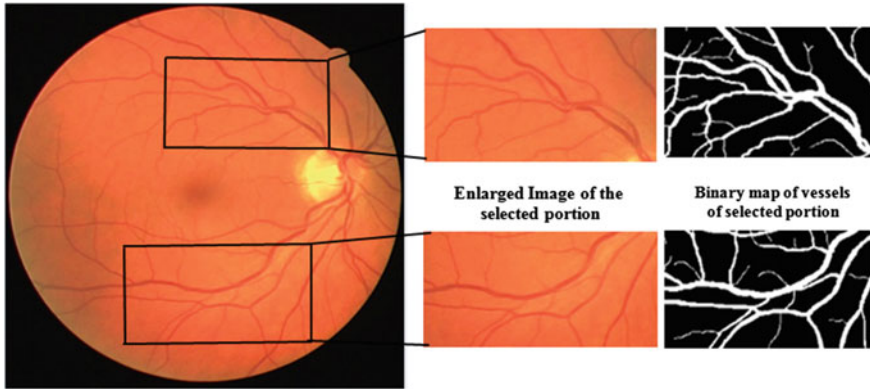
Fig. 4 Vessel segmentation: retinal image from DRIVE dataset and its binary vasculature mask
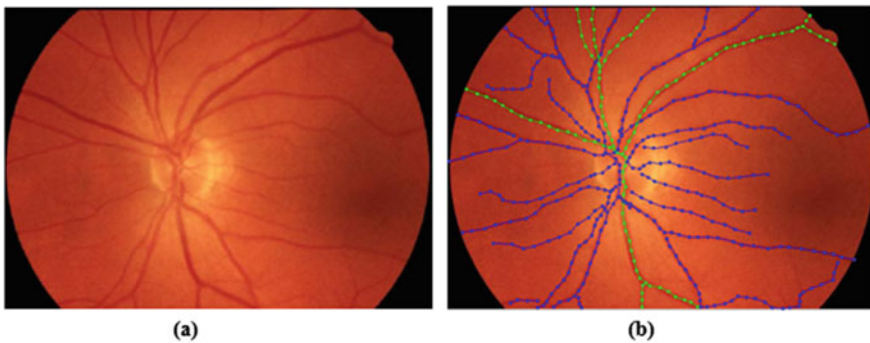

Fig. 5 Visualization of arteries and veins: **a** original image from STARE database; **b** discrimination of arteries and veins

2.4 Postprocessing

After the first three steps, a number of measurements on retinal vessels are computed; (i) calculation of arteriovenous ratio, (ii) identification of junction points and (iii) curvature-based measurements.

(i) Calculation of arteriovenous ratio

Arteriovenous ratio indicates the presence of narrowing of retinal blood vasculature arteries due to hypertension [4]. If the computed value of arteriovenous ratio is less than one then it signifies that retinal blood vasculature arteries are narrow as compared to retinal blood vasculature veins. Arteriovenous ratio is computed for the thickest arteries and veins present in retinal blood vasculature. It is calculated by using Central Artery of Retina (CAR) and Central Vein Retina (CVR) parameters

[57, 58]. CAR and CVR are computed on retinal arteries and veins, respectively. These values are computed by formulas derived by Parr [57, 59] and Hubbard [60].

(ii) Identification of Junction Points

Identification of junction points is utmost important for the detection of retinal abnormality related to hypertension. Variations in number of junction points signifies the progression of disease. In the first steps, all the possible junction points, crossings and ending points of the retinal blood vasculature are labeled [61]. This is done after eliminating the false points that might be detected. Junction points are the points having three pixels in the vicinity of the 3×3 matrix. End points are the points having only one pixel in the vicinity of the 3×3 matrix. Figure 6 shows the junction points and the ending points in the retinal image acquired from standard open-source benchmark database.

Geometrical attributes related to junction points are computed as they contribute highly in analyzing the blood pressure in the body. Murray [61] and Zamir [62] estimated the optimal values of these parameters for the optimum functioning of the system. Figure 7 depicts the parameters associated with junction points such as branch angle (θ), angle between parent branch and large daughter branch (θ_1), angle between parent branch and small daughter branch (θ_2) and blood vasculature diameters (d_0, d_1, d_2). The blood vasculature diameters d_0 is the diameter of the parent branch while d_1 and d_2 are the diameters of the large and small daughter branches, respectively. These retinal blood vasculatures computed are used to compute various junction points features related to parent branch and the daughter branches.
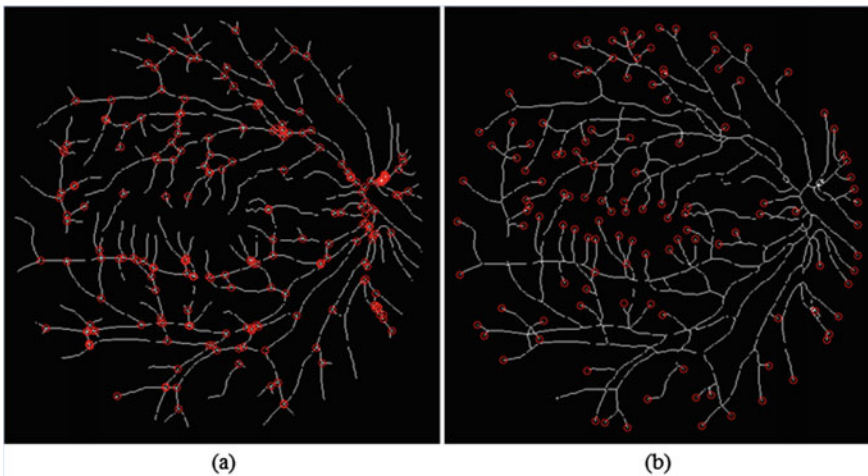

(a) (b)

Fig. 6 Significant points **a** junction points, **b** end points

Fig. 7 Diagrammatical
description of geometrical
attributes
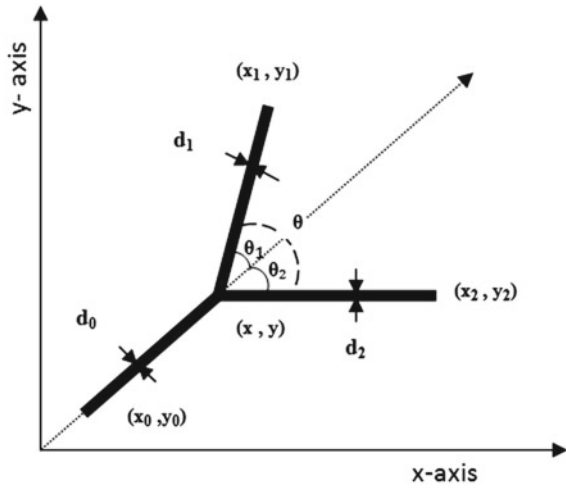
1. Junction Point Index (α): It computes the degree of symmetry and asymmetry in the junction points of the retinal blood vasculature [63]. Mathematically, it is calculated as

$$\alpha = \frac{d_2}{d_1}; \quad 0 < \alpha \leq 1 \tag{1}$$

2. Asymmetry factor (β): It also computes the degree of symmetry and asymmetry but in terms of cross-sectional areas

$$\beta = \frac{d_2^2}{d_1^2}; \quad 0 < \beta \leq 1 \tag{2}$$

3. Ratio of diameters (σ_1, σ_2): Computed using daughter branches (d_1, d_2) with the parent branch (d_0).

$$\sigma 1 = \frac{d_1}{d_0}, \sigma 2 = \frac{d_2}{d_0} \tag{3}$$

4. Area ratio (γ): It is mathematically computed by using the cross-sectional areas of the daughter and parent branches.

$$\gamma = \frac{d_1^2 + d_2^2}{d_0^2} \tag{4}$$

5. Junction exponent (z): It is computed using the repetitive method [63].

$$d_0^z = d_1^z + d_2^z \tag{5}$$

6. Branch angle (θ): It is defined as the angle formed by two daughter branches at each junction point.

$$\theta = \theta_1 + \theta_2 \tag{6}$$

(i) Curvature-based measurement

The curvature-based measurements signify the grading level of the retinal abnormality related to hypertension. Thus, it is necessary to analyze the changes in these measurements. The curvature-based measurement is evaluated on a curve $C = [x(t), y(t)]$ on a interval $[t_1, t_2]$ using RETTORT dataset using following formulas.

1. Ratio of arc length over chord length (k): It is mathematically computed using the formulas derived by [64–66].

The arc length $Lx(C)$ for the curve C is given by,

$$Lx(C) = \int_{t_1}^{t_2} \sqrt{x'(t)^2 + y'(t)^2}\,\mathrm{d}t \tag{7}$$

The cord length $Ly(C)$ of curve C is given by,

$$Ly(C) = \sqrt{[x(t_2) - x(t_1)]^2 + [y(t_2) - y(t_1)]^2} \tag{8}$$

Therefore, $k = \frac{Lx(C)}{Ly(C)}$ (9).

2. Parameters for curvature: In order to compute the variations in the blood vasculature orientations, absolute curvature *(tc)* and squared curvature *(tsc)* are used.

The curvature $k(t)$ of curve C at t is given by,

$$k(t) = \frac{x'(t)y''(t) - x''(t)y'(t)}{\left[y''(t)^2 + x''(t)^2\right]^{3/2}} \tag{10}$$

The absolute curvature $tc(C)$ of curve C is,

$$tc(C) = \int_{t_1}^{t_2} |k(t)|.\mathrm{d}t \tag{11}$$

The squared curvature $tsc(C)$ of curve C is,

$$tsc(C) = \int\limits_{t_1}^{t_2} \left|k(t)^2\right| . \mathrm{d}t \tag{12}$$

The value of above parameters is one if the blood vasculature is linear in shape and the value increases if more curvature is involved. Thus, the more curved the blood vasculature the larger is its curvature-based measurement. τ_1 is a basic dimensionless measurement used to determine the tortuosity index. Its use has been criticized as it may give a false representation of a vessel's tortuosity by underestimating its value in case of vessels having high number of curvatures. Nevertheless, any parameter from τ_1 to τ_8 can be used to measure curvature. The different variety of curvature-based measurements makes the comparison of various parameters easy.

3   Experimental Results

The designed method is assessed on images of retina taken from DRIVE, STARE, INSPIRE-AVR and RETTORT databases containing both healthy and pathology images [67–70]. The analysis of the designed method is carried out in four steps. The first step computes the retinal blood vasculature segmentation results on both DRIVE and STARE database; the second step is to evaluate the arteriovenous ratio on the selected ROI using INSPIRE-AVR database; the third step evaluates the geometrical measurements taken on total of 643 bifurcation points from both DRIVE and STARE datasets. The fourth step evaluates the curvature-based parameters on the RETTORT database.

3.1   Retinal Blood Vasculature Segmentation Results

The retinal blood vasculature segmentation results got by the proposed method for both healthy and abnormal images of retina from standard datasets DRIVE and STARE datasets obtained after taking complement of homogenized images. Figure 8d-f depicts the processed images of retina. It can be seen that the contrast between retinal blood vasculature and background is enhanced. The retinal blood vasculature present in each direction are enhanced further by using top-hat transform. Then, a sum of top-hat transformation is obtained for different directions rotated at 22.5° as shown in Fig. 8g–i respectively. The final blood vasculature extracted images in Fig. 8j–l are obtained by using repetitive thresholding technique corresponding to retinal images in Fig. 8g–i respectively. The average of the performance parameters like TPR, FPR, specificity, PPV and accuracy $\pm$ standard deviation for the DRIVE

database is 0.7293, 0.0298, 0.9701, 0.7084 and 0.9489 ± 0.0071 respectively. The performance parameters for the STARE database in terms of TPR, FPR, specificity, PPV and accuracy ± standard deviation are 0.7295, 0.0463, 0.9536, 0.6018 and 0.9356 ± 0.1539 respectively.

Tables 2 and 3 detail the results of the designed blood vasculature technique with DRIVE and STARE datasets presented in the literature. Tables 2 and 3 show that the designed technique exhibits comparable results in terms of sensitivity, specificity, PPV and accuracy than the other methods mentioned in the comparison along with non-expert segmentation. Tables 2 and 3 clearly depict that the designed technique outdid the other techniques presented in the comparison table in terms of sensitivity, specificity, PPV and accuracy. Also, the method presented performs superiorly than the second non-specialist observer in terms of specificity and accuracy.

3.2 Arteriovenous Ratio Results

Tables 4 and 5 detail the arteriovenous ratio results using INSPIRE-AVR dataset. The whole dataset is evaluated in terms of average, minimum, maximum and standard deviation along with the error obtained with respect to the standard reference of two experts. The average, minimum, maximum and standard deviation are 0.6492, 0.4968, 0.7898 and 0.0629 respectively for the INSPIRE-AVR database. The mean accuracy of arteriovenous ratio for the designed method with respect to expert 1 and expert 2 is 97.44% and 98.43% respectively. There is no significant difference between arteriovenous measurements of the designed method and medical specialists 1 and 2.

3.3 Junction Points and Curvature-Based Index Geometrical Parameters Results

A total of 643 junction points from standard open-source benchmark databases are used to assess the geometrical parameters related to junction points and curvature-based index. Tables 4, 5, 6 and 7 show the overall results of all the geometrical parameters involved. Zamir [66] estimated the optimum value of bifurcation angle (θ) for normal retina is when $75° < \theta < 102°$. 61.68% of junction angles fall in the optimum range by the proposed method. Blood vasculature diameters are defined as $d_0 \geq d_1 \geq d_2$.

The total number of retinal vasculatures decreases with hypertension, and also there is a decrease in vasculature diameters for $d1$ and $d2$ for hypertensive patients. In our study, only 1% of the measurements were found when $d_0 \leq d_1$, which occurs due to thinning of parent blood vasculature due to hypertension. The curvature-based measurements are carried on an open-source benchmark dataset RETTORT.
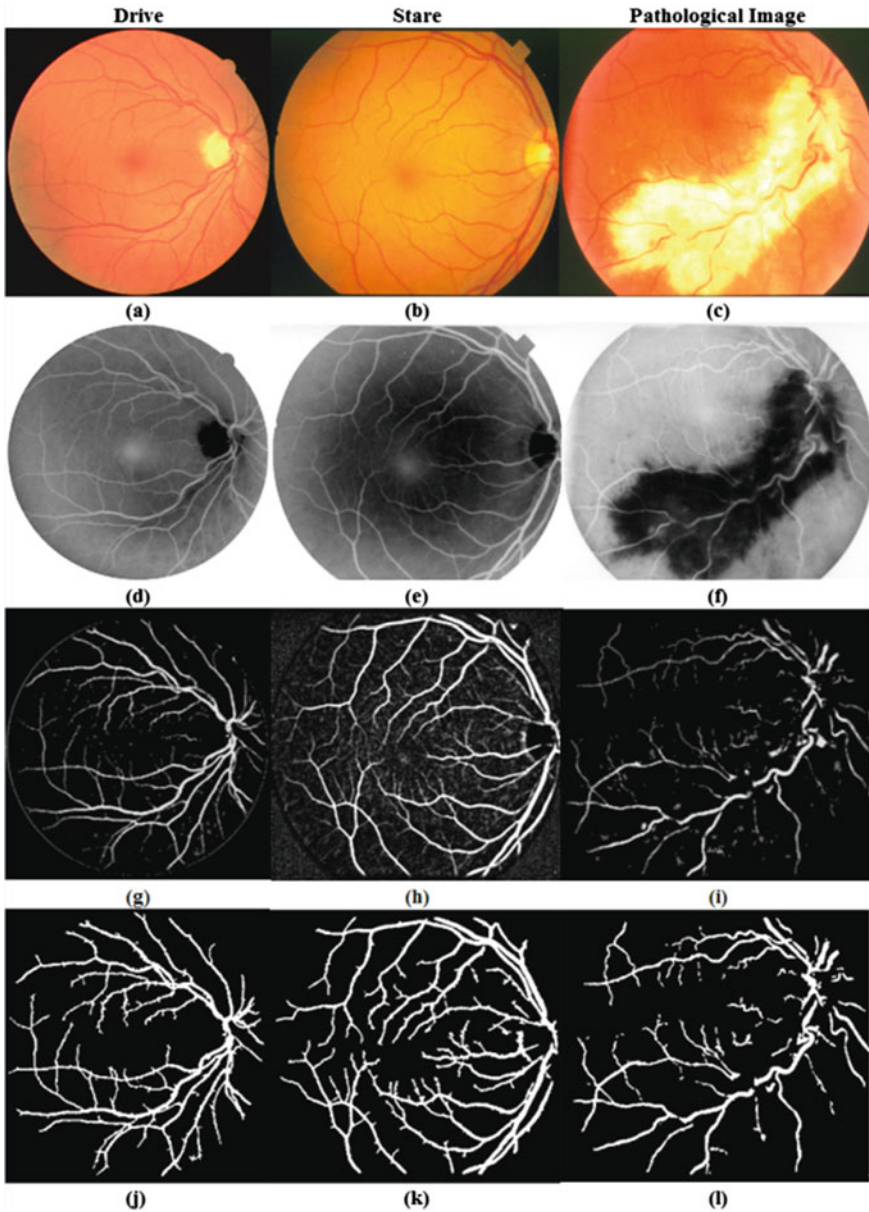
Fig. 8 Output images obtained by the designed technique; **a-c** original images of retina, **d-f** output image after taking the complement of homogenized image, **g-i** sum of top-hat transform output, **j-l** final segmented images

Table 2 Performance comparison of vessel segmentation methods (DRIVE images)

Sr. No.	Methods	TPR	FPR	SP	Acc	Std dev
1	2nd human observer	0.7763	0.0276	0.9702	0.947	0.0048
2	Chaudhuri et al. [13]	0.3357	–	–	0.8773	0.0232
3	Perez et al. [14]	0.785	–	0.949	0.932	–
4	Zana and Klein [16]	0.6971	–	–	0.9377	0.0077
5	Jiang et al. [17]	0.659	–	0.96	0.928	–
6	Mendonca and Campilho [20]	0.7344	–	0.9764	0.9452	–
7	Perez et al. [22]	0.7246	–	0.9655	0.9344	–
8	Perez et al. [23]	0.66	–	0.9612	0.922	–
9	Espona et al. [25]	0.6634	–	0.9682	0.9316	–
10	Anzalone et al. [27]	–	–	–	0.9419	–
11	Vlachos and Dermatas [31]	0.747	–	0.955	–	0.929
12	Al-Diri et al. [32]	0.7282	–	0.9551	–	–
13	Rossant et al. [38]	0.7031	–	0.9701	0.9433	–
14	Fraz et al. [39]	0.7224	–	0.9711	0.9469	–
15	Wang et al. [41]	–	–	–	0.9461	–
16	Chakraborti et al. [43]	0.7205	–	0.9579	0.937	–
17	Proposed methodology	0.7298	0.0298	0.9701	0.9489	0.0071

Table 3 Performance comparison of vessel segmentation methods (STARE images)

Sr. No.	Methods	TPR	FPR	SP	Acc	Std dev
1	2nd human observer	0.8951	0.06157	0.9384	0.9348	–
2	Hoover et al. [15]	0.6751	–	0.9567	0.9267	–
3	Vermeer et al. [18]	–	–	–	0.9287	–
5	Mendonca and Campilho [20]	0.6996	–	0.973	0.944	–
4	Perez et al. [22]	0.7506	–	0.9569	0.941	–
5	Perez et al. [23]	0.779	–	0.9409	0.924	–
6	Al-Diri et al. [32]	0.7521	–	0.9681	–	–
7	Gonzalez et al. [42]	0.7887	0.0367	–	0.9441	–
8	Chakraborti et al. [43]	0.6786	–	0.9586	0.9379	–
9	Proposed method	0.7295	0.0463	0.9736	0.9456	0.15391

Table 8 illustrates the measurements taken on the arteries and veins separately. These results are used to compare the shape-based features used to analyze curvature-based parameters.

Table 4 Geometrical measurements on DRIVE database

	θ	d_0	d_1	d_2
Min	29.0412	2	1.59	1.06
Max	157.1114	10.3	7.49	6.51
Mean	82.503	4.9504	3.9925	3.0108
Median	86.1215	4.94	3.94	3
Standard dev.	26.2662	1.2886	1.1564	0.9609
Variance	689.9157	1.6606	1.3373	0.9234

Table 5 Geometrical measurements on STARE database

	θ	d_0	d_1	d_2
Min	39.8713	1.88	0.94	0.81
Max	279.0756	7.97	7	5
Mean	82.3917757	4.31888235	3.59347059	2.36505882
Median	109.2546543	4.085	3.68	2.06
Standard dev.	23.61605434	1.33480776	1.2323664	1.01443268
Variance	557.7180225	1.78171176	1.51872694	1.02907367

Table 6 Geometrical measurements on DRIVE database depicting dependency of various parameters on each other

	α	β	σ_1	σ_2	γ
Min	0.2936	0.0862	0.3957	0.2542	0.2923
Max	1.8432	3.3976	0.996	0.984	1.8507
Mean	0.7669	0.615	0.8081	0.6134	1.0594
Median	0.7773	0.6043	0.8141	0.6122	1.028
Standard dev.	0.1638	0.2751	0.1131	0.132	0.2837
Variance	0.0268	0.0756	0.0128	0.0174	0.0804

Table 7 Geometrical measurements on STARE database depicting dependency of various parameters on each other

	α	β	σ_1	σ_2	γ
Min	0.2596	0.0674	0.4563	0.206	0.2506
Max	1	1	0.998	0.9539	2.1568
Mean	0.668	0.484	0.8306	0.5487	1.8481
Median	0.6852	0.4695	0.8507	0.5243	1.0299
Standard dev.	0.1948	0.2614	0.1296	0.1641	1.0589
Variance	0.0379	0.0683	0.0168	0.0269	0.1027

Table 8 Median values of the tortuosity measure on 30 arteries and veins form RETTORT database

Tortuosity notation	Median value for arteries	Median value for veins
τ_1	0.792	0.658
τ_2	0.208	0.342
τ_3	0.922	0.762
τ_4	0.919	0.762
τ_5	0.918	0.752
τ_6	0.913	0.773
τ_7	0.939	0.802
τ_8	0.928	0.712

4 Discussion and Conclusion

A novel technique is to recognize hypertension through retinal blood vasculature segmentation and computation of attributes such as arteriovenous ratio, number of junction points and curvature-based measurements in retinal images from standard open-source benchmark datasets DRIVE, STARE, INSPIRE-AVR and RETTORT datasets. Experimental results determine the high performance of the designed technique as compared to the existing methods in literature. It can be observed that the performance metrics such as sensitivity, specificity, positive predictive value and accuracy are high as compared to many methods reported in literature.

DRIVE dataset shows the average accuracy, specificity and positive predictive value of the designed method are 0.9489, 0.9703 and 0.7084 with a sensitivity and specificity of 0.7298 and 0.0298, respectively. In terms of average accuracy, the proposed method is improved than second observer, and the methods proposed in [13, 14, 16, 17, 20, 22, 23, 25, 27, 38, 39, 41, 43, 68]. This algorithm has achieved higher sensitivity than the methods proposed in [13, 16, 17, 22, 23, 25, 32, 38, 39, 43]. For the STARE dataset, the average accuracy, specificity produced by the proposed algorithm is 0.9456, 0.9736 with a sensitivity and specificity of 0.7295 and 0.0463 respectively. There is an enhancement in the results of the proposed method as compared to earlier methods in accuracy and sensitivity from the methods proposed in [15, 18–20, 22, 23, 32, 43]. It can be seen that the designed method attains high accuracy and specificity overall but not competitive sensitivity in few cases. This may be due to false positive produced in the pathological images. Computation of all the parameters in unhealthy images comprising of lesions is planned for future work.

The mean arteriovenous ratio a result of the designed technique on the INSPIRE-AVR dataset is 0.6492 with the highest and the lowest value being 0.4967 and 0.7897 respectively. It was also seen that the designed technique achieved a high mean accuracy of 97.44%, 98.43% for medical specialists 1 and 2 respectively. Additionally, RETTORT dataset was used to compute the curvature-based measurements of retinal blood vasculature. Finally, it can be highlighted that the designed technique used for

the measurement of junction points and curvature-based index plays a very significant role for ophthalmologists for the analysis and effective planning of retinal abnormality related to hypertension.

References

1. Chobanian AV (2001) Control of hypertension—an important national priority. N Engl J Med 345:534–535
2. Kearney P, Whelton M, Reynolds K, Muntner P, Whelton P, He J (2005) Global burden of hypertension: analysis of worldwide data. The Lancet 365:217–223
3. Kabedi NN, Mwanza J-C, Lepira FB, Kayembe TK, Kayembe DL (2014) Hypertensive retinopathy and its association with cardiovascular, renal and cerebrovascular morbidity in Congolese patients: cardiovascular topic. Cardiovasc J Afr CVJA 25:228–232
4. Grosso A (2005) Hypertensive retinopathy revisited: some answers, more questions. Br J Ophthalmol 89:1646–1654
5. Keith NM, Wagener HP, Barker NW (1939) Some different types of essential hypertension. Am J Med Sci 197:332–343
6. Tso MO, Jampol LM (1982) Pathophysiology of hypertensive retinopathy. Ophthalmology 89:1132–1145
7. Chapman N, Dell'Omo G, Sartini MS, Witt N, Hughes A, Thom S et al (2002) Peripheral vascular disease is associated with abnormal arteriolar diameter relationships at bifurcations in the human retina. Clin Sci 103:111
8. Azzopardi G, Petkov N (2013) Automatic detection of vascular bifurcations in segmented retinal images using trainable COSFIRE filters. Pattern Recogn Lett 34:922–933
9. Cavallari M, Stamile C, Umeton R, Calimeri F, Orzi F (2015) Novel method for automated analysis of retinal images: results in subjects with hypertensive retinopathy and CADASIL. Biomed Res Int 2015:1–10
10. Martinez-Perez M, Highes A, Stanton A, Thorn S, Chapman N, Bharath A et al (2002) Retinal vascular tree morphology: a semi-automatic quantification. IEEE Trans Biomed Eng 49:912–917
11. Grisan E, Foracchia M, Ruggeri A (2008) A novel method for the automatic grading of retinal vessel tortuosity. IEEE Trans Med Imaging 27:310–319
12. Heneghan C (2002) Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis. Med Image Anal 6:407–429
13. Chaudhuri S, Chatterjee S, Katz N, Nelson M, Goldbaum M (1989) Detection of blood vessels in retinal images using two-dimensional matched filters. IEEE Trans Med Imaging 8:263–269
14. Martínez-Pérez ME, Hughes AD, Stanton AV, Thom SA, Bharath AA, Parker KH (1999) Retinal blood vessel segmentation by means of scale-space analysis and region growing. In: Medical image computing and computer-assisted intervention—MICCAI'99 Lecture Notes in Computer Science, pp 90–97
15. Hoover A, Kouznetsova V, Goldbaum M (2000) Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans Med Imaging 19:203–210
16. Zana F, Klein J-C (2001) Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. IEEE Trans Image Process 10:1010–1019
17. Jiang X, Mojon D (2003) Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images. IEEE Trans Pattern Anal Mach Intell 25:131–137
18. Vermeer K, Vos F, Lemij H, Vossepoel A (2004) A model based method for retinal blood vessel detection. Comput Biol Med 34:209–219

19. Fraz M, Barman S, Remagnino P, Hoppe A, Basit A, Uyyanonvara B et al (2012) An approach to localize the retinal blood vessels using bit planes and centerline detection. Comput Methods Programs Biomed 108:600–616
20. Mendonca A, Campilho A (2006) Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. IEEE Trans Med Imaging 25:1200–1213
21. Al-Rawi M, Qutaishat M, Arrar M (2007) An improved matched filter for blood vessel detection of digital retinal images. Comput Biol Med 37:262–267
22. Martinez-Perez ME, Hughes AD, Thom SA, Bharath AA, Parker KH (2007) Segmentation of blood vessels from red-free and fluorescein retinal images. Med Image Anal 11:47–61
23. Martinez-Perez ME, Hughes AD, Thom SA, Parker KH (2007) Improvement of a retinal blood vessel segmentation method using the insight segmentation and registration toolkit (ITK). In: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society
24. Wang L, Bhalerao A, Wilson R (2007) Analysis of retinal vasculature using a multiresolution hermite model. IEEE Trans Med Imaging 26:137–152
25. Espona L, Carreira MJ, Ortega M, Penedo MG (2007) A snake for retinal vessel segmentation. In: Pattern recognition and image analysis. Lecture Notes in Computer Science, pp 178–185
26. Yang Y, Huang S, Rao N (2008) An automatic hybrid method for retinal blood vessel extraction. Int J Appl Math Comput Sci, vol 18
27. Anzalone A, Bizzarri F, Parodi M, Storace M (2008) A modular supervised algorithm for vessel segmentation in red-free retinal images. Comput Biol Med 38:913–922
28. Farnell D, Hatfield F, Knox P, Reakes M, Spencer S, Parry D et al (2008) Enhancement of blood vessels in digital fundus photographs via the application of multiscale line operators. J Franklin Inst 345:748–765
29. Lam B, Yan H (2008) A novel vessel segmentation algorithm for pathological retina images based on the divergence of vector fields. IEEE Trans Med Imaging 27:237–246
30. Espona L, Carreira M, Penedo M, Ortega M (2008) Retinal vessel tree segmentation using a deformable contour model. In: 2008 19th international conference on pattern recognition
31. Vlachos M, Dermatas E (2010) Multi-scale retinal vessel segmentation using line tracking. Comput Med Imaging Graph 34:213–227
32. Al-Diri B, Hunter A, Steel D (2009) An active contour model for segmenting and measuring retinal vessels. IEEE Trans Med Imaging 28:1488–1497
33. Muramatsu C, Hatanaka Y, Iwase T, Hara T, Fujita H (2010) Automated detection and classification of major retinal vessels for determination of diameter ratio of arteries and veins. In: Medical imaging 2010: computer-aided diagnosis
34. Ng J, Clay S, Barman S, Fielder A, Moseley M, Parker K et al (2010) Maximum likelihood estimation of vessel parameters from scale space analysis. Image Vis Comput 28:55–63
35. Amin MA, Yan H (2010) High speed detection of retinal blood vessels in fundus image using phase congruency. Soft Comput 15:1217–1230
36. Delibasis KK, Kechriniotis AI, Tsonos C, Assimakis N (2010) Automatic model-based tracing algorithm for vessel segmentation and diameter estimation. Comput Methods Programs Biomed 100:108–122
37. Miri MS, Mahloojifar A (2011) Retinal image analysis using curvelet transform and multistructure elements morphology by reconstruction. IEEE Trans Biomed Eng 58:1183–1192
38. Rossant F, Badellino M, Chavillon A, Bloch I, Paques M (2011) A morphological approach for vessel segmentation in eye fundus images, with quantitative evaluation. J Med Imaging Health Inform 1:42–49
39. Fraz MM, Remagnino P, Hoppe A, Uyyanonvara B, Rudnicka AR, Owen CG et al (2012) An ensemble classification-based approach applied to retinal blood vessel segmentation. IEEE Trans Biomed Eng 59:2538–2548
40. Xu Y, Géraud T, Najman L (2013) Two applications of shape-based morphology: blood vessels segmentation and a generalization of constrained connectivity. Lecture notes in computer science mathematical morphology and its applications to signal and image processing, pp 390–401

41. Wang Y, Ji G, Lin P, Trucco E (2013) Retinal vessel segmentation using multiwavelet kernels and multiscale hierarchical decomposition. Pattern Recogn 46:2117–2133
42. Salazar-Gonzalez A, Kaba D, Li Y, Liu X (2014) Segmentation of the blood vessels and optic disk in retinal images. IEEE J Biomed Health Inform 18:1874–1886
43. Chakraborti T, Jha DK, Chowdhury AS, Jiang X (2014) A self-adaptive matched filter for retinal blood vessel detection. Mach Vis Appl 26:55–68
44. Wang S, Yin Y, Cao G, Wei B, Zheng Y, Yang G (2015) Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. Neurocomputing 149:708–717
45. Oliveira WS, Ren TI, Cavalcanti GDC (2012) An unsupervised segmentation method for retinal vessel using combined filters. In: 2012 IEEE 24th international conference on tools with artificial intelligence
46. Youssef D, Solouma NH (2012) Accurate detection of blood vessels improves the detection of exudates in color fundus images. Comput Methods Programs Biomed 108:1052–1061
47. Kaur J, Mittal D (2015) Segmentation and measurement of exudates in fundus images of the retina for detection of retinal disease. JBEMi J Biomed Eng Med Imaging, vol 1
48. Kumari K, Mittal D (2015) Automated drusen detection technique for age-related macular degeneration. JBEMi J Biomed Eng Med Imaging, vol 1
49. Morales S, Naranjo V, Navea A, Alcaniz M (2014) Computer-aided diagnosis software for hypertensive risk determination through fundus image processing. IEEE J Biomed Health Inform 18:1757–1763
50. Narasimhan K, Neha V, Vijayarekha K (2012) Hypertensive retinopathy diagnosis from fundus images by estimation of AVR. Procedia Eng 38:980–993
51. Rani A, Mittal D (2015) Measurement of arterio-venous ratio for detection of hypertensive retinopathy through digital color fundus images. JBEMi J Biomed Eng Med Imaging, vol 2
52. Thongnuch V, Uyyanonvara B (2006) Automatic detection of optic disc from fundus images of ROP infant using 2D circular hough transform. Sirindhorn International Institute of Technology, Thammasat University, Thailand
53. Niemeijer M, Xu X, Dumitrescu AV, Gupta P, Ginneken BV, Folk JC et al (2011) Automated measurement of the arteriolar-to-venular width ratio in digital color fundus photographs. IEEE Trans Med Imaging 30:1941–1950
54. Kumari K, Mittal D (2015) Drusen quantification for early identification of age related macular degeneration. Adv Image Video Process AIVP, vol 3
55. Witt N, Wong TY, Hughes AD, Chaturvedi N, Klein BE, Evans R et al (2006) Abnormalities of retinal microvascular structure and risk of mortality from ischemic heart disease and stroke. Hypertension 47:975–981
56. Mittal D, Kumari K (2015) Automated detection and segmentation of drusen in retinal fundus images. Comput Electr Eng, 47:82–95
57. Parr J, Spears G (1974) General caliber of the retinal arteries expressed as the equivalent width of the central retinal artery. Am J Ophthalmol 77:472–477
58. Knudtson MD, Lee KE, Hubbard LD, Wong TY, Klein R, Klein BE (2003) Revised formulas for summarizing retinal vessel diameters. Curr Eye Res 27:143–149
59. Parr J, Spears G (1974) Mathematic relationships between the width of a retinal artery and the widths of its branches. Am J Ophthalmol 77:478–483
60. Hubbard LD, Brothers RJ, King WN, Clegg LX, Klein R, Cooper LS et al (1999) Methods for evaluation of retinal microvascular abnormalities associated with hypertension/sclerosis in the atherosclerosis risk in communities study. Ophthalmology 106:2269–2280
61. Murray CD (1926) The physiological principle of minimum work applied to the angle of branching of arteries. J Gen Physiol 9:835–841
62. Zamir M (1979) Arterial bifurcations in the human retina. J Gen Physiol 74:537–548
63. Al-Diri B, Hunter A (2009) Automated measurements of retinal bifurcations. In: IFMBE proceedings world congress on medical physics and biomedical engineering, 7–12 Sept 2009, Munich, Germany, pp 205–208
64. Hart W, Goldbaum M, Cote B, Kube P, Nelson M (1999) Measurement and classification of retinal vascular tortuosity. Int J Med Inform 53:239–252

65. Lotmar W, Freiburghaus A, Bracher D (1979) Measurement of vessel tortuosity on fundus photographs. Albrecht Von Graefes Archiv for Klinische Und Experimentelle Ophthalmologie Albrecht Von Graefes Arch Klin Ophthalmol 211:49–57
66. Zamir M (1976) Optimality principles in arterial branching. J Theor Biol 62:227–251
67. DRIVE database.Available at: http://www.isi.uu.nl/Research/Databases/DRIVE/. Accessed 08 July 15
68. STARE database. Available at: http://www.ces.clemson.edu/~ahoover/stare/. Accessed 08 July 15
69. INSPIRE-AVR dataset. Available at: http://www.medicine.uiowa.edu/eye/Datasets/. Accessed 08 July 15
70. RET-TORT dataset. Available at: http://bioimlab.dei.unipd.it/Retinal%20Vessel%20Tortuosity.htm. Accessed 08 July 15

A Comprehensive Review of Big Data Analysis Techniques in Health-Care

Sharad Kumar Tiwari, Jaskirat Kaur, Parveen Singla, and P. N. Hrisheekesha

Abstract With the development of the big data technologies in health-care, there is a big challenge to manage health care records such as meta-data, sensor data, imaging data and other health records. To handle, control and drive this big data, the traditional handing techniques, software and available hardware are not sufficient. Therefore, the current methodologies need to be augmented with advanced big data computation and system requirements. This huge data also provides opportunities for the health-care researchers to perform the data analysis on this available huge data. Hence, a modern solution is utmost required for diagnosis cost reduction and improving the healthcare facilities. In this research paper, a comprehensive review of big data techniques illustrated the different healthcare big data analysis algorithms, frameworks, tools, and techniques that are deployed on various platforms. Finally, convergence point of all the above-mentioned platforms in the form of Smart-health is established.

Keywords Health-care analytics · Big data · Data management · Learning health-care system · Smart-health

S. K. Tiwari (✉)
Electronics and Communication Engineering Department, Vignan Foundation for Science Technology and Research, Vadlamudi 522213, India
e-mail: drskt_ece@vignan.ac.in

J. Kaur
Department of Electronics and Communication Engineering, Punjab Engineering College (Deemed to be University), Chandigarh, India
e-mail: jaskiratkaur@pec.edu.in

P. Singla
Electronics and Communication Engineering Department, Department of Research and Development, Chandigarh Engineering College, Landran, Mohali 140307, India
e-mail: parveen.rise@cgc.edu.in

P. N. Hrisheekesha
Chandigarh Group of Colleges, Mohali 140307, India

401

1   Introduction

Health industry generates enormous data related to the patient's information, drug-disease information details, and other health-care issues [1]. The big data related to the health-care involves computerized metadata, electronic generated patient record, diagnosis and complied reports by the physicians, medical images, physiological data in tabular form, insurance as well as administrative data, medical news data, data generated from the mobile health-care systems, data from social media platforms and data of medical journals. The medical big data storage provides a facility to improve the health-care quality among the population in a cost-effective manner. The health-care system generates approximately a zettabyte (a trillion gigabytes) of data annually, and this amount is doubling every two years. International Data Corporation (IDC) approximated that the size of digital-universe was 130 Exabytes (EB) in 2005. The digital-universe expanded with a huge factor and becomes 16,000 EB by 2017. IDC made a predicted that digital-universe will expand to 40,000 EB by the end of 2020 [2]. According to the Shafqat et al. [3], health data of USA was around 150 EB in 2011 and it can expand beyond zettabytes (ZB) and yottabytes (YB). Because of this vast data, the health-care domain researchers of this domain use computer techniques and algorithms to extract valuable information and transform the data into some knowledge database. With the help of decision-making frameworks, the big heterogeneous medical data are now increasing the intuitive abilities of medical practitioners. To deal with this much big data, the highly recognized algorithms are required, which correlated all the associated factors and can conclude a decision. Therefore, it's a big opportunity for the AI in health-care researchers to expand the foundation for medical decision-making system [4]. Concept wise the basic architecture for the analysis of the big data in health-care are not very different from the traditional health-care framework architecture [1]. The major difference both of the architectures are way of performing the data pre-processing. Processing in the big data requires to be break into various nodes as the part of the distribution process [5]. The big data analysis in the health-care needs to refine the data as it can come from multiple sources [6].

This paper reviews various health-care framework loaded with advanced analytical algorithms and methodologies, used for the diagnosis, prediction and treatment of health-care diseases. In this review paper, we have analyzed few of the main frameworks in the health-care.

2   Health-Care Analytics Frameworks

We explored a number of frameworks for big data analytics in health-care. Few important frameworks are discussed in the following subsections:

2.1 Health-Care Informatics and Big Data

There are number of standard developments in the big data health-care informatics; the data are collected from the various number of sources considering security considerations [7–9]. Data analytics from the sources in the health-care informatics is not only limited to the diagnosis but also includes personalized medicine genomics, proteomics and long-term tracing of patients [10]. The big data technology is providing the health-care facility to people on their cellular network [8]. But still for in the big data for health-care many issues to be considered [7]. Handling the big data of the order yottabytes itself a big challenge. Health-care data can come from multiple sources, like real-time medical imaging, patient care devices, wearable health devices (smart watch, band), mobile devices and social media [11]. The theory of six V is also applicable for health-care big data as well, these V's are named as: volume, velocity, variety, veracity, and variability. Challenges with the big data in variety and heterogeneity have been already addressed by Zhao et al. [12] and Torino [13] respectively. Apart from these the size is also challenging aspect in the big data analysis. Heterogeneity in the data can arises from the structured and unstructured data entry from the multiple sources [13]. Also, the variety occurs when the data collected from the unmanaged sources like Twitter and other social media platforms; therefore, variety handling is more challenging. To avail the full advantage of the big data in health-care, it is essential to critically analyze the data. Apart from these challenges in the big data other complications in the data also occurs, like data ownership, governance and other legal technicalities [10]. As the internet is reaching to everyone, the communication between the patient and doctor is also expanding. Even nowadays, the patients suffering from chronic and acute diseases like diabetes, heart diseases, cancer, thyroid, etc. share and discuss their experience on social media and other platforms [1].

2.2 Internet of Things (IoT) Cloud Architecture in Health-Care

Use of IoT in health-care technology is a revolution; it provides a technological invention for health-care applications. Even the IoT is changing the face of health policies. Some e-health regulations and policies are under consideration which involves the IoT, big health-care data and intelligent systems to improve the health-care facilities [14]. To handle the health-care-related work in IoT, one network is established named as IoT health-care network (IoThNet). IoThNet contains the topologies, computing platforms and architectures [14].

A big data management facility is needed to process this colossal data from various IoT sensor networks. The variable data collecting devices like smart-watches or equipment collect the physiological signals are now can be handled by the different mobile applications [8]. Mobile applications also track the other metadata like diet

routine, lifestyle habits, etc. A simple sports-based mobile application can collect the data more than 100 million from all over the globe.

2.3 Big Data of Health-Care from Social Media (SM) Platforms

The extracting and processing of big data on health-care is very challenging. Analysis of any associations within the data or any relevant pattern could help to improve the health-care service by more personalized medicine recommendations; also this can reduce the overall cost of health-care services [1]. The analytical results of big data collected from the various SM platforms can help to develop a patient focus health service then a disease centered health service. Also, this will open a way for customizing a patient wise solution for better health-care [15]. The big data analytics tools like Apache Hadoop are managing large enterprises like Facebook or Google, and its applications can also be applied in health-care-related businesses and institutions [16]. The Twitter Application Programming Interface (API) have a specific library named, twitter4j. This library is also integrated with a web-based application using NoSQL Data-base Management System (DBMS) and extract the health data as input [16].

2.4 Integrative Health-Care Analytics Framework

The Integrative health-care analytics frameworks allow the various health-care services for the doctor and patient interaction. These frameworks incorporate the different feature selection, clustering, prediction and classification algorithms. Gemini is one such integrative health-care analytics framework, which has two modules; analytics and profiling. The profiling module stores patients profiles including the demographics collected from different data sources and analytics gives the predictions and conclusions [17].

2.5 Cloud-Based Platform for Health-Care

Diseases are spreading at a greater speed. Timely and accurate identification, prediction, confirmation and treatment of these diseases is utmost required. Thus, the integration of Predictive Analytical Decision Support System (PADSS) in a cloud-based health-care platform is Message Oriented Middleware (MOM) [18]. It helps in real-time channelized data sharing between health organizations using Health Level Seven (HL7) platform consisting Fast health-care Interoperability Resources (FHIR). HL7

is a non-profitable international body responsible for the development of interoperability standards that gives the specifications for health-care software applications. FHIR in HL7 enables clinical and administrative data exchange on international standards between health-care applications. Table 1 shows a detailed analysis of different health-care systems proposed so far.

3 Management of Big Data in Health-Care

The big data is generated at a rapid rate from different sources. The biggest challenge with big data is to handle a large volume of information. Handling it becomes more crucial when dealing with biomedical and health-care-related data. To make data available for further analysis and research, this huge data to be stored in the appropriate file format so that later it can be easily readable and accessible. One more main challenge with the health-care data to implementation of the high-end tools and protocols for effective data acquisition as well as high-end hardware for fast computation. To achieve these complex tasks, experts from various background like biomedical, statistics, computer science and mathematics needs to work together. Analytical developers developed cloud-based software tools and applications to collect the data from different sensors and store this data into the cloud. Also, this stored data should be made available for data engineers and researchers. In the big data of health-care handling the heterogeneity can be one more challenge. The heterogeneity of the big data may lead to less informative analysis with the conventional tools. Heterogeneity of data is another challenge in big data analysis. The huge size and highly heterogeneous nature of big data in health-care renders it relatively less informative using conventional technologies. Therefore, various big data analytics tools and platforms like Apache Spark, Hadoop, etc. are developed and discussed as follows:

3.1 Apache Hadoop

The most effective way to analyze a big data is to distribute the data and analyze it parallel in multiple nodes. Although, sometimes the huge amount of data needs thousands of computing machines for the fast computation. Also, distribution of the data and failures handling in parallel is also a complex task. Therefore, to handle this distribution problem an open source application Hadoop is proposed [23]. MapReduce algorithm used for the processing and generating the huge datasets implemented by the Hadoop. Yao et al. [24] proposed a big medical data processing system based on hadoop. Authors have process the health-care big data to explore the features of the Hospital information system. A five-node Hadoop cluster is proposed to execute the algorithm. To handle the big data from the countries having huge population like India, Augustine [25] has proposed a Hadoop-based big data analytics framework. Many specific Hadoop-based applications for specific diseases in health-care

Table 1 Detailed analysis of different health-care systems

Citations	Technique proposed	Application areas	Platform/ algorithms	Testing and validation
[7]	Big data health-care Informatics	Applications in disease diagnosis and prognosis	Big data algorithms and taxonomy	Proposed a new hypothesis for the diagnosis and treatment
[8, 14]	IoT big data cloud architecture for health-care	IoT application for the various disease diagnosis, prognosis, treatment and follow-up	New protocols connected to the data layer like IPv6, 6LoPAN	IoT-based health-care techniques are proposed along with the existing techniques
[16]	Management big data of health-care from Twitter	Diagnosis and follow-up different diseases through Twitter	Hadoop	Data collected from the Twitter API through various protocols
[4, 19]	Health-care system design	Clinic based on website or mobile applications	Decision support system for the big data	Big data collected from various sources and validation done on the completely unknown data to the designed system
[20]	Predictive analytics	Diagnosis of diabetes	Hadoop	Concept wise explained
[18]	Health-care platform (Cloud based)	Assistive system for the medical practitioners	Analytical Decision Support System (PADSS) is presented	Google cloud platform is accessed to design a decision-making system
[21]	Big data of health-care from the health insurance companies	Patient claim system	Hadoop	Condition-based support system
[13]	Multi-level disease diagnosis	Mobile application for disease diagnosis	Neural network and clustering techniques for patient classification	K-means clustering and K modified algorithms are applied for the testing and validation
[22]	Hybrid decision-making system for anesthesia	Anesthesia control and assessment of	Low-level and high-level frameworks	Taxonomy

are also proposed by the researchers. Rodge [26] proposed a software to process the patient information from different ships. A hybrid Hadoop five method is proposed to arrange and organize the big data collected from different ships. Then an Apache Hive warehouse system is made on the top of Hadoop to identify and analyze the traumatic Brain Injury (TBI) and other injury cases. In a similar manner for the disease prediction the Hadoop framework are implemented. Kalyankar [27] proposed a machine learning and Hadoop-based framework for the predictive analysis of the diabetes. Intelligent Care System is also proposed by Rathore et al. [28]. Authors have demonstrated IoT-based and Hadoop-based shared big data sharing system for the health-care applications. In the proposed system we have involved an architecture with advanced and enhanced data processing features, where the data is collected and generated from various sensors like physiological measuring devices, wearable health monitoring tools, etc.

3.2 Apache-Spark

Apache-spark is also an alternative to the Hadoop. It is an application for the distributed processing of the big data. It includes the high-level libraries that supports the SQL queries, machine learning techniques, graph processing methods and spark streaming of the data [29]. Apache Spark able to handle more complex tasks with less efforts. Spark uses the resilient distributed data-set for the processing of the data; this makes apache spark much faster than Hadoop for multi-analysis [30]. One disadvantage with the apache spark that it requires a big memory while working with big data [31]. Therefore, sometimes it's not a cost-effective solution. However, apache-spark provides a real-time framework for data processing and supports most of the programming languages. Also, it has a good fault tolerance capability and great scalability while analyzing the big data. Nazari and Shahriari [32] presented a detailed systematic review article on the big data analysis for the health-care in big data platforms like Apache-Hadoop, apache-spark AND Apache-Flink. Authors have presented a detailed comparison of all three big data analysis platforms. For the big data of health-care, the cluster computing capabilities of Apache-spark can provide the required scalability for the processing. The Lambda architecture can be used with the help of Apache-spark which can do the batch processing and real-time data mining both. Apache-spark also enables the personalized medical prescription and decision making for disease diagnosis for the health-care big data analysis. Accurate clinical decision making like prognosis, anomaly or biomarker detection, potential disease threat can also be performed with the help of Apache-spark.

3.3 Tools for Image Analysis

Magnetic resonance imaging (MRI), computed tomography (CT), X-ray, optical coherence tomography (OCT), fundus, ultrasound, molecular imaging, photo-acoustic imaging, positron emission tomography (PET), mammograms, etc. are some of the most popular medical imaging modalities. Normally, these are high-definition large images. Storing and communication of these images is a big challenge for the medical professionals. To help medical professionals, an effective systems are developed. Picture Archiving and Communication System (PACS) is one such most popular and effective techniques to store the medical images and reports [33]. To extract the hidden information many software tools are developed, these tools perform the operations like regression, segmentation, visualization, diffusion, classification, and reconstruction to help medical practitioners. For example, Statistical Parametric Mapping (SPM) is a powerful software which used to analyze and process different types of brain images like MRI, PET, EEG, fMRI (Functional MRI), etc. [33]. Similarly, significant state arts are available for diagnosis diabetic [34]. Other software like GIMIAS, Elastix, and MITK are also available which supports and process almost all imaging modalities.

4 Big Data Health-Care Analytics Techniques

As in the previous section different platforms for the analysis of big data in health-care, broadly, these techniques are divided into two major parts from the literature; data mining and machine learning techniques [35]. These two parts consist of the different algorithms, deep neural networks and many other techniques and approaches. Also, various hybrid techniques are also proposed for the disease prediction from the big data of health-care [36–38]. The most effective techniques and tools are designed to develop and formulate relationship and pattern from the data. A number of data mining technique like case studies, rule-based methods, artificial neural networks (ANN), classification and regression are used in the health-care applications [39]. Cloud Security Alliance (CSA) is a secure architecture for the big data analysis which was proposed by big data work group [5]. CSA architecture proposed by the big data group is explained in Fig. 1. According to this, Electronic Health Record (EHR) becomes an input for the further analysis, and data level security is performed by the FLUME and SCOOP.

Data mining tools and algorithms for big data for health-care analysis give the following advantages:

- Secure data mining for health-care system
- Reducing the cost and other expenses for the diagnosis and treatment
- Accurate health prediction
- Development of a secure and accurate health-care information system
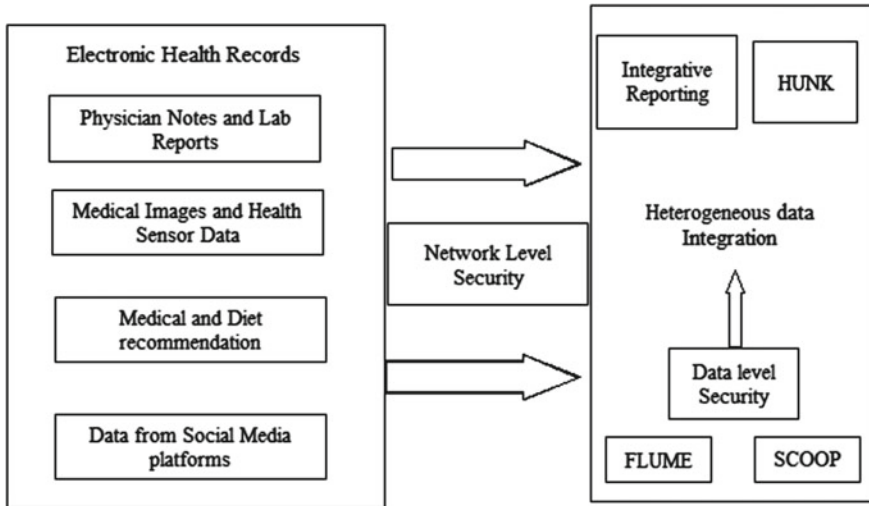- E-governance health-care system for remote population

Fig. 1 Architecture proposed by the big data group [5]

- Accuracy in e-health insurance system.

For the big data analysis for health-care for classification, machine learning (ML) and data mining techniques overlap at some points. Data mining is perceiving the previously unknown information while machine learning algorithms evaluate or predict results based on previously known information utilizing it as a training set [40]. Based on the big data for health-care analysis the following tasks can be performed for the general population.

- Recommended tests for the patients based on the symptoms.
- Disease prediction based on the health-related problems.
- Most common symptoms for chronic population.
- Estimation of the expenditure of different diseases at different hospitals and countries.

4.1 Clustering Techniques Classification of Big Data for Health-Care

Clustering is simply a data analytics technique to divide the data into a group of homogenous features based on the clandestine information and features. Cluster organization of the data shows the organization of the data. Clustering is mainly an unsupervised task, like every machine learning technique the heart of the analysis is the selection of the algorithm and it completely depends on the type of the data structure. The clustering algorithms for the big data in health-care are divided as density-based clustering, partitional clustering, rule based or fuzzy clustering,

hierarchical clustering, grid-based clustering, graph clustering, etc. These clustering algorithms are shown in Table 1 and discussed in detail in the following section:

4.1.1 Partitional Clustering

This clustering technique requires initially number of clusters. It classifies the data into defined number of clusters in such a way that every observation belongs to exactly one cluster; however, in fuzzy partitioning, one observation may belong to more than one cluster. *K*-means is one of the most used partitional algorithm [41]. Within each cluster, it tries to minimize the dissimilarity between each element. There are many variations (like K-medoids, weighted *k*-means, etc.) are also proposed by the researchers for the analysis of the big data in health-care. Gheid et al. [42] proposed a clustering algorithm based on multiparty additive cryptography-free scheme. Overlapping clusters are very common with the medical and health-care big data therefore to Khanmohammadi et al. [43] combined overlapping *k*-means clustering with k-harmonic means for effecting medical clustering. The output of the k-harmonic means clustering is used to initialize the cluster centers of overlapping k-means clustering method. Pattern discovery is also one important task with the health-care big data. Ramzi et al. [44] proposed a K-means clustering approach for the pattern recognition in the healthcare data.

4.1.2 Rule-Based and Fuzzy Clustering

The rule-based clustering algorithms extract the information with the help of "if–then" rule based, where "if" gives the condition and "then" gives the conclusion. If can consists more than one condition and based on these condition/s "then" produces the result. In the health-care simple example of if–then as: "*if person feels the fatigue then Diabetes can be there*".

$$\text{Symptom/s}^- \rightarrow \text{Possible disease}$$

Other advanced, cloud-based rules-based clustering algorithms are also presented for different applications of big data analysis. The fuzzy clustering techniques are also known as soft clustering. As discussed earlier in fuzzy clustering, each data point can belong to more than one cluster, the fuzzy clustering contains a continuous interval [0 1] from level 0 and 1. Fuzzy c-means, fuzzy mountain, and c-shells are some of the most popular fuzzy clustering algorithms. In c-means, the membership of each data point is calculated by optimizing the main objective function. This method automatically produces the clusters that correlated the data attributes [45]. Effective analysis of lung infection using fuzzy decision support system is suggested by Tiwari et al. [46]. Jindal et al. [47] presented a cloud-based fuzzy rule-based classifier for providing health-care e-service. Processing, clustering and retrieval of the health-care big data are performed on the cloud.

4.1.3 Hierarchical and Distribution-Based Clustering

Hierarchical cluster seeks to construct hierarchy of the clusters (sub-clusters) and can be seen as a sequence of partitional clustering methods. The pair-wise distance among the observations the hierarchy of the clusters is calculated. The planning for the hierarchical cluster can be broadly classify as bottom-down and bottom-up. These techniques are also known as divisive and agglomerative clustering techniques. The major disadvantage with this clustering is that back tracking is not possible once sub-clusters are been made. Also, it suffers with high computational complexity. Hierarchical clustering approach is useful when there are lots of sub-clusters are possible within a single cluster for an example the health insurance data. To cluster this Hillerman et al. [48] presented a cluster-based algorithm to evaluate the suspected health-care claims. Authors have presented an analytic hierarchy process for clustering based on different criteria.

Distribution based is an iterative clustering which is faster than other available clustering algorithms. It automatically produces the clusters according to the features, dependences and attributes of the data. The density-based clustering approach identifies the characteristic clusters in the data based on density. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is the most widely used density-based clustering algorithm and does not require the number of clusters as a parameter [45]. Vijayakumar et al. [49] used a hidden Markov model along with the Gaussian mix clustering technique in order to model the DNA copy change.

4.2 Machine Learning (ML) Techniques

By exploring the enormous big data of healthcare, the ML techniques are now being used for providing the advanced patient care and results in form of enhanced business outcome. In 2018, around 30% of healthcare service providers were using cognitive analytics techniques for analysis of the patient data. Supervised learning, unsupervised learning, reinforcement learning and hybrid learning come under the machine learning techniques.

4.2.1 Supervised Learning Models

The supervised learning the models are trained with the help of known target values, these can be either classification or regression models. The output is predicted with the help of known predictions, and based on the suitable training of the model, the output is predicted. The predicted values are compared with the known output values and then parameters of the classifier or regression model are been adjusted to minimize the error. Multiple supervised learning-based model are proposed for the analysis and diagnosis from the medical big data. Diabetes is a chronic disease, globally 463 million people are subfreezing with diabetes [50]. The EHR generates

Table 2 Machine learning algorithms for analysis of the healthcare big data

Author and year of publication	Healthcare data set used	Summary
Gupta and Garg (2020) [51]	Wisconsin Breast Cancer Dataset (WBCD)	Authors have proposed six different supervised ML algorithms and compared with one deep learning-based Adam Gradient Descent Learning model to diagnosis the malignant condition of the tumor
Luo (2016) [52]	Electronic health records	The EHR of all 50 states of the USA, the authors have predicted the type 2 diabetes for next year. Author has utilized the Associative Classifier technique for the prediction
Seenivasagam et al. (2020) [53]	Sputum cell images	Machine learning algorithm is streamlined with apache spark framework to classify different stages of the lung cancer
Zhang et al. (2017) [54, 55]	Electronic health records	Authors have constructed and evaluated the performance of the different machine learning-based classification algorithm for the predication of the type 2 diabetes. Classifiers like support vector machines, k-nearest-neighbors, random forest, logistic regression, etc. were considered
Taylor et al. (2015) [56]	Real-life hospital data	Authors have considered 500 clinical variables from the data available of 4 different hospitals and constructed a random forest model to propose a clinical decision-making system
Al-Shargabi et al. (2019) [57]	WBCD	Authors have proposed a multi-layer perceptron (MLP) algorithm for the classification of the breast cancer and achieved more than 97% accuracy

a huge data of diabetes. Saeedi et al. [50] presented a semi-automatic framework for the diabetes diagnosis by EHR. Authors have utilized total 3 years of the EHR data from the 10 different hospitals; the data was collected and handled by District Bureau of Health in Changning, Shanghai. Table 2 showing the research work done for the medical big data analysis using machine learning techniques.

4.2.2 Unsupervised Learning Models

In the unsupervised learning techniques, the model is trained with the help of known inputs only. Unsupervised learning models organize the data points by itself with the help of different clusters. Zhao et al. [12] have explained how a random sequencing-based unsupervised algorithm can be designed for the heterogeneous big data. Pereira et al. [58] presented a learning representation for the time series data for the detection of the anomaly. In supervised learning, labeling of data is huge and tiring task, especially in case of medical data for primary annotation a medical expert required which

itself crucial. For learning representation, authors have used variational recurrent autoencoder technique. For this unsupervised task, authors have choose ECG5000 electrocardiogram dataset. Ritu et al. [59] presented a novel framework to integrate the security aspects with the big data analysis and its effects on the decision making. The unsupervised learning techniques provide an advantage to the researcher that they can directly start working with the data without the help of any medical expert for the diagnosis a prediction; however, it is one ignorant part of the analysis and needs to be explore more.

4.3 Neural Networks Algorithms

The concept of ANN in machine learning computes the computational calculations on the bases of the mathematical models which try to function like human brain instead of the conventional computations. An ANN-based model consists of adaptable computational elements called neurons having some weights, these weights can be re-adjusted according to the errors. ANN is initially designed to mimic the human brain behavior in terms of decision taking, and it computes in the cognitive fashion with the help of categorization, adaptation and step-by-step learning. ANN computation is an analytical process; in the health-care domain, it can be used in almost all applications and areas. It can be applicable in the medical image as well as signal analysis, metadata-based diagnosis system, clinical information interpretations, disease prognosis analysis, drug delivery systems and in many more health-care areas [60].

ANN works as a computational tool with interconnected processing neurons responsible for information processing and produces the external stimuli [61]. Pap net is a commercially available application, which is based on ANN to assist in the screening of pap (cervical) smears for pap-test. A pap-test is used to check the cells collected from the uterine cervix to recognize cancerous changes [62, 63]. An early diagnosis can cure cervical cancer. Traditionally, the laboratories, with the help of microscopic lens and identify a few dozen of affected, abnormal cells from more than 50,000 cells in pap smear. With the help of Pap net [64], a significant accuracy is observed in the accuracy for identifying the cancerous cells.

Alexander et al. [65] presented a detailed study for the big data analysis for the heart attack prediction. Similarly, Gao [66] used deep learning techniques for diagnosing the abnormalities. For ECG measurements of total 11,572 subjects (1120 with heart disease and 10,452 without any history of heart attack) were evaluated. The results of neural network were compared with commonly used ECG diagnosis techniques. Neural networks were found 15.5% more sensitive than other state-of-the-art methods also, more sensitive in diagnosing techniques also [67]. ANN algorithms are the most accurate ones in predicting the chronic diseases like myocardial infarction, coronary artery disease, different types of tumors, liver cirrhosis, brain diseases; bowel disease, etc. ANN is also widely used for drug delivery system also. The National Cancer Institute (NCI) [64] has developed a neural network for drug delivery. The detail of the neural networks is explained in the following sections.

4.3.1 Neuro-Fuzzy (NF) Models

The healthcare domain is effects under the influence of the big data because of its volume, complexity, heterogeneity and dimension. The results of NF models on the medical big data give positive directions to the healthcare data analytics. The NF models for the big data analysis are slightly different from the normal common neural network in terms of weight connection and activation functions. Taher et al. [68] proposed a NF model for the dimension reduction and classification of the different medical datasets like thyroid dataset [69] and breast cancer dataset [70]. The huge complex medical data have details on the behavior actions for the decision making, forecasting for health instances, fraud detection in the medical insurance, etc. Vidhya and Shanmugalakshmi [71] proposed a multi-disease diagnosis system using a modified adaptive NF inference system (M-ANFIS). Authors have collected this huge medical data from multiple resources which include EHR, government agencies, patients portals, pharmaceutical platforms, search engine and social networking platforms, etc. in the fast-growing internet uses for society, people have started searching everything related to their healthcare on internet and it produces a medical huge data. Therefore, the medical data privacy has become crucial. Shi et al. [72] presented a privacy risk assessment model with the help of medical big data using Adaptive NF models. Authors have analyzed the major indicators which can affect the security and privacy of the medical big data. User's search history, interaction records are some of the key factors which were addressed.

4.3.2 Recurrent Neural Networks (RNN)

RNN is a method which learns with its past selections. To mimic the physician's clinical behaviors, an RNN-based model is proposed which is called Doctor AI [73]. Doctor AI diagnosis the future potential disease along with the doctor's perspective for the occurrence of the disease. The model is trained and tested on EHR data of around 250 K patients collected during 8 years. Kim et al. [74] proposed a RNN-based model for the imputation of the missing values in the healthcare big data. Authors had carried out an experiment for 12 consecutive years in 13 different medical fields. Total 11,500 subjects were examined every year, and total 7400 subjects missed the routine examination sometimes. RNN is very much useful for the chronic disease like diabetes or mental health cohorts. Talebi et al. [75] proposed a model of long short-term memory (LSTM)-based model to identify the healthcare trajectories. Authors have presented model named Deep Care to predict the future medical condition based on the EMR data.

4.4 Evolutionary Algorithms

In complex medical problems, the evolutionary computation [19, 76–78] helps to determine many medically optimized solutions. The conventional medical diagnosis methods search in a huge and complex space for decision making. For big data analysis, the evolutionary methods determine a relationship between the data attributes [79]. Evolutionary algorithms provide a high scalability and flexibility. These algorithms are also capable to solve global optimization problems as well; however, these evolutionary algorithms are not much popular among the researchers.

Synchronism is achieved in an algorithm when all fittest nodes are sent and accepted at one time as in synchronous island model [77]. All islands wait till the last island undergoes its evaluation of sub-problems to move on to next generation. It is thus known to slow down the throughput of an algorithm. To update the cells of a Cellular Evolutionary Algorithms (CEA) [77] sequentially having a population in a 2-D grid, there are many ways and the most general one is independent random ordering of updates in time while randomly choosing the next cell to be updated. Other three methods are fixed line sweep, fixed random sweep and new random sweep.

5 Limitations and Strengths

In order to handle the massive medical big data, the healthcare experts are in great hope from technical experts to develop a system that can manage and investigate the data for the right solution. To analyze the medical big data, the technologies like graph analysis, Hadoop, Apache-spark, Natural Language Processing (NLP), etc. are popular among the data scientists and medical experts. Developing correlations between the factors and features for decision making is one main purpose of these algorithms. Therefore, a great opportunity is there to promote the data science in the health-care, also, it is now becoming a foundation for the learning-based health-care system [61].

Handling a huge amount of healthcare data is itself a big challenge [20]. Five years ago only the health-care system of USA had estimated data was 150 Exabyte, now it have become to the range of zetabytes and yottabytes. The visualization of patient information in a summarized and correct way to the practitioners is also not so apparent, on the same time prediction and diagnosis are also vague. The main challenge that body of every patient reacts differently also, the symptoms and biomarkers are mandatory but sufficient for the accurate disease diagnosis [61]. The traditional methodologies are almost same for the big data analysis in health-care. The only difference between these tools is user interface console; however, these tools are not so user friendly for the medical experts and patients as well. On the same time, it involves the extensive and complex programming, which required continuous help

of software engineers. In medical big data analytics, complexity starts with the data acquisition and remains until the data is processed and analyzed.

Authentication is required when data comes from nodes and sensors. This is a very challenging job to map this type of unstructured HL7 coding standard [13]. To promote the e-health services, standardization of the whole process is must; however, there is still need of the collective recognition from the standardization bodies like Protocol for Smart Objects (IPSO), etc. [14]. On the same time, it's time now for a research breakthrough to handle the securities issues [14]. With the sensor nodes, network configuration and synchronization of time with online analysis is also one major area which required efforts and step-forward research [61].

6   Conclusion and Future Work

Presently, a range of health-care and biomedical tools such as genomics, mobile biometric sensors, and smart-phone apps are available which produce a large data. Hence, it is obligatory to be familiar with the possible predictions and analysis that can be achieved using this data. For instance, the study and analysis of such big data can give additional insights like technical, scientific, therapeutic and other types of advancements and development in health-care.

This review concludes that with the help of big data analytical tools for health-care are helping for developing the patient specific and personalized medical tools. The combined analysis of electronic health and medical records as well as other medical data incessantly helping to develop an improved prognostic framework. In this review paper, we mainly focus on the already developed big data analytical platforms and used for the application of health-care. The big data of health-care have its own major issues because of huge user generated data, but maximum of these issues are known. Since in this paper we have consolidated various of health-care big-data mining techniques; therefore, disease wise and/or disease dig data analytical systems can be streamlined and in future for the real time problems used directly. For an example, big data from the diabetes patients can be can be narrow down for the scope of diseases which are related to diabetes, like liver disease, Alzheimer's disease, etc. These more appropriate and focused studies will definitely help researchers to identify the limitations and gaps for the future work.

The medical data are growing exponentially, and this huge data from the various application areas demands to design and develop innovative techniques and strategies to analyze and interpret such massive data within a limited time-frame. The big health care data extracted and analysis by the integration of sensors designed for the specific part of the body have witnessed a great growth in last few years only. Now, development of a detailed model of the body by combining the data of physiological parameters, data of medical imaging and omics techniques may be the next big area of research. This idea will enhance our knowledge related to the disease prognosis and will definitely help to develop new more accurate diagnosis tools. There are

great opportunities in all step of this extensive big data analysis process to introduce systemic improvement in the health-care research.

References

1. Raghupathi W, Raghupathi V (2014) Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2:1–10
2. Yangam SH (2020) Big data in healthcare: COVID-19, EHRs, uses and challenges. Sch J Econ Bus Manag 7(12):414–430
3. Shafqat S, Kishwer S, Ur R, Junaid R (2018) Big data analytics enhanced healthcare systems: a review. J Supercomput 76:1754–1799
4. Agenda R (2014) Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. Health Aff 33(7):1163–1170
5. Cardenas AA, Manadhata PK, Rajan SP (2013) Big data analytics for security
6. Dinov ID (2016) Methodological challenges and analytic opportunities for modeling and interpreting big healthcare data. Dinov GigaSci 5(12):1–15
7. Andreu-Perez J, Poon CCY, Merrifield RD, Wong STC, Yang GZ (2015) Big data for health. IEEE J Biomed Health Inform 19(4):1193–1208
8. Cort R, Bonnaire X, Marin O, Sens P (2015) Stream processing of healthcare sensor data: studying user traces to identify challenges from a big data perspective. 52:1004–1009
9. Cao P, Badger EC, Kalbarczyk ZT, Ravishankar K, Withers A, Slagell AJ (2015) Towards an unified security testbed and security analytics framework. pp 1–2
10. Archenaa J, Anita EAM (2015) A survey of big data analytics in healthcare and government. Procedia—Procedia Comput Sci 50:408–413
11. Pentland A (2013) Big data and health: revolutionizing medicine and public health
12. Zhao J, Papapetrou P, Asker L, Boström H (2017) Learning from heterogeneous temporal data in electronic health records. J Biomed Inform 65:105–119
13. Torino PDI (2020) Data mining techniques for complex user-generated data
14. Islam SMR, Kwak D, Kabir H (2015) The internet of things for health care: a comprehensive survey. IEEE Access 3:678–708
15. Huang T, Lan L, Fang X, An P, Min J, Wang F (2015) Promises and challenges of big data computing in health sciences reference: to appear in: revised date: graphical abstract. Big Data Res 2(1):2–11
16. Cunha J, Silva C, Antunes M (2015) Health Twitter big data management with Hadoop framework. Procedia—Procedia Comput Sci 64:425–431
17. Ling ZJ et al (2014) GEMINI: an integrative healthcare analytics system. VLDB Endow 7(13):1766–1771
18. Neto S, Ferraz F (2015) Disease surveillance big data platform for large scale event processing. In: International conference on internet computing (ICOMP), no. November, pp 69–80
19. Kaggal VC et al (2016) Toward a learning health-care system—knowledge delivery at the point of care empowered by big data and NLP. Biomed Inform Insight 8:13–22
20. Ng K, Ghoting A, Steinhubl SR, Stewart WF, Malin B, Sun J (2014) PARAMO: a parallel predictive modeling platform for healthcare analytic research using electronic health records. J Biomed Inform 48:160–170
21. Evaluation A (2002) Providing concept-oriented views for clinical data using a knowledge-based system. J Am Med Inform Assoc 9(3):294–305
22. Talbi E (2016) A taxonomy of hybrid metaheuristics. J Heuristics 8:541–564
23. Borthakur D (2007) The Hadoop distributed file system: architecture and design
24. Yao Q, Tian Y, Li P, Tian L (2015) Design and development of a medical big data processing system based on Hadoop. J Med Syst 39:22–32

25. Augustine DP (2014) Leveraging big data analytics and Hadoop in developing India's healthcare services. Int J Comput Appl 89(16):44–50
26. Rodger JA (2016) Discovery of medical big data analytics : improving the prediction of traumatic brain injury survival rates by data mining patient informatics processing software hybrid Hadoop hive. Inform Med Unlocked 1:17–26
27. Kalyankar GD (2017) Predictive analysis of diabetic patient data using machine learning and Hadoop. In: International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), pp 619–624
28. Rathore MM, Paul A, Ahmad A, Anisetti M, Jeon G (2017) Hadoop-based intelligent care system (HICS): analytical approach for big data in IoT. ACM Trans Internet Technol 18(1):8.1–8.24
29. Zaharia M, Xin R, Wendell R (2016) Apache Spark: a unifed engine for big data processing. ACM Commun 59:56–65
30. Gopalani S, Arora R (2015) Comparing apache spark and map reduce with performance analysis using k-means. Int J Comput Appl 113(1):8–11
31. Saouabi M, Ezzati A (2017) A comparative between Hadoop MapReduce and apache spark on HDFS. In: 1st international conference on internet of things and machine learning, pp 1–4
32. Nazari E, Shahriari MH (2019) BigData analysis in healthcare: apache Hadoop, apache spark and apache. Front Health Inform 8:92–101
33. Strickland NH (2000) CURRENT TOPIC PACS (picture archiving and communication systems): filmless radiology. Arch Dis Child 83:82–86
34. Kaur J, Mittal D, Singla R (2021) Diabetic retinopathy diagnosis through computer-aided fundus image analysis: a review. Arch Comput Methods Eng, pp 1–39
35. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I (2017) Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J 15:104–116
36. Sengur A, Turkoglu I (2008) A hybrid method based on artificial immune system and fuzzy k-NN algorithm for diagnosis of heart valve diseases. Expert Syst Appl 35:1011–1020
37. Zheng B, Yoon SW, Lam SS (2014) Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. Expert Syst Appl 41(4):1476–1482
38. Purwar A, Singh SK (2015) Expert systems with applications hybrid prediction model with missing value imputation for medical data. Expert Syst Appl 42(13):5621–5631
39. Kaur H, Wasan SK (2006) Empirical study on applications of data mining techniques in healthcare. J Comput Sci 2(2):194–200
40. Wu X et al (2008) Top 10 algorithms in data mining. Knowl Inf Syst 14:1–37
41. Lerman I, da Costa JP, Silva H (2002) Validation of very large data sets clustering by means of a nonparametric linear criterion. In: Classification, clustering, and data analysis, pp 147–157
42. Gheid Z, Challal Y, Laboratoire M (2016) Efficient and privacy-preserving k-means clustering for big data mining. In: IEEE Trustcom/BigDataSE/ISPA, Tianjin, pp 791–798
43. Khanmohammadi S, Adibeig N, Shanehbandy S (2017) An improved overlapping k-means clustering method for medical applications. Expert Syst Appl 67:12–18
44. Haraty RA, Dimishkieh M, Masud M (2015) An enhanced k-means clustering algorithm for pattern discovery in healthcare data. Int J Distrib Sens Networks 2015:1–11
45. Xu D, Tian Y (2015) A comprehensive survey of clustering algorithms. Ann Data Sci 2(2):165–193
46. Tiwari SK, Walia N, Singh H, Sharma A (2015) Effective analysis of lung infection using fuzzy rules. Int J Bio-Sci Bio-Technol 7(6):85–96
47. Jindal A, Member S, Dua A, Member S, Kumar N, Member S (2018) Providing healthcare-as-a-service using fuzzy rule-based big data analytics in cloud computing. IEEE J Biomed Health Inform 22(5):1605–1618
48. Hillerman T, Carlos J, Souza F, Carla A, Reis B, Carvalho RN (2017) Applying clustering and AHP methods for evaluating suspect healthcare claims. J Comput Sci 19:97–111
49. Vijayakumar GMV (2018) Framework for cancer diagnosis using hidden Markov model and GM clustering. Wirel Pers Commun 102(3):2099–2116

50. Saeedi P et al (2019) Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the international diabetes federation diabetes atlas, 9th ed. Diab Res Clin Pract, vol 157, Nov 2019
51. Gupta P, Garg S (2020) Breast cancer prediction using varying parameters of machine learning models. Procedia Comput Sci 171:593–601
52. Luo G (2016) Automatically explaining machine learning prediction results: a demonstration on type 2 diabetes risk prediction. Health Inf Sci Syst 4(1)
53. Sujitha R, Seenivasagam V (2020) Classification of lung cancer stages with machine learning over big data healthcare framework. J Ambient Intell Humaniz Comput, pp 1–11
54. Zheng T et al (2017) A machine learning-based framework to identify type 2 diabetes through electronic health records. Int J Med Inform 97:120–127
55. Zheng T, Zhang Y (2017) A big data application of machine learning-based framework to identify type 2 diabetes through electronic health records. Commun Comput Inf Sci 731:451–458
56. Taylor RA et al (2016) Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach. Acad Emerg Med 23(3):269–278
57. Al-Shargabi B, Alshami F, Alkhawaldeh R (2019) Enhancing multi-layer perception for breast cancer prediction. Int J Adv Sci Technol, pp 1–11
58. Pereira J, Silveira M (2019) Learning representations from healthcare time series data for unsupervised anomaly detection. In: IEEE international conference on big data and smart computing (BigComp)
59. Chauhan R, Kaur H, Chang V (2021) An optimized integrated framework of big data analytics managing security and privacy in healthcare data. Wirel Pers Commun 117(1):87–108
60. Miah SJ, Hasan J, Gammack JG (2016) On-cloud healthcare clinic: an e-health consultancy approach for remote communities in a developing Country. Telematics Inform 34:311–322
61. Agenda R (2014) Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. Health Aff 33:1163–1170
62. Leary TJO et al (1998) PAPNET-assisted rescreening of cervical smears rescreening strategy. Jama Am Med Assoc 279(3):235–237
63. Schreiber K, Elgert PA, Mango L (1994) Evaluation of the PAPNET™ cytologic screening system for quality control of cervical smears. Am J Clin Pathol 101:220–231
64. Montavon G (2020) Introduction to neural networks. Lect Notes Phys 968:37–62
65. Alexander CA, Solutions H, Wang L (2017) Big data analytics in heart attack prediction. J Nurs Care 6(2):1–10
66. Gao X (2021) Diagnosing abnormal electrocardiogram (ECG) via deep learning. In: IntechOpen, pp 1–16
67. Martı JP, Laguna P, Lyon A, Minchole A, Rodriguez B (2018) Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances. J R Soc Interface 15:1–18
68. Azar AT, Hassanien AE (2015) Dimensionality reduction of medical big data using neural-fuzzy classifier. Soft Comput 19(4):1115–1127
69. UCI Machine Learning Repository. [Online]. Available: http://archive.ics.uci.edu/ml/index.php. Accessed: 09 Feb 2021
70. Mangasarian OL, Street WN, Wolberg WH (1995) Breast cancer diagnosis and prognosis via linear programming. Oper Res 43(4):570–577
71. Vidhya K, Shanmugalakshmi R (2020) Modified adaptive neuro-fuzzy inference system (M-ANFIS) based multi-disease analysis of healthcare big data. J Supercomput 76(11):8657–8678
72. Shi M, Jiang R, Zhou W, Liu S (2020) A privacy risk assessment model for medical big data based on adaptive neuro-fuzzy theory. Secur Commun Networks
73. Choi E, Bahadori MT, Schuetz A, Stewart WF (2016) Doctor AI: predicting clinical events via recurrent neural networks. Mach Learn Healthc 56:1–18
74. Kim HG, Jang GJ, Choi HJ, Kim M, Kim YW, Choi J (2017) Recurrent neural networks with missing information imputation for medical examination data prediction. In: 2017 IEEE international conference on big data and smart computing, BigComp 2017, pp 317–323

75. Talebi SM, Ayatollahi A, Moosavi SMS (2021) A novel iris segmentation method based on balloon active contour. Wirel Pers Commun 117:87–108
76. Hamet P, Tremblay J (2017) Artificial intelligence in medicine. Metabolism 69:S36–S40
77. Alba E, Giacobini M, Tomassini M, Romero S (2002) Comparing synchronous and asynchronous cellular genetic algorithms
78. Kaggal VC et al (2016) Toward a learning health-care system—knowledge delivery at the point of care empowered by big data and NLP. 8:13–22
79. Suresh A, Kumar R, Varatharajan R (2020) Health care data analysis using evolutionary algorithm. J Supercomput 76(6):4262–4271

Chatbots: Understanding Their World, Classification, and Their Current State

Amit Singh , **K. K. Sharma** , **Parveen Kumar** ,
and Sandeep Kumar Vats

Abstract An extraordinary growth is being witnessed for mobile devices and their ever-growing computational powers. Also, the chip demand is very high. This demand is not only coming from the software industry, but almost all the other sectors also contribute to this enormous growth. With the recent innovations in chip manufacturing, complemented with very high demands, compute power costs are getting reduced at a significant rate. The computation power of chips fuels the artificial intelligence (AI) technologies to harness them. AI is a very computation resource demanding technology, and, in many cases, it requires specialized computational resources. Due to deep mobile penetration in the society, AI technologies such as Data Science and NLP are taking great pace. On the other hand, chatbots are gaining momentum beyond customer care and typical use cases. This paper analyzes the history, growth, and new trends of chatbots. Especially the impact of AI on bot technologies and the road ahead. This paper will take a look at the chatbot journey. It also examines where chatbots are making an impact on human lives.

Keywords Chatbot · Chatbot architecture · Sophia · AI · Rule-based · NLP · Data science · Computational power · XML · PDA

A. Singh · K. K. Sharma (✉) · S. K. Vats (✉)
Computer Science and Engineering Department, Sanskriti University, Mathura, Chatta, UP, India
e-mail: drkks57@gmail.com

S. K. Vats
e-mail: sandeepsvats2@gmail.com

A. Singh
e-mail: amit.singh@masterkeys.in

P. Kumar (✉)
Centre for Railways Information System, Chanakyapuri, Delhi, India
e-mail: psehrawat05@gmail.com

1   Introduction

Often, a question arises, what is chatbot, and why is it relevant for a discussion now?
In simple words, a *chatbot* is a software system used for conversation or to perform
an automated task.

A chatbot or bot is generally pre-programmed to answer a few questions. Alter-
natively, a bot may make a few *decisions* based on questions or commands. Their
response could be very intuitive and amusing, which is AI-based. Amazon's Alexa
is one good example (Fig. 1).

There is extraordinary growth in mobile devices [1]. This gives much-needed
computational power in the hands of users, and at the same time, per megabyte
storage prices are on the decline. This provides an opportunity for technologies to
exploit the situation. Chabot technologies are not behind and are growing in terms of
sophistication or complexity in analyzing input data and providing conversation-like
output.

2   The Era of Chatbots

In current times, chatbots have fetched much interest from various organizations.
Now, it is no longer merely a software tool for customer support. Its applications
are being observed in wide spectrum of use cases. Their usage includes but is not
limited to education, service delivery, information retrieval, e-commerce, marketing,
hospitals, government benefits, sales, and much more.

Modern chatbots are a point of attraction due to their applicability and suitability
to large set of use cases. Essentially, these software agents work together to provide
a service or interactive conversation using user-friendly interfaces. These agents
work in the background to provide data and service to the end-user interface. As

technologies are advancing, interfaces can be as simple as a conversation chat or as sophisticated as Sophia, as shown in Fig. 2. Sophia is a robot who was given citizenship since she can converse like humans! [2]. An esteemed organization, Indian Space Research Organization (ISRO) is working on a project to send a humanoid robot to space called Vyom Mitra. These chatbot agents work in the background to provide multidimensional output, actions, and service to the complex end-user interface. Before 2017, it did not get enough attention for analytical studies to find why people are using chatbots. Peter and Asbjørn, did excellent work in this area. In his paper, Peter states that, "The study identifies key motivational factors driving chatbot use. The most frequently reported motivational factor is productivity; chatbots help users to obtain timely and efficient assistance or information. Chatbot users also reported motivations pertaining to entertainment, social and relational factors, and curiosity about what they view as a novel phenomenon. The findings can help developers facilitate better human–chatbot interaction experiences in the future. Possible design guidelines are suggested, reflecting different chatbot user motivations." These were exciting findings, and they lay the foundation for having improved chatbot experiences in the time to come [3].
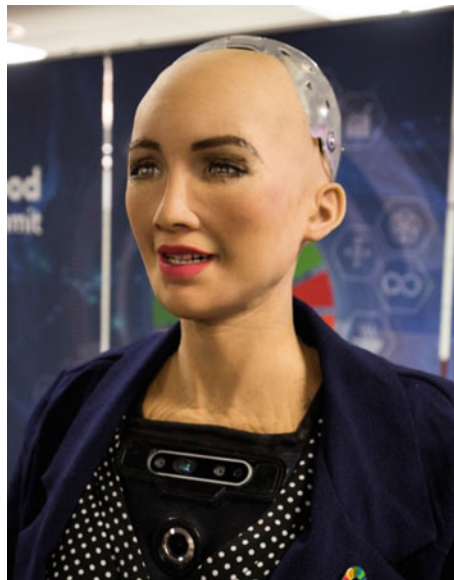
Fig. 2 Robot 'Sophia'

3   Chatbot Journey

A formal Chatbot journey was started in 1956. Here are few mentions of critical milestones in this journey.

1. Chatbots are around us for many decades; a notable example we found is ELIZA. ELIZA is regarded as the world's first chatbot. It was developed by Joseph Weizenbaum in 1956 [3]. Interestingly, it was emulating a psychotherapist by design. It also had a profound knowledge base in this specific domain. The major limiting factor at that time was hardware. Hardware in that era did not go great compared to today, neither in computing power nor storage perspective. These were a few primary factors that led this technology for decades to remain lagged.
2. In 1998, a markup language was well-defined by World Wide Web Consortium called extensible markup language (XML) [4]. As Wikipedia describes "XML is a markup language and file format for storing, transmitting, and reconstructing arbitrary data. It defines a set of rules for encoding documents in a format that is both human-readable and machine-readable" [5]. XML was free [6] and open-slandered [7]. It was an essential milestone as XML simplified interactions between machines and humans.

3. After a vacuum of a few years in chatbot technology, in early 2000, there were some exciting developments–Richard S. Wallace developed Artificial Linguistic Internet Computer Entity (ALICE). It is worth it to mention in this journey of chatbots. *This was, in fact, first use of AI in the form of artificial intelligence markup language (AIML).* Concepts of natural language processing (NLP) were used.

AIML is derived from XML and offers similar features. AIML enables programmers to input "*knowledge of dialogue patterns*" into chatbots. ALICE is now open-source software technology and widely used [8]. "ALICE bot was using *conversation* patterns. This information was stored in AIML files. ALICE relies on a vast number of basic 'categories or rules matching input patterns to output templates" [8, 9].

4   The Current State

Broadly, there are three types of chatbots nowadays, which are as follows:

(A)   Rule-based bot
(B)   AI-based bot
(C)   Hybrid, which combines the above two.
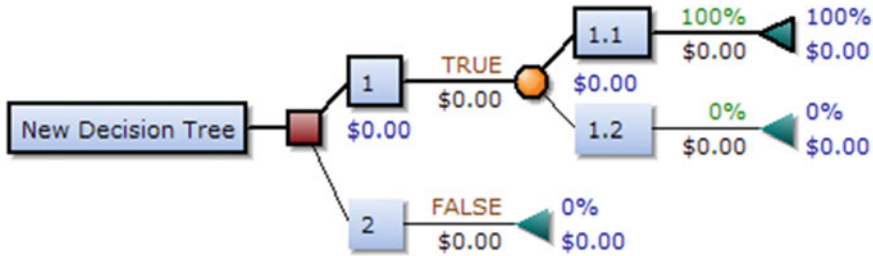
Fig. 3 Decision tree elements

With the progression of technology in time, hybrid categories are getting popular. For example, for home automation, we are using rule-based chatbots, but AI is being used when understanding the user's intent and optimizing user experience. AI leads to an impression of smartness in these implementations, which is often a pleasant experience for the consumer.

Let us have a closer look at the rest of the two types of chatbots.

(A) **Rule-based bot**

A decision tree[5] is a graphical structure that is like a tree-type model.

It is often used in algorithms that require conditions to process while traversing a tree, as shown in Fig. 3.

Rule-based type of chatbots uses decision trees. In bots, this technology is being used to create *adigital assistant*. Using this technique, a bot can answer from the frequently asked questions (FAQ) collection or for automation, and it can perform a set of tasks, e.g., switching off a preconfigured light at a certain time.

(B) **AI-Based bot**

This is the area where we see lots of technological momentum. The growing mobile device industry[11] over the last decade has generated an undeniable and heavy demand for AI chatbots to interact with users. Their popularity and adoption are increasing rapidly. The entire landscape of how we communicate has evolved due to the involvement of these mobile devices' capabilities. Multiple learning opportunities have presented themselves as well.
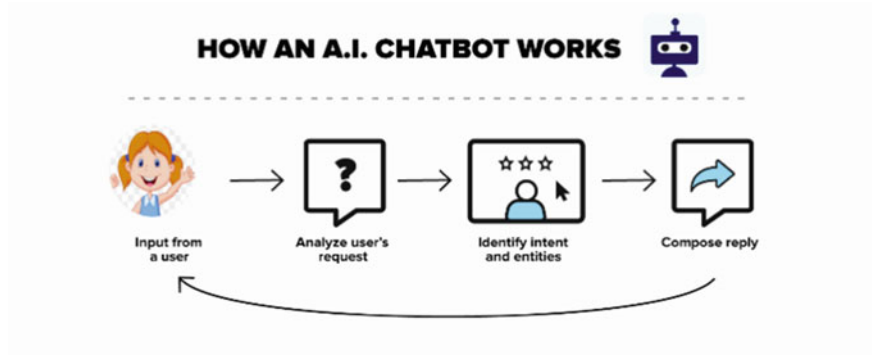
Fig. 4 AI chatbot in action

Personal digital assistants are the current trend of voice recognition and artificial intelligence. Some are "Siri" from Apple, "Alexa" from Amazon, Microsoft's "Cortana", or "Assistant" from Google. With the help of machine learning, they can perform basic day-to-day tasks just like a human (such as email prioritization, highlighting the most important content and interactions) to increase the overall efficiency of users (Fig. 4).

"Domain-specific" text-based chatbots are used for "target-specific functionalities" such as marketing content for publishing sites, asking for feedback, booking hotels, making a restaurant reservation, and so on. A specific flow or set of conditions is used to respond to queries from a client. This enables them to respond to pre-determined requests in any specific field. However, answering general questions presents a problem if the chatbot is not adequately trained. Examples of PDA include but are not limited to: Alexa from Amazon, Microsoft Cortona, Google Assistant, and many more. They use mean hardware coupled with cloud and AI technologies and provide a very rich experience to end users. Asking to play consumers their *favorite* song, driving *home,* or asking to *read their email,* all look very easy and intuitive using these. They are all examples of *relative* experiences for different people. Each one of us has a different favorite song, but these PDAs know that each one of us has a favorite song. Of course, this does require training for PDA bots. With the power of the cloud, the Internet, and modern AI technologies, this area is growing more and more personalized.

5 Applications of Chatbots

Few may say modern hybrid chatbots are still relatively new, but their popularity is skyrocketing amongst major businesses looking to automate customer service, marketing, or sales. The modern generation finds it very obvious to ask Alexa for a song or the latest news or asking google to call a number while driving. More and more people are ready to engage with chatbots, and interestingly, modern generation prefers this method over human representatives. For a future with chatbots, we must first understand these chatbots in depth, the different types of chatbots, and how they work.

What business sense they make—The need of the hour is accuracy and legitimacy at our fingertips. However, humans have their limits, which can only be taken further with the helping hand of the technology. Chatbots have been adopted by many of the forward thinking companies.

Increase revenue through the Web site—Organizations make lots of efforts to increase their web traffic, which eventually can get translated into a profit line. It is very expensive to receive lots of traffic via human interaction. To get a good sales number, organizations tend to deploy expensive resources and costly search engine optimization (SEO) techniques to drive traffic to their Web sites. All this can be optimized using chatbots with AI. It is believed that their traffic conversion is anywhere between 1 and 5%.
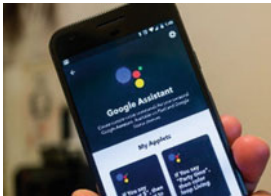
Provide business leads—In an ideal world, we wish we could verify every lead and check for compatibility seamlessly before spending precious organizational resources on it. In reality, many of the organizations find it impossible to get to an authentic business opportunity, especially if the web traffic is high. Modern bots can apply various modern techniques, for example, using advanced qualification logic and improving sales bottom line.

Handling Customer Pressure—Bots are the perfect solution to high-volume customer support inquiries, also called FAQs. At times, especially where customers often get irritated with standard knowledge bases that are hard to traverse to pinpoint required information. For example, finding the latest toy exclusively available on a popular e-commerce Web site.

Provide easy services to citizens—Government originations are using chatbots to provide services to citizens. This may start with a rule-based approach but continue

growing to provide an enhanced experience that could cover from FAQ in interactive mode. For example, Maitri [10] bot at Indraprastha Gas Limited (IGL), India Web site.

6   Chatbots Architecture and Their Limitations

This paper researched and was not able to come across open-source architecture and hence assumes that Application Programming Interfaces (API) provided by chatbot companies are closed source, and they may not necessarily follow simulation of the human mind.

Big IT organizations are all working to enhance chatbot technologies, which are often a mix of rule-based and AI-based approaches. Most successful chatbots are closed architecture; we simply do not know how they function. They expose a few APIs [11] to work on their chatbots. Chatbot development has many dimensions and strategies. Chatbot applications may use their own local architecture, which is often cumbersome and time-consuming. Another development approach is using a widespread chatbot API [12]. API approach is really fast to deploy to production if we use popular APIs. However, we lose flexibility in this trade-off between local architecture or API-driven development. Businesses and organizations may miss the opportunity to benefit from fast-paced development since technology implementations are limited by API usage only.

7   Opportunities Ahead

WhatsApp, a major chatting platform, had integrated BHIM APIs, a payment platform. This has created a very new dimension for chatbots. Conversational AI platforms such as Gupshup, Haptik, Verloop.io, and Yellow.ai are busy creating chatbots for various brands to provide an edge in selling their products to customers. Chatbots, AI, and FinTech technologies are in convergence at the current time. This is an exciting time for chatbot technologies.

Further, as Economic Times dated 01 Dec 2021 quotes, "Indians can now use WhatsApp to order groceries from billionaire Mukesh Ambani's JioMart via a new 'tap and chat' option, as his Reliance Industries Ltd. challenges the domination of Amazon.com Inc. and Walmart Inc.-owned Flipkart" [13].

8 Conclusion

Chatbot technology started at a very slow pace; however, ever-increasing compute power of the chips and their usage in mobile phones worked as an enabler, and we witness tremendous growth in this area. AI technologies are also progressing at a greater pace now. Their convergence is helping chatbots diversify their usage in the area not previously thought of. This process will propel new dimensions of chatbots. Thus, it would be easy to conclude that chatbots have a bright future. Businesses, organizations, and consumers can be the winners, simultaneously. This paper explored this journey and its classification. In the future, the work can be done on chatbot architecture in time to come, how their growth is still limited by their closed architectures, and how the convergence of technologies requires a new architecture for chatbots to facilitate exponential growth.

References

1. Smartphones Market—Growth, Trends, COVID-19 Impact, and Forecasts (2022–2027) https://www.mordorintelligence.com/industry-reports/smartphones-market. Retrieved on 15 Feb 2022
2. https://en.wikipedia.org/wiki/Sophia_(robot). Retrieved 18 Jan 2022
3. Brandtzaeg PB, Følstad A (2017) Why people use chatbots. In: Kompatsiaris I et al (eds) Internet science. INSCI 2017. Lecture Notes in Computer Science, vol 10673. Springer, Cham. https://doi.org/10.1007/978-3-319-70284-1_30. Retrieved 14 Jan 2022
4. Extensible Markup Language (XML) 1.0 from www.w3.org. Retrieved 14 Jan 2022
5. Weizenbaum J (1983) ELIZA—a computer program for the study of natural language communication between man and machine. Commun ACM 26:23–28. https://doi.org/10.1145/357980.357991. Retrieved 14 Jan 2022
6. XML as defined by Wikipedia. https://en.wikipedia.org/wiki/XML. Retrieved 11 July 2022
7. XML 1.0 Specification. https://www.w3.org/TR/1998/REC-xml-19980210.html. World Wide Web Consortium. Retrieved 16 Aug 2022
8. XML and Semantic Web W3C Standards Timeline (PDF). Dblab.ntua.gr. Archived from the original (PDF). Retrieved 14 Jan 2022
9. https://www.researchgate.net/publication/289684788_ALICE_chatbot_Trials_and_outputs ALICE Chatbot: Trials and Outputs. Retrieved on 15 Feb 2022
10. https://en.wikipedia.org/wiki/Decision_tree. Retrieved on 15 Feb 2022
11. Intelligent Chatbot APIs. https://nordicapis.com/15-intelligent-chatbot-apis/. Retrieved on 16 Sept 2022
12. API for Messenger Platform. https://developers.facebook.com/docs/messenger-platform/. Retrieved on 16 Sept 2022
13. https://economictimes.indiatimes.com/industry/services/retail/mukesh-ambani-turns-to-whatsapp-to-break-amazons-grip-on-indian-grocery-buyers/articleshow/87993512.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst. Retrieved on 02 Dec 2021
14. https://www.iglonline.net/english/Default.aspx. at home page left bottom. Retrieved on 15 Aug 2022
15. Dwarakanath N (2020) Gaganyaan mission: meet Vyommitra, the talking human robot that ISRO will send to space. India Today, 22 Jan 2020. Retrieved 15 Sept 2020
16. Google, Microsoft, Amazon, Alexa, Siri, "OK Google" and alike are respective trademark or companies of their respective owners

A Systematic Assessment on 3D-Based Deep Learning Models and Challenges in FER

Rajesh Singh and Anil Vohra

Abstract Face expression recognition (FER) is a widely emerging research area in today's era. The previous 2D-based techniques contain some drawbacks like illumination in data, variation in pose, and use ordinary feature extraction processes on scaled images. So, to overcome these kinds of hazards, 3D-based techniques have come to provide optimal solutions. The 3D-based techniques provide solutions for solving the face as well as video recognition problems with less pose variation problems and with optimal accuracy. This paper makes an attempt to provide a systematic review of various 3D techniques for FER with pros and cons, as well as different categories of emotions. The paper presents and analyses various existing FER-based models proposed by researchers and compares them with performance parameters. Furthermore, the paper provides a review of the different datasets used for FER.

Keywords Facial expression recognition · Emotions · 2D and 3D techniques · CNN and RNN

1 Introduction of FER

Facial expression recognition (FER) uses mathematical methods to analyse features of a person's face in images and videos [1]. It is a computer-based technology that is used to find solutions on the basis of 2D models. But the solutions based on 2D models were not completely acceptable for real-world situations because they present some problems of pose dissimilarity and illustration related to the nature of data [2].

But we all know technology has developed. Now there is 3D facial data, which helps in both still images and video sequences. This has also improved the accuracy of the FER system. Furthermore, we will analyse the limits and strengths of conventional and deep learning techniques [3]. When we talk about FER, a topic of sensation recognition, it includes the examination of human facial expressions in

R. Singh (✉) · A. Vohra
Department of Electronic Science, Kurukshetra University, Kurukshetra, India
e-mail: rsdeshwal@gmail.com

cross-media forms. Moreover, emotional recognition automatically processes human emotions not only from facial expressions but also from verbal expressions, body movements and gestures. We all know that facial expression is a way to express which is increasing day by day due to the broad set of its hidden applications such as supervision, safety and communication [4].

1.1 FER's Challenges and Opportunities

In the last several decades, several attempts have been made to improve FER algorithms for both practical applications and theoretical analysis. As per the FER literature, the main focus is to challenge the rough conditions of the environment where there are many opportunities and challenges that are listed and discussed here [5–7].

(i) Occlusion: It has been a discriminative model on several scales, with rapid utilisation for face recognition despite numerous impediments. Acknowledgement frameworks for halfway blocked faces have many details dependent on many inclusive hurdles and are arranged in a sequence having an approach of word reference.

(ii) Facial advances: In order to elaborate excellence, faces are recognised in addition to the event, which pictures contrast by constant variety of countenance. The process from coarse to fine was used for distinguishing the achievements of several postures in 3D faces. Several features extraction for less and huge details base requires the right way for warming photograph recognition, which requires the right amount of strategy to the framework of warm face technology. Warm photograph recognition and Gabor framework need primary consideration, which depends on climate for brightening, mainly in low resolution.

(iii) Identity bias: It is difficult to filter out age, gender and ethnicity from unstructured datasets.

(iv) Data: It is a challenging task to find and generate useful data from unstructured datasets.

(v) Pose variation: When an FER base dataset is processed, it is difficult to handle pose variations and it will impact the efficiency of the model.

2 Primary Emotions

In 1971, one of the great researchers named Ekman defined a set of emotions. According to him, universal emotions are wrath, repulsion, fear, happiness, surprise and sadness. Ekman and Friesen named these emotions as universal, irrespective of their culture and language. Ekman differentiated these emotions into seven categories. According to his view, human beings usually use a wider range of facial expressions for everyday communication in terms of these expressions [8] (Table 1).

Table 1 Emotions and descriptions [8]

Emotion	Description
Anger	It expresses negative feelings
Disgust	It expresses physical sense of human
Fear	It expresses nervousness of human when human feels unsafe
Happiness	It expresses happiness nature of human
Sadness	It expresses sad emotion of human
Surprise	It expresses unwanted situation of human

In 1978, Ekman and Friesen defined AUs as the actions of muscles generally seen at that time when an individual tries to produce a facial expression. They divide the facial features into two groups: the top of the face and the bottom of the face. As compared to the lower face, the upper face has little interaction with facial motion [9]. The boundaries of the lower face describe the movement in the nasolabial and nasal root regions, whereas the boundaries related to the top of the face define the movement and shape of the eyes, the brows and the cheeks. From the start, more than 7000 different combinations of action units have been seen, which gives philosophers a chance to look at the details of facial expression [10, 11] (Fig. 1).

Today, the most popular research topic for philosophers is still automatic AU recognition. Due to the lack of a FACS codes database, there is a disparity in favour of basic emotions, which is at the expense of AU recognition [11].
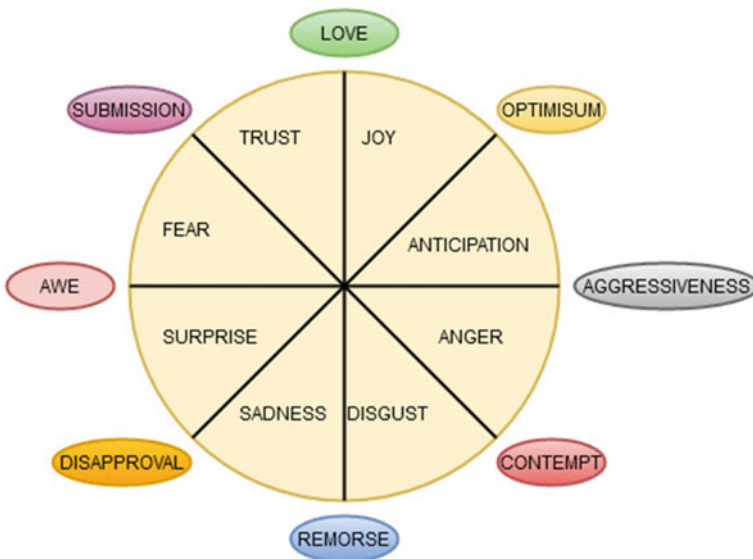
Fig. 1 Types of emotions [8]
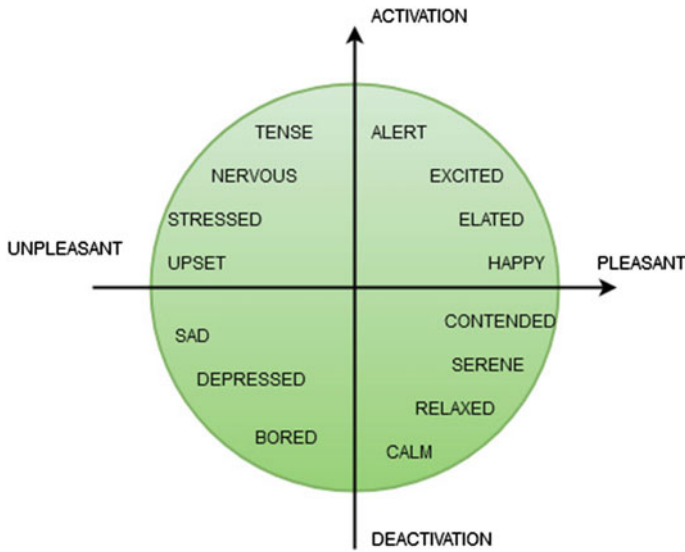
Fig. 2 Circumflex model of affects [12]

The 'Wheel of emotions', a model made by Plutchik, describes those emotions which can be stated as different strengths and can combine to form numerous emotions. Plutchik's eight basic emotions, called 'primary' are joy, trust, fear, surprise, sadness, anticipation, anger and disgust. There is also the existence of additional emotions such as aggression, optimism, love and submission. All these emotions are mixed to form primary emotions [12] (Fig. 2).

2.1 Parrot's Classification of Emotion

According to Parrot's model of emotion, emotions are categorised into two categories, i.e. primary and secondary emotions. Every secondary emotion is the induction of a primary emotion. He divided primary emotions into six parts, i.e. love, joy, sadness, anger, surprise and fear. All these primary emotions are further classified into secondary emotions [13], e.g. love can be divided into affection, lust, longing and joy, which can be divided into cheerfulness, contentment, optimism and relief. 'Surprise' is allocated or is the same as 'surprise'. Anger is split into irritability, exasperation, disgust and torment. Sadness is further classified into suffering, disappointment, shame, neglect, and the last one, fear, is divided into horror, nervousness.

Parrot's approach has identified a hundred feelings. He organised it into four parts, according to his circumflex model of affect, namely:

Activation, which further consists of four parts, i.e. alert, excited, elated and happy.

Pleasant, it further consists of four parts, i.e. relaxed, calm, contented and serene.

Fig. 3 Emotions classifications by Parrots' [13]
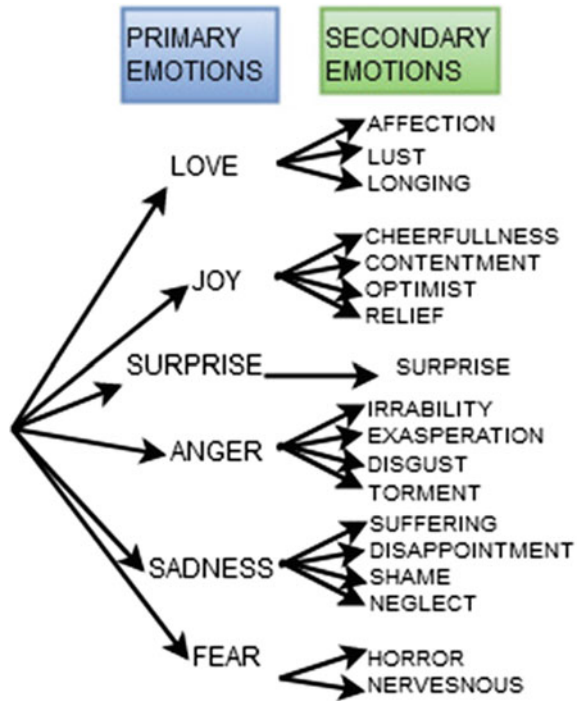
Deactivation, it also consists of further three parts, i.e. sad, depressed and bored. Unpleasant, it further consists of four parts, i.e. upset, nervous, stressed and tense.

The first layer is made up of six primary emotions, which can be sub-divided into different forms of sensitivity. The secondary emotions are the induction of the first ones (Fig. 3).

3 Facial Expression Recognition Techniques

In the present section, the primary methods for 3D facial expression recognition are given and compared based on features, algorithms that have many modals, and models. Thereafter, the main techniques of deep learning are applied in FER for extracting features and classification tasks as highlighted and differentiated on the basis of 3D and 2D.

3.1 Feature-Based Versus Model-Based Techniques

The present typical approaches for FER are based on static data in 3D and can be of two types, i.e. model-based algorithms and feature-based algorithms. Feature-based algorithms concentrate on extracting the spatial connection of a few interesting landmarks from the input data, such as geodetic and Euclidean distances, geometric facial surface information, local shape and gradient [13]. These traits are typically calculated on the area around primary facial landmarks, such as the eyes and mouth, which contain important details for emotion recognition.

Methods that are based on features are direct methods but possess two important demerits. Firstly, many such kinds of approaches require a set of accurately located areas for the extraction of features, which is another step of the procedure. In the past many years, this was considered a tough task in real-world applications as it needed the manual locations of areas for localization of features or registration of surfaces for marking this step, which has been automated. Also, a model's performance is directly related to the power of discrimination of features of faces that are adopted rather than shapes. Many of the tasks can be found in numerous publications that use the areas manually and label them as either Bosphorus dataset or BU-3DFE. Such experiments have shown their outstanding performance for identifying expressions universally, but many of the characteristics are yet to show sufficient power in order to distinguish subtle faces of AUs.

However, these days, it is not impossible to acquire 3D scans and also to take out geometric traits of several areas surrounding landmarks on faces. Instances of many famous expression features are the distance from 3D facial landmarks, 3D facial curves, facial geometry images, the distance between surface patches that are locally extracted and some normal maps. Another important technique exploits descriptors of 3D faces that are extracted from deep maps through the usage of mathematical operators: shape index, curvedness, mean curvature, first principal curvatures, etc. On the other hand, a generic face model is used and created for determining the neutral expression of emotions and measures the vector feature which is made by the deformation of coefficient of shapes. Such an approach requires bringing into the subsequent tracking model to form 3D face scans through a registration step.

3.2 Multi-modal Algorithms

Such algorithms that merge the outcomes from 2D and 3D data have not appeared since the early 2000s. Recently, in this regard, direct approaches have been used to combine features obtained independently from 2D or 3D methods, such as landmark location, curvature, texture information and facial shapes, in order to recognise and measure the intensity of expressions. Research shows that blending features acquired in separate modalities assists in catching the usual traits of deformation of the face

and also expands the accuracy of recognition despite the information on texture and has an impact on pose variations and illumination [14].

Earlier, 3D features were used to incorporate 2D texture information. After learning the SVM models from 3D and 2D data independently, classification is performed, and fusion of outcomes is undertaken independently for improving performance. Thus, the reduction technique of feature dimensionality is used for reducing the size of feature vectors while also retaining the quality.

Different visual cues which get combined with typical characteristics in improvising ways to estimate expressions of faces, such as head orientation, eye gaze, mouth fidgeting, FEs frequency or duration, and motion of the head and body. Many researchers have already shown that such a context can be improvised by recognising emotions and also the posture of the body, which makes it even more vital than that of FE, which is ambiguous. Also, such a voice can give many indications of emotions through acoustic properties like rhythm, amplitude, duration changes or pitch range [14].

Many studies were performed when 3D data of faces was combined with acoustic cues and physiological cues. Investigations are being done into future integration of visible and non-visible modalities such as functional data emerging from devices that are wearable, which is a possible arena of research in the upcoming years.

3.3 Deep Learning in FER

Deep learning is defined as a machine learning technique which came into theory for the first time in the 1980s and is recently considered as practise as it requires a huge amount of data which is labelled and also has sufficient processing power. Lately, effective learning methods are employed in a variety of tasks, which includes recognising facial expressions, which is a challenging issue for machine learning as users express their feelings in many different manners. DNN, or deep natural networks, are in use for the classification of pictures of human faces into many categories of emotions in the whole approach and thus overcome many obstacles in the typical ways of reaching a position of higher accuracy recognition than many activities of a human being [15].

Deep FER, which is automatic, includes three separate steps: deep feature learning, deep feature classification and preprocessing. Preprocessing is essential prior to training the neural network to learn important traits in recognising them. After completion of this initial phase, another deep learning method like CNN or RNN can be applied to FER, wherein such a phase is then categorised into the primary category of emotions. By using deep networks, the extraction of features and their categorization are performed in a direct way, whereas these are separate in the typical methods. Another way is to use a neural network, which is utilised for the extraction of features, and thus, to extract representations, independent classifiers like SVM or support vector machine are applied.

3.4 CNN

Many deep learning methods utilise the architecture of neural networks, and there is another renowned network that recognises faces or objects in CNN, or convolutional neural networks [16].

CNN is comprised of an inside layer, an outside layer and a maximum of 150 secret layers in between. That is why it is called 'deep'. However, old neural networks mainly have 2–3 layers. Such layers have the intention of finding out and learning characteristics; activation; repeating convolution; pooling operations; or ReLU, or rectified linear unit, where every secret layer enhances the sophistication of the characteristics of the images. The summary of the three steps is as below:

Convolution has been a mathematical operation which uses the function of a scanner and comprises the application of a small matrix which is known as a kernel in the particular matrices and the images. Every convolutional filter is turned into a different translation step, which also turns on many photograph features.

In the post-convolutional layer, there is always an activation function in many convNets. The essential purpose of this layer is to introduce nonlinearity in the system and also use functions of nonlinear like ReLU, sigmoid or tanh. Such rectified linear units go back to 0 when they do not receive any negative input. However, x gets back to its original value with any positive feedback.

Such pooling does nonlinear sampling, reduces the size of the outside matrices and also several parameters which can be understood by the network. Max Pooling and Average Pooling are the most common operations.

3.5 RNN-LSTM

RNN, or recurrent neural network, is a kind of modern superficial neural network which is mainly used in recognising speech. On the other hand, traditional neural networks have inputs and outputs that are not dependent on one another, where RNNs use information from sequences. This network depicts behaviour that is temporally dynamic and also utilises the inside memory for calculating output, which depends on many old computations.

A deviation of recurrent net and LSTM, which is long short-term memory, is used [17]. Such networks use many openings for regulation and controlling the cell state and also overcome the problems of absent gradients and exploding issues that are common in training RNNs. Such LSTMs retain memory for more time than RNNs and also enable it for modelling into long-term memory in an order that is also in major areas of application that involves recognising speech, modelling of language and also analysing video (video-based expression recognition tasks).
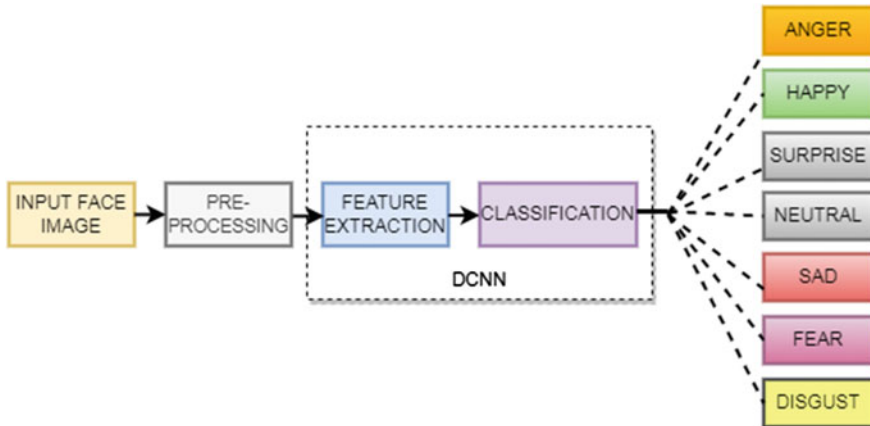
Fig. 4 CNN-based 3D FER system [5]

3.6 3D-CNN

The 3D CNN is used for video datasets. It consists of two phases, such as feature extraction and classification. The model is pretrained and tested with extracted features to process new datasets. Figure 4 depicts how the performance is computed based on the number of epochs. In the prepressing step, the frames from videos are filtered until no frames are left. After that, arrays of frames are created from the videos. In the next step, after all the preprocessing, the 3D CNN model is applied [5].

4 Comparative Analysis and Discussion

This section presents a comparison of feature-based algorithms and model-based algorithms in terms of various parameters such as database used, whether a 2D or 3D method is used, dynamic and expressions. Table 2 presents the comparison of feature-based algorithms, whereas Table 3 presents model-based algorithms. Furthermore, Table 4 presents a comparison of various deep learning models.

5 Datasets

There are a lot of datasets correlated to FER, available for extensive and comparative demonstrations; many of them are used for conventional FER appeal with the decision methods, others are based on this system of deep learning with the facial expression

Table 2 Comparison of existing feature-based models

Author	Dataset	2D	Dynamic	Expressions
Berretti et al. [18]	BU-3DFE	No	No	6 BE
Maalej et al. [19]	BU-3DFE	No	No	6 BE
Berretti et al. [20]	BU-3DFE	No	No	6 BE
Le et al. [7]	BU-4DFE	No	Yes	3 (happiness, sadness and surprise)
Li et al. [21]	BU-3DFE	No	No	6 BE
Sandbach et al. [22]	BU-4DFE	No	Yes	3 (anger, happiness and surprise)
Li et al. [23]	BU-3DFE	No	No	6 BE
Rabiu et al. [24]	BU-3DFE and UPM-3DFE	No	No	6 BE
Sandbach et al. [25]	BU-4DFE	No	Yes	6 BE
Rajamanoharan et al. [26]	Bosphorus and D3DFACS	No	Yes	22 AUs

Table 3 Comparison of existing algorithms based on models

Author	Dataset	2D	Dynamic	Expressions
Ramanathan et al. [27]	Private	No	No	(happiness, sadness and anger)
Gong and Wang [28]	BU-3DFE	No	No	6 BE
Zhao et al. [29]	BU-3DFE	No	No	6 BE
Zhao et al. [30]	Bosphorus	No	No	16 AUs
Fang et al. [31]	BU-3DFE and BU-4DFE	No	Yes	6 BE
Chen et al. [32]	AVEC	Yes	Yes	6 BE

Table 4 Deep learning models

Deep learning models	Dataset	Data type	Expression
VGG 16	JAFFE	Image	6 BE * 4 levels of intensity + Neutral
AlexNet	FERG-DB+ FRGG	Image	6 BE + Neutral
GoogLeNet	JAFFE+ CK+	Image	6 BE + Contempt and neutral + AUs
EfficentNet	FRGC ver2.0	Image	6 BE * 4 levels of intensity + Neutral
ResNet	FRGC ver2.0+ CK+	Video Type	6 BE * 4 levels of intensity + Neutral
Xception	FERG	Image + video	6 BE * 4 levels of intensity + Neutral
Inception	CK+	Image + video	6 BE + Neutral

Table 5 Comparison on different existing datasets

Dataset	2D data	3D data	Expressions	Occlusions	Head poses
JAFFE	Yes	No	6 BE + Neutral	No	No
FRGC ver2.0	No	Yes	6 BE	No	No
MMI	Yes	No	6 BE + Neutral	No	No
BU-3DFE	No	Yes	6 BE * 4 levels of intensity + Neutral	No	No
Bosphorus	Yes	No	6 BE + Neutral + 28 AUs	Yes (4)	Yes (13)
BU-4DFE	No	Yes	6 BE + Neutral	No	No
CK+	Yes	No	6 BE + Contempt and neutral + AUs	No	No
UPM-3DFE	No	Yes	6 BE + Neutral	No	No
FERG-DB	Yes	No	6 BE + Neutral	No	No

recognition conclusion. Previously, researchers studied facial expressions using 2D data that could be static or dynamic, posing numerous challenges. A 3D-based analysis of facial emotions will help to handle and facilitate notable pose variations and precise facial behaviour. But it consists of difficulties such as a high computational cost and data loss. The most common datasets used in different existing papers are discussed in Table 5.

6 Literature Survey

This article by Huang et al. [33] is a comprehensive overview of FER-based methods. This report provides a well-organised summary of the most current developments in FER. Existing FER methods were divided into two broad categories: traditional methods and those based on deep learning. Following this, a broad structure of a typical FER method was suggested, and the technologies that may be used in each part of the process were reviewed. Regarding deep learning techniques, four different types of cutting-edge FER strategies based on neural networks are introduced and analysed. To shed light on these innate issues, Li et al. [34] offered a thorough overview of deep FER, including everything from datasets to methods. First, we presented some of the most popular datasets from the aforementioned literature, together with some generally recognised criteria for selecting and evaluating such data. In addition, the typical steps involved in a deep FER system's pipeline were outlined, together with the relevant background information and ideas for appropriate implementations at each step. Using 3D data and 2D depth photographs, Hariri and Farah [35] sought to discern emotional facial expressions in isolation of the subject's identity. Because 3D FER is such a fine-grained recognition challenge, the performance of the FER might suffer if the 3D photographs are mapped into 2D depth images that lack some geometric properties of the expressive face.

However, 2D depth pictures have been effectively utilised to convolutional neural networks (CNN), which have enhanced handcrafted-based approaches in computer vision and pattern recognition applications. We show in this work that handmade features and deep learning features are complementary for 3D FER by combining the two. Covariance descriptors have shown promising results in their capacity to compactly describe a variety of characteristics from various sources. Because of this, it was recommended to employ covariance matrices of features (both manually created and deep) rather than features independently. Given that covariance matrices are a special kind of manifold space produced by symmetric positive definite (SPD) matrices, we have concentrated on generalising the RBF kernel to the manifold space for 3D FER by means of supervised support vector machine classification. When it comes to FER, Khan [36] provided a comprehensive overview. Literature is compiled from a variety of high-quality studies published in the last decade. The methods of traditional machine learning (ML) and many types of deep learning (DL) are the basis of this analysis. In addition, we examine and contrast the outcomes of several publicly accessible FER datasets for assessment measures. To shed light on the potential research deficit in this area, this study gives a comprehensive assessment of FER utilising conventional ML and DL techniques. In order to effectively categorise emotions from face photographs, Alsharekh [37] presented a DL approach based on a CNN model. To better analyse the compiled expressions the Viola–Jones (VJ) face detector generates, a new approach with an enhanced network architecture has been presented. In order to find the best possible model, a series of tests were conducted before the suggested model's internal architecture was finalised. This project's outcomes were achieved by a combination of subjective and objective efforts.

7   Conclusion

This systematic review provides an exploration and grouping of various existing 3D techniques that are focused on FER based on human emotions. We provide pros and cons of 2D- and 3D-based techniques such as feature based, CNN, DNN and RNN, and further compare different datasets on different parameters such as occlusion, emotion and type of data. We provide a systematic review of existing FER-based 3D modular techniques. Furthermore, a comparison between existing models has been presented with various parameters. After that, provide various datasets used by existing authors and compare them. In the future, it is intended to continue working on it and propose a mechanism based on transfer learning for FER on 3D datasets and compare performance with existing CNN and DNN models.

References

1. Nonis F, Dagnes N, Marcolin F, Vezzetti E (2019) 3D approaches and challenges in facial expression recognition algorithms—a literature review. Appl Sci 9(18):3904
2. Ko BC (2018) A brief review of facial emotion recognition based on visual information. Sensors 18(2):401
3. Li H, Sun J, Xu Z, Chen L (2017) Multimodal 2D+ 3D facial expression recognition with deep fusion convolutional neural network. IEEE Trans Multimedia 19(12):2816–2831
4. Bejaoui H, Ghazouani H, Barhoumi W (2017) Fully automated facial expression recognition using 3D morphable model and mesh-local binary pattern. In: International conference on advanced concepts for intelligent vision systems, Sept 2017, pp 39–50. Springer, Cham
5. Martinez B, Valstar MF (2016) Advances, challenges, and opportunities in automatic facial expression recognition. In: Advances in face detection and facial image analysis, pp 63–100
6. Kawulok M, Celebi E, Smolka B (eds) (2016) Advances in face detection and facial image analysis. Springer
7. Le V, Tang H, Huang TS (2011) Expression recognition from 3D dynamic faces using robust spatio-temporal shape features. In: Proceedings of the face and gesture 2011, Santa Barbara, CA, USA, 21–25 Mar 2011, pp 414–421
8. Ekman P, Friesen WV (1971) Constants across cultures in the face and emotion. J Pers Soc Psychol 17(2):124
9. Mishra B, Fernandes SL, Abhishek K, Alva A, Shetty C, Ajila CV, Shetty D, Rao H, Shetty P (2015) Facial expression recognition using feature based techniques and model based techniques: a survey. In: 2015 2nd international conference on electronics and communication systems (ICECS), Feb 2015, pp 589–594. IEEE
10. Danelakis A, Theoharis T, Pratikakis I (2015) A survey on facial expression recognition in 3D video sequences. Multimedia Tools Appl 74(15):5577–5615
11. Sariyanidi E, Gunes H, Cavallaro A (2014) Automatic analysis of facial affect: a survey of registration, representation, and recognition. IEEE Trans Pattern Anal Mach Intell 37(6):1113–1133
12. Deshmukh S, Patwardhan M, Mahajan A (2016) Survey on real-time facial expression recognition techniques. Iet Biometrics 5(3):155–163
13. Pantic M, Rothkrantz LJM (2000) Automatic analysis of facial expressions: the state of the art. IEEE Trans Pattern Anal Mach Intell 22(12):1424–1445
14. Corneanu CA, Simón MO, Cohn JF, Guerrero SE (2016) Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: history, trends, and affect-related applications. IEEE Trans Pattern Anal Mach Intell 38(8):1548–1568
15. Li S, Deng W (2020) Deep facial expression recognition: a survey. IEEE Trans Affect Comput
16. Revina IM, Emmanuel WS (2021) A survey on human face expression recognition techniques. J King Saud Univ-Comput Inf Sci 33(6):619–628
17. Ekman PE, Davidson RJ (1994) The nature of emotion: fundamental questions. Oxford University Press
18. Berretti S, Bimbo AD, Pala P, Amor BB, Daoudi M (2010) A set of selected SIFT features for 3D facial expression recognition. In: Proceedings of the 2010 20th international conference on pattern recognition, Istanbul, Turkey, 23–26 Aug 2010, pp 4125–4128
19. Maalej A, Amor BB, Daoudi M, Srivastava A, Berretti S (2010) Local 3D shape analysis for facial expression recognition. In: Proceedings of the 2010 20th international conference on pattern recognition, Istanbul, Turkey, 23–26 Aug 2010, pp 4129–4132
20. Berretti S, Ben Amor B, Daoudi M, Del Bimbo A (2011) 3D facial expression recognition using SIFT descriptors of automatically detected keypoints. Vis Comput 27:1021–1036
21. Li H, Morvan JM, Chen L (2011) 3D facial expression recognition based on histograms of surface di_erential quantities. In: International conference on advanced concepts for intelligent vision systems, Ghent, Belgium, 22–25 Aug 2011. Springer, Berlin, Germany, pp 483–494

22. Sandbach G, Zafeiriou S, Pantic M, Rueckert D (2011) A dynamic approach to the recognition of 3D facial expressions and their temporal models. In: Proceedings of the face and gesture 2011, Santa Barbara, CA, USA, 21–25 Mar 2011, pp 406–413

23. Li H, Chen L, Huang D, Wang Y, Morvan J (2012) 3D facial expression recognition via multiple kernel learning of multi-scale local normal patterns. In: Proceedings of the 21st international conference on pattern recognition (ICPR2012), Tsukuba, Japan, 11–15 Nov 2012, pp 2577–2580

24. Rabiu H, Saripan MI, Mashohor S, Marhaban MH (2012) 3D facial expression recognition using maximum relevance minimum redundancy geometrical features. EURASIP J Adv Sig Process 2012:213

25. Sandbach G, Zafeiriou S, Pantic M, Rueckert D (2012) Recognition of 3D facial expression dynamics. Image Vis Comput 30:762–773

26. Rajamanoharan G, Zafeiriou S, Pantic M (2012) Binary pattern analysis for 3D facial action unit detection. In: Proceedings of the British machine vision conference (BMVC), Guildford, UK, 3–7 Sept 2012

27. Ramanathan S, Kassim A, Venkatesh YV, Wah WS (2006) Human facial expression recognition using a 3D morphable model. In: Proceedings of the 2006 international conference on image processing, Atlanta, GA, USA, 8–11 Oct 2006, pp 661–664

28. Gong B, Wang Y (2009) Automatic facial expression recognition on a single 3D face by exploring shape deformation. In: Proceedings of the 17th ACM international conference on multimedia, Beijing, China, 19–24 Oct 2009, pp 569–572

29. Zhao X, Huang D, Dellandrea E, Chen L (2010) Automatic 3D facial expression recognition based on a Bayesian belief net and a statistical facial feature model. In: Proceedings of the 2010 20th international conference on pattern recognition, Istanbul, Turkey, 23–26 Aug 2010, pp 3724–3727

30. Zhao X, Dellandréa E, Chen L, Samaras D (2010) AU recognition on 3D faces based on an extended statistical facial feature model. In: Proceedings of the 2010 fourth IEEE international conference on biometrics: theory, applications and systems (BTAS), Washington, DC, USA, 27–29 Sept 2010, pp 1–6

31. Fang T, Zhao X, Ocegueda O, Shah SK, Kakadiaris IA (2012) 3D/4D facial expression analysis: an advanced annotated face model approach. Image Vis Comput 30:738–749

32. Chen H, Li J, Zhang F, Li Y, Wang H (2015) 3D model-based continuous emotion recognition. In: Proceedings of the 2015 IEEE conference on computer vision and pattern recognition (CVPR), Boston, MA, USA, 8–10 June 2015, pp 1836–1845

33. Huang Y, Chen F, Lv S, Wang X (2019) Facial expression recognition: a survey. Symmetry 11(10):1189

34. Li S, Deng W (2020) Deep facial expression recognition: a survey. IEEE Trans Affect Comput

35. Hariri W, Farah N (2021) Recognition of 3D emotional facial expression based on handcrafted and deep feature combination. Pattern Recogn Lett 148:84–91

36. Khan AR (2022) Facial emotion recognition using conventional machine learning and deep learning methods: current achievements, analysis remaining challenges. Information 13(6):268

37. Alsharekh MF (2022) Facial emotion recognition in verbal communication based on deep learning. Sensors 22(16):6105

Assessment on Different IoT-Based Healthcare Services and Applications

Rashi Rastogi and Mamta Bansal

Abstract The Internet of things (IoT) is the most advanced IT technology. All industries are attempting to capitalise on the benefits of this technology by incorporating it into their operations. The most common application area that has seen the benefits of implementing this technology and is benefiting greatly from its evolution is healthcare. This paper provides an analysis of services, software, and methodologies established on the foundation of IoT for the healthcare system, known as the Internet of health things (IoHT). IoT is heavily overlapping with the healthcare industry, transforming it into a smarter healthcare system. This paper attempts to perform a systematic assessment on various existing IoT-reliant healthcare systems based on parameters such as application area, hardware software requirements, and algorithms used. We present the IoT healthcare architecture, various healthcare services, applications, and different technologies in detail. Furthermore, future perspective challenges have been discussed.

Keywords Internet of things · Internet of health things (IoHT) · Healthcare system · Disease and monitoring

R. Rastogi (✉) · M. Bansal
Department of CSE, Shobhit Institute of Engineering and Technology (Deemed-to-be-University), Meerut, India
e-mail: rastogi.rashi4@gmail.com

M. Bansal
e-mail: mamta.bansal@shobhituniversity.ac.in

R. Rastogi
Sir Chottu Ram Institute of Engineering & Technology, Ch. Charan Singh University, Meerut, India

1 Introduction

Over the years, the medical management industry has showcased instant development and has been a foremost contributor to profit and business. In recent years, the healthcare revolution has shown fast growth and has been a major contributor to revenue and employment [1]. A couple of years back, the analysis of illness and deviations in human anatomy was only practicable after having a physical examination in a hospital. On the whole, convalescents had to linger in the health centre throughout their course of medication. The leading edge that has been attained over time has now authorised the identification of numerous ailments and keeps track of health using miniaturised tools like smart watches. Furthermore, automation has transformed hospital-centric medical management into a patient-centric organisation [2]. For instance, quite a few medical examinations (like blood glucose level, blood pressure count, pO_2 level, etc.) can be carried out at home without the assistance of a healthcare executive. Further, medical facts can be passed on to healthcare centres from rural areas using modern telecommunication facilities [3].

IoT has increased self-sufficiency while also varying the human's ability to connect with the external environment. IoT, by means of advanced mechanisms and processes, has grown into a leading provider of international engagement. It links a huge number of appliances, wireless sensors, home appliances, and electronic gadgets to the Internet. The implementation of IoT can be done in the fields of agriculture, automobiles, homes, and healthcare. The flourishing demand of the IoT is due to its dominance in presenting higher accuracy, lower cost, and its capability to anticipate forthcoming events in an efficient way. Furthermore, improved understanding of operating systems and applications, as well as improved mechanisation of mobile and computers, easy accessibility of Wi-Fi mechanisation, and improved Internet economy, has all contributed to the rapid revolt of IoT. The IoT gadgets (sensors, actuators, and so on) have been united with other substantial gadgets to record and interchange data by utilising the distinct communication conventions such as Bluetooth, ZigBee, IEEE (Wi-Fi), and so on [4]. In medical management applications, sensors either implanted or wearable on the human body are utilised to gather mental data like warmth, pressure rate, electrocardiogram (ECG), electroencephalogram (EEG), and many more like that from the patient's anatomy [5]. In addition, ecological data, namely temperature, humidity, date, and time, can also be reported. Collected statistics can be useful in creating substantial and accurate reasoning about the health issues of patients. Because a large number of facts are obtained from a variety of sources (sensors, mobile phones, e-mail, software, and applications), fact retention and ease of access play an important role in the IoT system. The details of the above-stated sensing instruments are put forward to doctors, caretakers, and approved associations [6]. The sharing of the aforementioned information with wellness programme providers via cloud/server allows for quick analysis of patients and medical intervention if necessary. The mutual support between the users, patients, and linked coursework is kept in existence for efficient and reliable communication. In particular, IoT structure makes use of a user's configuration that performs as a

progress report for medical caregivers and brings off user authority, data visualisation, and trepidation [7].

Many countries have accepted the latest machinery and courses of action, making the most of the functional ability of IoT in medical management structures. This changed the existing investigation in the medical management structure into an overly encouraging field to walk around in. Numerous nations have implemented new infrastructure and regulations to increase the usefulness of IoT in healthcare delivery. Due to this, healthcare research is now a more fruitful area of study.

1.1 IoT-Based Healthcare System Architecture

There have been many studies conducted for the framing and crafting of what the architecture of H-IoT should look like. One such five-stage framework is illustrated in Fig. 1 [8].

(i) *Physiological Sensor:* The projected architecture begins with physiological sagacity. In this phase, sensing of a variety of physiological actions and constraint by a refined electrophysiological sensor like pulsation velocity, body warmth, blood force, ECG, EEG, EMG, respiration pace, etc. The improvement of a flourishing IOT-supported healthcare system works on the basis of the physiological sensor.

(ii) *Processing and controlling:* Data that is pre-acquired through sensors and processed by the use of a microcontroller or processing unit is initially processed. Single-circuit integrated, small, and self-contained computers are called microcontroller boards with memory and a range of secondary devices
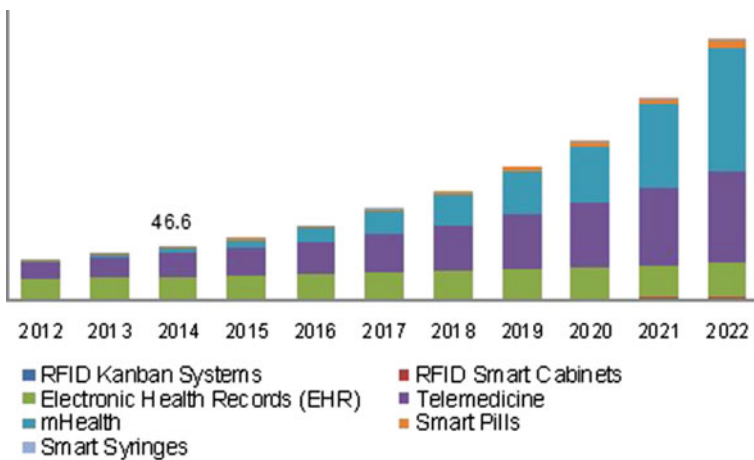
Fig. 1 Five-stage IoT-based smart healthcare system [11]

for input and output roles. Even with a restricted range of functions and power, it is eligible to carry out a variety of functions that are sovereign rational and arithmetical [9].

(iii) *Communication and transmission:* Union of the network to the applications is only possible through IoT's foundation, which is communication technology, enabling the coupled devices to transport information over the set of connections. The exchange of information among systems and devices configured in the healthcare setup is also done with the same medium. IoT's application and services identification can be short-range communications or long-range communications.

(iv) *Storage and computing:* It is an important aspect of the IoT healthcare device in terms of pre-attained data from the sensor. An on-demand facility and a great deal of space are provided by cloud computing. The information of each patient is required to be stored in order to re-access it as and when the patient re-visits the healthcare station, and subsequently, the information needs regular updating on every visit.

(v) *Data analytics and decision support system:* Once collected through sensors from patients, physiology-based data must be kept safe in order to be used in the future [10].

1.2   Contribution of Paper

The purpose of this paper is to provide a comprehensive review of the technologies, services, and applications that make the Internet of things possible in the healthcare sector and to summarise the progress of state-of-the-art studies in this field. Figure 2 depicts the rate of growth of the worldwide smart healthcare market from 2012 to 2022 [12].

1.3   Classification of Smart Healthcare System

The healthcare system is characterised as a conservative healthcare arrangement and a smart healthcare system, which aims at supporting and supplementing the systems that were followed traditionally in terms of ambient assisted living, remote healthcare systems, wearable healthcare systems, and smart phone-based healthcare systems. The intelligence and advancement are evident through IoT-based SHCS as all the systems are inter-linked and advanced medical techniques collect data and thereafter forward it to the cloud, where eventually, the data can be reached out to from any place, anytime. The experts in these fields no longer make personal visits to the patients and facilitate remote diagnosis and trace assets of medicine [13]. Figure 3 depicts the categorization of smart healthcare services. The full forms of used terms are as follows:
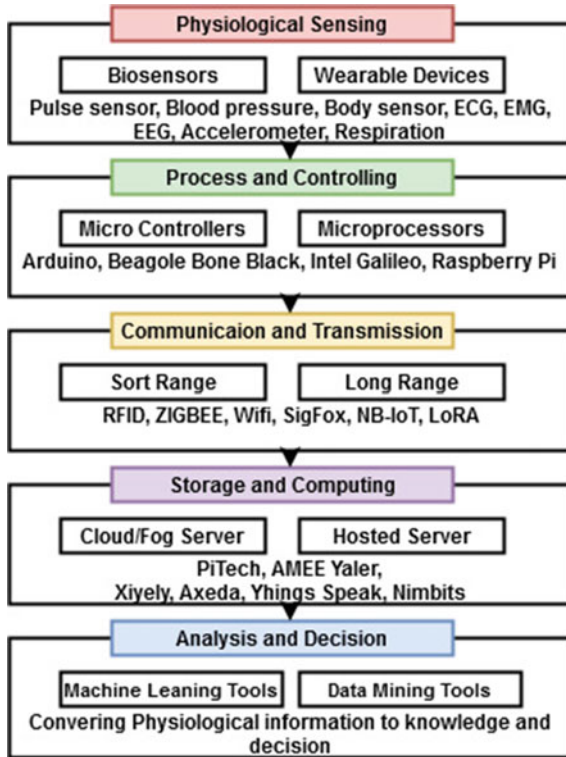
Fig. 2 Smart healthcare system based on the product [12]
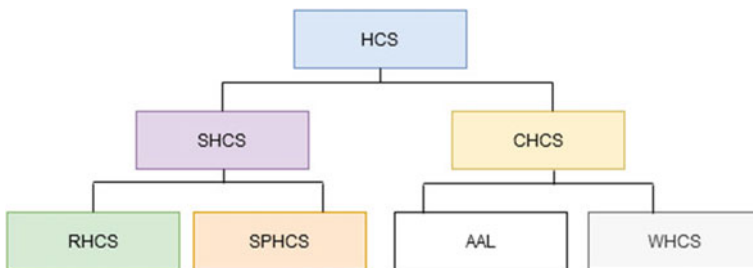
Fig. 3 Categorization of smart healthcare services [13]

HCS: Healthcare system
SHCS: Smart healthcare system
CHCS stands for conventional healthcare system
RHCS stands for remote healthcare system

SPHCS stands for smart phone healthcare system
ambient-assisted living
WHCS: Wearable healthcare system

Real-time information is conveyed to experts of the domain from a remote location with the help of multiple updated and advanced communication technologies. As the name designates, SPHCS employs a smart phone and smart applications for patient health monitoring either in a hospital setting or in home surroundings [14]. The goal of the individual healthcare system is to develop the self-rule of elderly, disabled, or incapacitated people in their space of living by providing services as a personal assistant [15].

1.4 IoT Technologies

It is not an easy task to integrate two entirely different domains to draw out the best outcome for both the chosen fields. Collaborating IoT with healthcare was too much of a tough nut to crack [16]. It required the adoption of multiple technologies to achieve the targeted outcome. The development of H-IoT is extremely complex, as this particular path of development flourishes the capability and efficiency of IoT. Such skills can roughly be characterised into three sets, namely identification, communication, and location technology [17]. Figure 4 depicts the technologies used in the healthcare system [18].

(i) *Identification Technology*: The work begins with the identification of sensors placed and allocated in the areas that are remote and distant from one another. It is basically the criteria to access the patients' information through a sanctioned

Fig. 4 IoHT technologies [18]
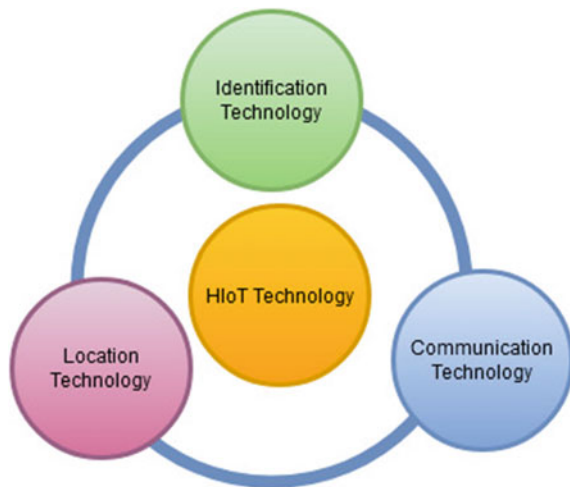
node. It is then followed by the allotment of a UID to each of the sanctioned authorised sensors so as to ease the process of identification and thereafter assist the exchange of unmistakable data.

(ii) *Communication Technology*: Technologies under this category support and facilitate smooth links between the various bodies involved in the H-IoT set of connections. There are two distinct types that we can divide this kind of technology into; namely short-range and medium-range communication technologies. Short-range ones are utilised to access links between objects in a restricted space or body area network. However, the other category assists communication over wider ranges, say the one between a station of base and a BAN's central sensor. In the case of short-range, the range can range from a few centimetres to a few metres.

(iii) *Location Technology*: The healthcare sector too has multiple bodies spread over different domains. Real-time location systems or location technologies are what help to judge and relocate the placement of something considered under the range of healthcare's set of connections and also serve as a support system for tracking the stage at which the treatment plan lies along with ascertaining the availability of resources at hand. The global positioning system is one such technology used in this specific domain, which takes the support of satellites to track specific things. But its applicability is possible as long as the object, and the four distinct satellites have a clear line of sight between them. It can be put to use to track the position of mobile healthcare services like ambulances, patients, or caregivers on their way to healthcare centres.

2 Services and Application of IoT

2.1 Services

There are many services being added on daily basis that add to the development of healthcare which has turned the tables upside down in regard to the solutions for the existing variety of problems. With the uplift in the demand, these all now have an essential role to play in entire H-IoT framework. Each and every service that relates to this domain brings to the forefront a big deal of healthcare solutions along with them [19].

(i) *Ambien-Assisted Living*: Not every aged person has someone to look after them, and, in that case, AAL serves the purpose right. It is a branch of IoT integrated with AI to assist ageing people. It works to make sure that elderly too have some independence in the premises of their homes so as to live in the way they want along with concerning about the safety and comfort too. It avails the elderly with real-time monitoring technique to ensure human-like serving assistance at the times of emergency.

(ii) *Mobile IoT*: It is now not always necessary to move physically to a healthcare centre to register for an appointment or wait for longer hours to get access of the service. Introduction of M-IoT has eradicated all such hindrances, and the healthcare officials can now get an access to the patients' information through the online data mode where the patient has already added all the required details.

(iii) *Wearable Devices*: Nowadays, many of the health assessments can be done through a device tied to the wrist of an individual. The innovation has made it possible. The sensors so attached to these devices are the ones that collect the information and ease the process of data collection. The collected data is influenced by the environment and the health conditions of the patients. The information so collected is then uploaded on the healthcare servers for further use. Some devices also avail connectivity to mobile phones as well, and the cost of healthcare is hence reduced to a great extent.

(iv) *Community-Based Healthcare Services*: In this domain of healthcare-based on IoT; the operations and the services are carried out and rendered to specific set of people or a group belonging to a particular community. It might also be based upon some confined areas of a city or town say a tiny area of residence, a hotel or so. The administrating medical experts at such centres or such networks are supposed to judge and take care of the people coming under their area of work. Multiple set of connections work together in this type of service to provide better services.

(v) *Cognitive Computing*: In this field of study and eventually practice; the system sensors are framed and formulated so as to replicate the processing of a human brain. The sensors so installed work in a way to first analyse the problem and then imitate the way a human brain works to resolve a problem and come out with possible solutions. With the advanced sensors and AI; it is possible to draw out conclusions or solutions from a complex and large set of healthcare data. Cognitive computing enhances the functioning in the further steps and corresponds effectively to the surroundings.

(vi) *Adverse Drug Reaction*: Every human body reacts differently to a certain drug or medication. In case of overdose or even when two medicines are taken at the same time; the effect can be noticed and is evaluated in terms of adverse drug reaction. This reaction can be observed in one go or even might take some regular intakes of doses to reflect the side effects of medicine taken. ADR does not rely upon the kind of medicine or even the ailment it has been taken for, but it does vary according to every different person. In this particular type of service; to ascertain each prescribed medicine for a patients' terminal; a distinctive bar code is brought into practical implication.

(vii) *Blockchain*: Anything that is available in fragments cannot yield an outcome that is reliable. Thus, in the field of healthcare as well; the need for secured and complete transfer of patient information between entities involved in the process of care-giving is immensely high. The issue of broken or fragmented

information across the domains of healthcare can lead to delayed or inappropriate treatment given to the patient. Blockchain assists to remove such barriers and create a link amongst the present data storeroom of the network.

2.2 Applications

(i) *ECG Monitoring.* It is a system that allocates the abnormalities in the functioning of the heart by evaluating the rate at which the heart muscles rhyme. It denotes the activity of heart in electrical format that is caused because of atria and ventricles' repolarisation and depolarization.

(ii) *Glucose Level Monitoring.* Amongst many other diseases existing among humans, diabetes is the most common one. The monitoring of level of glucose in the blood is done in case when the same sustains at a higher level for an elongated duration of time and the condition so formed is called diabetes. Type-I, type-2, and gestational diabetes are the commonly found types.

(iii) *Temperature Monitoring.* The first thing that a patient's condition is accessed upon is the temperature of the human body as it is indicative of preservation of homeostasis. It is an integral part of numerous investigative procedures. The doctors keep a regular check of the ups and downs in patient's body temperature to draw conclusions and make decisions on the health of the patient and provide the required treatment thereafter. It can be a warning sign in a few sicknesses like sepsis, trauma, and more like that.

(iv) *Blood Pressure Monitoring.* The evaluation of the pressure of blood in human body has become one basic step whenever a patient visits a doctor no matter what illness he/she is there for. The advancements in the IoT applications have brought a revolution in the way blood pressure is evaluated now as compared to the way it was done in the past. It is supposed to be accessed in a way where at least a person records it for the patient.

(v) *Oxygen Saturation Monitoring.* An essential parameter to be accessed while judging the overall health status of a patient is to see where the level of oxygen of the patient actually lies in medical terms. In healthcare diagnosis, it has become an important thing to analyse as it works on the basis of real-time monitoring and eradicates the traditionally existing issues in its evaluation process. The development of IoT devices related to this domain has accelerated its potential application in the field of medicine.

(vi) *Asthma Monitoring*: It is about monitoring the persistent difficulty in breathing and other issues that might hinder the airways; commonly termed as asthma. The windpipe might face shrinkage due to swelling at that specific area. The health issues that co-exist with this condition can include breath shortness, pain in the chest, continued cough, and many such. We still cannot be predictable about the time a patient might suffer from an asthma attack and the only solution we have at hand for this is the nebulizer or the inhalers meant for the same.

(vii) *Mood Monitoring*: Mood tracking provides vital information regarding a person's emotional state and is used to maintain a healthy mental state. It also assists healthcare professionals while dealing with various mental diseases such as depression, stress, bipolar disorder, and so on.

(viii) *Medication Management*: The elderly usually ignores the suggested medication due to several reasons. They might forget about it as per ageing effect or even might neglect deliberately. But, faithfulness towards the medicines being allocated by the doctors as a part of the treatment plan is highly recommended or else the illness may sustain for a longer duration or even might turn out to be more risky to the health of the individual. Medicines should not be missed by person of any age if he or she is under any medical treatment.

(ix) *Wheelchair Management*: IoT developments are now centred towards a new field of medical science that might prove out to be a boon to those with disability in moving and particularly the ones that are a result of brain damage. Wheelchair becomes an essential one for a patient with disability as it helps the patient to feel bit boundless mentally and physically too. But, a normal wheelchair cannot serve the purpose entirely in case of disability caused by damage in brain functioning so the above mention advancement needs to be paced up to ensure the achievement of this goal.

(x) *Rehabilitation System*: The solicitation of IoT rehabilitation is evident in many domains like while treating cancer, stroke, and injury caused by sports and few other disabilities too. It is a helping hand to the physical medicines prescribed to the patients as per their ailment. It helps the patients with disability; in restoration of their functional ability. Rehabilitation includes identification of the difficulty and thereafter assisting the patient to gain a normal life all over again.

(xi) *Other Notable Applications*: With the advancements in technology and also the way IoT has grown; its potential application has widened even throughout the fields that initially were untouched by its existence. The H-IoT applications are accelerating every day. Its use is no more restricted to the above-mentioned domains only; rather, it has crossed that line (Fig. 5).

3 Literature Survey

See Table 1.

3.1 Inferences Drawn from Literature

In this paper, a literature review of various IoT-based smart healthcare systems from 2018 to 2022 has been presented. The literature survey covers the applications, services, and technologies used for smart healthcare systems. The authors proposed

Fig. 5 Application of IoT [19]

various solutions such as HIDS and IoMRT. They analysed prose mechanisms using UCI datasets. Some researchers have proposed IoMT-based architectures.

4   Challenges, Limitations, and Future Scope

(i) *Servicing and Maintenance Cost*: Upgradation indeed is necessary for everything that we are surrounded with and so is the case with H-IoT-based devices. Belatedly, there are swift hi-tech progressions in this field that ask for the same every now and then. This system like any other system demands a great deal of maintenance, updating, servicing, and cost to be invested upon it to ensure required changes. The cost thus falls upon both the entities including the providers and the users of the particular apparatus. Setups that call for a low maintenance are highly preferable as the complete IoT system is network of multiple inter-linked sensors and devices. So as to keep the costs from crossing the budget line, it is required to keep a check on it [37].

Table 1 Existing techniques

Ianculescu et al. [20]	2018	Explained the use of several 'smart' products used for the detection and treatment of skin-related diseases
Ud Din et al. [21]	2019	Designed 5-layered architecture for IoT systems comprising of several components of IoT
Baig et al. [22]	2019	Review undertaken that presents development in wearable technology, the limitations, and challenges associated with wearable devices
Bhawiyuga et al. [23]	2019	Recommended an IoT system architecture that incorporates a middleware that functions as a bridge between wearable devices and cloud platform
Boulos et al. [24]	2019	Built up a model called geo AI that merges AI and geographic information of patients to provide better solutions in several healthcare segments
Darwish et al. [25]	2019	Developed an IoT cloud assimilation platform to healthcare, and analysed the associated issues and challenges
Das et al. [26]	2019	Developed a framework that is useful specifically for visually challenged patients using IoMT
Deshpande et al. [27]	2019	Carried out a detailed analysis of various tools and techniques used to analyse the data generated via 'smart devices' for decision-making purposes
Dhanvijay and Patil [28]	2019	Proposed real-time monitoring system and architectural framework for H-IoT
Dwivedi et al. [29]	2019	Proposed IoT topology and a conceptual model to tackle risks to the confidentiality and privacy of clinical data using hybrid technology
Guntur et al. [30]	2019	Proposed a conceptual paradigm, called IoMRT, which uses healthcare robotics. Analyse the subject and challenges concerned with the same
Yoon [31]	2019	Described the worth of using blockchain technology in healthcare sector
Koshimizu et al. [32]	2020	Developed a model that uses deep neural network to variation of blood pressure and predict the progress of the disease
Arulanthu and Perumal [33]	2020	Built OMDSS for the prediction of CKD by using classifier techniques to provide efficient healthcare and evaluated the performance of different classifier techniques using UCI dataset
Javaid and Khan [34]	2021	Studied various IoT technologies that were beneficial in COVIDpandemic
Johri et al. [35]	2022	People's understanding of skin disorders, awareness of IoT-based smart skin monitoring systems, and awareness of the benefits of adopting IoT-based smart skin monitoring systems were all critically examined. The study employs a quantitative technique
Saif et al. [36]	2022	Offers a hybrid intelligent intrusion detection system (HIIDS) based on machine learning and metaheuristic algorithms for Internet of things-based applications such as healthcare

 (ii) *Power Consumption*: Self-power generating setups are the ones that sciences are looking forward to. The count of medical devices that work on batteries is very high. Each one needs battery to work. As and when a sensor is turned on it is highly restricted to not to make any alterations in the battery. So, to let such systems work efficiently batteries that are high on power were installed. Energy breeding systems that are renewable to the core can be an option. The problem of global warming can be dealt with this proposed solution.

(iii) *Standardisation*: Authenticity in the medical field is the first and foremost priority for the operators and also to those at the receiving end of the service. With a wide range of entities producing health equipment's which allege legitimacy of their produce but end up violating the given norms that they are supposed to follow in terms of the quality, structure, and design. Therefore, it all needs a set of people so committed to make sure that no adulteration takes place in this regard, and maintenance of standard is highly assured through devices that rely upon set of rules of communication, statistics aggregation, and access boundary. The standardisation can be pulled off through formation of cluster of people or organisations that can merge with the experts of research to ensure standardised and validated devices with no lack in its genuineness.

(iv) *Data Privacy and Security*: The way monitoring is done presently which is entirely different from the way it was done in the past. And supposedly, it is an outcome or after-effect of cloud computing. But just like while mining for gold many more impurities too come along in the process; this advancement has also brought along many loopholes for the medical systems [34].

 (v) *Scalability*. Everything this world consists of; is supposed to walk hand in hand with the changes that happen around. And, so is the case with IoT. The devices so formed and designed to be used in IoT should inculcate some alterations as per the environment. The resources at hand can be efficiently used if they are adjoined with devices that are scalable. A big list of setups is used; such as sensors, which facilitate transfer of information. It thus makes things sorted for the present along with the future. If a system is low at uniformity, it ends up at less scalability as well; thus, efficient management is essential.

(vi) *Identification*. Exchange of information in terms of identity of all the entities involved in the process of medical assistance is fundamental.

Table 2 exhibition assessment between the formerly programmed papers from the point of application area, detection device, wireless communication technology, HW, and SW desires to control and sustain remote monitoring [38].

5   Conclusion

The current assessment looked into many areas of the H-IoT system. The architecture, components, and inter-component communication of an H-IoT system are discussed in detail. This study also examines current healthcare services that have been explored

Table 2 Comparative analysis between existing systems [38]

System	Application area	Detection device	Wireless communication technology	Algorithm	HW and SW need to operate
Voice pathology monitoring	Automatic voice pathology monitoring	Smart sensors like microphones	Bluetooth and radio access network or 5G wireless technology	Parallel deep learning	Smartphone downloads the app, cloud computing, server, and AlexNet
IAAS	Collision-resistant alarm system for blind, elderly, and patients with eye diseases	Wearable RFID	Radio frequency identification (RFID)	Locally weighted linear regression (LWLR) algorithm	COTS UHF RFID, R420 RFID reader, 8 dBi directional antennas, RFID tags, software
Wearable ECG	Electro-cardiograph monitoring system	Wearable territory of sensors	Bluetooth, 4G LTE, RF, wireless communication	Threshold training algorithm	MCU, ECG, SoC, analogue front end, BLE, cloud
Sensors heart disease mentoring	Heart disease mentoring	Sensors wear in wrist or carried by patients	Bluetooth, GSM, and GPRS, ADSL	Not specified	Blood pressure, ECG, SpO_2 sensors
Diabetes management based on IoT	Self-management of diabetes	Sensors used at regular intervals	GSM network and/or Wi-Fi network, Internet, Bluetooth	Not specified	Blood sugar, weight scale, and blood pressure
Aiding ambient-aided livelihood	Putting off fatness	Sensors attached to the body	Wi-Fi, Internet, Bluetooth	Administered learning	HR sensor, smart watch or smartphone

for potential IoT-based solutions. Internet of things (IoT) technology has helped healthcare professionals monitor and diagnose a wide range of health issues, measure a wide range of health indicators, and provide diagnostic services in previously inaccessible locations, thanks to these ideas. In doing so, it has reorganised the healthcare industry around patients rather than hospitals. We also addressed several H-IoT system applications and their recent trends. After that a comparative analysis of various existing healthcare systems has been presented with different performance parameters such as hardware and software requirements, algorithm, technologies, and application areas. Furthermore, the challenges and issues linked with the design, manufacture, and use of the H-IoT system have been addressed. These challenges will serve as the foundation for future advancement and research focus in the next few years. Furthermore, we present a comparative analysis between existing systems such as diabetes management sensors for health monitoring with various performance parameters.

In the future, it is intended to continue working on it and propose a mechanism to detect diseases in an automatic manner with high accuracy using machine learning.

References

1. Pradhan B, Bhattacharyya S, Pal K (2021) IoT-based applications in healthcare devices. J Healthc Eng 2021
2. Yassein MB, Hmeidi I, Al-Harbi M, Mrayan L, Mardini W, Khamayseh Y (2019) IoT-based healthcare systems: a survey. In: Proceedings of the second international conference on data science, E-learning and information systems, pp 1–9
3. Al-Rawashdeh M, Keikhosrokiani P, Belaton B, Alawida M, Zwiri A (2022) IoT adoption and application for smart healthcare: a systematic review. Sensors 22(14):5377
4. Ali Z, Hossain MS, Muhammad G, Sangaiah AK (2018) An intelligent healthcare system for detection and classification to discriminate vocal fold disorders. Future Gener Comput Syst 85:19–28
5. Khan M, Han K, Karthik SJWPC (2018) Designing smart control systems based on internet of things and big data analytics. Wirel Pers Commun 99(4):1683–1697
6. Nachankar PJ, Somani MG, Singh DM, Katkar SN (2018) IOT in agriculture. Decis Mak 1(3)
7. Menon VG, Jacob S, Joseph S, Sehdev P, Khosravi MR, Al-Turjman F (2022) An IoT-enabled intelligent automobile system for smart cities. Internet Things 18:100213
8. Mathew PS, Pillai AS, Palade V (2018) Applications of IoT in healthcare. In: Cognitive computing for big data systems over IoT. Springer, Cham, pp 263–288
9. Jagadeeswari V, Subramaniyaswamy V, Logesh R, Vijayakumar V (2018) A study on medical Internet of Things and Big Data in personalized healthcare system. Health Inf Sci Syst 6(1):1–20
10. Ahad A, Tahir M, Yau KLA (2019) 5G-based smart healthcare network: architecture, taxonomy, challenges and future research directions. IEEE Access 7:100747–100762
11. Dang LM, Piran MJ, Han D, Min K, Moon H (2019) A survey on internet of things and cloud computing for healthcare. Electronics 8(7):768
12. Froiz-Míguez I, Fernández-Caramés TM, Fraga-Lamas P, Castedo L (2018) Design, implementation and practical evaluation of an IoT home automation system for fog computing applications based on MQTT and ZigBee-WiFi sensor nodes. Sensors 18(8):2660
13. Birje MN, Hanji SS (2020) Internet of things based distributed healthcare systems: a review. J Data Inf Manag 2(3):149–165

14. Kadhim KT, Alsahlany AM, Wadi SM, Kadhum HT (2020) An overview of patient's health status monitoring system based on Internet of Things (IoT). Wirel Pers Commun 114(3):2235–2262
15. Selvaraj S, Sundaravaradhan S (2020) Challenges and opportunities in IoT healthcare systems: a systematic review. SN Appl Sci 2(1):1–8
16. Aftab H, Gilani K, Lee J, Nkenyereye L, Jeong S, Song J (2020) Analysis of identifiers in IoT platforms. Digit Commun Netw 6(3):333–340
17. Syed L, Jabeen S, Manimala S, Alsaeedi A (2019) Smart healthcare framework for ambient assisted living using IoMT and big data analytics techniques. Future Gener Comput Syst 101:136–151
18. Maskeliūnas R, Damaševičius R, Segal S (2019) A review of internet of things technologies for ambient assisted living environments. Future Internet 11(12):259
19. Geetha Poornima K, Krishna Prasad K (2020) Application of IoT in predictive health analysis—a review of literature. IJMTS 5(1):185–214
20. Ianculescu M, Alexandru A, Coardoa D, Coman OA (2018) General reports smart wearable medical devices—the next step in providing affordable support for dermatology practice. ICI, pp 41–46
21. Ud Din I, Almogren A, Guizani M, Zuair M (2019) A decade of internet of things: analysis in the light of healthcare applications. IEEE Access 7:89967–89979
22. Baig MM, Afifi S, Gholam Hosseini H, Mirza F (2019) A systematic review of wearable sensors and IoT-based monitoring applications for older adults—a focus on ageing population and independent living. J Med Syst 43(8)
23. Bhawiyuga A, Pramukantoro ES, Kirana AP (2019) A web of thing middleware for enabling standard web access over BLE based healthcare wearable device. In: 2019 IEEE 1st global conference on lifesciences and technologies (LifeTech)
24. Kamel Boulos MN, Peng G, VoPham T (2019) An overview of GeoAI applications in health and healthcare. Int J Health Geogr 18(1)
25. Darwish A, Hassanien AE, Elhoseny M, Sangaiah AK, Muhammad K (2019) The impact of the hybrid platform of internet of things and cloud computing on healthcare systems: opportunities, challenges, and open problems. J Ambient Intell Hum Comput 10(10):4151–4166
26. Das A, Rad P, Choo KKR, Nouhi B, Lish J, Martel J (2019) Distributed machine learning cloud teleophthalmology IoT for predicting AMD disease progression. Future Gener Comput Syst 93:486–498
27. Deshpande P (2018) Predictive and prescriptive analytics in big data era. Adv Intell Syst Comput 810:123–132
28. Dhanvijay MM, Patil SC (2019) Internet of Things: a survey of enabling technologies in healthcare and its applications. Comput Netw 153:113–131
29. Dwivedi A, Srivastava G, Dhar S, Singh R (2019) A decentralized privacy-preserving healthcare blockchain for IoT. Sensors 19(2):326
30. Guntur SR, Gorrepati RR, Dirisala VR (2019) Robotics in healthcare: an internet of medical robotic things (IoMRT) perspective. Machine learning in bio-signal analysis and diagnostic imaging, pp 293–318
31. Yoon HJ (2019) Blockchain technology and healthcare. Healthc Inform Res 25(2):59–60
32. Koshimizu H, Kojima R, Kario K, Okuno Y (2020) Prediction of blood pressure variability using deep neural networks. Int J Med Inform 136:104067
33. Arulanthu P, Perumal E (2020) An intelligent IoT with cloud centric medical decision support system for chronic kidney disease prediction. Int J Imaging Syst Technol
34. Javaid M, Khan IH (2021) Internet of Things (IoT) enabled healthcare helps to take the challenges of COVID-19 pandemic. J Oral Biol Craniofac Res 11(2):209–214
35. Johri A, Bhadula S, Sharma S, Shukla AS (2022) Assessment of factors affecting implementation of IoT based smart skin monitoring systems. Technol Soc 68:101908
36. Saif S, Das P, Biswas S, Khari M, Shanmuganathan V (2022) HIIDS: hybrid intelligent intrusion detection system empowered with machine learning and metaheuristic algorithms for application in IoT based healthcare. Microprocessors Microsyst 104622

37. lataSahu M, Atulkar M, Ahirwal MK (2020) Comprehensive investigation on IoT based smart HealthCare system. In: 2020 first international conference on power, control and computing technologies (ICPC2T). IEEE, pp 325–330
38. Bhatia K, Singh M (2019) Towards development of portable instantaneous smart optical device for haemoglobin detection non-invasively. Heal Technol 9(1):17–23
39. Su H, Ovur SE, Li Z, Hu Y, Li J, Knoll A, Ferrigno G, De Momi E (2020) Internet of things (IoT)-based collaborative control of a redundant manipulator for teleoperated minimally invasive surgeries. In: 2020 IEEE international conference on robotics and automation (ICRA). IEEE, pp 9737–9742

Energy-Efficient Solar Powered Raspberry Pi Pesticide Sprayer Robot for Agriculture Applications

P. A. Harsha Vardhini, N. Sindhu, G. Janardhana Raju, Kalyana Srinivas, and M. H. S. Vishnu Sai

Abstract Agriculture being the primary and major source for supporting the basic needs of people, it is also evident that agriculture stand as the back bone of our country India. Now, agriculture/farming is playing a predominant role in Indian economy too. Three-fourth of India's population is directly or indirectly dependent on farming. For ages, farmers accustomed with conventional and manual methods for seed sowing to harvesting in the farming process. Later with the advent of automation, new agricultural methods emerged to support farmer at all stages of farming. As a stage of farming, spraying pesticides to protect the crop and reap a good harvest plays a major role. Effective spraying techniques improved the benefits, and this works projects the design implementation of such a pesticide sprayer. Energy-efficient systems have gained the significance in all major fields. Systems with renewable energy sources gained the preference in many applications. An energy-efficient solar pesticide spraying Raspberry Pi system for agriculture/farming applications design implementation is illustrated.

Keywords Agriculture · Solar energy · Energy-efficient system · Raspberry Pi

1 Introduction

Replenishment of renewable energy from natural resources is at a very high rate. Existence of these resources despite of consumption and the rate of replenishment of the natural resources led to sustainable technologies to prevail in major applications.

P. A. Harsha Vardhini (✉) · M. H. S. Vishnu Sai
Vignan Institute of Technology and Science, Deshmukhi, Telangana, India
e-mail: pahv19@rediffmail.com

N. Sindhu · G. Janardhana Raju
NNR Educational Society Group of Institutions, Ghatkesar, Telangana, India

K. Srinivas
VNR Vignana Jyothi Institute of Engineering & Technology, Hyderabad, Telangana, India

Since few decades, solar energy is highly desired as a large sources of renewable energy. Solar fuel compared with other energy sources is highly abundant. Solar energy is free of pollution, as it does not contain pollutants or by-products that harm the atmosphere. India is an agricultural-based country about three-fourth of India's population is directly or indirectly depends on farming. For ages, farmers accustomed with conventional and manual methods for seed sowing to harvesting in the farming process including pesticide spraying. As in Fig. 1, conventional methods of pesticide spraying affect the health of farmer as many pesticides are made of chemicals/toxins. These pesticides in gaseous state turn in to liquid form as the farmer unlocks the spraying system. Though it is for protecting the crop from insects and bugs, the toxins in the form of vapor/droplets of pesticides do enter the human respiratory system by ingestion and inhalation that affects the health of farmer and also the farmer family accompanying in pesticide spraying process as in Fig. 2. There is a high risk of human life due to the contamination of this toxic sprayed liquid as depicted by Fig. 3 [2–4].

Later with the advent of automation, new agricultural methods emerged to support farmer at all stages of farming. Effective spraying techniques improved the benefits,

Fig. 1 Manual spraying of pesticides in farms
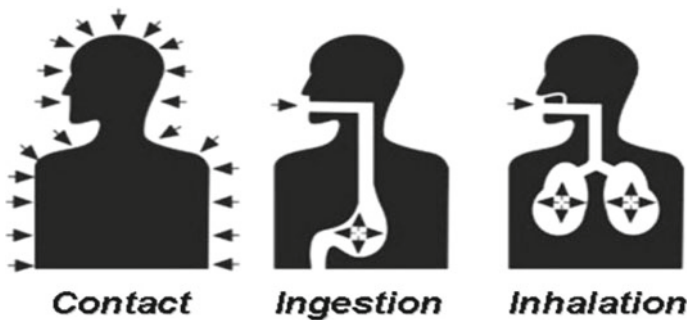
Fig. 2 Main routes of pesticide entry in human beings

Fig. 3 Picture of relation of toxicity, exposure, and risk
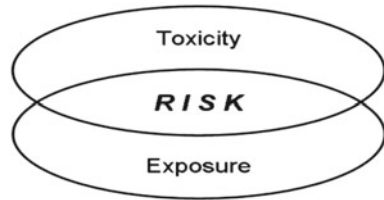
and this works project the design implementation of such a pesticide sprayer. In this work, solar powered automatic pesticide sprayer design implementation is illustrated. A sprayer of this type is a great way to use and is interminably utilized in the agribusiness field and likewise utilized for a few purposes. This has more favorable circumstances than siphon petroleum motor sprayer. It is thus free contamination pump compatible with spray pump petrol engine. This pump can be controlled from far places by giving instructions through personal computer or a mobile phone.

2 Solar Powered Pesticide Spraying Raspberry Pi System

Pesticide spraying system is implemented with Raspberry Pi powered by solar energy. Figure 4 depicts the design implementation of energy efficient self-driven pesticide spraying system including various modules [5, 6].
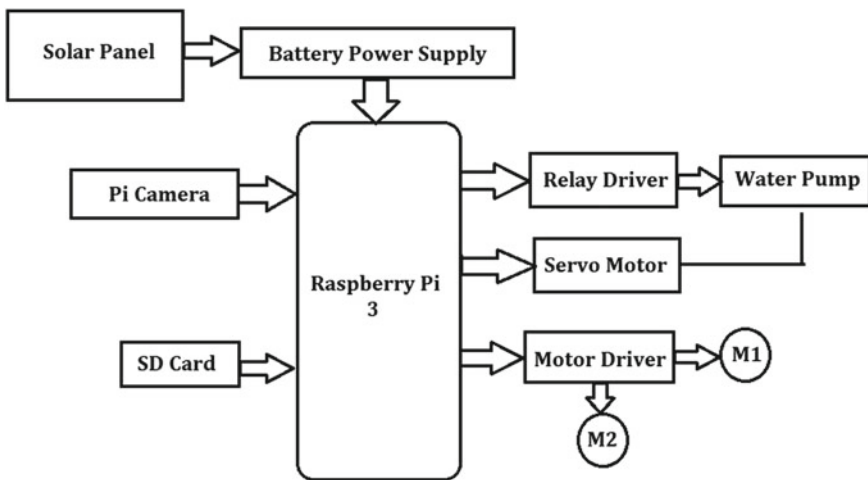
Fig. 4 Pesticide spraying system
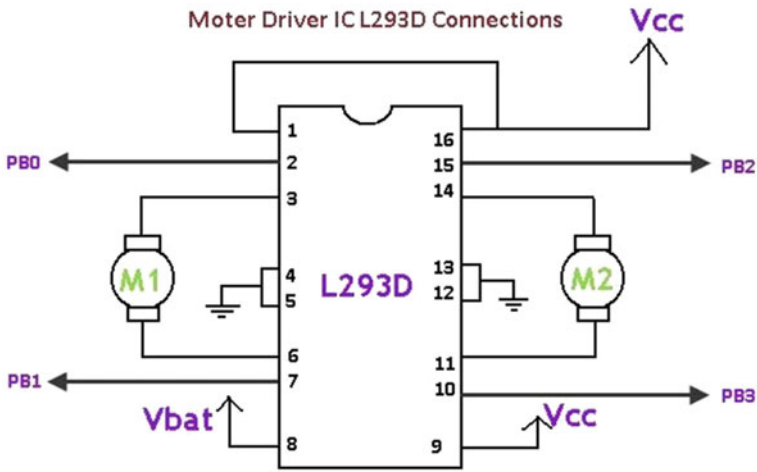
Fig. 5 Schematic interfacing of the L293D

2.1 Regulated Power Supply

Controlled power supply is a built-in circuit; a steady DC is transformed into unchecked AC as shown in Fig. 4. Using a rectifier this transforms AC supply to DC. The aim is provide a system with a steady "or less normal current" voltage that has to be controlled within some power supply limits.

2.2 L293d Motor Driver

L293D is engine dual H-bridge pilot with IC. Motor drivers serve as current amplifiers, picking up a control signal with low current and producing a signal of higher current, use this signal of higher current to power the engines. L293D has two H-bridge driver circuits built into it. L293d is double h motor bridge pilot, so can link two DC motors with one IC that can be operated In direction both clockwise and clockwise, if you have fixed motion direction motor you can Link up to 4 DC Engines using all four I/O motors. Additionally, it includes the emf performance diodes inside the IC for circuit safety from the back as depicted in Fig. 5.

2.3 Solar Cell/Plate

One sun photovoltaic cell system converting solar power to electricity with photovoltaic effect. The word solar cell is sometimes reserved for devices specifically to

Fig. 6 Solar cell/plate
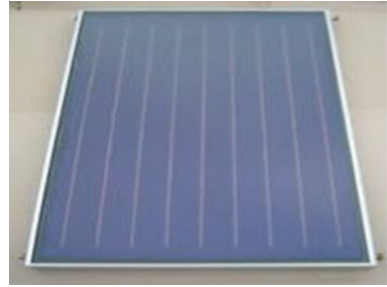
absorb energy from sunlight by using term photo voltaic cell when it is the root not specified as depicted in Fig. 6. Photovoltaic is the field of science and work related to the energy use by solar cells [7–9]. For commercially usable multi-crystalline solar cells, efficiencies in converting solar cells are 14–19%.

2.4 Raspberry Pi Camera Module

Raspberry Pi camera board as in Fig. 7 directly connects the Raspberry Pi CSI connector. With the new v1.3, it is able to offer a crystal clear image with 5mp resolution and audio capture 1080p hd 30 fps. Board features a 5MP (2592 × 1944 pixels) Omni vision 5647 sensor in a fixed focus module. Through means of a ribbon cable with 15 pins, the module Raspberry pi binds CSI serial interface dedicated to 15 pin camera developed specifically for cameras interfacing [10, 11].

Fig. 7 Raspberry Pi camera module
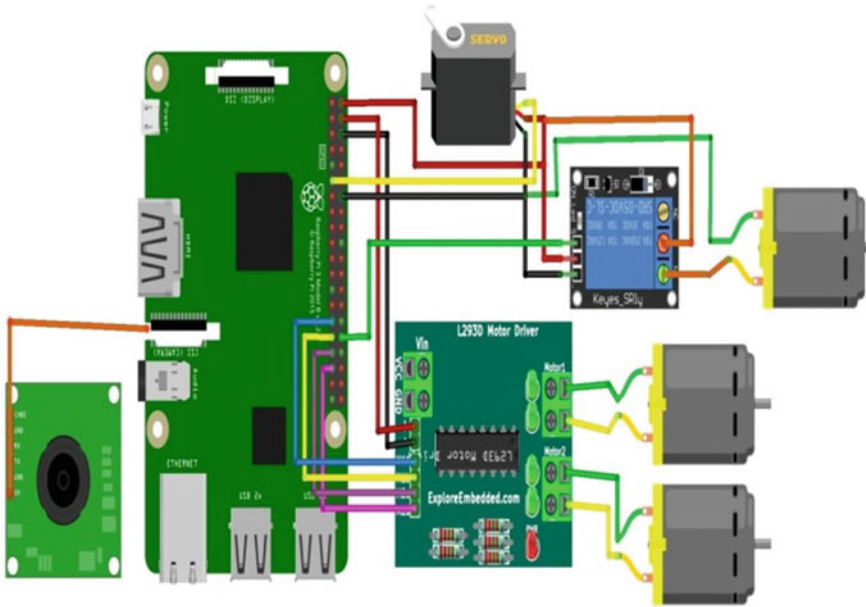
Fig. 8 Schematic diagram of connections Raspberry Pi system basic prototype of energy-efficient self-driven pesticide spraying system

Motor Servos: Engine servo is an electrical motor system able to move and if you wish turn an icon of excellence accuracy either angle, or distance, using the motor servos. We may obtain a very high torque servo motor in a small and lightweight box which is used in many applications such as toy cars, RC helicopters and aircraft, robots, computers, and so on [12] (Fig. 8).

3 Prototype of Pesticide Spraying System—Implementation and Results

3.1 Software Description

Guido van Rossum is in charge of designing Python. In 1989, Guido van Rossum started introducing Python. Python is a very basic programming language, and Fig. 9 depicts the features of Python programming.

The client is a Mac device that handles requests through HTTP using the standard network protocol used by the worldwide Web transmission of knowledge. The word can apply to the program as a whole or specifically the software which accepts and oversees HTTP requests. Computers are Web servers which deliver Web pages
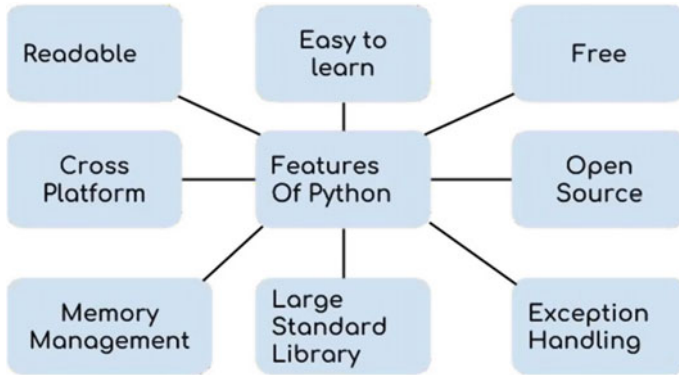
Fig. 9 Features of Python programming language

(serves up). Installation of software and servers linking the system to the Internet will transform any device on a Web server. There are many Web server software applications, including public domain software and commercial packages.

3.2 Install Raspbian with the Pi Raspberry—Implementation and Results

Raspbian is relatively simple to install on a farm noise Pi. Launch Raspberry and write the disk picture onto microbes card and boot the Pi Raspberry onto that microbes card. For this, you will need one "go with at least 8 GB" microbes card, device with a hole and a Pi Raspberry, of course and simple peripherals outside "a joystick, a keyboard, a panel, and a power source." It is not the only way to install Raspberry "more about this in a second," but practicing is a valuable practice since so many other Pi Raspberry operating systems may also be enabled. When you learn how to compose a microbes card image disk, you'll open up a ton of fun possibilities for Raspberry Pi projects [13–16]. Step-by-step implementation of the practical output is given below:

1. logging-in into the remote desktop as in Fig. 10 using user ID and password created earlier as shown in Figs. 11 and 12.
2. Opening the previously installed PUTTY software as depicted in Fig. 13 and login using credentials.
3. Now, we will be able to see the home page of Raspberry Pi computer and run all the supporting files that are already in the memory of Raspberry Pi. Now, we are able to see controlling of the page robot and spraying system as depicted in Fig. 14.
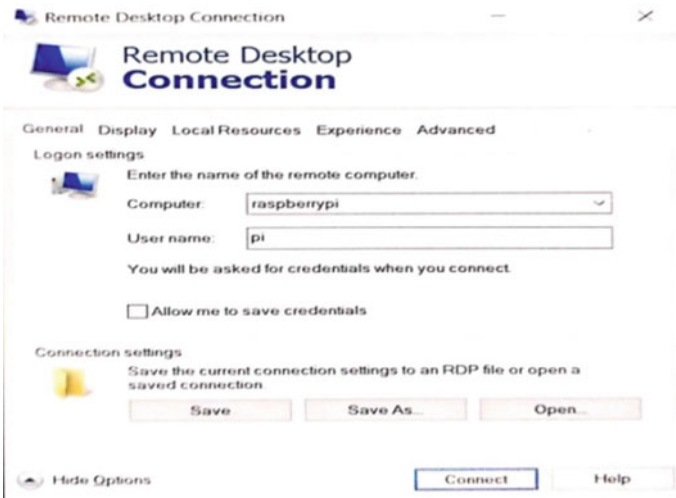
Fig. 10 Page control robot

Fig. 11 Connecting to Raspberry Pi
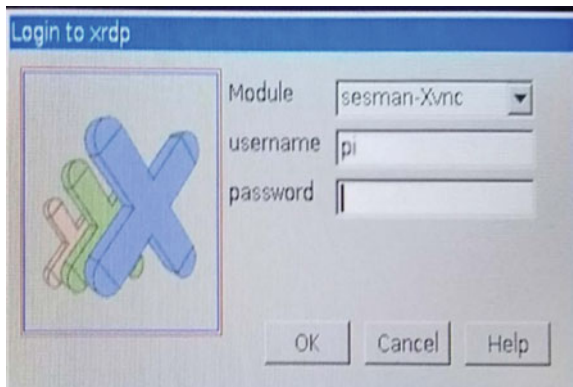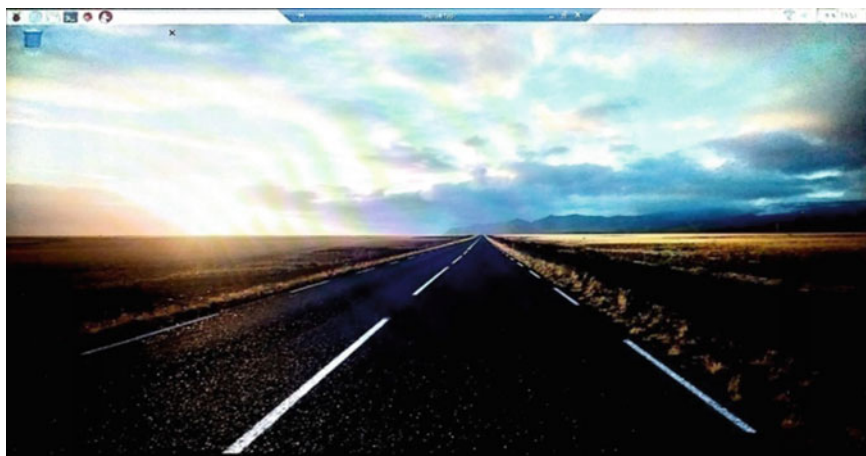
Fig. 12 Login page for putty
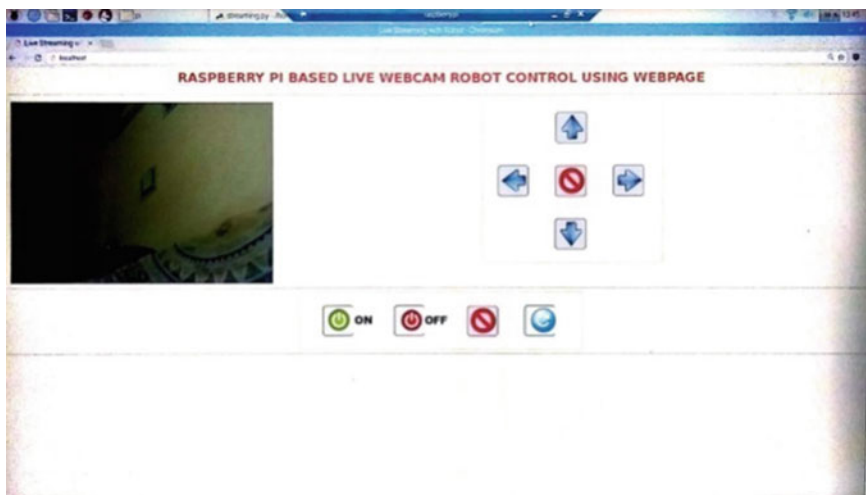
Fig. 13 Webpage for controlling robot

4 Conclusion

Self-driven Raspberry Pi pesticide spraying system powered by solar energy is providing more effective and safety way of pesticide spraying on to various crops. Implemented device enhances the human convenience. This system with pesticide sprayer moves as a robot in fields and the pipes on both sides to spray the pesticides on the plants. In this device, tiny tanks for pesticides, and the generator will then automatically continue to spray. This machine even has a Webcam that can monitor and view the live footage. This system for other types of applications can be a basic prototype with very few modifications. Future scope of the work is to extend the application with advanced technologies in the field of agriculture but also to bring the technology in a very convenient way close to the reach of the farmers in financial terms.

References

1. Harsha Vardhini PA, Vishnu Sai MH, Matta B (2022) Human footstep power harvesting systems using piezoelectric sensor technology. In: 2022 international mobile and embedded technology conference (MECON), 2022, pp 450–454. https://doi.org/10.1109/MECON53876.2022.975 2131
2. Upender P, Soumya M (2014) Embedded computer control system for service mobile robots. Int J Mag Eng Technol Manag Res (IJMETMR) 1(10):187–190
3. Upender P, Sainath K, Sritej M (2019) IoT based smart electronic trolley. J Appl Sci Comput (JASC) 12: 814–819. ISSN: 1076-5131
4. Harsha Vardhini PA, Ravinder M, Srikanth P, Supraja M (2019) IoT based wireless data printing using Raspberry Pi. J Adv Res Dynam Control Syst (JARDCS) 11(SI 4):2141–2145
5. Milan P, Jigna D, Kamlesh C, Bataviya M, Raviya K, Dhaval M (2018) Automatic pesticide spraying robot: design, development and fabrication based on farming. Int J Trend Res Dev 2:635–637

6. Vasanth N, Akash G, Srikanth KR, Pavan STN, Sinha R (2017) Solar powered automatic pesticides sprayer. In: 2017 international conference on energy, communication, data analytics and soft computing (ICECDS), Chennai, 2017, pp 3438–3441
7. Manimegalai M, Mekala V, Prabhuram N, Suganthan D (2018) Automatic solar powered grass cutter incorporated with alphabet printing and pesticide sprayer. In: 2018 international conference on intelligent computing and communication for smart world (I2C2SW), Erode, India, 2018, pp 268–271. https://doi.org/10.1109/I2C2SW45816.2018.8997301
8. Harsha Vardhini PA (2017) Efficient irrigation system powered by solar panel setup with tracking mechanism. Int J Emerg Technol Innov Res 4(3):238–241
9. Vardhini PH, Hanku V. Energy efficient implementation of IoT based home irrigation system using Raspberry Pi. In: 1st international conference on advanced technologies in engineering management & sciences, pp 218–223
10. Mahender G, Upender P (2016) Improved pre-failure healing system for movable sensor networks. Int J Innov Technol (IJITECH) 04(08):1342–1345
11. Harsha Vardhini PA, Ravinder M, Srikanth Reddy P, Supraja M (2019) Power optimized Arduino baggage tracking system with finger print authentication. J Appl Sci Comput J-ASC 6(4):3655–3660
12. Vardhini D, Chandra Babu K, Rani D (2019) Raspberry Pi based automated and efficient irrigation system with add-on field security. Int J Adv Sci Technol 28(19):192–196
13. Ren J, Dong W, Su Y, Zhu Z, Zeng X, Sun X (2017) Automatic displacement detection of target spray system based on micro inertial sensor. In: 2017 32nd youth academic annual conference of Chinese association of automation (YAC), Hefei, 2017, pp 1131–1134
14. Harsha Vardhini PA, Janardhana Raju G (2020) Design of internet of things based smart and efficient water distribution system for urban and agriculture areas. J Comput Theor Nanosci 17(9–10):4688–4691. https://doi.org/10.1166/jctn.2020.9301
15. Sosa R, Vazquez L, Santana I, Rubio E, Duran-Faundez C (2018) Automated sprayer system for variable rate application of pesticides. In: 2018 IEEE international conference on automation (ICA-ACCA), Concepcion, 2018, pp 1–6

Performance Comparison of Routing Protocols of WBAN for Emergency Handling in Homogeneous Network

Ramanpreet Kaur, Bikram Pal Kaur, Ruchi Pasricha, Jaskirat Kaur, Parveen Singla, and Harliv Kaur

Abstract The healthcare industry is booming with latest technological advancements in order to provide promising health solution. This has been possible due to research carried in the field of wireless body area networks which aim at continuous monitoring of patients through advanced technologies like deep learning, cloud computing, edge computing. Many researchers have contributed towards the development of various energy efficient routing protocols for successfully delivering packets to the destination. The successful delivery of packets at the destination node depends on the selection of routing protocol. This paper presents the performance assessment of routing protocols—SIMPLE, iM-SIMPLE, M-ATTEMPT, AMERP for handling critical data in WBAN based on residual energy, dead node analysis and packets received. The first node dead, half node dead and last node dead analysis has also been presented for all the four protocols. It has been presented that AMERP outperforms the three other routing protocols—SIMPLE, iM-SIMPLE and M-ATTEMPT for all the presented parameters.

Keywords Average network lifetime · Emergency traffic · Healthcare · Residual energy · Routing protocols · WBAN

R. Kaur (✉) · B. P. Kaur · R. Pasricha · P. Singla
CEC Landran, Mohali, India
e-mail: ramanpreet.71@gmail.com

J. Kaur
Department of Electronics and Communication Engineering, Punjab Engineering College (Deemed to be University), Chandigarh, India
e-mail: jaskiratkaur17@gmail.com; jaskiratkaur@pec.edu.in

H. Kaur
NJIT, Newark, NJ, USA

1   Introduction

The past decade has seen the implementation of latest technologies leading to the development of promising solutions in Wireless Body Area Networks (WBANs). The major aim of wireless body area networks is to regularly monitor the health of the patients at home/hospitals/remote areas by the medical practitioners [1]. This has been achieved with the help of the servers installed in hospitals/clinics. WBAN deals with the sensor nodes placed on or around the body or implanted inside the human body. These nodes sense the various physiological parameters of the body, record their values, and transmit these recorded values to the server wirelessly and followed by which the server responds with the diagnose [1, 2]. There are number of challenges that need to be dealt with while implementing a WBAN. One of the major challenges is the energy consumption by the sensor nodes. The energy consumption of the complete network should be as minimum as possible to enhance the lifetime of the network [3]. During the past few years, many researchers have contributed towards the development of energy efficient routing protocol for WBAN using the different technologies and algorithms [4, 5]. The energy consumption also increases if postural movement of sensor nodes is considered. One such protocol is the iM-SIMPLE (Improved Stable Increased-Throughput Multi-hop Protocol for Link Efficiency) protocol developed by Javaid, N., which considered postural movement, homogeneous network, both single hop and multi hop communication to calculate network lifetime, residual energy, throughput, dropped packets [6]. Other protocols in the series include Mobile Supporting Adaptive Threshold-based Thermal-aware Energy Efficient Multi-hop Protocol (M-ATTEMPT) [7] and Stable Increased-Throughput Multi-hop Protocol for Link Efficiency (SIMPLE) [8]. While M-ATTEMPT considered both normal and emergency packets, supported heterogeneous WBAN and used both single and multi-hop communication, but consumption of energy is higher during hotspot detection wherein a separate route is followed. SIMPLE protocol considered homogenous network with no body movements and supported both multi hop and single hop communication. These protocols had adopted older technologies like Integer Linear Programming, multi hop technologies for calculating the cost function to select the forward node which can transmit data to the sink node. Another protocol—Adam Moment Estimation Optimized Mobility Supported Energy Efficient Routing Protocol (AMERP) presented by researchers has been evaluated for handling emergency traffic in WBAN for homogeneous network [9, 10].

It has been concluded from the related work that most of the routing protocols did not consider the mobility aspect of all sensor nodes simultaneously. The handling of critical or emergency packets by the protocol is another desirable feature of any routing protocol. This is because medical emergency for transmitting and receiving node may arise at any time. Many routing protocols aim at handling normal traffic among sensor nodes with minimum postural movements. The objective of the present paper is to provide the comparative analysis of the four routing protocols for homogeneous environment in which all sensor nodes are characterized by same configuration and all sensor nodes except the sink nodes are supporting the mobility aspect. The performance comparison is done for the parameters-residual energy, packets received and dead node analysis.

The present paper is organized as—Sect. 1 highlights the aim of Wireless Body Area Networks and its routing protocols considered for present study, Sect. 2 presents the architecture of WBAN and related work. Section 3 presents the protocol AMERP for emergency handling and Sect. 4 highlights the comparison of AMERP with SIMPLE, iM-SIMPLE, M-ATTEMPT protocols. Section 5 concludes the paper with future scope.

2 WBAN Architecture

The architecture for WBAN design for communication purpose is divided into three tiers and is shown in Fig. 1 [5, 11]. Intra-WBAN includes connecting nodes within two meters of range in and around human body. It also includes inside WBAN correspondence and among WBANs and different tiers. The various physiological sensors placed on human body forward the data to a personal device in Intra WBAN which in further processes, transmits the data packets to an access point placed in Inter WBAN.

The Inter-WBAN involves communication between personal device and other access points and interconnects WBANs with other networks like Wi-Fi, cellular, etc. The Inter WBAN supports both Infrastructure mode architecture and ad hoc-mode architecture [12]. In Infrastructure mode as shown in Fig. 2, the access point acts a database server related to the application. This mode provides security control, provides central management, and is most widely used deployment strategy in hospitals. In ad hoc mode architecture as shown in Fig. 3, different access points are present to provide flexibility and extend the WBAN's coverage area from 2 to 100 m, suitable for short- and long-term infrastructure.

The beyond WBAN Communication uses a gateway to connect Inter and Intra WBAN. It involves application specific communication for use in metropolitan areas. The end or destination device like medical servers, hospitals, remote patient monitoring etc. can be linked to Inter WBAN through internet, wired or wireless network. The Beyond WBAN involves communication of medical data of the patient which can be accessed by authorized medical professionals only. The security of the medical data of the patient is also an important concern in Beyond WBAN tier.
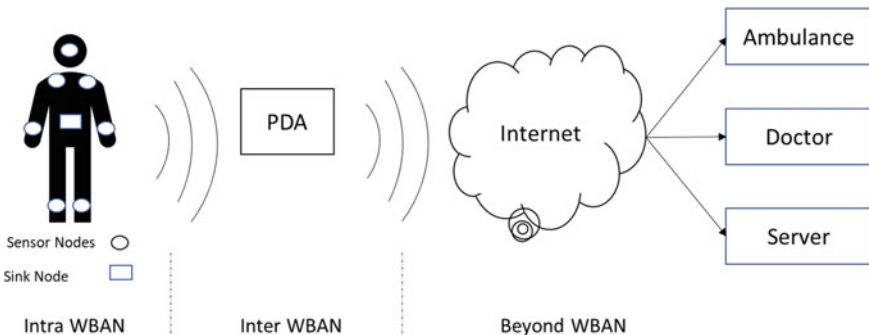
Fig. 1 WBAN architecture
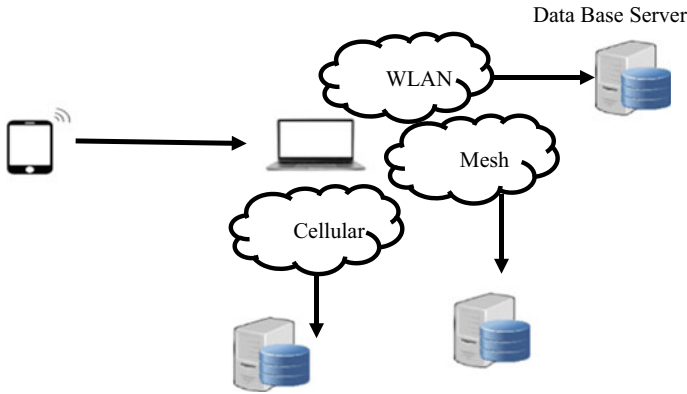
Fig. 2 Infrastructure mode
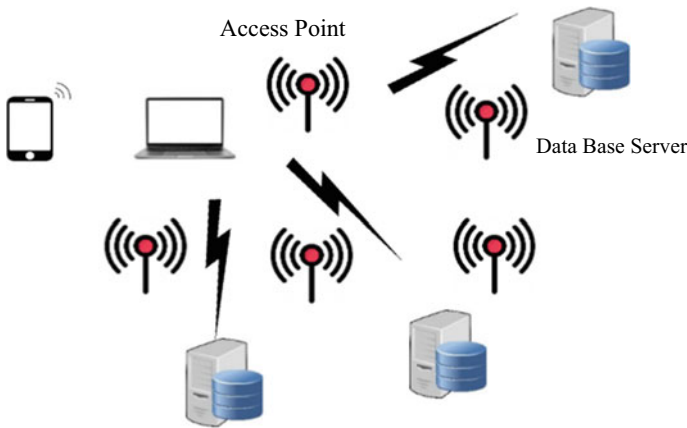
Fig. 3 Adhoc mode

SIMPLE Protocol developed in [8] considered multi hop and single hop for homogeneous network. The performance is calculated based on the cost function which depends on residual energy of node and distance between node and sink. iM-SIMPLE [6] protocol adopted multi hop communication approach for the improvement in performance parameters and used a personal digital assistant (PDA) on human body as an access point. M-ATTEMPT protocol [7] used single hop for emergency packets and multi hop communication for normal packets. AMERP developed in [9, 10] used Adam optimizer [13] to predict the forward node and is able to handle emergency traffic as well.

3 Proposed Methodology

Adam Moment Estimation Optimized Mobility Supported Energy Efficient Routing Protocol (AMERP) developed in [9] has been extended for handling critical or emergency data in [10]. The performance of the protocol has been compared with the SIMPLE (Stable Increased-Throughput Multi-hop Protocol for Link Efficiency), iM-SIMPLE (Improved Stable Increased-Throughput Multi-hop Protocol for Link Efficiency), M-ATTEMPT (Mobile Supporting Adaptive Threshold-based Thermal-aware Energy Efficient Multi-hop Protocol) protocols for emergency data handling. Figure 4 shows the positions of the eight sensor nodes and one stationary sink node that have been considered and placed on the body. The mobility of eight sensor nodes has also been considered. The eight nodes measure the physiological parameters of the human body-blood pressure, coordinates of the position, glucose level, EMG, lactic acid, body temperature, pressure, and acceleration, respectively.

The respective distance of sensor nodes from the sink is considered as − 44.7, 36.0, 30.0, 22.4, 30.0, 31.6, 51.0, 70.7 (all in centimeters) as shown in Fig. 4. For the homogeneous configuration, the initial energy of all nodes is considered to be same and is 0.5 J for each node except the sink node. The total initial energy of homogeneous network is 4 J (8 nodes * 0.5 J) [9].

The energy equations considered for the calculation for transmission and reception of packets from nodes to sink and vice versa are given in Eqs. 1 and 2 based on first order radio model [14].

Fig. 4 Position of nodes
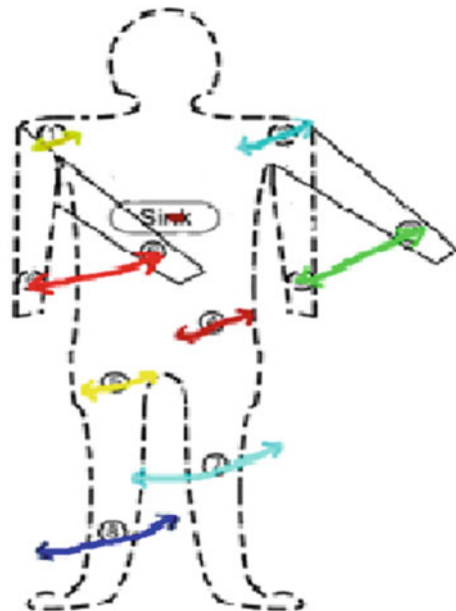
$$E_{\text{txed energy}} = E_{\text{tx per bit}} * k1 + \varepsilon_{\text{amp}} * n1 * k1 * d^{n1} \qquad (1)$$

The energy received by the node is

$$E_{\text{rxed energy}} = E_{\text{rx per bit}} * k \qquad (2)$$

where, $E_{\text{txed energy}}$ is the energy transmitted, $E_{\text{tx per bit}}$ is the energy used in transmitting per bit of energy, $k1$ is packet size, ε_{amp} is radio amplifier constant, $n1$ path loss coefficient, $E_{\text{rxed energy}}$ received energy, $E_{\text{rx per bit}}$ energy per bit during reception, and d is taken as the distance of the receiving node from the transmitting node. The parameters of radio amplifier nRF2410A transceiver has been considered.

4   Results and Discussions

This section presents the comparison of SIMPLE, iM-SIMPLE, M-ATTEMPT, AMERP for emergency data handling in Wireless body area networks in terms of packets received, residual energy, dead node analysis and FND, HND and LND analysis. As stated in Sect. 3, the sink node is kept as stationary node and initial energy of all the nodes except sink node is same for homogeneous configuration. It is assumed that power consumed during sensing is negligible as compared to power consumed during communication of packets among sensor nodes.

4.1   Packets Received

The successful reception of packets for different simulation rounds for homogeneous network with emergency traffic has been plotted in Fig. 5. Since emergency packets carry the critical information about the patient's health, maximum of the transmitted packets must reach the destination. These packets need to be transmitted to the sink node directly without any forward node selection. But it has been observed that few emergency packets are also dropped because the few sensor nodes are having insufficient energy level to hold the packets. As a result of which, the nodes die out. Higher the count of active nodes, higher will be the count of received packets. The maximum number of packets received respectively for SIMPLE, iM-SIMPLE, M-ATTEMPT are 1.5×10^5 bytes, 2.8×10^5 bytes, 1.8×10^5 bytes, approximately. The number of packets received is 4.9×10^5 bytes approx. after 3300 rounds in AMERP.
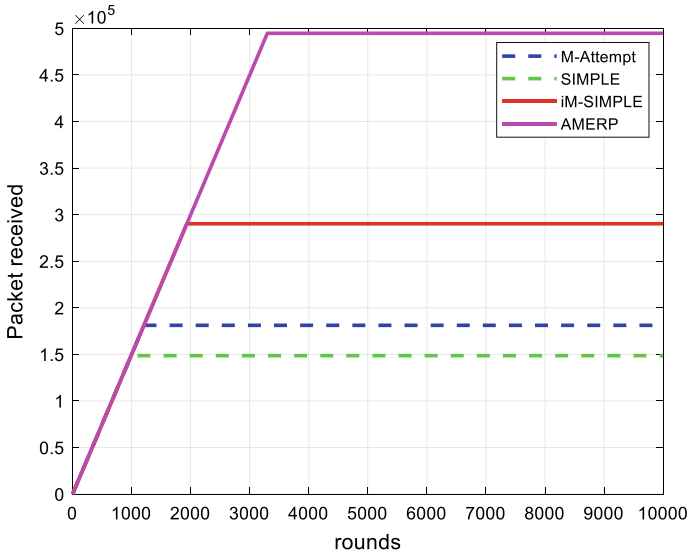
Fig. 5 Packets received for SIMPLE, iM-SIMPLE, M-ATTEMPT and AMERP for emergency traffic

4.2 Dead Node Analysis (Average Network Lifetime)

Emergency packets are transmitted directly by the node without selecting any forward node. These packets need to be delivered to the destination first without following any optimum path. As shown in Fig. 6, in AMERP, all nodes die out around 3300 rounds. This is because of emergency packets that need to be delivered faster to the sink node. All nodes in iM-SIMPLE die out at 2719 rounds. In SIMPLE and M-ATTEMPT protocol, all the nodes die out after nearly 1000 and 1459 rounds, respectively.

4.3 Residual Energy

Since the emergency packets need to be delivered directly to the sink node without the selection of forward node, the energy consumed for these packets will be much higher than the normal traffic. While for AMERP, the value at 3300 rounds is 0 J. This means that all nodes have died out till this round in AMERP. For iM-SIMPLE, the residual energy of the network becomes zero at 2719 rounds. In SIMPLE and M-ATTEMPT protocol, the residual energy becomes zero after nearly 1000 and 1459 rounds, which means that all the nodes have died out after these number of rounds as shown in Fig. 7. This is because emergency packets consume more energy and need to be delivered to the sink node without selection of any forward node.
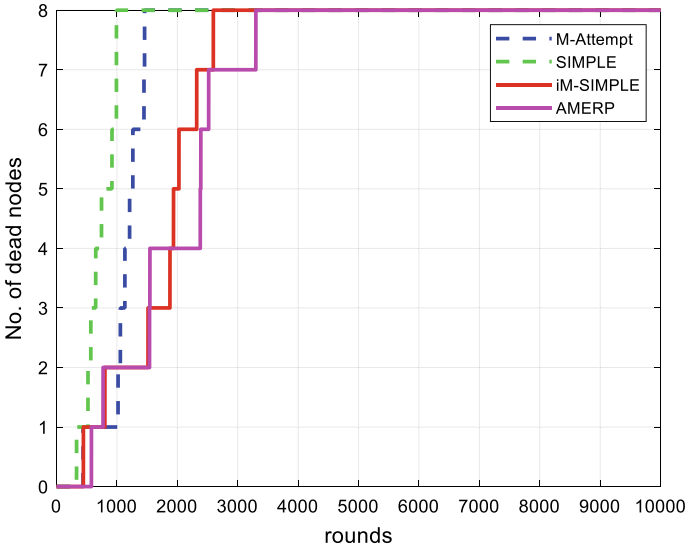
Fig. 6 Dead node analysis for SIMPLE, iM-SIMPLE, M-ATTEMPT and AMERP for emergency handling
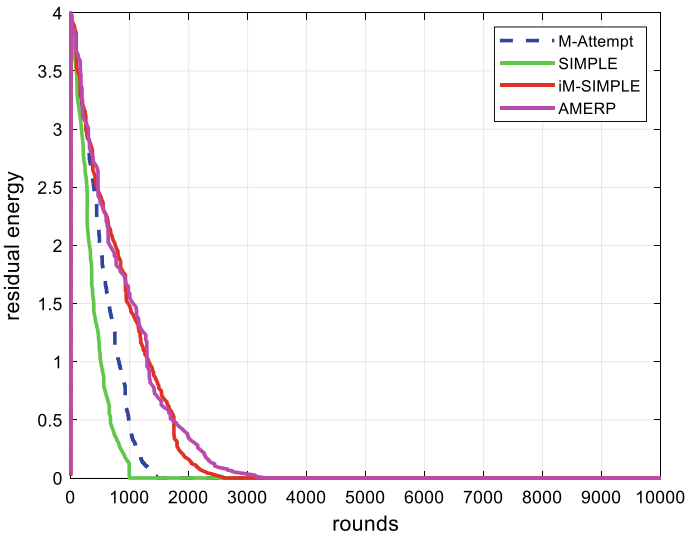
Fig. 7 Residual energy for with SIMPLE, iM-SIMPLE, M-ATTEMPT and AMERP for emergency handling

4.4 FND, HND and LND Analysis

The FND, HND, LND analysis for the emergency traffic and homogeneous network is shown in Fig. 8. And for AMERP, the first node dies at 520 rounds. The FND for SIMPLE and M-ATTEMPT occurs at 368 rounds and 478 rounds, respectively. For FND, iM-SIMPLE follows M-ATTEMPT. The HND analysis shows that iM-SIMPLE performs better than M-ATTEMPT, AMERP and SIMPLE protocols. The three existing protocols—iM-SIMPLE, M-ATTEMPT, SIMPLE do not support longer network lifetime for emergency packets as LND for them occurs at 2719 rounds, 1459 rounds and 1000 rounds. This means that the emergency aware protocol AMERP provide longer network lifetime than the three protocols.

As presented in Table 1, it is stated that AMERP performs outperforms the other three routing protocols. Since emergency packets in WBAN are critical packets that need to be delivered to the sink node directly without the intervention of any forward node and maximum number of packets are received for AMERP. It is restated that a greater number of alive nodes will increase the network lifetime. AMERP support longer network lifetime as the nodes die later in AMERP as compared to SIMPLE, iM-SIMPLE, M-ATTEMPT.
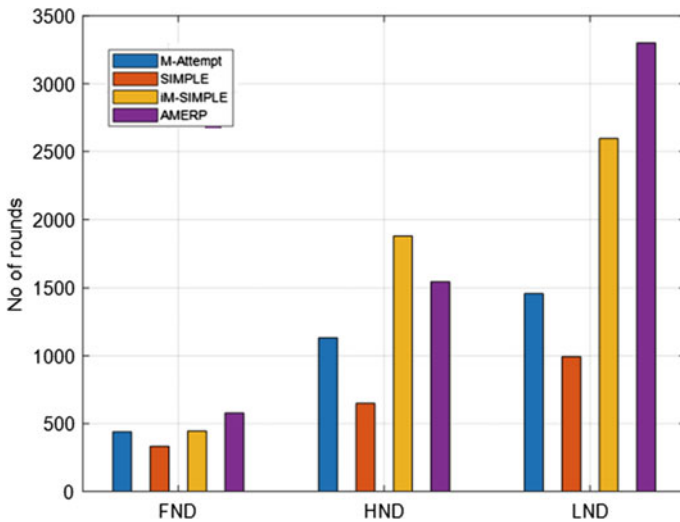
Fig. 8 FND, HND, LND analysis for SIMPLE, iM-SIMPLE, M-ATTEMPT and AMERP and AMERP for emergency handling

Table 1 Comparison of SIMPLE, iM-SIMPLE, M-ATTEMPT, AMERP

Parameter/protocol	SIMPLE	iM-SIMPLE	M-ATTEMPT	AMERP
Packets received (in bytes)	1.5×10^5	2.8×10^5	1.8×10^5	4.9×10^5
First node dead (in rounds)	368	478	467	520
Half node dead (in rounds)	732	1847	1158	1558
Last node dead (in rounds, residual energy becomes zero)	1000	2719	1459	3300

5 Conclusion and Future Scope

Wireless Body Area Networks has been a promising area of research for the advanced healthcare industry. This paper presents that routing protocols developed for WBAN must be energy efficient so that longer network lifetime can be supported. This paper presented performance comparison of routing protocols—SIMPLE, iM-SIMPLE, M-ATTEMPT, AMERP for the emergency handling in homogeneous network. The parameters like packets received, residual energy, Dead node analysis has been presented for the three protocols. It has been presented that AMERP for the emergency packets performs better than SIMPLE, iM-SIMPLE, M-ATTEMPT. The AMERP emergency aware provides longer network lifetime up to 3300 rounds and maximum number of packets received for emergency aware AMERP is 4.9×10^5 as compared to other protocols SIMPLE, iM-SIMPLE, M-ATTEMPT. In future scope, the presented protocols will be studied for heterogeneous configuration of nodes and advanced optimized algorithmic approach will be used for optimizing performance parameters.

References

1. Movassaghi S, Abolhasan M, Lipman J, Smith D, Jamalipour A (2014) Wireless body area networks: a survey. IEEE Commun Surv Tutor 16(3):1658–1686. https://doi.org/10.1109/SURV.2013.121313.00064
2. Khan JY, Yuce MR (2010) Wireless Body Area Network (WBAN) for medical applications. New developments in biomedical engineering, chap 31. IntechOpen, pp 591–628
3. Kaur R, Kaur BP, Singla R, Kaur J (2020) Sensors technology and power consumption issue in healthcare environment. J Crit Rev 7(19)
4. Shah SN, Jhaveri RH (2016) Recent research on wireless body area networks: a survey. Int J Comput Appl 142(11):975–8887
5. Salayma M, Al-Dubai A, Romdhani I, Nasser Y (2017) Wireless Body Area Network (WBAN): a survey on reliability, fault tolerance, and technologies coexistence. ACM Comput Surv 50(1). https://doi.org/10.1145/3041956
6. Javaid N, Ahmad A, Nadeem Q, Imran M, Haider N (2015) iM-SIMPLE: iMproved stable increased-throughput multi-hop link efficient routing protocol for Wireless Body Area Networks. Comput Hum Behav 51:1003–1011. https://doi.org/10.1016/j.chb.2014.10.005

7. Javaid N, Abbas Z, Fareed MS, Khan ZA, Alrajeh N (2013) M-ATTEMPT: a new energy-efficient routing protocol for wireless body area sensor networks. Procedia Comput Sci 19:224–231. https://doi.org/10.1016/j.procs.2013.06.033
8. Nadeem Q, Javaid N, Mohammad SN, Khan MY, Sarfraz S, Gull M (2013) SIMPLE: stable increased-throughput multi-hop protocol for link efficiency in Wireless Body Area Networks. In: Proceedings of the 8th international conference on broadband, wireless computing, communication and applications, BWCCA 2013, Compiegne, France, pp 221–226. https://doi.org/10.1109/BWCCA.2013.42
9. Kaur R, Kaur BP, Singla RP, Kaur J (2021) AMERP: Adam moment estimation optimized mobility supported energy efficient routing protocol for wireless body area networks. Sustain Comput: Inform Syst 31:100560. https://doi.org/10.1016/j.suscom.2021.100560
10. Kaur R, Kaur BP, Singla RP, Kaur J, Singla P (2022) Performance evaluation of AMERP for emergency handling in homogeneous network. In: Proceedings of AIP conference (ICCS 2021), Punjab, India
11. Bilstrup K (2008) A preliminary study of wireless body area networks. Technical Report No. IDE0854, Halmstad University, Halmstad, Sweden
12. Mehta G, Singla P, Mittal (2018) Design and evaluation of a wireless body sensor system for smart home health monitoring. Int J Comput Sci Mob Comput 7(11):167–172
13. Kingma DP, Ba JL (2015) Adam: a method for stochastic optimization. In: Proceedings of the 3rd international conference on learning representations, ICLR 2015, San Diego, CA, USA, pp 1–15
14. Heinzelman W, Chandrakasan A, Balakrishnan H (2000) Energy-efficient communication protocols for wireless microsensor networks. In: Proceedings of the IEEE 33rd Hawaaian international conference on systems science (HICSS), USA, pp 1–10

Deep Learning-Based Ensemble Framework for Brain Tumor Classification

Aman Gupta, Aman Sharma, and Rajni Mohana

Abstract Brain tumor is one of the severe diseases which badly impacts the human body. The fatality rate is quite high for this deadly disease. Doctors are the main gateway for manual inspection of the MRI scans for the brain tumor which is in fact a traditional approach. The sheer amount of data through these scans are immeasurable; hence, this procedure is quite unfeasible. Therefore, it is essential to have reliable and precise automatic classification strategies in place to lessen or reduce the fatality rate of the patients. Tumor classification approaches have been developed to aid neurologists in saving time and achieving demonstrated accuracy. In this study, we proposed an ensemble approach for classifying the brain tumor into different types like gliomas, meningioma and pituitary tumors. The accurate and precise distinction among the three types of brain tumors signify a very crucial phase of the medical diagnostic procedure and later successful examination of patients. Our study suggests that our Deep Learning Voting Ensemble Model outperforms various base deep learning models that are used individually for the application of brain tumor classification.

Keywords Brain tumor classification · Deep learning · Ensemble model

A. Gupta (✉) · A. Sharma · R. Mohana
Jaypee University of Information Technology, Solan, Himachal Pradesh, India
e-mail: amangupta202001@gmail.com

A. Sharma
e-mail: aman.sharma@juit.ac.in

R. Mohana
e-mail: rajni.mohana@juit.ac.in

1 Introduction

Although biopsy is the gold standard for tumor grading, because it is intrusive, it is also lethal to the brain. The early detection of a tumor is important for correct identification. Prompt action is additionally essential. Early diagnosis may be difficult work because of the problem of classifying traditional and abnormal tissues correctly. In this case, automatic diagnostic systems will make it easier to overcome the constraints of traditional techniques and manual diagnosis. Brain tumor connotes a distorted mass of tissue whereby the cells duplicate suddenly and endlessly at intervals in the brain tissues [1]. Brain tumor is not region explicit nor will it have a specific form nor size. Supported development and separation of cells once contrasted with typical cells the tumor is distinguished into two types: benign and malignant tumor. For benign tumors, the care is very straightforward by eliminating the actual infectious sites. Then again, malignant tumors are perilous a lot that they will get in your mind even when they are removed [2]. The dataset for the deep learning model contains MRI scans as among all other imaging modalities, and MR images are preferred because they are the only non-invasive technique that provides useful information on brain tumor kind, size, shape and position in 2D and 3D formats [3]. During the course of this research, it was endeavored at classifying the brain tumor and contrasting the outcomes classification of brain tumor with use of pretrained Keras models like VGG16, DenseNet201, EfficientNetB1 and convolutional neural network architecture with the proposed majority voting-based ensemble model of deep learning models. The voting-based ensemble model is computationally less expensive as compared to other ensemble techniques.

Objectives of the research study:

- Voting-Based Ensemble Classifier is used which adds diversity to the final classifier.
- Ensembling helps in improving the prediction accuracy of brain tumor classification.
- Result analysis based on performance parameter, i.e., sensitivity shows that the final trained model is robust toward healthcare applications.

2 Literature Review

Kang et al. [4] used three magnetic resonance images datasets to validate various different kinds of pretrained models for feature extraction, ML-based classifiers and the effectiveness of an ensemble of deep features for brain tumor classification. He stated in his paper that the ensemble model can make the accuracy of the predictions significantly higher. Tandel et al. [5] used four different clinically applicable MR Images datasets for their study. These datasets were used in training and testing of five deep learning-based models and 5 ML-based models. They proposed a majority voting-based ensemble model. The accuracy of the models increased considerably

and achieved good results for brain tumor classification. Ensemble model was able to utilize the potential of multiple models combined together. Abiwinanda et al. [6] stated that manually diagnosing the type of tumor is a time-consuming and error-prone process. They used pretrained models to classify glioma, meningioma and pituitary tumors. Sajjad et al. [7] proposed a novel CNN-based multi-grade brain tumor grading scheme. Deep learning techniques were used to segment the tumor regions in MRI scans. Then they used data augmentation techniques to have more data for training and testing of the models. To identify the grade of the tumor, the data was fed into a pretrained CNN model. This study made use of two distinct datasets. With augmented data, the suggested approach achieved 90.67% accuracy. Cheng et al. [8] used T1-weighted contrast-enhanced MR images for their study. They augmented tumor regions via image dilation to be used as the region of interest. The intensity histogram, GLCM and bag-of-words model are three feature extraction approaches with which they evaluated the efficacy of the proposed method. The accuracy of the proposed method was found out to be 91.28% on T1-weighted MR images. Singh et al. [9] used a novel normalized histogram and segmentation approach based on the K-means clustering algorithm. They used various image denoising filters to reduce noises present in the MR images. Naive Bayes and SVM are used to efficiently predict brain tumors. Noreen et al. [10] used a MR images dataset for training and testing the proposed models. They used both deep learning as well as machine learning algorithms to extract features and classify them into various types of tumors. Their ensemble model showed promising accuracy.

3 Dataset

The dataset used for the proposed project is through Kaggle [11]. The dataset has MR Images of the brain. The project is intended to have a multi-class classification due to which the dataset has four classes, three classes of tumor and glioma, pituitary and meningioma and one for no tumor. Some visual differences are definitely observed between the types which resulted from biological differences between them. The dataset is divided into training and testing sets. In the training set, there are 2870 files which consist of 826, 822, 827 and 395 files for glioma tumor, meningioma tumor, pituitary tumor and no tumor, respectively, and for testing set, there are 394 files which consist 100, 115, 105 and 75 files for glioma tumor, meningioma tumor, pituitary tumor and no tumor, respectively. The training-to-testing set ratio used in the proposed model is 88%:12%.

4 Methodology

The proposed methodology for this study is divided into three phases. Each phase signifies the progress of the model.

4.1 Data Preprocessing and CNN Model Building

This phase of making an ensemble model mainly focused on data processing of the dataset and building a CNN model from scratch. Data processing is considered a crucial part because raw data or dataset found in real life usually have inconsistencies like formatting, human errors and can also be incomplete. Data preprocessing is one of the ways to resolve these issues and make datasets more complete and efficient to perform data analysis on.

Data processing performed on all deep learning models was similar. The dataset used for this model was magnetic resonance images (MRI) which consisted of various types of brain tumor and also a class for no tumor, i.e., healthy MRI scans. The major steps involved in the data processing were to read the images, resizing all the images to a standard size.

4.2 Ensembling the Deep Learning Models

In this phase, we made a voting-based ensemble model using various different deep learning models which are mainly used for image classification.

We used VGG-16, DenseNet201, CNN model (that we made from scratch) and EfficientNet B1.

4.2.1 VGG-16

K. Simonyan and A. Zisserman proposed a CNN model called VGG16 in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition" [18]. It is the model that performed well in the 2014 ILSVRC contest. It is one of the finest vision model designs that have ever been proposed. It had 3×3 filter convolution layers having stride that was equal to 1, used similar padding and max pooling layer of 2×2 filter. The architecture has two fully connected layers in the end, which was followed by a softmax for output.

4.2.2 DenseNet 201

Dense convolutional network (DenseNet) is a CNN that connects every other layer in a feed forward manner. DenseNet works on the thought that convolutional networks could be considerably deeper, a lot more accurate and efficient to train if they need shorter connections between layers closer to the input and those close to the output. DenseNet-201 is 201 layers deep. They significantly mitigate the gradient vanishing issue generated by an excessively deep network [12]. It takes advantage

of the condensed network to provide models that are easy to train and very parametrically efficient due to the potential of feature reuse by multiple layers, which enhances performance [13]. Feature propagation is strengthened, and it also reduces the number of parameters substantially. It has demonstrated outstanding performance on a variety of datasets, including ImageNet and CIFAR-100 [14].

4.2.3 EfficientNetB1

EfficientNet, first introduced in Tan and Le, 2019, is one of the most efficient models that achieves state-of-the-art accuracy on both ImageNet and common image classification transfer learning tasks. EfficientNet models offer a strong balance of efficiency and accuracy across a wide range of scales. It demands a channel size to be multiples of 8.

4.2.4 CNN Model

In this study, we made our own CNN model from scratch to have a high accuracy as well as efficiency. Five convolution layers are included with a size of kernel kept equivalent of 1 * (5 × 5), 2 * (3 × 3), 2 * (2 × 2), and the strides were set to (2, 2) for all layers. To reduce training time, Max Pooling is used as it decreases the no. of parameters being used. To reduce overfitting, Dropout is used which is a regularization method [15].

4.2.5 Ensemble Model

Sometimes a single model is not enough for classification predictive models. Ensemble is a machine learning technique that generates an optimized prediction model by integrating many basic models [8]. For this study, the majority of votes method is selected as the ensemble model as it is generally used in classification problems. In this study, we used four models to make predictions for each data point. Predictions from each model are considered a "vote" [16]. This ensemble model is based on VGG-16, EfficientNetB1, DenseNet201 and CNN as shown in Fig. 1.

4.3 Script (Pseudo Code)

Pseudo codes for different parts of the algorithm:
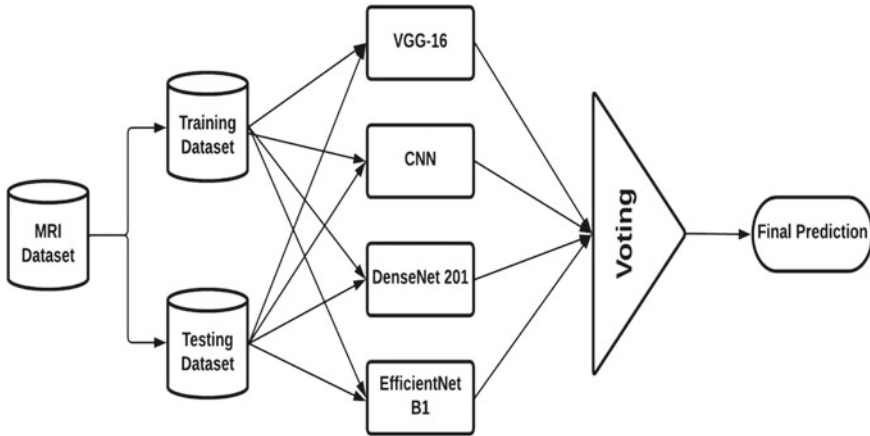
Fig. 1 Voting ensemble model based on various deep learning models

Pseudo Code 1 Preprocessing

Input: Image dataset

Output: Class_labels: [0, 1, 2, 3]

1. func create_training_dataset():   #Function for creating training data array
2. class_number <- category_index   #Getting category labels for Images
3.   For image in images
4.   img_array <- read_image
5.   new_img_array <- resize_image(img_array)   #resizing images
6.   training_data <- append(new_img_array,class_number)
7. create_training_dataset()

Pseudo Code 2 Model Building

1. Create CNN model with filters (64 or 128 or 256), kernel size(3 × 3 or 5 × 5) and pool size(2 × 2).
2. Train the CNN model with the brain Tumor dataset.
3. Import VGG-16 model from Keras library.
4. Train this pretrained model on the dataset.
5. Import DenseNet201 model from Keras library.
6. Train this pretrained model on the dataset.

Table 1 Hyperparameters used

Model	Learning rate	Epochs	Optimizer	Batch size
VGG-16	0.01	20	Adam	40
EfficientNetB1	0.01	12	Adam	32
DenseNet201	0.01	30	Adam	32
CNN	0.001	40	Adam	40

7. Import EfficientNetB1 model from Keras library.
8. Train this pretrained model on the dataset.
9. Save these trained models and their weights.
10. Make a list of these models with their names.
11. Pass these models into the Voting classifier function.
12. Fit this ensemble model.
13. Get majority voted class as output.

4.4 Parameter Setting

In this section, we discussed various factors that were used to improve the accuracy of our voting ensemble model. VGG-16, EfficientNetB1, DenseNet201 and CNN were with highest accuracy. These four were chosen as base learners for the ensemble model. Hyperparameter tuning on the following model was done to further improve the accuracy which, as shown in Table 1, can in turn improve the accuracy of voting classifiers.

5 Results

In this study, performance evaluation was performed with Accuracy, Sensitivity, Specificity and F1 score. Each of these performance measurements is represented by:

$$\text{Accuracy} = \frac{\text{T.P.} + \text{T.N.}}{\text{T.P.} + \text{T.N.} + \text{F.P.} + \text{F.N.}} \quad (1)$$

$$\text{Sensitivity} = \frac{\text{T.P.}}{\text{T.P.} + \text{F.N.}} \quad (2)$$

$$\text{Specificity} = \frac{\text{T.N.}}{\text{T.N.} + \text{F.P.}} \quad (3)$$

Table 2 Precision, recall and F1 score for different tumor types by CNN model

Tumor type	Precision (%)	Recall (%)	F1 score
No tumor	90	89	89
Glioma tumor	94	89	92
Meningioma tumor	86	86	86
Pituitary tumor	93	98	95

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

here,

T. P.   True Positive
T. N.   True Negative
F. P.   False Positive
F. N.   False Negative.

Tables 2, 3 and 4 and Fig. 2 the graphs show the performance metrics such as precision, recall and F1 score for VGG-16, DenseNet201 and CNN models, respectively for various types such as no tumor, glioma tumor, meningioma and pituitary tumors. We can see from Fig. 2a that the precision and F1 score was very low for glioma and meningioma tumors. It performed very poorly on this dataset. Recall score was also low for the VGG-16 model.

Accuracy and sensitivity of the models can be seen from the graph in Fig. 3. For some models, we can see that the sensitivity and accuracy was almost equal.

In case of brain tumor classification, there should be high sensitivity of the model. As sensitivity is calculated as shown in Eq. 2, so there should be less false positives. If there is a high number of false positives means less sensitivity, that could be dangerous for the patient.

Table 3 Precision, recall and F1 score for different tumor types by DenseNet201 model

Tumor type	Precision (%)	Recall (%)	F1 score
No tumor	74	99	85
Glioma tumor	93	92	93
Meningioma tumor	97	87	92
Pituitary tumor	91	72	80

Table 4 Precision, recall and F1 score for different tumor types by VGG-16 model

Tumor type	Precision (%)	Recall (%)	F1 score
No tumor	97	76	85
Glioma tumor	65	97	78
Meningioma tumor	75	99	85
Pituitary tumor	100	26	41
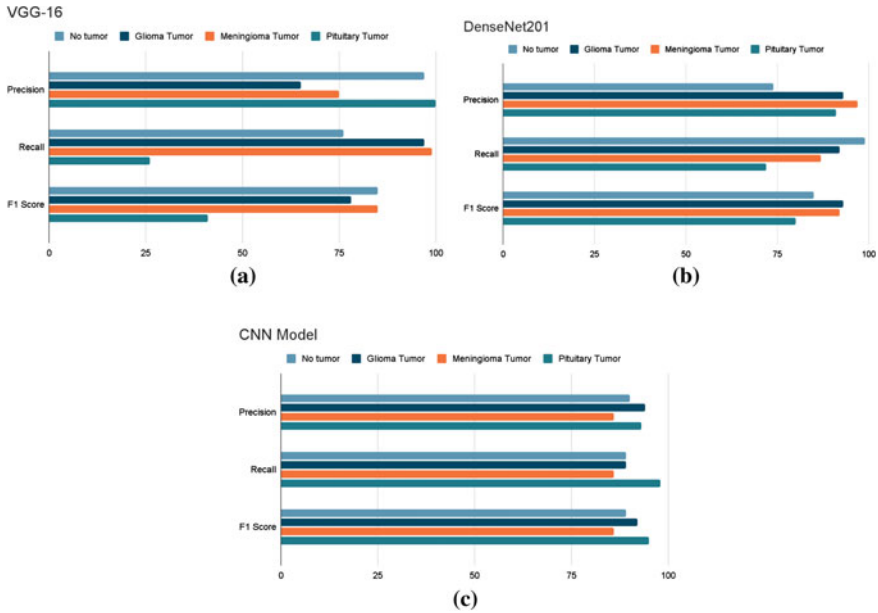
Fig. 2 Various performance metrics for VGG-16, DenseNet20 and CNN models (*X*-axis: percentage values and *Y*-axis: performance metrics)
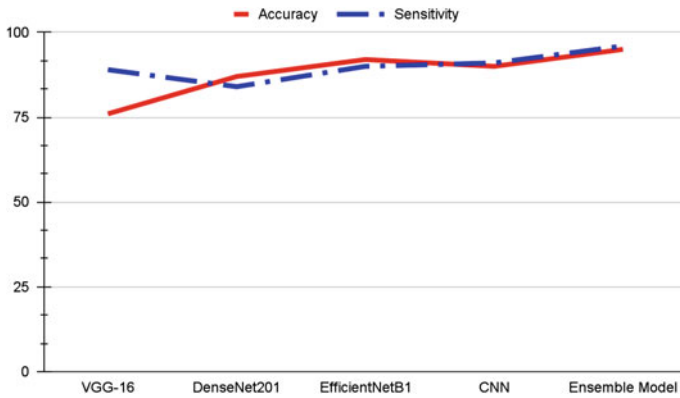
Fig. 3 Accuracy and sensitivity of various model

Table 5 shows the accuracy and sensitivity of all models. As we can see from this table, the sensitivity as well as the accuracy of the ensemble model is the highest among all other models.

With the help of above experimental results that are based on performance parameters, it is deduced how well the deep learning models and majority voting-based ensemble model performed on different criteria. As we can see that the sensitivity of

Table 5 Accuracy and sensitivity of various deep learning models

Model name	Accuracy (%)	Sensitivity (%)
VGG-16	76	89
CNN	90	91
DenseNet201	87	84
EfficientNetB1	92	90
Ensemble Model	95	96

the individual models was not enough to say that they are well-performing models. VGG-16 was the poorest performing model in this case. The results show that the ensemble model has outperformed every other deep learning model.

6 Conclusion and Future Scope

In light of the preceding section, it can be concluded that the accuracy and sensitivity of the ensemble model of various deep learning models are higher than any deep learning model. The processing of each step has an impact on the key performance indicators, mainly on the accuracy calculated at the end. The ensemble model was successful in classifying brain tumors with high accuracy. In future, we will try other deep learning models as base learners for developing ensemble techniques to further improve the prediction accuracy and robustness of the predictive model for brain tumor classification.

References

1. Isina A et al Review of MRI based brain tumor image segmentation using deep learning methods. Procedia Comput Sci 102:317–324
2. Abbas et al (2019) Automatic brain tumor detection in medical imaging using machine learning. In: 2019 international conference on information and communication technology convergence (ICTC). https://doi.org/10.1109/ictc46691.2019.893974
3. Pereira et al (2018) Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans Med Imaging 35:1240–1251
4. Kang J, Ullah Z, Gwak J (2021) MRI-based brain tumor classification using ensemble of deep features and machine learning classifiers. Sensors 21(6):2222. https://doi.org/10.3390/s21062222
5. Tandel GS, Tiwari A, Kakde OG (2021) Performance optimisation of deep learning models using majority voting algorithm for brain tumour classification. Comput Biol Med 135:104564. ISSN 0010-4825
6. Abiwinanda N, Hanif M, Hesaputra ST, Handayani A, Mengko TR (2019) Brain tumor classification using convolutional neural network. In: World congress on medical physics and biomedical engineering 2018. Springer, pp 183–189
7. Sajjad M, Khan S, Muhammad K, Wu W, Ullah A et al (2019) Multi-grade brain tumor classification using deep CNN with extensive data augmentation. J Comput Sci 30:174–182

8. Cheng J, Huang W, Cao S, Yang R, Yang W et al (2015) Correction: enhanced performance of brain tumor classification via tumor region augmentation and partition. PLoS ONE 10:e0144479
9. Singh G, Ansari M (2016) Efficient detection of brain tumor from MRIs using K-means segmentation and normalized histogram. In: 2016 1st India international conference on information processing, Delhi, India, pp 1–6
10. Noreen N, Palaniappan S, Qayyum A, Ahmad I, Alassafi MO Brain tumor classification based on fine-tuned models and the ensemble method. Comput Mater Continua. https://doi.org/10.32604/cmc.2021.014158
11. Bhuvan S (2020) [Brain Tumor Classification (MRI)] [Data Set]. Kaggle
12. Tao et al (2020) NSCR-based DenseNet for lung tumor recognition using chest CT image. Hindawi, 16 Dec 2020, 2020/6636321/
13. Huang G et al (2017) Densely connected convolutional networks [Paper presentation]. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708. arxiv.org/abs/1608.06993
14. Jaiswal A et al (2020) Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. J Biomol Struct Dyn 39:5682–5689
15. Srivastava N et al (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1):1929–1958
16. Singh S (2018) A comprehensive guide to ensemble learning (with Python codes). Analytics Vidhya, 2018. [Online]. Available https://www.analyticsvidhya.com
17. Hong P et al (2007) Accuracy of classifier combining based on majority voting. In: 2007 IEEE international conference on control and automation, Guangzhou, China, pp 2654–2658
18. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv 1409.1556

Analysis of Optimized Algorithms for Quality of Service in IoT Communication

Uma Tomer and Parul Gandhi

Abstract The Internet of Things (IoT) has evolved over the last decade to connect a massive number of objects. These connected objects provide a vast number of services to enhance the daily life of end users. The Internet of Things is based on integrating several processes, like networking, sensing, identification, and computation. Many IoT applications like health, agriculture, logistics, retail, transport, smart cities, remote control, process automation, smart metering, etc., can be grouped into different fields. The Internet of Things (IoT) was used for automating human life during the provision of its services. This incredible technology has the potential to improve and ease human life significantly. Quality of Service Enhancement metrics must be clearly defined to enhance the popularity of any service. This paper puts forward a detailed analysis of algorithms for optimizing Quality of Service (QoS) in IoT. This review identifies and describes various IoT quality of service metrics enhancement through optimization, keeping in mind that the three pillars of IoT that are computing, communication and things. Author presents a review of highlights of different securities and privacy issues and conclude that optimization provides advantages in different applications of IoT considering the significance of QoS metrics such as power consumption, life time, and latency.

Keywords IoT · Optimization · QoS · QoS metrics · Computing · Communication · Energy consumption

U. Tomer (✉) · P. Gandhi
FCA, Manav Rachna International Institute of Research and Studies, Faridabad, Haryana, India
e-mail: uma.tomer@gmail.com

P. Gandhi
e-mail: parul.fca@mriu.edu.in

1   Introduction

In the twenty-first century, IoT is among the rapidly growing techniques, and it is ready to impact the masses. It has proven helpful in many fields, including ecological monitoring, environmental changes, Medicaid, smart homes, smart cities, and so on. These IoT devices have their own digital recognition, so it can easily connect to the environment, enabling them to share data, allowing them to interact and communicate intelligently. Things can link without requiring contact among two humans or a person and a digital device, as we see with computers, cell phones, laptops, and tablets, thanks to the IoT. To meet the needs of IoT users, service providers have created a variety of uses. For the betterment of services, customers requested Quality of Service (QoS) which varies from person to person and also according to different area of IoT. To describe a consumer's expectations and service providers, evident explanation of applications quality metrics to adjust correspondingly. The researcher should focus on defining the QoS metrics for describing the expectations of Internet of Things service users. The IoT eco system's communication networks are in charge of transporting real-time information and uses worldwide [1]. The new IoT generation has been developed to achieve service quality (QoS). In the IoT surroundings, bandwidth is the essential source, and its management is necessary for greater QoS. Furthermore, in the IoT environment, the demand for multimedia services has increased. Different features and QoS methods are required for various services in the IoT surroundings [2]. Security in the IoT is still insufficient. Because of IoT, devices are enabled with low-level resources and low-level computing, making security a challenge. In general, a lightweight optimization to work with IoT that ensures safety and privacy against attackers is required. In the IoT field, this optimization will provide confidentiality, data integrity, and authentication. Static program analysis uses various computer science areas that analyze source code to decide a program's properties at compile time [3]. The static analysis tool approximates the real program behavior by generating an internal model of the source code. Also, mathematically robust methods and heuristics to recognize particular coding errors may correspond to run-duration errors like memory leaks and crashes or security vulnerabilities. Issues with IoT protection in every device with its availability, integrity, and confidentiality are common. IoT has several problems that create security a difficult task, e.g., the heterogeneous nature of nodes with Internet access with fewer embedded security devices [4]. Optimization has various benefits, including security, privacy, data recovery, removing redundant data, increasing data rate, and improving application and network response times [5].

2 Architecture of IoT

The IoT includes network connectivity and interconnectivity, machine-human interfaces, and intelligent devices and sensors. All of this, when combined, will provide a business enablement aspect, giving the world of Information technology a new dimension [6]. The composition of architecture of IoT can be broadly decomposed into components: Internet, to promote a network-aligned vision of the IoT, and Things, which focuses on artifacts that can be integrated into a shared structure. However, as a collaborative system, the Internet of Things consists of three fundamental components: things, communication, and computing.

2.1 Things

Devices with sensors and the ability to connect to the Internet are included in the category of things. Intelligent machines, monitors, persons, and all other objects that are aware of their surroundings and can communicate with other entities at any time fall into this category.

2.2 Communication

At the scale of IoT deployment, the network must provide communication, security, privacy, and computation. The major networking element is individually placed to support communication among devices and with the outside world using the network's protocols.

2.3 Computing

The mined data gathered from things must be analyzed, and calculation is required. These calculating services could be used by edge devices, fog nodes, and cloud nodes. Calculating should allow for intelligent deductions and suggestions within a defined system, forming the link between real application and user behavior analysis.

3 QoS Metrics

To supply services to IoT, QoS helps in managing a system's capabilities and resources. It allows service providers to give their customers plain visibility of their services, including their performance and usability. Service Level Agreements (SLAs) between IoT service providers and IoT customers can be implemented using QoS. By using precisely described QoS metrics, customers will be able to identify best services of IoT which will also help with service quality optimization (Fig. 1).

3.1 QoS in Communication

It is the responsibility of communication networks in the IoT ecosystem to transport real-time data and applications. In communication, delay-sensitive as well as delay-tolerant applications are possible. Furthermore, to meet the requirements of various IoT applications, it is essential to increase the network services value to improve their quality. In the following sections, author goes through the various quality metrics that can be found in a network structure.

Jitter: Undesirable outcome induced by an unbalanced deviation in the data packet's travel time intervals is termed as jitter. The delay among all packets varies, instead of remaining constant. Jitter in a network will be due to, network congestion, improper queuing, or configuration error. Packet loss and network congestion can also be caused by jitter. For a quality network connection, the service providers of network must give sufficient bandwidth and probable latency; this can help to reduce jitter [7].

Throughput and Efficiency: A metric for how many data packets is passed or received over a network is throughput. The amount of available bandwidth that a

Fig. 1 QoS-based metrics

network has, is calculated in (bps). The throughput of the network reduces as network latency enhances.

Monetary Cost: Only bandwidth is used to estimate network quality. Furthermore, as bandwidth increases, the prices increase as well. Users want a network including reduced cost and increased bandwidth, but there is a trade-off [8].

Availability: When a network is required for use, its availability is described as the percentage of active and operational duration. It is usually computed in "scales of nine." Both the Internet connection and the VPN tunnel have an impact on communication availability [9].

Bandwidth: The amount of information sent through a network in a provided amount of period is referred to as bandwidth. It is measured in (Mbps). It is preferable to access a network with a large amount of available bandwidth.

Security and Privacy: It is not easy to meet customer's requirements for security and privacy. Specific techniques have been developed to obtain authentication and privacy in a better way [10]. Some Privacy Enhancing Technologies (PET) have been developed. Simultaneously, network protection refers to the steps taken to ensure that the communication channel's accessibility and integrity and the data passing through it are secured. Different encryption methods are utilized to keep the network protected.

Interoperability: In a communication network, interoperability refers to a system's and technology's ability to function together while having various software systems and technical frameworks.

Reliability: The connection among the destination and origin node pair can be used to represent communication reliability. It is only considered trustworthy if the communication medium assures that the information will be delivered to the intended user without security or loss breach. It is known to be reliable if the communication network gives acknowledgment only if information delivery fails.

Service Level Agreement: SLAs are provided by network service providers and define the mediation of services among the service providers and the customers as stated in the contract. The customer will require a tool to assess ability of the service providers to use their services. SLA can be analyzed using the Ocular IP [11] tool.

3.2 QoS in Things

The IoT promises a high number of radio frequency identification (RFID)-enabled sensor systems that are linked to an intelligent network implicitly. The quality parameters of items (sensors) defined as adding a new dimension to IoT applications are described below.

Interoperability: It is described as a sensor's ability to share resources and configuration regardless of their configuration or architecture. Other sensors may be used with them. There is no global standard for measuring interoperability. Compatibility measure, potential measure, and performance measure are the metrics used to determine the interoperability [12].

Flexibility: Sensor flexibility refers to a device's ability to change its waveform and its software and hardware configuration. Whenever a device in a band is not in use, it may change its waveform to reduce peak operation duration. As a result, a most flexible device operating on a single radio band may be utilized to serve various reasons with varying data rate, range, latency, and other specifications.

Availability: The available resources provide adequate amount of energy to the sensors to access information. This can be defined as quality aspect of things. Device sensing efficiency is improved by the availability of signals outside of the minimal range.

Response Time: The duration it takes for the sensors to change their output state in response to a change in the input parameter. Low-response time sensors are ideal for all applications.

Reliability: Reliability refers to a system's ability to self-configure and recover in modifying surroundings. In sensor technique, a bit of error can arise, causing the system to incorrectness. Reliability should be considered during system implementation and design to ensure the high-term applicability of devices or systems.

Sensitivity: Various sensors can be used in various applications, with its sensitivity specifications. The degree to which the result modifies is the measured physical input parameters change [13]. Sensors with low and sustained sensitivity are preferred over sensors with less sensitivity.

Security: Google Cloud Internet of Things Core is a public service given by Google that provides complete stack protection for billions of dispersed heterogeneous devices [14]. It provides the connection, management, and ingestion of data from various sources in a secure and reliable method. Safe communication requires security credentials such as proper user registration, identity authorization, and authentication.

Power Consumption: Charging the device might probably use electricity. Things that use battery or solar energy, on the other hand, are more appealing to users.

Mobility Support: This broad term can encompass all aspects of the Internet of Things device functionality. As per Raja and Su [15], sensor node mobility is divided into weak mobility and strong mobility. A node is considered to have poor mobility if it fails or goes out of service for any reason, while nodes that move freely in the network due to physical factors such as wind or water are considered to have strong mobility.

3.3 QoS in Computing

Since QoS scales and improves the efficiency of applications that generate large amounts of data, cloud computing is an essential element of the IoT. To reduce the various strains of the network, computations may also be executed at the fog or edge. Considering the significance of computing for IoT, author established a few QoS metrics that help customers to relate service providers (edge, fog, or cloud) based on these quality metrics.

Scalability: Scalability in computing applications is defined as achieving maximum throughput with the shortest response time. By 2025, the IoT will be the primary source of large information creation. More users will require greatly scalable computation power, which can be scaled by addition resources, including overhead.

Reliability: By increasing the reliability, the running duration of the application over the leased cloud is also increased. The system disrupts with the extension in number of components connected to the computing system. To make the most of the cloud's abundant resources, it is important to understand how it falls. Mean time between failures (MTBF) will be provided by computing service providers as a measure of reliability.

Pricing: Pricing is another factor when deciding which computing services to use. The three service parameters storage, compute, and network are used to determine computing cost. End-users will prefer only those service providers with best QoS at the economical price.

Response Time: It is the duration taken to respond to a request. It is dependent on both the infrastructure and the application that has made the request.

Privacy and Security: The cloud node or edge node is a physical element of IoT which is vulnerable to attacks and threats includes disabling network providing, illegal access to the consumer's private data, inserting inaccurate information, and so on. The integrity and confidentiality of the information at the computing node must be guaranteed when it comes to computer security [16].

4 Enhancement of QoS in IoT Using Optimization

The Internet is being utilized to share information between a slew of small, resource-constrained devices that make IoT so popular with its technological advancements. The IoT network obtains a huge amount of data from these devices. To allow the best use of the available network, solutions are needed for different network-compared issues such as congestion, security, heterogeneity, reliability, scalability, QoS, energy conservation, and routing in the IoT. In this section, an all-inclusive study of QoS optimization in the Internet of Things is discussed. The review concludes that quality is always enhanced using optimization for IoT communication.

4.1 IoT Network's QoS Attributes

An IoT network's Quality of Service parameter is examined from different dimensions and perspectives, including packet loss rate, jitter, latency, interference avoidance, and bandwidth. As a result, QoS must be defined differently for every technology. Liang et al. [17] proposed DRX/DTX as a discontinuous transmission/reception (DTX/DRX) process for 3GPP LTE-A, which ensures packet missed rate,

traffic bit rate, and packet delay in IoT applications while storing energy in a Quality of Service context. To make the most of the resources in long-term evolution (LTE) air interface, Piri and Pinola [18] checked packets of various sizes in the LTE uplink. Smaller packets are used with nearly half the throughput of larger packets, as per the results. This result allows joining of multiple packets at the Internet of Things gateway's mobile end to increase the Quality of Service parameters, including, jitter, bandwidth, latency, and packet loss usage for an increased number of the small packet.

4.2 QoS Optimization

In general, QoS optimization is characterized as a technology that improves network efficiency in any surroundings. Since large amounts of data from various sources is constantly being embedded into the network, it is quite significant in IT. Higher elimination of redundant data improves application and network response time, data recovery, and data rates. This section deals with the requirement of network optimization in the Internet of Things. Various algorithms for QoS optimization in the Internet of Things are discussed and further compared with various QoS parameters that explains network parameters to support these algorithms.

4.2.1 Review of QoS Optimization in IoT

Internet of Things QoS optimization has got more attention as billion of IoT devices are suggested to link to the global network in the future. As a result, operators and researchers must develop practical solutions for optimizing IoT networks to decrease the impact of the IoT-produced traffic on other network services and allow efficient use of network resources. The traffic produced by IoT devices differs from that generated by mobile networks because of the heterogeneity of benefits and device categories. For screening the operation of IoT services and devices, Internet of Things traffic is regulated. While the IoT applications generate less data, device integration into the application generates more traffic due to control plane messages. As a result, non-application traffic places an important extra load on the network. An efficient mechanism to resolve and optimize control plane information from the Internet of Things devices is required to alleviate the burden (Table 1).

The QoS optimization and IoT is the recent trend in the IT industry, and in this section, author considered the articles which highlighted QoS parameter in IoT for optimization.

Table 1 Optimization for QoS enhancement in IoT

S. No	Authors	Contribution	Merits
1	Hasan et al. [19]	Canonical particle multi-swarm optimization algorithm	Delay and energy expenditure are reduced
2	Jaiswal et al. [20]	An ecological routing protocol for IoT applications depends on wireless service networks (WSNs)	This protocol improves delivery ratio, end-to-end obstructions, packet network lifetime, and best-concerning energy or power saving
3	Suseela et al. [21]	PBMRA algorithm	The PBMRA algorithm improved overall throughput and packet rate
4	Chandra et al. [22]	A scheme for improving QoS by handover management	The re-authentication delay is reduced. To enhance security and seamless handover is obtained
5	Singh et al. [23]	Nature-inspired grasshopper optimization algorithm is dependent on optimization routing scheme	Reduced packet loss, reduced time delay, and increased throughput
6	Hamidouche et al. [24]	A genetic algorithm depends on methods for routing and clustering in WSNs	This method is used to extend the life of a sensor and improve QoS
7	Makkena et al. [25]	Clustering technique	Thus, the Quality of Service in WSN is improved
8	Chen et al. [26]	Adaptive retransmit mechanism is implemented for delay in differentiated services (ARM-DDS)	ARM-DDS scheme reduces data transmission delay along with energy consumption
9	Subashini et al. [27]	Smart duty cycle scheduling protocol (SDCS)	By avoiding data redundancy and transmission, the network's lifetime can be increased. Besides, the throughput and packet delivery ratio have also improved
10	Tang et al. [28]	Eco-friendly and dependable routing algorithms depend on DS evidence theory (DS-EERA)	In the meanwhile, it will reduce packet loss and increase data transmission efficiency
11	Preeth et al. [29]	Efficient clustering and immune-inspired routing (FEEC-IIR) is based on adaptive fuzzy rule	Various QoS parameters can be enhanced such as packet loss ratio, throughput, end-to-end impede, network lifetime, and ratio of packet delivery

(continued)

Table 1 (continued)

S. No	Authors	Contribution	Merits
12	Schormans et al. [30]	Quantum particle swarm optimization (QPSO)	Network lifetime
13	Sarang et al. [31]	Energy-efficient asynchronous Quality of Service (AQSen) MAC protocol, known as AQSen-MAC	To enhance the network throughput and packet delivery ratio of up to twelve percent in the network
14	Cao et al. [32]	For battery-powered IoT applications, mobility-aware network lifetime maximization	This method can obtain up to 169.52 percentage network period
15	Baroudi et al. [33]	A genetic algorithm (GA-TBR)	Reduce delay
16	Jazebi et al. [34]	Routing scheme of IoT utilizing SFLA	Enhance the network lifetime and reduce energy consumption
17	Kulkarni et al. [35]	Multi-objective hybrid routing algorithm is used to enhance QoS	To increase the packet delivery ratio
18	Restuccia et al. [36]	Swarm intelligence-depend sensor selection algorithm (SISSA),	Enhanced network period
19	Rani et al. [37]	Proactive routing algorithm and optimization technique	Improve the network lifetime
20	Alghamdi et al. [38]	The hybrid model of route optimization	Improved the QoS

4.2.2 Energy Consumption

The QoS optimization and IoT are the recent trend in the IT industry, in this section; we mainly consider the articles which highlighted QoS parameter, IoT, and optimization. Hasan et al. [19] investigated the connectivity pattern of sensors in the Industrial Internet of Things (IIoT) through canonical particle multi-swarm optimization (CPMSO) algorithm. This algorithm facilitates efficient deployment by constructing a network that can withstand failure while maintaining a high quality of service through delay, energy consumption, and throughput. To decrease energy consumption and delay, a multi-swarm optimization approach was used to determine the deployment pattern of disjoint path routes. Chandra et al. [22] defined a handover management scheme for 3GPP-compliant network subscribers roaming between WiMAX and WLAN to improve QoS. The time taken to re-authenticate has been reduced. During the handover among WiMAX and WLAN HO, the signaling cost and re-authentication delay were significantly reduced. The time taken to re-authenticate

has been reduced. This enhances the security, and also seamless handover is reached. Singh et al. [23] described the latest optimization routing approach depending on the grasshopper optimization algorithm that enhances sensor networks' energy consumption while also improving service support quality by decreasing time delay and reducing packet loss.

Chen et al. [26] examined a researcher investigating ARM-DDS scheme to meet various application delays. First, we look at how different retransmit mechanisms affect delays, energy consumption, and how parameter optimization can help. In many applications, data routing delays are guaranteed to be within a specific range, and the network's energy consumption is reduced. The ARM-DDS scheme improves network energy while reducing data transmission delay by 12.1%. Tang et al. [28] explained a routing algorithm depends on DS evidence theory, both energy-efficient and reliable (DS-EERA). Firstly, three attributes are indicated as proof, residual energy, traffic, proximity of its path along with shortest path of neighboring nodes by DS-EERA. DS-EERA is a promising technology that can help to extend the life of a network. In the meantime, it will achieve depreciated rate of packet loss and enhance data transmission reliability. Preeth et al. [29] introduced a WSN-assisted IoT system for adaptive fuzzy rule-based energy-efficient clustering and immune-inspired routing (FEEC-IIR) protocol. The AF-MCDM, a fuzzy TOPSIS, and AHP approach are presented as an energy-effective clustering algorithm for optimal cluster head selection. The IIR optimization algorithm is utilized to improve the information delivery dependency for routing. For reducing energy consumption, a cost-effective method, cluster-depend routing, is used. This approach enhanced the parameters of QoS such as network lifetime, packet loss ratio, jitter, end-to-end delay, bit error rate (BER), channel load, throughput, packet delivery ratio, buffer occupancy with comparison of energy consumption and extant clustering and routing approach.

Baroudi et al. [33] developed a genetic algorithm (GA-TBR) to gather state information within the smart grid WSN environment and then optimize route selection to ensure the required QoS. With the ad hoc on-demand distance vector routing protocol, this algorithm will choose routes with the least amount of delay. Jazebi et al. [34] presented a routing scheme for the IoT utilizing shuffled frog leaping algorithm (SFLA). The shuffled frog leaping algorithm is utilized for Internet of Things by the routing scheme. Using the shuffled frog leaping optimized algorithm (RISA), IoT routing scheme is used to determine content depend path among the destination and source nodes. Using an appropriate information aggregation scheme, RISA can increase network period and decrease energy consumption. This approach will enhance power consumption, increase network lifetime, increase throughput, and increase packet delivery ratio (PDR). Kulkarni et al. [35] presented QoS-assured multi-objective hybrid routing algorithm (Q-MOHRA) which is a novel heuristic routing algorithm for heterogeneous WSNs. When compared to secure hierarchical routing protocol (SHRP), it enhances packet delivery ratio by 24.31%, and when compared to dynamic multi-objective routing algorithm (DyMORA), it enhances packet delivery ratio by 11.86%. Using mean energy consumption, Q-MOHRA increases DyMORA by a factor of 8.27%. Alghamdi et al. [38] presented a hybrid route optimization model that improves service quality in multi-hop sensor networks.

An algorithm is provided for selecting an optimized route from source to destination by taking into account various network parameters and bandwidth. All available routes from source to destination are ranked as per their weights. A network simulator is used to test the algorithm.

4.2.3   Network Lifetime

An energy-effective routing protocol for WSN provides IoT benefit containing unfairness in the network including increased traffic load. To choose an optimal path, this protocol assumes three factors: lifetime, traffic intensity, and reliability at the next-hop node. Jaiswal et al. [20] presented a rigorous simulation utilizing NS-2. Compared to other protocols, this protocol saves more energy, has a higher packet delivery ratio, has a more extended network lifetime, and has a shorter end-to-end delay. Suseela et al. [21] developed a priority-based multipath routing algorithm (PBMRA). In an experimental setting, this algorithm is compared to other existing algorithms such as ad hoc on-demand distance vector (AODV), dynamic source routing protocol (DSR), destination sequenced distance vector (DSDV), and optimized link state routing protocol (OLSR). In terms of overall throughput, packet received rate, and path reliability, the PBMRA algorithm outperforms existing algorithms.

Hamidouche et al. [24] presented an approach based on genetic algorithms for clustering and routing in WSNs. This method aims to make the sensor's life longer and improves service quality. In terms of different performance measures, such as consumption of energy and the amount of packets accepted by the base station, this algorithm outperforms existing algorithms. Makkena et al. [25] presented a clustering technique that automatically elects the cluster heads based on their residual energy and sensor utility to its cluster members. There are an equal number of cluster members for each cluster head. In the case of obstacles, the route is optimized using the shortest path algorithm of Dijkstra that decreases average energy, packet lifetime, hop count, and packet delay. As a result, wireless sensor network's (WSN) service quality has improved. Subashini et al. [27] investigated an innovative duty cycle scheduling protocol at schedule dynamic circuit services (SDCS) that helps networks to last longer by removing energy efficiency transmission conflicts, data redundancy, and traffic conflicts. By decreasing the interval among packets, the end-to-end delay is reduced, and the packet delivery ratio increases by eliminating packet miss. The packet delivery ratio and throughput have also enhanced as a result of the performance. Both the end-to-end time delay and energy consumption are decreased to 0.69 s and 0.076 J, correspondingly. Schormans et al. [30] combined the quantum particle swarm optimization (QPSO) with the improved non-dominated sorting genetic algorithm to obtain the Pareto optimal front (NSGA-II). This algorithm achieves the trade-off between energy efficiency and QoS provisioning by sacrificing about 10% network lifetime while improving 15% outage performance.

Sarang et al. [31] presented an energy-effective Asynchronous Quality of Service (AQSen) Multiple Access Control protocol, known as AQSen-Multiple Access

Control. To strengthen the efficiency of energy, packet delivery ratio, and network throughput, the AQSen-Multiple Access Control assumes various kinds of data packets and employs 2 novel methods: scheduling and self-adaptation specify the receiver adapts its service cycle based on the residual energy to ensure the current. The AQSen-Multiple Access Control protocol decreases utilization of energy at the receiver end up to 13.4%, reduces consumption per bit by up to three percent, and enhances network throughput and packet distribution ratio by up to twelve percent in the network. Cao et al. [32] introduced a technique for battery-powered Internet of Things. They introduce a mobility-aware network for period maximization that approximates real-duration calculation below the QoS constraint. The method scheme is divided into two stages: online and offline. By utilizing the mixed-integer linear programming (MILP) method, an optimal transportability-aware task is used to increase the network period, which is extracted at the offline stage. Adoration-effective and cross-entropy technique is presented at the online stage to adapt task execution for fluctuating Quality of Service requirements. When compared to bench-marking solutions, this technique came as an improvement on network lifetime by up to 169.52%.

Restuccia et al. [36] described the problem of optimizing a WSN's lifetime in the presence of uncontrollable and random sink mobility while adhering to QoS constraints, which is still unsolved. Followed by SISSA, a novel swarm intelligence-based sensor selection algorithm is described that develops the lifetime of a network while complying to QoS constraints which were predetermined. As per experimental and analytical results, SISSA provides 56% of the lifetime through ideal approach which makes it scalable and energy-efficient, in all network parameter sets considered. Rani et al. [37] developed a model to improve the network lifetime of WSN; the researcher applied a hybrid optimization method including a proactive routing algorithm (WSN). In the destination sequence distance vector (DSDV) routing protocol, the bacteria foraging optimization soft computing methods, and genetic algorithm are used separately, followed by the hybridization of the genetic algorithm and bacterial foraging optimization in the same routing protocol. Throughput, routing overhead, congestion, end-to-end delay, bit error rate, and packet delivery ratio are utilized as parameters for simulation. The hybrid DSDV routing protocol outperforms the standard DSDV routing protocol and is better suited for small WSNs.

5   Conclusion

Several IoT advancements have already been established in the literature, as well as in real-time applications and networks. One of the main challenges that Internet of things will face in the coming years is optimization for quality of service. In this review paper, an encyclopedic analysis on the optimization for QoS in IoT is presented. IoT parameters including congestion, scalability, reliability, heterogeneity, and security are presented along with state-of-art optimization methods. The review

wraps up with associated issues and challenges for QoS in IoT optimization. Eventually, different issues and challenges of optimization for QoS in Internet of Things are delineated. The conclusion of this paper emphasizes the significance of QoS optimization in the Internet of Things, which provides numerous perks which includes brisk data rates, security issues, data recovery with improved QoS by eliminating iterating data, and enhanced application and network response times.

References

1. Sandeep KV, Abdulhayan S (2020) Implementation of data integrity using MD5 and MD2 algorithms in IoT devices. PalArch's J Archaeol Egypt/Egyptol 17(7):7388–7395
2. Hameed A, Alomary A (2019) Security issues in IoT: a survey. In: 2019 international conference on innovation and intelligence for informatics, computing, and technologies (3ICT). IEEE, pp 1–5
3. Huuck R (2015) IoT: the internet of threats and static program analysis defense. In: EmbeddedWorld 2015: exhibition and conferences, p 493
4. Zejun R, Xiangang L, Runguo Y, Tao Z (2017) Security and privacy on internet of things. In: 2017 7th IEEE international conference on electronics information and emergency communication (ICEIEC), July 2017
5. Srinidhi NN, Kumar SMD, Venugopal KR (2019) Network optimizations in the internet of things: a review. Eng Sci Technol Int J 22(1):1–21
6. Singh M, Baranwal G (2018) Quality of service (QoS) in internet of things. In: 2018 3rd international conference on internet of things: smart innovation and usages (IoT-SIU). IEEE, pp 1–6
7. Jitter (VoIP). [Online]. Available https://www.techopedia.com/definition/3041/jittervoip-voice-over-internet-protocol-voip
8. Salih YK, See OH, Ibrahim RW, Yussof S, Iqbal A (2014) A network selection indicator based on golden relation between monetary cost and bandwidth in heterogeneous wireless networks. Res J Appl Sci Eng Technol 7(3):478–483
9. Gridelli BS (2014) How to calculate network availability? [Online]. Available https://netbeez.net/blog/how-to-calculate-networkavailability/
10. Weber RH (2010) Internet of Things—new security and privacy challenges. Comput Law Sec Rev 26(1):23–30
11. OcularIP. [Online]. Available https://www.localbackhaul.com/ocularip/
12. Ducq Y, Chen D (2008) How to measure interoperability: concept and approach. In: 2008 IEEE international technology management conference, pp 1–8
13. Kalantar-zadeh K (2013) Sensors 11–29
14. CLOUD IOT CORE. [Online]. Available https://cloud.google.com/iot-core/
15. Raja A, Su X (2009) Mobility handling in MAC for wireless ad hoc networks. Wirel Commun Mob Comput 9(3):303–311
16. Habib SM, Ries S, Mühlhäuser M (2010) Cloud computing landscape and research challenges regarding trust and reputation. In: 2010 7th international conference on ubiquitous intelligence and computing and 7th international conference on autonomic and trusted computing (UIC/ATC 2010), pp 410–415
17. Liang J-M, Chen J-J, Cheng H-H, Tseng Y-C (2013) An energy-efficient sleep scheduling with QoS consideration in 3GPP LTE-advanced networks for internet of things. IEEE J Emerg Sel Topics Circuits Syst 3(1):13–22
18. Piri E, Pinola J (2016) Performance of LTE uplink for IoT backhaul. In: 2016 13th IEEE annual consumer communications and networking conference (CCNC). IEEE, pp 6–11

19. Hasan MZ, Al-Rizzo H (2019) Optimization of sensor deployment for industrial internet of things using a multiswarm algorithm. IEEE Internet Things J 6(6):10344–10362
20. Jaiswal K, Anand V (2019) EOMR: an energy-efficient optimal multi-path routing protocol to improve QoS in wireless sensor network for IoT applications. Wirel Pers Commun 1–23
21. Suseela S, Eswari R, Nickolas S, Saravanan M (2020) QoS optimization through PBMR algorithm in multipath wireless multimedia sensor networks. Peer-to-Peer Netw Appl
22. Chandra I, Prabha KH, Sivakumar N (2018) Optimization of QoS parameters using scheduling techniques in heterogeneous network. Pers Ubiquitous Comput 22(5–6):943–950
23. Singh A, Sharma A Optimizing energy efficiency in wireless sensor networks on various QoS parameters using grasshopper optimization algorithm
24. Hamidouche R, Aliouat Z, Gueroui AM (2018) Genetic algorithm for improving the lifetime and QoS of wireless sensor networks. Wirel Pers Commun 101(4):2313–2348
25. Makkena T, Kothapalli A, Swain G (2017) Improving quality of service in wireless sensor networks. Int J Pure Appl Math 116(5):147–152
26. Chen Y, Liu W, Wang T, Deng Q, Liu A, Song H (2019) An adaptive retransmit mechanism for delay differentiated services in industrial WSNs. EURASIP J Wirel Commun Netw 2019(1):258
27. Subashini S, Mathiyalagan P (2020) A cross layer design and flower pollination optimization algorithm for secured energy efficient framework in wireless sensor network. Wirel Pers Commun 1–28
28. Tang L, Lu Z, Fan B (2020) Energy efficient and reliable routing algorithm for wireless sensors networks. Appl Sci 10(5):1885
29. Preeth SKSL, Dhanalakshmi R, Kumar R, Shakeel PM (2018) An adaptive fuzzy rule based energy efficient clustering and immune-inspired routing protocol for WSN-assisted IoT system. J Ambient Intell Humanized Comput 1–13
30. Schormans JA (2015) QoS-aware energy efficient cooperative scheme for cluster-based IoT systems. IEEE Syst J
31. Sarang S, Stojanović GM, Stankovski S, Trpovski Ž, Drieberg M (2020) Energy-efficient asynchronous QoS MAC protocol for wireless sensor networks. Wirel Commun Mob Comput 2020
32. Cao K, Xu G, Zhou J, Wei T, Chen M, Hu S (2018) QoS-adaptive approximate real-time computation for mobility-aware IoT lifetime optimization. IEEE Trans Comput-Aided Des Integrated Circ Syst 38(10):1799–1810
33. Baroudi U, Bin-Yahya M, Alshammari M, Yaqoub U (2019) Ticket-based QoS routing optimization using genetic algorithm for WSN applications in smart grid. J Ambient Intell Humaniz Comput 10(4):1325–1338
34. Jazebi SJ, Ghaffari A (2020) RISA: routing scheme for Internet of Things using shuffled frog leaping optimization algorithm. J Ambient Intell Humanized Comput 1–11
35. Kulkarni N, Prasad NR, Prasad R (2018) Q-MOHRA: QoS assured multi-objective hybrid routing algorithm for heterogeneous WSN. Wirel Pers Commun 100(2):255–266
36. Restuccia F, Das SK (2015) Lifetime optimization with QoS of sensor networks with uncontrollable mobile sinks. In: 2015 IEEE 16th international symposium on a world of wireless, mobile and multimedia networks (WoWMoM). IEEE, pp 1–9
37. Rani S, Balasaraswathi M, Reddy PCS, Brar GS, Sivaram M, Dhasarathan V (2020) A hybrid approach for the optimization of quality of service metrics of WSN. Wirel Netw 26(1):621–638
38. Alghamdi TA (2020) Route optimization to improve QoS in multi-hop wireless sensor networks. Wirel Netw 1–7
39. Jadaun A, Alaria SK, Saini Y (2021) Comparative study and design light weight data security system for secure data transmission in internet of things. Int J Recent Innov Trends Comput Commun 9(3):28–32. https://doi.org/10.17762/ijritcc.v9i3.5476
40. Ashish VK, Alaria SK, Sharma V (2022) Design simulation and assessment of prediction of mortality in intensive care unit using intelligent algorithms. Math Stat Eng Appl 71(2):355–367. https://doi.org/10.17762/msea.v71i2.97

41. Alaria SK, Jadaun A (2021) Design and performance assessment of light weight data security system for secure data transmission in IoT. J Netw Sec 9(1):29–41
42. Khandelwal R, Mukhija MK, Alaria SK (2021) Numerical simulation and performance assessment of improved particle swarm optimization based request scheduling in edge computing for IOT applications. New Arch-Int J Contemporary Arch 8(2):155–169

IoT-Based Measurement and Prediction of Water Quality Index: A Review

Pooja Kumari, Baban Kumar S. Bansod, and Lini Mathew

Abstract Under the impact of populace development, contamination, and environmental change, drinking water is turning into a scant asset. Water quality impacts human wellbeing or health and the climate. The water quality index (WQI) is a fundamental pointer for good water executives. A water quality index (WQI) considers a few quality factors all the while. Precise water quality expectation assumes a crucial part in further developing water management and pollution control. Different scientists have utilized a scope of WQI model applications to survey the water nature of streams, lakes, and supplies. The essential point of this review was to fundamentally audit the furthermost ordinarily utilized WQI models and figure out which were the utmost reliable. In this research, the main aim is to examine distinct WQI models. The majority of models included four main parts, while the details varied greatly. Key concerns impacting model accuracy include uncertainty and eclipse issues and application of WQI model.

P. Kumari · L. Mathew
Department of Electrical Engineering, National Institute of Technical Teacher's Training and Research, Chandigarh, India
e-mail: lini@nitttrchd.ac.in

B. K. S. Bansod (✉)
Department of Material Science and Sensor Application (MSSA), CSIO-CSIR, Chandigarh, India
e-mail: babankumarcsio@gmail.com

1 Introduction

Water is fundamental factor for life on the planet. In any case, numerous nations face freshwater shortage [1]. Sadly, nations, where freshwater is more open, are undependable from water-connected problems. Water effluence has been accounted for quite a long time as a developing concern [2, 3]. The World Water Council (WWC) predicts a 40% to half expansion in the total populace in the following 50 years [4]. Water superiority administration necessitates the assortment and examination of enormous water superiority informational indexes that can be challenging to evaluate and integrate. Different instruments have been created to evaluate water superiority information; the water quality index (WQI) model is one such device. In an overview led in 2018, it was viewed that as half of the labor force in India is engaged with the rural area, yet its commitment is simply 16% to GDP [5]. As of late, roused by the advancement in the Internet of Things (IoT), different IoT-based answers for water observing have been contrived [6]. Be that as it may, these frameworks are over the top expensive, or their models are shut to public use [7]. The use of the WQI was first presented by Robert et al. [8] and changed through various experts. By and large, WQI plans incorporate broad estimations and in this way demand significant investment and exertion. Moreover, customary techniques for working out the WQI require a lot of physical and compound information, normally at everyday spans. Thus, elective methodologies are expected to work out WQI in a computationally proficient and precise manner. Such improvement can help ecological specialists while checking and assessing water superiority.

Most pieces of the WQI model were made considering capable evaluation and close by rules [9]. An internal quality of water all around contains four cycles or parts. In the first place, the water superiority limits of interest are picked. Second, the water superiority data is scrutinized, and for each water superiority limit, the obsessions are changed over into a singular-regarded dimensionless rundown. Third, the weighting factor for each water superiority is not completely, firmly established, and fourth, a last single-regarded water superiority stills up in the air by an assortment capacity using the records and weighting components to all water superiority limits [10].

The survey recognized six segments in which Segment I contains the introductory piece of the water quality index, Segment II presents concise history on WQI models, Segment III presents literature review of various papers, Segment IV portrays IoT in water quality index observing, utilized strategies of ANN, and its application, Segment V involves basic discoveries, suggestions, and future work, and Segment VI contains the end part of this review.

2 A Brief History on WQI Models

According to the researcher, there are eleven fundamental water quality index models developed. The first fundamental model was developed by the Horton Index in 1960 which was also known as the first WQI model in which he considered 10 water superiority boundaries in many waterbodies, USA. After five years, it had first modification and produce National Sanitation Foundation (NSF) Index in 1965. Similarly, after some years, it had the second modification which leads to Dinius Index in 1972 and Oregon Index in 1980 (which is known as modified version of NSF Index, West Java Index in 2017 [11]).

The second well-known model was SRDD Index which was developed in Scotland in 1970. This WQI got modification in 1979 which is known as Bascaron Index, 1979; House Index in 1986; and Dalmatian Index in 1999.

Similarly, some more models were known as Ross Index which was presented in 1977, Environmental Superiority Index which was presented in 1982, the Smith Index which was obtainable in 1990, Dojildo Index (1994), British Colombia WQI (1995), Liou Index (2003), Said Index (2004), Malaysian Index (2007), and Almedia Index (2012).

This paper reviews various researches but identifies basic 21 water quality index models which are used on global basis which can be seen in Fig. 1.
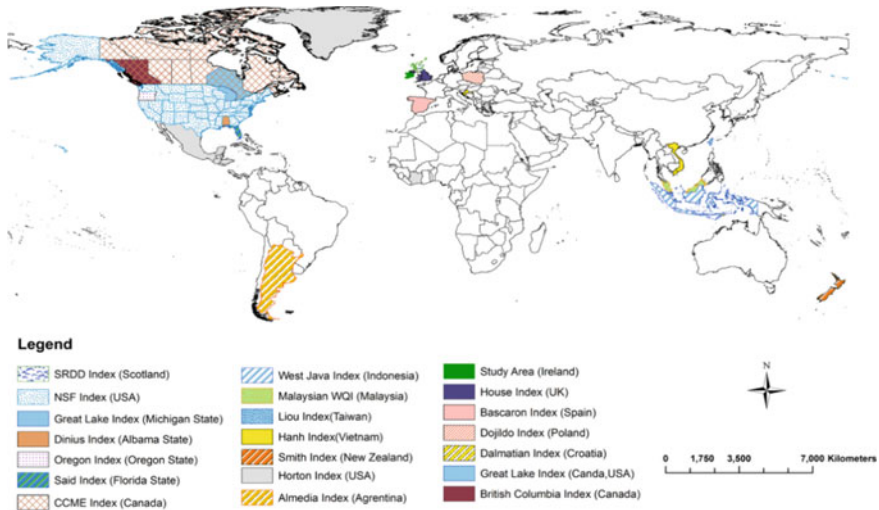
Fig. 1 Globally used WQI models range: 1960–2020 [12]

3   Literature Review

Ahmad et al. used oversaw machine learning and AI estimations to assess the water quality index (WQI), where a particular record was used to summarize the general water quality and water quality class (WQC) [13]. His proposed systems and tendency assisting with a learning speed of 0.1 and polynomial backslide with a degree of 2 expected the WQI most effectively, and this WQI is not totally, permanently established with a mean absolute error (MAE) of 1.9642 and 2.7273. For this present circumstance, it has the most important request accuracy of 85.07%.

Hassan et al. describe the exhibition of AI procedures, for example, random forest (RF), neural network (NN), multiple linear regression (MLR), support vector machine (SVM), and bagged tree model (BTM), anticipating water superiority parts from an Indian water superiority dataset [14]. To do this, the greatest notable factors of the informational collection were obtained, like biochemical oxygen demand (BOD), dissolved oxygen (DO), total coliform (TC), nitrate, pH, and temperature. The outcomes uncovered that the useful models executed well in foreseeing water superiority boundaries; in any case, the best exhibition was attached to MLR with higher accuracy. Further examination will be finished to fabricate models that join the proposed strategy with other profound learning procedures and ways to deal with work on the productivity of the choice cycle.

Verma et al. employed Raspberry Pi, which is a strong IoT gadget, for running the Internet of Things [15]. The water superiority list framework created during this task will help in robotizing the checking of water superiority around us, which is presently a significant risk to regular prosperity in the event that it is disregarded.

Yadav et al. fostered an order model for the expectation of the class of water system that is irrigation water quality index (IWQI) [5]. The IWQI utilized at this time is a total of five boundaries which are individually the convergence of sodium adsorption ratio (SAR), electrical conductivity (EC), chloride, bicarbonate, and sodium particles. For arrangement, they utilized seven-order calculations and chose three boundaries out of five: EC, chlorine, and sodium to anticipate the IWQI class. Unpredictable forest classifier performed the best, followed by gradient boosting and neural networks. Additionally, the dataset covering different kinds of water bodies can be used to chip away at the model. The proposed work used here can be facilitated into IoT-based green structures to inspect water framework water predominance, which will be with more speed and even more fiscally reachable and stood out from manual examination place testing. It can likewise assist with concluding the water test in view of yield and soil properties.

Vasanthavigar et al. determined the water quality index for pre-monsoon (PRM) which shows a higher level of low superiority contrasted with post-monsoon (POM), demonstrating compelling particle draining, overutilization, and anthropogenetic exercises of emanating release for modern, rural, and homegrown use [16]. Higher SAR, RSC, total hardness, and percentage of sodium show that three-fourth of tests remain inadmissible for water system resolutions contrasted with POM because of extended water home time, disintegration of creation minerals lithology, and the

expansion of synthetic composts to water system water. The water superiority in the review region is gradually arriving at a disturbing level, so great arranging is fundamental in this undertaking to protect the delicate biological system.

Reza et al. showed that normal sources (groundwater impacted by neighborhood lithology, enduring, float, and permeation) as well as anthropogenic sources sully groundwater in the review region [17]. Nonetheless, in spite of the transcendent effect of coal mining and other modern exercises, an examination of tests from 24 open and cylinder wells demonstrates that most groundwater tests are normally fit for drinking, with the exception of when the WQI surpasses the permissible furthest reaches of 50. The investigation discovered that groundwater in certain spots needs some level of treatment before it is polished off and furthermore should be safeguarded from the risks of winning contamination.

Sillberg et al. fostered an AI-based approach coordinating attribute realization (AR) and support vector machine (SVM) calculation toward group the water nature [18]. The RA decided the main elements to work on the nature of the stream utilizing the straight capability. In the classification, the qualities that contributed the most were: ammoniacal nitrogen (NH3-N), trichlorobenzene (TCB), fecal coliform bacteria (FCB), BOD, DO, and salt, expanding the contributing qualities in the scope of 0.80–0.98, contrasted with 0.25–0.64 for total dissolved solids (TDS), turbidity, TN, suspended solids (SS), NO3-N, and conductivity. The straight SVM technique gave the best arrangement results addressed by a precision of 0.94, a normal exactness of 0.84, a normal review of 0.84, and a typical F1-score of 0.84. Approval showed that AR-SVM is a strong technique for distinguishing stream water superiority with an exactness of 0.86–0.95 when applied to three to six highlights.

Bui et al. fostered random tree and bagging (BA-RT) AI technique. Their exploration tried four free mining calculations which involve retrieval-based pre-training (REPT), M5 prime tree model (M5P), random forest models (RF), and real-time models (RT) and 12 half breeds to foresee month to month WQI in a sticky northern environment "Iran' [19]. To anticipate IRAQIs, they found that waste coliforms and complete solids had the best and least effect. Here, the ideal information blends contrasted across calculations, yet factors with low relationships performed more regrettable. The half and half calculations worked on their prescient power from a few, yet not every free model and the mixture BA-RT beat different models by accomplishing $R^2 = 0.941$ utilizing a 10-overlap cross-approval procedure, outflanking 15 free models and mixture calculations.

Zhang et al. worked on a hybrid artificial neural network (HANN) model utilizing hereditary calculation for expectation of consumption water handling plants in China [20]. The model was constantly ready, unendingly endorsed using month to month data from 45 DWTPs in China that consolidate eleven data factors for water prevalence and practical execution. The ANN model showed more significant limit and consistency in guessing full-scale water creation from DWTPs blending with water prevalence and useful components. His assumption shows that the HANN model better its show from 0.71 to 0.93 by extending the arrangement data gave, as affirmed by the way that the model can achieve the main degree of execution.

An IoT-based rainfall harvesting solution will be presented by Ranjan et al. [21]; the model has a 60–40 split between the tanks. The rain sensor is located at the top of the system and can tell whether it is raining or not. Using a computer, the machine determines the rainwater's pH and checks to see whether the water's pH is over 5. The servo motor on the hinge turns clockwise with the tank full on the right. The tank is filled on the left side if the pH is less than 5 and the hook spins in the other direction. To isolate pH five or more significant, i.e., pH less than 5, the drinking water must be separated from acidic water. All of this is doable, and it can be done using the Wi-Fi module on the NodeMCU.

Konde et al. proposed that the old method of obtaining water samples and sending them to a laboratory for testing and evaluation may be used in the future [22]. This strategy is time-consuming, inefficient, and wasteful of human resources. They have installed a real-time water quality monitoring gadget that uses a variety of sensors to measure water quality. The presence of pollutants may be seen as a variation in this parameter's magnitude. Data from the sensors is sent over Wi-Fi to the microcontroller, which then sends the data to a smartphone or computer. Using this technology, water supply may be regulated to provide a safe drinking water environment.

With a focus on the convergence of IoT and blockchains, Pincheira et al. suggested an IT platform that includes a compensation plan for more effective agricultural water conservation efforts [23]. For a blockchain water distribution system, the framework proposes a software solution that allows restricted IoT devices to directly exchange data on a public block chain network without any third-party assistance. The suggested method is implemented using off-the-shelf hardware components for the desired purpose. Storage, program size, connection costs, and energy consumption are thoroughly benchmarked.

S. Gupta et al. presented a methodology for evaluating water quality parameters including turbidity, pH, and temperature automatically. Because of its low power consumption and built-in Wi-Fi, the ESP32 was chosen for underwater communication [24]. The IoT-based model included communication modules, a turbidity meter, and a pH sensor. In addition, a machine understanding algorithm based on K-means was employed to assess water quality using pre-determined parameters. The locomotive model has been created to continually monitor water quality in big and small water bodies. The readings are available on a website that the Central Pollution Control Board (CPCB) can access. A robot may be used to monitor water quality from any location. The designed model is low-cost, and the robot can communicate from underwater using high-speed Wi-Fi, making this project self-sufficient and efficient.

Smart water meters that use the Internet of Things (IoT) and cloud computing with machine learning algorithms may be developed and implemented in any location, according to Ray et al. [25] in the year 2020. Clean water conservation has become a global issue in recent years. Singaporeans might save water for future generations by utilizing a smart water meter. Using cloud technology, sensors will monitor the hydraulic data and give autonomous control and alarm alerts. This study's findings may be used to take action if they are analyzed critically.

4 IoT in Water Quality Index Monitoring

The Internet of Things (IoT) is an incorporated framework that has been generally useful from shrewd production network, savvy urban communities, and brilliant power matrices to savvy wearable gadgets [26]. The IoT has produced impressive interest as a way to deal with interfacing regular items through sensors and making an organization to send and get information at an inferior cost. This advancement can advance helpful functional proficiency in huge sewer frameworks. For instance, IoT ecological sensors have turned into an important device for demonstrating and checking natural peculiarities [27]. Essentially, water quality can be observed through an organization of remote sensors.

Similarly, Mauro et al. [28] fostered a self-controlled multiparameter sensor gadget that can gage residue type and micrometer thickness. This gadget offered strong, minimal expense frameworks that empowered prescient upkeep, early admonition highlights, and viable water quality administration processes.

Notwithstanding, the AI approach is viewed as a compelling elective strategy for foreseeing complex nonlinear frameworks on the grounds that factual and deterministic methods for assessing water quality information are insufficient because of dynamic ecological variables and boundaries that can influence water quality. As of late, AI has given systems to hardware treatment and determination, execution expectation, functional enhancement, issue analysis, sewage treatment plant (STP) execution assessment, energy cost demonstrating, and decisive reasoning that include a lot of information for direction.

4.1 Artificial Neural Networks (ANN) and Its Applications

The fundamental design of a neural network comprises three principal layers of interconnected fake neurons; the layers are an information layer, a secret layer, and a result layer. A neural network utilizes interconnected neurons, which bear the heaviness of the organization, as handling components [29].

To frame an ANN, the virtual organization passes the information from the dataset to the neurons in the main layer. These neurons, basically called hubs, have two primary capabilities, which are to change the heaviness of the hub by gaining adaptively from the information and changing over it into a drive and communicating this motivation to others hitches. At the point when the entire framework plays out the educational experience, every neuron gets a contribution from the neuron in the past layer, and its result is sent as contribution to the neuron in the following layer of the organization. The exactness of the outcomes produced by the organization in the preparation cycle works on as the organization learns. This method is rehashed until a halting standard is fulfilled.

Ongoing reports have shown a huge expansion in the use of ANN calculation techniques in water quality observing and estimating.

For instance, some authors applied ANN procedures in the improvement of a multi-stage anticipated model for petrol wastewater bioremediation. Moreover, some authors assessed ANN procedures to foresee tide levels, and verifiable site information and water level were utilized as an info model. The outcomes showed that the enhanced review models in view of ANN delivered acceptable execution during the preparation and approval period. The review recommends that the model can possibly anticipate galactic flowing stream drift and water level. Furthermore, by just utilizing water level information, the ANN created a steady stream gauge without depending on location water level records, as the outcome has showed phenomenal execution as stream figures can be made as long as two days ahead of time [30].

Some authors assessed the capability of ANN and SVM to foresee the degree of absolute nitrogen (TN) in wastewater. The data was normalized to 1 and—1 and applied as information and result data. Similarly, MATLAB writing computer programs were used to build the ANN and SVR model, and the overall improvement computation was used to choose the ideal of these two models. The result of the survey shows that both ANN and SVM systems are strong in predicting water quality. The SVM model showed that better precision execution appeared differently in relation to ANN. In any case, ANN had the choice to depict the association between TN center in wastewater and actual science-related conditions and final products to the extent that mindfulness assessments. Accordingly, ANN was seen as more trustworthy and reasonable than SVM for heading and cycle control [31].

5   Basic Findings, Suggestions

Structure of the WQI model is shown beneath in Fig. 2. This incorporates four primary advances depicted beneath.

Stage 1 is the choice of water quality boundaries: At least one water quality boundary is chosen to be remembered for the evaluation.

Stage 2 Generation of boundary lists: Parameter fixations are switched over completely to unitless files.

Stage 3 Assign weighting values to the boundaries: The boundaries are weighted by their significance for the assessment.

Stage 4 Calculation of the water quality list utilizing a collection capability: The sub-files of the singular boundaries are joined utilizing the loads to give a solitary by and large record. A rating scale is by and large used to sort/rank water quality in light of the general record esteem.
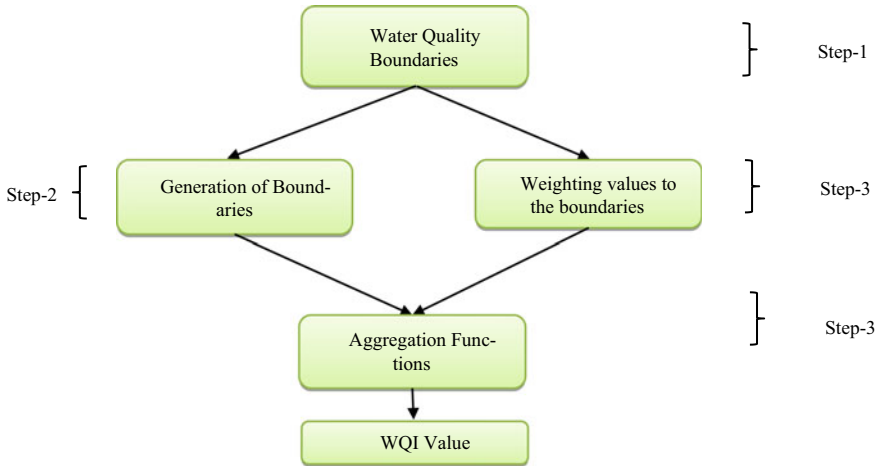
Fig. 2 General flowchart of WQI models [12]

5.1 Calculation for WQI

$$M_i = \frac{p}{D_i}$$

where M is the weightage, D_i is the recommended ith standard, and p is the constant of proportionality

Specific eminence assessment is provided by the given formula

$$Q_i = \frac{100 \times V_i}{D_i}$$

We can clearly understand the maximum tolerable rate and the weightage of the water in Table 1 which describes acidity of water, total dissolved solids, and total hardness to determine the water quality index. Similarly, in Fig. 3, we can see the number of parameters used in the different types of the models like *Horton Index* have 8 types of parameters, *NSF Index* have 11 parameters, *SRDD Index* have 10 parameters, *Dinius Index* have 11 parameters, *ROSS Index* have 4 parameters, *Dojildo Index* have 26 parameters, and *Liou Index* have 13 types of different parameters.

The aggregation capability is the last step of the WQI model. It is applied to add up to the age of the limits into a single water quality file score. Most models have used either added substance capacities or multiplicative abilities or a blend of the two. The different assortment abilities are discussed quickly here.

To WQI the researchers have done various work with various models which can be seen in Table 2, the Horton Index used 8 parameters based on parameters significance and data availability, the parameters value used as sub-index value, and sub-index ranges from 0 to 100 assigned, fixed and unequal system (4 for DO and 1 for other

Table 1 Unit weightage of parameters based on the Indian drinking water standard (IS: 10500, 1993)

Parameters	Uppermost tolerable rate for water (D_i)	Unit weightage (M_i)
Acidity or alkalinity	7.5	0.02808068
TDS	500.0 (0–1200)	0.00042121
Total hardness	300.0	0.00070201
Cl^-	250.0	0.00084242
Na^+	45.0	0.00468011
F^-	1.0	0.21060514
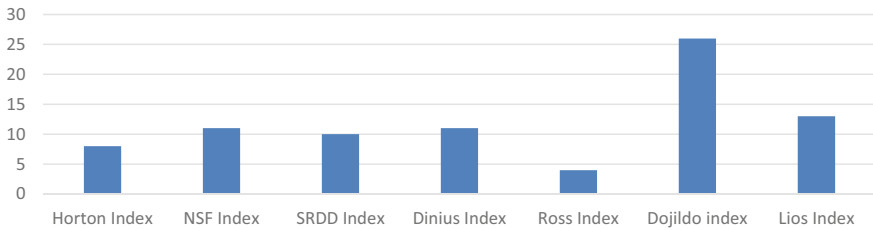
Fig. 3 Number of parameters in different model

parameters) suggested and the aggregation technique is used as used simple additive mathematical functions as given below in Eqs. (1) and (2). A comparative study on some parameter is shown in Table 3.

Similarly, the NSF Index used 11 parameters by using Delphi technique with the quality standard guideline and scale that ranged from 0 to 1; the expert panel judgment, and sum of weight value is equal to 1 given, used two mathematical functions • first one is additive formula (Eq. 2), second one is multiplicative formula (Eq. 3)

$$\text{WQI} = \frac{w_1 s_1 + w_2 s_2 + w_3 s_3 + \cdots + w_n s_n}{w_1 + w_2 + w_3 + \cdots + w_n} m_1 m_2 \tag{1}$$

$$\text{Brown WQI} = \sum_{i=1}^{n} w_i s_i \tag{2}$$

$$\text{WQI} = \prod_{i=1}^{n} s_i^{w_i} \tag{3}$$

SRDD Index showed 10 parameters by using Delphi technique and also used expert opinion, and it ranged from 0 to 100 recommended by SRDD; panel based and sum of weight value equal to 1 recommended by SRDD, additive mathematical function adopted (Eq. 4), multiplicative formula that was used for NSF (Eq. 3).

Table 2 Outline of most normal WQI models

Name of WQI models	No. of parameters	Generation of boundaries	Weighting values	Aggregation techniques
Horton Index (1960)	8	0–100	Fixed and unequal system	Mathematical function
NSF Index (1965)	11	0–1	Master board judgment	Adaptive and multiplicative function
SRDD Index (1970)	10	0–100	Equal to 1	Adaptive and multiplicative function
Dinius Index (1972)	11	Boundaries are directly assigned	Unequal weight	Multiplicative function
Ross Index (1977)	4	–	Equal to 1	Mathematical equation
Dojildo Index (1994)	26	NA	NA	Square root of the harmonic mean
British Index (1995)	Open choice	NA	Unequal weight	Simple mathematical formula
Liou Index (2004)	13	NA	Equal weighting system	Hybrid function

Table 3 Comparative study on some parameters

WQI	Temperature	pH	Alkalinity	Total hardness	Total nitrogen
Horton Index (1960)	–	4	1	–	–
NSF Index (1965)	0.10	0.11	–	–	–
SRDD Index (1970)	0.05	0.09	–	–	0.08
Dinius Index (1972)	–	0.77	0.063	0.065	0.090
Ross Index (1977)	–	0.12	–	–	–
Dojildo Index (1994)	0.07	–	–	–	0.16

$$\text{SRDD - WQI} = \frac{1}{100}\left(\sum_{i=1}^{n} w_i s_i\right)^2 \tag{4}$$

Dinius Index showed 11 parameters by using Delphi technique, parameters value directly assigned as sub-index value also used unequal weight and the sum of weighting value is equal to 10; multiplicative function used (Eq. 3).

Table 4 Status of WQI

WQI (scale of 100)	Status
0–25/100	Brilliant
26–50/100	Respectable
51–75/100	Deprived
> 75s/100	Unfortunate

Ross Index showed 4 general WQ parameters using Delphi method, expert panel judgment-based sub-index system, expert based and sum of weight value is equal to 1 given; used additive mathematical equation (Eq. 1).

Dojildo Index showed 26 parameters by using open (additional group) and close system (basic parameters group), and the aggregation technique is adopted square root of the harmonic mean function (Eq. 5)

$$\text{WQI} = \sqrt{\frac{n}{\sum_{i=1}^{n} \frac{1}{s_i^2}}} \tag{5}$$

British Colombia Index used common monitoring parameters which are open choice system but used at least 10 parameters. Simple specific mathematical formula is used for aggregation.

Lius Index used 13 parameters, and these parameters were selected based on environmental and health significance, and the parameters actual concentration is directly used as sub-index, weighting factors were generated by the using rating curves that were illustrated based on the standard guideline of WQ variables, and Liou WQI model proposed hybrid (additive and multiplicative) functions (Eqs. 1 and 2).

Table 4 shows the status of water quality index (WQI) on the scale of 100 from bad to excellent.

6 Discussion

6.1 Eclipsing Issues in Models

The fact that WQI models cannot solve the eclipsing issue is one of its major flaws. Ott [32] used the word "eclipsing" to explain how the final model output conceals the underlying nature of the water quality [32]. Inappropriate sub-indexing procedures, parameter weightings that do not accurately represent the relative impacts of parameters, or ineffective aggregation methods may all contribute to the eclipsing issue. Consider a WQI model, for instance, where the weight values for the two parameters are equal to 0.5 and the sub-index values are $I1 = 50$ and $I2 = 110$. The ultimate WQI would be 80 if a simple additive aggregation function was used. Even if one

or more of the criteria may not achieve their recommended values, this index value ($I = 80$) may nevertheless indicate good water quality. In this case, the aggregation process obscures or "eclipses" the parameter failure.

Important data on water quality may be lost during the aggregation process, according to several experts (Abbasi and Abbasi 2012) [29]. Eclipsing issues are well explained by Otte (1978) and Smith [33]. WQI models have eclipse problems that have been found by many studies [33–35]. They point out that it is generated as a result of the employment of complex mathematical operations during the aggregate step [33].

Many scholars have worked to reduce eclipse concerns; for instance, the Smith Index suggested utilizing the minimal operator index aggregation function (Abbasi and Abbasi, 2012). The final WQI score is modified if a single determinant's many sub-index values are used to carry out the various aggregation processes [33]. According to Smith [33], the test's dissolved oxygen and fecal coliform determinants are what give the weight values of 0.30 and 0.12 their greatest and lowest ranges, respectively. When these weighting values are applied, they create various WQI scores for the same determinant, which are then used to determine the final WQI score (Fig. 4). The eclipsing issue is the name given to this kind. Smith [33] recommended that the ultimate WQI score be achieved in this situation without exceeding the minimum operator function (Eq. 6).

$$\text{WQI} = \text{Min}(S_i, S_{i+1}, S_{i+2}, \ldots I_{\text{subn}}) \tag{6}$$

The eclipses basic problem is resolved by the minimal operator function, but that much uncertainty is still generated throughout the aggregation process.

when using three different aggregating functions and modifying only one sub-index rating (Isub), with all others remaining fixed at their maximum value, for
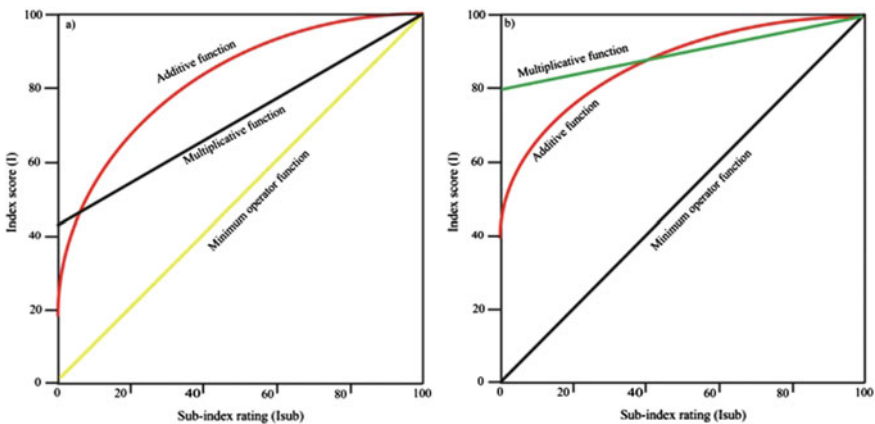
Fig. 4 Change in the final WQI score

(a) changing dissolved oxygen concentration and (b) changing fecal coliform concentration.

6.2 Uncertainty in the Model

The alteration of the parameter threshold's potential impact on the corresponding sub-index and end index values is the main topic of an investigation of index uncertainty. Any model must include uncertainty as a basic component, and it may be tied to certain model parameters. Studies have shown that the origins of the WQI model's final indices of uncertainty are related to each other [36, 37]. The parameter selection, sub-indexing method, and parameter weighting all contributed to the model's uncertainty [36, 38]. Table 8 displays the main causes of eclipses and uncertainty in the WQI system. A significant source of uncertainty has been identified as the aggregation function [33]. Smith demonstrated the functional unpredictability of the WQI modeling aggregation (Fig. 3). Juwana et al. [36] examined the uncertainty and sensitivity of several aggregation functions that employed various weight schemes and discovered that the calculation (arithmetic and geometric) used had the greatest impact on the final index values. Numerous research have been conducted to pinpoint the origins of uncertainties and to measure it. In these works, ambiguity in parameter selection procedures has been eliminated using a variety of statistical techniques, including correlation matrix, main component assessment, clustering algorithms, and discriminant analysis. In order to reduce uncertainty in the selecting and weighting of the parameter, several WQI models employed expert opinion. Juwana et al. [36] estimated the uncertainty and sensitivity of the different aggregation functions using the Monte Carlo simulation approach for the coefficients of variation and correlation. In order to handle the final WQI results confidently and utilize them to manage water resources effectively and keep them in excellent condition, a WQI model's design should include identifying and measuring uncertainty.

6.3 Choosing a Parameter

The number of variables that WQIs utilize to evaluate the quality of the water differs greatly across models. A few models, such the CCME Index and Said Index, simply used four characteristics to evaluate the quality of the water. Given that the quality of the water is dependent on a wide range of natural and man-made elements, it would seem that this is a little number to accurately portray the whole image of the water quality. On the opposite extreme, there are 26 factors needed for the WJ index, and it is doubtful that all of the measured data would be readily accessible. Although the majority of WQI model parameter lists are chosen based on the opinions of expert panels, one must also take the availability and caliber of the monitoring data into account.

While some models allow for user parameter selection, others do not. Users may choose WQ parameters with ease in open systems, such as the CCME model, by providing their own explanation. Users are limited to taking into account the parameters suggested by the model in fixed systems like the NSF, SRDD, Ross, and Bascaron models. Flexible mixed systems like the Dojildo Index (1994) are available. The choice of parameters should be made using expert knowledge and experience. Researchers have analyzed expert panel views using statistical methods. The Delphi procedure is a well-liked method for getting an expert panel's recommendations for the optimum parameter choices, although some research showed that it generated data uncertainty and decreased model accuracy. The goal of the evaluation (such as to evaluate for drinking water quality, bathing water quality, and shellfish production) should be taken into consideration while choosing parameters. The use of statistical methods to assist parameter selection has gained popularity recently. For instance, Ma et al. [39] utilized the Spearman's rank correlation coefficient to assess the interaction between variables and eliminated those that had no discernible link with the others.

A key consideration in the parameter selection procedure is the availability of data. Detailed information on physical, chemical, biological, hazardous, and pesticide factors is required by certain WQI models [40]. Because water quality monitoring systems are expensive and labor-intensive, they provide significant challenges for underdeveloped nations. Recent studies have identified the monitoring program as the primary cause of errors and uncertainty due to improper site selection and planning, inaccurate measurement caused by faulty equipment calibration or sample contamination, poor sampling techniques, inconsistent recording and data transcription, management and storage capability, and insufficient data collection (Department of Water, 2009). Due to a lack of sufficient funding, such as financial assistance, experienced human resources, and efficient water management boards, many developing nations are also unable to build contemporary, sophisticated laboratory facilities [41].

When fresh information or fresh proof of a parameter's significance becomes available, models should be updated. Temperature and total phosphorus were added to the Oregon WQI model throughout updates to make it better (Cude, 2001). Principal component analysis (PAC) has been utilized by several studies to identify parameter relevance [42–46], while cluster analysis has also been employed by others (CA). Fecal coliforms are included in several of the examined WQI models because they have historically been employed as an indicator organism for fecal pollution and microbiological water quality [46]. As a result, international organizations like the World Health Organization [47] and the EU, via the Water Framework Directive (WFD, 2000), advocate the use of E. coli as a more accurate indication of fecal contamination and microbiological water pollution than fecal coliforms. fecal coliforms above E. coli. Therefore, we should be included in recently created WQI models. coli in addition to, or instead of, fecal coliforms, especially if they are being used to evaluate drinking or bathing waters.

6.4 Calculation of the Parameter Sub-index

Many WQI models, including the CCME index, omit the sub-index calculation, despite the fact that it has an impact on the final WQI; however, other WQI models, including the Oregon, British Columbia, House, SRDD, Stoner's, and Smith Index, use the measured parameter values directly as sub-index values. Many of those who compute sub-indexes have developed the sub-index rule for parameters with the advice of specialists, and some of the sub-index producing techniques are rather complicated. Care must be taken while establishing the methods for obtaining the sub-index values to ensure that the resulting values do not obscure the significance or effect of the parameter. A significant drawback of sub-indexing systems, according to Swamee and Tyagi [48], is that they obscure the original understanding of water quality. To create suitable sub-indexing criteria that are more uniformly applied across WQI models, local guideline values for water quality should, if practical, be matched with international guideline values (e.g., WHO and EU Water Framework Directive).

6.5 Weighting Parameters

Another important part of the WQI model is the parameter weighting, which accounts for the relative impact of a water quality measure on the final WQI. However, other models do not use any weightings at all, like the CCME, Smith, and Dojildo models. The most common weightings are unequal because they may differentiate between the effects of various characteristics. Based on the opinions of an expert panel, several models acquired parameter weight values (e.g., the NSF, House, and SRDD models). The environmental relevance of the parameter, the suggested guideline levels, and the applications/uses of the water body have traditionally served as the foundation for the expert panels' weightings. The different weightings used to the same characteristics in different models show how difficult it is to determine the proper weight values and how a parameter's significance varies depending on the assessment's goals. An example is dissolved oxygen, to which several models have assigned weight values ranging from 4, 0.17, 0.18, 0.2, 4, 0.16, 0.10, 8, 0.167, and 0.22. The AHP approach has been used to assess parameter importance, which lowers uncertainty brought on by improper parameter weighting.

6.6 Aggregation Function

Researchers have used a wide variety of aggregation strategies. The most common functions have been simple multiplicative or additive ones. There has been some success with the modified aggregation strategies that several scholars have suggested

to aggregate parameter sub-indexes with reduced uncertainty [49–53]. Fuzzy interface systems and artificial neural networks are also used in computer-based aggregation techniques that have had some success in this area [54–61].

Additionally, various aggregation algorithms provide different index scores for the same water quality data. As a consequence, a variety of differences in the water quality classes were seen. The output score for the WQI model does not match the water quality classes. As a result, it is difficult to determine what the realistic water quality conditions are. These kinds of uncertainty are reflected in the WQI model's aggregation process's shortcomings. It is essential to follow certain rules while creating the optimum WQI model to evaluate actual surface water quality circumstances. After that, a reliable WQI model may be developed to assess the surface water's quality with confidence.

7 Conclusion

This survey presents a top-to-bottom conversation of the uses of ANNs in the hydrological field. Appropriately, considering the picked composing used for this review, we find that different kinds of ANN estimations for assumption and seeing of water quality limits give pleasing results. In addition, this review revealed that it is doable to consolidate neural networks with other existing progressions, for instance, ANN-based survey models, geographic information structure (GIS)-based ANNs, and far-off sensor networks with ANNs.

Although several various versions have been produced to date, WQI models have been frequently employed for water quality evaluation due to their relative simplicity and clearly understandable output. This review was done to look at the mathematical methods and structures employed in WQI models. The research discovered that while the architecture of most models was generally similar, the finer specifics of the four key components differed substantially. The research also emphasized the eclipsing and uncertainty problems brought on by the model building process. The key findings from the review are as follows:

- The four phases of the majority of WQI models are as follows: (1) choosing the water quality parameters, (2) figuring out the parameter sub-indices, (3) figuring out the parameter weightings, and (4) aggregating the sub-indices to calculate the overall water quality index. Model applications are mostly site- or region-specific, despite the fact that the majority of models have been constructed in a general way that makes them readily transferable to other locations. The type of waterbody (river, lake, estuary, or groundwater), its current or intended uses (such as drinking water, industrial use, bathing, and fisheries), local water quality guidelines and assessment protocols, and the availability of data all play a significant role in the selection of parameters, sub-indexing rules, and weightings.
- The number and kind of water quality parameters that have been included into WQI models, the weights assigned to certain parameters, and the standards (such

as guideline levels) used to create sub-index values all vary significantly. As a result, there is relatively little consistency across models, making it difficult to contrast applications to various research fields. WQI models may become more appealing instruments for water quality evaluation if their structure and procedures are simplified, such as by including international guideline values (such as WHO, EU WFD, or comparable). For greater usage, updating models to take into account relevant new factors is also essential; an example of this would be adding E. coli is the preferred sign of fecal contamination and a gauge of the microbiological water quality, as well as a source of toxins, such as nitrogen and phosphorus, which are crucial for eutrophication. When conducting new research, attention must be made to choose the right model, decide if a new or modified model is required, and guarantee that the model is used correctly.

- Two of the major problems that have an impact on the accuracy of model outputs are eclipsing and uncertainty. Here, all four WQI model phases might be helpful. To date, the selection of parameters, creation of sub-indexing methods, and choice of suitable weightings have all significantly depended on expert panel judgments. While this is better than relying just on one person's judgment, it still has the potential to bring uncertainty into to the models. In more recent times, computer-based methods like fuzzy interface systems and artificial neural networks have been utilized to lessen the uncertainty brought on by the final aggregation process, as well as mathematical methods like principal component analysis and cluster analysis to better guide the choice of parameters and their weightings. To increase the assurance around the correctness of the final calculated indices, the usage of these strategies should be sought. For each WQI application, model uncertainty should at the very least be evaluated and quantified.

8   Future Work

Future work in hydrology ought to zero in on uses of other delicate registering advances, for example, profound learning apparatuses, hereditary calculations (HC), arbitrary backwoods, outrageous learning machines, and irregular timberlands. Some researches will propose the stage to reach out to take advantage of the data gathered from the versatile sensor hubs to change the goal inspecting areas adaptively after some time by consolidating natural model variation. Also, the future work suggests applying hybrid machine learning techniques and deep learning methods to forecast WQI by separating them into wet and dry period datasets. We may create a formula or utilize an alternative method to calculate the water quality index (WQI). It is advised that more research be done to include biological and meteorological aspects into the construction of the water quality model as this will guarantee that a larger variety of contamination source is reached.

References

1. Yaroshenko I, Kirsanov D, Marjanovic M, Lieberzeit PA, Korostynska O, Mason A, Frau I, Legin A (2020) Real-time water quality monitoring with chemical sensors. Sensors 20(12):3432
2. Giudicianni C, Herrera M, di Nardo A, Carravetta A, Ramos HM, Adeyeye K (2020) Zero-net energy management for the monitoring and control of dynamically-partitioned smart water systems. J Clean Prod 252:119745
3. EarthEcho Water Challenge (2021) Available online: https://www.monitorwater.org/. Accessed on 19 May 2021
4. Water Crisis|World Water Council (2021) Available online: https://www.worldwatercouncil.org/en/water-crisis. Accessed on 19 May 2021
5. Yadav RK, Jha A, Choudhary A (2021) IoT based prediction of water quality index for farm irrigation. In: 2021 international conference on artificial intelligence and smart systems (ICAIS), IEEE, pp 1443–1448
6. Banna MH, Imran S, Francisque A, Najjaran H, Sadiq R, Rodriguez M, Hoorfar M (2014) Online drinking water quality monitoring: review on available and emerging technologies. Crit Rev Environ Sci Technol 44(12):1370–1421
7. Abba S, Namkusong JW, Lee J-A, Crespo ML (2019) Design and performance evaluation of a low-cost autonomous sensor interface for a smart IoT-based irrigation monitoring and control system. Sensors 19(17):3643
8. Brown RM, McClelland NI, Deininger RA, Tozer RG (1970) A water quality index-do we dare. Water Sewage Works 117(10)
9. Thangaratinam S, Redman CWE (2005) The Delphi technique. Obstet Gynaecol 7(2):120–125
10. Debels P, Figueroa R, Urrutia R, Barra R, Niell X (2005) Evaluation of water quality in the Chillán River (Central Chile) using physicochemical parameters and a modified water quality index. Environ Monit Assess 110(1):301–322
11. Horton RK (1965) An index number system for rating water quality. J Water Pollut Control Fed 37(3):300–306
12. Md Uddin G, Nash S, Olbert AI (2021) A review of water quality index models and their use for assessing surface water quality. Ecolo Ind 122:107218
13. Ahmed U, Mumtaz R, Anwar H, Shah AA, Irfan R, García-Nieto J (2019) Efficient water quality prediction using supervised machine learning. Water 11(11):2210
14. Hassan Md M, Hassan Md M, Akter L, Rahman Md M, Zaman S, Hasib Md K, Jahan N, et al (2021) Efficient prediction of water quality index (WQI) using machine learning algorithms. Human-Centric Intell Syst 1(3–4):86–97
15. Verma R, Ahuja L, Khatri SK (2018) Water quality index using IOT. In: 2018 international conference on inventive research in computing applications (ICIRCA), IEEE, pp. 149–153
16. Vasanthavigar M, Srinivasamoorthy K, Vijayaragavan K, Rajiv Ganthi R, Chidambaram S, Anandhan P, Manivannan R, Vasudevan S (2010) Application of water quality index for ground-water quality assessment: Thirumanimuttar sub-basin, Tamilnadu, India. Environ Monitor Assess 171(1):595–609
17. Reza R, Singh G (2010) Assessment of ground water quality status by using water quality index method in Orissa, India. World Appl Sci J 9(12):1392–1397
18. Sillberg CV, Kullavanijaya P, Chavalparit O (2021) Water quality classification by integration of attribute-realization and support vector machine for the Chao Phraya River. J Ecol Eng 22(9)
19. Bui DT, Khosravi K, Tiefenbacher J, Nguyen H, Kazakis N (2020) Improving prediction of water quality indices using novel hybrid machine-learning algorithms. Sci Total Environ 721:137612
20. Zhang Y, Gao X, Smith K, Inial G, Liu S, Conil LB, Pan B (2019) Integrating water quality and operation into prediction of water production in drinking water treatment plants by genetic algorithm enhanced artificial neural network. Water Res 164:114888
21. Ranjan V, Reddy MV, Irshad M, Joshi N (2020) The Internet of Things (IOT) based smart rain water harvesting system. In: IEEE, 2020 6th international conference on signal processing and communication, ICSC2020, pp 302–305

22. Konde S, Deosarkar DS (2020) IOT based water quality monitoring system. SSRN Electron J 6

23. Pincheira M, Vecchio M, Giaffreda R, Kanhere SS (2020) Exploiting constrained IoT devices in a trustless blockchain based water management system. In: IEEE international conference on blockchain and cryptocurrency, ICBC2020, pp 1–7

24. Gupta S, Kohli M, Kumar R, Bandral S (2021) IoT based under water robot for water quality monitoring. IOP Conf Ser Mater Sci Eng 1033:012013. https://doi.org/10.1088/1757-899x/1033/1/012013

25. Ray A, Goswami S (2020) IoT and cloud computing based smart water metering system. 2020 international conference on power electronics and IoT Applications in renewable energy and its control, PARC 2020, pp 308–313

26. Thombre S, Islam RU, Andersson K, Hossain MS (2016) Performance analysis of an IP based protocol stack for WSNs. In: 2016 IEEE conference on computer communications workshops (INFOCOM WKSHPS). IEEE, pp 360–365

27. Edmondson V, Cerny M, Lim M, Gledson B, Lockley S, Woodward J (2018) A smart sewer asset information model to enable an 'Internet of Things' for operational wastewater management. Autom Constr 91:193–205

28. Mauro MDM, Pani GP, Tizzoni Tizzoni M, Carminati MC, Foschi JF, Mezzera Mezzera L, Zanetto FZ, Antonelli MA, Turolla AT (2020) A self-powered wireless water quality sensing network enabling smart monitoring of biological and chemical stability in supply systems

29. Singh TN, Sinha S, Singh VK (2007) Prediction of thermal conductivity of rock through physico-mechanical properties. Build Environ 42(1):146–155

30. Hidayat H, Hoitink AJF, Sassi MG, Torfs PJJF (2014) Prediction of discharge in a tidal river using artificial neural networks. J Hydrol Eng 19(8):04014006

31. Guo H, Jeong K, Lim J, Jo J, Kim YM, Park J-P, Kim JH, Cho KH (2015) Prediction of effluent concentration in a wastewater treatment plant using machine learning models. J Environ Sci 32:90–101

32. Ott WR (1978) Environmental indices: theory and practice. Arbor Sci Publ, AnnAnn Arbor, Michigan 371

33. Smith DG (1990) A better water quality indexing system for rivers and streams. Water Res. https://doi.org/10.1016/0043-1354(90)90047-A

34. Steinhart CE, Schierow L-J, Sonzogni WC (1982) An environmental quality index for the great lakes. JAWRA J Am Water Resour Assoc. https://doi.org/10.1111/j.1752-1688.1982.tb00110.x

35. Sutadian AD, Muttil N, Yilmaz AG, Perera BJC (2016) Development of river water quality indices—a review. Environ Monit Assess 188:1–29. https://doi.org/10.1007/s10661-015-5050-0

36. Juwana I, Muttil N, Perera BJC (2016) Uncertainty and sensitivity analysis of west Java water sustainability index - a case study on citarum catchment in Indonesia. Ecol Indic 61:170–178. https://doi.org/10.1016/j.ecolind.2015.08.034

37. Seifi A, Dehghani M, Singh VP (2020) Uncertainty analysis of water quality index (WQI) for groundwater quality evaluation: application of Monte-Carlo method for weight allocation. Ecol Indic 117. https://doi.org/10.1016/j.ecolind.2020.106653

38. Sutadian AD, Muttil N, Yilmaz AG, Perera BJC (2016) Development of river water quality indices—a review. Environ Monit Assess 188:1–29. https://doi.org/10.1007/s10661-015-5050-0

39. Ma Z, Li H, Ye Z, Wen J, Hu Y, Liu Y (2020) Application of modified water quality index (WQI) in the assessment of coastal water quality in main aquaculture areas of Dalian, China. Mar Pollut Bull 157. https://doi.org/10.1016/j.marpolbul.2020.111285

40. Ongley ED, Booty WG (1999) Pollution remediation planning in developing countries. Water Int 24:31–38. https://doi.org/10.1080/02508069908692131

41. Debels P, Figueroa R, Urrutia R, Barra R, Niell X (2005) Evaluation of water quality in the Chillán River (Central Chile) using physicochemical parameters and a modified water quality index. Environ Monit Assess (2005). https://doi.org/10.1007/s10661-005-8064-1

42. Abrahão R, Carvalho M, Da Silva WR, Machado TTV, Gadelha CLM, Hernandez MIM (2007) Use of index analysis to evaluate the water quality of a stream receiving industrial effluents Water. SA

43. Gazzaz NMNM, Yusoff MKMK, Aris AZAZ, Juahir H, Ramli MFMF (2012) Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. Mar Pollut Bull 64:2409–2420. https://doi.org/10.1016/j.marpolbul.2012.08.005

44. Sun W, Xia C, Xu M, Guo J, Sun G (2016) Application of modified water quality indices as indicators to assess the spatial and temporal trends of water quality in the Dongjiang River. Ecol Indic 66. https://doi.org/10.1016/j.ecolind.2016.01.054

45. Wu Z, Wang X, Chen Y, Cai Y, Deng J (2018) Assessing river water quality using water quality index in Lake Taihu Basin, China. Sci Total Environ. https://doi.org/10.1016/j.scitotenv.2017.08.293

46. Odonkor ST, Ampofo JK (2013) Escherichia coli as an indicator of bacteriological quality of water: an overview. Microbiol Res (Pavia). https://doi.org/10.4081/mr.2013.e2

47. Ashbolt NJ, Grabow WOK, Snozzi M (2001) Indicators of microbial water quality. In: Fewtrell L, Bartram J (eds) World Health Organization (WHO). Water quality: guidelines, standards and health. IWA Publishing, London, UK. ISBN: 1 900222 28 0

48. Swamee PK, Tyagi A (2000) Describing water quality with aggregate index. J Environ Eng. https://doi.org/10.1061/(ASCE)0733-9372(2000)126:5(451)

49. Hurley T, Sadiq R, Mazumder A (2012) Adaptation and evaluation of the Canadian council of ministers of the environment water quality index (CCME WQI) for use as an effective tool to characterize drinking source water quality. Water Res 46:3544–3552. https://doi.org/10.1016/j.watres.2012.03.061

50. Şener Ş, Şener E, Davraz A (2017) Evaluation of water quality using water quality index (WQI) method and GIS in Aksu River (SW-Turkey). Sci Total Environ 584–585:131–144. https://doi.org/10.1016/j.scitotenv.2017.01.102

51. Wu Z, Wang X, Chen Y, Cai Y, Deng J (2018) Assessing river water quality using water quality index in Lake Taihu Basin, China. Sci Total Environ. https://doi.org/10.1016/j.scitotenv.2017.08.293

52. Hallock D (2002) a water quality index for ecology's stream monitoring program. Environ Assess Progr. https://doi.org/10.1021/jp905202x

53. Khan AA, Paterson R, Khan H (2004) Modification and application of the Canadian council of ministers of the environment water quality index (CCME WQI) for the communication of drinking water quality data in Newfoundland and Labrador. Water Qual Res J Canada. https://doi.org/10.1007/s10661-005-9092-6

54. Kloss FR, Gassner R (2006) Bone and aging: effects on the maxillofacial skeleton. Exp Gerontol 41:123–129. https://doi.org/10.1016/j.eswa.2012.02.141

55. Lermontov A, Yokoyama L, Lermontov M, Machado MAS (2009) River quality analysis using fuzzy water quality index: Ribeira do Iguape river watershed. Brazil Ecol Indic 9:1188–1197. https://doi.org/10.1016/j.ecolind.2009.02.006

56. Li R, Zou Z, An Y (2016) Water quality assessment in Qu River based on fuzzy water pollution index method. J Environ Sci (China) 50:87–92. https://doi.org/10.1016/j.jes.2016.03.030

57. Mahapatra SS, Nanda SK, Panigrahy BK (2011) A cascaded fuzzy inference system for Indian river water quality prediction. Adv Eng Softw. https://doi.org/10.1016/j.advengsoft.2011.05.018

58. Nikoo MR, Kerachian R, Malakpour-Estalaki S, Bashi-Azghadi SN, Azimi-Ghadikolaee MM (2011) A probabilistic water quality index for river water quality assessment: a case study. Environ Monit Assess. https://doi.org/10.1007/s10661-010-1842-4

59. Ocampo-Duque W, Ferré-Huguet N, Domingo JL, Schuhmacher M (2006) Assessing water quality in rivers with fuzzy inference systems: a case study. Environ Int. https://doi.org/10.1016/j.envint.2006.03.009

60. Ocampo-Duque W, Osorio C, Piamba C, Schuhmacher M, Domingo JL (2013) Water quality analysis in rivers with non-parametric probability distributions and fuzzy inference systems:

application to the Cauca River, Colombia. Environ Int. https://doi.org/10.1016/j.envint.2012.11.007

61. Peche R, Rodríguez E (2012) Development of environmental quality indexes based on fuzzy logic. a case study. Ecol Indic 23:555–565. https://doi.org/10.1016/j.ecolind.2012.04.029
62. Feiyuan YX, Encheng (2018) A biological sensor system using computer vision for water quality monitoring. IEEE Access 37:151–161
63. Zhang K, Dong Z (2019) Microwave sensing of water quality. IEEE Access 10:724–732
64. Chowdurya MSU, Emranb TB, Ghosha S, Pathaka A, Manjur Alama M, Absara N, Anderssonc K, Hossaind MS (2019) IoT based realtime river water quality monitoring system, vol19. Elsevier, pp 1483–1493
65. Koditala NK, Pandey PS (2018) Water quality monitoring system using IoT and machine learning. IEEE Trans 64:1172–1173
66. Olatinwo SO, Joubert TH (2019) Energy efficient solutions in wireless sensor systems for water quality monitoring: are view. IEEE Sens J 19:259–271
67. Olatinwo SO, Joubert TH (2019) Enabling communication networks for water quality monitoring applications: a survey. IEEE Access 34:2562–2571
68. Pappu S, Niharika PV, Soundarya Pappu AV, Vudathaand P, Niharika AV (2018) Intelligent IoT based water quality monitoring system. IEEE J 12:5447–5454
69. Shakhari S, Banerjee I (2019) A multi-class classification system for continuous water quality monitoring. Elsevier 36:1172–1181
70. Khan Y, See CS (2017) Predicting and analyzing water quality using machine learning: a comprehensive model. Conference 20:944–952

Data Integration in IoT Using Edge Gateway

AjayVeer Chouhan, Atul Sharma, and Chander Diwaker

Abstract The Internet of Things (IoT) is useful for collecting data from various edge computing scenarios, where each subnet can communicate with other subnets with the help of gateways. This subnet is the wireless sensor network (WSN), i.e. a widely used network for data collection from sensors. Sensors accumulate the data from different geographical regions and send those observations to the base station. The base station is a place having a server node that can fetch and issue instructions to other member nodes for data collection according to the requirements. Normally, the data obtained from sensors is unstructured and not integrated. Therefore, it becomes a tedious task to aggregate or integrate this kind of data and to offer secure data forwarding. Here, a novel mechanism for data integration and secure data transmission between sensor nodes and the base station through edge gateways is proposed. In this mechanism, the Message Queuing Telemetry Transport (MQTT) protocol is implemented to send data from sensors to the edge gateways. The MQTT protocol handles data obtained from multiple sensors and edge servers or gateways.

Keywords Communication · Edge computing · Data aggregation · Data integration · Gateways · Security

1 Introduction

The Internet of Things (IoT), edge computing (EC), and 5G play a vital role in data transmission for a distributed data environment, because handling large amounts of information received from multiple locations in unstructured form is challenging [1]. To resolve this kind of hazard, IoT, EC, etc. can be used, especially when data is stored at local servers instead of global servers [2]. Initially, while using EC in a distributed

A. Chouhan (✉) · A. Sharma · C. Diwaker
Department of CSE, UIET, Kurukshetra University, Kurukshetra, India
e-mail: ajay.chouhan15@gmail.com

A. Sharma
e-mail: asuiet@kuk.ac.in

environment, edge servers try to store data locally to the devices that access that data instead of long-distance locations. It helps the system to reduce latency time and increase the network's throughput. EC handles the continuous growth of large amounts of data with IoT cloud storage [3]. Huge amounts of data reach edge servers from different sources. This unstructured data also has duplicate data, but edge gateways have to send only the relevant data to the IoT cloud so that the bandwidth of the network can be saved. It also helps to reduce the end-to-end delay, but security of data is still an issue as it is transmitted with the help of local edge servers towards the cloud. This is because the information received from various sources towards local edge servers may not be secure, and it is also forwarded to global edge servers. This kind of data transmission increases the chances of security breakage in data transmission in a distributed environment [4].

The fog IoT devices have a larger gap than mist and dew, but it is still chosen because it has less delay and more processing power than mist and dew [5]. The WAN is the core and includes traditional network components like bridges, routers, etc. It creates a hierarchy of networks with dew, mist, and fog on the lower levels [6]. To perform secure data transmission in a distributed environment, in this paper a mechanism is proposed that uses MQTT protocol for secure data transmission in edge gateway [7]. Here, EC-based multi-edge computing for secure data transmission is used to increase throughput and reduce end-to-end delay [8].

Figure 1 depicts the architecture of edge computing. In the first step, the data is collected and processed by the IoT network locally using edge devices. After data is processed, it is sent to cloud servers so that it can be used for as many computations and memory storage as possible. The paper is organised into four sections. In Sect. 1, introduction, motivation of work, and system model have been presented. In Sect. 2, proposed scheme of data integration in wireless edge computing (DIWSN) mechanism has been presented. Section 3 presents results, and Sect. 4 presents conclusion of work [9].

1.1 Motivation

The following questions related to reliable service for real-time IoT applications are behind this work.

- What if an IoT device is unable to connect to a cloud server because of a network or cloud fault?
- What happens if a fault occurs at any edge level, whether it be an extreme edge or a middle edge?
- In a dynamic environment, how may edge resources be safely assigned to IoT applications?

The main focus of this paper is to handle and overcome the abovementioned issues by proposing a mechanism based on data integration in wireless edge computing.
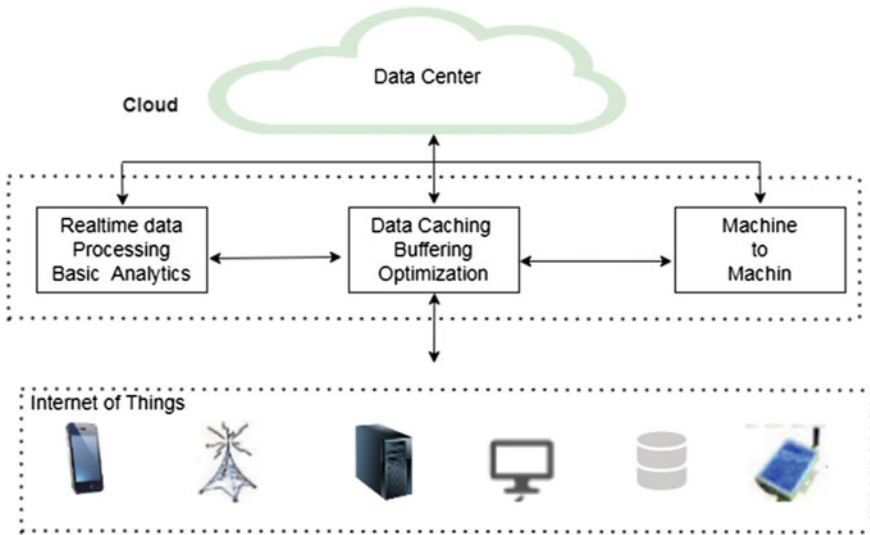
Fig. 1 Edge computing architecture

We provide a system model that uses linear programming as a problem-solving technique.

1.2 System Model

In this section, a mathematical model is designed for multi-edge computing, working on observing area Δ, with sensors [10]. Let $U = \{u_1, u_2, u_3, ..., u_m\}$ be the set of nodes. Suppose R be the set of sensors in the structure, with $R = \{r_1, r_2, r_3, ..., r_M\}$ where $M (M \geq 1)$ is the number of WSN sensors. For the moment, when every node is not arranged, for any node u_i and any sensor r_α ($r_\alpha \in R$, $1 \leq \alpha \leq M$), the node related to sensor function $h_{i\alpha} (u_i, r_\alpha)$ is as

$$h_{i\alpha}(u_i, r_\alpha) = \{1, \quad \text{if } u_i \mapsto r_\alpha, 4\,0, \quad \text{otherwise},\tag{1}$$

where $u_i \mapsto r_\alpha$ represents that node u_i is connected to sensor r_α.

Suppose, DA be the data collection task, with $DA = \{ta_1, ta_2, ta_3, ..., ta_n\}$ and task weight vector $W(DA) = \{w_1^{ta}, w_2^{ta}, w_3^{ta}, ..., w_n^{ta}\}$. The task weights meet restrictions with Eq. (2) as

$$\sum w_j^{ta} = w_1^{ta} + w_2^{ta} + w_3^{ta} + \cdots + w_n^{ta} = 1.\tag{2}$$

Every task $ta_j (ta_j \in DA, \; j \in n)$ happened in a proper space $(ta_j)(\delta_j \subset \Delta)$. Each task area is described with Eq. (3) as

$$\delta_i \cap \delta_j = \phi (1 \leq I \neq j \leq n), \tag{3}$$

$$\sum_{i \in n} \delta_i \leq \Delta, \delta_i \tag{4}$$

As different type of sensors has distinct weights for the task ta_j, so weight vector $w(ta_j, R)$ for ta_j and sensor set R is defined as:

$$w(ta_j, R) = (w_1^R(ta_j), w_2^R(ta_j), \ldots, w_N^R(ta_j)), \tag{5}$$

$$\sum_{\alpha \in N} w_\alpha^R(ta_j) = 1, \tag{6}$$

where $w_\alpha^R(ta_j)$ represents the weight value for sensor r_α for the task ta_j. Also, the sensor weight for task matrix W is represented by (7) as

$$W = \left(w_{11} \cdots w_{1N} \vdots \ddots \vdots w_{n1} \cdots w_{nN} \right)_{n \times N} \tag{7}$$

Here, the row represents tasks, and the column represents sensors.

The initial class is U, and next is DA; the matrix has one row for each element of Y and one column for each element of Y [11]. The incidence function $\beta_{ij}(u_i, ta_j)$ of task $ta_j (j \in n)$ and node $u_i (i \in m)$ can be expressed as

$$\beta_{ij} = \{1, if \; X_i \in \delta_j, 0, \quad if \; X_i \notin \delta_j,$$

The incidence matrix (B) connecting task and node is given as:

$$B = \left(\beta_{11} \cdots \beta_{1n} \vdots \ddots \vdots \beta_{m1} \cdots \beta_{mn} \right)_{m \times n} \tag{8}$$

Some definitions depicting the analysis are as follows:

Definition 1 Data collection time (DCT) is from the time point where node senses the data to the time point when the data server accepted all the data in multi-sensory WSN [12]. So, assume t_g for each node contains node sensing time (t_s), data transmission time (t_c), and stand by time in line (t_w) with data server, as shown in Eq. (9) 4.

$$t_g = t_s + t_c + t_w \tag{9}$$

where the DCT for $T_g(ta_j, u_i)$ is defined as:

$$T_g(\text{ta}_j, u_i) = \sum h_{i\alpha} t_s(u_i, s_\alpha, \text{ta}_j) + k(u_i)\frac{D(\text{ta}_j, u_i)}{U_c} + t_w(\text{ta}_j, u_i) \qquad (10)$$

Here, $t_s(u_i, s_\alpha, \text{ta}_j)$ is the αth sensing time of node u_i for the task ta_j, $D(\text{ta}_j, u_i)$ is the data volume of WSN node u_i for the task ta_j, U_c is transmission speed, and $t_w(\text{ta}_j, u_i)$ is the stand by time during the wireless transmission. The DCT of the task ta_j is found by the maximum $t_g(\text{ta}_j, u_i)$. Let $t_g(\text{ta}_j)$ be the data gathering time for task ta_j and is defined by Eq. (11) as

$$T_g(\text{ta}_j) = (T_g(\text{ta}_j, U)). \qquad (11)$$

In the meantime, the energy consumption of this plan is given by Eq. (12) as

$$E = \sigma t_s + \xi t_c + \tau t_w, \qquad (12)$$

where σ, ξ, and τ are the power consumed, transmission, and stand by time for the task, respectively.

Definition 2 Quality of data (QoD) index is the proportion of the relevant data to the sum of composed data, for each job. Thus, QoD index θ_j of the job ta_j is defined by Eq. (13) as:

$$\theta_j = \frac{D_u(\text{ta}_j)}{D_g(\text{ta}_j)} \qquad (13)$$

where $D_u(\text{ta}_j)$ is the valid data of job and $D_g(\text{ta}_j)$ is the sum of composed data of edge computing for the job ta_j. Thus, the total QoD index of the multitask can be formulated by (14) as:

$$Q(\text{TA}) = \sum_{j=1}^{m} \theta_j \qquad (14)$$

Here, task-driven model is sensitive, and energy utilisation is related to the DCT, by including sensing time, waiting time, and transmission time. It means more DCT leads to higher energy utilisation. Thus, we adjust DCT accordingly to limit the energy utilisation. The issue of maximum QoD of ta_j can be formulated by Max $(Q(DA))$ subject to $t_g(\text{ta}_j) \leq T_{\text{limit}}(\text{ta}_j)$,

$$S_i \cap S_j = \emptyset(1 \leq i \neq j \leq m), \qquad (15)$$

where $\sum_{i \in m} S_i < \Delta$

In Eq. (15), the data gathering time must meet the restriction of being less than the task assigned time frame $(T_{\text{limit}}(ta_j))$.

2   Proposed Scheme of Data Integration in Wireless Edge Computing (DIWSN)

This work provides a self-arranging edge framework to work on the standards of Software-Defined Networks (SDN). IoT is given to the controller so that the given framework can sort information using all of the current assets.

Every time the calculation model receiving the request cannot get solved within a time limit because of battery discharge, internet problem management, etc. So, problem-solving related to latency is a risk for these types of models. So, EC is used for time minimisation and to increase authenticity as well. Here, two types of junctions get contemplated; the first junction is for IoT, where a customer is finding the servers. And the second junction is called resources to work as a server [13].

The above-defined system can easily be understood by Fig. 2. In this diagram, Open Flow (OF) switches join the 1stjunction, i.e. IoT, and the 2ndjunction, i.e. resources. It is clear from this figure how OF switches work. Furthermore, Fig. 3 shows the step-by-step evaluation as a flow diagram of the system, i.e. how every step is followed in sequence.

The main aim is to form a framework that minimises the time utilisation of the process. So, to complete this aim, the given system passes the learning stage. During learning, all the data is collected from present resources and their potential. Using this data, the EC design of OF switches to pass the IoT request to the resources in a round-robin (RR) manner. EC finds out the extra time taken to complete the task between the 1st and 2nd junctions in this system. Due to this characteristic, EC is said to be self-arranging. Before arranging the requests, EC should know the free resources for calculation. Nodes can be used to consolidate or disperse perspectives to find resources.

Message Queuing Telemetry Transport (MQTT) is a lightweight utilisation-layer messaging protocol relying on the subscribe/publish/subscribe (sub/pub) model. IoT applications rely on the request-response model; therefore, it does not accomplish the requirements of IOT applications [14]. In the sub/pub model, multiple clients (sensors) can attach to a central server named "broker" and subscribe to a subject according to interest. Clients can also send messages to specific topics of interest via the broker. A broker is a common way for sensor devices to attach to and exchange information. MQTT utilises TCP link on the transport layer for links between brokers and sensors, which provides reliable communication.

Messages are always put out on subjects, which presents the end address for that message. A client may publish as well as subscribe to multiple subjects. Every client subscribed to a subject receives all the messages published for that topic. Subjects should follow a hierarchy utilising a slash (/) as a separator. This allows for logical arrangement and grouping for edge computing.

In the proposed mechanism, sensor nodes first send basic information about themselves to their base station, such as their location and messages that they have to
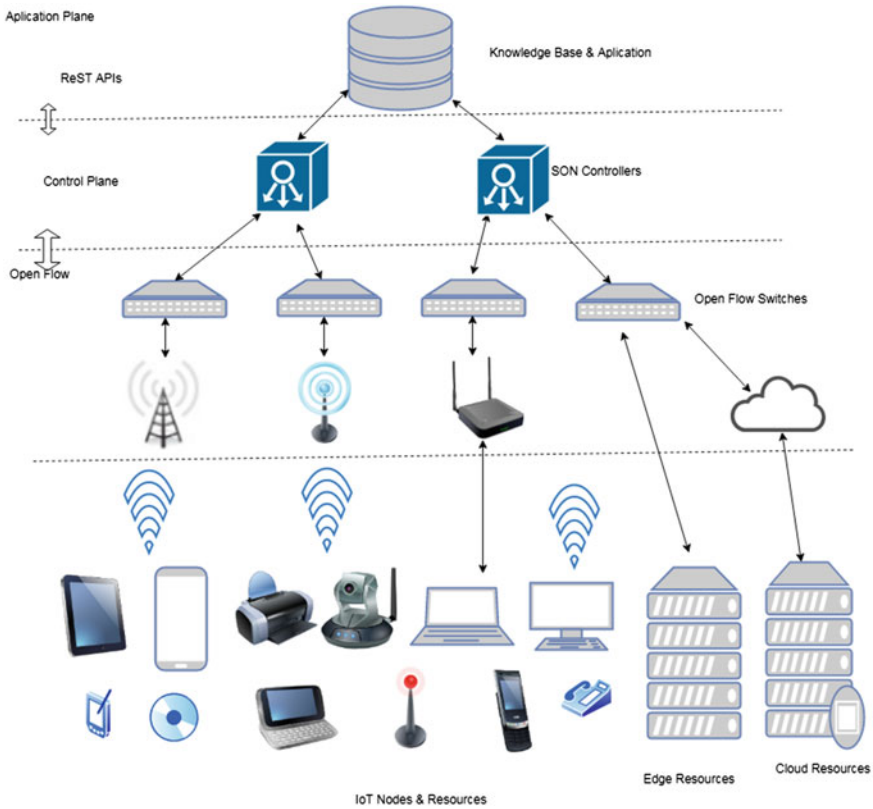
Fig. 2 Proposed framework

transmit and messages that they have received or delivered, and then the MQTT protocol is used to manage data aggregation at the multi-sensor level, and data collected or filtered from MQTT servers is forwarded to the edge server to further fulfil the necessary requirements of sensors [15]. The role of MQTT is to handle requests from multiple sensors with the help of an MQTT broker. Instead of the request and response mechanism used in TCP, MQTT uses a subscribe and publish mechanism for data transmission. The role of subscribe and publish is to manage the large amounts of data used in IoT. After that, the edge server calculates the jobs of sensor nodes and distributes them according to the load of sensor nodes and the information that sender nodes submit to the edge server. The data transmission takes place with the help of cryptography techniques such as public key and private key in which the sender node sends data by encrypting with its private key and, on the other hand, the receiver edge server node decrypts the data with the help of its private key. The edge completes data collection and communication processes after the edge server has completed its task. The sensor nodes directly communicate with edge servers
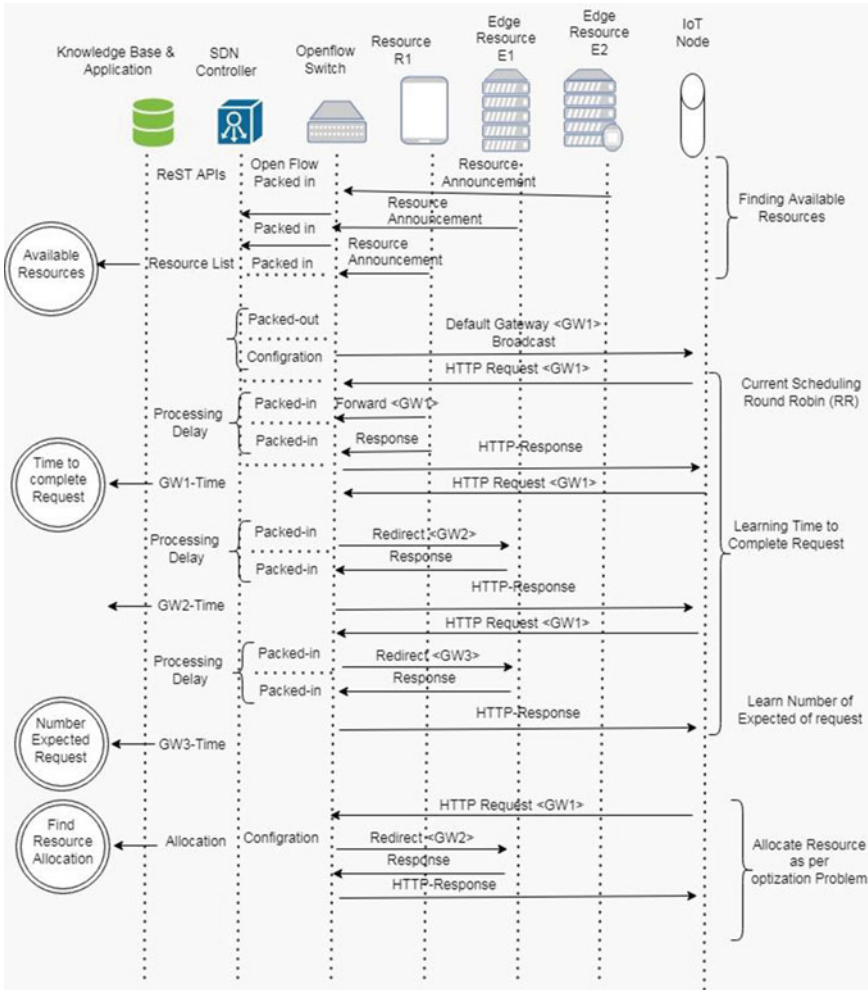
Fig. 3 Flow diagram of proposed mechanism

with the help of wired and wireless media. Consequently, the edge server computes overall results based on global information coming from the MQTT broker.

3 Results and Analysis

The proposed mechanism has been implemented using Python. The significance of the proposed work and assumptions of implementations mainly focus on secure data transmission between nodes with the help of publisher and subscriber nodes used
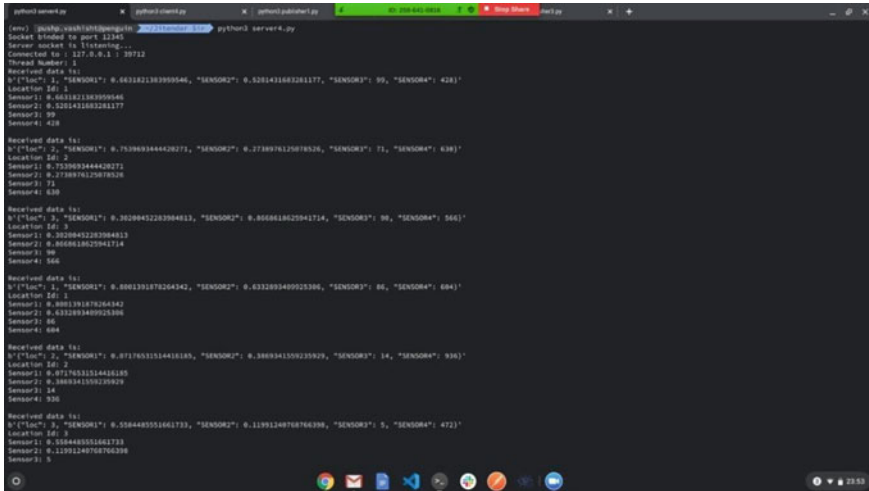
Fig. 4 Base station/server receiving aggregated and encrypted data using socket

in the MQTT protocol. Python is an open-source programming language. In the proposed framework, the MQTT publisher nodes collect data from various sources and provide this data to the MQTT subscriber node or edge gateway node for further data aggregation and processing. At this level, between subscriber and publisher, the MQTT protocol has been used for data transmission. When all the data is submitted to the edge gateway, then the data aggregation process will be started. At this level, data transmission takes place with the help of the socket protocol. At the edge gateway level, data that is aggregated by the edge gateway is transferred to the server or base station by encrypting the data at the edge gateway end, and the server receives that encrypted data and decrypts it with a private key to access that data. In this manner, secure information is transmitted between the edge gateway and the base station. The results show secure data transmission between edge gateways and base stations. Here, data generated by one mode is protected by a public key, and data is decrypted by the destination node using a private key. The key combination is generated with the help of hash functions. Figure 4 depicts the secure data transmission using the MQTT protocol.

4 Conclusion

Data aggregation in edge computing is a very challenging task. Data collected by various sensors is unstructured and redundant. To handle this kind of issue in this paper, a mechanism has been presented. Data aggregation has been resolved with the help of edge gateways using the MQTT protocol, and data is transmitted securely with the help of the socket protocol. In the MQTT protocol, multiple sensors can

publish and subscribe to their requirements to the MQTT broker. The data comes from the MQTT broker, which comes from various multi-sensors and is transmitted to edge servers. Edge gateways are responsible for collecting data, aggregating the received data, and forwarding it to base stations. The results show that information is successfully transmitted from the edge gateway to the base station in a secure manner with end-to-end encryption. Python is used to implement the proposed mechanism. The research work can be extended to the proposed work with some real-life applications. Some limitations of the purpose work include data management between IoWT and edge computing, as well as the role of social networks in IoWT. In future, we will try to propose a framework for interaction between entities in the LSE environment.

References

1. Flauzac O, Javier Gonzalez Santamaria C, Nolot F, Woungang I (2020) An SDN approach to route massive data flows of edge computings. Int J Commun Syst 33(7):e4309
2. Meslin A, Rodriguez N, Endler M (2020) Scalable mobile sensing for smart cities: the musanet experience. IEEE Internet Things J 7(6):5202–5209
3. Rafique W, Qi L, Yaqoob I, Imran M, Rasool RU, Dou W (2020) Complementing IoT services through software defined networking and edge computing: a comprehensive survey. IEEE Commun Surv Tutor 22(3):1761–1804
4. Verderame L, Merelli I, Morganti L, Corni E, Cesini D, D'Agostino D, Merlo A (2020) A secure cloud-edges computing architecture for metagenomics analysis. Futur Gener Comput Syst 111:919–930
5. Shi W, Cao J, Zhang Q, Li Y, Xu L (2016) EC: vision and challenges. IEEE Internet Things J 3(5):637–646
6. Grover J, Garimella RM (2018) Reliable and fault-tolerant IoT-edge architecture. In: 2018 IEEE sensors. IEEE, pp 1–4
7. Rani S, Saini P (2020) Fog computing: applications and secure data aggregation. In: Handbook of computer networks and cyber security. Springer, Cham, pp 475–492
8. Vatankhah Barenji A, Li Z, Wang WM, Huang GQ, Guerra-Zubiaga DA (2020) Blockchain-based ubiquitous manufacturing: a secure and reliable cyber-physical system. Int J Prod Res 58(7):2200–2221
9. Rahmani AM, Gia TN, Negash B, Anzanpour A, Azimi I, Jiang M, Liljeberg P (2018) Exploiting smart e-Health gateways at the edge of healthcare Internet-of-Things: a fog computing approach. Futur Gener Comput Syst 78:641–658
10. Grover J, Garimella RM (2019) Concurrency and synchronization in structured cyber physical systems. In: Cyber-physical systems: architecture, security and application. Springer, Cham, pp 73–99
11. Buyya R, Yeo CS, Venugopal S, Broberg J, Brandic I (2009) Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility. Futur Gener Comput Syst 25(6):599–616
12. Satyanarayanan M, Bahl P, Caceres R, Davies N (2009) The case for vm-based cloudlets in mobile computing. IEEE Pervasive Comput 8(4):14–23
13. Mahmud R, Kotagiri R, Buyya R (2018) Fog computing: a taxonomy, survey and future directions. In: Internet of everything. Springer, Singapore, pp 103–130
14. Mach P, Becvar Z (2017) Mobile edge computing: a survey on architecture and computation offloading. IEEE Commun Surv Tutorials 19(3):1628–1656
15. Satyanarayanan M, Simoens P, Xiao Y, Pillai P, Chen Z, Ha K, Hu W, Amos B (2015) Edge analytics in the internet of things. IEEE Pervasive Comput 14(2):24–31

16. Bonomi F, Milito R, Zhu J, Addepalli S (2012) Fog computing and its role in the internet of things. In: Proceedings of the first edition of the MCC workshop on Mobile cloud computing, pp 13–16
17. Satyanarayanan M (2017) The emergence of edge computing. Computer 50(1):30–39
18. Sardellitti S, Scutari G, Barbarossa S (2015) Joint optimization of radio and computational resources for multicell mobile-edge computing. IEEE Trans Sig Inf Proc Over Netw 1(2):89–103
19. Khan AM, Freitag F (2017) On participatory service provision at the network edge with community home gateways. Procedia Comput Sci 109:311–318

Impact of Node Radiation on Human Health Using IoMT

Hemanta Kumar Bhuyan, Lokaiah Pullagura, Biswajit Brahma, and Chavva Subba Reddy

Abstract This paper addresses the impact of radiation energy for data transmission on tissue using wireless body area network (WBAN). The WBAN is concerned not only with energy savings but also with the potential undesirable influence of node radiation on patient body. The specific absorption rate (SAR) is used to quantify the dose of radiation where an individual is exposed. Most current WBAN data routing algorithms do not account for radiation levels when choosing data paths. In this paper, we introduce a fuzzy-based SAR routing (FSR) algorithm for WBAN. The FSR protocol is a clustering-based method designed to lessen the harmful effects of radiation on humans. Cluster nodes delivered data packets to the cluster head (CH) instead of the hub. We have considered data transmission between nodes that reduced signal strength. But a weaker signal reduces radiation where large datasets produced and absorbed radiation, which neighboring tissues felt more strongly. The suggested approach considered both the signal's transmission strength and the node's susceptibility to RF radiation while choosing CH, with the goal of minimizing radiation. We considered fuzzy logic that selects CH. The chosen CH node had little transmission power and is positioned where RF radiation is absorbed least. The proposed technique performed less computing, has greater energy efficiency, and lower SAR than competing existing protocol.

Keywords Wireless body area network · SAR · Fuzzy logic · Routing protocol · Clustering · Throughput

H. K. Bhuyan (✉) · C. S. Reddy
Department of Information Technology, Vignan's Foundation for Science, Technology and Research (Deemed to Be University), Guntur, Andhra Pradesh, India
e-mail: hmb.bhuyan@gmail.com

L. Pullagura
Jain University, Bengaluru, Karnataka, India

B. Brahma
Mckesson Corporation, San Francisco, CA, USA

549

1 Introduction

Recent nano-meter-scale manufacturing technology allows for the production of tiny low radiation energy. One can create a tiny biosensor node [1] by integrating these devices on board and using a transducer tailored to a particular physiological parameter as shown in Fig. 1. When coupled with the appropriate transducer, a biosensor node may take readings of a wide range of vital parameters, like heartbeat rate, respiration rate, quantity of blood oxygen level, and beyond. A biosensor node is able to sense the values of physiological parameters and relay that information within a certain range of communication [2]. Because of their small size, biosensor nodes can be easily integrated into a variety of wearable forms, including textile fiber, clothing, and wrist bands [3]. Biosensor nodes can be placed in many locations on the patient's body to continuously record vital signs. Different biosensor nodes are identified through the human body in Fig. 1 to create a WBAN. These nodes use radio frequency signals to relay sensory information to a central node [4]. The patient's physiological data is transmitted from the hub node through Wi-Fi to an online cloud service. Therefore, a WBAN system offers a reliable environment for integrating IoT and mobile technology into healthcare delivery via telemedicine [4].

Typically, nodes in a WBAN are less massive than those in other networks. As a result, the WBAN nodes store a relatively negligible amount of energy. The Printed Circuit Board (PCB) of the ECO body motion sensor, for instance, is only 20×10 mm^2 in size [5]. In spite of the fact that the nodes in a WBAN network have limited resources, they devote over two-thirds of their available power to data transmission [1]. Based on the IEEE 802.15.6 standard, WBAN cannot tolerate any more than a 10% packet loss in transmission [6]. In WBAN, nodes send and receive data signals via intra- and extra-corporeal pathways. When electromagnetic radiation reaches the
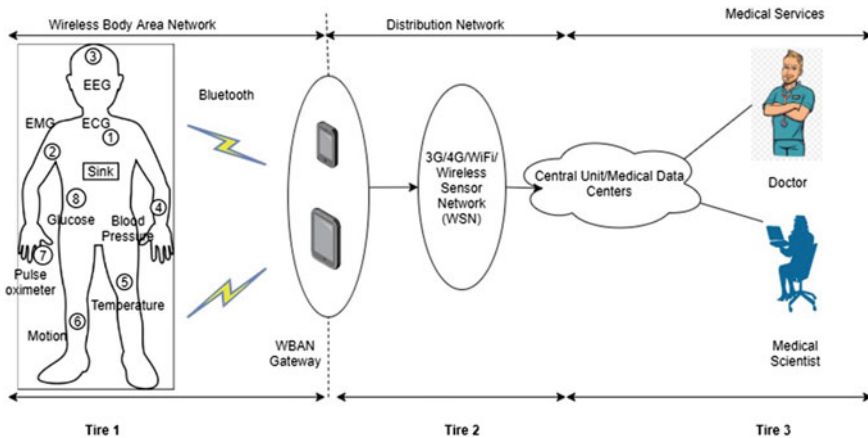
Fig. 1 Model of WBAN

human body, it generates heat, which is harmful to the human body. Possible negative consequences of radiation effect on humans' body [7–9].

The SAR of a node increases with both the strength of its signal and the quantity of data being sent [8, 10]. The maximum safe specific absorption rate (SAR) for a human body is 1.59 W/Kg, as set by the Federal Communications Commission (FCC). However, the skin, the head, and the trunk are more susceptible to radiation damage because of their high-water content [9]. To this end, a WBAN must employ a low-power, high-quality-of-service data routing strategy that priorities low-spectrum-access-requirement data pathways. The data collected by the sensor nodes is sent directly to the hub via a protocol called single hop data routing, which results in minimal transmission delays.

In multi-hop protocols, relay nodes in between sensor nodes collect data from the sensors and send it back to the hub. The SAR impact is mitigated by low-power trans-mission, and the skin near CH still perceives significant quantities of EM radiation. While all of the child nodes go dormant, the CH node stays online and gets all of the data. It gathers up the data packets and sends them all in one go to the central node. Therefore, the tissues of the body closed to CH are subjected to advanced levels for a longer period of time. Developing a WBAN program also includes measures to mitigate radiation's negative effects are no easy task. In the second part of this work, we provide an in-depth analysis of the current state of the art in multi-hop. As per above model, it can be concluded that while constructing clustering procedures for WBANs, the vast majority of the studies to date ignore the radiation energy.

In this work, the FSR protocol is a clustering-based method designed to lessen the bad effects of radiation on humans. The "entropy-based fuzzy-TOPSIS MADM approach" is used into FSR in order to choose the most promising cluster heads while maintaining a low SAR. The objective of the paper is as follows. If a cluster node meets all three of the following conditions, it will be chosen as the cluster head.

(a) First, the selected CH node needs to use the least amount of transmission power to send the data packet to the hub.
(b) Second, the chosen CH node should be situated in an area of the body that is relatively unaffected by radio frequency (RF) energy.
(c) Third, the node with the highest residual energy in the CH tree should be chosen.

Varying areas of the body are given different SAR weights in the proposed FSR protocol. The entropy-based fuzzy-TOPSIS MADM function is then applied to the node characteristics by hub. TOPSIS is an acronym for "Technique for Preference Ordering through Statistical Idealism." Multi-attribute decision-making, or MADM for short.

1.1 Problem Definition

The topic of WBAN protocol design has generally ignored SAR reduction as an important criterion. Ahmed et al. [7], for instance, analyzed the potential adverse

effects of the WBAN system's EM radiation on human health under a variety of conditions. Node SAR evaluations based on transmission distance, antennae current, hub node location, sent power, and received signal strength were modeled mathematically. In this system, a static relay node's body location is optimized using a particle swarm pptimization (PSO)-based technique to allow sensor nodes to broadcast to it using the minimum amount of transmission power. Transmission power increases the volume of electromagnetic radiation. Also, due to the intensive nature of data relay activities, the onboard power requirements of a static relay node might quickly escalate. That point could mean the end of network functionality. In this work, we introduce a FSR routing protocol for WBAN that makes use of fuzzy-TOPSIS. The suggested approach includes measures to mitigate or prevent the harmful effects of CH node SAR on humans. The following are the contributions of the proposed work:

(a) The proposed protocol uses a low absorbing rate using the fuzzy-TOPSIS entropy algorithm.
(b) Comparing the PSO-based technique to the fuzzy-based entropy algorithm, it reduces the computing complexity.
(c) In contrast to the static relay method used by the PSO algorithm, the proposed protocol extends the lifespan of the network.
(d) The EADT technique is used in the proposed protocol to further increase the network's energy competence by preventing nodes data from sending out duplicate packets.

Remaining part of this paper is arranged as follows. Section 2 provides references and explanations for the relevant contents. Models and assumptions made during protocol implementation are discussed in Sect. 3. In Sect. 4, we will go over the specifics of the proposed FSR protocol. The outcomes are discussed in Sect. 5. In Sect. 6, we draw final conclusions and discuss some potential future lines of inquiry.

2 Related Work

In this part, we take a look at the various advanced data routing strategies currently in use within WBANs. It was important to note the CH selection and SAR reduction methods used by each existing project as part of the survey. The survey results pointed in a new approach for innovative research on low SAR WBAN design, revealing a gap in the existing literature. In this section, we also discuss the contributions of current research and related problem statements. When a relay node's energy falls below a certain level, or when the node's temperature increases over certain limitations, it is replaced. Because the strain is spread out, the lifespan of the network is increased.

The CH node collects the packets, compresses the data, and then transmits it on to the hub. For each new data transmission round, the CH node is chosen by these protocols. This is done in an effort to evenly share the CH workload across the nodes in the network. A longer lifespan for the network is achieved by fairly distributing the CH load across all nodes. Candidates for cluster nodes are evaluated using a variety

of metrics, including their energy levels, their transmission powers, their proximity to the hub, etc. The node that acquires the optimal set of desirable characteristics is promoted to CH. To choose its CHs, the RK-EERP protocol [11] uses a MADM strategy based on a composite index (CI). In this step, we assign a cluster importance (CI) score to every node in the cluster. A node's CI can be calculated by dividing its distance from the hub by its residual energy. With each transmission round, the CH position goes to the node that has the lowest composite index (CI) as [12]. The node with the energy dissipation during data transmission is chosen as the CH by the EB-MADM protocol [13]. Bhuyan et al. [14–17] also used generic model for data sharing in network system.

For a CI-MADM to select a cluster head node, the node's properties must be precisely characterized. On the other hand, accurate real-time evaluation of node properties is not always possible. Fuzzy linguistic attribute values are used to rate WBAN nodes for their suitability for collected CH. For the purpose of node ranking, fuzzy mapping rules are used [18, 19]. Attributes of the node are accurate representations of the node's properties when changed to certain fuzzy grades. Therefore, the FM-MADM technique efficiently deals with unclear node properties that have only been roughly approximated. Consequently, the FM-MADM method provides each node in the network with an equivalent amount of CH load. Network longevity is improved by distributing the load more evenly than it would be with the CI-MADM method. Classification model is also considered as [20–22]

The fuzzy-TOPSIS MADM method is used to pick the CH nodes in the EB-fg-MADM protocol [23]. The weighted MADM matrix is a representation of triangular fuzzy numbers (TFN) that have been modified from the properties of nodes. The most sought-after values for TFN attributes are stored in FPIS. The FNIS stores the most undesirable values for TFN attributes. Despite the uncertainty in the estimated attribute values, fuzzy-TOPSIS provides reliable decision-making. Fuzzy model-based data analysis is also considered in [24–26]. It also allows for an unlimited number of candidate nodes and attributes to be considered in the CH. The appropriate attribute weights in EB-fg-MADM are calculated using a global search optimization heuristic based on the genetic algorithm (GA). A complexity of O (n3) can be seen in the GA. The entropy-based fuzzy-TOPSIS method, proposed by Reddy et al. in [27], applies Shannon's entropy method to fuzzy data to establish attribute weights. The computational complexity of the entropy-based strategy is shown to be at the lower O ($n \log (n)$) level.

According to the presented analysis of existing WBAN protocols, node features such as residual energy, closeness to a hub node, and neighbor node density have been taken into account by these protocols when choosing a relay/cluster head, in spite of the fact that node transmission power is taken into account by DSCB [12] as a factor for CH selection. Data sharing through distributed system is identified by [28, 29]
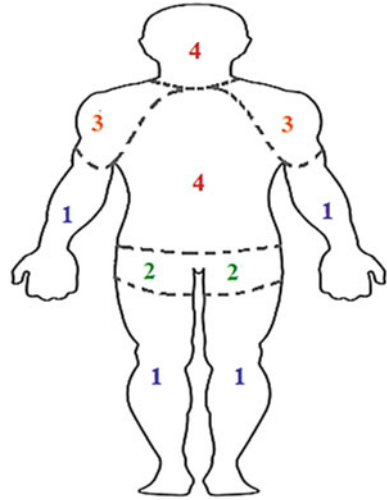
3 Framework for WBAN Model

The details of the many models and assumptions utilized in the construction and experimentation of the protocol are presented here.

3.1 Weights for SAR

More water in the brain and central nervous system makes these areas more vulnerable to the effects of electromagnetic radiation than, say, the limbs. Different areas of the body are given SAR weights between 1 and 4 based on their SAR sensitivity. The SAR is most significant when it is applied to the head or the trunk (i.e., 4). Extremities are assigned a SAR value of 1, the lowest value possible [8, 9]. The relative SAR weights of various body components are shown in Fig. 2.

3.2 Topology for the WBAN

The suggested protocol is implemented using a WBAN comprised of various nodes with single hub node. The topology of the network and the physical locations of the nodes that make it up are shown in Fig. 3. Here, it is expected that node $S4$ has been chosen as CH via the FSR protocol.

Low-bit rate data transmission is carried out by nodes $S7$ and $S8$ (below 1 kbps). In addition, they are situated on the areas of the body with the weakest node SAR weights (1 for both). In this case, nodes $S7$ and $S8$ are allowed direct data contact
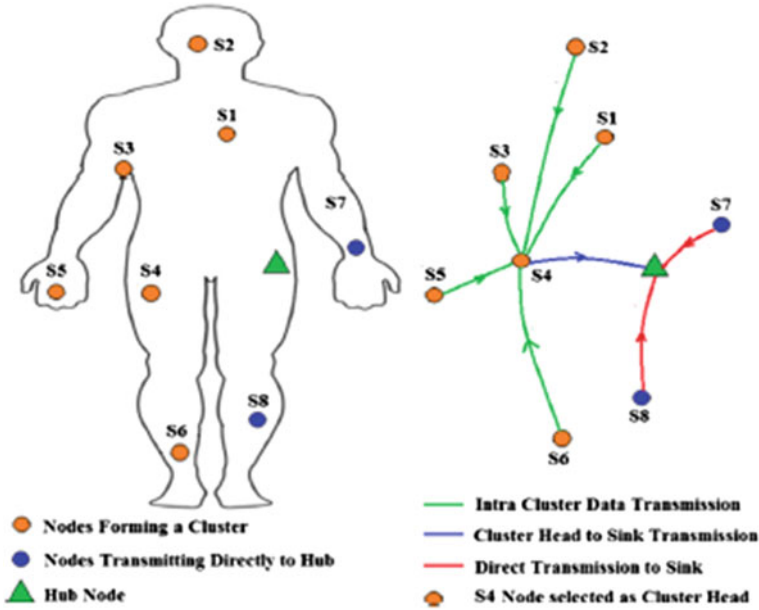
Fig. 3 WBAN topology with sensor positions

with the hub because their SAR impact is negligible. All the other nodes ($S1$ through $S6$) group together to create a cluster and communicate their packet status to the hub through the CH.

3.3 Transmission Energy Model for Node

As per our proposed work, energy expended by a node for W bits of data transmission is computed in Eq. 1 in the radiation model. Meanwhile, the energy expended by the receiving node increases as shown in Eq. 2 [30, 31] for the same number of data bits.

$$E_T(W, D) = E_{\text{TX - elect}} \times W + E_{\text{AMP}} \times \varepsilon \times D^\varepsilon \times W \tag{1}$$

$$E_R(W) = E_{\text{RX-elect}} \times W \tag{2}$$

where $E_{\text{Tx-elect}}$, $E_{\text{Rx-elect}}$, and E_{Amp} indicate the power needed to operate the node's transmitter, receiver, and amplifier circuits during a single-bit transfer. The separation between the sender and the receiver, denoted by D, is of critical importance. Human attrition coefficient is denoted by ε. There are extra responsibilities placed on the CH node, such as receiving data packets sent by child nodes and aggregating that data. Thus, Eq. 3 and 4 are used to approximate the transmission and reception energies

at a CH node.

$$E_{\text{TCH}}(W, D) = (E_{\text{TX - elect}} + E_{\text{AMP}} \times \varepsilon \times D^{\varepsilon}) \times (n + 1) \times W \times \delta \qquad (3)$$

$$E_{\text{RCH}}(W) = E_{\text{RX - elect}} \times W \times n \qquad (4)$$

where n number of nodes are sending packets to CH. CH's data aggregation ratio is represented by the letter δ. When evaluating the efficiency of routing protocols, sensor and processor power consumption are often disregarded.

3.4 Connecting Path Loss Model

Using on-body propagation channels, EM signals in a WBAN experience path loss as they travel. The narrow ISM band communication channel loss model is given by Eq. 5 [15], which operates in the 2.4 GHz range.

$$\text{PL}(D, F)\text{indB} = 10. \log_{10}\left(\frac{4\pi D_r f}{C}\right) + 10\eta \log_{10} \frac{D}{D_r} + G_{\sigma} \qquad (5)$$

where G is the route loss index, D is the transmission distance, D_r is the reference distance, PL is the path loss, and G_{σ} is a Gaussian random variable with zero mean and r standard deviation. In the current study, we evaluate the stability of a connection by measuring its "Received Signal Strength" (RSS). The route loss (in dB) is the amount of attenuation that the transmitted signal experiences on its way to the receiving node. Then, using Eq. 6, we can determine the P_r(in dBm) of the received signal.

$$P_r(\text{in dBm}) = P_t(\text{in dBm}) - \text{PL}(\text{in dB}) \qquad (6)$$

If the received signal strength P_r is lower than the receiver sensitivity P_{rsens}, the link quality is poor and the packet is dropped. If it does not, then the packet is considered to have successfully arrived at its final destination.

4 Proposed Low SAR Routing Scheme—FSR

4.1 Location-Based Approach

Each transmission round begins with the hub accumulating data about the network nodes. Node identifiers and current residual energy values are transmitted in each HELLO message. Nodes all use the same transmission power, P_t, to send out their

HELLO messages. In this context, *V* represents the group of connected devices. Hub now uses the following equation to estimate node-route *u*'s loss PL (*u*, hub).

$$PL(u, \text{hub}) \text{ in dB} = P_t \text{ in dBm} - P_{\text{rec}}(u, \text{hub}) \text{ in dBm} \tag{7}$$

As per WBAN model with Eq. (5), the hub is determined the distance as *D*(*u*, Hub) as Eq. (8).

$$D(u, \text{Hub}) = D_r 10^{\frac{PL(u,\text{hub}) - 20.\log_{10}\left(\frac{4\pi D_r f}{C}\right) - \chi\sigma}{10}} \tag{8}$$

Again, the hub maintains least transit power as P_{tmin} (*u*, hub) as *u*-node providing data packets to the one hub as Eq. (9).

$$P_{\text{tmin}}(u, \text{ hub}) = PL(u, \text{ hub}) \text{in dB} + P_{\text{rsens}} \text{in dBm} \tag{9}$$

As a result, the hub learns the following about node-*u* after receiving its HELLO message:

(a) The identifier for node-*u*.
(b) $E_r(u)$, residual energy for node (*u*).
(c) To send data from node-*u* to hub, the least transmit power required is P_{tmin} (*u*, hub).

Different data are preloaded about the locations on the human body where various nodes have been implanted, along with the corresponding specific absorption rates (SARs). Therefore, the SAR weight (SARW) of node-*u* is known to hub via node-identifier *u*'s (ID) (*u*). This allows the hub to collect data on the three properties of each node in the network (residual energy (Er (*u*)), minimum transmit power (P_{tmin} (*u*, hub), and SAR weight (SARW (*u*))).

4.2 Fuzzy-TOPSIS-Based Entropy Analysis for SAR-Low CH Selection

A cluster containing nodes *S*1–*S*6 is the desired end result. Since the CH is selected from among these nodes, we refer to them as "CH candidates." As part of the low SAR CH selection procedure, hub takes the values of three node attributes as Er (*u*), P_{tmin} (*u*, hub), and SARW (*u*) *V* (*u*,(*U*)) into account. Here, *U* represents the group of nodes that make up the cluster (i.e., CH candidates).

4.3 Fuzzy Domain Attribute Transformation

In order to classify the values of node attributes into fuzzy linguistic terms, FTF first converts the data to triangular fuzzy numbers (TFN). To do this, the FTF first converts the values of the numerical attributes of all CH candidate nodes to the interval [0, 10]. After the qualities have been normalized, they are translated into a fuzzy language phrase. Every linguistic fuzzy phrase has an associated TFN ($f_{min}, f_{mode}, f_{max}$). Several fuzzy language terms and their TFNs are explained in this section.

TFNs are created by transforming the normalized attribute values (Ern (u), P_{tmin} (u, hub), and $SARW_n$ (u) $\forall u \in U$) into their corresponding TFNs (Erf (u), $P_{tmin}f$ (u), and $SARW_f(u)$ $\forall u \in U$). In order to do the fuzzy transformation, the triangle fuzzy membership function is utilized. A normalized residual energy attribute value Ern is fuzzy transformed as follows:

$$\mu_f(E_m) = \begin{cases} \frac{E_m - f_{min}}{f_{mode} - f_{min}} & \text{if } f_{min} \leq E_{rn} \leq f_{mode} \\ \frac{f_{max} - E_{rn}}{f_{max} - f_{mode}} & \text{if } f_{mode} \leq E_{rn} \leq f_{max} \\ \text{else} \qquad 0 \end{cases} \tag{10}$$

5   Experimental Analysis

We considered our experimentation as per proposed method. The suggested FSR protocol is evaluated with the help of MATLAB tools. The results are validated and improved upon by comparing them to those of the PSO-based low SAR path discovery methodology [8]. The simulation is considered to measure the performance of the suggested FSR model based on network lifetime, stability period, node-to-node delay, and computational complexity.

5.1 SAR Network Efficiency

We can determine the specific absorption rate (SAR) that a node's EM radiation causes in living tissue based on Eq. 11.

$$SAR = \int \frac{\sigma(r)|E(r)|^2}{\rho(r)} dr \tag{11}$$

The electrical conductivity, r (S/m), and the tissue density, q (Kg/m3), are shown in the following equation. The electrical field induced is denoted by the symbol E (Volts/m). E is proportional to the Pt of the transmitting node (dBm). As a result, the SAR of a node depends on the strength of the signals it sends. The experimental

Table 1 CH selection in % for several nodes

Nodes	CH selection in %	Weighted node SAR (mW/Kg)
S1	16	0.7
S2	0	1.7
S3	5	0.5
S4	21	0.4
S5	12	0.28
S6	43	0.3

Table 2 % of selection of CH nodes with different SAR weights

	CH selection %		
Methods →	FSR (proposed)	RK-EERP	PSO-based relay
Range (0–0.5)	80	47	0
Range (0.5–1)	20	39	100
> 1	0	14	0

setup is measuring for the SAR of various CH candidate nodes [32]. Minimum transmit power is needed for various CH candidate nodes. The SAR sensitivity of a biological component is also a factor in how seriously radiation affects that part. When the suggested FSR protocol is put into place, the proportion of selected CHs is from among various CH candidate nodes. Table 1 explains the percentage of CH candidate nodes that were selected.

To ensure optimal signal quality, 37.55% of all transmission rounds are allocated to the CH nodes as Table 2. As much as 13.68% of the time, transmissions are prioritized toward the CH nodes. Low SAR path finding protocols based on PSO rely on a single relay node to send data in all transmission rounds. Bar charts demonstrating the percentage of nodes selected from CH that have various weighted SAR values. As shown by the results, the suggested FSR protocol's consideration of the SAR factor results in a higher proportion.

5.2 Durations of Network Stability and Availability

The network stability period is the amount of time during which all nodes in a WBAN are fully functioning, while the lifetime of a WBAN is the period of time during which it is operational before all of its nodes become inoperable. The number of still-alive nodes and the amount of energy left in the network after a given number of transmission rounds are depicted in Tables 3 and 4, respectively. They summarized data on how long different protocols maintain network stability and how long individual networks last. Results show that the proposed FSR protocol can maintain a network for 5360 rounds of transmission before it loses connectivity and 11,310 rounds of transmission before it dies.

Table 3 Number of alive nodes per round

	Alive node count		
Methods →	FSR (proposed)	RK-EERP	PSO-based relay
Rounds (5000)	8	8	8
Rounds (10,000)	2	0	0
Rounds (15,000)	0	0	0

Table 4 Network residual energy

	Residual energy		
Methods →	FSR (proposed)	RK-EERP	PSO-based relay
Rounds (0)	4	4	4
Rounds (5000)	1.3	0.2	1.1
Rounds (10,000)	0.1	0	0
Rounds (15,000)	0	0	0

In comparison, the RK-EERP technique has a lower stability and longevity. The static relay node bears the brunt of the relaying load for all nodes during each round of transmission. Therefore, it quickly drains its energy supply. When this happens, sensor nodes quickly drain their power supply because they must begin transmitting data to the hub in real time. By ensuring that the network's nodes all use the same amount of energy, dynamic cluster head selection extends its useful life. The EADT technique saves a lot of power by preventing the nodes from sending the duplicated data packets. For the first 5000 iterations of data transmission, the EADT system reduces WBAN energy consumption by 0.72 Joules. The proposed FSR protocol's 11,310-round network lifetime is significantly higher because to the addition of EADT and dynamic cluster head selection procedures.

5.3 Data Flow Rate

Data loss must be kept to a minimum during data transmission by nodes. To comply with IEEE 802.15.6 requirements for WBAN [6], the success rate of data packets must be at least 90%.

If a node determines that the packet it is now transmitting is identical to the last one it received, it will remove the packet from the transmission. In contrast,

during this cycle the hub will duplicate the same node's last packet. Because of this, a packet can be sent to the hub without worrying about it being lost because of a choppy connection. Therefore, the suggested FSR protocol can benefit from increased network throughput if the event-aware data transmission technique is used.

The proposed FSR protocol is able to provide 5.45104 data packets to the hub after 15,000 rounds of transmission. The packets successfully delivered in the same number of transmission rounds are 2.9310104 and 5.110104, respectively.

5.4 Comparative Experimental Analysis

When all nodes in a WBAN are operating normally, we can consider it as the network stability period. A WBAN's lifetime is the time it remains functional before all of its nodes stop working. Tables 3 and 4 illustrate the comparative result analysis among different methods with proposed method as per the percentage of active nodes and the remaining network energy after a certain number of transmission rounds. The number of still-alive nodes is active after 5000 rounds in proposed method where nodes are not sustained in existing methods. Similarly, the amount of energy left in the network after a given number of transmission rounds is depicted in Table 4.

Further, the comparative performance of various methods is determined using different parameters such as (a) CH selection % (as low, moderate, High SAR), (b) network stability period as number of transmission rounds, (c) network lifetime, (d) throughput, and (e) latency. The network stability period is more in proposed method (PSR) than other two methods (i.e., RK-EERP and PSO-based relay) as shown in Fig. 4.

RK-EERP protocol has less stability and lifetime periods as per number of rounds of transmission of data. Similarly, PSO-based SAR path needs more static relay
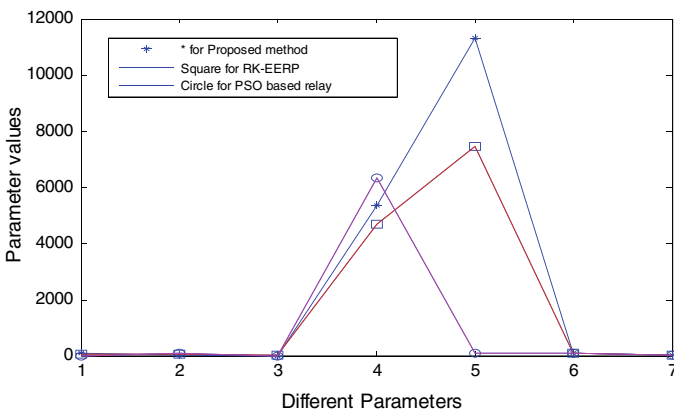
Fig. 4 Comparative result analysis of existing and proposed method

nodes. So, sensor nodes are unable to forward additional load. Thus, the PSO-based method indicates more stability period on more number of rounds. Thus, this protocol needs to take more loads for each node as per transmission round with more power consumption. But our proposed FSR protocol saves the energy strategy as per cluster head selection and event-aware data transmission strategies. These two approaches extended network lifetime and breaking the extra data packet transmission.

5.5   *Computational Complexity*

Fuzzy-TOPSIS MADM is a technique incorporated into the proposed FSR protocol. In this context, n represents the total number of nodes in the WBAN and m indicates the total number of node attributes that will be used to make the CH decision. As a result, the FSR protocol gives $O(n^2)$ complexity. The particle population size, denoted by N, and the maximum number of PSO iterations, denoted by R, are also variables in this expression. Thus, $O(n^2)$, $O(n^2)$, and $O(n^3)$ complexity levels are represented by the provided FSR, RK-EERP, and PSO-based low SAR path discovery procedures.

6   Conclusions

In this work, we described the development of a wireless body area networks where the hazardous effects of EM radiation are incorporated into the suggested protocol. In contrast to other methods, it does not favor the CH node with the highest SAR. As a result, the proposed method chooses CH nodes that route data with a low SAR for 80% of the total number of transmission rounds. The suggested protocol supports throughput and latency characteristics that meet or exceed IEEE 802.15.6 standards, and it extends the network's lifetime up to 11,310 transmission rounds. Developing the ML-aided behaviors prediction system necessitates first establishing a baseline for activity monitoring based on sensors. Research in the future work will concentrate on creating a healthcare-specific smart monitoring and decision-making support system that makes use of the Internet of Things and machine learning (ML). Hospital inpatients' activity levels will be tracked using sensors as the basis for the planned study. Through the use of machine learning and statistical analysis, sensor data will be used to recognize actions and train neural networks for behavior prediction.

References

1. Cavallari R, Martelli F, Rosini R, Buratti C, Verdone R (2014) A survey on wireless body area networks: technologies and design challenges. IEEE Commun Surv Tutor 16(3):1635–1657. https://doi.org/10.1109/SURV.2014.012214.00007
2. Patel M, Wang J (2010) Applications, challenges, and prospective in emerging body area networking technologies. IEEE Wirel Commun 17(1):80–88. https://doi.org/10.1109/MWC.2010.5416354
3. Liu J, Sohn J, Kim S (2017) Classification of daily activities for the elderly using wearable sensors. J Healthc Eng 7. https://doi.org/10.1155/2017/8934816
4. Wu T, Wu F, Redoute´ JM, Yuce MR (2017) An autonomous wireless body area network implementation towards IoT connected healthcare applications. IEEE Access 5:11413–11422. https://doi.org/10.1109/ACCESS.2017.2716344
5. Park C, Liu J, Chou PH (2005) Eco: an ultra-compact low-power wireless sensor node for real-time motion monitoring. In: IPSN 2005. Fourth international symposium on information processing in sensor networks, pp 398–403. https://doi.org/10.1109/IPSN.2005.1440956
6. Smith DB, Hanlen LW (2015) Channel modeling for wireless body area networks. In: Mercier P, Chandrakasan A (eds) Ultra-low-power short-range radios, 1st edn. Springer, Cham, pp. 25–55. https://doi.org/10.1007/978-3-319-14714-7_2
7. Ahmed G et al (2018) Rigorous analysis and evaluation of specific absorption rate (SAR) for mobile multimedia healthcare. IEEE Access 6:29602–29610. https://doi.org/10.1109/ACCESS.2018.2839909
8. Wu T, Lin C (2015) Low-SAR path discovery by particle swarm optimization algorithm in wireless body area networks. IEEE Sens J 15(2):928–936. https://doi.org/10.1109/JSEN.2014.2354983
9. Tuovinen T, Berg M, Yazdandoost KY, Hämäläinen M, Iinatti J (2013) On the evaluation of biological effects of wearable antennas on contact with dispersive medium in terms of SAR and bio-heat by using FIT technique. In: Proceeding of 7th international symposium on medical information and communication technology (ISMICT), pp 149–153
10. Cicioğlu M, Çalhan A (2019) Dynamic HUB selection process based on specific absorption rate for WBANs. IEEE Sens J 19(14):5718–5722. https://doi.org/10.1109/JSEN.2019.2906044
11. Khan RA, Xin Q, Roshan N (2021) RK-energy efficient routing protocol for wireless body area sensor networks. Wireless Pers Commun 116:709–721. https://doi.org/10.1007/s11277-020-07734-z
12. Ullah Z, Ahmed I, Razzaq K, Naseer MK, Ahmed N (2017) DSCB: dual sink approach using clustering in body area network. Peer-to-Peer Netw Appl. https://doi.org/10.1007/s12083-017-0587-z
13. Choudhary A, Nizamuddin M, Singh MK, Sachan VK (2019) Energy budget based multiple attribute decision making (EB-MADM) algorithm for cooperative clustering in wireless body area networks. J Electric Eng Technol. https://doi.org/10.1007/s42835-018-00006-8
14. Bhuyan HK, Dash SK, Roy S, Swain DK (2012) Privacy preservation with penalty in decentralized network using multiparty computation. Int J Adv Comput Technol (IJACT) 4(1):297–303
15. Kaur N, Singh S (2017) Optimized cost effective and energy efficient routing protocol for wireless body area networks. Ad Hoc Netw. https://doi.org/10.1016/j.adhoc.2017.03.008
16. Bhuyan HK, Kamila NK, Dash SK (2011) An approach for privacy preservation of distributed data in peer-to-peer network using multiparty computation. Int J Comput Sci Issues (IJCSI) 3(8):424–429. ISSN (Online version): 1694-0784, ISSN (Print version): 1694-0814
17. Bhuyan HK, Mohanty M, Das SR (2012) Privacy preserving for feature selection in data mining using centralized network. Int J Comput Sci Issues (IJCSI) 9(3):434–440. ISSN (Online version): 1694-0784, ISSN (Print version): 1694-0814
18. Shunmugapriya B, Paramasivan B (2021) Fuzzy based relay node selection for achieving efficient energy and reliability in wireless body area network. Wireless Pers Commun. https://doi.org/10.1007/s11277-021-09027-5

19. Chavva SR, Sangam RS (2019) An energy-efficient multi-hop routing protocol for health monitoring in wireless body area networks. Netw Model Analy Health Inf Bioinf 8:21. https://doi.org/10.1007/s13721-019-0201-9
20. Bhuyan HK, Reddy CV (2018) Madhusudan: sub-feature selection for novel classification. In: International conference on inventive communication and computational technologies (ICICCT), IEEE explore, 20–21 April 2018. https://doi.org/10.1109/ICICCT.2018.8473206
21. Bhuyan HK, Ravi VK (2021) Analysis of sub-feature for classification in data mining. IEEE Trans Eng Manag 1–15
22. Bhuyan HK, Raghu Kumar L, Ramakrishna RK (2019) Optimization model for Sub-feature selection in data mining. In: 2nd international conference on smart systems and inventive technology (ICSSIT 2019), IEEE explore, pp 1–6
23. Choudhary A, Nizamuddin M, Sachan VK (2020) A hybrid fuzzy-genetic algorithm for performance optimization of cyber physical wireless body area networks. Int J Fuzzy Syst 22:548–569. https://doi.org/10.1007/s40815-019-00751-6
24. Bhuyan HK, Chakraborty C, Pani SK, Ravi VK (2021) Feature and sub-feature selection for classification using correlation coefficient and fuzzy model. IEEE Trans Eng Manag 1–15
25. Bhuyan HK, Kamila NK, Jena LD (2016) Pareto-based multi-objective optimization for classification in data mining. Cluster Comput 19(4):1723–1745
26. Bhuyan HK, Kamila NK (2014) Privacy preserving sub-feature selection based on fuzzy probabilities. Cluster Comput 17(4):1383–1399
27. Suchith Reddy A, Rathish Kumar P, Anand Raj P (2019) Entropy-based fuzzy TOPSIS framework for selection of a sustainable building material. Int J Const Manag. https://doi.org/10.1080/15623599.2019.1683695
28. Chakraborty C, Mishra K, Majhi SK, Bhuyan HK (2022) Intelligent latency-aware tasks prioritization and offloading strategy in distributed fog-cloud of things. IEEE Trans Ind Inf 1–8
29. Bhuyan HK, Chakraborty C (2022) Explainable machine learning for data extraction across computational social system. IEEE Trans Comput Soc Syst 1–15
30. 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-(HICSS '00), Washington DC, USA, vol 2. IEEE Computer Society, pp 10. https://doi.org/10.1109/HICSS.2000.926982
31. Choudhary A, Nizamuddin M, Zadoo M (2022) Body node coordinator placement algorithm for WBAN using multi-objective swarm optimization. IEEE Sens J 22(3):2858–2867. https://doi.org/10.1109/JSEN.2021.3135269
32. Sudha MN, Benitta SJ (2016) Design of antenna in wireless body area network (WBAN) for biotelemetry applications. Intell Decis Technol 10:365–371

Machine Learning Techniques for Human Activity Recognition Using Wearable Sensors

Moushumi Das, Vansh Pundir, Vandana Mohindru Sood, **Kamal Deep Garg**, **and Sushil Kumar Narang**

Abstract The HAR refers to a system that monitors human behaviors and activities by collecting data from a wide range of distinct kinds of wearable sensors. This system is known as human activity recognition. HAR enables in improving Health, Quality of life, Biomedical, and Mental Health. Recently, there has been an increase in demand for a range of HAR techniques, including those that make use of deep learning, machine learning, and CNN algorithms. Specifically, this need has been on the rise in the United States. In the most recent few years, this field of research has achieved major advancements, and there is still a substantial amount of research being carried out. Human activity recognition (HAR) has seen tremendous growth in the disciplines of pervasive computing, human–computer interaction, and human behavior analysis. Recently, machine learning (ML) algorithms have proven effective to predict a variety of human actions using time-series data from wearable sensors and cell phones. Even though ML-based techniques were excellent at identifying activities, managing time series data remains difficult. Time-series data still has a lot of problems, including hard feature extraction, extremely biased data, etc. Furthermore, manual feature engineering is a key component of the bulk of

M. Das · V. Pundir · V. M. Sood (✉) · K. D. Garg · S. K. Narang
Chitkara University Institute of Engineering and Technology, Chitkara University, Chandigarh, Punjab, India
e-mail: vandana.sood@chitkara.edu.in

M. Das
e-mail: moushumi3813.be21@chitkara.edu.in

V. Pundir
e-mail: vansh3856.be21@chitkara.edu.in

K. D. Garg
e-mail: kamaldeep.garg@chitkara.edu.in

S. K. Narang
e-mail: sushilk.narang@chitkara.edu.in

HAR approaches. Since each measurement made by wearable sensors like gyroscopes and accelerometers has a timestamp associated with it, time series data is the format utilized for data on human activity. The raw sensor data must be processed to obtain the necessary temporal properties for HAR. In-depth feature engineering and data pre-processing are required for the majority of HAR approaches, which in turn necessitates subject-matter expertise. These techniques are time-consuming and application-specific. This paper begins with an in-depth discussion of HAR, then moves on to a discussion of the various machine learning methods, and then concludes with a discussion of wearable sensors and the applications that can be used with them.

Keywords Human activity recognition (HAR) · Machine learning (ML) · Wearable sensors · Sensors · Convolution neural networks (CNN) · Health · Quality of life · Biomedical · Mental health · Deep neural network

1 Introduction

HAR uses sensors to predict human movement. The raw sensor data is used to model the activities. HAR uses machine learning to operate automatically. Sensor data is utilized to analyze HAR's outputs. Complex jobs are simplified and easier to analyze [1]. HAR covers human-to-human interactions including behavior, gestures, communication, and interpersonal activities like moving, eating, and sleeping. These motions include gestures, atomic activities, human-to-human or object-to-human interactions, collective actions, behaviors, and events [2–4]. Through remote wound image analysis [5] or daily sensory data collection, research has advanced throughout COVID-19, especially in remote medical care. Body sensors and smartphone sensors can now monitor everyday activities. Human action recognition utilizing sensor inputs is a hot topic. HAR detects and identifies actions based on sensor data. HAR is used for gaming, security, and survival [6]. Recognizing habitual behaviors is key to managing patients and living well. Detecting activity helps doctors discover and treat major ailments in seniors [7]. HAR includes data collection, preprocessing, segmentation, feature extraction, and classification [6]. Table 1 provides us with a general description of the actions used in HAR like gestures, atomic actions, and behavior.

Table 1 General description of actions in HAR

Action	Description
Gestures	The movement of body parts according to the situation
Atomic actions	The movements are related to actions in complex activities
Interactions between humans and objects or between humans	The movements that include objects or other persons
Group actions	Actions performed within a group
Behaviors	The physical actions with emotions
Events	Events are the activities that show an impact on society

1.1 Applications

a. Smart Homes—Human Activity Recognition promotes independence. Sensors around residences make solitary and elderly living easier. Sensors help us live safely and securely. HAR provides safety and makes life fun by incorporating game features. It streamlines communication. Smart houses have two sensors. One is a watch-like sensor and the other is environmental [7–11].

b. Entertainment Systems—TI helps create a fake setting. Two-dimensional animations become three-dimensional. TI's application can colorize monochromatic video. Augmented reality applications use visual systems, non-visual sensors, and multimodal sensors. Visual augmented reality systems benefit smart home systems, healthcare monitoring systems, and public space security and surveillance to help active, aided living. Blind sensors Ball switches, RFID, and more advanced accelerometers are used in audio processing and computer vision sensing [12–14].

c. Healthcare—Wearable sensor localization, data preprocessing, and recognition are the three basic HAR components. Sensors and computers help monitor humans in healthcare. Tracking athlete training, recovery, and robot monitoring are primary uses. HERMES, CASAS, Mobiserv, SWEET-HOME, and HBMS are recent healthcare projects [15–18].

1.2 Challenges

a. Subject Sensitivity—Training and testing subjects greatly affect the accuracy of accelerometer-based activity recognition. Various people have different motion patterns. She/he may have diverse patterns even for the same subject. Training

and testing on the same subject are the most accurate, according to research. Multiple-subject training and testing have the second-highest accuracy.

b. Location Sensitivity—Due to an accelerometer's inherent characteristics, both in wearable sensors and mobile devices, the raw reading of the sensor is highly dependent on the orientation and placement of the sensor on the subject's body.

c. Activity Complexity—Complex user activities test the recognition model. The classification method has trouble recognizing motion during activity transitions. Multitasking may confound the classifier, which is trained on a one-activity-per-segment assumption.

d. Energy and Resource Constraints—Applications for activity recognition necessitate continuous sensing and energy-intensive online classification model updating. It may also need substantial computational resources (like mobile phone memories) for online updating.

e. Elderly and Youth—A growing need for aged care (both physically and emotionally) exists, in part due to the baby boomer generation's impending retirement. The current study on human activity monitoring has as one of its main objectives the creation of new tools and programs for the care of the elderly [19].

1.3 Contribution of the Work

This work identifies and addresses several challenges faced in HAR processing. The objectives of the work are:

a. Identification of human activities is among the most vitally significant applications in various fields. Because of their semantic similarity, human actions that are performed daily, such as walking, sitting, and jogging, as well as acts that are essential to human life, such as falling forward or backward, are present.

b. Along with the applications various limitations are being discussed that need to be improvised for better implementation.

c. The various ML techniques, the work, and the progress occurring in this field are discussed.

d. The use of wearable sensors is an important part of HAR, and many of the applications are being discussed.

The novelty of this paper is how can we tackle the issue of Human Activity Recognition using various ML-based techniques and use wearable sensors along with those techniques for better results.

The remaining sections are arranged as follows: Sect. 2 presents the related work done for HAR and the various techniques used for it. Section 3 discusses the ML Techniques in HAR, and Sect. 4 gives the idea about wearable sensors and their applications. Section 5 presents the discussion of the work done and Sect. 6 concludes and discusses the future scope.

2 Related Work

Numerous challenges have been faced by researchers in this area. There are still older studies available that analyze the HAR using the publicly accessible dataset. There are a few HAR datasets that are available to the public.

Ullah et al. [20] proposed at WISDM Accelerometer and gyroscope sensors are used for the modalities in this dataset. He further used the Bi-directional LSTM model.

Nweke et al. [21] using WISM data set, PAMPA2 dataset, and mHealth datasets proposed Hybrid CNN with LSTM classification.

Dua et al. [22] raw data from wearable sensors was used with minimum pre-processing, and no handmade feature extraction algorithms were applied. On the UCI-HAR, WISDM, and PAMAP2 datasets, the accuracies are 96.20, 97.21, and 95.27%.

Teng et al. [23] UCI HAR dataset, UniMib SHAR dataset, and PAMP Dataset were used. The results for the basic designs evaluated show that local loss outperforms global loss. Despite the reduced number of parameters, the local loss could approach state-of-the-art on a variety of HAR datasets at no additional expense.

Ta et al. [24] ELA approach suggested that combines a gated recurrent unit (GRU), a convolutional neural network (CNN) stacked over the GRU, and a deep neural network (DNN). The input samples for DNN were a feature vector including 561 time-domain and frequency-domain parameters. The whole linked DNN was utilized to fuse three models for activity classification.

Li et al. [25] applied convolutional and long-short-term memory classification on the UniMiB SHAR dataset. The approach was created to obtain characteristics that would characterize data's short- and long-term temporal dependencies.

Gupta et al. [26] CNN-GRU is a proposed hybrid deep neural network architecture that combines convolutional and gated recurrent units for human activity recognition. This model was successfully validated on the WISDM dataset and produced accuracy that is comparable to other cutting-edge deep neural network models generated with Auto ML, such as Inception Time and Deep Conv LSTM. Table 2 represents all the details of the related work in a tabular manner.

Based on the work that is connected to the one that was just mentioned, it has been noticed that CNN and LSTM models in HAR still require much investigation, which is its own distinct area of study. There have been a variety of models and methods proposed for managing the HAR. In many of these cases, machine learning techniques and wearable sensors have been combined to achieve improved accuracy. A new deep transition-aware HAR system is required. In addition, training transition data in a short period of time requires a significant amount of labor that must be correctly categorized. A lot of research is still in progress that is introducing new technologies in this field but machine learning techniques are becoming the most useful nowadays.

Table 2 Summary of related work in HAR

Reference	Year	Dataset	Classification method	Accuracy (%)	Advantages
[24]	2022	Input samples (from multiple sources)	Deep learning neural network	96	Improved accuracy and reduced computational time
[26]	2021	Wearable sensors	CNN model	95–97	No handcrafter features
[22]	2021	PAMP dataset WISDM	Novel hybrid DNN model	95	Better accuracy as compared to state-of-art
[23]	2020	UCI HAR	State-of-arts on HAR datasets	97.93	Local loss words are better than the global loss
[20]	2019	Gyroscope and accelerometer	Bi-directional LSTM model	80–90	Diversity of data
[21]	2018	Opportunity and UniMiB SHAR	Hybrid CNN with LSTM	91.73	Robust nature of the hybrid mechanism
[25]	2018	UniMiB-SHAR	Convolution, long and short-term memory	94	Identify data features characterizing both short and long-term time relationships

3 Machine Learning Techniques in HAR

HAR is a subfield of human behavior analysis, human–computer interaction, and ubiquitous computing [27]. In recent decades, various artifact reduction approaches have emerged. In addition to the theory that movement artifacts affect the signal, most of these strategies rely on signal qualities. When these techniques were devised, identifying other signals closely related to movement artifacts was difficult or inaccessible [28].

Numerous Machine Learning (ML) approaches have recently been used for HAR. Random Forest, Support Vector Machine (SVM), and k-Nearest Neighbors are a few of them (kNN). For instance, a feature selection technique based on game theory was established, while the HAR method published employed features based on collective empirical mode decomposition. Utilizing SVM and kNN classifiers, the features and feature selection approach were assessed. Sensor fusion and classifier fusion layers were the components of the entropy-based hierarchical fusion model that was developed for HAR. The weights in the model were calculated using the entropy weight approach. Utilizing ML approaches for activity recognition necessitates a significant amount of work in feature extraction, data preparation, and pre-processing, all of which need domain knowledge.

These approaches use custom feature learning algorithms, even though they obtain respectable recognition performances. Additionally, the characteristics that may be retrieved. These approaches are job-specific and cannot be applied to other jobs [23].
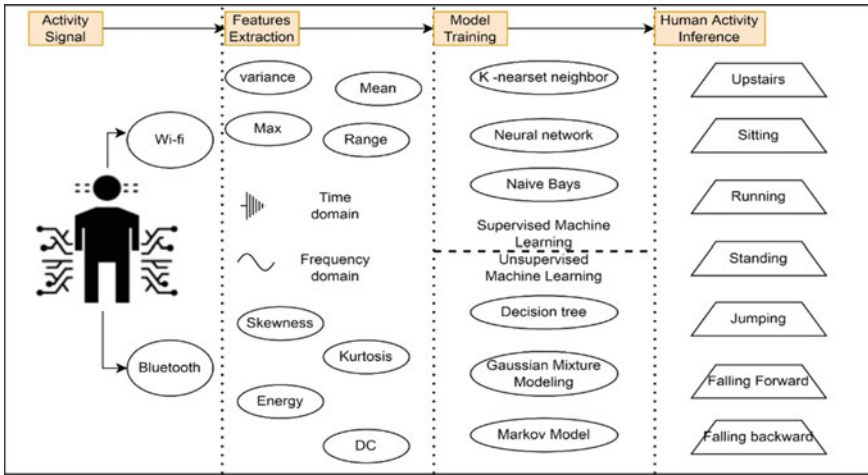
Fig. 1 Machine learning for human activity recognition

Figure 1 depicts the machine learning techniques used in HAR which include activity signal, features extraction, model training, and human activity inference.

Wearable Sensors play a crucial part in Machine Learning. Wearable sensors are essential for improving the performance of HAR systems so that the ML-based model can be used in HAR to provide accurate results.

4   Use of Wearable Sensors in HAR

Wearable sensors can detect and locate human actions. These sensors can be placed in bands, smartwatches, garments, or telephones. Both are choices. Accelerometers, gyroscopes, magnetometers, electromyography (EMG), electrocardiography (ECG), and others are IoT-based sensor devices. These safe sensors can be worn to record motions. The activity data a person provides correlates directly with the placement of wearable sensors on the body. The breastbone, lower back, and waist are suitable sensor locations [26].

a. Biometric User Identification—Recent studies suggest that mobile sensors and accessories can be used to monitor behavioral features. Gait recognition, touch dynamics, and keystroke dynamics are examples. Behavioral biometrics can be used to covertly verify and identify people using mobile devices like smartphones. Continuous or active authentication systems were developed using behavioral biometrics [29].

b. Data Pre-processing—Wearable sensor data are cleaned and normalized during data preparation to create a consistent dataset for training an identification model. This is done to gather body-worn sensor data. Outlier data values are also

removed. Eliminated sounds. Sensor data were pre-processed using a median filter and a third-order low-pass Butterworth filter with a cutoff frequency of 20 Hz frequency to reduce background noise [29].

c. Fitness and Lifestyle—Sitting, walking, lying down, climbing stairs, jogging, and running are physical activities. Regular exercise prevents obesity, diabetes, and heart disease. It boosts mental health. Wearable devices collect data on an individual's daily habits and health, such as exercise length and intensity. Fitbit measures smart device energy. Tracking activities help prevent chronic diseases. Driving, walking, cycling, and public transit are linked to obesity [30].

d. Healthcare and Rehabilitations—HAR has influenced healthcare and rehabilitation's ability to make an accurate diagnosis and capture vital information. Wearables let doctors evaluate and monitor patient health by recording, storing, and sharing data with other medical facilities. Pulmonary illness, including asthma and COPD, is a leading cause of death and morbidity. Recent research has used wearable technology to detect coughing, a pulmonary symptom. Electromyography (EMG) sensors are widely used to detect muscle movements and hand motions to improve prosthesis control [30].

e. Human Computer Interactions—Wearable HCI technology has made it easier to operate and connect with electrical devices, computers, and robotics. An inertial measurement unit wrist-worn wearable (IMU) can detect hand-shakes to control smart gadgets. The user can skip songs by shaking their hand instead of bringing up the screen and pressing a button. AR and VR have changed how we interact and experience the world. These apps may be able to produce cold or hot weather by altering the virtual environment and enabling more realistic human-virtual object interaction [30].

f. InertialMeasurement Unit (IMU)—IMUs have accelerometers, gyroscopes, and magnetometers. A three-axis accelerometer monitors linear motion and gravitational forces. Gyroscopes measure rotation (roll, yaw, and pitch). Magnetometers read Earth's magnetic fields. Magnetometers are used to determine posture and orientation based on the geomagnetic field and are not included in HAR data processing. HAR ignores geomagnetism. HAR uses accelerometers and gyros. 6-axis IMUs have accelerometers and gyroscopes. This creates a 9-axis IMU with a 3-axis magnetometer (IMU). Smartphones and other wearables make IMU data more common and accessible [30–33].

5 Discussion and Results

Deep learning, machine learning, and CNN are among of the methods employed in HAR. In this study, we have covered the role that machine learning techniques used in HAR. Due to the vast range of applications, it can be put to, including those in the fields of healthcare, intelligent surveillance systems, smart homes, rehabilitation, and many others has sparked a lot of interest. Human movements can be shown using a variety of data modalities, including acceleration, RGB, Wi-Fi signal, audio,

depth, and skeletal data, among others [33–38]. The research primarily focuses on approaches, offering fresh, practical methods.

The paper discussed HAR which is Human Activity Recognition and how wearable sensors can be used in reference to HAR, the various applications for it, and briefly explored how it may benefit us. For such sensors, several datasets and attributes are also mentioned. Then we discovered several AI-based models that were put out by various authors and included all of the relevant deep learning and machine learning models. It addressed a variety of relevant works utilizing AI approaches as well as upcoming research in that area. The difficulties and applications are also covered along with this. The paper provided an explanation of Machine Learning Techniques that can be used in HAR for further improvement. Moreover, in-depth information on Wearable sensors in HAR. The below-given graph in Fig. 2 is the comparative analysis of the accuracy of the various models and algorithms used for HAR until now.

Through the above comparison, the result comes out that the deep learning neural network models provide 96% accuracy using input samples as a dataset [24], and the CNN model gives out 96% accuracy on average with a dataset from wearable sensors [26], the hybrid DNN model gives 95% of accuracy on the basis of PAMP dataset [22], the state of the art method is 97.93% accurate using the UCI HAR data [23], LSTM model result only 85% accuracy which is the least useddata from gyroscope [20], the CNN model with LSTM provides 91.73% results with UniMib Shar data [21] whereas the long-short-term memory method gives with 94% for the UniMib SHAR as data [25]. Thus, on average, all the proposed methods provide are 93% accurate on an average and among them the State-of-Arts method is the most
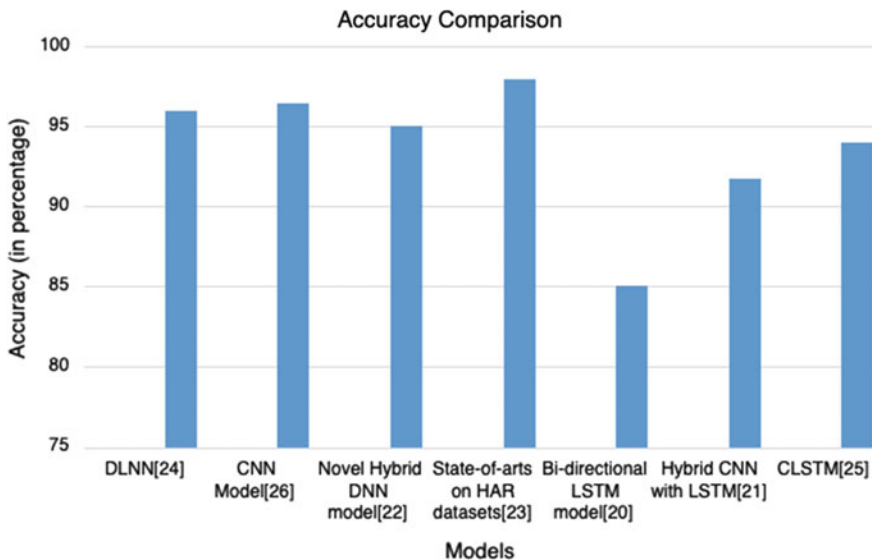
Fig. 2 Graph depictingthe accuracy of different models used with HAR

accurate method proposed for now. Thus, by examining all of these results, we intend to propose a classifier that will significantly improve on categorizing the action with much more accurate results by utilizing machine learning in wearable sensors.

6   Conclusion and Future Scope

Based on the information offered thus far, it appears reasonable to explore HAR as an important topic. Understanding human behaviors is a prominent research focus in computer vision. Different techniques employ a wide range of data formats and volumes, with the choice relying on how well the underlying algorithm manages varied and/or comprehensive types of data. Background noise, intricate camera movement, intraclass variation, and data collection issues make recognizing human activity more difficult. Many individuals have worked on constructing real-time intelligent surveillance systems, but the processing speed of video frames isn't up to standard, and there isn't a system that has a detection accuracy of 100% and a false detection rate of 0% for movies with complex backdrops. Because of computational restrictions, the majority of studies on this topic are unable to characterize human actions. There is still a great amount of difficulty in gathering and annotating data on human activity. Despite breakthroughs in human understanding methodologies, a number of challenges remain, including modeling human positions, coping with occlusions, and data annotation. The primary focus of research in the future should be on identifying characteristics that are unique to human activity. The extraction of low-level data to describe human motions is a complex process. A mechanism should be devised that would simulate human movement in a way that was not dependent on the subject's size or viewpoint. This would need to be done for all possible combinations of human positions.

References

1. Yang Z, Metallinou A, Narayanan S (2014) Analysis and predictive modeling of body language behavior in dyadic interactions from multimodal interlocutor cues. IEEE Trans Multimedia 16:1766–1778
2. Qiu S et al (2022) Multi-sensor information fusion based on machine learning for real applications in human activity recognition: state-of-the-art and research challenges. Inf Fusion 80:241–265
3. Chen K, Zhang D, Yao L, Guo B, Yu Z, Liu Y (2021) Deep learning for sensor-based human activity recognition. ACM Comput Surv 54(4):1–40
4. Jiang S, Kang P, Song X, Lo B, Shull PB (2022) Emerging wearable interfaces and algorithms for hand gesture recognition: a survey. IEEE Rev Biomed Eng 15:85–102
5. Wan S, Qi L, Xu X, Tong C, Gu Z (2020) Deep learning models for real-time human activity recognition with smartphones. Mobile Netw Appl 25(2):743–755
6. Li F, Shirahama K, Nisar M, Köping L, Grzegorzek M (2018) Comparison of feature learning methods for human activity recognition using wearable sensors. Sensors 18(3)

7. Chirra VRR, ReddyUyyala S, Kolli VKK (2019) Deep CNN: a machine learning approach for driver drowsiness detection based on eye state. Rev. d'Intell Artif 33(6):461–466

8. Jabbar R, Al-Khalifa K, Kharbeche M, Alhajyaseen W, Jafari M, Jiang S (2018) Real-time driver drowsiness detection for an android application using deep neural networks techniques. Procedia Comput Sci 130:400–407

9. Alshbatat AIN, Alhameli S, Almazrouei S, Alhameli S, Almarar W (2020) Automated vision based surveillance system to detect drowning incidents in swimming pools. In: 2020 Advances in Science and Engineering Technology International Conferences (ASET), pp 1–5

10. Verma KK, Singh BM, Dixit A (2019) A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system. Int J Inf Technol 1–14

11. Yang M, Rajasegarar S, Erfani SM, Leckie C (2019) Deep learning and one-class SVM based anomalous crowd detection. In: 2019 international joint conference on neural networks (IJCNN), pp 1–8

12. Whiteside D, Cant O, Connolly M, Reid M (2017) Monitoring hitting load in tennis using inertial sensors and machine learning. Int J Sports Physiol Perform 12(9):1212–1217

13. Mlakar M, Luštrek M (2017) Analyzing tennis game through sensor data with machine learning and multi-objective optimization. In: Proceedings of the 2017 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2017 ACM international symposium on wearable computers, pp 153–156

14. Reno V, Mosca N, Marani R, Nitti M, D'Orazio T, Stella E (2018) Convolutional neural networks based ball detection in tennis games. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 1758–1764

15. Ren L, Peng Y (2019) Research of fall detection and fall prevention technologies: a systematic review. IEEE Access 7:77702–77722

16. Vallabh P, Malekian R (2018) Fall detection monitoring systems: a comprehensive review. J Ambient Intell Humaniz Comput 9(6):1809–1833

17. Bet P, Castro PC, Ponti MA (2019) Fall detection and fall risk assessment in older person using wearable sensors: a systematic review. Int J Med Inform 130:103946

18. Nooruddin S, Islam MM, Sharna FA et al (2021) Sensor-based fall detection systems: a review. J Ambient Intell Human Comput 1–17

19. Sunny JT, George SM, Kizhakkethottam JJ, Sunny JT, George SM, Kizhakkethottam JJ (2015) Applications and challenges of human activity recognition using sensors in a smart environment. IJIRST Int J Innov Res Sci Technol 2:50–57

20. Ullah M, Ullah H, Khan SD, Cheikh FA (2019) Stacked LSTM network for human activity recognition using smartphone data. In: 8th European workshop on visual information processing (EUVIP). IEEE, pp 175–180

21. Xu C, Chai D, He J, Zhang X, Duan S (2019) InnoHAR: a deep neural network for complex human activity recognition. IEEE Access 7:9893–9902

22. Dua N, Singh SN, Semwal VB (2021) Multi-input CNN-GRU based human activity recognition using wearable sensors. Computing 103(7):1461–1478

23. Teng Q, Wang K, Zhang L, He J (2020) The layer-wise training convolutional neural networks using local loss for sensor-based human activity recognition. IEEE Sens J 20(13):7265–7274

24. Tan TH, Wu JY, Liu SH, Gochoo M (2022) Human activity recognition using an ensemble learning algorithm with smartphone sensor data. Electronics 11(3):322

25. Li F, Shirahama K, Nisar MA, Köping L, Grzegorzek M (2018) Comparison of feature learning methods for human activity recognition using wearable sensors. Sensors 18(2):679

26. Gupta S (2021) Deep learning based human activity recognition (HAR) using wearable sensor data. Int J Inform Manage Data Insights 1(2):100046

27. Challa SK, Kumar A, Semwal VB (2021) A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data. Visual Computer 1–15

28. VanHees VT, Gorzelniak L, Dean León EC, Eder M, Pias M, Taherian S et al (2013) Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PloS One 8(4):e61691

29. Mekruksavanich S, Jitpattanakul A (2021) Biometric user identification based on human activity recognition using wearable sensors: an experiment using deep learning models. Electronics 10(3):308
30. Zhang S, Li Y, Zhang S, Shahabi F, Xia S, Deng Y, Alshurafa N (2022) Deep learning in human activity recognition with wearable sensors: a review on advances. Sensors 22(4):1476
31. Mehta K, Sood VM, Singh C, Chabra P (2022) Machine learning based intelligent system for safeguarding specially abled people. In: 2022 7th international conference on communication and electronics systems (ICCES), pp 1199–1206. https://doi.org/10.1109/ICCES54183.2022.9835773
32. Rai V et al (2022) Cloud computing in healthcare industries: opportunities and challenges. In: Singh PK, Singh Y, Chhabra JK, Illés Z, Verma C (eds) Recent innovations in computing. Lecture notes in electrical engineering, vol 855. Springer, Singapore. https://doi.org/10.1007/978-981-16-8892-8_53
33. Mohindru V, Vashishth S, Bathija D (2022) Internet of Things (IoT) for healthcare systems: a comprehensive survey. In: Singh PK, Singh Y, Kolekar MH, Kar AK, Gonçalves PJS (eds) Recent innovations in computing. Lecture notes in electrical engineering, vol 832. Springer, Singapore. https://doi.org/10.1007/978-981-16-8248-3_18
34. Mohindru V, Singh Y, Bhatt R, Gupta AK (eds) (2021) Unmanned aerial vehicles for internet of things (IoT): concepts, techniques, and applications. Wiley
35. Mohindru V, Sharma A, Mathur A, Gupta AK (2021) Brain segmentation using deep neural networks. Int J Sens Wirel Commun Control 11(1):81–88
36. Mohindru V, Singla S (2021) A review of anomaly detection techniques using computer vision. In: Singh PK, Singh Y, Kolekar MH, Kar AK, Chhabra JK, Sen A (eds) Recent innovations in computing. ICRIC 2020. Lecture notes in electrical engineering, vol 701. Springer, Singapore. https://doi.org/10.1007/978-981-15-8297-4_53
37. Mohindru V, Bhatt R, Singh Y (2019) Reauthentication scheme for mobile wireless sensor networks. Sustain Comput Inform Syst 23:158–166
38. Mohindru V, Singh Y (2018) Node authentication algorithm for securing static wireless sensor networks from node clone attack. Int J Inf Comput Secur 10(2–3):129–148

A Non-conventional Patch-Based Shared Aperture Antenna with Parasitic Element for L/S Band

Jonny Dhiman and Sunil Kumar Khah

Abstract Advanced communication systems require antennas utilising the common aperture to provide multi-functions in different frequency band. We report design of antenna elements using shared aperture for L and S bands with a frequency ratio of 1.6:3.9. The performance of the design is enhanced by length optimised parasitic strip of L shape etched on the corner of the antenna surface. The antenna configuration shows good corroboration between the simulated and measured results. The designed antennas exhibit a return loss of less than -25 dB and VSWR less than 1.1 for respective resonant frequencies.

Keywords Microstrip patch antenna · Parasitic patch · Return loss · VSWR

1 Introduction

Requirement of multifunctionality of antenna system in communication is key issue for performing multiple tasks simultaneously. Sharing of time or aperture is the traditional way of achieving these properties in a system [1]. Modern communication systems require compact multifunctional antennas with dual band or multiband frequency operational capabilities. Miniaturisation of the systems is associated with the space constraint. This constraint is overcome by shared aperture technology [3–5]. Shared aperture method is one of the best approaches, where multi-antennas uses identical aperture for communication devoid of band interference.

Multiple efforts by different researchers have led to the development of shared aperture technology. Most of the designers have designed the shared aperture antennas for the frequency bands having operating frequency ratio of greater than 1:3. In all the previous works [6, 7], the wide frequency ratio did provide the spacing feasibility of the size, but in case of L and S bands the design provides us the narrow

J. Dhiman (✉) · S. K. Khah

Electromagnetic Analysis Lab, Department of Physics and Materials Science, Jaypee University of Information Technology, Waknaghat, Distt. Solan, HP 173234, India
e-mail: jonny.dhiman@gmail.com

577

frequency ratio of 1:2, which means S band elements will be spaced at about ½ the spacing of L band. For this consideration, L shape antenna is designed with non-conventional shape antenna (swastika). The aperture around the swastika is used for another antenna. Some researchers have worked on continuous bands [6], in all those works the antenna designed is a stacked element or multilayer structure, which is difficult to design. In the present design, same surface is used for designing of both elements. The surface utilisation and location of antenna on the aperture is the key for the present design.

This communication presents a shared aperture antenna design for L and S bands. The design utilised the aperture of the antenna for designing both the antennas without cross talk. Both antennas work almost independent of one another. The proposed antenna with a frequency ratio of nearly 1:2 is best suited for on-board communication systems, SAR applications. The work consists of design of the antenna followed by simulation and experimental results to verify the design.

2 Antenna Design

The key element of the design is the choice of radiating elements. For a system to work in nearly 1:2 frequency ratio on the same aperture, the choice of radiating element becomes more important. For the present design, a basic antenna of swastika shape is chosen to leave the periphery of the aperture for being used as a base for another antenna. The swastika was designed using basic square patch. Following [8], the dimensions of the square patch antenna are calculated. The square shape antenna is converted to swastika by creating the four notches along the perimeter of the patch keeping one notch on each length. The effective length of the patch is given as

$$L_{\text{eff}} = L - 2\Delta L_{\text{eff}},$$

where

L = Length or width of the patch = $\frac{c}{2 f_0 \sqrt{\varepsilon_e}}$

$$\Delta L_{\text{eff}} = 0.412 h \frac{(\varepsilon_e + 0.3)\left(\frac{W}{h} + 0.264\right)}{(\varepsilon_e - 0.258)\left(\frac{W}{h} + 0.8\right)}$$

$$\varepsilon_e = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r + 1}{2}\left[1 + 12\frac{h}{W}\right]^{-\frac{1}{2}}$$

The depth of the notches was optimised using the simulating the antenna design stepwise in multiples of quarter wavelength. The best possible design was for the L band resonating at 1.64 GHz. For designing the antenna for S band design, following [8] two microstrip lines were placed at top-right corner of the patch in the form of inverse L or shape 7, which resonates in S band.

The geometry of the present design is given in Fig. 1 with designed parameters listed in Table 1. The antenna is designed on the square surface of length (L) and width (W) for L and S bands. The L band antenna requires more area, so the basic design of L band is made from notching the square to form swastika. A circle of diameter 3 mm is designed in every quarter of the swastika. For S band, the periphery is used by joining two strips of same rectangular dimensions at top-right corner of the swastika in a shape like number 7. The 7-shape antenna is having two arms of same length L_t and L_r. On the opposite corner, a parasitic patch is designed in L shape with equal length arm L_1 and L_b, for the basic design all four lengths are equal. The parasitic patch near the radiating element enhances the properties and provides a wide band [9]. The same is observed in L band antenna. With increase in the parasitic length, the bandwidth and return loss of the antenna are enhanced. The same effect is not visible in S band antenna. Using the parasitic effects to optimise the antenna for enhanced properties, a variable length parameter (L_v) is incorporated in design. This parameter increases the length of both arms of parasitic patch. The length (L_v) was optimised for the best results, and in the present communication, two cases with different L_v are experimentally reported. For the present design, width or gap (a) of all the patches is taken as 10 mm.

The numeric values of the prototype antenna are presented in Table 2. Three different prototype antenna systems were fabricated (i) without parasitic patch, (ii) with parasitic patch of length 113.84 mm and (iii) with parasitic patch of length 117.84 mm and are shown in Figs. 2 and 3, respectively. The length parameter of the parasitic patch was optimised by simulating the design with microwave studio with different lengths of the parasitic patch. The design of the antenna was carried out for the two best cases of the simulated results. In the present case, the best case was obtained for 113.84 and 117.8 mm.
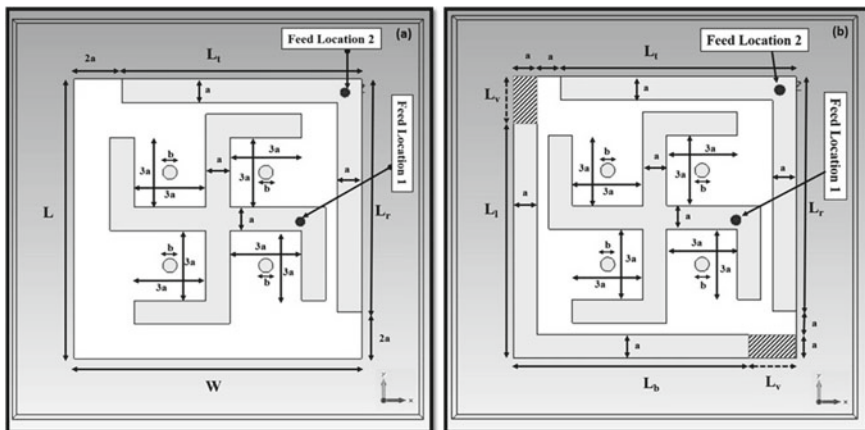
Fig. 1 a, b Geometry of microstrip antenna with swastika shape and 7-shaped patch antenna (without and with parasitic patch)

Table 1 Description and numeric design values of the antenna system

S. No.	Variable	Description	Design values (in mm)
1.	L	Length of the aperture	120
2.	W	Width of the aperture	120
3.	L_t	Length of top side of the 7-shaped microstrip patch	100
4.	L_r	Length of right side of the 7-shaped microstrip patch	100
5.	L_b	Length of bottom side of the parasitic patch	100
6.	L_l	Length of left side of the parasitic patch	100
7.	L_v	Variable length parameter of the parasitic patch	13.84 and 17.8
8.	a	Width or gap	10
9.	b	Diameter of the circle	3

3 Experimental Results

Three different antennas were designed for both L and S bands with and without parasitic patch. Two different lengths of parasitic patches were taken for fabrication. The simulated and measured results are reported in Table 2a for L band (swastika shape antenna) and in Table 2b for S band (7-shape antenna).

In the absence of parasitic patch, both the antennas report return loss value of around −30 dB. The same is increased to be greater than −40 dB in presence of parasitic patch on the aperture. The effect of parasitic patch near the radiating element of L band enhances the bandwidth and return loss of the swastika antenna of L band. Other significant antenna characteristics are also calculated and reported in the table. Voltage standing wave ratio (VSWR) value is approximately identical in all cases (i.e. ≈1). In addition, cross-polarisation and normalised impedance values are also reported. Directivity of more than 6 dBi is reported in all cases. Return loss values for without and with parasitic patch are plotted and shown in Fig. 4a–c, and the coupling between the radiating elements is minimum as reported by insertion loss, which is same for both antennas, the insertion loss is shown in Fig. 5.

Table 2 **a** Simulated and measured results for L band (Swastika shape antenna). **b** Simulated and measured results for S band (7-shape antenna)

(a)

Swastika-shaped patch antenna

Parameters versus configuration		Without coupling	With coupling	
			Length = 113.84 mm	Length = 117.94 mm
Resonant frequency (GHz)	Simulated	1.64	1.645	1.645
	Measured	1.629	1.574	1.585
Return loss (dB)	Simulated	−28.41	−34.48	−41.55
	Measured	−29.75	−49.8	−39.25
10 dB bandwidth (%)	Simulated	2.66	2.562	2.51
	Measured	5.4	4.19	8.075
Cross-polarization (dB)		−32.2	−33.3	−37.3
VSWR		1.07:1	1.03:1	1.01:1
Normalized impedance (R±jX)		1.0854 − 0.042	1.1060 + 0.015	1.081 + 0.020
Directivity (dBi)		6.446	6.45	6.419

(b)

Seven (7)-shaped patch antenna

Parameters versus configuration		Without coupling	With coupling	
			Length = 113.84 mm	Length = 117.94 mm
Resonant frequency (GHz)	Simulated	3.93	3.94	3.94
	Measured	3.265	3.87	3.909
Return loss (dB)	Simulated	−28.9	−40.45	−32.79
	Measured	38.15	−37.83	−52.6 10 dB
Bandwidth (%)	Simulated	1.787	1.7415	1.727
	Measured	3.307	2.686	2.58
Cross-polarization (dB)		−35	−29.7	−26
VSWR		1.07:1	1.01:1	1.046:1
Normalized impedance (R±jX)		1.074 + 0.0163	1.023 − 0.008	1.052 − 0.006
Directivity (dBi)		7.17	7.155	7.166
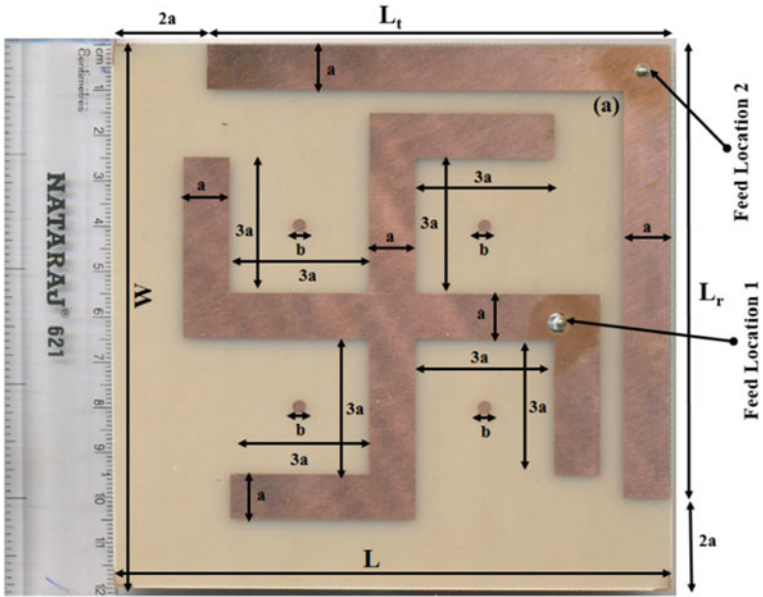
Fig. 2 Antenna system without parasitic patch
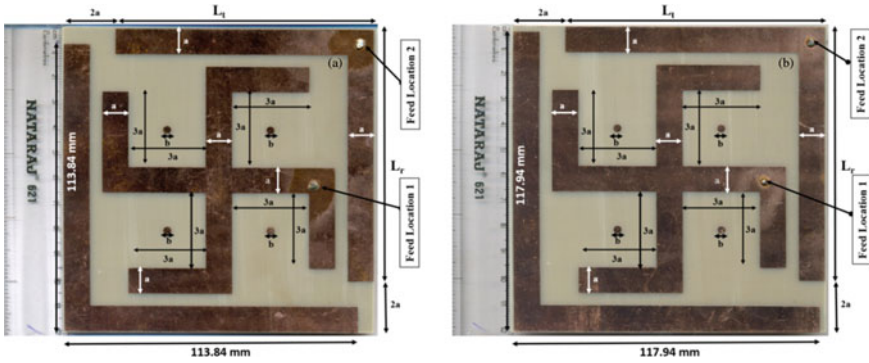
Fig. 3 a, **b** Antenna system with parasitic patch ($L_v = 113.84$ and 117.94 mm)
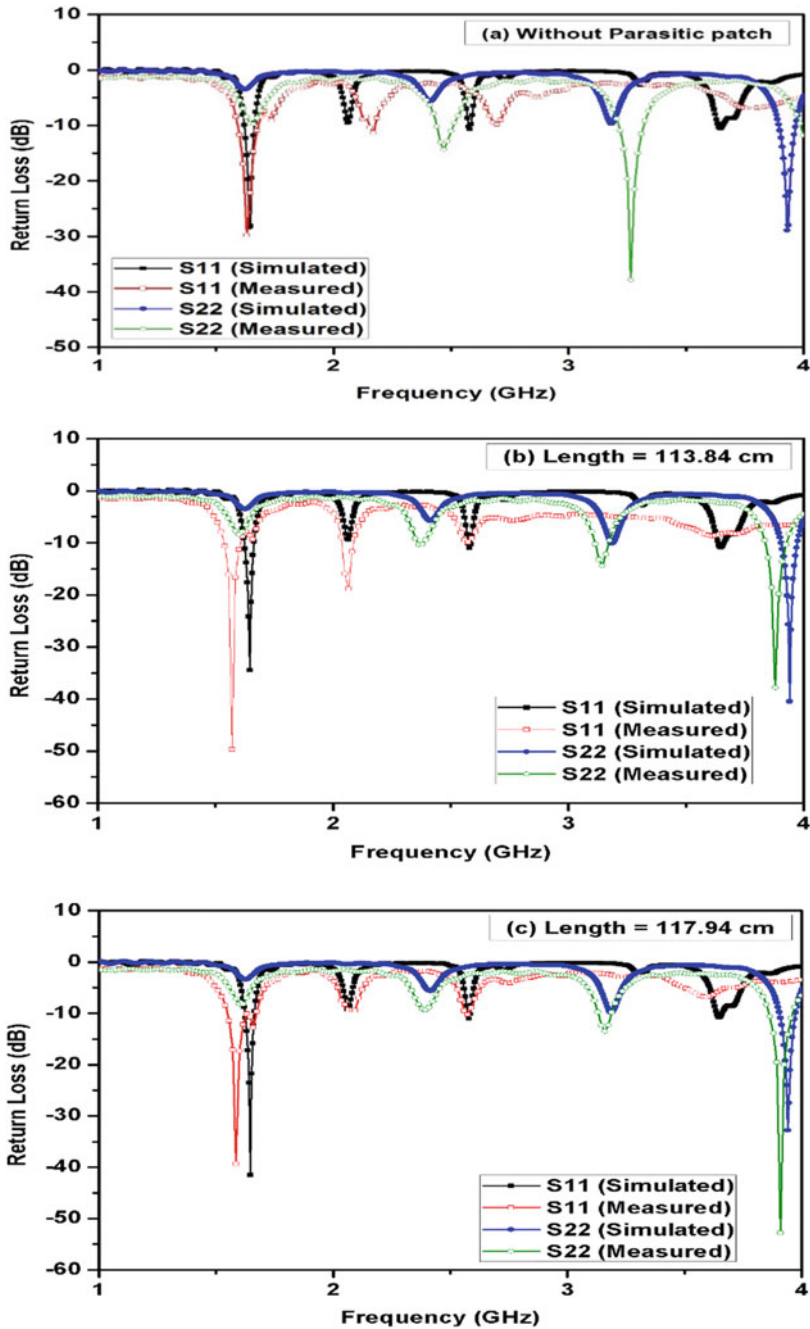
Fig. 4 **a** Return loss of shared aperture antenna without parasitic patch. **b** Return loss of shared aperture antenna with parasitic patch with length of 113.84 mm. **c** Return loss of shared aperture antenna with parasitic patch with length of 117.94 mm
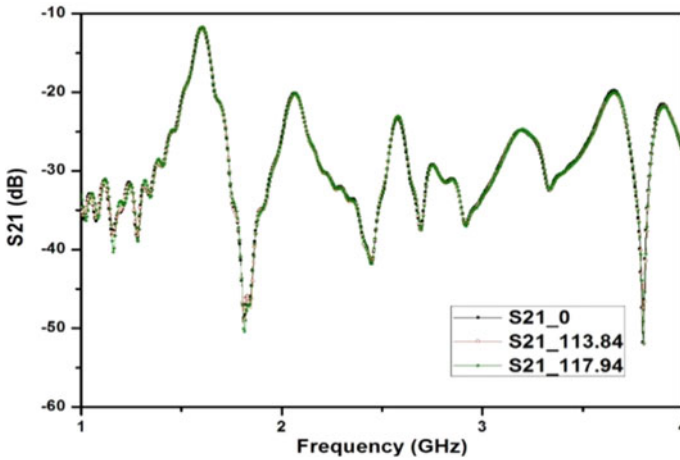
Fig. 5 S21 parameter for without and with parasitic patch

4 Conclusion

Combination of non-conventional patches is used to design shared aperture antenna in L and S bands. For enhancing the return loss and bandwidth, a parasitic patch is introduced near one radiating patch. The simulated and the experimental results are in agreement. Return loss of around -50 dB and four times increase in10 dB bandwidth is achieved in one of the experimental cases. Reduction in cross-polarisation was also observed, however other parameters remain almost same. The present designs will provide a window for the designers to design the antenna systems on mobile platforms performing simultaneously multiple tasks, such as communication, remote sensing, and electronic warfare. The proposed antenna system with a frequency separation around 1:2 is best suitable for the SAR applications at L and S bands.

References

1. Coman CI, Lager IE, Ligthart LP (2006) The design of shared aperture antennas consisting of differently sized elements. IEEE Trans Antennas Propag 54(2):376–383
2. Axness TA, Coffman RV, Kopp BA, O'Haver KW (1996) Shared aperture technology development. J Hopkins APL Tech Dig 17(3):285–394
3. Pozar DM, Targonski SD (2001) A shared-aperture dual-band dual-polarized microstrip array. IEEE Trans Antennas Propag 49(2):150–157
4. Chakrabarti S (2011) Development of shared aperture dual polarized microstrip antenna at L-band. IEEE Trans Antennas Propag 59(91):294–297
5. Zhong S-S, Sun Z, Kong L-B, Gao C, Wang W, Jin M-P (2012) Tri-band dual-polarization shared-aperture microstrip array for SAR applications. IEEE Trans Antennas Propag 60(9):4157–4165
6. Sharma DK, Kulshrestha S, Chakrabarty SB, Jyoti R (2013) Shared aperture dual band dual polarization microstrip patch antenna. Microwave Opt Lett 55(4):917–922

7. Zhou S-G, Yang J-J, Chio T-H (2015) Design of L/X-band shared aperture antenna array for SAR application. Microwave Opt Technol Lett 57(9):2197–2204
8. Garg R, Bhartia P, Bahl I, Ittipiboon A (2001) Microstrip antenna design handbook. Artech House
9. Au TM, Luk KM (1991) Effect of parasitic element on the characteristics of Microstrip Antenna. IEEE Trans Antennas Propag 39(8):1247–1251

Multi-modality Medical Image Fusion Employing Various Contrast Enhancement Techniques

Shruti Jain , Bandana Pal, and Rittwik Sood

Abstract Clinical imaging has a major role in healthcare applications. The blur and the noise of the picture are eliminated, which improves the contrast and provides information about the image. But to increase the precision of the image, it is first improved. The tasks of improving medical pictures are to intensify an area of concern. Because of such issues, image processing techniques like enhancement and fusion algorithms aid human visuals in making unbiased judgments of brain images using CT and MRI. The main aim of this research paper is to exploit the benefits of spatial enhancement methods, namely binarization, median filter, contrast stretching, and contrast limited adaptive histogram equalization (CLAHE). This helps in attaining greater quality images. The enhanced images are followed by the fusion of images using principal component analysis, discrete wavelet transform, and Laplacian re-decomposition (LRU) fusion methodology. The comparative analysis is done based on qualitative and quantitative parameters and found that the results using CLAHE and LRU outperform other techniques. Later, the model is validated on the available datasets and the results are in consensus.

Keywords Medical images · Enhancement techniques · Fusion · Quantitative analysis

S. Jain
Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh, India

B. Pal (✉)
Indira Gandhi Delhi Technical University for Women, Delhi, India
e-mail: bandanapal707@gmail.com

R. Sood
NIT Hamirpur, Hamirpur, Himachal Pradesh, India

1   Introduction

Medical imaging is a challenging region of study. In the present scenario, medical science, health treatment, computer-assisted diagnostics, and rehabilitation are heavily dependent on image processing techniques [1, 2]. The imageries of the concerned area of the human body are seen on the computer of medical practitioners for the diagnosis of diseases. Imaging throughout the medical industry allows surgeons to quickly examine the inner area of internal organs and enable accurate operations to access the concerned part [3, 4]. Imaging is essential to the preparation and control of interventional therapies and biopsies. Nowadays, most doctors face the challenge of diagnosing. This situation happens when more than one modality of illness needs to be diagnosed [5, 6]. Also, reasons like acquisition devices and inferior lighting conditions degrade the quality of images. This makes it difficult for an analyzer to accurately observe the internal structures of the human body [7–9]. With the evolution of technology, the assistance of hi-tech computers in the medical field depends occasionally on image fusion methods. Especially, in the medical field, the detection and diagnosis of a disease with a single image can be tedious and difficult for doctors. Therefore, for precise examination, physicians must combine multiple images of the concerned patients to aid in improved visualization of complex parts of the human body leading to cuts above surgeries and therapies [10, 11]. Also, exploration of the area of image fusion paves the path for additional improvements which can be considered one of the reasons to interest many researchers [12–14]. Image fusion aids in providing information that is more suitable for visual perception. Image fusion refers to the process of fusing two or more images into one fused image that has higher information in comparison with any of the input images [15]. Many fusion algorithms have been established in the past few decades. There are different types or levels of data depiction such as pixel level or low level, feature level or middle level, and high-level or decision level [16, 17]. Out of the three, pixel level is preferred over the others as it can mechanize in the spatial domain and transform domain as shown in Fig. 1. They emphasize working directly on the pixels of desired images.

In [18], authors have discussed different potential techniques to increase the accuracy of underwater images. A hybrid solution incorporating DCP and CLAHE methodologies for improved picture quality is presented. It has been found that the DCP dims the picture which leads to a low-contrast picture. So, to improve the contrast of the scene, CLAHE is used. In paper [19], low contrast, noise, and blurred images are improved using different filtering and contrast-enhancing techniques. It has been found that the median filter was less effective in comparison with the Wiener filter for denoising of images. Such approaches allow physicians and radiologists to diagnose accurately. In [20], authors have proposed an algorithm that enhances fingerprint images using binarization and thresholding techniques. In [21], authors carried out comparative analyses in three cases: linear contrast approaches, nonlinear contrast methods, and linear and nonlinear contrast methods which were concluded to be best for the analysis of remote sensing images. In [22], authors implemented four types of contrast stretching methods, namely modified global CS (MGCS), global CS
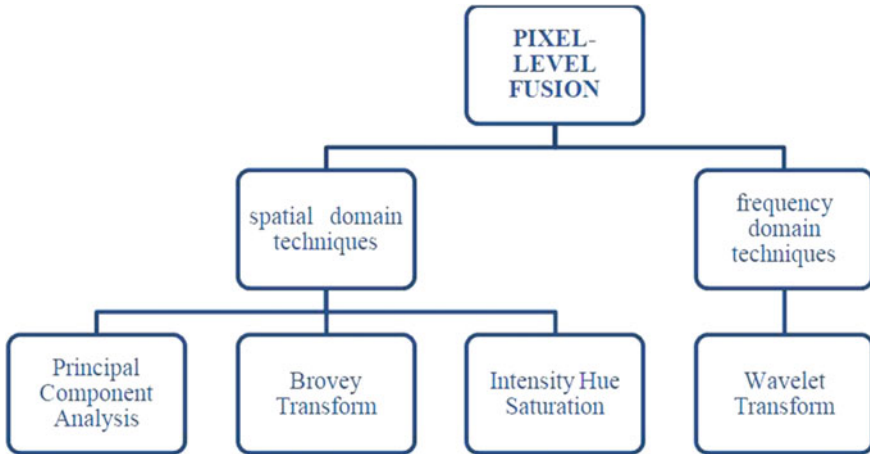
Fig. 1 Different pixel-level data fusion techniques

(GCS), linear CS (LCS), and modified linear CS (MLCS) methods on 150 pictures of malaria. It has been found that MLCS and MGCS result better in improving the images (contrast).

The fundamental aim of this paper is to analyze and examine different and recently introduced enhancements that will aid in contrast enhancement and noise suppression in brain images. In this paper, a novel image fusion technique for magnetic resonance image (MRI) and computed tomography (CT) brain images has been implemented which provides better quality, decreases vagueness, and minimizes the redundancy of fused images. Without fusion, images are distorted and unsharp due to inconsistent environmental conditions. With the proposed fused technique, quantity of acquired information has been enhanced. The need for improved diagnostic imaging has been raised to enable the experimenter to make an appropriate diagnosis.

In this paper, Sect. 2 explains the methodology; Sect. 3 explains the results of the different enhancement and fusion techniques and discusses the various evaluation parameters. Section 4 explains the concluding remarks followed by future work.

2 Methodology

In this paper, authors have used different enhancement techniques for enhancing the images and fusion methods which helps in getting a better quality of images. The proposed methodology used in this paper is shown in Fig. 2.

A benchmarked database is taken from [23] which provides a wide range of neuro-imaging anatomies. When the brain is imaged, it is viewed in two-dimensional (2D) slices. The database comprises MRI and CT images of both the left and right sides of the brain. Figure 3 shows four different sets of images from the dataset.
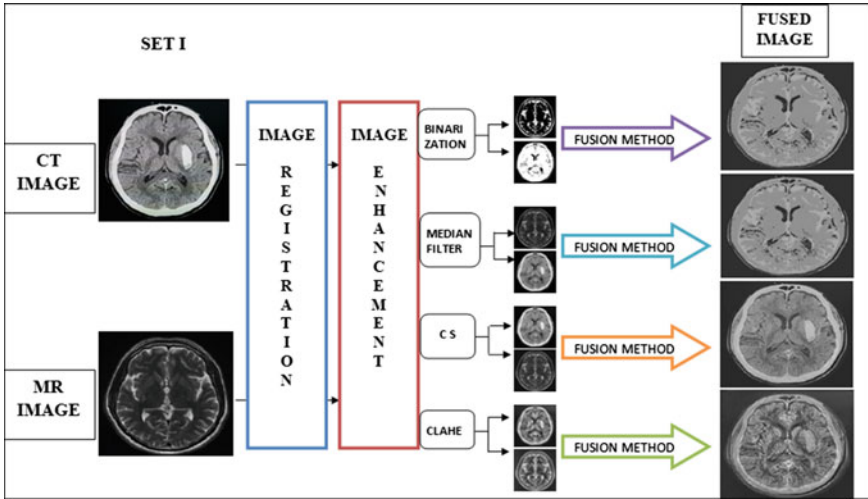
Fig. 2 Proposed methodology for image fusion
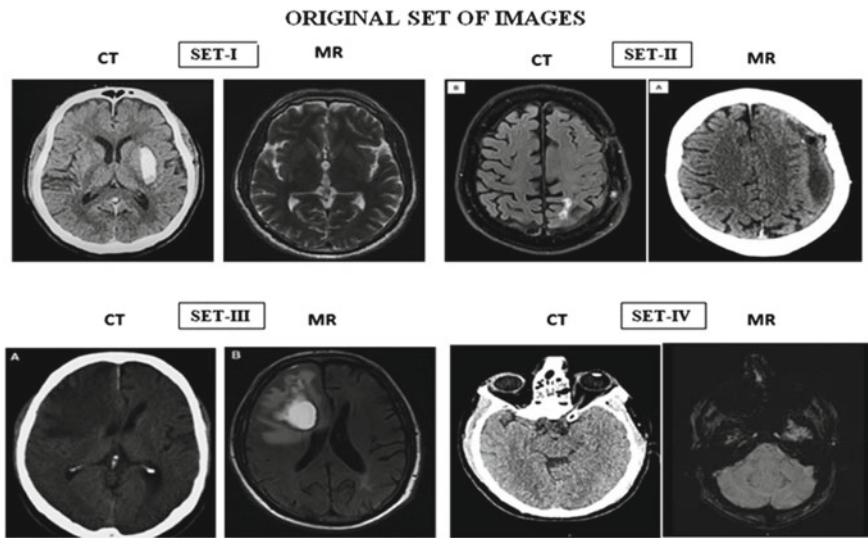
Fig. 3 Original sets of CT-MR images [23]

In Fig. 3, set I shows the brain image, set II shows the head region, set III shows the left frontal lobe, and set IV shows a small hemorrhagic region. The images are of equal dimensions. The different enhancement techniques were explored which improve the interpretation of images for better quality, minimize the impulsive noise, and sharpen the edges. Some of the widely used methods are the histogram-based nonlinear transfer function and frequency domain. To comprehend the enhancement algorithms

deeply, binarization (Bz), median filter (MF), contrast stretching (CS), and contrast-limited adaptive histogram equalization (CLAHE) methods are used [24, 25]. Image binarization reduces the information of an image from 256 to 2. This technique is mostly used if the object is to be extracted from an image. For color normalization and color illumination, median filters are used. CS is a simple image enhancement technique that is commonly known as normalization. It improves the contrast of the image using stretching. CLAHE is a modified adaptive histogram equalization in which the contrast amplification is limited to reduce the noise amplification problem.

There are different types of fusion techniques. But in this paper, *principal component analysis (PCA)* [26], *discrete wavelet transform (DWT)* [27, 28], *and Laplacian re-decomposition method (LRD)* methods are used. PCA is the fundamental concept to minimize numerous sets of data comprising several variables. It uses transformations based on an orthogonal component resulting in a set of linear values obtained by transforming a group of opinions from conceivably related variables [29, 30]. This transformation is commonly used for image encoding, random noise signal, and image rotation and elimination. Of the standard fusion algorithm, PCA has several components of choice. The algorithm for the implementation of PCA is shown in Fig. 4.

DWT is a spatial domain approach that produces distortion in the merged image. The discrete wavelet transformation has been a very useful fusion technique. Images used in the fusion of images should be recorded [31]. Mis-registration is a primary error source for image fusion. DWT is a two-channel and multi-scaled transformation method.

The conventional LRD method never illustrates the structural details of the source pictures. The core principle of regional Laplacian filtering is to use edge preservation and filters for sharpening the source images. LRD is an advanced and high-frequency technique that uses a simpler mapping feature which makes image gradient information better to emphasize the complex structures and suppress noise. In this method, the mechanism of identifying unnecessary and complementary information to minimize mutual interference is implemented. Thus, pictures of overlap and non-overlapping
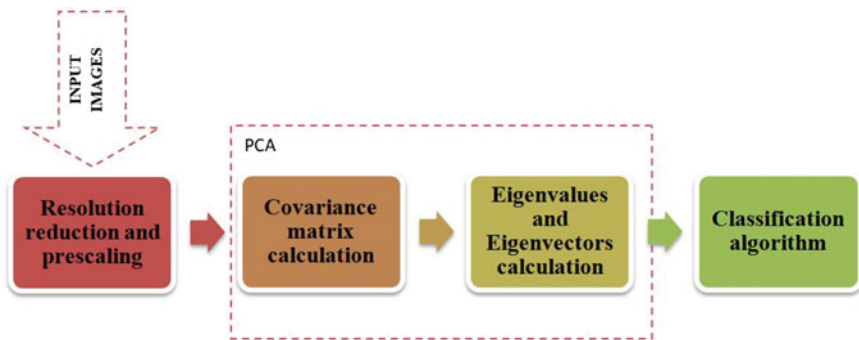
Fig. 4 Block diagram of PCA technique

domains are also acquired through this decomposition. Further, an inverse LRD fusion rule is adopted which recreates high-frequency images. Based on performance evaluation metrics, results are compared. Algorithm 1 shows the proposed steps used in this paper.

Algorithm 1: Algorithm for implementation of fusion of two images

I/P image: CT and MR images
O/P image: enhanced/fused image
Step 1—Select the medical images followed by conversion to grayscale
Step 2—Resizing of images
Step 3—Images are processed by the enhancement algorithm
Step 4—Resultant images are evaluated using *brisque*
Step 5—Image fusion is used by employing different transform techniques
Step 6—Evaluation of performance metrics

3   Results and Discussion

The implementation of the two modalities of brain images, namely MRI and CT, is considered in this paper as shown in Fig. 5 which was made to have equal dimensions.

After the image registration step, each enhancement methodology, i.e., Bz, MF, CS, and CLAHE, is implemented on the data images. Figures 6 and 7 show the comparative analysis of CT and MR images, respectively, based on brisque.

Figure 6 focuses on the experimental values of the CT images and indicates that the values, e.g., in set III, the brisque value for the source CT image is 32.94 which is less than 51.68 obtained in binarization which implies that this method in some cases might be beneficial but for cases, where the background and the region of areas' pixels are imbalanced, will instead produce poor results. In the case of the median filter, the value is 33.45 which is larger, and contrast stretching with a numeric value of 31.08 is surely lesser but in comparison with 23.96, the CLAHE method is identified to be more efficient. On assessing the individual CT-MR graphs, it is observed that CLAHE-enhanced images are furnishing smaller brisque values which justifies that

Fig. 5 Brain images **a** MR image, **b** CT image
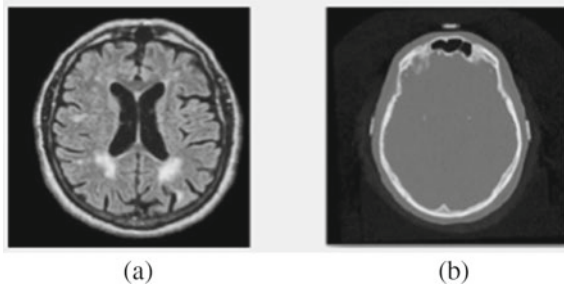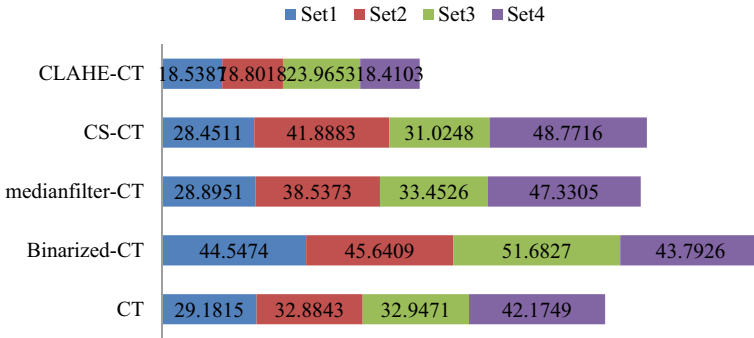
Fig. 6 Comparative analysis of CT images (b/w original image and enhanced image) based on brisque
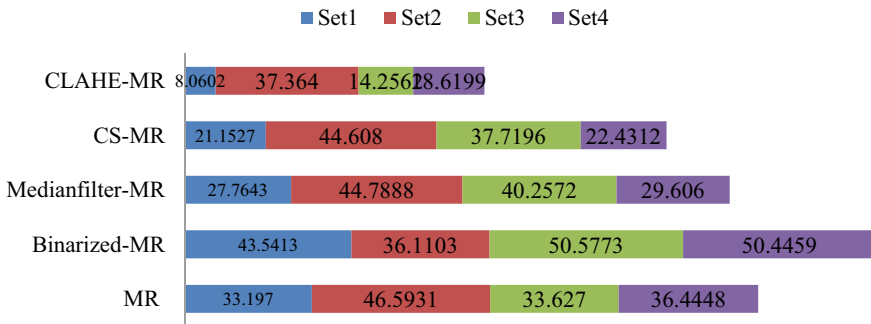
Fig. 7 Comparative analysis of MR images (b/w original image and enhanced image) based on brisque

the source images were initially having low perceptual quality. On the other hand, the values yielded by CS in some cases are performing equally as median filter and binarization. But the results obtained from the other three were not convincing in the sense that the values were not consistent. From the experimental values shown in Figs. 6 and 7, CLAHE outperforms the other methods. The enhanced images from CLAHE methods are further fused by PCA, DWT, and LRD fusion techniques. Table 1 presents the different performance evaluations of different fusion techniques.

In Table 1, I1 refers to the first input image, I2 refers to the second input image, and F refers to the resultant fused image. Considering, the complex yet efficient mechanism involved in LRD, it outperformed the other two fusion methods based on PSNR, SNR, and SSIM. On comparing the aforementioned fusion algorithms in Table 1, it is observed that better PSNR, SNR, and SSIM values are furnished by the LRD fusion technique. LRD could maintain more knowledge quantity in the fused outcome and outperform other common methods not only in the subjective visual perception but also in the quantitative perception. The analysis is further carried out to find a fusion technique that not only will suppress noise but preserve and

Table 1 Performance evaluation of different fusion techniques based on PSNR, SNR, and SSIM

Set no.	Fusion techniques	Performance metrics					
		PSNR		SNR		SSIM	
		I1 and F	I2 and F	I1 and F	I2 and F	I1 and F	I2 and F
Set 1	**PCA**	13.13	9.28	13.09	9.25	0.28	0.33
	DWT	16.98	16.97	16.95	16.94	0.49	0.31
	LRD	**20.85**	**19.51**	**20.82**	**19.48**	**0.62**	**0.54**
Set 2	**PCA**	13.92	7.65	13.89	7.62	0.23	0.16
	DWT	14.66	14.71	14.63	14.67	0.53	0.20
	LRD	**20.69**	**15.55**	**20.65**	**15.51**	**0.85**	**0.17**
Set 3	**PCA**	10.67	8.36	10.63	8.32	0.20	0.28
	DWT	16.98	17.02	16.94	16.98	0.50	0.45
	LRD	**23.82**	**17.05**	**23.78**	**17.01**	**0.81**	**0.58**
Set 4	**PCA**	13.33	11.36	13.29	11.32	0.40	0.29
	DWT	15.91	14.31	15.88	14.28	0.34	0.45
	LRD	**17.74**	**15.87**	**17.71**	**15.84**	**0.56**	**0.43**

enhance the already present details. Therefore, two commonly used methods, PCA and DWT, are considered along with a newly introduced method called the Laplacian re-decomposition method. From Table 1, considering set I, it is observed that PCA performs poorly based on PSNR, SNR, and SSIM. Its corresponding values are the lowest, whereas in DWT, the values obtained are better than PCA but LRD values are much more satisfactory both visually and quantitatively.

Validation of Model

Authors have validated the model employing LRD and CLAHE on other datasets shown in [32–34]. The different evaluation parameters have been evaluated, and the results are given in Table 2.

Table 2 Validation of model

Set no.	Fusion techniques	Performance metrics					
		PSNR		SNR		SSIM	
		I1 and F	I2 and F	I1 and F	I2 and F	I1 and F	I2 and F
Set 1	[32]	14.15	8.29	11.0	10.45	0.22	0.30
	[33]	16.91	16.31	16.88	15.28	0.34	0.45
	[34]	18.98	17.02	16.94	16.98	0.50	0.49
	LRD + CLAHE	**20.85**	**19.51**	**20.82**	**19.48**	**0.62**	**0.54**

4 Conclusion and Future Work

In recent years, enormous efforts have been made to design image processing algorithms to enhance spatial resolution. A new method is presented in this paper for fusing multi-modal medical images utilizing PCA, DWT, LRD, and different enhancement techniques. Several pairs of medical images are considered and tested, and it is found that the proposed algorithm employing the LRD technique and CLAHE enhancement technique provides improved visual and quantitative results, with high contrast and luminance. PSNR, SSIM, and SNR values have been evaluated. The model is validated using the existing model, and it has been found that the proposed model is in consensus. In the future, authors will try hybrid fusion techniques to get better-quality of images.

References

1. Mustafa WA, Mohamedand MM, Kader A (2017) Contrast enhancement based on fusion method: a review. In: 1st international conference on green and sustainable computing (ICoGeS) 2017. IOP Publishing; IOP Conf Ser J Phys Conf Ser 1019:1–7. https://doi.org/10.1088/1742-6596/1019/1/012025

2. Maragatham G, Md. Mansoor Roomi S (2015) A review of image contrast enhancement methods and techniques. Res J Appl Sci Eng Technol 9(5):309–326. ISSN: 2040-7459; e-ISSN: 2040-7467©Maxwell Scientific Organization, pp 1–18

3. Habeeb NJ, Al-Taei A, Fadhil-Ibrahim M (2005) Contrast enhancement for multi-modality image fusion in spatial domain. J Theor Appl Inf Technol 96(20):6926–6936. Ongoing JATIT & LLS ISSN: 1992-8645

4. Gohiya S, Rawat P (2017) A method of medical image enhancement using wavelet based fusion and analysis. Int J Adv Res Sci Eng (IJARSE) 6(9):822–829. ISSN: 2319-8354

5. Seema, Bansal G (2017) A review of image contrast enhancement techniques. Int Res J Eng Technol (IRJET) 04(08):1139–1142. e-ISSN: 2395-0056, Aug 2017. www.irjet.net, p-ISSN: 2395-0072

6. Kaur N, Baghla S, Kumar S (2015) A review: image enhancement and its various techniques. Int J Adv Sci Eng Technol 3(3):141–144. ISSN: 2321-9009, July 2015

7. Ačkar H, Almisreb AA, Saleh MA (2019) A review on image enhancement techniques. Southeast Euro J Soft Comput 8(1):42–48

8. Mehera B, Agrawala S, Pandaa R, Abrahamb A (2019) A survey on region based image fusion methods. Inf Fusion 48:119–132, 120–125

9. Masood S, Sharif M, Yasmin M, Shahid MA, Reh A (2017) Image fusion methods: a survey. J Eng Sci Technol Rev 10(6):186–194, 187–191

10. Abass HK. A study of digital image fusion techniques based on contrast and correlation measures. Ministry of Higher EducationAnd Scientific Research Al-Mustansiriyah University College of Science/Department of Physics, pp 1–9

11. Yin K, Yu W (2008) Image fusion based on the wavelet transform. J Inf Comput Sci 5(3):1379–1385

12. Rani A, Kaur G (2010) Image enhancement using image fusion techniques. Int J Adv Res Comput Sci Softw Eng 3(2):413–414

13. Pal B, Mahajan S, Jain S (2020) Medical image fusion employing enhancement techniques. In: 2020 IEEE international women in engineering (WIE) conference on electrical and computer engineering (WIECON-ECE), Bangladesh, 26–27 Dec 2020

14. Bopche S, Gade A. Implementation of image fusion using DWT and PCA. Int J Future Revolution Comput Sci Commun Eng 4(2):507–509. ISSN: 2454-4248

15. Liua Y, Chena X, Wang Z, Jane Z, Rabab W, Wange KWX (2018) Deep learning for pixel-level image fusion: recent advances and future prospects. Inf Fusion 42:158–159

16. Kokate RM, Shandilya VK. Enhancement techniques of medical images: a review. Int J Adv Res Comput Commun Eng 5(2):533–535, Feb 2016 ISSN (Online) 2278-1021, ISSN (Print) 2319-5940. https://doi.org/10.17148/IJARCCE.2016.52120

17. Jain S, Chauhan DS (2019) Computer aided design and synthesis for marker proteins of HT carcinoma cells: a study. In: Paul S (ed) Biomedical engineering and its applications in healthcare. Springer, Singapore, pp 399–419

18. Malathi, Manikandan A (2019) An enhancement of underwater images using DCP and CLAHE algorithm. Int J Eng Adv Technol (IJEAT) 9(2):2805–2813. ISSN: 2249-8958

19. Pal B, Mahajan S, Jain S (2020) A comparative study of traditional image fusion techniques with a novel hybrid method. In: 2020 international conference on computational performance evaluation (ComPE), Shillong, India, pp 820–825. https://doi.org/10.1109/ComPE49325.2020.9200017

20. Shetter A, Prajwalasimha SN, Swapna H (2018) Finger print image enhancement using thresholding and binarization techniques. In: Proceedings of the 2nd international conference on inventive communication and computational technologies (ICICCT 2018), pp 899–901. IEEE Xplore Compliant—Part Number: CFP18BAC-ART; ISBN: 978-1-5386-1974-2

21. Al-Amri SS (2011) Contrast stretching enhancement in remote sensing image. BIOINFO Sensor Netw 1(1):1–4

22. Abdul-Nasir AS, Mashor MY, Mohamed Z (2012) Modified global and modified linear contrast stretching algorithms: new colour contrast enhancement techniques for microscopic analysis of malaria slide images. Hindawi Publishing Corporation; Comput Math Methods Med 2012:1–16. https://doi.org/10.1155/2012/637360

23. Departments of Radiology and Neurology at Brigham and Women's Hospital, Harvard Medical School, the Countway Library of Medicine, and the American Academy of Neurology. http://www.med.harvard.edu/aanlib/home.html

24. Makandar, Halalli B (2015) Breast cancer image enhancement using median filter and CLAHE. Int J Sci Eng Res 6(4):462–465. ISSN 2229-5518

25. Jain S, Sachdeva M, Dubey P, Vijan A (2019) Multi-sensor image fusion using intensity hue saturation technique. In: Luhach A, Jat D, Hawari K, Gao XZ, Lingras P (eds) Advanced informatics for computing research. ICAICR 2019. Communications in computer and information science, vol 1076. Springer, Singapore, pp 147–157

26. Salau AO, Jain S, Eneh JN (2021) A review of various image fusion types and transform. Indonesian J Electr Eng Comput Sci 24(3):1515–1522

27. Vijan A, Dubey P, Jain S (2019) Comparative analysis of various image fusion techniques for brain magnetic resonance images. Int Conf Comput Intell Data Sci (ICCIDS 2019) Procedia Comput Sci 167:413–422

28. Prashar N, Sood M, Jain S (2019) Morphology analysis and time interval measurements using mallat tree decomposition for CVD Detection. In: Luhach A, Singh D, Hsiung PA, Hawari K, Lingras P, Singh P.= (eds) Advanced informatics for computing research. ICAICR 2018. Communications in computer and information science, vol 955. Springer, Singapore, pp 171–181

29. Zhuang L, Guan Y (2018) Adaptive image enhancement using entropy-based subhistogram equalization. Hindawi Comput Intell Neurosci 2018:1–13. https://doi.org/10.1155/2018/3837275

30. Campos GFC, Mastelini SM, Aguiar GJ, Mantovani RG, de Melo LF, Barbon Jr S (2019) Machine learning hyperparameter selection for contrast limited adaptive histogram equalization. J Image Video Proc 59:1–20. https://doi.org/10.1186/s13640-019-0445-4

31. Bhardwaj C, Jain S, Sood M (2020) Automated diagnostic hybrid lesion detection system for diabetic retinopathy abnormalities. Int J Sens Wirel Commun Control 10(4):494–507

32. https://www.shutterstock.com/de/image-photo/ct-scan-brain-showing-intracerebral-haemorrhage-722186461
33. https://n.neurology.org/content/neurology/79/21/e183/F1.medium.gif
34. https://radiologyassistant.nl/neuroradiology/brain/anatomy

CPSSD: Cyber-Physical Systems for Sustainable Development—An Analysis

Shivani Gaba, Alankrita Aggarwal, Shally Nagpal, Pardeep Singh, and Rajender Kumar

Abstract Nowadays, various emerging technologies such as machine learning, IoT, and sensor-based learning are spreading in many areas like health care, manufacturing, and security. But an emerging technology named cyber-physical systems (CPS) directs the secure communications and feedbacks among physical and cyber mechanisms. CPS is going to play a significant role for next-generation healthcare systems for sustainable development. Cyber-physical systems are implemented to achieve efficient healthcare systems, and new architectures have been designed to monitor the devices in actual times. Apart from all the challenges cyber-physical systems face, they are moving toward the best sustainable development in healthcare systems. In this paper, we will study the survey of cyber-physical systems for sustainable development. Then, how CPS can be helpful for sustainable development for health care and then the critical procedural challenges, other interaction challenges, and future areas are also discussed. After reviewing the various systems and challenges, the study will focus on the recent technical challenges faced by cyber-physical systems and other future options available for CPS to better social and natural environmental levels.

Keywords Cyber-physical systems (CPS) · Health care · Sustainable computing

S. Gaba (✉) · S. Nagpal · R. Kumar
Panipat Institute of Engineering and Technology, Panipat, Haryana, India
e-mail: sgsgknl@gmail.com

A. Aggarwal
Chandigarh University, Mohali, India

P. Singh
Graphic Era Hill University, Dehradun, India

1 Introduction

Cyber-physical systems (CPS) are feedback-based systems which are adaptive and predictive in nature. CPS require improved design tools that enable design methodology which supports scalability and complexity management. Cyber-physical systems (CPS) direct the secure communications and feedbacks among physical and cyber mechanisms or can be considered as a new prototype for connection between physical and cyber mechanisms. The cyber components refer toward sensing and communiqué systems; i.e., it is the digital environment which is communicated and managed by computer programs where as the physical components consist of broad areas of systems and sensors. Cyber-physical systems incorporates programming, equipment, sensors, actuators, and inserted frameworks associated with human–machine interfaces and different frameworks. Various sensors, actuators, and control gadgets are associated by an organization to frame a perplexing framework for gaining, handling, ascertaining, and examining actual climate data and applying the outcomes to the natural climate [1, 2].

The improvement of CPS innovation is the way of working on personal satisfaction more productively than any time in recent memory, yet the dangers are turning out to be an ever-increasing number of intense as far as security. Also, the CPS experience issues surveying risks and weaknesses brought about by communications, and new security issues are arising [2, 3].

As IoT devices and sensors are connected and restricted over the network, which may result in various security issues, this is a severe and significant concern that vibrates the fundamentals of cyber-physical systems by directly intimidating the lives of people and the world. Therefore, it is a necessity to have a deep understanding of all weaknesses, fear, attacks via the security and privacy of CPS controls [1].

1.1 Motivation

The motivation of writing the paper is that as we know that every resource is limited to use and for the sustainable development, it is mandatory for analysis of cyber-physical systems when it comes to environmental sustainability. So we are discussing the analysis of cyber-physical systems for environmental sustainability.

1.2 Contributions

The various contributions of the paper are described below:

- In this paper, we will study the survey of cyber-physical systems in healthcare systems.

- Then, how CPS can be helpful for sustainable development for health care and then the critical procedural challenges, other interaction challenges, and future areas are also discussed.
- After reviewing the various systems and challenges, the study will focus on the recent technical challenges faced by cyber-physical systems and other future options available for CPS to better social and natural environmental levels.

1.3 Organization

The organization of paper is defined as: Sect. 2 defines architecture of cyber-physical systems. Section 3 discusses the planned approach for sustainable development. Section 4 defines the comparative analysis of cyber-physical systems, and finally conclusion is given in Sect. 5.

2 Architecture of Cyber-Physical Systems (CPS)

Cyber-physical system's design is the foundation of innovative work, and cyber-physical system's models should be altered and coordinated based on the current framework, network framework, and PC framework structure. Reflection and demonstrating of correspondence, calculation, and essential elements in various scales and time sizes are also expected to oblige the improvement of CPS. The CPS framework structure model is partitioned into three layers: user, data framework, and basic framework. The basic framework is made out of an enormous number of implanted frameworks, sensor organizations, smart chips, and so on, assuming responsibility for the assortment and transmission of data and the execution of control signals, as it is the establishment of the CPS. The data framework layer is mostly answerable for transmitting and handling the information gathered from the existing framework, the center of the CPS. The client layer primarily finishes the work, such as information question, methodology, and security assurance under human–PC connection climate, which regular CPS activities should ensure. CPS run as shut circle control [4]. The design of CPS is displayed in Fig. 1 and explained below:

A. *Sensor Networks*

Use a collection of sensors and real-time installed frameworks for ongoing information obtaining.

- It ensures the privacy of transmission of information.
- Minimizing the organization's energy utilization by the energy the board.
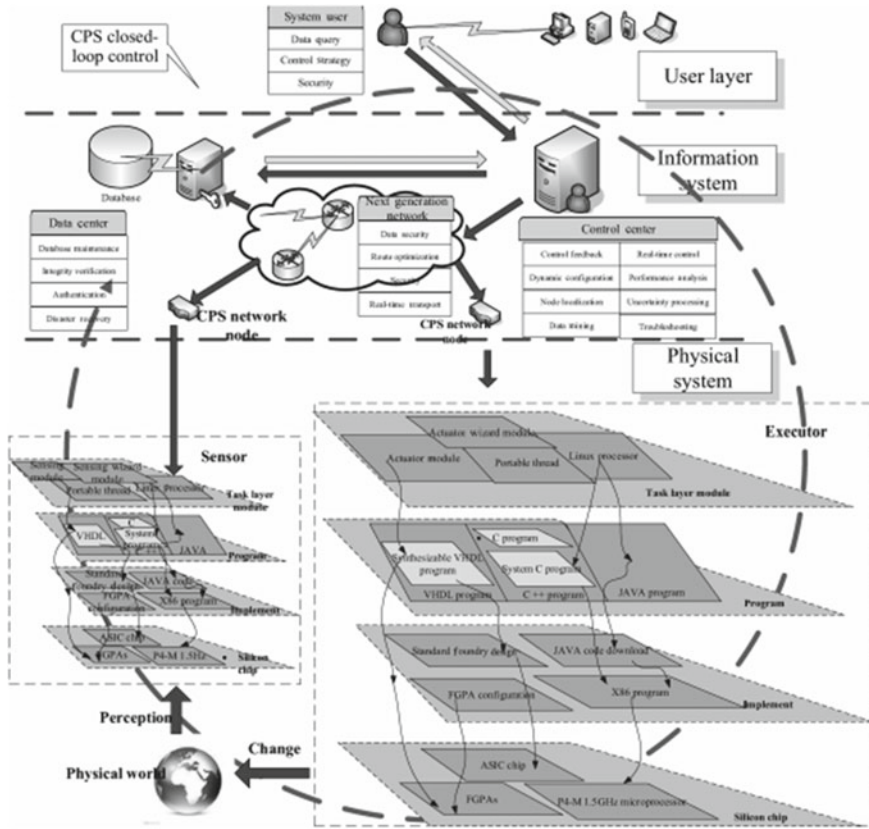
Fig. 1 Architecture of cyber-physical systems [1]

- Apply constant information security innovation to ongoing handling.

B. *Next-Generation Network Systems*

- It uses an unapproachable way for hacking and guard improvement over various organization assaults.
- Exploits advanced encryption calculation and validation innovation to guarantee the wellbeing of information transmission.
- Acknowledges the brisk trade of information transmission by upgrading existing steering calculations [5].

C. *Data Center*

- Sensor network sends information to a server farm for capacity through cutting-edge network frameworks.
- Server farm checks the confirmation and respectability of got information and stores it if they pass the assessment.

- In any case, it makes an impression on the control community.
- The server farm is likewise liable for routine support of the information base and speedy reaction to guidelines sent by control focus like a question. Standard crisis medicines are additionally expected to keep information base from breakdown [6].

D. *Control Center*

- Control focus is the main piece of cyber-physical systems.
- It gets clients' request directions and afterward sends inquiry orders to the server farm after character validation.
- It orders the inquiry results as indicated by control methodologies, reports back to the client in the event that they meet the necessities, in any case, discovers the area of the innovation, and sends control guidelines to actuators for comparing handling. Clients' necessities can powerfully change the control focus arrangement strategy.
- Leads figure investigation and execution examination of CPS conduct through information mining innovation and vulnerability preparing innovation.
- Distinguishes the organization and hub disappointment through deficiency analysis innovation and direct comparing handling.
- Guarantees the continuous control handling of CPS through constant control innovation.

E. *Actuator Networks*

- Obtain manage directions from control focus and send control guidelines to comparing hubs.

F. *System User*

- The system client incorporates an assortment of WEB workers, singular hosts, and outside gadgets.
- It is answerable to the communication with cyber-physical systems, for sending request direction for controlling focused information of input [2, 3].

In this architecture, CPS would run under shut circle control, and the continuous ability, security, and framework execution are completely thought to be, so it can begin to meet future CPS necessities. A few researchers likewise led explores on the framework design of CPS with various contemplated subjects and various application viewpoints. Progressed power framework is an unpredictable constant framework that contains network and actual segments. Each part might work well autonomously, however not when they are joined together because the impedance might cause mistakes, for example, the infringement of Nyquist rate in the recurrence area.

3 Planned Approach for Environmental Sustainability

The planned approach for environmental sustainability is shown in Fig. 2. The steps for ecological sustainability are divided into six categories such as commit, creating the structure for the support of sustainability in the environment, financing, and supporting the environmental sustainability, set the goals, measure them and evaluate the changes done, after that celebrate and share the successes, and finally continue to access and identify new opportunities [7].

A. *Make the Commitment*

Whenever the concept of change comes to mind, commitment is the primary paradigm. As drivers are behind the picture for the completeness of environmental sustainability in hospitals [8], the motive behind the commitment is:

- Economy
- Communal social accountability
- Taking care of pollution
- Employees satisfaction increases
- Foresting public image.

B. *Creating a structure for supporting environmental sustainability*

Suffering sustainability drives in crisis facilities and care structures require participation from pioneers across various workplaces, from senior bosses to division-level allies. Clinical facilities and care systems use various developments to consolidate sustainability and may think about setting up social events, for instance,

- Sustainability Leadership Council: This social event of senior pioneers can help critical level drives and allocate money-related resources. Since this social occasion is revolved around long-stretch moves close, it may need to meet once a quarter or less as often as possible.

Fig. 2 Environmental sustainability

Fig. 3 Sustainability endeavors [15]
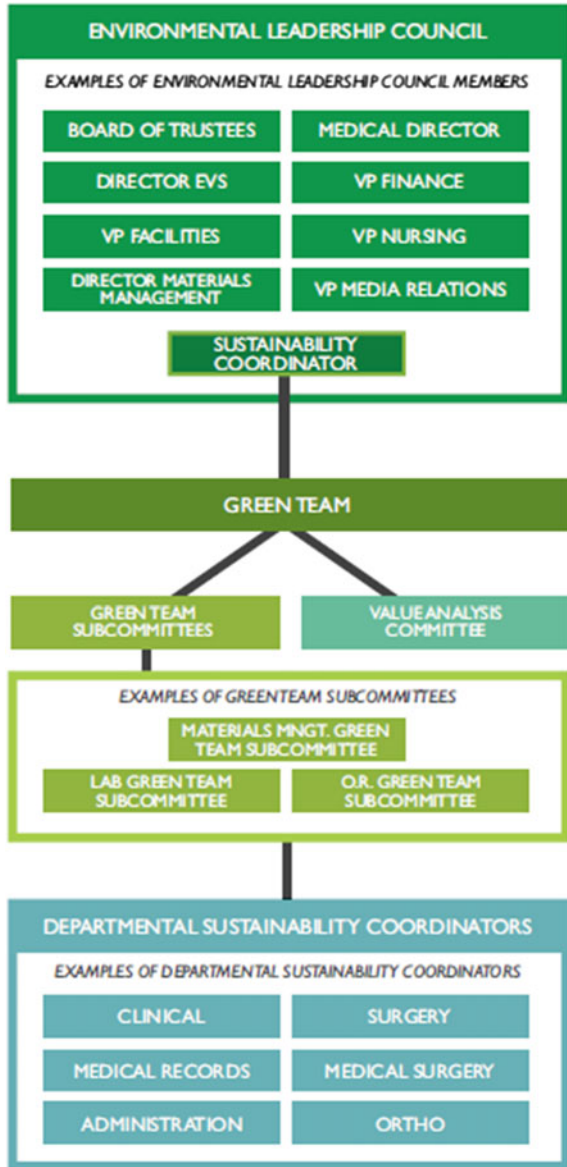
- Sustainability Committee: This gathering involves division boss-level specialists responsible for completing sustainability programs for any typical reason. Some green groups are parceled into subcommittees focused on energy, water, waste, and purchasing [9, 10].

- Value Analysis Committee: This social event gives a proficient method to pick things and advantages and watch out for creation networks execution issues like cost, utility, and suitability.
- Departmental Sustainability Coordinators: Division-level facilitators are tremendous for pushing and completing sustainability attempts [11, 12].

Figure 3 shows how sustainability endeavors can become coordinated into an emergency clinic's construction. While a few gatherings and panels incorporate staff individuals who should be included by the idea of their jobs and obligations, the overseer of ecological administrations ought to be associated with reusing.

C. *Support and Finance*

These include:

- *Awards, refunds, and gifts*: Many utilities have award or discount programs to assist with financing energy projects. Sustainability systems are likewise appealing to expected contributors, including people, associations, and organizations [13].
- *Shared-reserve funds arrangements*: Under shared-investment funds arrangements, an outsider consents to back, plan, and introduce energy projects, with the expenses paid from energy reserve funds that outcome from the ventures.
- *Power buy arrangements*: Under power buy arrangements, an outsider possesses, introduces, and works a force delivering resource, for example, a sustainable power source. This way, the clinic consents to buy the energy created from the plant.
- *Fossil fuel by product balances*: Various intentional business sectors are accessible to sell carbon balances and environmentally friendly power credits.
- *Set goals, measure, report, and evaluate change*

Finally set goals, measure, report, and evaluate the changes. Medical care pioneers can utilize their association's sustainability articulation as a plan to lay out quantifiable objectives for sustainability endeavors. Medical clinics and care frameworks should start by estimating pattern levels of energy use, water utilization, and the waste stream [14].

A. *Celebrate and share successes*

Praising victories is a critical inspiration that helps keep sustainability endeavors pushing ahead over the long run. Prizes and festivities for workers can move restored center around sustainability drives.

B. *Continue to access and identify new opportunities*

Sustainability is an excursion of permanent execution improvement. Numerous medical clinics and care frameworks start with less complicated activities before continuing to more mind-boggling sustainability projects. Indeed, even emergency clinics driving the sustainability development look for extra freedoms to turn out to be more productive.

Fig. 4 Analysis of current approach with proposed scheme
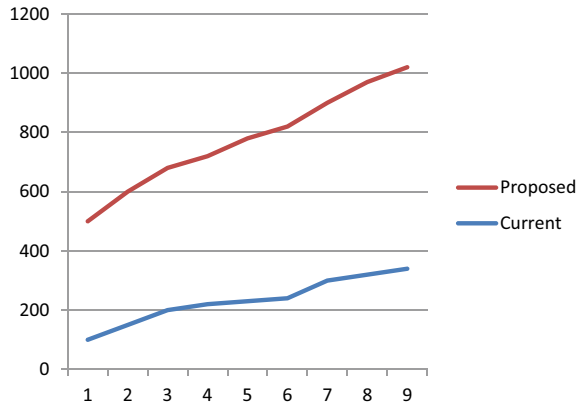
4 Experimental Result and Analysis

To examine the relevance of the proposed way to deal with CPS sustainability, creators demonstrated a designated ruinous sway on the CPS network foundation. The analysis is shown in Fig. 4, and this is an analysis of result if we will apply the planned approach so that can be considered as best approach for sustainable environment.

5 Conclusion

As cyber-physical systems are coming across many fields, it is having various applications in engineering domains. Here in this paper, we have analyzed the cyber-physical systems for sustainable environment that how cyber-physical systems are enhancing their research in sustainable development also. We have discussed six-step process for sustainable development and have compared it with previous paradigms as well, and we get to now after our approach the results enhanced. By understanding the essential significance of natural sustainability, emergency clinics, and care frameworks, the country over can further develop local area wellbeing, assemble its public picture, smooth out office activities, also, work on monetary execution—all vital perspectives to flourishing in the medical care climate of today, what is more, tomorrow.

References

1. Kim NY, Rathore S, Ryu JH, Park JH, Park JH (2018) A survey on cyber physical system security for IoT: issues, challenges, threats, solutions. J Inf Proc Syst 14(6):1361–1384
2. Liu Y, Peng Y, Wang B, Yao S, Liu Z (2017) Review on cyber-physical systems. IEEE/CAA J Automatica Sin 4(1):27–40

3. Sun Y, McMillin B, Liu XQ, Cape D (2007) Verifying noninterference in a cyber-physical system the advanced electric power grid. In: Proceedings 7th international conference quality software, Portland, OR, pp 363−369

4. Ilic MD, Xie L, Khan UA, Moura JMF (2010) Modeling of future cyber-physical energy systems for distributed sensing and control. IEEE Trans Syst Man Cybernet A: Syst Human 40(4):825–838

5. Al Faruque M, Regazzoni F, Pajic M (2015) Design methodologies for securing cyber-physical systems. In: Proceedings of the 10th international conference on hardware/software codesign and system synthesis, Amsterdam, pp 30–36

6. Eisenbarth T, Kumar S, Paar C, Poschmann A, Uhsadel L (2007) A survey of lightweight-cryptography implementations. IEEE Des Test Comput 24:522–533. https://doi.org/10.1109/MDT.2007.178

7. Isozaki Y et al (2016) Detection of cyber attacks against voltage control indistribution power grids with PVs. IEEE Trans Smart Grid 7(4):1824–1835

8. Aggarwal A, Dhindsa KS, Suri PK (2021) Performance-aware approach for software risk management using random forest algorithm. Int J Softw Innov (IJSI) 9(1):12–19

9. Nagpal S, Aggarwal A, Gaba S (2022) Privacy and security issues in vehicular Ad Hoc networks with preventive mechanisms. In: Proceedings of international conference on intelligent cyber-physical systems. Springer, Singapore, pp 317–329

10. Gaba S, Budhiraja I, Makkar A, Garg D (2022)Machine learning for detecting security attacks on blockchain using software defined networking. In: 2022 IEEE international conference on communications workshops (ICC workshops), pp 260–264. https://doi.org/10.1109/ICCWorkshops53468.2022.9814656

11. Kosek AM (2016) Contextual anomaly detection for cyber-physical security in smart grids based on an artificial neural network model. In: Proceedings cyber-physical security and resilience in smart grids (CPSR-SG), pp 1–6

12. Zegzhda PD, Poltavtseva MA, Lavrova DS (2017) Cyber-physic system systematization and security evaluation. In: Problems of information security. Computer systems (Problemy Informatsionnoi Bezopasnosti,Komp'yuternye Sistemy), no. 2, pp 127–138

13. Gaba S, Budhiraja I, Kumar V, Garg S, Kaddoum G, Hassan MM (2022) A federated calibration scheme for convolutional neural networks: models, applications and challenges. Comput Commun

14. Shivani Gaba DK, Shifali Singla A (Aug 2019) Genetic improved quantum cryptography model to optimize network communication. Special Issue, 8(9S):256–259

15. American society for healthcare engineering (2014)

16. Johnson T (2010) Fault-tolerant distributed cyber-physical systems: two case studies, masters thesis, University of Illinois, Department of Electrical and Computer Engineering, Urbana, USA

17. Schneier B (1999) Attack trees. Dr Dobbs J 24(12):21–29

18. Ten C, Liu C, Manimaran G (2008) Vulnerability assessment of cybersecurity for SCADA systems. IEEE Trans Power Syst 23(4):1836–1846

Automated Plant Health Assessment Through Detection of Diseased Leaflets

Preeti Sharma [iD]

Abstract The diseases in plants have a great impact upon the yield and quality of agricultural products. The study of such diseases helps to ascertain the reasons of losses affecting the economy of a country's growth to a great scale. This helps to draw the attention toward the different visual patterns on plant leaves against various ailments to monitor the health for a sustainable agriculture output. It is indeed a great challenge to monitor the health and detect diseases on plants manually as it calls for tremendous amount of time, efforts and expertise. The proposed method focuses on techniques of image processing for detecting various diseases on plant leaves. Steps like image acquisition, feature extraction and segmentation help to classify the plants as healthy or unhealthy and alert the farmers also.

Keywords Image processing · Image capture · Segmentation · Feature extraction

1 Introduction

India is an agricultural country where majority of the population income depends mainly upon farming. Agriculture contributes as a large source of employment and source of earnings through exports. The plants, fruits and even medicinal herbs offer serviceable benefactions to the human race. Though farmers have wide range of options of sowing crops, yet a degradable performance is observed when the agricultural output is affected by various diseases [1]. It is estimated that 35–45% of the crop production is lost due to untimely detection of diseases in plant leaflets. Though most of the observations are strictly based upon the physical examination of the leaves, this leads to extensive delay in time and treatment [2]. Such approaches also produced results with low accuracy greatly affecting the economic growth of our country as well as a significant income drop in the income of the farmers [3]. With the advent of the technology and latest techniques, the detection of diseases

P. Sharma (✉)
Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India
e-mail: preeti.sharma@chitkara.edu.in

in plants has been automated to improve food security [4]. These solutions have proved to as early and timely detection approaches with better accuracy than manual interventions [5, 6]. This paper helps to introduce techniques of image processing for detection of diseases in different plants.

2 Related Work

In this section, different methods of detection of plant diseases have been discussed using various image processing techniques.

In 2015, Rama Krishnan et al. [2] used the backpropagation algorithm for detecting the diseases in groundnut leaflets. Though colored images gave a higher accuracy, but limited diseases were detected. Padol et al. [4] used K-means clustering approach for detecting diseases in grape leaves in 2016. Though Support Vector Machine (SVM) was also used to classify the unhealthy leaves, the scope of accuracy improvement was also limited. Shaikh et al. [5] used Bi-level Thresholding technique with Gray-level Co-occurrence Matrix (GLCM) in Hidden Markov Model (HMM) for detecting diseases in citrus plant leaves. The results proved good accuracy in classifying and extracting the diseased parts. Kshir Sagar et al. [6] used MATLAB for implementing Multi-SVM and K-means algorithms for extracting feature from diseases of plants and fruits utilizing GLCM. Gandhi et al. [7] contributed solutions for detecting diseases in more than 150 crops using convolutional neural networks with Deep Learning approach in 2019. The model worked well with mobile and laptop mode but still required a physical intervention of farmers for image capturing. In 2020, Rama Thunnisa et al. [8] and Gomathy et al. [9] used SVM and clustering approaches, respectively, to detect diseases in vegetable plants. Saktidasan et al. [10] also used combination of Principal Component Analysis (PCA) and SVM with K-means algorithm to obtain better results. Singh et al. [11] to used latest approaches of Random Forest classifiers and Deep Learning for early plant disease detection with large datasets. But still, a limited accuracy was achieved in the results. A brief review of the recent work done is given below in Table 1.

3 Requirement of Hardware and Software

3.1 Components

- ESP-32 Microcontroller.
- Soil Moisture Sensor.
- DHT11 Temperature and Humidity Sensor.
- Submersible 3-6 V DC Pump.

Table 1 Brief review of work done

Ref. No.	Year/ Authors	Tech. used	Advantages	Disadvantages	Application(s)
[1]	[2015], Sheng et al.	K-means	Faster extraction of affected areas with higher accuracy using Itti's method	Miscalculation under contrast conditions	Soybean leaves' disease detection
[2]	[2015], Rama Krishnan et al.	Backpropagation	Use of colored images with up to 97.41% of detection efficiency	Limited scope of detection	Groundnut leaves' disease detection
[3]	[2015], Bhange et al.	SVM and K-means	Two separate methods suggested	Almost 80% accuracy achieved	Pomegranate disease detection
[12]	[2015], M. Ranjan et al.	Artificial neural network (ANN), Hue saturation value features, backpropagation	Different colors were used to assign blocks for different diseases	Up to 80% accuracy achieved	Cotton leaves' disease detection
[13]	[2016], R. Patil et al.	K-means	Technique of segmentation and feature extraction was done on the basis of six clustering processes	Accuracy not mentioned in detection results	Detection of diseases in grape leaves
[14]	[2016], R. Anand et al.	K-means	ANN's used	Limited scope	Brinjal leaves' disease detection
[15]	[2017], Rukaiyya, P. Shaikh et al.	Bi-level Thresholding, Gray-level Co-occurrences Matrix (GLCM)	Hidden Markov Model (HMM)	Scope not extended beyond disease detection in citrus leaves	Citrus leaf unhealthy region detection
[16]	[2018], V. Sawarkar et al.	SVM, fuzzy classifier, color analysis, neural networks, KNN	SVM classifiers and ANN	Limited results	Rose plant disease detection
[6]	[2018], Kshirsagar, et al.	K-means, GLCM, multi-SVM	Technique works well with both leaves and fruit-related diseases	Limited set of leaves and fruits were used	General plants disease detection

(continued)

Table 1 (continued)

Ref. No.	Year/ Authors	Tech. used	Advantages	Disadvantages	Application(s)
[17]	[2019], Sharath D. M. et al.	Canny edge detection (CNN)	Calculation for the affected edges was done in pixels, and based on pixel count, the percentage of infection in fruits was determined	Large datasets for future scope	Disease detection in pomegranate plants for bacterial blight
[7]	[2019], R. Gandhi et al.	Convolutional neural networks using deep learning	Two model's inception and mobile net were deployed which could work in both mobile phone and computer	Physical deployment of farmers required	More than 152 crops solution possible
[18]	[2019], S. Vignesh wari et al.	SVM classifiers	Detail study on the diseases of rice plant, size of images in dataset, etc.	Limited methods were suggested	Machine learning (ML) techniques for rice plant disease detection in agricultural research
[19]	[2019], L. Annabel et al.	ANN, classification, disease detection, SVM, ML	A brief summary of various techniques used for classifying and detecting various bacterial, fungal and viral plant leaf diseases	Poor recognition rate and classification accuracy	Leaf disease detection in different plants
[8]	[2020], Rahama Thunnisa U. et al.	K-means clustering, SVM	Clustering algorithm helped in achieving highly accurate results in less time	Less datasets were used	Detection of vegetable diseases and classification
[9]	[2020], B. Gomathy et al.	Image segmentation, classification, neural network, clustering	A review of different techniques for detecting plant diseases was presented	Limited results with accuracy were mentioned	Plant diseases' detection and classification techniques

Table 1 (continued)

Ref. No.	Year/ Authors	Tech. used	Advantages	Disadvantages	Application(s)
[10]	[2020], N. Vasudevan et al.	K-means clustering, Random Forest classification, image segmentation	Detection of the plant diseases in the early stage with comparison between two different algorithms for better results	Limited results with accuracy	Disease detection and recognition using K-means clustering
[11]	[2020], Singh D. et al.	Deep learning	Large datasets including 2598 data points	No accuracy mentioned	Visual detection of plant diseases
[20]	[2021], Sharma P. et al.	Application of internet of things	Water saving as irrigation is automated	Other sensors for measuring soil pH value, flow of air, fertilizer, etc. were not used	Irrigation automation

3.2 Software Utilized

- Blynk application.
- Arduino IDE.
- MATLAB.

4 Working

The circuit details of the project and its layout are shown in Figs. 1 and 2, respectively. The basic steps of the suggested methodology are given in Fig. 3. The flow description for the implementation steps is also shown in Fig. 4.

These steps are explained as:
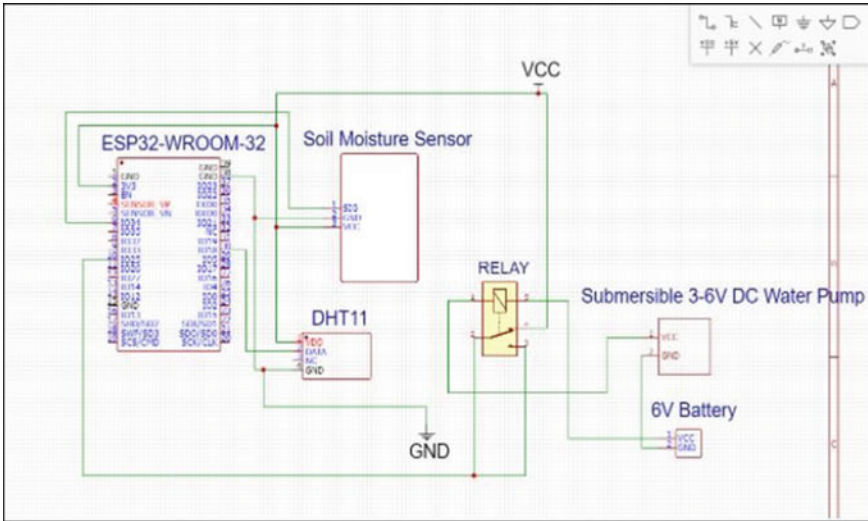
Fig. 1 Circuit details
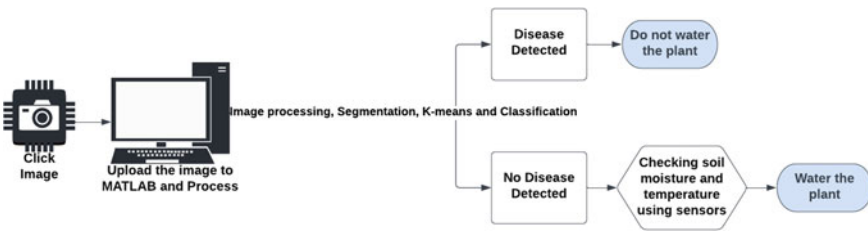
Fig. 2 Schematic representation

4.1 Acquiring the Image

First of all, colored images of the leaflets were taken through a high-resolution mobile camera. These were uploaded to a folder [13, 21, 22].

4.2 Initialization of the Image

This step is difficult as it is highly prone to noisy effects. The image is curtailed into size 256×256 using the Otsu technique of thresholding. This helps in classifying the given pixels into twin categories.
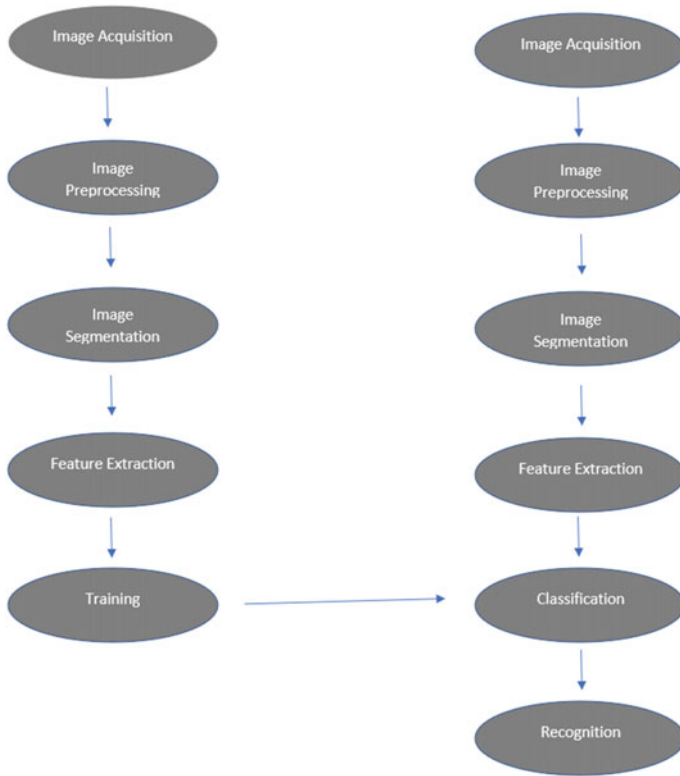
Fig. 3 Suggested methodology

4.3 Segmenting the Image

The information in the colored images is divided into three characteristics, namely 'L', 'a' and 'b', relating to light, red or green and blue or yellow values. After this conversion, K-means algorithm was utilized to segment the image into three clusters [5, 6, 14], as follows:

(a) The given data were assigned a number for cluster. Here K = 3.
(b) The mean was selected.
(c) The length amidst the points of data and mean were calculated utilizing Euclidean distance (utilizing the maximum distance as the criteria).
(d) The points of data nearer to the mean must be unchanged.
(e) The points of data closer to the mean value must be moved to adjoining cluster [9].

Fig. 4 Flow description of Implementation steps
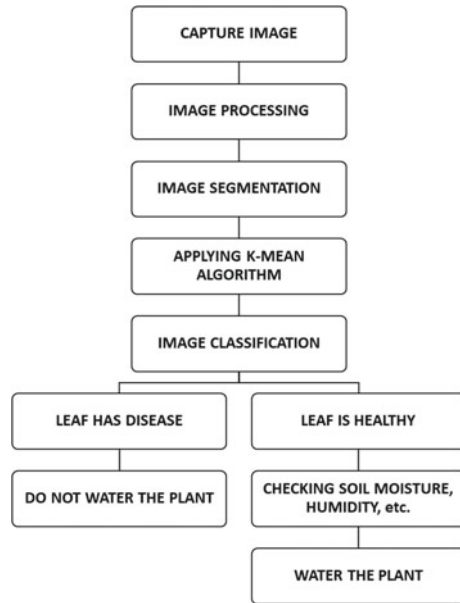
4.4 Extraction of Features

Each cluster selected in the previous step is utilized for extracting the features. The colored or RGB images are converted into grayscale digital images to express the intensity of the leaf diseases in the range 0–1. Only a limited number of pixels are selected that are necessary and adequate to characterize the entire segment of an image. The contrived area within the histogram of data or the image is used to indicate its frequency of occurrence. The affected area (in %) in an image relates to the ratio of the region of plant disease to the total area of the leaf used and reflects the image quality of the healthy plants. A GLCM was used to describe the steadiness of an image using the dimensional connection from the pixels of the image. Features like such as contrast, energy, homogeneity, correlation, mean and skewness were reclaimed from the matrix [14]. These are given as:

(1) Contrast estimates the extremity between a picture element and its adjoining pel over a proper image. For a constant image, the value is zero.
(2) Energy quantifies the level of fidelity between squared elements totalized in a matrix with levels amidst zero and one. For a constant image the value is one.
(3) Homogeneity weighs the level of affinity amidst the pixels. For a constant image, the value is one.
(4) Correlation evaluates the relationship amidst a pel value with its nearby values betwixt −1 and 1. [15].

(5)　Mean estimates the average of the samples over a finite number of samples.
(6)　Skewness is a measure of lack of symmetry.

4.5　Classification

SVM has been used as a two-fold classifier for classifying the consistency in different pattern acknowledgement applications. The notion of SVM is to generate a hyper-plane betwixt sets of data for indicating the classes they belong to [14]. Specimens nearer to the brink are chosen for resolution of the hyper-plane, as shown in Fig. 5.

For the experiment, a collection of normal and abnormal leaflets of Sigonia, Bhindi (Abelmoschus esculentus), Brinjal (Solanum melongena) and Karela (Momordica charantia) plants were taken. These are shown in Figs. 6 and 7, respectively, and denoted as plants P1, P2, P3, P4 and P5. Generally, the plants are affected by common diseases, such as Alternaria alternata, Anthracnose, Bacterial blight and Cercospora leaf spot. For detecting the most diseased plants, a mean of all pretentious areas was calculated for each plant to discriminate between the normal and damaged leaves in terms of accuracy of the algorithm used. The results helped to automatically water the healthy plants through the Blynk app software in the plant health tracking system once the soil parameters are checked using hardware components.

The complete project images are shown in Figs. 8a–d.

Fig. 5　Support Vector Machine classifier [10]
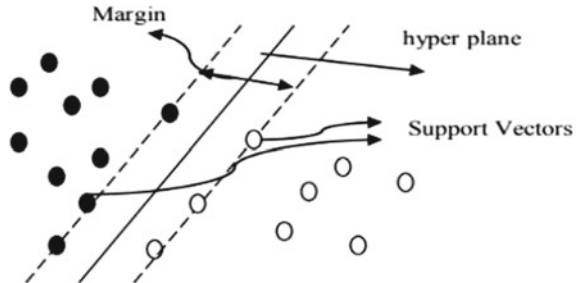
Fig. 6 Normal leaflets

Fig. 7 Damaged leaflets

5   Results

Highly pretentious diseased areas in plants were observed utilizing the k-mean clustering approach with SVM. The moisture, temperature and humidity parameters of the soil were continuously checked through the hardware components. The Blynk app successfully displayed the alerts for automatic watering of the plants. The detection of plant diseases with desirable accuracy is given in Table 2.

A brief comparison of the proposed technique with the existing methods is also summarized in Table 3 below.
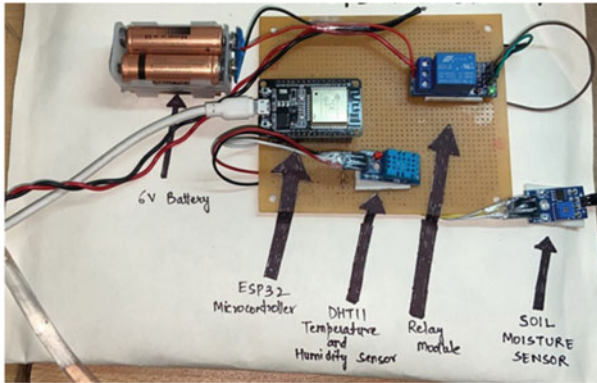
(a)
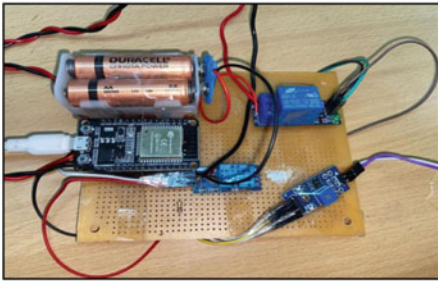
(b)

(c)
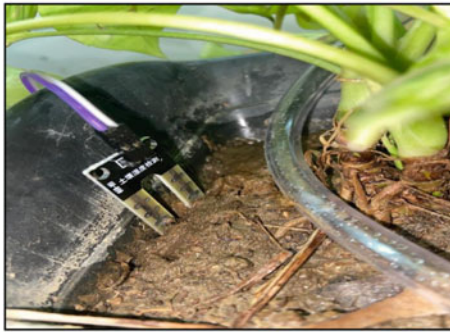
(d)

(e)

Fig. 8 a–d Complete project images

Table 2 Accuracy (in %) of the project for detecting diseases upon various abnormal leaves of plants

Plant No.	S. No.	Unhealthy plants (with disease)					Healthy plants
	Leaves tested	Alternaria alternata	Anthracnose	Bacterial blight	Cercospora leaf spot	Accuracy (%) in affected Areas (unhealthy plants)	
P1	L1	✔	✔			64.8063	
	L2		✔			41.33	
	L3		✔			30.27	
	L4						18.113
	L5						15.001
P2	L1				✔	50.8546	
	L2						15.0062
	L3			✔		50	
	L4	✔		✔		27	
	L5		✔	✔		69	
P3	L1						19.17
	L2	✔			✔	31	

(continued)

Table 2 (continued)

Plant No.	S. No.	Unhealthy plants (with disease)					Healthy plants
	Leaves tested	Alternaria alternata	Anthracnose	Bacterial blight	Cercospora leaf spot	Accuracy (%) in affected Areas (unhealthy plants)	
	L3		✓			48.5	
	L4		✓	✓		51.239	
	L5						17.301
P4	L1		✓		✓	47.23	
	L2	✓			✓	30.07	
	L3	✓		✓		27.93	
	L4						20.05
	L5			✓	✓	36.75	
P5	L1	✓		✓		28.87	
	L2	✓	✓		✓	47.41	

(continued)

Table 2 (continued)

Plant No.	S. No.	Unhealthy plants (with disease)					Healthy plants
	Leaves tested	Alternaria alternata	Anthracnose	Bacterial blight	Cercospora leaf spot	Accuracy (%) in affected Areas (unhealthy plants)	
	L3			✓	✓	52	
	L4	✓	✓	✓		51.239	
	L5						19.431

Table 3 Comparison of accuracy (in %) of the proposed technique with the existing techniques

	[9]	[10]	[17]	Proposed technique
Accuracy obtained	Good (range not mentioned)	Very limited	Less	40–60% approximately
Diseases detected	One	Two	Two	Four
Future scope	Large datasets needed	More agricultural plants can be used for detection	Other Vegetables can also be used	Technique may be utilized for testing diseases in more agricultural plants, vegetables and flowers

6 Conclusion

A nominal board utilizing sensors for detecting the real-time diseases in plants is introduced in the paper. A majority of common diseases in plants have been detected by the proposed technique with high accuracy. The technique currently limits the results obtained only from a few plants. Also, manual intervention of farmers is necessary for acquiring the images for calculating the detection accuracy.

However, the scope of the utility can be extended to other crops, fruits or vegetables also with the possibility of identification of more diseases utilizing better algorithms. The manual process of capturing the images may also be mechanized to extend the benefits of the proposed solution to a large part of the society and the country.

References

1. Gui J et al (2015) A new method for soybean leaf disease detection based on modified salient regions. Int J Multimedia Ubiquit Eng 10(6):45–52
2. Ramakrishnan M (2015) Groundnut leaf disease detection and classification by using back probagation algorithm. In: 2015 international conference on communications and signal processing (ICCSP). IEEE
3. Bhange M, Hingoliwala HA (2015) Smart farming: pomegranate disease detection using image processing. Procedia Comput Sci 58:280–288
4. Padol PB, Yadav AA (2016) SVM classifier based grape leaf disease detection. In: 2016 Conference on advances in signal processing (CASP). IEEE
5. Dhole SA, Shaikh RP (2016) Review of leaf unhealthy region detection using image processing techniques. Bull Electr Eng Inform 5(4):451–453
6. Kshirsagar G et al. Plant disease detection in image processing using MATLAB. Int J Recent Innov Trends Comput Commun 6(4)
7. Gandhi R et al (2018) Plant disease detection using CNNs and GANs as an augmentative approach. In: 2018 IEEE international conference on innovative research and development (ICIRD). IEEE
8. Rahamathunnisa U et al (2020) Vegetable disease detection using k-means clustering and SVM. In: 2020 6th international conference on advanced computing and communication systems (ICACCS). IEEE
9. Gomathy B, Nirmala V (2019) Survey on plant diseases detection and classification techniques. In: 2019 international conference on advances in computing and communication engineering (ICACCE). IEEE
10. Sankaran KS, Vasudevan N, Nagarajan V (2020) Plant disease detection and recognition using K means clustering. In: 2020 international conference on communication and signal processing (ICCSP). IEEE
11. Singh D et al (2020) PlantDoc: a dataset for visual plant disease detection. In: Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, pp 249–253
12. Ranjan M et al (2015) Detection and classification of leaf disease using artificial neural network. Int J Tech Res Appl 3(3):331–333
13. Patil R et al (2016) Grape leaf disease detection using k-means clustering algorithm. Int Res J Eng Technol (IRJET) 3(4):2330–2333
14. Anand R, Veni S, Aravinth J (2016) An application of image processing techniques for detection of diseases on Brinjal leaves using k-means clustering method. In: 2016 international conference on recent trends in information technology (ICRTIT). IEEE
15. Prakash RM et al (2017) Detection of leaf diseases and classification using digital image processing. In: 2017 international conference on innovations in information, embedded and communication systems (ICIIECS). IEEE
16. Sawarkar V, Kawathekar S (2018) A review: rose plant disease detection using image processing. IOSR J Comput Eng (IOSR-JCE) e-ISSN:2278–0661
17. Sharath DM et al (2019) Image based plant disease detection in pomegranate plant for bacterial blight. 2019 international conference on communication and signal processing (ICCSP). IEEE
18. Daniya T, Vigneshwari S (2019) A review on machine learning techniques for rice plant disease detection in agricultural research. System 28(13):49–62
19. Annabel LSP, Annapoorani T, Deepalakshmi P (2019) Machine learning for plant leaf disease detection and classification—a review. In: 2019 international conference on communication and signal processing (ICCSP). IEEE

20. Sharma P, Kanika K, Kaur M (2021) IOT based automated irrigation system. J Cardiovasc Dis Res 12(4):35–40
21. Trivedi NK, Gautam V, Anand A, Aljahdali HM, Villar SG, Anand D, Goyal N, Kadry S (2021) Early detection and classification of tomato leaf disease using high-performance deep neural network. Sensors 21:7987. https://doi.org/10.3390/s21237987
22. Ramesh TR, Lilhore UK, Poongodi M, Simaiya S, Kaur A, Hamdi M (2022) Predictive analysis of heart diseases with machine learning approaches. Malays J Comput Sci 132–148. https://doi.org/10.22452/mjcs.sp2022no1.10

1D CNN Model: BERT-Based Approach for Depression and Suicide Identification

S. P. Devika, M. R. Pooja, and Vinayakumar Ravi

Abstract Depression was worldwide psychological well-being issue, with severe belongings leading toward self-harm and suicide. An automated mechanism for detecting depression can greatly aid in clinical diagnosis and early management for depression. We propose novel text detection for depression diagnosis approach based on linguistic information from social media like Reddit posts. The suggested technique was composed of components: an attention layer to handle linguistic content, a One-Dimensional Convolutional Neural Network (1D CNN) to handle text, and a fully linked network that integrates the model output to assess depressed state. The study begins with a thorough examination of the process of creating a big labeled dataset derived from a spare dataset using an semi-supervised learning strategy to defeat data abundance. Dataset was reviewed rigorously in order to create a text classifier with many classes and comprehend the elements and attitudes associated with the various target groups. The study creates text analysis models and deep learning neural networks in order to add an evaluation tool in online suicide risk monitoring website trends. The findings can be put into a mental health report professionals to better grasp the essential trend.

Keywords Depression · Suicide · Reddit · CNN · BERT

S. P. Devika (✉) · M. R. Pooja
Department of Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysuru, Karnataka, India
e-mail: devikasp5@gmail.com

M. R. Pooja
e-mail: pooja.mr@vvce.ac.in

V. Ravi
Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia

1 Introduction

Suicide and sadness are on the rise across world. The World Health Organization states that (WHO) over 8,00,000 individuals commit suicide each year. That equates to every 40 s, one person dies. In accordance with Centers for Disease Control and Prevention (CDC), here was increase in the number of persons who suicide attempt or have suicidal thoughts rather of committing suicide. Suicide was a huge civic health issue which affects individuals of the all ages. Suicide was a personal decision that each person makes. However, the researchers believe that it is tied to things such as violence, health challenges, and economic worries. According to the WHO, the leading cause of death among 15–19 year olds is suicide, which might be induced by bullying, child abuse, or violence. External abuses can occur that have a negative influence on a person's mental health and result in depression.

Depression and suicide was able to be precipitated by psychological distress. Although the causes of suicide are unknown, there are several preventive mechanisms in place to help prevent them. Various groups are taking preventative measures and protective techniques to reduce the number of suicide deaths. Platforms for social media are the simplest in the direction of use and connect with. It was a means through which public may state their views and weaknesses. It is also evolved into forum wherever cyberbullying and kindness coexist. The rise of social media has evolved into a hub for cyberbullying and trolling. The platform has come to represent both cyber-aggression and friendship. Reddit is a social forum where users may ask questions, get answers, and exchange information. Reddit now has over 330 million active users worldwide and was rapidly approaching the size of Twitter. It always publishes text, photographs, videos, links, or polls. Each post has a title as well as substance. Reddit's material is classified and referred to as Subreddit. Reddit does not take ownership of or exercise control over those who do not commit suicide, but rather provides a forum for individuals to express themselves anonymously so that their peers can understand them and support or empower one another. As a result, previous study has focused on categorizing individuals on social media sites such as Tweet and Subreddit as examples having a high or low risk of suicide conduct or categorizing users as suicidal or non-suicidal poets.

To investigate patterns in the Reddit postings, several machine learning and deep learning techniques were used. The study will concentrate on statistically answering the research question: What factors determine the suicide risk classifications? And, how effective are deep learning and Natural Language Processing algorithms? Do in classifying suicide risk categories?

2 Literature Review

A substantial quantity work has occurred written resting on identifying online suicide risks' platforms. Artificial intelligence (AI) has a statistically significant role in diagnosing mental health disorders. With the abundance of available data via social media, such as Reddit, AI systems including machine learning (ML) and suicidal thoughts can be prevented or detected using Natural Language Processing (NLP) has emerged. There is a connection between the human intellect and its expression itself via language and psychological well-being discoveries. Suicidal jeopardy investigation was still topic of significant a keen interest in data analytics. This part discusses various studies that have been undertaken to estimate suicide risks.

Over the last decade, most NLP research has stressed the necessity of extracting sentiments from text data to disclose thoughts and conclusions about a topic or issue. NLP has shown to be promising extraction method aspects from text such as polarity, sentiment, and views that might reflect an individual's thinking. The study underlines the significance of the attributes collected from the web platform's textual data. For the categorization process, several data are utilized by ML algorithms that are collected via social media sites.

Majority of studies on suicide risk detection have concentrated on investigating internet material. The author claims that the increased use of technology and online social media has a detrimental effect on depression and suicide. Platforms like Reddit, Facebook, and Twitter all have a lot of text data, which have resulted in some interesting research on suicide and mental health issues. These microblogging services' platforms contain tens of millions of users ideas which may exist examined via the content and serve as an illustration of entire populace. Sample's findings know how to be applied to the entire population. In text data, the research also emphasizes the association among depression and suicide.

According to report, social networks and discussion boards have grown commonplace. Social Network Mental Disorders (SNMDs) such as cyberbullying and netcompulsion have developed from the use. According to the author, these have an early influence on mental health and the development of suicidal ideas. As a result, early response in such cases is required to avoid fatalities. To diagnose mental health issues, the study uses data from social media and healthcare records.

3 Proposed Approach

Our method combines an attention layer and a 1D CNN model in the direction of summaries' input from several modalities addicted to embeddings. Instead of manually selecting and organizing features, 1D CNN uses discriminative information, key characteristics generated from text descriptions, and phrase input. The horizontal concatenation of embeddings achieves multimodal information integration. The concatenated capture of text embeddings, both short-term and long-term

linguistic and acoustic properties, are allowing depressed patients and healthy persons to be distinguished. Finally, concatenated embeddings from 1D CNN models were fed into two one-layer FC networks, one of which generates a label depression and the other rates the severity of depression.

3.1   Text Features

Because depressed text data are few and difficult to get, training word embeddings from scratch is difficult. On huge datasets, an alternate method was to employ a pre-trained language representation. Bert conditions are used to pre-train deep bidirectional word representations in all layers using both left and right contexts. As a result, it was used in the suggested technique.

3.2   One-Dimensional Convolutional Neural Network

In CNN, convolutional layers combine across two-dimensional breadth and height. It was not relevant, however, when Mel spectrograms are used as inputs, where the width and height reflect time and frequency, respectively. When working with Mel spectrograms in the identification of depression issue, one-dimensional convolution is preferable than two-dimensional convolution. Convolution along the frequency axis is employed in the suggested model. As a result, the model may create characteristics that represent a short-term temporal connection. To put it another way, 1D convolution across frequency axis seeks toward gain distinctive Mel spectrogram persistence. One-dimensional convolution, in particular, changes reducing the size of the filter that 1 along an axis that is not the convolution axis. For example, 1D convolution was used on the filter size for Mel spectrograms which should be (1, k) along the frequency axis, where k was the size of the kernel along the frequency axis. Amount of nodes in the totally connected layers FC1, FC2, and FC3 is represented by the "Out characteristics" of FC1, FC2, and FC3 in Table 1.

1D Convolutional layers, two pooling layers, as well as a three-layer FC network are all included which comprise the proposed 1D CNN model. The final activation function and kernels stride vary based on the evaluation task, as in second convolution layers. A linear activation function as well as a two-dimensional kernel stride is utilized for depression and suicide severity evaluation task (regression). Softmax with a single kernel stride is used to employ depression forecasting challenge (Fig. 1).

Table 1 1D CNN model parameter settings

	Classification/Regression	
	Kernel-[1, 5]	
Conv-1	Stride-1	ReLU
	Filter size-32	
Max-pool-1	Kernel-[0.1, 2] Stride-[1, 0.3]	
	Kernel-[1, 5]	
Conv-2	Stride-1/2	ReLU
	Filter size-32	
Max-pool-2	Kernel-[0.1, 3] Stride-[1, 3]	
FC1	Out feature-128	ReLU
FC2	Out feature-128	ReLU
Dropout	0.5	
FC3	Out features-2	Softmax

Fig. 1 Proposed model diagram
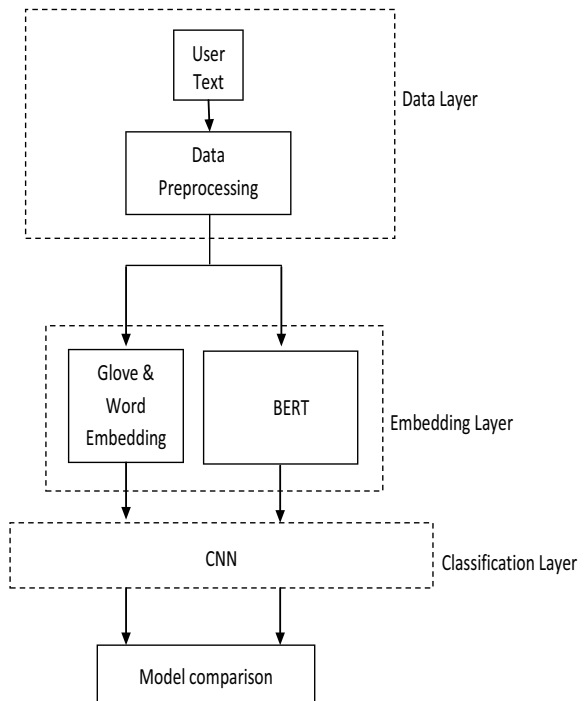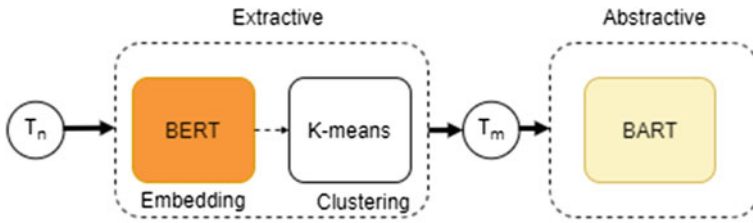
Fig. 2 Diagram demonstrating the summarization of user posts

4 Result and Discussion

The application of deep neural networks (DNNs), such as Convolutional Neural Networks (CNNs), has seen significant advancements in identifying mental illness on social media. In Natural Language Processing (NLP) tasks like text categorization and sentiment analysis, deep learning has shown excellent results. Numerous academic papers have focused on using deep learning models to analyze user content and textual data (Fig. 2).

The following diagram depicts the stages involved in implementing CNN-based techniques for identifying depression. The id of the context, the timestamp, the username, and the comments are among the properties. The system setup comprises an Intel(R) core(TM) i5 processor, 8 GB of RAM, and a 64-bit operating system. The flow of algorithms was shown as below:

1. The user's post was captured, and then preprocessing was performed to clean the data to be set all NA values to zero.
2. Using algorithms to differentiate positive and negative remarks as well as to forecast whether or not the user is depressed or suicidal.
3. With 80–20 criteria, cross-validation divides the first dataset for training and testing.

Accuracy and loss

Average difference is squared among the definite and projected values. It was mostly used to determine how the proper model output was. The lower the number, the better the fit.
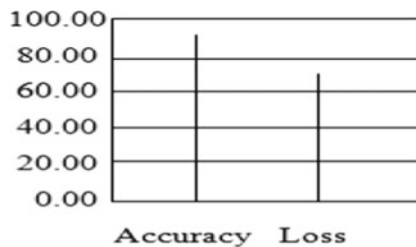
5   Conclusion

In this study, we developed a reliable framework for identifying depression and suicide using BERT and CNN, two deep learning techniques. The recommended method increases the accuracy of the intelligent healthcare systems' detection of psychological well-being illnesses, including depression and suicide. They selected different characteristics from both domains and divided them into four categories in order to account for user behavior. Deep learning-based text categorization algorithms have recently been created to help identify whether a single tweet has a tendency toward depression. In this study, deep learning algorithms are used to develop a deep learning technique for timely identification of depression and suicide using social media monitoring. Using Word2Vec in conjunction with the CNN Model, an accuracy of 99% with a loss of 67% has been recorded. Better results from future studies seek to use BERT in the role of text deep learning processing model techniques in the direction of detect depression and suicide over a specific user over time. The purpose of this work was to conduct a relative analysis based on machine learning methods in support of depression using prediction of a context-free model. As in future, a contextual bidirectional model for prediction might be developed to improve performance and give better decision assistance tools to consumers.

References

1. Shetty SC (2020) A deep learning approach for suicide risk assessment using Reddit
2. Baek JW, Chung K (2020) Context deep neural network model for predicting depression risk using multiple regression
3. Basha N, Ziyan Sheriff M, Kravaris C, Nounou H, Nounou M (2020) Multiclass data classification using fault detection-based techniques
4. Tadesse MM, Lin H, Xu B, Yang L (2019) Detection of depression-related posts in Reddit social media forum
5. Trotzek M, Koitka S, Friedrich CM (2020) Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences
6. Salekin A, Eberle JW, Glenn JJ, Teachman BA, Stankovic JA (2018) A weakly supervised learning framework for detecting social anxiety and depression
7. Esposito A, Raimo G, Maldonato M, Vogel C, Conson M, Cordasco G (2020) Behavioral sentiment analysis of depressive states
8. Ji S, Pan S, Li X, Cambria E, Long G, Huang Z (2020) Suicidal ideation detection: a review of machine learning methods and applications
9. Liu G, Guo J (2019) Bidirectional LSTM with attention mechanism and convolutional layer for text classification
10. Wu J-L, He Y, Yu L-C, Lai KR (2020) Identifying emotion labels from psychiatric social texts using a bi-directional LSTM-CNN model
11. Dinkel H, Wu M, Yu K (2019) Text-based depression detection: what triggers an alert
12. Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding

Enhanced Kidney Stone Detections Using Digital Image Processing Techniques

Surjeet Singh and Nishkarsh Sharma

Abstract On some days, kidney stones can become a big problem and if not detected early, then it will cause complications and sometimes surgery is in addition to what is needed to discover the stone. Here, to see a stone that is visible very well, credits to the image processing because by processing the image there is a bent to promote accurate and automatic results how to find a stone. Due to the presence of noise, the separation of kidney stones is inaccurate. Kidney stones have become more common in recent years as a result of a variety of causes (Akkasaligar et al. in 2017 International conference on smart technologies for smart nation (SmartTechCon). IEEE, pp 353–356, 2017 [1]). Using human and operational tests to generate findings for huge databases is tough. Identifying the precise position of the kidney stone during surgery is quite difficult. Kidney stone disease is one of the world's most dangerous diseases. A digital image and data processing methods are used to build an automated kidney stone identification (Karthick and Kumar in Int J Res Appl Sci Eng Technol (IJRASET). https://doi.org/10.22214/ijraset.2019.10020, 2019 [2]). On CT scans and MRIs, there is a lot of noise, which adds to low accuracy. Neural network-based artificial intelligence approaches have produced impressive outcomes. Artificial intelligence, such as neural networks, has shown to be quite useful in this field. As a result, in this project, the Digital Image Processing is being used (Ranjitha in Int J Recent Technol Eng. https://doi.org/10.35940/ijrte.b2469.078219, 2019 [3]).

Keywords Kidney stone · CT scans · MRIs · Neural network · Artificial intelligence · Digital image processing

S. Singh (✉) · N. Sharma
Department of Computer Science and Engineering, Jaypee University of Information Technology, Waknaghat 173234, India
e-mail: surjeetknmit@gmail.com

N. Sharma
e-mail: 181258@juitsolan.in

1 Introduction

The doctor often utilizes the hand approach to detect the stone in the X-ray image, but our method is totally automatic, which saves time and reduces the risks of making a mistake. This paper describes a method for detecting kidney stones that includes image processing processes. The kidney's primary role is to maintain electrolyte balance in the blood. The kidney is a bean-shaped organ that can be found on either side of the spine as depicted by Fig. 1 [1]. Globally, kidney stones are on the rise, and most individuals who suffer from the sickness known as concretion disease are unconscious of it because it gradually deteriorates the tissues before symptoms appear. Application for image processing includes pattern identification, medicinal applications, colour processing, and image sharpening and restoration. Ultrasound images have low intensity and pepper noise, which is a negative. The task of finding kidney stones is difficult. Speckle noise is a natural feature of medical ultrasound imaging, and it often tends to lower image quality and contrast, decreasing the visualization modality's diagnostic utility [3]. There have been 16,000 publications published in the last five years on kidney stone diagnostics using various filters in Google Scholar and IEEE Explore. Find photos with low intensity and irregular noise using digital image processing [2]. The kidney stone is located using ultrasound images, and it is then eliminated by shattering it into smaller pieces that can flow through the urinary tract. Kidney stones are hard to find because of their low resolution and speckle noise.

This problem is solved by employing the appropriate imaging techniques and filters. Speckle noise is common in ultrasound images, which cannot be eliminated with standard filters. As a result, the median filtering algorithm is proposed, which reduces speckle noise. A median filter is used to eliminate noise and detect the stone region in the pre-processed image.

Fig. 1 Stone in kidney [1]

Table 1 Comparison of various current methods

Ref. No.	Algorithm/Techniques	Importance	Measuring
[6–8]	Threshold comparison	CT scan image filtering before digitization	Accuracy, specificity, sensitivity
[9–11]	KNN, SVM	Image resolution and stone detection in kidney scans	Recall precision
[12, 13]	*K* nearest neighbour	There are several framework established used for segmentation	Accuracy
[14–16]	Gray level co-occurrence matrix future extraction	The image is improved by filtering the low-contrast areas	Accuracy, specificity, sensitivity
[1, 17]	Segmentation	The size and position of the stones are determined with the help of segmentation	Accuracy, precision, sensitivity
[18–20]	Image processing techniques	Ultrasonography pictures are used to compare filters	Accuracy and sensitivity

2 Literature Review

Due to their low resolution and impulse noise, kidney stones are a very difficult task to detect [4, 5]. The solution to this problem is to employ the appropriate imaging methods and filters. The median filtering technique is suggested to minimize the speckle noise typically present in ultrasound images, that can't be removed by conventional filters [5]. A median filter is used to pre-process the image to identify the stone region while removing noise. The most of kidney stone disease sufferers are unaware they have the condition because it slowly affects organs before exhibiting symptoms. Numerous kidney stone types, including renal calculi, struvite stones, and staghorn stones, were examined [6]. In this section we have compared different types of approaches which are already used by the different researchers in their publications. This section combines the different approaches taken to detect kidney stones and to track the accuracy using the methodologies represented in Table 1.

Furthermore, we have taken references from different authors and researchers and have understood their approaches to perform the experiment.

3 Approach and Methodology

The goal of this study is to detect the kidney stone from the dataset of ultrasound images (which covers 50% of image with Kidney stone and rest 50% of images without kidney stones. This study will be carried based on digital image processing techniques such as applying different filters on the dataset and moreover, this study

Input Images ➡ Image Filters ➡ Image Processing ➡ Comparison ➡ END

Fig. 2 Proposed model for kidney stone detection

has been conducted based on research obtained through various resources. We have divided this study into 2 parts where Part 1 is identified as Median Filter and the PART 2 is identified as Edge Detection Filter. We have used MATLAB to simulate and develop the code. After successfully obtaining the results from both the filters on the dataset image, we will compare the obtained images on the MATLAB itself to track the accuracy and sensitivity obtained. Furthermore, we will compare the obtained results in MS-Excel. The result picture will be presented in the command window once the code has been processed. To collect the results and determine the classification performance using the technique, the trial will be repeated with various renal ultrasound images. Kidney stone ultrasound scans are used as input images, which are separate variables. The output variables will be accuracy and sensitivity.

$$\text{Detection rate} = (\text{No. of output images/Total input images}) * 100$$

Figure 2 represents the step by step process which were carried in this study.

A. *Dataset*

Nearly 114 kidney images (Ultrasound Image) from various hospitals in the US are included in the kidney image database. In the database, there are both normal and aberrant photos. Images from the database are randomly picked and processed to the stone detection method.

B. *Data pre-processing*

Preparing relevant data for usage is a step in the data extraction process that involves pre-processing. Real-world data is frequently incomplete, unreliable, devoid of certain structures or behaviors, and subject to a plethora of errors. Pre-processing data is an established remedy for such problems. To prepare it for subsequent processing, raw data is preprocessed.

C. *Input images from dataset*

We discovered that the dataset is so vast so, at random we picked up a raw image and loaded it in the MATLAB for further processing as shown in Fig. 3a.

Code

```
clc
close all;
clear all;
```
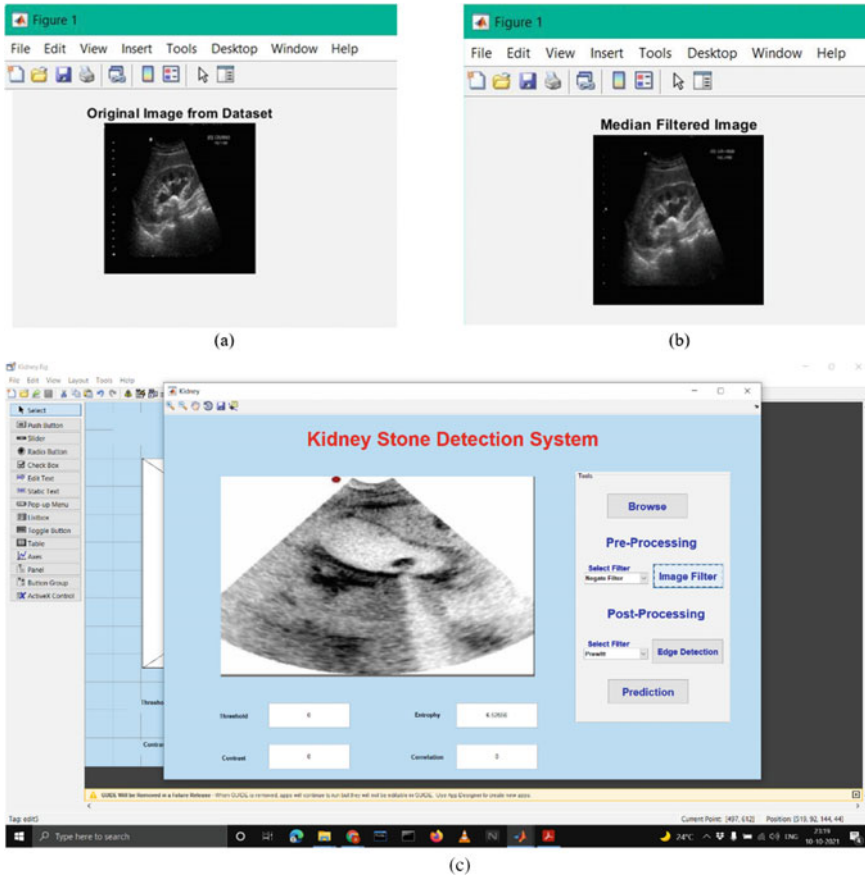
Fig. 3 **a** Image from dataset, **b** median filtered image, **c** negate filter image

```
a=imread("kidney.jpg");
imshow(a);
subplot(3,2,1);imshow(a);title('original Image from dataset');
```

D. *Image filter (median filter)*

A non-linear adaptive filtering method known as the median filter is used to eliminate noise from images and signals. A frequent pre-processing method to enhance the outcomes of future processing is noise reduction (for example, edge detection on an image) [21–24]. Because it maintains outlines while lowering noise in some circumstances (though see the rationale below), median filtering is frequently used in digital image processing. It can also be useful in signal processing. The essential idea behind the median filter is to iteratively replace each element in the signal with the median of the entries immediately adjacent to it. The "window" is a neighborhood pattern that moves entry by entry over the full signal. The simplest evident

window for one-dimensional signals is merely the first few preceding and following entries, whereas the window for two-dimensional (or higher-dimensional) data must encompass all entries within a certain radius or ellipsoidal region (i.e. the median filter is not a separable filter). We will now pass the median filter on the image which was taken at random from the dataset in the previous step. After loading the image, we will pass the median filter on the image and further we will try to change the intensity of the median filter and observe the better filtered image in the MATLAB as shown in Fig. 3b. We will perform this operation throughout the dataset in the step by step manner and we will record different types of results on MS-Excel in terms to calculate accuracy.

Code

```
clc
close all;
clear all;
[file, pathname]=uigetfile('*.jpg','Load Image 1');
cd(pathname);
a=imread(file);
%subplot(3,2,1);imshow(a);title('original Image from dataset');
resized_image=imresize(a,[256 256]);
I=0.21*resized_image(:,:,1) + 0.72*resized_image(:,:,2) + 0.07*resized_image(:,:,3);
%MEDIAN FILTER
J=medfilt2(I)
subplot(2,2,1);imshow(J);title('Median Filtered Image')
```
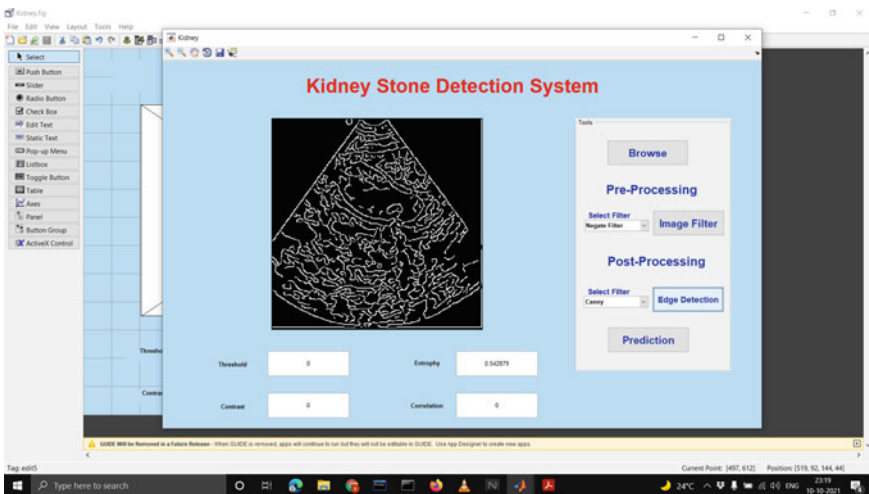
Algorithm

- Load the input images from the available database.
- Applying pre-processing filters on images such as negate, contrast, median filter that are widely used in the field of image processing.
- We have then chosen/customized our filter values and their sensitivity according to our requirement.
- We have then loaded and processed our image on console and obtain the values of each image that we have generated followed by further processing of image in the post-processing step.

E. *Image filter (edge detection)*

Edge detection is the term for a group of mathematical methods for identifying slopes or boundaries in digital images when there are abrupt changes in the lighting, or more precisely, where there are interruptions. Finding signal discontinuities in one-dimensional signals is a task known as step detection, whereas finding them in time-varying signals is known as change detection. Edge detection is a crucial approach in image processing, machine vision, and computer vision, particularly in the areas of pattern extraction and recognition. Edges serve as a representation of

an object's boundaries. Edge detection is therefore an important step in any object detection or recognition process. Gradient image approximation is the foundation of straightforward edge detection kernels. An object's edges hold the majority of the shape information. As a result, we first identify the edges of a photo before using these filters to enhance the edges and improve the brightness and transparency of the image. Edge detection is frequently used to measure, spot, and pinpoint changes in an image's grayscale. The edges make up the fundamental aspect of an image based on approximate gradient images. The edges and lines of an object are the most visible. An object's structure may be determined using edges and lines. As a result, edge extraction is a crucial approach in graphics processing and feature extraction. We will now pass the Edge detection filter on the image which was taken at random from the dataset in the previous step. After loading the filtered image(s) we will now pass the edge detection filters (CANNY, SOBEL and PREWITT) on the image as shown in Fig. 4 and further we will try to change the intensity of the filters (if required) and observe the better filtered image in the MATLAB. We will perform this operation throughout the dataset in the step by step manner on the median filtered images and we will record all types of results on MS-Excel in terms to calculate accuracy in the following step.

Code

```
clc
close all;
clear all;
[file, pathname]=uigetfile( '*.jpg','Load Image 1');
cd(pathname);
a=imread(file);
```


Fig. 4 Edge detection filtered image

```
%subplot(3,2,1);imshow(a);title('original Image from dataset');
resized_image=imresize(a,[256 256]);
I=0.21*resized_image(:,:,1)  +  0.72*resized_image(:,:,2)  +  0.07*resized_
image(:,:,3);
%Sobel filter
BW1=edge(I,'sobel');
%Prewitt filter
BW2=edge(I,'Prewitt');
%Canny
BW3=edge(I,'Canny');
subplot(2,2,1);imshow(BW1);title('Sobel Image');
subplot(2,2,2);imshow(BW2);title('Prewitt Image');
subplot(2,2,3);imshow(BW3);title('Canny Image');
```

Algorithm

- Load the input images from the pre-processed image database.
- Applying post-processing filters on images such as sobel, canny and other edge detection customised filters that are widely used in the field of image processing.
- We have then chosen/customized our filter values and their sensitivity according to our requirement.
- We have then loaded and processed our image on console and obtain the values of each image that we have generated followed by further processing of image and then pixel values has been recorded for further analysis.

4   Result-Analysis

In order to calculate and track the results for this study, we used the raw dataset images and after successfully passing them from median filter and Edge detection filters, we have obtained the most accurate and sensitive images which we will further compare to track which is the most accurate process to detect kidney stones in an ultrasound image. As discussed earlier, we have recorded the SNR and PSNR values of the dataset images on MATLAB [25], after successfully tracking the stones using different filters in the experiment, or even not detecting any stones in the ultrasound images. We have obtained both the values before the experiment and after the experiment on MATLAB. To obtain the most accurate results and compare between the images and their accuracy and sensitivity we have used MS-Excel (for comparison and creating graphs and tables for prediction). Table 2 is obtained when we recorded randomly 18 iterations from the dataset to track and make suitable prediction based on analysis.

This table illustrates the effectiveness and sensitivity of the median filter and edge detection filter, respectively. From the filtered picture dataset, 18 cycles were taken. The dataset includes 2 classes (with kidney stones and without kidney stones). The performance graph for the identification probability of kidney in terms of accuracy

Table 2 Samples taken from datasets

S. No.	Median filter		Edge detection	
Parameter	Accuracy	Sensitivity	Accuracy	Sensitivity
1	84.3	80.3	80.2	80
2	89.4	85.1	85.5	81.2
3	87.2	84.1	84	83.7
4	88.5	86	82.3	80.6
5	90.5	84.2	84.1	81.2
6	87.2	83.1	86.3	84.8
7	89.6	85.1	85.5	81
8	84	81.5	81.8	80.2
9	86.3	82.3	82.2	84.4
10	85.6	84.2	81.5	80.1
11	91.5	88.5	86.2	84.3
12	89.2	86.1	84.2	82.3
13	89.4	85.2	86.3	82.1
14	90.6	84.2	84.2	81.2
15	84.2	81.2	81.5	80.4
16	87.2	84.2	84.1	83.3
17	84.3	80.2	80.1	80.6
18	89.5	84.5	86	85.3

and sensitivity is shown in Fig. 5a, and it compares the median filter and edge detection filter over the course of 18 iterations.

The responsiveness curve for the detection of 18 image dataset is shown in Fig. 5b, along with an evaluation of the median filter and edge detection filter. The amount of samples taken, mean precision and sensitivity values for the 114 samples, and standard deviation were all calculated using MS-Excel and are shown in Table 3. The accuracy and sensitivity of the median filter and edge detection filter are shown in the table along with the standard error mean. For 114 samples, the median filter and rank filter algorithms' mean accuracy value, standard deviation, and standard error mean are determined [1, 2, 6].

(a)

(b)

(c)
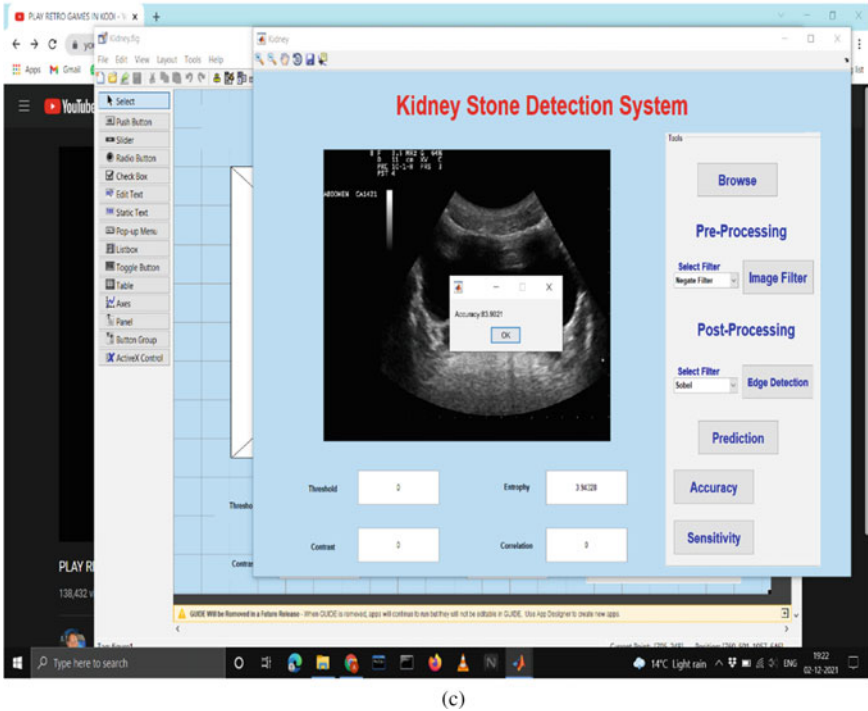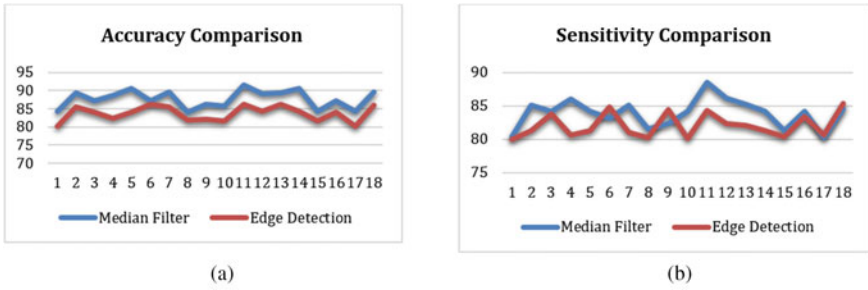
Fig. 5 **a** The performance graph that contrasts edge-based segmentation filter and median filter. **b** The sensitivity graph compares the quality of the project's edge detection filter (**c**) and median filter

Table 3 Group statistics: statistical analysis of median filter and edge detection filter

Group		No. of samples	Mean	Std. deviation	Std. mean error
Accuracy	Median	114	87.694	2.454161465	0.578451405
	Rank	114	83.667	2.102379604	0.495535625
Sensitivity	Median	114	83.889	2.150756061	0.506938065
	Rank	114	82.039	1.793497696	0.422731461

5   Conclusion

The study's objective was to examine the various methods used by researchers to find kidney stones in human bodies. The main objective here was to achieve bigger and more significant results. We were able to obtain some incredibly positive results after just 18 iterations. We discovered that the median filter outperforms the edge detection filter in terms of accuracy and sensitivity. Based on the findings and tabular data, median filters have a higher accuracy (87.6%) and sensitivity (83.8%) than edge detection filters, which have a lower accuracy (83.6%) and sensitivity (82.0%) for detecting kidney stones in ultrasound pictures. Future work on this area can make use of artificial intelligence algorithms, techniques, or approaches. The study's main flaw, meanwhile, is the healthcare system's difficulties. With time, these restrictions will be overcome, but in the meanwhile, we'll be able to combine more intricate strategies for more significant results.

Author Contribution All writers contributed equally to this research, and all authors reviewed and approved the final manuscript.

References

1. Akkasaligar PT, Biradar S, Kumbar V (2017) Kidney stone detection in computed tomography images. In: 2017 International conference on smart technologies for smart nation (SmartTechCon). IEEE, pp 353–356
2. Ranjitha M (2019) Segmentation of ultrasound abdominal images to extract region of interest. Int J Recent Technol Eng. https://doi.org/10.35940/ijrte.b2469.078219
3. Karthick N, Kumar PV (2019) A review on brain tumour detection using magnetic resonance imaging. Int J Res Appl Sci Eng Technol (IJRASET). https://doi.org/10.22214/ijraset.2019.10020
4. Rahman T, Uddin MS (2013) Speckle noise reduction and segmentation of kidney regions from ultrasound image. In: 2013 International conference on informatics, electronics and vision (ICIEV). IEEE, pp 1–5
5. Hosseini Z, Bibalan MH (2018) Speckle noise reduction of ultrasound images based on neighbor pixels averaging. In: 2018 25th National and 3rd international Iranian conference on biomedical engineering (ICBME). IEEE
6. Hafizah WM, Supriyanto E (2011) Comparative evaluation of ultrasound kidney image enhancement techniques. Int J Comput Appl 21(7):15–19
7. Thein N, Nugroho HA, Adji TB, Hamamoto K (2018) An image preprocessing method for kidney stone segmentation in CT scan images. In: 2018 International conference on computer engineering, network and intelligent multimedia (CENIM). IEEE, pp 147–150
8. Heygster G (1982) Rank filters in digital image processing. Comput Graph Image Process 19(2):148–164
9. Verma J, Nath M, Tripathi P, Saini KK (2017) Analysis and identification of kidney stone using Kth nearest neighbour (KNN) and support vector machine (SVM) classification techniques. Pattern Recognit Image Anal 27(3):574–580
10. Akshaya M, Nithushaa R, Raja NSM, Padmapriya S (2020) Kidney stone detection using neural networks. In: 2020 International conference on system, computation, automation and networking (ICSCAN). IEEE, pp 1–4

11. Vikas AL (2021) Kidney stone detection using image processing and neural networks. Ann RSCB 25(6):13112–13119
12. Velmurugan V, Arunkumar M, Gnanasivam P (2017) A review on systemic approach of the ultra sound image to detect renal calculi using different analysis techniques. In: 2017 Third international conference on biosignals, images and instrumentation (ICBSII). IEEE, pp 1–7
13. Shaharuddin NA, Mahmud WMHW, Ibrahim N (2016) An overview in development of computer aided diagnosis (CAD) for ultrasound kidney images. In: 2016 IEEE EMBS conference on biomedical engineering and sciences (IECBES). IEEE, pp 35–39
14. Arabi PM, Pratibha TP, Dahal A (2017) Distinguishing staghorn and struvite kidney stones using GLCM and pixel intensity matrix parameters. Int J Adv Network Appl 8(4). Special issue—NCBSI-2016
15. Pratt WK (2013) Introduction to digital image processing. CRC Press
16. Arabi PM, Joshi G, Nigudgi S (2019) Categorizing kidney stones using region properties and pixel intensity matrix. In: Cognitive informatics and soft computing. Springer, Singapore, pp 11–18
17. Suganyadevi S, Renukadevi M (2021) Segmentation of kidney stone region in ultra sound image by using region partition and mounting segmentation algorithm (RPM). Inf Technol Ind 9(1):512–518
18. Rohith A, Premkumar S (2021) Detection of kidney stones in ultrasound images using median filter compared with rank filter. Rev Gestão Inovação Tecnol 11(4):1096–1111
19. Sadeghi M, Shafiee M, Memarzadeh-Zavareh F, Shafieirad H (2012) A new method for the diagnosis of urinary tract stone in radiographs with image processing. In: Proceedings of 2012 2nd international conference on computer science and network technology. IEEE, pp 2242–2244
20. Ebrahimi S, Mariano VY (2015) Image quality improvement in kidney stone detection on computed tomography images. J Image Graph 3(1):40–46
21. ALhussieny YA (2017) Using MATLAB to get the best performance with different type median filter on the resolution picture. Int J Comput Commun Instrum Eng (IJCCIE) 4(1). ISSN 2349-1469, EISSN 2349-1477
22. Suresh MB, Abhishek MR (2021) Kidney stone detection using digital image processing techniques. In: 2021 Third international conference on inventive research in computing applications (ICIRCA). IEEE, pp 556–561
23. Ihsan R, Marqas R (2020) A median filter with evaluating of temporal ultrasound image for impulse noise removal for kidney diagnosis. J Appl Sci Technol Trends 1(2):71–77
24. Goel R, Jain A (2020) Improved detection of kidney stone in ultrasound images using segmentation techniques. In: Advances in data and information sciences. Springer, Singapore, pp 623–641
25. https://www.ni.com/en-in/innovations/white-papers/11/peak-signal-to-noise-ratio-as-an-image-quality-metric.html#:~:text=The%20term%20peak%20signal%2Dto,the%20quality%20of%20its%20representation

Prediction on Illegal Drug Consumption Promotion from Twitter Data

Shailesh S. Shetty, Chandra Singh, Supriya B. Rao, Usha Desai, and P. M. Srinivas

Abstract Illegal drug consumption is one of the major problems faced by the world. With the increase in deaths from consumption of drugs, it is important to identify the users and understand why they use drugs. Because of the increased accessibility to the internet across the globe, many people are connecting to social media at a higher frequency. By extracting data from the twitter, we can predict if the user consumes illicit or semi-legal drugs. This can be used to track the spread of these posts in specific communities.

Keywords Illegal drug consumption · FDA adverse event monitoring system (FAERS) · Adverse drug reaction (ADR)

1 Introduction

One of the world's biggest issues is the use of illegal drugs. It's critical to recognize drug users and comprehend their motivations for using drugs in light of the rise in drug-related fatalities. Many individuals are connecting to social media more often due to the growing accessibility of the internet globally. One can determine whether a person uses illegal or partially legal substances by pulling data from Twitter [1, 2]. This may be used to monitor how these posts are spreading areas. We aim to investigate using all the syntax, slang and emoji embedded in tweets using which we can build an accurate classifier to identify illicit drug users. We use machine learning methods for automatic classification that can identify tweets that are indicative of drug

S. S. Shetty · S. B. Rao · P. M. Srinivas
Department of CSE, Sahyadri College of Engineering & Management, Mangalore, India

C. Singh (✉)
Department of ECE, Sahyadri College of Engineering & Management, Mangalore, India
e-mail: chandrasingh146@gmail.com

U. Desai
Department of Electronics and Communication Engineering, SR University, Warangal, Telangana, India

abuse [3, 4]. To build a system which is capable of enabling the government bodies to function efficiently with regard to helping missing people find their next of kin by indicating their relationship with that of their family members. To provide a safer and smooth experience in social media platforms by generating suggestions of people related to them. Although the current implementation of this model is constrained to only a particular region as the images (NSL) are obtained from a specific area, what makes the project unique is the scalability fact or which tends to grow exponentially [5, 6]. As humans descend along the generation tree, enormous amounts of minute changes are observed, and adaptability tends to be a major challenge to the current existing systems [1, 7].

2   A Background Study

To map the chemical space underlying addiction, we use multidimensional scaling (MDS) to integrate the chemical configurations of addictive entities into a low-dimensional environment. After the embedding is obtained, the etiological agents underlying various types of addiction can be linked to each other and to other chemical entities. In which, Pop Data used to create complex models of addiction progression, where data from the popular social media networks us taken [8]. Implementing natural-language processing and machine-learning approaches in a web-based program to help scientists and software developers exploit serendipitous drug use social media. It's trained, tuned, and tested our models with a 15714-sentence gold standard dataset 447 [2.8%] representing serendipitous substance use. In addition, authors compared deep neural networks to support vector machine, random forest, and algorithms with AdaBoost M1. Context knowledge helped to reduce the false-positive rate of models for deep neural networks [9].

Wu et al. [10] focuses on building a computational pipeline to capture, store, and interpret tweets to identify messages concerning adverse drug reactions, identified as drug side effects induced by a drug at a regular dosage during daily use. The FDA Adverse Event Monitoring System (FAERS) provides a database of adverse drug incidents documented by healthcare providers and customers. The pervasive electronic social networks, such as Twitter, will complement adverse drug effects with information. Short Twitter posts, or tweets, are frequently used to share a view on drugs, and to elicit and collect customer input. Adverse drug reaction (ADR) can be defined as adverse effects resulting from ingestion of prescription pharmaceutical products. These responses are frequently absent in clinical trials and witnessed by patients in the real world. If identified at an early stage and corrected, it may be useful for other patients with the same disease. Several health departments in various countries seek to collect the patients' real-time drug reactions through polls, monitoring systems, etc. Patients' direct guidance on the basis of their experience is more important and helpful when making measures to reduce adverse drug reactions [11]. An ADR is an unwanted or dangerous reaction [7] that happens when a product or product combination is assembled under ordinary use conditions, and is believed to be linked to

the medication. The reaction could be a known side effect of the medication, or it could be recent and unrecognized beforehand. Fast identification and documentation of adverse drug reactions is important so that risks are immediately identified, and correct institutional motions are made to ensure proper use of solutions. Suspected ADRs should be accounted for by every friendly user, including medicines, hair, anticorps, radiographic distinction services, essential and home-developed products. Authors suggested the use of drug side effect biological network algorithm to discover side effect linked processes from the Drug Process-Side Effect Network [12–15].

3 Problem Statement

Drug abuse is a serious public health problem that effects almost every community and family in some way. It is usually celebrated and promoted on social media. In this proposed work, we try to find out tweets that are related to drug promotion. Here we aim to collect twitter data and predict if the user is promoting or consuming any illegal drugs.

4 Design

The software configuration consists mostly of 3 modules to manage tweets linked to the drug. They are divided into 4 pre-processing states where they are visualization, generate tokens, dealing with null values and split train test. When the user receives the feedback from twitter info, the algorithm first visualizes the tweets where the characters are interpreted and produces the necessary tokens as shown in Fig. 1.

The generated tokens are tested to insure that the null values in the datasets are tested. These datasets are trained to measure dividing the datasets step by step before they achieve the necessary accuracy limit. We train these datasets using 2 different algorithms approach they are as follows: (1) Train with Gaussian Naive Bayesian (NB) Algorithm (2) Train with Multinomial NB Algorithm. We check these data sets with these two approaches to the algorithm, from which we construct test models where the highest accuracy has been achieved as shown in Fig. 2.

The data for training the model is collected from the user input. The data is first loaded into the dictionary as the section where the data is washed is made where the letters, special characters and numeric are categorized and rendered into other sets of clusters. The distance of each word is separated or reduced to null to ensure that it is reprocessed. After a long-term analysis to predict the outcome using the Gaussian NB classifier and the multinomial naive bayes classifier, we get the appropriate outcome with the highest precision from testing the data sets that we can predict the performance as shown in Fig. 3.

From the used memory where the training data is loaded for the training phase. This loaded dataset is pre-processed, and the data collected is analyzed by replacing
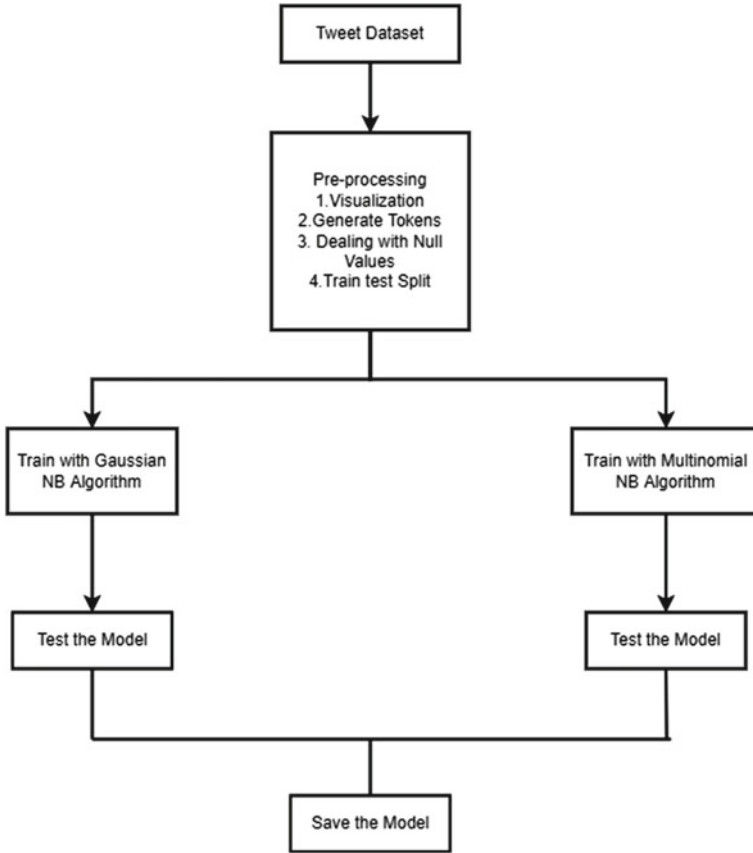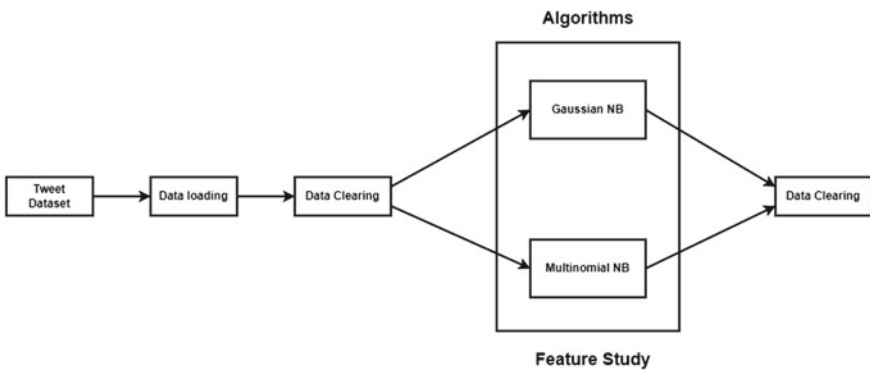
Fig. 1 System architecture diagram
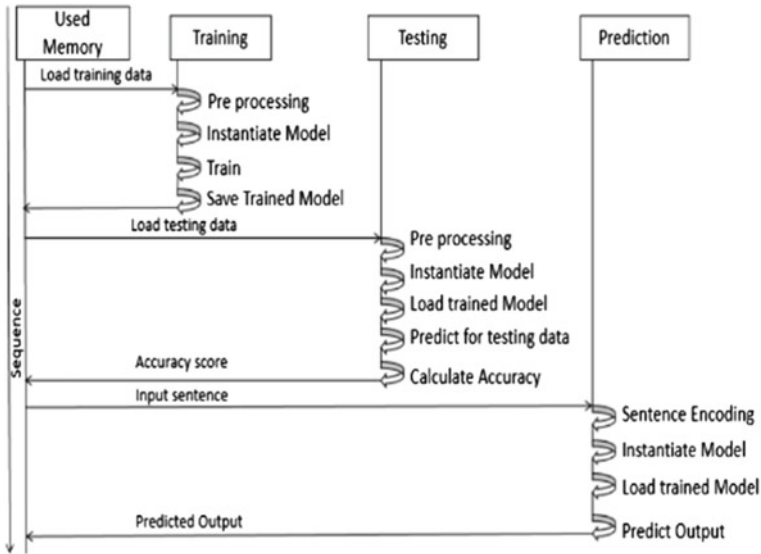

Fig. 2 Data flow design

Fig. 3 Sequence diagram for the proposed system

it with zero if there are any null values. Similarly Non-numeric data likewise is replaced with zero. Through this we get instantiate model, from which we practice this model twice when the precision limit is reached and save it as a trained model. Test data is loaded in the next step and re-process it for processing. The instantiate model is referenced and loads the trained model for training from which we can extract the estimation for testing results. After the accuracy score is obtained, we determine the precision of the trained data collected, then we return it to the memory sequentially. In the last step, the input sentence is loaded along with the trained model and the encoding sentences together with the instantiate model. The given formula is retraced with the trained formula from which we will determine the expected outcome if the consumer feedback facilitates drug use from the dates we received. In the usage case example where the interaction between the user and the web page takes place, the user enters the sentence in the text box whether the user submits and stores the document, the submitted content is evaluated and the backend interaction is calculated in order to validate the sentence. The stored sentence is analyzed and determined if it contains any special characters, which can be classified to be used if it can be predicted. The use of special characters can cause erroneous results which redirect back to the initial stage with appropriate error message. If the sentence is correct to check the sentence with Gaussian and multinomial naive bayes algorithms as shown in Figs. 4 and 5.

The flow diagram is like the rest of the previous diagram as usage case diagram. The user provides input data from twitter tweets i.e.: data collection [16, 17]. These data is lodged sequentially from the used memory. Pre-processing is initiated with the removal of the null value and non-numeric values. These training data are converted
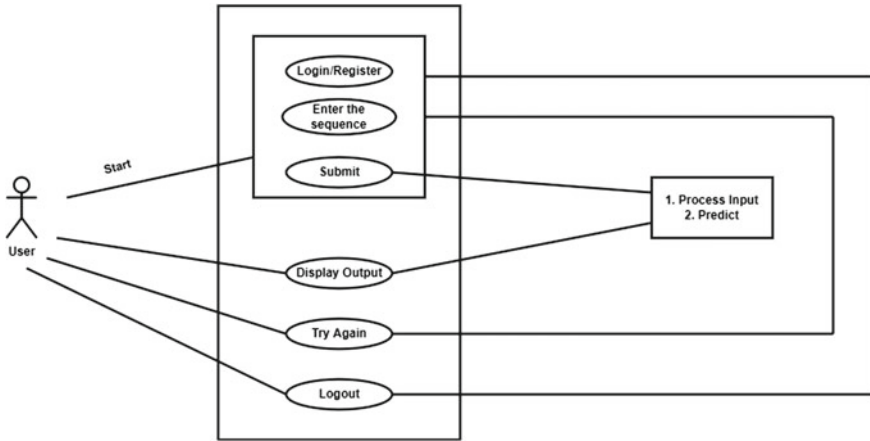
Fig. 4 Use case diagram for customer
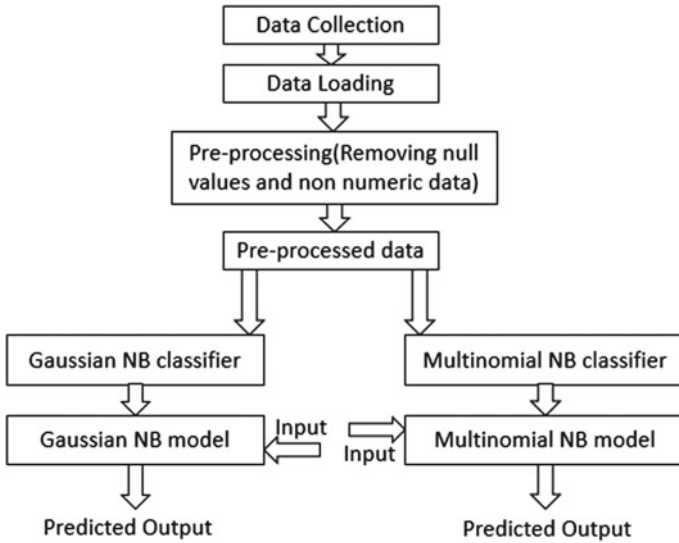
Fig. 5 Flow diagram of proposed work

to the trained model. The sentence is later checked with the Gaussian and Multino-mial Naive Bayes Classifier. During the process the highest number of accuracies is calculated from the previous installed model by giving user input.

5 Implementation

Implementation is the final step toward system development. The real execution of the specified requirements will take place during this phase, the model developed, algorithms developed, and various other factors needed to develop the system. The implementation phase leads to the conversion of the latest system design into a functional system. An implementation plan is drawn up before the system's actual execution is developed.

5.1 Pseudocode for Drug Prediction

```
def process(params):
  filename = 'trained_model.sav'
  clfMulti = pickle.load(open(filename, 'rb'))
  feature = pd.read_csv('train.csv')
  col = col.drop(['id', 'class'])
  tweet = params['text']
  tweet = tweet.split()
  inp = [0 for i in range(len(col))]
  for num,i in enumerate(col):
    cnt = 0
    for j in tweet:
      if i == j;
      cnt+ = 1
      inp[num] = cnt
  inp = np.array(inp).reshape(1,−1)
  ans = clfMulti.predict(inp)
  print(ans[0])
  if ans[0] == 'Y':
    return 'Yes, It is related to drugs!'
  else:
    return 'No, It is not related to drugs!'
```

5.2 Algorithm

1. Fetching of dataset from Twitter.
2. Normalization of dataset.
3. Building a model using the Gaussian NB Algorithm and Multinomial NB Algorithm.
4. Training the model using the training data.
5. In pre-processing, the fetched data is processed.

Table 1 Test cases

TC#	Description	Expected result	Actual result	Status
TC-1	User registration	Positive	Positive	Pass
TC-2	Login	Positive	Positive	Pass
TC-3	Login wrong password	Negative	Negative	Pass
TC-4	Train accuracy	Expected more than 80%	Positive	Pass
TC-5	Test accuracy	Expected more than 80%	Positive	Pass
TC-6	Entering submit without sentence	Error message	Positive	Pass
TC-7	Entering submit with sentence	Related to drug	Not related to drug	Fail
TC-8	Entering submit with sentence	Related to drug	Related to drug	Pass
TC-9	Entering submit with sentence	Not related to drug	Related to drug	Fail
TC-10	Entering submit with sentence	Not related to drug	Not related to drug	Pass

6. If there are any null values, it is replaced with zero.
7. Similarly non-numeric data is replaced with zero. Classifier Module.
8. Gaussian NB Algorithm.
9. Multinomial NB Algorithm.

5.3 Types of Testing

Unit testing here, using different modules, each part of the system is checked. From the system the output from each test case should be as expected. If errors are found, they should be corrected according to appropriate action. Once it is successful, different model will be built which will form the system. The integration testing here combines and tests individual models in group. The integration testing phase is tested by units tested and combined in unit test phase. In this phase compatibility of each is checked (Table 1).

6 Results and Discussion

The dataset for this project is a collection of tweets extracted from the twitter using twitter API. The overall dataset is split into training set and testing set. This step is necessary since we train the model on our train set and later test it using the unseen test dataset. By doing this, the model performance on the entirely new set of data is verified. Here train set comprises of 3166 samples and test set has 1076 samples.

That is the dataset is split in a 75–25% ratio. The confusion matrix is shown in Table 2. The accuracy, recall, precision, and specificity of different Naïve Bayes algorithms are shown in Fig. 6. Our study shows the Multinomial Naïve Bayes algorithm has better performance compared to Gaussian Naïve Bayes algorithm. The Multinomial NB algorithm provided accuracy around 89.86% whereas Gaussian NB model gave the accuracy of 88.3%.

Five parameters, accuracy, sensitivity, specificity, precision, and recall are used in experimental analysis where TP stands for true positive, TN is true negative, FP is false positive, and FN is false negative.

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{\text{TP} + \text{TN} + \text{FP} + \text{FN}},$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}},$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

Table 2 Confusion matrix for illegal drug prediction

Parameter	Gaussian NB	Multinomial NB
TP	912	936
TN	50	26
FP	75	83
FN	39	31
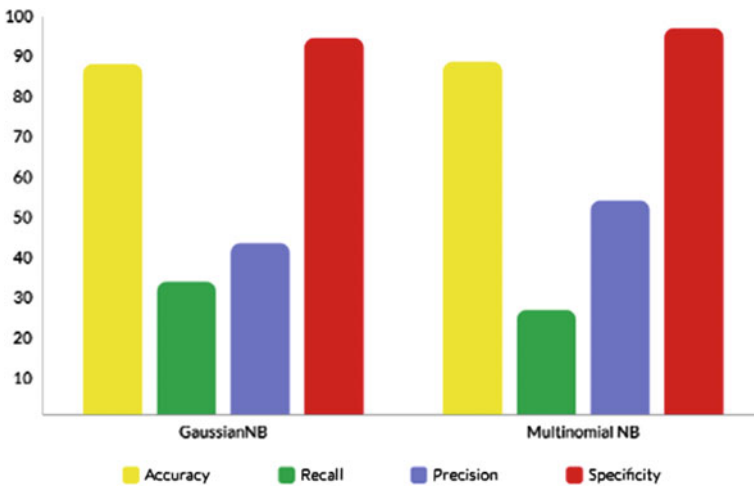
Fig. 6 Accuracy, specificity, sensitivity, precision, and recall for illegal drug prediction

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

Plays the user interface of the project where the user will register and login shortly after registration. Soon after the login the next page is open where the user can type the required sentences or tweets from twitter to analyze. Right below it the submit button is present which we can get the predicted output from the web. Output of the given input is displayed were it shows the no drug related content present.

7 Conclusion and Future Work

The proposed system will be able to accomplish the tasks such as effectively train the dataset, compare tweets that contain posts about drugs, predict users or promoters of illegal drugs on Twitter, and provide a graphical user interface that can display users who have tweeted about the use of illegal drugs. The machine learning algorithms such as Gaussian NB and Multinomial NB has been applied on the dataset and when results are compared it is found that Multinomial NB provided better accuracy. In future work more datasets can be collected and then experimented with various algorithms to find out the best among them.

References

1. Freifeld CC, Brownstein CC, Menone CM, Bao W, Felice R, Kass-Hout T et al (2014) Digital drug safety surveillance: monitoring pharmaceutical products in Twitter. Drug Saf 37(5):343–350
2. National Poisoning Data System (2017). http://www.aapcc.org/data-system/. Accessed 16 Jan 2017
3. US Food and Drug Administration (FDA) (2017). http://www.fda.gov/. Accessed 16 Jan 2017
4. Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality (formerly the Office of Applied Studies) (2010) The DAWN report: highlights of the 2009 drug abuse warning network (DAWN) findings on drug-related emergency department visits, Rockville, MD
5. Sarker A, O'Connor K, Ginn R, Scotch M, Smith K, Malone D, Gonzalez G (2016) Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from twitter. Drug Saf 39(3):231–240
6. Gittelman S, Lange V, Crawford CAG, Okoro CA, Lieb E, Dhingra SS, Trimarchi E (2015) A new source of data for public health surveillance: Facebook likes. J Med Internet Res 17(4):98
7. S. E. U. D. Process-Side. Drug-side effect network biological processes and side effects using drug process-side effect network
8. Eshleman R, Jha D, Singh R (2017) Identifying individuals amenable to drug recovery interventions through computational analysis of addiction content in social media. In: 2017 IEEE international conference on bioinformatics and biomedicine (BIBM). IEEE, pp 849–854
9. Ru B, Li D, Hu Y, Yao L (2019) Serendipity—a machine-learning application for mining serendipitous drug usage from social media. IEEE Trans Nanobiosci 18(3):324–334
10. Wu L, Moh TS, Khuri N (2015) Twitter opinion mining for adverse drug reactions. In: 2015 IEEE international conference on big data (big data). IEEE, pp 1570–1574

11. Sornalakshmi K, Sujatha G, Hemavathi D, Sindhu S. Detecting adverse drug reaction (ADR) mentions from social media
12. Shah N, Srivastava G, Savage DW, Mago V (2019) Assessing Canadians health activity and nutritional habits through social media. Front Public Health 7:1
13. Desai U, Nayak CG, Seshikala G, Martis RJ (2017 July) Automated diagnosis of coronary artery disease using pattern recognition approach. In: 2017 39th Annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp 434–437
14. Desai U, Nayak CG, Seshikala G (2017) Application of ensemble classifiers in accurate diagnosis of myocardial ischemia conditions. Prog Artif Intell 6:245–253
15. Desai U, Nayak CG, Seshikala G (2016) An application of EMD technique in detection of tachycardia beats. In: 2016 International conference on communication and signal processing (ICCSP), pp 1420–1424
16. Hanson CL, Cannon B, Burton S, Giraud-Carrier C (2013) An exploration of social circles and prescription drug abuse through twitter. J Med Internet Res 15(9):e189
17. Gold J, Pedrana AE, Sacks-Davis R, Hellard ME, Chang S, Howard S et al (2011) A systematic examination of the use of online social networking sites for sexual health promotion. BMC Public Health 11:583

Two-Wheeler Shock Absorber Analysis with Polyurethane Insert

V. G. Pratheep [iD]**, T. Tamilarasi** [iD]**, J. S. Joe Idaya Pranesh, D. Balachandar, A. Vimal Kumar, and S. Siva Subramanian**

Abstract Rider discomfort brought on the uneven road surfaces leads to back strain. Over time, the damping system will also sustain damage from these roadways. In the end, it makes long rides more difficult and necessitates the repair of the damping system. To give passengers a comfortable ride, the recommended methodology focuses on minimizing the impacts of cushioning and dissipating the force or kinetic energy generated. The purpose of this research is to help shock absorbers by filling the space between the suspension springs with polyurethane material. This enhances the effectiveness of the shock absorbers. The material's excellent plasticity makes it easier to create the necessary shape filling. In this research work, the polyurethane insert is to be developed and analysis the different property and placed inside the spring of the suspension.

Keywords Shock absorber · Suspension · Polyurethane insert

1 Introduction

A shock absorber is a component of the suspension system that works to lessen or neutralize the impact of absorbing bumps and vibrations on a vehicle [1–3]. It aids in enhancing the comfort and quality of the ride. Enhancing ride quality is essential to offering a comfortable ride. As can be seen in Fig. 1, the shock absorber design is crucial to the suspension system [4–7]. The shock absorber in the suggested system receives assistance from polyurethane material, which fills the space inside the suspension spring and contributes to a superior cushioning effect for a comfortable ride. The great plasticity of polyurethane aids in giving the filler the desired shape. Because the material and spring of the shock absorber in the suspension system are both elastic, the filling can then be placed and fitted exactly.

V. G. Pratheep (✉) · T. Tamilarasi · J. S. Joe Idaya Pranesh · D. Balachandar · A. Vimal Kumar · S. Siva Subramanian
Kongu Engineering College, Perundurai, India
e-mail: dkbalu4112@gmail.com

Fig. 1 Shock absorber

Figure 2 shows how uneven road surfaces create rider discomfort and back strain. This, in turn, deteriorates the damping system over time. It also causes discomfort after long rides and leads to the premature replacement of the damping system owing to damage.

The proposed solution is to aid the shock absorber by filling the gap inside the suspension with polyurethane material [8–10]. This contributes to a more comfortable riding experience while also increasing shock absorber efficiency.

Fig. 2 Riders back pain

2 Materials and Objective

Suspension springs are typically constructed of steel alloys. Chrome silicon, beryllium copper, chrome vanadium, low-tempered high carbon, and stainless steel are some of the most prevalent alloys. Polyurethane is utilized because it has a high yield strength and can help the spring.

3 Method

Solid Works is utilized for shock absorber design and analysis. The realized idea is to help shock absorbers by using polyurethane material as a filler component inside the suspension spring gap. This serves to absorb impulse force caused by uneven roadways and effectively dissipate kinetic energy to give a smooth riding experience while also boosting shock absorber effectiveness. The major goal is to create and analyze a shock absorber that gives a better riding experience for passengers while also increasing the suspension system's lifetime. To achieve our goal, we use polyurethane, which has good elastic properties and can absorb and release impulsive forces. The proposed concept is to use polyurethane material to assist shock absorbers and improve shock absorber efficiency. The polyurethane component will be installed beside the springs. The extraordinary versatility of polyurethane material allows the filler component to be manufactured in the desired shape. Because of the elastic properties of both the suspension's material and spring, this filler component is exactly positioned and fitted. In harsh conditions, this design remains stable with moisture and swells minimally. The use of polyurethane as a shock absorber provides superior cushioning against unexpected damping. Elastomeric polymers include polyurethane. The material is extremely versatile, long-lasting, flexible, adaptable, and tough. Polyurethane has a wide temperature range (-62 to $150\,^{\circ}C$) and is resistant to water, oil, and grease.

4 Model

See Fig. 3.

5 Analysis Procedure

The proposed process is to investigate the materials, pick materials based on qualities, design the component with an efficient design by calculating the parameters listed below, and then test the simulated results with ANSYS. The simulation analysis

Fig. 3 Shock absorber

Insert (Polyurethane)

Spring (Beryllium copper)
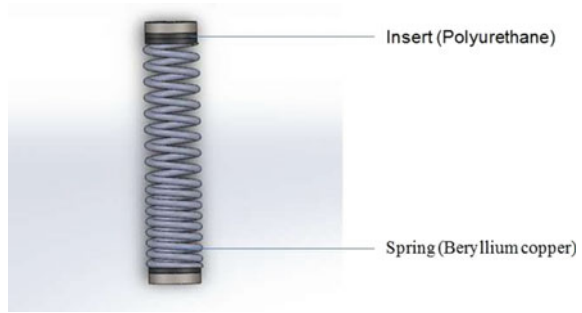
determines the functional property of the designed component and compares it to the findings of the current structure.

5.1 Calculation

Material: Beryllium copper (modulus of rigidity) $G = 50{,}000 \, \text{N/mm}^2$. Mean diameter of a coil, $D_{\text{m}} = 33.3$ mm.

Diameter of wire, $d = 6.7$ mm. Total no. of coils, $n_{\text{t}} = 17$.

No. of active turns, $n = 15$.

Outer diameter of spring coil, $D_0 = D_{\text{m}} + d = 40$ mm; weight of bike $= 113$ kg.

Let weight of one person $= 75$ kg. Weight of two persons $= 75 \times 2 = 150$ kg. Total weight (T),

$$T = \text{Weight of bike} + \text{persons} = 263 \text{ kg}$$

Weight act on rear suspension $= 65\%$ of total weight 65% of $263 = 171$ kg.

Considering dynamic loads, it will be double, $W = 342$ kg $= 3355$ N.

For single shock absorber weight, $W = W/2 = 1677$ N.

Spring index, $C = D_{\text{m}}/d = 5$.

Free length of the spring (L_{f}) $= 210$ mm.

Solid length of the spring (L_{s}) $= d \times n_{\text{t}} = 6.7 \times 17 = 114$ mm. Wahl's stress factor (k),

$$k = (4C - 1)/(4C - 4) + 0.615/C = 1.219.$$

Deflection of spring under load (δ),

$$\delta = (8WC^3n)/(G.d) = 75.08 \text{ mm}.$$

Max. shear stress in the spring (τ),
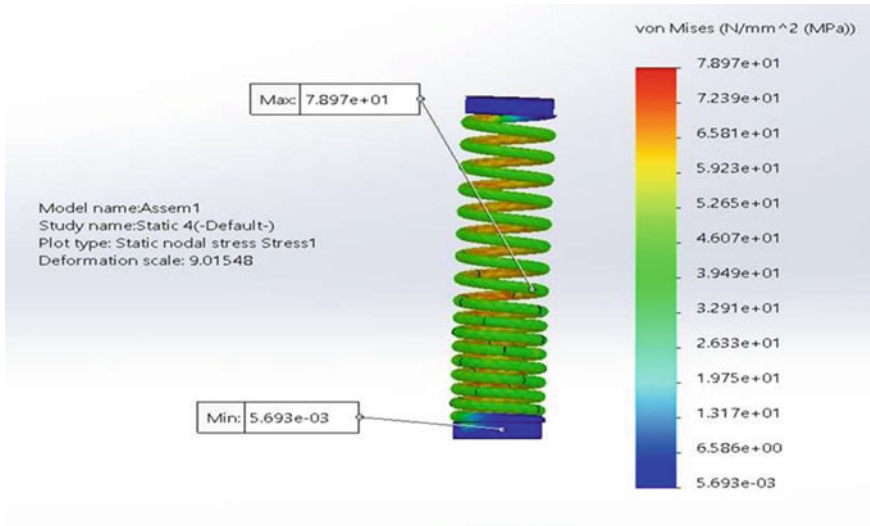
Fig. 4 Stress analysis without insert

$$\tau = (k8WD\_O)/(\pi.d^3) = 692.32 \text{ MPa}.$$

Spring stiffness (K),

$$K = (Gd^4)/(8D_m^3 n) = 22.676 \text{ N/mm}^2.$$

5.2 Stress Analysis Without Polyurethane

Figure 4 shows an applied force of 100 N without a polyurethane insert. The maximum tension on the spring is 78.9 MPa. The spring's minimum stress is 56.9 MPa.

5.3 Stress Analysis with Polyurethane

Figure 5 shows a 700 N applied force with a polyurethane insert size of 10 mm on top and 5 mm on bottom of the spring. The maximum spring tension is 73.9 MPa. The spring's minimum stress is 0.00359 MPa. When compared to earlier stress analyses, the model developed with the insert is significantly lower.
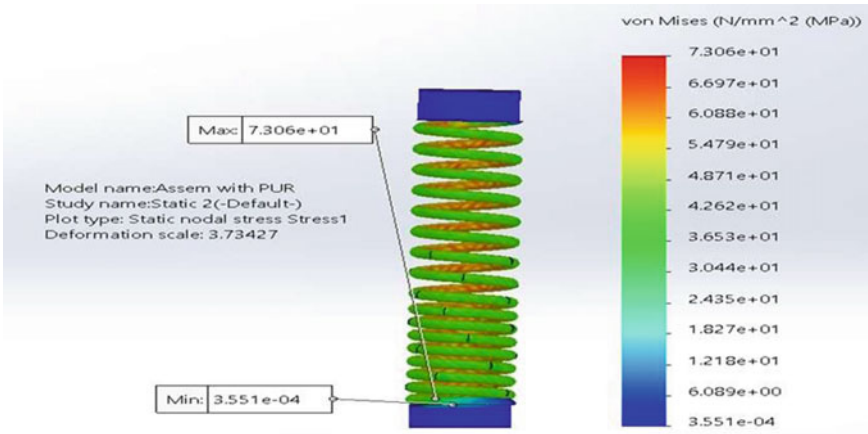
Fig. 5 Stress analysis with insert

Fig. 6 Deformation analysis without insert
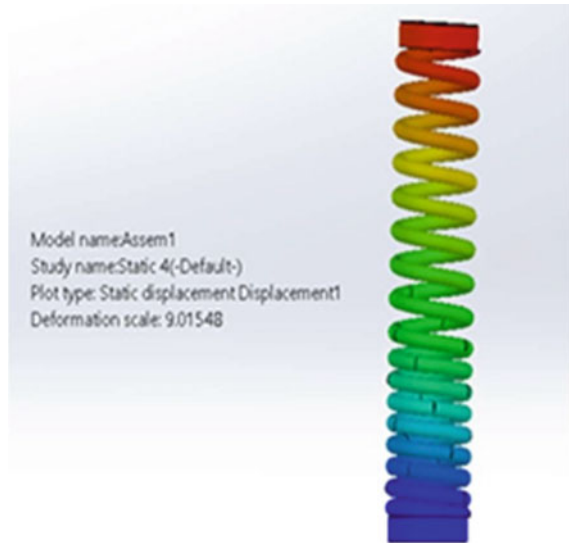
5.4 Deformation Analysis Without Polyurethane

5.5 Deformation Analysis with Polyurethane

When comparing Figs. 6 and 7, it is clear that the distortion in Fig. 6 is less severe. As a result, it is clear that the proposed model produces less deformation (Fig. 8).
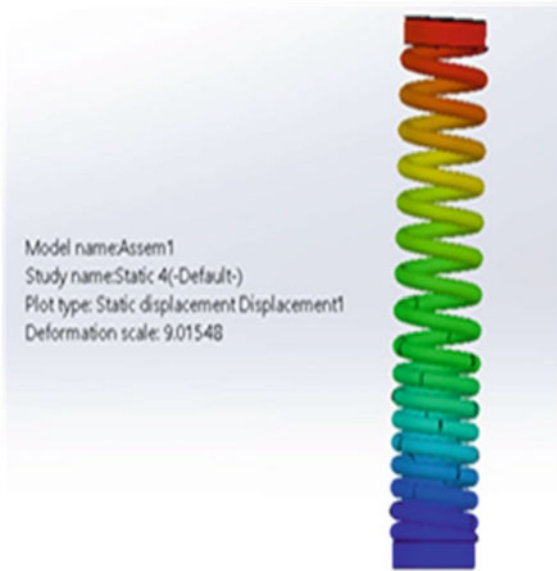
Fig. 7 Deformation analysis with insert

Fig. 8 Stress analysis with insert

Model name:Assem with PUR
Study name:Static 2(-Default-)
Plot type: Static displacement Displacement1
Deformation scale: 7.09511
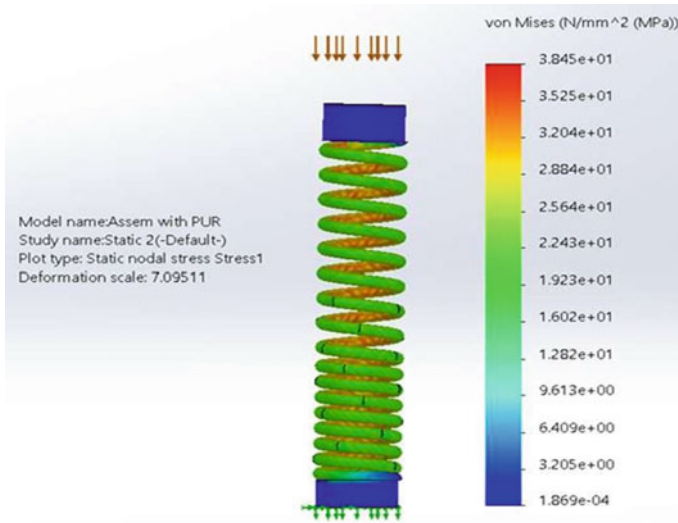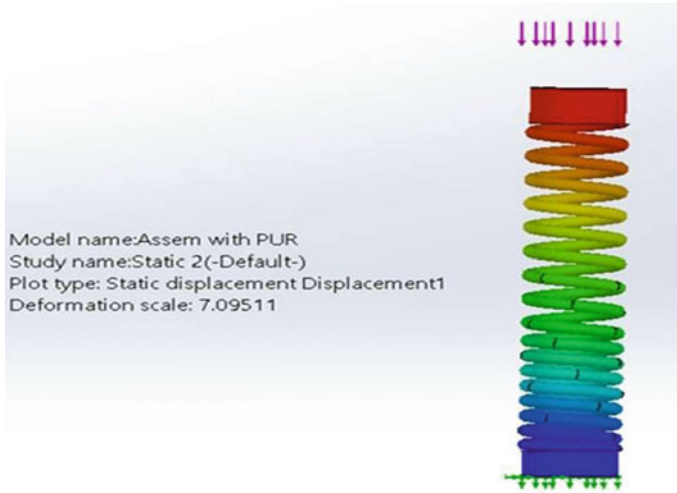
Fig. 9 Deformation analysis with insert

5.6   Stress Analysis with Higher Thickness of 10 mm at One Side

A 10 mm insert is used on the bottom end, and a 700 N force is applied. When compared to Fig. 5, this causes more stress. As a result, this is not ideal for a two-wheeler.

In Fig. 9, a 10 mm insert is put on both sides of the spring. The spring is subjected to a 700 N force. This results in 86 MPa stress, which is higher as shown in Fig. 5.

6   Results and Discussions

When the stress created on the spring with and without polyurethane is compared, the stress created on the spring with polyurethane is less than that created without the insert. This demonstrates that the insertion of a polyurethane insert significantly reduces the abrupt damping force and provides a considerably smoother ride. When the deformation scale with and without polyurethane is compared, it is clear that the spring with polyurethane insert has less deformation, which will aid in a steady and smooth ride as well as the ability to withstand rapid damping force.

7 Conclusion

Observing the results, it is apparent that the insert of polyurethane material reduces the tension on the spring and allows it to sustain higher levels of stress. As a result, injecting polyurethane can improve the shock absorber. Finally, the shock absorber with polyurethane insert is more effective and efficient.

8 Future Scope

An environmentally friendly substance can be used instead of polyurethane. The main problem in implementing the proposed model is maintaining homogeneity in polyurethane material. Future versions can be modified to use polyurethane without disassembling the shock absorber.

References

1. Manjunatha TS, Abdul Budan D (2012) Manufacturing and experimentation of composite helical springs for automotive suspension. Int J Mod Eng Res 1(2)
2. Raghu Kumar B, Viajay Prakassh V, Ramesh N (2013) Static analysis of mono leaf spring with different composite materials. J Mech Eng Res 5(2):32–37
3. Sidharamanna G, Shankar S, Vijayarangan S (2006) Mono composite leaf spring for light weight vehicle—design, end joint analysis and testing. Mater Sci 12(3)
4. Mehdi B, Majid B (2012) Optimization of steel helical spring by composite spring. Int J Multi Sci Eng 3(6)
5. Pyttel B, Ray KK, Isabell B, Tiwari A, Kaoua SA (2012) Investigation of probable failure in helical compression springs used in fuel injection system of diesel engines. J Mech Eng Civ Eng 2(3):24–29
6. Kumar K, Aggarwal ML (2012) A finite element approach for a multi leaf spring using CAE tools. Res J Recent Sci 1(2):92–96
7. Singh N (2013) General review of mechanical springs used in automobiles suspension system. Int J Adv Eng Res Stud 115–122
8. Subramaniam T, Velliangiri D (2016) A study of modeling and finite element analysis of automotive vehicle wheel rim assembly for the deformation and various stress condition. Int J Adv Eng Technol 7(1):23–29
9. Shivaraj Singh W, Srilatha N (2018) Design and analysis of shock absorber: a review. Mater Today Proc Part 1 5(2):4832–4837
10. Adhith Kumaar SB, Pushkaran S, Puviyarasan M, Kabilan P (2018) Displacement and stress analysis of suspension system using ANSYS. Int J Eng Res Technol 7(3):283–286
11. Tamilarasi T, Varun S, Saravana Kumar M, Sivaprakash S (2016) Design and development of pipeline cleaning robot. Int J Eng Res 5(4):315–317
12. Shanmugam A, Muralidharan N, Pratheep VG, Thangarasu SK, Tamilarasi T, Dhileephan K (2021) Design and fabrication of paver block laying machine. Mater Today Proc
13. Pratheep VG, Tamilarasi T, Ravichandran K, Shanmugam A, Thangarasu SK, Prenitha A (2022) Design and development of medicine retrieval robot for pharmaceutical application. In: Computational intelligence in machine learning. Springer, Singapore, pp 301–307

14. Pratheep VG, Tamilarasi T, Dhileephan K, Antony AJ, Heeraj A, Akash M (2021) Design and fabrication of multitasking transit robot. IOP Conf Ser Mater Sci Eng 1057(1):012034
15. Thangarasu SK, Thomas AT, Shanmugam A, Pratheep VG, Tamilarasi T (2021) Design and fabrication of automatic banana leaf cutting machine. Mater Today Proc

Design of a Novel Smart Device for Human Safety

Mukul Anand, Zeeshan Naseem, Shikhar Trivedi, Abhishek, and Pardeep Garg

Abstract In this paper, a novel idea of enhancing human security by introducing a smart band that can sense the heartbeat and motion of the victim has been proposed. With the help of this design, the live location of the victim can be sent to the Save Our Soul (SOS) contacts. A few devices that work on this technology are having the limitation in the sense that they do not have a proper false alarm control functioning. A proper false alarm control function is designed in the proposed work so that the receiver will get only appropriate information. Another key contribution is that the designed device is in the form of an ornament wearable by both men and women, which makes it less suspicious and high on demand. This device is the need of humankind as the criminal cases are increasing day by day, and it can protect the victim from criminals and the mishaps. The designed low-cost anti-crime decorative band using System on Chip (SOC) technology has a Wi-Fi module to monitor heart rate, movement, and location of the person via sensors. A message on the saved device in the event of the tragedy can be sent via this device, and the victim's live location can be tracked via Global Positioning System (GPS). It focuses on people of 8+ age groups who may get involved in mishaps. These situations can be mitigated by this crime adornment (belt).

Keywords Node MCU microcontroller · Heart rate sensor · Blynk API · GPS module · Vibration sensor

1 Introduction

Over the years, there have been many incidents of crime and violence against men and women. Cases like robbery, rape, kidnapping, sexual assault, and murder are increasing at an alarming rate. In many cases, criminals shoot the victim in a remote

M. Anand · Z. Naseem · S. Trivedi · Abhishek · P. Garg (✉)
Department of Electronics and Communication Engineering, Jaypee University of Information Technology, Waknaghat 173234, India
e-mail: garg.pardeep22@gmail.com

or isolated area leaving the victim defenseless [1, 2]. It is clear that there is a definite need for the protection all across the world. It is now possible to use the benefits of current technology wisely to solve social problems. There is a need for the use of the latest technologies, the Internet of Things (IoT) to eradicate the fear-filled lifestyle of people [3]. IoT is a natural system of connected objects accessible online. Various smart apps such as smart watches and mobile-based application have been introduced to empower human safety [4, 5]. In this paper, a wearable device based on Node MCU called smart ornament to protect humans in a remote or crowded place is presented. The main purpose is to send an urgent message and help message to SOS contacts and the police, in order to avoid wrongdoing and to help the needy person by providing the real-time evidence for immediate action against perpetrators or crime [6, 7]. However, it is an important point to note that new technologies have influenced humans in almost every phase of life. Thus, it is now possible to use the benefits of current technology wisely to solve social problems. In today's world, an increasing number of criminal cases are taking place. According to some data, every 3 min a chain snatching, molestation, or a break-in is taking place in India. However, this is just the recorded data and God knows its truth, so scientists should focus more on improving people's safety. There is always need of such devices which can track the victim's location in an emergency [8–10]. The solution in the form of a smart device to help the society has been proposed in this paper. The proposed work's design contains sensors in an ornament that will protect victims. The sensors used in this Arduino technology often monitor the heart rate, motion, and location of the victim. This plays a major role in preventing crime, murder, and rape in the near future. Heart rate sensor is used to detect the heart rate of the victim. It also locates people in danger using GPS. This device uses the cloud to send messages to the preset SOS contacts of the victim. By using this technology, persons who are in danger can be detected; their location can be accessed even if their phones are confiscated [11–15]. The organization of the paper is as follows: Sect. 1 provides a brief introduction to human's safety and status. The related work corresponding to safety strategies and strategies used in the past has been discussed in Sect. 2. The proposed approach has been presented in Sect. 3. Section 4 represents the results and discussion. The final phase is shown in Sect. 5 which concludes the proposed project and suggests an open space for future development in the system.

2 Related Work

Many researchers have worked on the use of Internet of Things (IoT) that leads to the smart security technology for human safety. This section presents the related technological work carried out for the human security, their methodologies, and their drawbacks. Karmakar et al. have proposed a system involving an EMG sensor and pulse sensor named SafeBand. These bands are capable of sending SOS signals after decoding the movement of a muscle caused by hand gestures. SafeBand works on EMG muscle sensor, pulse sensor, Arduino Uno, and ESP8266 Wi-Fi module and

is powered by a rechargeable battery; their drawback is that they have a low-end pulse sensor that needs to be connected to the fingertips which makes it hard to use and wearing EMG sensor for too long can cause bleeding [1]. Harikiran et al. have proposed an IoT smart band that is preloaded with the human's reaction like anger, fear, and anxiety which will send a preloaded message if there is any behavioral change; their drawback is that everything is preloaded, but things can vary when the real situation arrives [2]. Miriyala et al. have proposed an IoT system with tear gas and camera containing goggles for the live streaming that is easily detectable by any person, and it will put the victim in danger [3]. Helen et al. have proposed a system to detect the heart rate by observing epinephrine hormone through which the device gets activated, and it will make a call to the SOS contact. Calling feature takes time to connect and may not be identified by the emergency contact person that this is an emergency call; it can be hazardous for the victim [4]. Thiyagarajan and Ravendra have proposed an IoT-based band for the women safety which will be triggered by pressing a button on the device. The drawback of the idea is pressing any button in that situation will alter the offender, and they can drop off the device [5]. Eeshwaroju et al. have proposed a solution for smart healthcare system, smart education system, and smart jobs system; the proposed work is mainly concerned to the health monitoring and education system [6]. Raju et al. have proposed a smart home design to control the electronic appliances in home and forth remotely using smart phone by using Blynk android app which gave us the idea for automation [7]. Raksha et al. have proposed a system that depends on Raspberry Pi, pulse sensor, temperature sensor, GPS, and camera that will measure the vital changes and the predefined value which will send an alert message using Wi-Fi module if the predefined value of the sensors will increase; their drawback is that the values are predefined which is not feasible as it can differ from person to person [8]. Said et al. have proposed a system based on advanced wireless connections, wearable sensors, and machine learning technologies. It detects the abnormal behavior by detecting heart rate increase and abnormal motion; the main focus of the proposed design is on crowdsourcing real-time detection method for abnormal crowd behavior in mass gatherings. The drawback of this method is that the data received by the proposed design can easily be manipulated as the mobile crowd sensing detects the mobile network traffic at a particular area [9]. Venkatesh et al. have proposed a contraption which is a consolidation of different devices as a wearable "wrist band" that unendingly talks with a sensible phone that moves toward the net; when a woman senses danger during this endeavor, she needs to keep the trigger of the contraption close. The drawback of this approach is that the user has to tap twice on the screen in order to send the location and message which makes it a bit difficult for a person who cannot reach their wrist [10]. Islam et al. have proposed an intelligent collaborative security model to minimize security risk at the healthcare centers, but the proposed system is inconsistent [11]. The proposed method in this paper will overcome some of the limitations of existing methods by the alarm raising technique. Existing methods have used a switch or a screen to be tapped in order to raise the alarm, whereas the proposed design can raise the alarm by simply monitoring the vitals or by doing a gesture. Existing methods have a big limitation to stop the false alarm, whereas the proposed design gives its user a span

of 15 s between triggering and sending the SOS message if the device is triggered
by mistake, they can simply stop it by making a gesture, and the SOS message will
not be sent. Existing devices make a beep sound between triggering and sending
the SOS message which can alert the criminal, whereas the proposed design only
vibrates between the triggering and sending of SOS message which cannot raise any
alert to the criminal.

3   Proposed Work

This paper proposes a very effective way to solve human security-related problems.
Initially, other band or safety devices had some limitations because of the false alarm,
and they were restricted either to healthcare or women safety. The proposed device
is unisex and easy to operate. The other devices are in the form of smart watches,
but proposed device is in an ornament, thus avoiding the attention of the criminals.
The block diagram of proposed design is shown in Fig. 1.

Node MCU is an open-source platform based on ESP8266 which can connect
objects and let data transfer using the Wi-Fi protocol. In addition, by providing some
of the most important features of microcontrollers such as general-purpose input–
output (GPIO) pins, pulse width modulation (PWM) pins, and analog to digital
converter (ADC), it provides more options for other useful sensors. The MPU6050
sensor module is a complete gyroscope device for tracking six-axis movements.
It includes a three-axis gyroscope, three-axis accelerometer, and digital motion
processor all in a small packet. It has an inter-integrated circuit (I2C) bus interface
for microcontrollers. The basic heart rate sensor consists of a light emitting diode
(LED) and a detector such as a light resistor or photodiode. The heartbeat causes
fluctuations in blood flow to various parts of the body. When tissue is illuminated by
a light source, i.e., LED light, the reflected light is received by the light detector. The
output of the detector is in the form of an electrical signal and is equal to the heart
rate. The NEO-6M GPS module is a well-performing complete GPS receiver with a
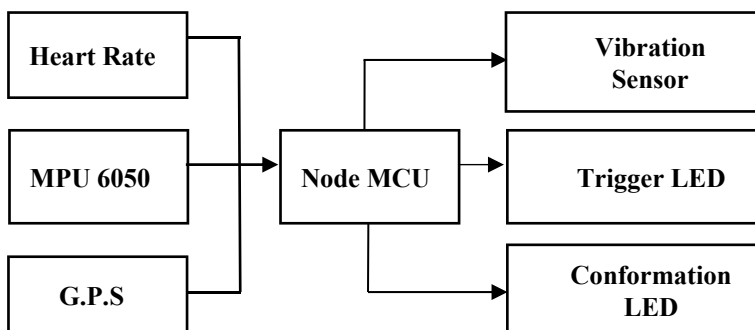
Fig. 1 Block diagram of proposed design

built-in $25 \times 25 \times 4$ mm ceramic antenna, which provides a strong satellite search capability. With the power and signal indicators, the location of this module can be monitored.

In the proposed system, Node MCU has been used as main controller; the heart rate sensor is connected to the Node MCU to detect the heart rate of the victim. The MPU6050 gyroscope is used to track the wrist's motion and GPS to detect the location of the bearer. When the panic state is detected, an alarm is triggered. The bearer can also trigger the alarm manually by rotating their wrist three times. The connected vibration sensor will vibrate, and first LED will glow indicating that the panic state alarm has been triggered. The bearer can cancel the alarm within 15 s if it is falsely triggered. After 15 s, in case of non-cancelation of the alarm, a SOS message embedded with the live location of the bearer will be sent to their registered SOS contacts. The acknowledgment of sent SOS message will be delivered to the victim by the flash of second LED.

4 Results and Discussion

In this modern world, the street crimes are increasing day by day. The recent data shows that the street crimes had a reported increase of 44%. When any mishap happens, every human goes through a panic state. This panic state needs to be detected and analyzed for any potential mishaps. The heart rate sensor in the proposed design constantly monitors the heart rate of the person. In case of any panic situation, the elevated heart rate of the person is detected; this initializes the panic mode of the band's software. The bearer can trigger the alarm by rotating their wrist three times. The rotation motion is registered and analyzed using the MPU6050 gyroscope sensor. If three continuous rotations of the wrist are registered, then it is a case for potential danger. However, it might be possible that the bearer might have triggered the alarm involuntarily. This is a case of false alarm and will cause unnecessary resource and personnel allocation that can be allotted to some good use. To avoid such a scenario, Node MCU controller waits for 15 s. In these 15 s, the controller constantly flashes one LED and vibrates the vibration sensor which creates the haptic sensation to the bearer that the alarm might have been triggered. If the bearer might have accidentally triggered the alarm, then they can disable it by the alarm status button present on the Blynk interface. In case of a genuine alarm trigger, after the 15 s wait timer, an SOS message with the bearer's live location will be sent to the registered SOS contacts of the bearer. The confirmation of the message delivery will be provided to the bearer by flashing of the second LED three times. The prototype of proposed design is shown in Fig. 2.

When both the conditions are satisfied (rotating wrist three times and increased heartbeat), then the panic state will be triggered and a timer of 15 s will start with periodic vibrations. The graph for the increased heartbeat and alarm trigger point is shown in Fig. 3.
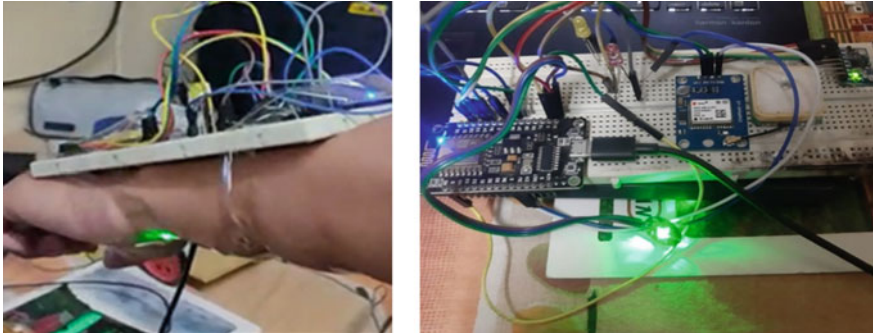
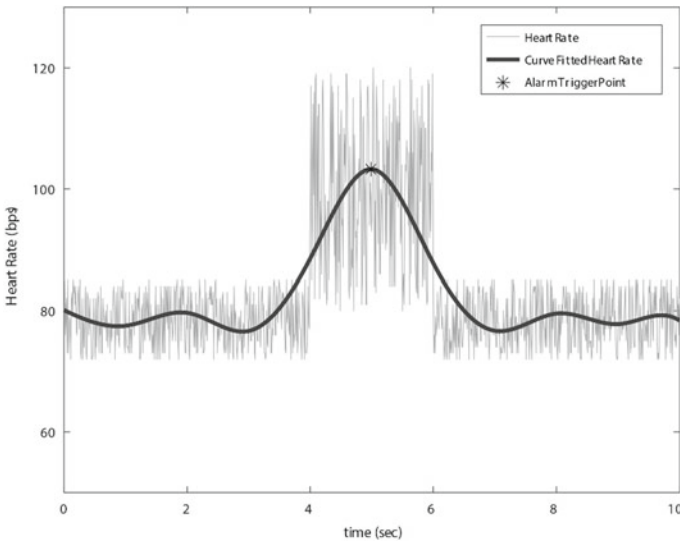**Fig. 2** Prototype of the proposed design

Fig. 3 Graph for increased heartbeat and the alarm trigger point

The increased heart rate graph shows the triggering of the device, and trend line indicates the raised heartbeat in panic situation. And in the case of false alarm, the bearer can switch it off by using the push option sent via the Blynk cloud server (the option will be displayed on the smart band to turn the alarm off). After the 15 s timer if the alarm is not turned off, it will send a SOS message with the victim's live location to the present SOS contacts and the second LED will flash. The Blynk application interface is shown in Fig. 4.

If the notification is left unseen, the live location of the bearer will be sent to the SOS contacts. Received E-mail after the panic state detection is depicted in Fig. 5.

Coordinates of the victim are shown on Google Maps as represented in Fig. 6.
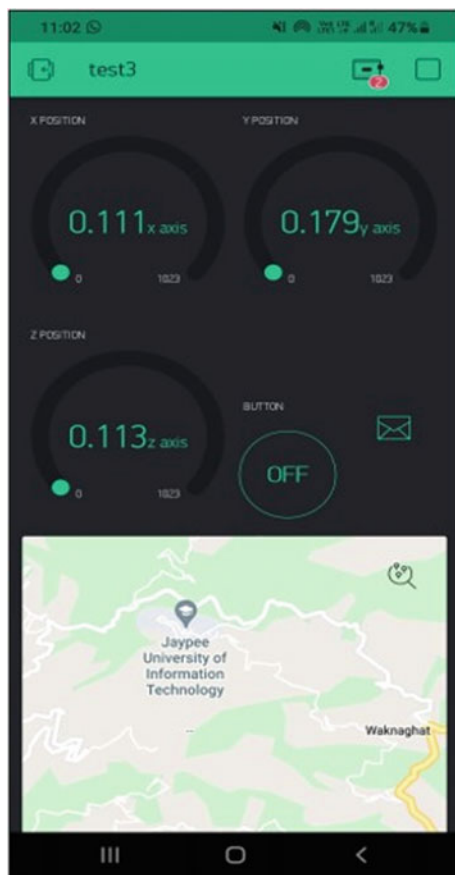
Fig. 4 Blynk application interface
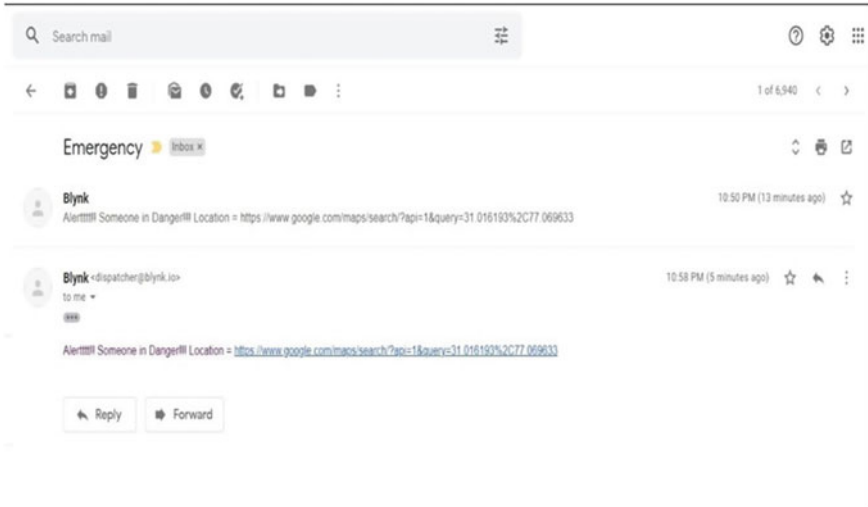
Fig. 5 Received E-mail after the panic state detection
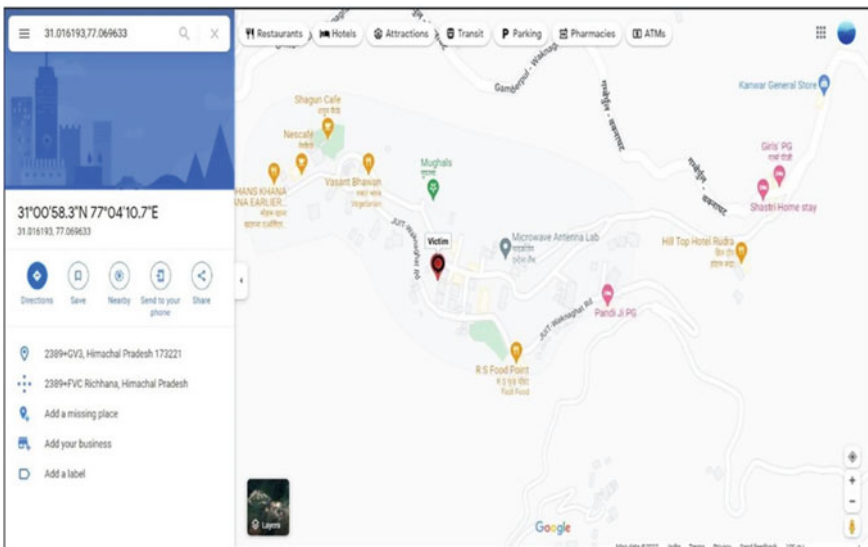
Fig. 6 Coordinates of the victim

5 Conclusion

The idea presented in this paper will effectively help in declining criminal cases like street robbery and other heinous crimes. The anonymity of a security band is a must so that it avoids the attention of the criminals. The proposed design can be

miniaturized and concealed into an ornament, thus making it invisible to the criminal. The efficient use of power is a major factor because the device should function the whole day. The proposed design is power efficient as it only uses the specific sensors when they are required, thus eliminating extra power consumption making the device last longer. It can be concluded from the paper that the designed low-cost anti-crime ornament can help in reducing the crime rates significantly due to its concealable design and efficient power management. It can also be used as a remote for home automation and a health device. The future work can be the miniaturization of the proposed design into other ornaments like ring and goggles.

References

1. Karmakar A, Ganguly K, Banerjee PS (2022) SafeBand: IoT-based smart security band with instant SOS messaging. In: Mandal JK, Buyya R, De D (eds) Proceedings of international conference on advanced computing applications. Advances in intelligent systems and computing, vol 1406. Springer, Singapore

2. Harikiran GC, Menasinkai K, Shirol S (2016) Smart security solution for women based on internet of things (IOT). In: International conference on electrical, electronics, and optimization techniques, pp 3551–3554

3. Miriyala GP, Sunil P, Yallapalli RS, Pasam VRL, Kondapalli T, Miriyala A (2016) Smart intelligent security system for women. Int J Electron Commun Eng Technol 7(2):41–46

4. Helen A, Fathila MF, Rijwana R, Kalaiselvi VKG (2017) A smart watch for women security based on IoT concept 'watch me'. In: 2nd International conference on computing and communications technologies, pp 190–194

5. Thiyagarajan M, Ravendra C (2015) Integration in the physical world in IoT using android mobile application. In: International conference on green computing and internet of things, pp 8–10, Oct 2015

6. Eeshwaroju S, Jakkula P, Ganesan S (2020) IoT Based empowerment by smart health monitoring, smart education and smart jobs. In: International conference on computing and information technology, pp 1–5

7. Raju KL, Chandrani V, Begum SS, Devi MP (2019) Home automation and security system with node MCU using internet of things. In: International conference on vision towards emerging trends in communication and networking, pp 1–5

8. Raksha S, Reddy YR, Meghana EI, Reddy KM, Panda PK (2021) Design of a smart women safety band using IoT and machine learning. Int J Contemp Archit 8(1)

9. Alsalat G, Said K, El-Ramly M, Aly A (2018) Detection of mass panic using internet of things and machine learning. Int J Adv Comput Sci Appl 9(5)

10. Venkatesh K, Parthiban S, Kumar PS, Vinoth Kumar CNS (2021) IoT Based unified approach for women safety alert using GSM. In: Third international conference on intelligent communication technologies and virtual mobile networks, pp 388–392

11. Islam SMR, Kwak D, Kabir MH, Hossain M, Kwak KS (2015) The internet of things for health care: a comprehensive survey. IEEE Access 3:678–708

12. Sethuraman TV, Rathore KS, Amritha G, Kanimozhi G (2019) IoT Based system for heart rate monitoring and heart attack detection. Int J Eng Adv Technol 8(5). ISSN: 2249-8958

13. Bhardwaj V, Joshi R, Gaur AM (2022) IoT-Based smart health monitoring system for COVID-19. SN Comput Sci 3:137

14. Raguvaran K, Thiyagarajan J (2015) Raspberry PI based global industrial process monitoring through wireless communication. In: International conference on robotics, automation, control and embedded systems, pp 1–6

15. Manocha A, Kumar G, Bhatia M (2021) IoT-Inspired machine learning-assisted sedentary behavior analysis in smart healthcare industry. J Ambient Intell Humanized Comput

Absorbance and Emission Studies of ZnO Nanostructures

Priyanka Sharma, Sanjiv Kumar Tiwari, and Partha Bir Barman

Abstract Zinc Oxide nanoparticles (NPs) and nanorods (NRs) were prepared by the wet chemical technique at low temperatures. The morphology of the obtained samples was depicted from the field-emission electron microscopy (FESEM) images. The optical studies such as absorbance and emission were conducted. Thereafter, it was concluded that the morphology has negligible impact on the absorbance spectra but significantly affects emission spectra's intensity ratios.

Keywords ZnO · Seeds · Nanoparticles · PL emission · Absorbance · Nanorods

1 Introduction

ZnO is extensively studied for its electrical, photoelectronic, catalytic, and photochemical capabilities, which include sensors, hydrogen storage devices, ultraviolet (UV) laser diodes, and UV optical detectors [1–4] due to its unique optoelectronic properties. At ambient temperature, ZnO has wide band gap energy of 3.37 eV and a large exciton binding energy of 60 meV. For the development of ZnO nanostructures (NSs), a variety of synthesis techniques are available, including chemical synthesis, chemical vapor deposition, pulsed laser deposition [5, 6], electron beam sputtering, etc. [7–9]. ZnO is chemically and thermally stable n-type semiconductor. It is an intrinsic semiconductor with a variety of defect states present within its band gap. Due to the presence of defects, it has a wide range of applications. The presence and the concentration of defects affect largely the emission properties of the material. The defect-related emission studies of ZnO NRs at various low-temperature ranges are also documented in the literature [10]. Zhang et al. in their review depict the synthesis and the application of the UV and Vis defect-related emission optical properties of the ZnO in the field of biomedical sciences such as drug delivery, biological imaging,

P. Sharma · S. K. Tiwari · P. B. Barman (✉)
Department of Physics and Materials Science, Jaypee University of Information Technology, Waknaghat, Solan, H.P. 173234, India
e-mail: pb.barman@juit.ac.in

antibacterial activity [11]. Further, Haque et al. and other researchers performed the synthesis and photocatalytic activity of the ZnO NPs toward methylene blue dye, which is very harmful to our environment [12, 13].

In the present work, the synthesis and optical properties such as absorbance and emission of ZnO NPs and NRs were presented. The novel aspect of the current work is that morphology has little impact on ZnO absorbance spectra whereas has a large influence on defect intensity ratios in PL-emission spectra.

2   Materials and Methods

ZnO nanoparticles and nanorods were prepared by the wet chemical method. Firstly, Zinc Oxide NPs (seeds) were prepared by thermal decomposition of the zinc acetate hexahydrate in ethanol as per the following chemical equation:

$$Zn(CH_3COO)_2.2H_2O + CH_3CH_2OH \rightarrow ZnO(s) + H_2O. \tag{1}$$

Thereafter, another nanostructure which is nanorods was prepared by using zinc nitrate hexahydrate as a source for Zn^{2+} ions and HMTA as a capping agent and a source of OH- ions. It was due to the presence of HMTA, that the morphology of the nanowires was obtained. Growth time and temperature were both fixed to 8 h and 90 °C, respectively. The complete synthesis process involved in the present study is described in the literature [14].

3   Results and Discussions

3.1   *ZnO Nanoparticles (NPs)*

3.1.1   Absorbance Spectroscopy

Figure 1a represents the FESEM image of the ZnO NPs prepared by the above-mentioned wet chemical method. FESEM images depict the clear spherical morphology of the NPs, and the NPs were observed to be clustered together somewhere in the sample. Thereafter, the distribution of the NPs was obtained using ImageJ software. Figure 1b represents the histogram (distribution) plot for the NPs of the ZnO. The average particle size depicted in the histogram was 26.2 nm.

Further, the absorbance spectra of the NPs were recorded, which are shown in Fig. 2a. The absorbance maximum was observed at a wavelength of ~358 nm (corresponds to the energy value of 3.46 eV), which provides clear evidence of ZnO formation. Thereafter, the band gap was calculated using the Tauc plot represented in Fig. 2b and given by Eq. (2):
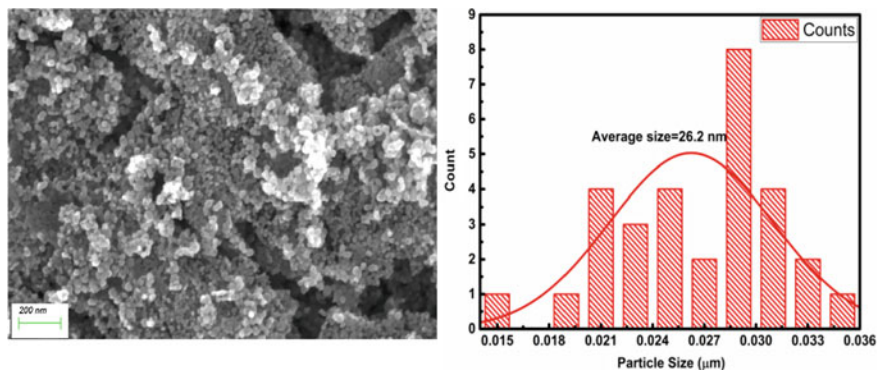
Fig. 1 **a** FESEM image of the ZnO NPs and **b** represents the histogram plot and the average size of the NPs
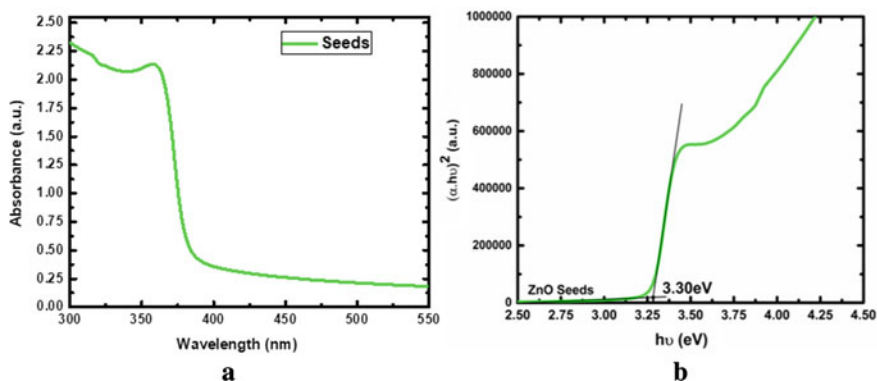

Fig. 2 **a** Absorbance spectra of the ZnO NPs and **b** represent the Tauc plot of the NPs

$$\alpha h\upsilon = A\left(h\upsilon - E_g\right)^{(1/2)} \quad \text{if } h\upsilon \geq E_g, \tag{2}$$

where α, $h\upsilon$, and A represent the absorption coefficient, energy of the incident photon, and a constant which depends upon the refractive index of the material. The band gaps of the NPs were observed to be 3.30 eV.

3.1.2 PL-Emission Spectroscopy

Further, the PL-emission spectra of the ZnO NPs were recorded to study the various impurities and defects' states within the band gap of the NPs, as ZnO is a semi-conductor having variety of intrinsic defects. The presence of intrinsic defects in the NPs samples can be observed from the PL-emission spectra as represented in Fig. 3. The variety of intrinsic defects found to be present in the ZnO materials are
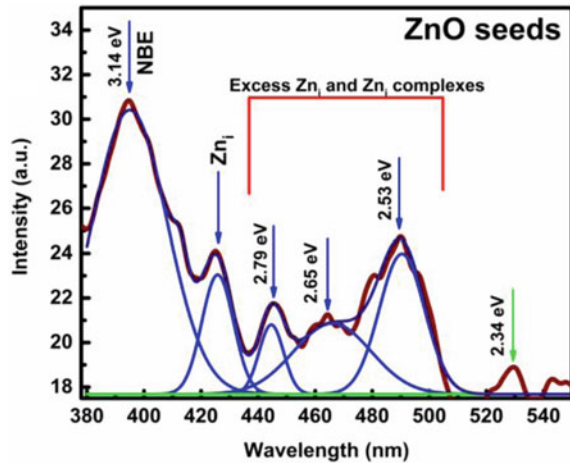
zinc vacancy, zinc interstitials, oxygen vacancy, oxygen interstitials, and antisites of oxygen and zinc, as reported in the literature [15, 16]. The various peaks observed from the PL-emission spectra were near-band edge (NBE) emission, zinc interstitial, zinc complexes, and oxygen vacancy peaks. The PL-emission spectra represented in Fig. 3 depict that the intensity of NBE emission is higher as compared to the defects present within the band gap of the ZnO NPs. The ratio of the intensity of various defects is less compared to the near-band edge (NBE) emission.

3.2 ZnO Nanorods (NRs)

3.2.1 Absorbance and PL-Emission Spectroscopy

Figure 4a shows the FESEM images of the ZnO NRs. FESEM images reveal the formation of NRs with hexagonal cross sections and clustering of NRs which further form flower-type morphology. Figure 4b depicts the diameter distribution (histogram) plot for the given FESEM image. The average diameter of the NRs was found to be ~410 nm from the histogram plot. Thereafter, the absorbance of the NRs was recorded as the function of wavelength and maximum absorbance was observed at the wavelength ~355 nm (3.54 eV) (Fig. 5). The band gap was calculated from the Tauc plot which came out to be ~3.27 eV using Eq. (2). Further, defects in the rods can be found by recording the PL-emission spectra. Figure 6 represents the PL-emission spectra of the NRs and it contains the defect states similar to as observed for the ZnO NPs. The only difference lies in the intensity ratios of the defect states (Figs. 3 and 6).
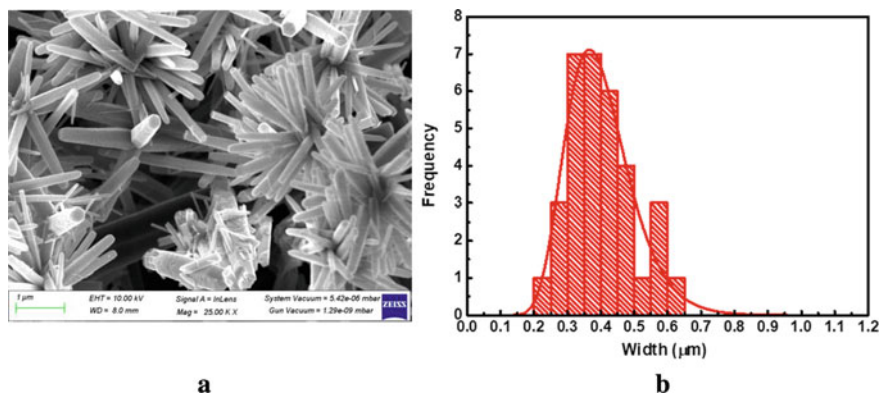
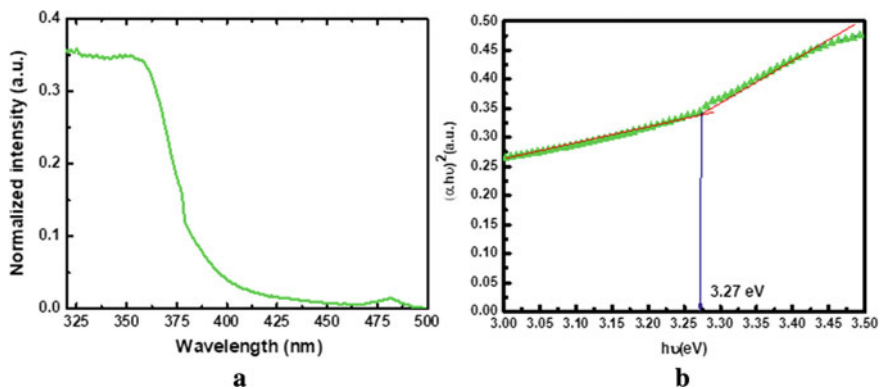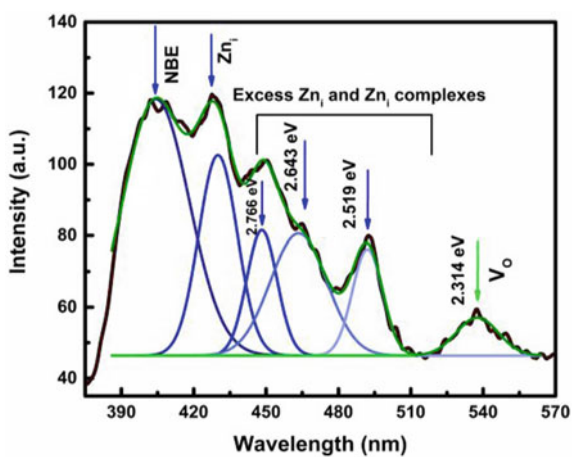**Fig. 4** **a** FESEM image of the ZnO NRs and **b** histogram for the respective image

Fig. 5 **a** Absorbance spectra of the ZnO NRs and **b** Tauc for the respective spectra

Fig. 6 PL-emission spectra of the ZnO NRs

3.3 Effect of Morphology (Seeds and Rods) on the Absorbance and Emission Properties

The absorbance and emission spectra were recorded for ZnO NPs as well as for ZnO NRs. It was observed that the absorbance spectra for both the samples show quite resemblance. The absorbance peak maxima for NPs and NRs were observed at ~358 nm and ~355 nm, respectively. As a result, it may be inferred from this finding that the morphology and dimensions of the sample barely affect the UV–vis absorbance spectroscopy. Moreover, the PL-emission spectra of both the samples were also recorded and depict the same defects at the approximately same position in the band gap. The only parameters observed to differ are the position of the NBE peak and the intensity ratio of the defect peaks. As it was clear from the PL-emission spectra of both the samples that the defect peak at wavelength ~425 nm, which corresponds to the zinc interstitials, increases as the morphology changes from NPs to NRs. Further, the position of NBE peak was observed to shift toward lower energy (3.14–3.06 eV) as the morphology transformed from NPs to NRs.

3.4 Conclusion

ZnO NPs and NRs were successfully prepared by the wet chemical method at low-temperature conditions. The ZnO NPs with an average particle size of ~26 nm and NRs with an average diameter of ~400 nm were synthesized with the reported chemical method. The FESEM images were recorded to confirm the morphology of the prepared samples. Further, the absorbance spectra and emission spectra of the samples were recorded to study their optical properties. The results from the current study reveal that the morphology and the dimensions of the nanostructures had a negligible impact on the absorbance spectra. PL-emission spectra clearly depict the effect of morphology on the intensities of various defect states. The intensity of Zn interstitial and other defects gets enhanced as the morphology changes to NRs from the NPs. The reason for this is found to be due to the surface effect of NRs. Further, studies on the impact of coating metal layer on the surface of NSs can be investigated.

References

1. Sandhu A (2007) The future of ultraviolet LEDs. Nat Photonics 1:38. https://doi.org/10.1038/nphoton.2006.36
2. Suh D, Lee SY, Hyung JH, Kim TH, Lee SK (2008) Multiple ZnO nanowires field-effect transistors. J Phys Chem C 112:1276–1281. https://doi.org/10.1021/jp709673s
3. Huang MH, Mao S, Feick H, Yan H, Wu Y, Kind H, Weber E, Russo R, Yang P (2001) Room-temperature ultraviolet nanowire nanolasers. Science (80-)(292):1897–1899. https://doi.org/10.1126/science.1060367

4. Wang X, Summers CJ, Wang ZL (2004) Growth of aligned ZnO nanorods for nano-optoelectronics and nanosensor arrays. Nano 4:2–5. http://pubs.acs.org/doi/abs/10.1021/nl035102c

5. Wang SF, Tseng TY, Wang YR, Wang CY, Lu HC, Shih WL (2008) Effects of preparation conditions on the growth of ZnO nanorod arrays using aqueous solution method. Int J Appl Ceram Technol 5:419–429. https://doi.org/10.1111/j.1744-7402.2008.02242.x

6. Huber TE, Celestine K, Graf MJ (2003) Magnetoquantum oscillations and confinement effects in arrays of 270-nm-diameter bismuth nanowires. Phys Rev B Condens Matter Mater Phys 67:1–11. https://doi.org/10.1103/PhysRevB.67.245317

7. Park SH, Seo SY, Kim SH, Han SW (2007) Homoepitaxial ZnO film growth on vertically aligned ZnO nanorods. J Cryst Growth 303:580–584. https://doi.org/10.1016/j.jcrysgro.2006.12.079

8. Zhao L, Lian J, Liu Y, Jiang Q (2006) Structural and optical properties of ZnO thin films deposited on quartz glass by pulsed laser deposition. Appl Surf Sci 252:8451–8455. https://doi.org/10.1016/j.apsusc.2005.11.054

9. Choopun S, Tabata H, Kawai T (2005) Self-assembly ZnO nanorods by pulsed laser deposition under argon atmosphere. J Cryst Growth 274:167–172. https://doi.org/10.1016/j.jcrysgro.2004.10.017

10. Zhou J, Nomenyo K, Cesar CC, Lusson A, Schwartzberg A, Yen CC, Woon WY, Lerondel G (2020) Giant defect emission enhancement from ZnO nanowires through desulfurization process. Sci Rep 10:1–10. https://doi.org/10.1038/s41598-020-61189-7

11. Zhang ZY, Xiong HM (2015) Photoluminescent ZnO nanoparticles and their biological applications. Mater (Basel) 8:3101–3127. https://doi.org/10.3390/ma8063101

12. Haque MJ, Bellah MM, Hassan MR, Rahman S (2020) Synthesis of ZnO nanoparticles by two different methods and comparison of their structural, antibacterial, photocatalytic and optical properties. Nano Express 1. https://doi.org/10.1088/2632-959X/ab7a43

13. Karthik KV, Raghu AV, Reddy KR, Ravishankar R, Sangeeta M, Shetti NP, Reddy CV (2022) Green synthesis of Cu-doped ZnO nanoparticles and its application for the photocatalytic degradation of hazardous organic pollutants. Chemosphere 287:132081. https://doi.org/10.1016/j.chemosphere.2021.132081

14. Sharma P, Tiwari SK, Barman PB (2022) Abnormal red shift in photoluminescence emission of ZnO nanowires. J Lumin 251:119231. https://doi.org/10.1016/j.jlumin.2022.119231

15. Fang Z, Wang Y, Xu D, Tan Y, Liu X (2004) Blue luminescent center in ZnO films deposited on silicon substrates. Opt Mater (Amst) 26:239–242. https://doi.org/10.1016/j.optmat.2003.11.027

16. Janotti A, Van De Walle CG (2005) Oxygen vacancies in ZnO. Appl Phys Lett 87:1–3. https://doi.org/10.1063/1.2053360

A Hybrid Multi-Objective Optimization Algorithm for Scheduling Heterogeneous Workloads in Cloud

Iqbal Gani Dar and **Vivek Shrivastava**

Abstract Cloud computing provides a platform for hosting and providing services to users in an on-demand manner over the Internet. Numerous challenges are faced due to the dynamic nature of the cloud environment. One such key challenge involves the scheduling of tasks submitted by various users at the same time on virtual machines (VMs). Task scheduling involves the optimized resource allocation to various tasks while considering several performance metrics such as the turnaround time and the throughput. To improve these performance measures, a variety of single-criteria and multi-criteria decision-making (MCDM) methods have been suggested. In this study, we have proposed a hybrid algorithm BWM_ELECTRE (Best Worst Method EliminationEt Choice Translating REeality). The performance of this algorithm has been analyzed and evaluated in terms of the turnaround time and the throughput performance measures. The efficiency of BWM_ELECTRE has been compared with the best-known MCDM method BWM_TOPSIS (Best Worst Method Technique for Order Preference by Similarity to Ideal Solution). A dataset of 2400 cloudlets has been considered to compare the performance of these algorithms. Experiments have been carried out in an open-source simulation tool CloudSim 3.0.1. The results validate that BWM_ELECTRE outperforms the BWM_TOPSIS algorithm by 1.27%, and 2.08% in terms of the average turnaround time and the throughput, respectively.

Keywords BWM · Cloud computing · CloudSim · ELECTRE · TOPSIS

I. G. Dar (✉) · V. Shrivastava
Department of Computer Science, International Institute of Professional Studies, DAVV, Indore 452001, India
e-mail: iqbal.gani.dar-rs@iips.edu.in

V. Shrivastava
e-mail: vivek.shrivastava@iips.edu.in

1 Introduction

Over the past half-century, there has been a significant progress in Information and Communications Technology (ICT) field. This has led to the perception that computing will one day be called the fifth utility after water, gas, electricity, and telephony. This computing utility will provide computing services that are considered vital for meeting the everyday demands of every person. To make this perception practically achievable, various computing models have been put forward, of which the most recent one is cloud computing (CC). Enterprises and users access applications on the cloud sitting anywhere on the globe on demand and pay per use [1].

As a commercial computing platform, CC provides three types of services to users, namely Platform as a Service (PaaS), Software as a Service (SaaS), and Infrastructure as a Service (SaaS) [2]. Currently, the increasing trend of users adopting CC has led to increased benefits to both cloud users and cloud service providers (CSPs). However, this growth has also led to a more complex resource provisioning and resource scheduling process. The literature proposes that task scheduling problems (TSPs) can be formulated and solved mathematically by using MCDM methods. The BWM is one of the most recent MCDM methods that is highly reliable and consumes less time for decision-makers in comparison with various other MCDM methods such as analytic hierarchy process (AHP) [3]. MCDM offers a group of techniques for selecting a set of alternatives based on evaluating several conflicting criteria for prioritizing the alternatives efficiently [4].

In this study, we introduce a hybrid model, namely BWM_ELECTRE for formulating the TSP in CC. For performance evaluation, we have compared this hybrid model with the best-known MCDM method BWM_TOPSIS.

Our contributions in this paper are as follows:

1. Designing an efficient task scheduling strategy based on the integration of BWM and ELECTRE methods.
2. Finding the turnaround time and the throughput of the BWM_ELECTRE model.
3. Designing a systematic task scheduling policy based on the integration of BWM and TOPSIS methods.
4. Finding the turnaround time and the throughput of the BWM_TOPSIS model.
5. Comparing the BWM_ELECTRE and BWM_TOPSIS models on CloudSim 3.0.1 toolkit where the comparison metrics are the turnaround time and the throughput.

The rest of this paper is organized as follows. Section 2 provides a review of the related works done. Section 3 describes the proposed work. Section 4 presents the results and discussion. Finally, Sect. 5 concludes the paper.

2 Related Work

Cloud infrastructure that was once restricted to a single provider data center is now expanding. This has resulted in more tasks from users being scheduled in VMs. As a result, TS becomes a key issue that affects the performance of CC services. Therefore, to implement users' tasks, cloud data centers should meet different criteria in applying the cloud resources and consider the multiple requirements of users. The TSP is an active research field in a heterogeneous CC environment. However, there is no exact optimal solution that optimizes all criteria in TSP [4]. Concerning TS and resource allocation, many different techniques and models have been proposed. Some of the investigated techniques include the adoption of nature-inspired algorithms such as energy aware, time, and throughput optimization heuristic (EATTO) based on the bat algorithm [5], improved Whale optimization algorithm (WOA) for cloud task scheduling (IWC) [6], chaotic squirrel search algorithm (CSSA) [7], modified fire fly algorithm [8], adaptive particle swarm optimization (PSO)-based task scheduling (AdPSO) [9], hybrid bio-inspired algorithm [10], harmony-inspired genetic algorithm (HIGA) [11], chaotic Darwinian chicken swarm optimization (CDCSO) [12], and game theory [13]. However, there is still a lack of tools and literature for determining the effect of these techniques on the user workload and computing resources. The adoption of a certain technique is governed by the objective to accomplish a certain task in the CC environment.

The application of mathematical techniques such as MCDM methods for handling TSPs is gaining attention. MCDM process selects the best alternative among a set of all the alternatives. The computation is performed mainly by multiple criteria. MCDM evaluates many alternatives under a finite number of conflicting criteria. Many researchers have applied MCDM methods to TS and resource allocation problems Table 1 provides a summary of the most recent previous works done in task scheduling.

Table 1 Summary of the related work

Author (Year)	Performance metrics	Tools used
Khorsand et al. [1]	Energy, makespan	CloudSim 3.0.3
Guo [2]	Completion time, deadline violation rate, resource utilization	CloudSim
Alhubaishy et al. [3]	Time, cost	Mathematically proved
Gyani et al. [4]	Review paper on MCDM	Mathematically proved
Gu et al. [5]	Energy consumption, execution time, throughput	Python
Chen et al. [6]	Execution time, load, price cost	MATLAB R 2018b

Table 1 (continued)

Author (Year)	Performance metrics	Tools used
Sanaj et al. [7]	Energy, time, cost, resource utilization, service-level agreement (SLA) violation	CloudSim
Bacanin et al. [8]	Cost, makespan	Python, Java
Nabi et al. [9]	Makespan, throughput	Eclipse IDE 3.0 and CloudSim 3.0.3
Domanal et al. [10]	Execution time, response time, resource utilization	Python
Sharma et al. [11]	Makespan, energy	CloudSim 4, MATLAB 2013b
Kiruthiga et al. [12]	Energy, execution cost, load balancing index, task completion time, throughput, response time, CPU utilization, bandwidth utilization	CloudSim 3.0.3
Yang et al. [13]	Response time, resource utilization, throughput	Mathematically proved
Devarasetty et al. [14]	Deployment cost, QoS performance	C++
Alla et al. [15]	Makespan, degree of imbalance, resource utilization	CloudSim
Kumar et al. [16]	QoS	Mathematically proved
Youssef [17]	Computational complexity, reliability	Mathematically proved
Dar et al. [18]	Number of cloudlets executed, turnaround time, throughput	CloudSim 3.0.1

3 Proposed Work

3.1 Problem Formulation

The number of users in a cloud computing system is growing and so are the tasks sent from these users to the cloud. The execution and management of all these tasks at the same time need a huge amount of resources. In a cloud-based system, a task (cloudlet) has many attributes associated with it such as cloudlet ID, length, deadline, rank, and energy. A particular task scheduling algorithm should consider these attributes for the efficient scheduling of tasks. Some criteria need to be taken into consideration, and these attributes can be taken either as single objective or multi-objective functions. Our contribution in this work is a novel technique for task scheduling that takes into account the priority issue and aims to improve performance by minimizing turnaround time while maximizing throughput. We first calculate the weights of tasks' criteria using BWM. Then ELECTRE is applied to prioritize the tasks. The task length (size), deadline, and available resources are the criteria taken
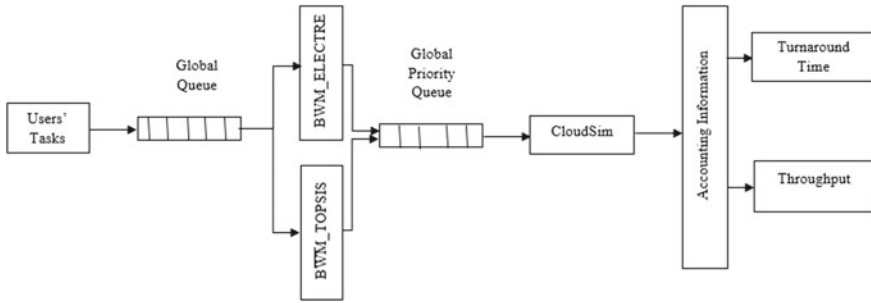
Fig. 1 Proposed model

into account in this work. These metrics are undoubtedly the most crucial factors that can satisfy the needs of users and CSPs alike. Figure 1 shows the proposed model of our work.

3.2 BWM

BWM is a recent MCDM method that has attracted many researchers for formulating decision problems and finding optimal solutions based on pairwise comparisons among various alternatives. The BWM was introduced by Rezaei to avoid some limitations of AHP, such as being inconsistent and time consuming when applied. Also, BWM needs only $2n$-3 comparisons, whereas AHP needs $n*(n-1)/2$ comparisons [3]. These advantages of the BWM have motivated us to propose a TS model for enhancing resource allocation in CC. A set of n criteria are taken, and a pairwise comparison of these criteria is performed using a 1/9 to 9 scale. Other scales such as 0.1 to 1.0 or 1 to 100 can also be used. The resulting matrix would be

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix},$$

where a_{ij} shows the relative preference of criterion i to criterion j. $a_{ij} = 1$ shows that i and j have the same importance. $a_{ij} > 1$ shows that i is more important than j with $a_{ij} = 9$ showing the extreme importance of i to j. The relative preference of j to i is shown by a_{ji}.

The steps of BWM are described below to calculate the weights of the criteria.

Step 1. Identify the set of preferred decision criteria. In this step, a set of criteria $(c_1, c_2, \ldots, c_n)$ are considered that should be used by the decision-makers (experts)
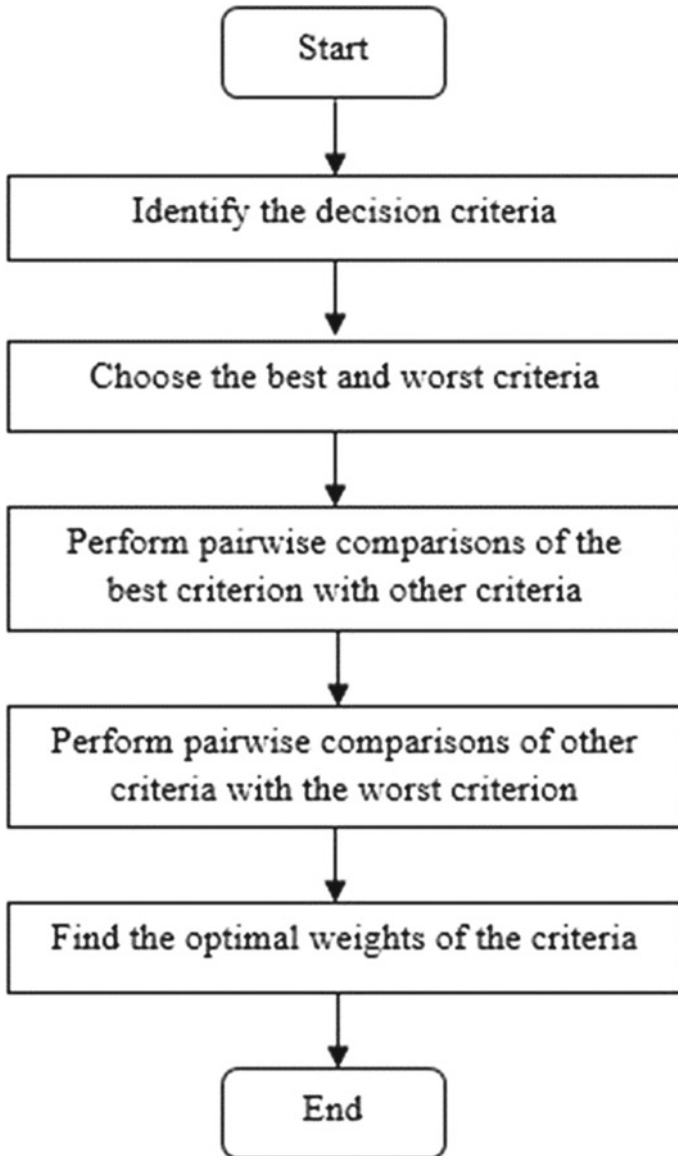
Fig. 2 Steps in BWM

to arrive at a decision. In this study, the most important decision criteria related to the tasks considered are length, deadline, and the number of resources.

Step 2. Choose the best, i.e., most desirable or most important criteria and the worst, i.e., least desirable or least important criteria.

Step 3. Determine the preference of the best criterion over all the other criteria using a number between 1 and 9 where 1 represents equal importance.

Step 4. Determine the preference of all the criteria over the worst criterion using a number between 1 and 9.

Step 5. Find the optimal weights. A linear programming model (LPM) is required to be solved to find the optimal weights of the criteria. The LPM is formulated using the following formula. Figure 2 summarizes the steps of BWM.

$$\text{Min}\, \xi L$$
$$\text{Subject to}$$
$$Wb - \alpha Bj Wj| \leq \xi L, \ \text{for all } j$$
$$|Wj - \alpha j w W w| \leq \xi L, \ \text{for all } j$$
$$\sum W_j = 1$$
$$W_j \geq 0, \text{for all } j$$

3.3 ELECTRE

In this study, ELECTRE III is used to prioritize the tasks. The ELECTRE III method generates a priority score for each task. ELECTRE III offers a useful foundation for comparison based on the assessment of several conflicting criteria [15]. There are two steps in the ELECTRE III method: For each pair of items a and b, a valued outranking relation is created in the first phase by assessing the validity of the assertion that "a is at least as strong as b" or that "a outperforms b." In the subsequent phase, a ranking of all items is computed using these valued outranking relations. The following four steps can be used to assess the assertion "a outranks b" represented by $\sigma(a, b)$:
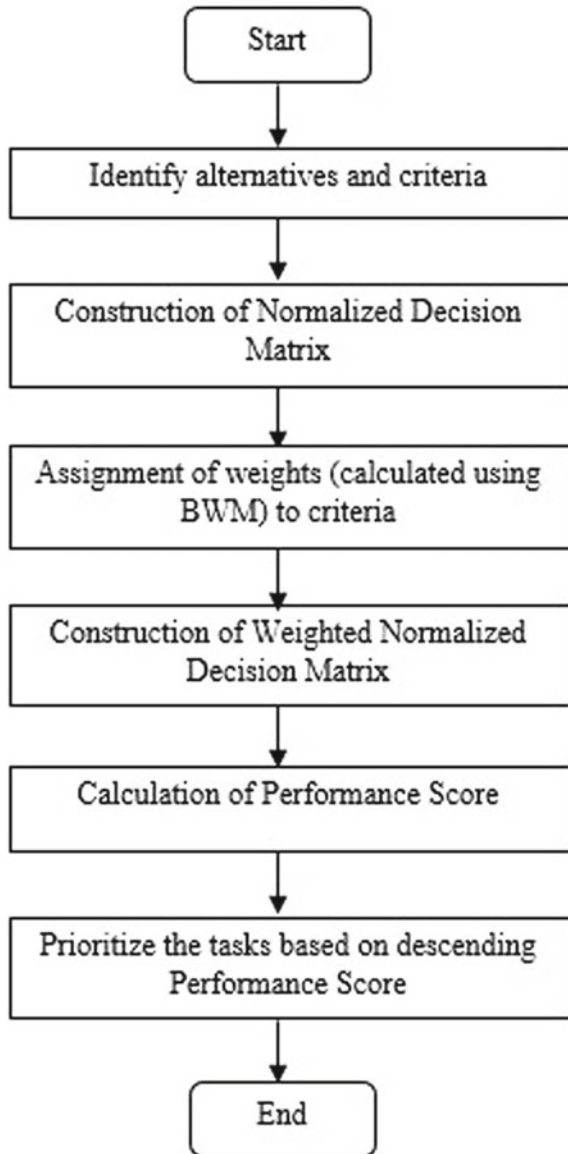
Step 1. For each pair of elements a and b, find the partial concordance index as follows:

$$C_j(a, b) = \begin{cases} 0 & \text{if } g_j(b) - g_j(a) \geq p_j \\ 1 & \text{if } g_j(b) - g_j(a) \leq q_j \ , \\ \frac{p_j + g_j(a) - g_j(b)}{p_j - q_j} & \text{otherwise} \end{cases} \quad (1)$$

where $g_j(a)$ and $g_j(b)$ denote the assessment of items a and b, respectively, and according to the criterion, $j.q_j$ is the threshold for indifference and p_j is the threshold for preference.

Step 2. For each pair of elements a and b, find the global concordance index (GCI) C (a, b) as follows:

Fig. 3 Steps in ELECTRE
method

$$C(a, b) = \frac{\sum_{j=1}^{n} w_j \times C_j(a, b)}{\sum_{j=1}^{n} w_j},$$ (2)

where w_j stands for weight criterion j. The degree of arguments that assert that "a outranks b" is measured by the GCI C (a, b).

Step 3: For each pair of items a and b, find the partial discordance index (PDI) $D(a, b)$:

$$D_j(a, b) = \begin{cases} 0 & \text{if } g_j(b) - g_j(a) \geq p_j \\ 1 & \text{if } g_j(b) - g_j(a) \leq v_j \\ \frac{g_j(b) - g_j(a) - p_j}{v_j - p_j} & \text{otherwise} \end{cases} \tag{3}$$

where v_j is the veto threshold, which stands for the upper bound on how much compensation decision-makers will tolerate. The strength of arguments against the assertion that "a outranks b" according to criteria j is evaluated by the PDID (a, b).

Step 4. Find the credibility index for the assertion that "a outranks b":

$$\sigma(a, b) = \begin{cases} C(a, b) \, if \, D_j(a, b) \leq C(a, b) & \forall j \\ C(a, b) \times \prod_{D_j(a,b) \succ C(a,b)} \frac{1 - D_j(a,b)}{1 - C(a,b)} \end{cases}. \tag{4}$$

The ELECTRE III method's next step involves using the model to rank alternatives using the credibility matrix. Figure 3 summarizes the steps involved in ELECTRE.

3.4 TOPSIS

The set of steps involved in TOPSIS are as follows [17].

Step 1. Formation of the decision matrix (DM) and choose the criteria's weight. In this step, three criteria have been identified related to tasks that are length, deadline, and the number of resources. Weights calculated using BWM are assigned to these criteria such that $w_1 + w_2 + w_3 = 1$, where w_1, w_2, and w_3 are the weights of length, deadline, and the number of resources. Criteria can be beneficial criteria (a larger value is desired) or non-beneficial (a lower value is desired).

Step 2. Compute the normalized decision matrix (NDM).

The values in the DM are converted to a normalized scale because different criteria have different units. Standardized formulas can be used to normalize values. The following are some of the most often used techniques for determining the normalized value n_{ij}:

$$n_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x^2 ij}} \tag{5}$$

$$n_{ij} = \frac{x_{ij}}{\max x_{ij}} \tag{6}$$

$$n_{ij} = \begin{cases} \frac{x_{ij} - \min x_{ij}}{\max x_{ij} - \min x_{ij}} \\ \frac{\max x_{ij} - x_{ij}}{\max x_{ij} - \min x_{ij}} \end{cases} \tag{7}$$

for $i = 1, \ldots, m; j = 1, \ldots, n$.

Step 3. Compute the weighted (NDM). The following formula is used to calculate the weighted normalized value v_{ij}:

$$v_{ij} = w_j n_{ij} \tag{8}$$

for $i = 1, \ldots, m; j = 1, \ldots, n$,

where jth criterion has weight w_j and sum of weights equals 1.

Step 4. Find the best solutions (positive and negative). The optimal positive solution maximizes the favorable criteria while minimizing the unfavorable ones, whereas the ideal negative solution maximizes the unfavorable criteria while minimizing the favorable ones.

The form of the positive ideal solution (PIS) A^+ is:

$$A^+ = \left(v_1^+, v_2^+, \ldots, v_n^+\right) = \left(\left(\max v_{ij} | j \in I\right), \left(\min v_{ij} | j \in J\right)\right). \tag{9}$$

The form of negative ideal solution (NIS) A^- has the form:

$$A^- = \left(v_1^-, v_2^-, \ldots, v_n^-\right) = \left(\left(\min v_{ij} | j \in I\right), \left(\max v_{ij} | j \in J\right)\right), \tag{10}$$

where I is associated with favorable criteria and J with the unfavorable criteria, $i = 1, \ldots, m; j = 1, \ldots, n$.

Step 5. Compute the separation measures from the PIS and the NIS.

Various distance measures can be applied with the TOPSIS method. The distance of each alternative from the PIS is given as

$$d_i^+ = \left(\sum_{j=1}^{n} \left(v_{ij} - v_j^+\right)^p\right)^{1/p}, i = 1, 2, \ldots, m. \tag{11}$$
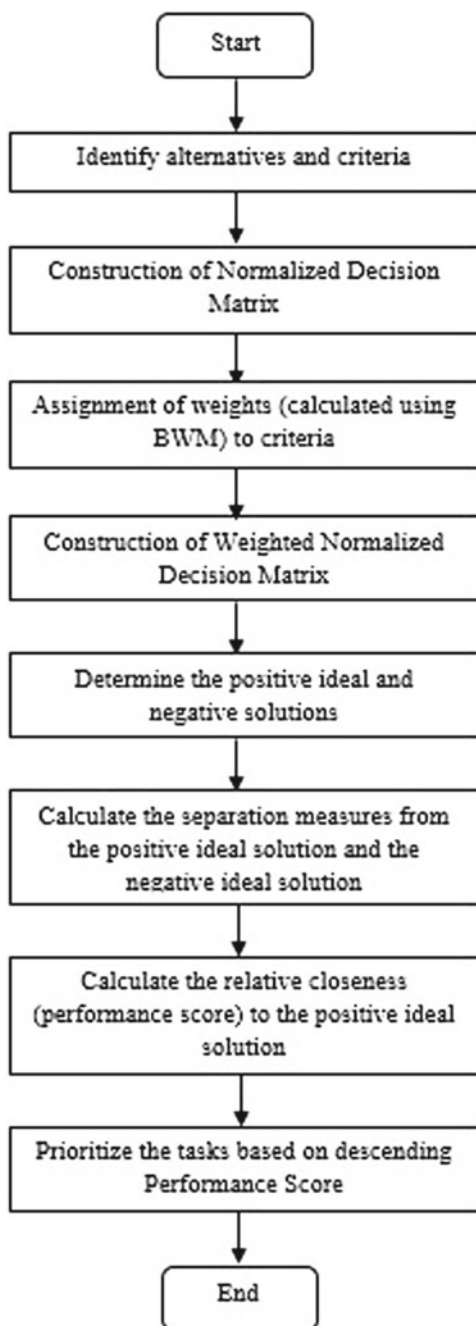
The distance of each alternative from the NIS is given as

$$d_i^- = \left(\sum_{j=1}^{n} \left(v_{ij} - v_j^-\right)^p\right)^{1/p}, i = 1, 2, \ldots, m, \tag{12}$$

where $p \geq 1$. The most popular conventional n-dimensional Euclidean measure is available for $p = 2$.

$$d_i^+ = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^+\right)^2}, i = 1, 2, \ldots m. \tag{13}$$

Fig. 4 Steps in TOPSIS method

$$d_i^- = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^- \right)^2}, i = 1, 2, \ldots m. \qquad (14)$$

Step 6. Find the relative closeness to the PIS. The definition of the ith alternative's relative closeness to A^+ is

$$R_i = \frac{d_i^-}{d_i^- + d_i^+}, \qquad (15)$$

where $0 \leq R_i \leq 1$, $i = 1, 2, \ldots, m$.

Step 7. Order the preferences. A set of alternatives can be ranked in descending order by R_i value. Figure 4 shows the steps in TOPSIS.

4   Simulation Results and Discussion

The performance of the proposed algorithms has been evaluated in the CloudSim 3.0.1 toolkit where the performance measures are turnaround time and throughput. The objective of the proposed hybrid model is to meet the multi-criteria QoS requirements from both cloud users' and providers' perspectives by minimizing the turnaround time for the user and maximizing the throughput for the providers. CloudSim is an open-source simulation that has been written in Java language. It enables modeling and simulation of CC and application provisioning environments. This tool currently provides support for both modeling and simulation of single and internetworked clouds [1]. The simulation is done using two data centers, two hosts, with different parameter values for VM and cloudlet used as given in Table 2. We have used a synthetically generated dataset of 2400 workloads. Table 3 shows performance metrics and comparison (Figs. 5 and 6).
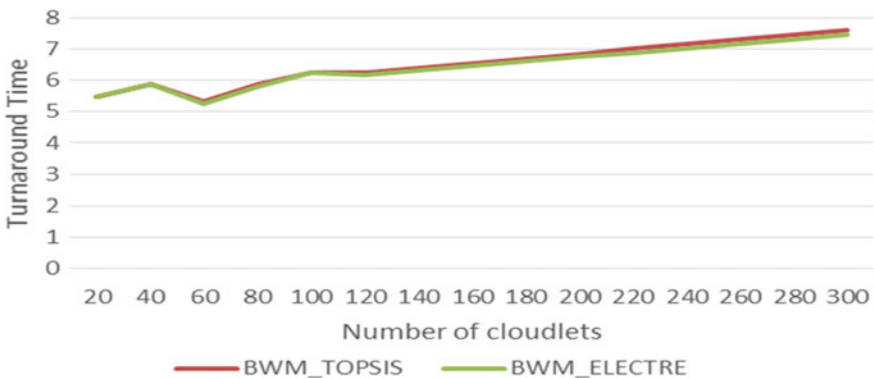

Fig. 5   Average turnaround time

Table 2 Simulation parameters

Parameters	Values
VM parameters	
RAM	512 MB
MIPS	1000
Bandwidth	1000
Number of CPUs	1
VMM	Xen
Cloudlet parameters	
Length	1000
File size	300
Output size	300
Number of CPUs	1

Table 3 Performance metrics and comparison

Performance metrics	Comparison	
	BWM_TOPSIS	BWM_ELECTRE
Average turnaround time	6.534336	6.451023
Throughput	1656	1691
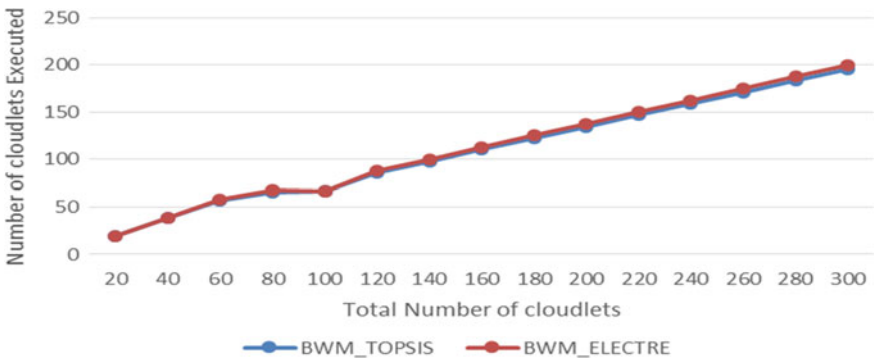
Fig. 6 Throughput comparison

5 Conclusion and Future Scope

Cloud computing is a popular computing model that provides computing as a service and enables users to share a pool of virtualized resources on pay per use basis. Presently, there are more advantages for both customers and providers due to the rising trend of users migrating to CC. However, this rise has also led to a more difficult work scheduling and resource provisioning process. Numerous researches in this area ignore multiple criteria that are present in task scheduling in the real

world and instead concentrate only on one. This does not satisfy multi-criteria QoS requirements from both cloud users' and providers' points of view set in the Service Level Agreement (SLA). We have addressed this issue in this paper and proposed a hybrid multi-criteria task scheduling algorithm by integrating BWM and ELECTRE. To evaluate the performance of BWM_ELECTRE, we have compared it with BWM_TOPSIS. We prioritize tasks based on three criteria, namely cloudlet length, deadline, and the number of resources. The experimental findings show that BWM_ ELECTRE outperforms BWM_TOPSIS in terms of the turnaround time and the throughput. Adding more criteria to the scheduling process to make efficient scheduling decisions may be another direction of research. Further, a comparative study of other MCDM techniques for task scheduling can be carried out in future.

References

1. Khorsand R, Ramezanpour M (2020) An energy-efficient task-scheduling algorithm based on a multi-criteria decision-making method in cloud computing. Int J Commun Syst 33(9):1–17. https://doi.org/10.1002/dac.4379
2. Guo X (2021) Multi-objective task scheduling optimization in cloud computing based on fuzzy self-defense algorithm. Alexandria Eng J 60(6):5603–5609. https://doi.org/10.1016/j.aej.2021.04.051
3. Alhubaishy A, Aljuhani A (2020) The best-worst method for resource allocation and task scheduling in cloud computing. In: Proceedings of international conference on computer applications and information security CONFERENCE, IEEE. https://doi.org/10.1109/ICCAIS48893.2020.9096877
4. Gyani J, Ahmed A, Haq MA (2022) MCDM and various prioritization methods in AHP for CSS: a comprehensive review. IEEE Access 10:33492–33511. https://doi.org/10.1109/ACCESS.2022.3161742
5. Gu Y, Budati C (2020) Energy-aware workflow scheduling and optimization in clouds using bat algorithm. Futur Gener Comput Syst 113:106–112. https://doi.org/10.1016/j.future.2020.06.031
6. Chen X, Cheng L, Liu C, Liu Q, Liu J, Mao Y, Murphy J (2020) A WOA-based optimization approach for task scheduling in cloud computing systems. IEEE Syst J 14(3): 3117–3128. https://doi.org/10.1109/JSYST.2019.2960088
7. Sanaj MS, Prathap JPM (2020) Nature inspired chaotic squirrel search algorithm (CSSA) for multi objective task scheduling in an IAAS cloud computing atmosphere. Eng Sci Techol Int J 23(4):891–902. https://doi.org/10.1016/j.jestch.2019.11.002
8. Bacanin N, Zivkovic M, Bezdan T, Venkatachalam K, Abouhawwash M (2022) Modified firefly algorithm for workflow scheduling in cloud-edge environment. Neural Comput Appl 34(11):9043–9068. https://doi.org/10.1007/s00521-022-06925-y
9. Nabi S, Ahmad M, Ibrahim M, Hamam H (2022) AdPSO: adaptive PSO-based task scheduling approach for cloud computing. Sensors 22(3):1–22. https://doi.org/10.3390/s22030920
10. Domanal SG, Guddeti RMR, Buyya R (2020) A hybrid bio-inspired algorithm for scheduling and resource management in cloud environment. IEEE Trans Serv Comput 13(1):3–15. https://doi.org/10.1109/TSC.2017.2679738
11. Sharma M, Garg R (2020) HIGA: Harmony-inspired genetic algorithm for rack-aware energy-efficient task scheduling in cloud data centers. Eng Sci Technol Int J 23(1):211–224. https://doi.org/10.1016/j.jestch.2019.03.009

12. Kiruthiga G, Vennila MS (2020) Energy efficient load balancing aware task scheduling in cloud computing using multi-objective chaotic darwinian chicken swarm optimization. Int J Comput Netw Appl 7(3):82–92. http://dx.doi.org/10.22247/ijcna/2020/196040
13. Yang J, Jiang B, Lv Z, Choo KKR (2020) A task scheduling algorithm considering game theory designed for energy management in cloud computing. Futur Gener Comput Syst 105:985–992. http://dx.doi.org/10.1016/j.future.2017.03.024
14. Devarasetty P, Reddy S (2021) Genetic algorithm for quality of service based resource allocation in cloud computing. Evol Intell 14(2):381–387. https://doi.org/10.1007/s12065-019-00233-6
15. Ben AH, Ben AS, Ezzati A, Touhafi A (2021) A novel multiclass priority algorithm for task scheduling in cloud computing. J Supercomput 77(10):11514–11555. https://doi.org/10.1007/s11227-021-03741-4
16. Kumar RR, Kumari B, Kumar C (2021) CCS-OSSR: a framework based on hybrid MCDM for optimal service selection and ranking of cloud computing services. Cluster Comput 24(2):867–883. https://doi.org/10.1007/s10586-020-03166-3
17. Youssef EA (2020) An integrated MCDM approach for cloud service selection based on TOPSIS and BWM. IEEE Access 8:71851–7186. https://doi.org/10.1109/ACCESS.2020.2987111
18. Dar I, Shrivastava V (2022) Scheduling tasks using multi-criteria decision making (MCDM) algorithms in cloud infrastructure. Stoch Model Appl 26(3):535–541

EEG-Based Emotion Recognition Using SVM

Ram Avtar Jaswal and Sunil Dhingra

Abstract Emotional recognition and detection has become one of the fastest emergent areas which has encouraged the researchers to do more research in this area. In this paper, we described an EEG-based emotion recognition method which identifies the human emotional states. The proposed method used the DEAP database and developed a supervised machine learning algorithms to recognize the emotional from EEG signal. The Welch's transformer is used to preprocess the EEG signal to obtain the alpha, delta, theta, beta waves and gamma waves. Others attributes such as mean, variance, standard deviation, kurtosis and skewness are also collected. After using SVM and XGbooster supervised machine learning algorithms, appropriate emotional state has been detected. The proposed method achieved the 96.2% response rate for happy, 94.8% for neutral, 93% for sad and angry. The proposed method has been compared with existing method.

Keywords Emotions · Electroencephalography · Deep convolutional neural network

1 Introduction

Emotions are the basis of human daily life and play a significant part in human intellect such as in coherent decision making, human intelligence, understanding and interpersonal communication. The emotion state of a person can be clearly recognized on the basis of the audio/visual signals (external expressions), physiological signals (internal expressions) and person perception. Human–computer interaction (HCI) systems are getting attention of researchers to recognition the emotions of

R. A. Jaswal (✉) · S. Dhingra
Department of Instrumentation, Kurukshetra University, Haryana, India
e-mail: ramavtar.jaswal@gmail.com

S. Dhingra
e-mail: sdhingra@kuk.ac.in

human being [1]. Recorded physiological signals of participants play an important role in this area, and after analyzing these signals, emotions of human can be recognized. Physiological signals have been recorded from central and autonomic nervous system. In central nervous system, signal from spinal cord and brains has been recorded, whereas in autonomic system physiological signal also contains other regulating physiological functions such as heart rate and arousal. There are different methods by which physiological signals have been recorded. The common signals that measured the emotions are galvanic skin response (GSR), electromyography (EMG), hear rate (HR), respiration rate (RR) and electroencephalography (EEG). In GSR, the level of emotional increases with respect to the level of arousal, whereas in EMG, it is negatively associated the valence emotions with frequency of muscle tension. In case of heart rate, the rate increases with negative emotions like fear, whereas RR changes irregularly with emotions like anger. Electroencephalography (EEG) is a one of techniques which is used the neuro-imaging that measure potential voltage changes through electrical impulses in triggering neurons, detect them in the scalp and record them with a device with electrode samples [2–4]. EEG signals are classified according to frequency and can be separated in five frequency bands: delta band (0.5–3 Hz), theta band (4–7 Hz), alpha band (8–13 Hz), beta band (14–30 Hz) and gamma band (above 30 Hz). These frequency band can be related with specific types of activities such as finger movements, falling asleep, active thinking and problem solving. EEG, which used in brain–computer interface (BCI), facilitates communication between computer and human being independently [5]. To understand the human state of mind, application of affective computing becoming increasingly popular day by day.

2 Related Work

The method of expressing emotions is the bipolar model proposed by Russell in [6] with arousal and valence dimensions. The value range of valence extends from −ve to +ve and range of aroused from "not-aroused" to "excited". This 2D model can recognize emotional labels and even also define emotions without isolated emotional label. However, the feelings of fear and anger had not varied in model as both emotional have the same −ve valence level and high-arousal values.

Mehrabian and Russell in [7] proposed 3D Pleasure (Valence)-Arousal-Dominance (PAD) model. The model's "pleasure-displeasure" level corresponds to the value of 2D model which measure the level of pleasure-emotional. The range of "Arousal Non-arousal" corresponds to the value of the Arousal and "Dominance Submissiveness" correspondence to the dimension of emotional control. The dominant dimension finds more sensitive labels in 3D model. As surprise and happiness have high arousal and positive valence value, still model is able to differentiate in these emotional using the degree of dominance.

Emotional cognition has received a lot of attention over the past decade as it is directly related to physiology, psychology, marketing and therapy. Previous studies

in [8–13] proposed different types of frequency domains, statistical properties, time domains and various feature extraction methods. In [14], author proposed a method combine six statistical properties in time domain of EEG signals to identify human emotions. The proposed method removed the noise and repetitive channels using the Relief-F and PCA algorithms. The method applied the SVM training to DEAP data and identified emotions. The resultant accuracy was 81.87%.

In the [15], authors tried to diminish the size of the aspect vectors to improve classification accuracy. The proposed method used primary component analysis (PCA) to rule out symptoms. PCA and SVM method applied to SEED dataset which classify positive, negative, and neutral emotions, and resultant accuracy was 85.85%.

In [16], authors experimented different kind of classifications on various functions. The authors collected statistical properties, entropy, energy density, frequency properties and wavelet energy. Researchers selected 14 electrodes from the DEAP database and collected 168 properties for every subject. Authors applied the decision tree, SVM, KNN classifiers, and accuracy result was 77.54% for valence, 79% for dominance and 78.88% for arousal.

The authors in [17] discussed multi-classification as a way to improve accuracy and had an excellent collaboration. First, the researchers collected several properties of the time–frequency field and the time–frequency properties of EEG signals. The researchers use PSO algorithm for feature selection. Researchers achieved the 76.67% accuracy rate.

Author proposed the DGCCA-AM method in [18] which provides a combination of the recognition characteristics of emotions into five categories. Authors used the short-time Fourier transform (STFT) and non-overlap of Hanning windows for 4 s to collect the differential entropy (DE) characteristics of the EEG signals. The proposed method is analyzed using a public dataset SEED-V. The experimental recorded an accuracy of 82.11%.

Authors in [19] projected a method based on the asymmetry index (AsI) to estimate the level of emotion in streaming music videos. AsI estimated the level of emotional by measuring the amount of mutual information in two frontal lobes. The low emotional signal segments are filtered out after calculating the ASI and then wavelet translation is applied. After extracting the features, SVM classification has been applied and obtained accuracy rate of 70.5%.

Authors in [20] projected decision tree and a rule-based classification algorithms which used EEG signals to identify emotions in four categories, i.e., happiness, sadness anger and relaxation. The proposed method collected the characteristics of the frequency and time domain of DEAP dataset. Then rules classification model generates 10 emotion classification rules and achieves an average precision of 81.64% for the relaxed emotion class.

Many studies have done extensive research on the recognition of emotions and have shown that the classification of emotions is more precise. In [21], authors used deep learning to categorize four emotions and applied functions to obtain the time, entropy and frequency domain. The precision of the time domain attributes was 93.48%, the precision of the frequency domain attributes was 92.44% and the precision of the entropy-based attributes was 93.75%.

Authors in [22] classified the emotions into two levels of arousal, valence. The proposed method involves the analysis of non-stationary nonlinear time series of EEG signals. The authors applied artificial neural network (ANN) classification on DEAP dataset. The results of arousal accuracy are 75% and valence accuracy of 72.87%.

In [23], authors applied STFT transform to obtain the high-order crossing (HOC), HHS spectrum of the delta, alpha, beta and gamma bands from DEAP dataset. Then authors used the mRMR algorithm to remove non-informative features. The results show that the overall accuracy is less than 60% even mRMR improves the accuracy of the results.

3 Proposed Work

The proposed method consists of data acquisition, preprocessing, feature extraction and classification step. The proposed method used EEG signals the DEAP dataset [24]. The preprocessed DEAP dataset are analyzed using various functions of time and frequency domains. The proposed method develops a supervised machine learning model to divide emotional states into four categories. Figure 1 illustrations the proposed framework to recognize the emotions.

3.1 Data Acquisition

The first step of proposed method is data acquisition. Depending on the capacity of the system, different sampling rates can be used to obtain EEG signals in system. In the proposed method, DEEP dataset has been used to analyze the emotion through EEG signals. Brain signals of 32 subjects were retrieved from EEG device at 500 Hz sampling rate in the DEAP database. The dataset has two type of physiological signal data: preprocessed and raw data. In raw data, different results may be obtained by using different preprocessing function. To maintain consistency, the proposed
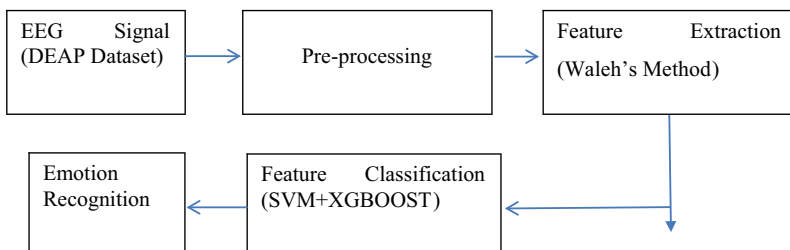

Fig. 1 Proposed method framework

method uses the Python preprocessed data. Figure 1 shows the placement of the EEG electrodes in dataset. The preprocessed dataset consists of 32 EEG signals (128 Hz) and 8 peripheral signals. There are two type of array contain in data: first for data and other for the emotional label. The data is applied to the classification algorithm independently of the preprocessing step. The method used the Python's pickle library to read EEG data. Using the Python's numpy library, data can be encoded in the different files which can be used to perform various binary classifications to distinct categories of valence, arousal, dominance and liking emotions.

3.2 Preprocessing and Feature Extraction

The values of arousal and valence are divided into two classes: high arousal (HA) and low arousal (LA). The combinations of valence and arousal can be converted to emotional states: high-arousal positive valence (Excited, Happy), low-arousal positive valence (Calm, Relaxed), high-arousal negative valence (Angry, Nervous) and low-arousal negative valence (Sad, Bored). These two-dimensional emotional space can be divided into four areas: HAHV, HALA, LAHV and LVLV. Table 1 gives the mean and standard deviation value of HAHV, LAHV, HALV and LALV. The scatter plot of the HAHV, LAHV, HALV and LALV is shown in Fig. 2.

Table 1 Mean, standard deviations (STD) value

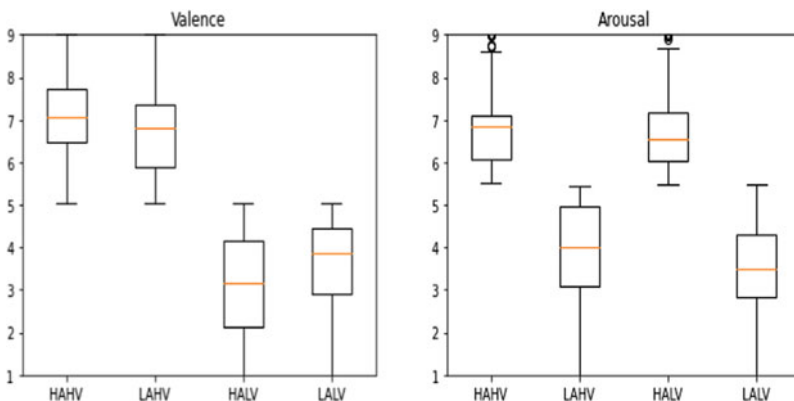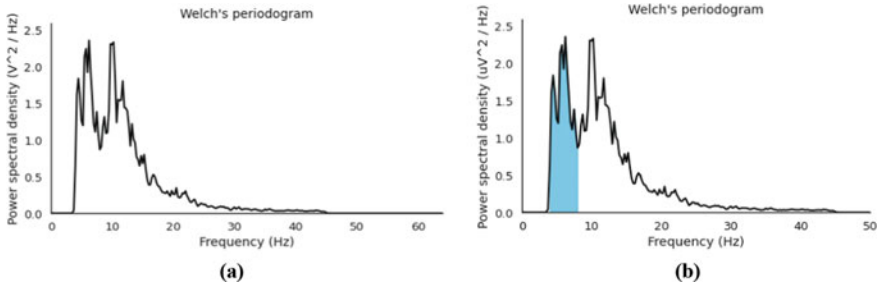|	Valence	Arousal
HAHV	Mean 7.12 STD 1.04	Mean 6.81 STD 0.84
LAHV	Mean 6.67 STD 1.09	Mean 3.87 STD 1.09
HALV	Mean 3.19 STD 1.23	Mean 6.79 STD 0.96
LALV	Mean 3.57 STD 1.12	Mean 3.47 STD 1.19

Fig. 2 Values of HAHV, LAHV, HALV, LALV

Fig. 3 Power spectral density **a** frequency band, **b** theta band

The frequency band of EEG signal is divided into four bands: theta (θ: 4–7 Hz), alpha (α: 8–13 Hz), beta (β: 14–30 Hz) and gamma (γ: 31–50 Hz). The Hanning window is applied to each EEG channel and power spectral density (PSD) calculated by Welch's method which is shown from Figs. 3, 4, 5, 6, 7, 8 and 9.
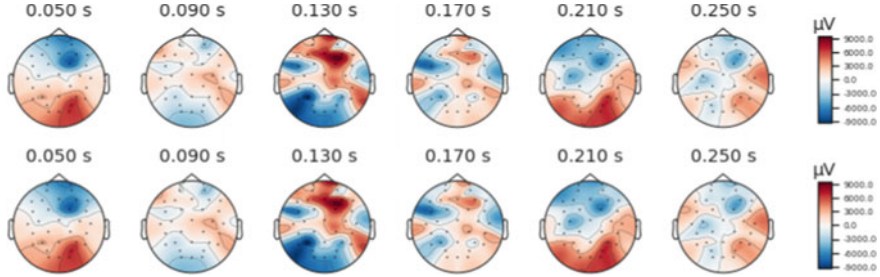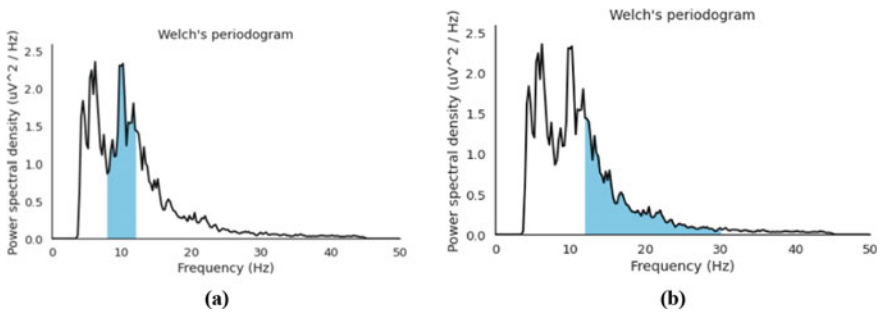

Fig. 4 Theta band

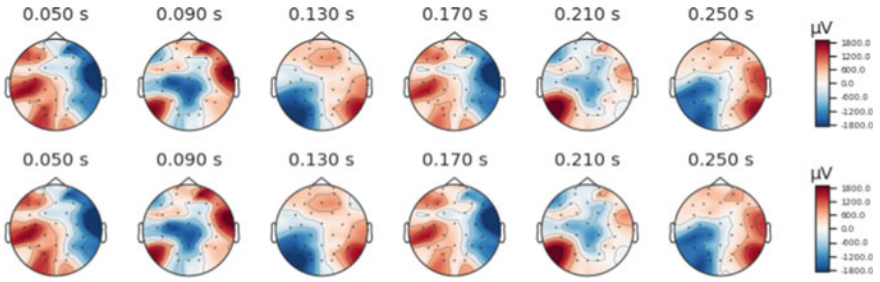**Fig. 5** Power spectral density **a** alpha band, **b** beta band
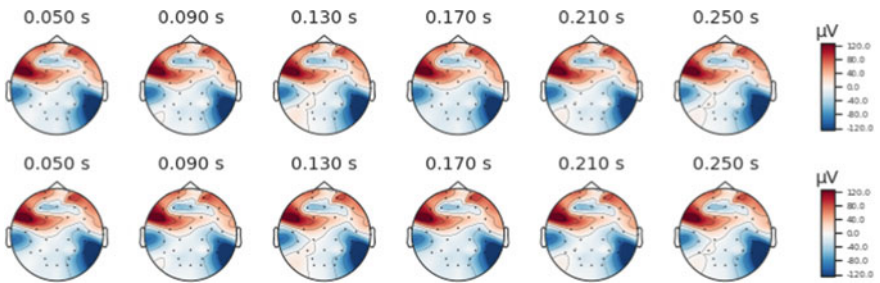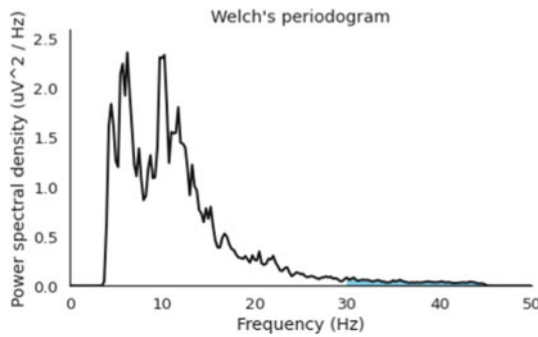
Fig. 6 Alpha band

Fig. 7 Beta band

Fig. 8 Power spectral density-gamma band

3.3 Emotions Classification

The classification first uses sample emotion data for training and then classifier recognizes human emotions after training. For classification, SVM with XGBooster architecture has been used in the proposed method.
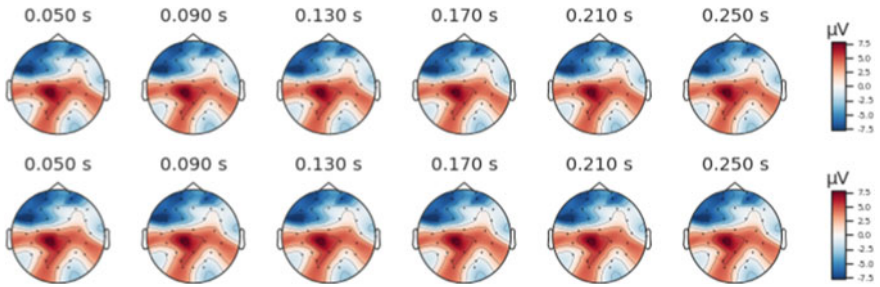
Fig. 9 Gamma band

Table 2 Confusion matrix

Predictable emotion	Experimental emotion				Accuracy (%)
	Happy	Neutral	Sad	Angry	
Happy	96.2	0.24	1.1	1.1	96.2
Neutral	0.3	94.8	1.2	2.2	94.8
Sad	2.4	2.46	93.4	3.4	93.4
Angry	1.2	2.5	4.3	93.3	93.3

4 Results and Discussion

The confusion matrix technique is used to check the performance of classification algorithms. If the number of observations for each class is not equal, or the dataset contains multiple classes, the accuracy of the classification will only be misleading. The confusion matrix calculation can provide a good understanding of the types of errors generated by the classification model. This matrix can be used for problems that can be understood by two classes, but it can also be easily applied to problems with three or more class values by adding rows and columns to a messy matrix. Table 2 gives the values of the confusion matrix used to test the accuracy of the data used. For example, if user is sad, the correct percentage is 93.4%.

5 Comparison with Existing Methods

Finally, Table 3 compares the proposed work with other work using the same DEAP dataset. The proposed work gives better results with an accuracy rate of 94.425%. However, Ahirwal and Kose classification gives better results than Nawaz et al. and Li et al. However, due to the number of levels and calculations, method can be time-consuming. The results given by SVM are better than to ANN.

Table 3 Comparison of proposed method

Authors	Algorithm classification	Accuracy (%)
Nawaz et al. [16]	SVM, KNN, DT	77
Ahirwal and Kose [21]	ANN	93
Li et al. [17]	SVM	76.67
Proposed method	SVM	94.425

6 Conclusion and Future Scope

In this paper, we have proposed an emotion recognition system-based EEG data. The proposed method performed well in terms of accuracy. It can be seen from the results that the happy state achieves 96.2% results compared to other states. In a neutral state, 94.8% recognition rate was observed. However, detection rate of angry and sadness has been reduced by 1.1 and 1.2%, respectively. In the future, we will consider using more physical parameters for future development and improvement of the system which can increase the recognition rate of the system. Currently only four basic emotions are calculated, more emotions, i.e., fear, stress, etc., may be calculated in the future.

References

1. Jerry S, Krusienski JW (2012) Brain-computer interfaces in medicine. Mayo Clin Proc 87:268–279
2. Rattanyu K, Mizukawa M (2011) Emotion recognition using biological signal in intelligent space. In: Human-computer interaction. Springer, pp 586–592
3. Al-Nafjan A, Hosny M, Al-Wabil A, Al-Ohali Y (2017) Classification of human emotions from electroencephalogram (EEG) signal using deep neural network. Int J Adv Comput Sci Appl 8:419–425
4. Bhatti AM, Majid M, Anwar SM, Khan B (2016) Human emotion recognition and analysis in response to audio music using brain signals. Comput Human Behav 65:267–275
5. Zheng WL, Zhu JY, Lu BL (2017) Identifying stable patterns over time for emotion recognition from EEG. IEEE Trans Affect Comput 10:417–429
6. Russell JA (1980) A circumplex model of affect. J Pers Soc Psychol 39(6):1161–1178
7. Russell JA, Mehrabian A (1977) Evidence for a three-factor theory of emotions. J Res Pers 11(3):273–294
8. Al-Nafjan A, Hosny M, Al-Ohali Y, Al-Wabil A (2018) Recognition of affective states via electroencephalogram analysis and classification. In: Proceedings of international conference on intelligent human systems integration. Springer, Cham, Switzerland, pp 242–248
9. Tripathi S, Acharya S, Sharma RD, Mittal S, Bhattacharya S (2017) Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset. In: Proceedings of 29th IAAI conference, pp 4746–4752
10. Chen J, Hu B, Wang Y, Dai Y, Yao Y, Zhao S (2016) A three-stage decision framework for multi-subject emotion recognition using physiological signals. In: Proceedings of IEEE international conference on bioinformatics and biomedicine (BIBM), pp 470–474

11. Zhuang N, Zeng Y, Yang K, Zhang C, Tong L, Yan B (2018) Investigating patterns for self-induced emotion recognition from EEG signals. Sensors 18(3):841
12. García-Martínez B, Martínez-Rodrigo A, Cantabrana RZ, García JMP, Alcaraz R (2016) Application of entropy-based metrics to identify emotional distress from electroencephalographic recordings. Entropy 18(6):221
13. Lan Z, Sourina O, Wang L, Scherer R, Müller-Putz GR (2019) Domain adaptation techniques for EEG-based emotion recognition: a comparative study on two public datasets. IEEE Trans Cogn Dev Syst 11(1):85–94
14. Liu Y, Fu G (2021) Emotion recognition by deeply learned multi-channel textual and EEG features. Future Gener Comput Syst 119
15. Rahman MA, Hossain MF, Hossain M, Ahmmed R (2020) Employing PCA and t-statistical approach for feature extraction and classification of emotion from multichannel EEG signal. Egypt Inform J 21:23–35
16. Nawaz R, Cheah K, Nisar H, Yap V (2020) Comparison of different feature extraction methods for EEG-based emotion recognition. Biocybern Biomed Eng 40(3):910–926
17. Li Z et al (2020) Enhancing BCI-based emotion recognition using an improved particle swarm optimization for feature selection. Sensors 20:3028
18. Liu W, Qiu J, Zheng WL, Lu BL (2019) Multimodal emotion recognition using deep canonical correlation analysis. arXiv preprint arXiv:1908.05349
19. Ren F, Dong Y, Wang W (2019) Emotion recognition based on physiological signals using brain asymmetry index and echo state network. Neural Comput Appl 31:4491–4501
20. Pane ES, Hendrawan MA, Wibawa AD, Purnomo MH (2017) Identifying rules for electroencephalograph (EEG) emotion recognition and classification. In: 5th international conference on instrumentation, communications, information technology, and biomedical engineering (ICICI-BME). IEEE, pp 167–172
21. Ahirwal MK, Kose MR (2018) Emotion recognition system based on EEG signal: a comparative study of different features and classifiers. In: Second international conference on computing methodologies and communication (ICCMC). IEEE, pp 472–476
22. Mert A, Akan A (2018) Emotion recognition from EEG signals by using multivariate empirical mode decomposition. Pattern Anal App 21:81–89
23. Ackermann P, Kohlschein C, Bitsch JA, Wehrle K, Jeschke S (2016) EEG-based automatic emotion recognition: feature extraction, selection and classification methods. In: IEEE 18th international conference on e-health networking, applications and services (Healthcom). IEEE, pp 1–6
24. Koelstra S, Muehl C, Soleymani M, Lee J-S, Yazdani A, Ebrahimi T, Pun T, Nijholt A, Patras I (2012) DEAP: a database for emotion analysis using physiological signals (PDF). IEEE Trans Affect Comput 3(1):18–31