

Cluster-based intelligent recommendation system for hybrid healthcare units

Edited by

Sultan Ahmad, Surbhi Bhatia Khan, Amin Ul Haq
and Hikmat A. M. Abdeljaber

Published in

Frontiers in Medicine



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-5929-1
DOI 10.3389/978-2-8325-5929-1

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Cluster-based intelligent recommendation system for hybrid healthcare units

Topic editors

Sultan Ahmad — Prince Sattam Bin Abdulaziz University, Saudi Arabia

Surbhi Bhatia Khan — University of Salford, United Kingdom

Amin Ul Haq — University of Electronic Science and Technology of China, China

Hikmat A. M. Abdeljaber — Applied Science Private University, Jordan

Citation

Ahmad, S., Khan, S. B., Ul Haq, A., Abdeljaber, H. A. M., eds. (2025). *Cluster-based intelligent recommendation system for hybrid healthcare units*.

Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5929-1

Table of contents

- 05 **Exploring the mechanism of agarwood moxa smoke in treating sleep disorders based on GC–MS and network pharmacology**
Nianhong Chen, Yucheng Xia, Weiyang Wu, Siyu Chen, Mingming Zhao, Yanting Song and Yangyang Liu
- 18 **Feature elimination and stacking framework for accurate heart disease detection in IoT healthcare systems using clinical data**
Wang Jian, Jian Ping Li, Amin Ul Haq, Shakir Khan, Reemiah Muneer Alotaibi, Saad Abdullah Alajlan and Md Belal Bin Heyat
- 36 **Efficient differential privacy enabled federated learning model for detecting COVID-19 disease using chest X-ray images**
Rawia Ahmed, Praveen Kumar Reddy Maddikunta, Thippa Reddy Gadekallu, Naif Khalaf Alshammari and Fatma Ali Hendaoui
- 52 **Intelligent diagnosis system based on artificial intelligence models for predicting freezing of gait in Parkinson's disease**
Abdullah H. Al-Nefaie, Theyazn H. H. Aldhyani, Nesren Farhah and Deepika Koundal
- 73 **Enhanced cardiovascular disease prediction through self-improved Aquila optimized feature selection in quantum neural network & LSTM model**
Aman Darolia, Rajender Singh Chhillar, Musaad Alhussein, Surjeet Dalal, Khursheed Aurangzeb and Umesh Kumar Lilhore
- 91 **Intelligent tumor tissue classification for Hybrid Health Care Units**
Muhammad Hassaan Farooq Butt, Jian Ping Li, Jiancheng (Charles) Ji, Waqar Riaz, Noreen Anwar, Faryal Farooq Butt, Muhammad Ahmad, Abdus Saboor, Amjad Ali and Mohammed Yousuf Uddin
- 104 **Enhanced feature selection and ensemble learning for cardiovascular disease prediction: hybrid GOL2-2T and adaptive boosted decision fusion with babysitting refinement**
S. Phani Praveen, Mohammad Kamrul Hasan, Siti Norul Huda Sheikh Abdullah, Uddagiri Sirisha, N. S. Koti Mani Kumar Tirumanadham, Shayla Islam, Fatima Rayan Awad Ahmed, Thowiba E. Ahmed, Ayman Afrin Noboni, Gabriel Avelino Sampedro, Chan Yeob Yeun and Taher M. Ghazal
- 122 **Utilizing deep learning models in an intelligent eye-tracking system for autism spectrum disorder diagnosis**
Nizar Alsharif, Mosleh Hmoud Al-Adhaileh, Mohammed Al-Yaari, Nesren Farhah and Zafar Iqbal Khan

- 137 **Hybrid healthcare unit recommendation system using computational techniques with lung cancer segmentation**
Eid Albalawi, Eali Stephen Neal Joshua, N. M. Joys, Surbhi Bhatia Khan, Hadil Shaiba, Sultan Ahmad and Jabeen Nazeer
- 153 **Multivariate pattern analysis of medical imaging-based Alzheimer's disease**
Maitha Alarjani and Badar Almarri
- 172 **A deep neural network-based approach for seizure activity recognition of epilepsy sufferers**
Danial Khurshid, Fazli Wahid, Sikandar Ali, Abdu H. Gumaei, Samah M. Alzanin and Mogeib A. A. Mosleh
- 188 **Systematic survey on data security in wireless body area networks in IoT healthcare system**
Wang Jian, Alia Tabassum and Jian Ping Li
- 197 **DeepCGAN: early Alzheimer's detection with deep convolutional generative adversarial networks**
Imad Ali, Nasir Saleem, Musaed Alhussein, Benazeer Zohra, Khursheed Aurangzeb and Qazi Mazhar ul Haq
- 208 **Deep learning infused SIRVD model for COVID-19 prediction: XGBoost-SIRVD-LSTM approach**
Hisham Alkhalefah, D. Preethi, Neelu Khare, Mustafa Haider Abidi and Usama Umer
- 220 **Utilizing deep learning models in an intelligent spiral drawing classification system for Parkinson's disease classification**
Nesren Farhah
- 235 **Enhancing healthcare recommendation: transfer learning in deep convolutional neural networks for Alzheimer disease detection**
Purushottam Kumar Pandey, Jyoti Pruthi, Saeed Alzahrani, Anshul Verma and Benazeer Zohra
- 247 **Comparing circular and flexibly-shaped scan statistics for disease clustering detection**
Lina Wang, Xiang Li, Zhengbin Zhang, Haoxun Yuan, Pengfei Lu and Yaru Li



OPEN ACCESS

EDITED BY

Amin Ul Haq,
University of Electronic Science and
Technology of China, China

REVIEWED BY

Riaz Ullah Khan,
University of Electronic Science and
Technology of China, China
Jalaluddin Khan,
K L University, India

*CORRESPONDENCE

Yangyang Liu
✉ yyliau@implad.ac.cn;
✉ yangyang_liu56@outlook.com
Yanting Song
✉ songyt@hainanu.edu.cn

[†]These authors have contributed equally to
this work

RECEIVED 13 March 2024

ACCEPTED 18 April 2024

PUBLISHED 09 May 2024

CITATION

Chen N, Xia Y, Wu W, Chen S, Zhao M,
Song Y and Liu Y (2024) Exploring the
mechanism of agarwood moxa smoke in
treating sleep disorders based on GC–MS and
network pharmacology.
Front. Med. 11:1400334.
doi: 10.3389/fmed.2024.1400334

COPYRIGHT

© 2024 Chen, Xia, Wu, Chen, Zhao, Song and
Liu. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Exploring the mechanism of agarwood moxa smoke in treating sleep disorders based on GC–MS and network pharmacology

Nianhong Chen^{1,2†}, Yucheng Xia^{2,3†}, Weiyan Wu^{4†}, Siyu Chen⁴,
Mingming Zhao³, Yanting Song^{1*} and Yangyang Liu^{2*}

¹Key Laboratory of Tropical Biological Resources of Ministry of Education, School of Pharmaceutical Sciences, Hainan University, Haikou, China, ²Hainan Provincial Key Laboratory of Resources Conservation and Development of Southern Medicine, Key Laboratory of State Administration of Traditional Chinese Medicine for Agarwood Sustainable Utilization, International Joint Research Center for Quality of Traditional Chinese Medicine, Haikou, China, ³College of Traditional Chinese Medicine, Hainan Medical University, Haikou, China, ⁴Chengmai County Hospital of Traditional Chinese Medicine, Haikou, China

Background: Agarwood moxibustion is a folk therapy developed by individuals of the Li nationality in China. There is evidence that agarwood moxa smoke (AMS) generated during agarwood moxibustion therapy can treat sleep disorders via traditional Chinese medicines' multiple target and pathway characteristics. However, the specific components and mechanisms involved have yet to be explored.

Objective: GC–MS (Gas Chromatography–Mass Spectrometry) and network pharmacology were used to investigate AMS's molecular basis and mechanism in treating sleep deprivation.

Method: GC–MS was used to determine the chemical composition of AMS; component target information was collected from TCMSP (Traditional Chinese Medicine Systems Pharmacology), PubChem (Public Chemical Database), GeneCards (Human Gene Database), and DisGeNet (Database of Genes and Diseases) were used to identify disease targets, and JVenn (Joint Venn) was used to identify the common targets of AMS and sleep disorders. STRING was used to construct a protein interaction network, Cytoscape 3.9.1 was used to build a multilevel network diagram of the "core components–efficacy targets–action pathways," the targets were imported into Metascape and DAVID for GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) analyses and Autodock was used for molecular docking. This research used a network pharmacology methodology to investigate the therapeutic potential of Agarwood Moxa Smoke (AMS) in treating sleep problems. Examining the target genes and chemical constituents of AMS offers insights into the molecular processes and targets of the disease.

Result: Nine active ingredients comprising anti-inflammatory substances and antioxidants, such as caryophyllene and p-cymene, found seven sleep-regulating signaling pathways and eight targets linked to sleep disorders. GC–MS was used to identify the 94 active ingredients in AMS, and the active ingredients had strong binding with the key targets. Key findings included active components with known medicinal properties, such as p-cymene, eucalyptol, and caryophyllene. An investigation of network pharmacology revealed seven signaling pathways

for sleep regulation and eight targets linked to sleep disorders, shedding light on AMS's effectiveness in enhancing sleep quality.

Conclusion: AMS may alleviate sleep disorders by modulating cellular and synaptic signaling, controlling hormone and neurotransmitter pathways, etc. Understanding AMS's material basis and mechanism of action provides a foundation for future research on treating sleep disorders with AMS. According to the study, Agarwood Moxa Smoke (AMS) may improve sleep quality by modifying cellular and synaptic signaling pathways for those who suffer from sleep problems. This might lead to the development of innovative therapies with fewer side effects.

KEYWORDS

agarwood moxa smoke (AMS), agarwood moxa stick, agarwood moxibustion therapy, sleep disorders, GC–MS, network pharmacology, molecular docking

1 Introduction

Sleep disorder syndrome, also known as sleep deprivation, occurs when there is a disruption in the onset and maintenance of sleep, leading to inadequate sleep quality that does not meet the patient's physiological needs and seriously impacts their daytime activities. This syndrome refers explicitly to circumstances in which it is difficult or impossible to fall asleep, usually when the amount of time spent asleep is inadequate, when it is easy to wake up from sleep, when it is challenging to fall back asleep after waking up, or even when one is awake the entire night (1). Epidemiological surveys indicate that sleep disorders are positively correlated with age and affect approximately 30% of the global population annually. Sleep disorders have become a widespread social problem (2), and the incidence rate of sleep problems in Chinese adults is 42.5%, with a peak incidence rate of 38.2% (3). Sleep disorders not only decrease the quality of life but also the use of benzodiazepine-type Western medicines for treatment, which, although effective quickly, increase the risk of mental and physical diseases, as well as drug dependence and abuse, is dangerous (4, 5). Therefore, there is an urgent need for more effective treatment methods with fewer side effects.

Chinese individuals frequently treat sleep issues with traditional practices such as moxibustion and massage. Traditional Chinese medicines adhere to dialectical treatment and holistic concepts to provide symptomatic treatment for patients with sleep disorders (6). Traditional Chinese medicine's external therapy approach has fewer adverse effects and is better tolerated by patients (7). This kind of moxa stick is referred to as agarwood moxa stick in this article, and the use of agarwood moxa sticks in moxibustion treatment is known as agarwood moxibustion therapy. Agarwood moxibustion therapy is a traditional Chinese medicine originating from individuals of the Li nationality. This therapy is frequently used to treat a variety of chronic, crippling illnesses, as well as diseases brought on by dampness, wind, and cold; furthermore, this treatment can significantly improve the quality of sleep. Agarwood has long been used as incense in China due to its calming and sleep-promoting properties (8). In traditional Chinese medicine, sleep disorders are thought to be caused by an imbalance among the kidneys, liver, spleen, and heart. The practical volatile components produced by the combustion of agarwood and moxa velvet can be absorbed through the skin, protect the spleen and

kidneys, relieve pain, calm the mind, and promote immunity, thereby reducing insomnia symptoms (9, 10). Researchers have examined the components and pharmacological effects of moxa smoke (11, 12), and their findings show that it can lower the levels of aspartic acid and glutamate in the brain.

After burning the agarwood and moxa velvet, the volatile components also have a pleasant, calming effect that can help extend the time spent asleep (13). Sleep problems have resulted from the disruption of healthcare systems, everyday routines, and sleep habits caused by the COVID-19 pandemic. Sleep disturbances are made worse by elements like elevated stress, anxiety, schedule adjustments, and social isolation. Sleep has been further disrupted by uncertainty, fear of infection, and health worries. Screen use has grown during lockdowns and quarantine periods, which has a detrimental effect on sleep quality. Controlling the virus's transmission and reducing its downstream impacts on public health, such as sleep problems, depend heavily on early forecasting and detection. Addressing these issues and controlling the transmission of the virus are crucial for mitigating the downstream impacts on public health, including sleep problems. Early intervention measures play a pivotal role in this effort, as they enable timely treatment to address both the direct and indirect effects of the pandemic on sleep and overall well-being (14). In some regions of China, agarwood has been used for the prevention of COVID-19, partly based on its reported aromatic purifying properties (15). Therefore, using agarwood moxibustion can not only to some extent achieve early prevention of diseases, but also provide early intervention for sleep problems caused by the spread of COVID-19 (16). While the direct effects of agarwood moxibustion in combating COVID-19 require further scientific validation, its traditional use underscores its potential benefits in public health crises. However, more research needs to be done investigating the components of moxa smoke that are produced during moxibustion treatment and determining how these components affect insomnia. Therefore, to explore the connection between the chemical components of AMS and sleep and to further elucidate the mechanism of AMS-mediated treatment of sleep disorders, in this study, we investigated the practical components of AMS using GC–MS in conjunction with network pharmacology.

2 Materials

2.1 Medicinal material

The agarwood moxa sticks used in this study were made in the laboratory, with agarwood purchased from Haikou, Hainan, and moxa velvet from Qichun, Hubei. The agarwood slices were crushed, run through a sieve, and mixed evenly with moxa velvet in specific proportions. Then, the samples were laid flat on smokeless paper, rolled tightly, fixed, and finally, a 1.8 cm diameter agarwood moxa stick was made.

2.2 Instrument

A Manual SPME Holder (Lot: 155193, Merck, USA), an Agilent gas chromatography-mass spectrometer (Model: 5977B/8860, Agilent Technologies, USA), and a PDMS SPME solid-phase microextraction head (Lot: 163109, Merck, USA) were utilized.

2.3 Databases and software

The Traditional Chinese Medicine Systems Pharmacology database and analysis platform (TCMSP)¹ (17); the PubMed database² (18); the Swiss target prediction database³ (19); the Pharmacochemical Database (ChEMBL)⁴ (20); the Universal Protein database (UniProt)⁵ (21); the GeneCards database⁶ (22); the DisGeNet database⁷ (23); the Database for Annotation, Visualization and Integrated Discovery v6.8 (DAVID)⁸ (24); the Jvenn website⁹ (25); the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING)¹⁰ (26); the Metascape database¹¹; and the RCSB Protein Data Bank database (RCSB PDB)¹² (27) were utilized. Cytoscapev3.9.1 software, Autodock book software, and Pymol software were also utilized.

In summary, the application of Agarwood Moxa Smoke (AMS) in treating sleep disorders is fraught with difficulties, such as problems with data completeness and quality, access to extensive databases, and intricate integration of disparate information. New methods may resist traditional medical practices, and integrating research results into clinical practice may encounter obstacles from regulatory bodies, medical professionals, and patients. Researchers, healthcare professionals, regulatory agencies, and traditional medicine practitioners must

work together to remove these obstacles and encourage broader adoption of AMS-based therapy.

3 Methods

3.1 Collection of AMS

The agarwood moxa stick was ignited and precipitated in a homemade glass collection tank; after 3 min, when the flue gas was complete, a manual sampler was inserted, and the handle was pressed to extend the extraction head. After 10 min of extraction, the extraction head was retracted, and the manual holder was removed. The manual holder was immediately inserted into the gas chromatograph sample inlet (temperature 230°C) for 3 min of analysis without splitting the sample.

3.2 GC–MS analysis

An HP-5MS elastic quartz capillary column (30 mm*0.25 mm, 0.25 µm) was utilized. The carrier gas was high-purity helium, with a volume flow rate of 1 mL/min, a sample inlet temperature of 250°C, and a detector temperature of 300°C. Programmed heating was performed as follows: after maintaining a column temperature of 50°C for 1 min, the temperature was raised at a rate of 15°C/min to 143°C, kept for 10 min, raised at a rate of 1°C/min to 155°C, raised at a rate of 25°C/min to 225°C, maintained for 7 min, raised at a rate of 2°C/min to 250°C, and finally maintained for 10 min. The electron bombardment (EI) energy was 70 eV, the ion source temperature was 250°C, the solvent delay was 5 min, and the scanning range was 50–500 amu.

3.3 Prediction of the sources and targets of the chemical components in AMS

The traditional Chinese medicine system pharmacology analysis platform (TCMSP) (17) was utilized to search and identify the chemical components in AMS based on the chemical composition data derived from the GC–MS analysis. Traditional Chinese medicine (TCM) uses Agarwood Moxa Smoke (AMS), combining conventional knowledge and cutting-edge scientific techniques. It provides focused treatment for sleep disturbances, which may result in fewer adverse effects and more successful results. By bridging the gap between conventional knowledge and contemporary understanding, scientific validation via GC–MS analysis and network pharmacology increases acceptability and trustworthiness. Because agarwood moxa stick is a traditional Chinese medicine mixture, its smoke contains numerous complex chemical components of traditional Chinese medicine. Based on Lipinski's Rules of Five, the potential effective active ingredients were identified. For chemical components not found in the TCMSP database, the PubChem database (18) was used to search the Smiles number of the element. Next, an AMS chemical component library was created by searching the pertinent target databases Swiss Target Prediction (19) and the ChEMBL database (20). Finally, the UniProt database (21) was used to standardize target gene names, and only

- 1 <https://old.tcm-sp-e.com/tcm-sp.php>
- 2 <https://pubchem.ncbi.nlm.nih.gov/>
- 3 <http://www.swisstargetprediction.ch/>
- 4 <https://www.ebi.ac.uk/chembl/>
- 5 <http://www.uniprot.org>
- 6 <https://www.genecards.org/>
- 7 <https://www.disgenet.org/>
- 8 <https://david.ncifcrf.gov/>
- 9 <http://jvenn.toulouse.inra.fr/app/example.html>
- 10 <https://cn.string-db.org/>
- 11 <https://Metascape.org/>
- 12 <https://www.rcsb.org/>

human (*Homo sapiens*) target genes were retained for subsequent analysis.

3.4 Identification of targets in sleep disorders

Due to the mind-tranquilizing effects of both moxa velvet and agarwood in agarwood moxa sticks, this study further utilized the Gene Cards (22) and DisGeNet databases (23) to search for genes associated with sleep disorders; “Sleep disorders” was used as the keyword to search for targets in sleep disorders, and gene names were standardized by the DAVID v6.8 database (24). Only pertinent values ≥ 1.5 were chosen as the primary targets for sleep disorders to ensure the validity of the data.

3.5 Construction of the protein–protein interaction network

On the Jvenn platform (25), the targets of AMS components were compared to the disease targets related to sleep disorders, and the shared targets were identified. The shared targets were considered potential targets of AMS for treating sleep disorders. The shared target gene set was input into the STRING database (26). Then, a protein–protein interaction (PPI) network was constructed with *Homo sapiens* as the target species and an average confidence level of 0.4 as the threshold for the interaction score. The PPI network was visualized with Cytoscape v3.9.1 software.

3.6 Construction of the “core components-targets-action pathways” network

Cytoscape 3.9.1 software created a multilevel network of “core components, targets, and action pathways” to link AMS’s disease-related genes, core components, and targets. A network was established with circles representing disease-related genes, diamonds representing core components, and triangles representing AMS targets; this network was used to evaluate the mechanism of action of AMS in the treatment of sleep disorders.

3.7 GO functional analysis and KEGG pathway enrichment analysis

The targets for treating sleep disorders corresponding to the chemical components of AMS were input into the Metascape and David databases (24). The databases conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses to obtain biological information about potential targets and analyze AMS’s potential mechanism of action in treating sleep disorders. GO analysis results are divided into the categories of biological processes (BPs), cellular components (CCs), and molecular functions (MFs). According to the *p* values, the top 8 results from GO analysis were selected, and a histogram of enrichment quantity

statistics was drawn; for KEGG analysis, the top 7 pathways were selected for visualization.

3.8 Molecular docking validation

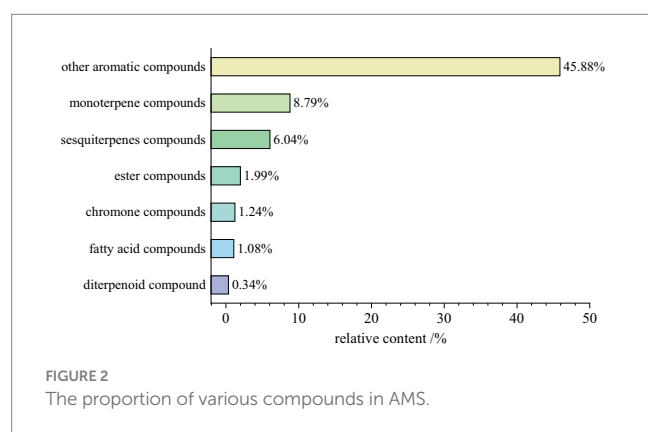
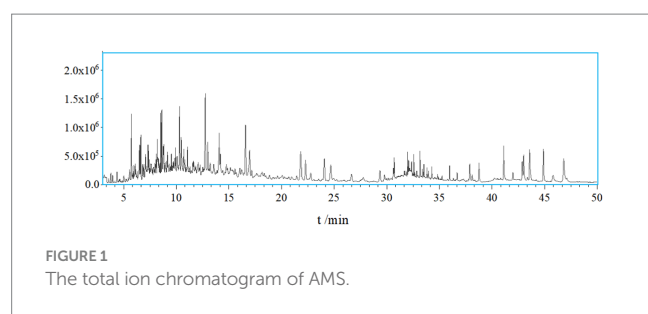
The active ingredients related to treating sleep disorders by AMS were docked to core targets to forecast and evaluate the protein–molecule interactions and binding energy. After using the TCMSP and PubChem databases to obtain mol2 files of compound structures, the RCSB PDB database (27) was used to obtain PDB files of the core target structures. Using Autodock software (28) for docking and PyMOL software (29) for visualization and processing, binding energy was used as an indicator to evaluate the binding activity and docking effects of ligand–protein interactions. Generally, $-1.2 \text{ kcal}\cdot\text{mol}^{-1}$ binding energy indicates strong binding between the protein and ligand.

4 Results

When optimizing computer models to investigate how Agarwood Moxa Smoke (AMS) affects sleep problems, parameter tuning is an essential component. Researchers may increase their models’ efficiency, accuracy, and resilience by fine-tuning their parameters, providing more trustworthy outcomes. Molecular Docking parameter tweaking, route analysis parameters, threshold selection, network visualization, docking method settings, and binding site flexibility are essential tactics. Network quality and dependability are affected by threshold selection, while interpretability is improved by network visualization. Performance assessment, hyperparameter optimization, cross-validation, and stringent validation techniques are all part of overall model optimization.

4.1 Chemical composition of AMS

A total of 138 chemical components were identified in AMS. Among these volatile components, there were 20 compounds with a relative abundance of more than 1%, including phenol, 3-methyl-phenol, p-cymene, azulene, endo-borneol, α -terpineol, 4-ethyl-2-methoxy-phenol, caryophyllene, caryophyllene oxide, bis (2-ethylhexyl) phthalate, etc. The components of the smoke and the combustion products of agarwood moxa sticks are complex, and volatile aromatic compounds are one of the main components of AMS. The study looks at how Agarwood Moxa Smoke (AMS) affects sleep problems; however, it has trouble analyzing proprietary or unbalanced datasets. The dependability of findings can be affected by imbalanced datasets, resulting in biased model performance and decreased predicted accuracy. Class distributions can be balanced using ensemble techniques, undersampling, and oversampling. Sensitive information in proprietary datasets makes data access, exchange, and validation difficult. Real-world data assessment requires collaboration with data owners. Scalability testing, result evaluation, and robustness assessment are examples of testing capabilities. Eugenol, 4-phenyl-3-buten-2-one, 4-phenyl-2-butanone, and n-hexadecanoic acid may originate from (30). Sesquiterpene



compounds may degrade into monoterpenes and small-molecule volatile compounds after high-temperature cracking. The identified chemicals were mainly phenolic compounds containing methoxy groups and monoterpenoids; these chemicals are related to the aroma produced after the combustion of agarwood moxa sticks (Figures 1, 2; Table 1).

4.2 Identifying the targets of AMS

After searching the TCMSP database based on GC–MS results identified 94 effective standardized active ingredients for sleep disorders. By integrating target data, the potential targets of the 94 chemical components of AMS were predicted, with a total of 514 target sites. Using “Sleep disorders” as the search term, 17 gene targets were obtained by compiling the sleep disorder-related disease genes identified in multiple databases according to the described screening criteria (Figure 3, Table 2).

4.3 Construction and analysis of the PPI network

The 94 components of AMS and sleep disorders shared 17 nonoverlapping targets. Based on the STRING background network, a PPI network was constructed based on the potential targets of AMS in treating sleep disorders. The network contained 17 nodes and 34 edges, with an average node degree value of 4. According to the enrichment analysis, the PPI network was significantly enriched with a p -value < 0.47 . Based on the number of active components, the core targets were identified as the glucose transporter 1 (GLU1) gene SLC2A1 (31), the monooxygenase A gene MAOA (32), the

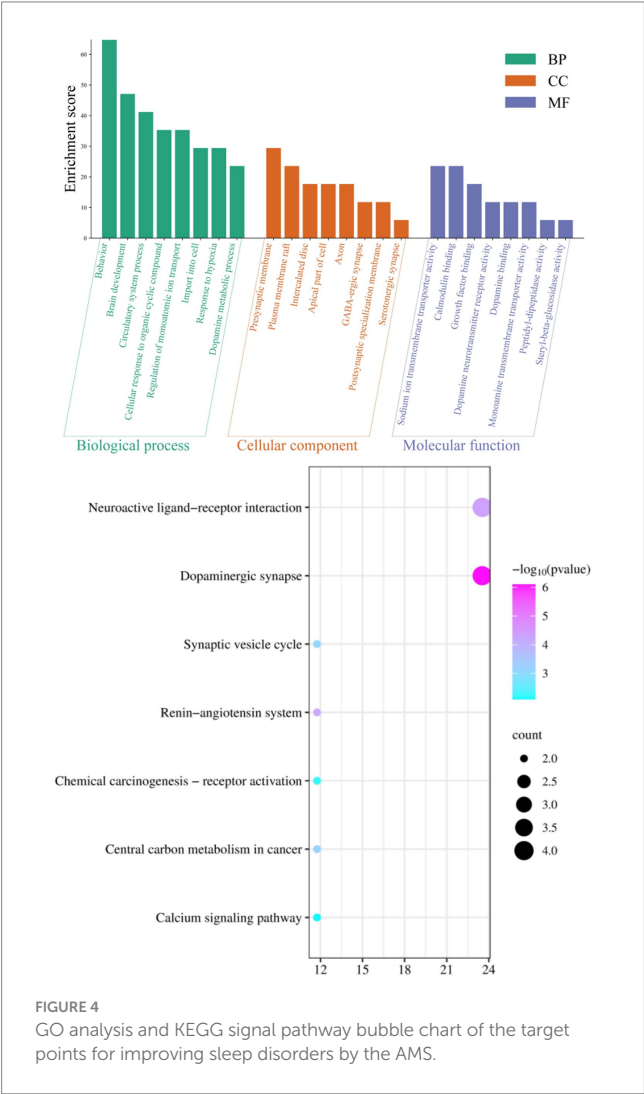
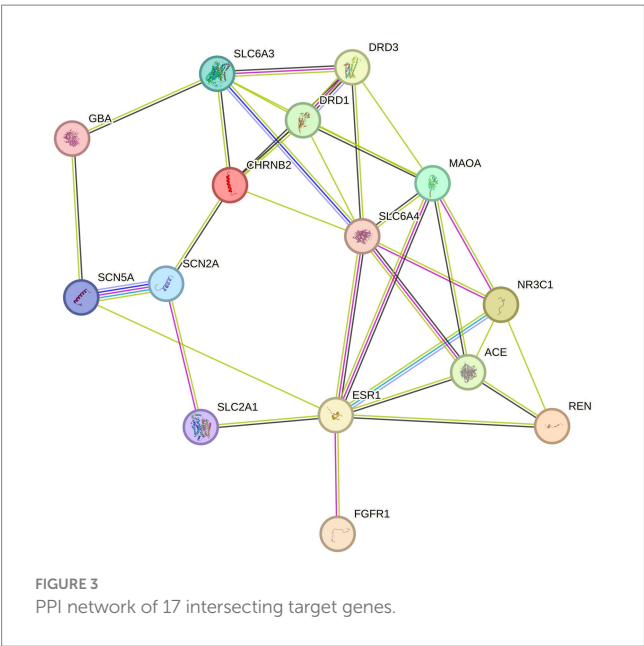
synaptic-related genes SCN2A (33), and the dopamine transporter receptor SLC6A3 (34).

4.4 GO and KEGG analysis

The integration of various datasets, network design, route analysis, molecular docking, scoring functions, and docking validation present computational overhead for the study, which looks into the therapeutic benefits of Agarwood Moxa Smoke (AMS) on sleep disorders. The more datasets and interactions there are, the greater the processing cost. Computational resources are needed throughout network building for data processing, visualization, and analysis. Pathway analysis entails processing substantial amounts of biological data to find essential pathways connected to the therapy of AMS. Due to the potential requirement to analyze several scoring methods, scoring functions incur additional computational complexity. Robust studies need a delicate balance between computational complexity and analytical depth. To gain a deeper understanding of the mechanism of action of AMS in the treatment of sleep disorders, in this study, we conducted GO and KEGG analyses of the 514 potential targets of AMS in the treatment of sleep disorders. The eight biological processes with the most significant enrichment by the components of AMS (sorted by p -value) included behavior ($p = 2.34 \times 10^{-15}$), brain development ($p = 3.98 \times 10^{-9}$), circulatory system processes ($p = 5.01 \times 10^{-9}$), cellular response to organic cyclic compounds ($p = 2.10 \times 10^{-7}$), regulation of monoatomic ion transport ($p = 6.31 \times 10^{-7}$), import into the cell ($p = 8.32 \times 10^{-8}$), response to hypoxia ($p = 3.80 \times 10^{-7}$), and dopamine metabolic processes ($p = 1.62 \times 10^{-9}$). In terms of cell components, the treatment of sleep disorders with AMS mainly involves the presynaptic membrane ($p = 2.57 \times 10^{-8}$), plasma membrane rafts ($p = 4.37 \times 10^{-7}$), intercalated disks ($p = 2.82 \times 10^{-6}$), apical parts of the cell ($p = 2.04 \times 10^{-3}$), axons ($p = 5.01 \times 10^{-3}$), GABAergic synapses ($p = 1.00 \times 10^{-3}$), postsynaptic specialization membranes ($p = 2.19 \times 10^{-3}$), and serotonergic synapses ($p = 1.12 \times 10^{-3}$). In terms of molecular function, the target genes were mainly enriched in protein cell activity and binding, including aspects such as sodium ion transmembrane transporter activity ($p = 1.62 \times 10^{-6}$), calmodulin binding ($p = 4.27 \times 10^{-6}$), growth factor binding ($p = 5.37 \times 10^{-5}$), dopamine neurotransmitter receptor activity ($p = 2.95 \times 10^{-6}$), dopamine binding ($p = 6.17 \times 10^{-6}$), monoamine transmembrane transporter activity ($p = 1.95 \times 10^{-5}$), peptidyl-dipeptidase activity ($p = 1.12 \times 10^{-3}$), and steryl-beta-glucosidase activity ($p = 1.12 \times 10^{-3}$) (Figure 4, Table 3).

Enrichment through KEGG pathway analysis showed that the effect of AMS on sleep disorders was significant. The seven KEGG pathways with considerable enrichment were dopaminergic synapse ($p = 7.94 \times 10^{-7}$), the neuroactive ligand–receptor interaction ($p = 4.47 \times 10^{-5}$), the renin-angiotensin system ($p = 7.41 \times 10^{-5}$), central carbon metabolism in cancer ($p = 7.08 \times 10^{-4}$), the synaptic vesicle cycle ($p = 8.71 \times 10^{-4}$), chemical carcinogenesis–receptor activation ($p = 6.17 \times 10^{-3}$), and the calcium signaling pathway ($p = 7.94 \times 10^{-3}$), suggesting that AMS can regulate metabolic pathways related to sleep and restore the function of metabolic pathways that were impacted by insomnia (Figure 4).

The glucose transporter gene SLC2A1, the monooxygenase A gene MAOA, the synaptic-related genes SCN2A, dopamineD1 receptor DRD1 and dopamineD3 receptor DRD3, the nicotinic acetylcholine $\beta 2$ receptors CHRN2, the 5-hydroxytryptamine transporter receptor



SLC6A4, and the dopamine transporter receptor SLC6A3 are all associated with multiple pathways and promote neurotransmitter transport and neuronal excitability. The regulation of these targets by the

active ingredients of AMS can simultaneously alter numerous signaling pathways related to sleep disorders, reflecting the “multi-component multi-target multi-pathway” approach. We mapped the dopamine receptor family DRD1, DRD3, and other targets for improving sleep disorders with the components of AMS to the dopaminergic metabolic signaling pathway and constructed a metabolic pathway map (Figure 5).

4.5 Construction of the “components-targets-action pathway” network diagram

With the help of Cytoscape 3.9.1 software, the active ingredients and their targets, as well as the targets involved in sleep disorders, were collected, and a network diagram of the “components-targets-action pathway” of AMS was constructed. The “component-target-pathway” network diagram of AMS intuitively displays the corresponding targets of each active ingredient in AMS. The blue circular nodes represent the 94 functional targets involved in sleep disorders, and the pink diamond nodes represent the 17 related genes that can help regulate sleep disorders. The connection between the component and the target indicates that the element can regulate the target (Figure 6).

4.6 Molecular docking results

The 5 main components (p-cymene, (+)-2-bornanone, endo borneol, caryophyllene oxide, and eugenol) of AMS related to the 5 key targets (SLC2A1, MAOA, SCN2A, DRD1, and DRD3) in the “component-target-pathway” network. The protein crystal structure of the target was obtained from the PDB database (PDB ID numbers: 6THA, 6EZZ, 4RLY, 7JVP, and 7CMV). The molecular docking results showed that the binding energies of (+)-2-bornanone with DRD3 and SCN2A and eugenol with DRD1 and SLC2A1 were less than $-5 \text{ kJ} \cdot \text{mol}^{-1}$, indicating that the identified components have good binding with the targets and the main active components of AMS can improve sleep disorders via multiple targets. Some receptor-ligand binding patterns are shown in Figure 7 and Table 4.

According to the study, qing moxa smoke contains 294 chemicals that may be hazardous to the kidney, liver, and heart. It highlights the complexity of the components of moxa smoke. It recommends that to protect the health of patients and practitioners, moxibustion rooms should have artificial or mechanical ventilation (35). The samp8 mice were divided into six groups, and the anti-aging effects of moxa smoking were investigated. In comparison to the normal control group, the results demonstrated a considerable drop in cerebral 5-ht, da, and ne levels. On the other hand, 5-ht and ne levels were greater in groups l2, m1, and m2, whereas da levels were higher in l2 and m2 (36). However, Table 5 shows the comparison between existing and proposed techniques as compared with the AMS results.

5 Discussion

The GC–MS analysis results show that there are 9 main active ingredients in AMS, including p-cymene (1.09%), eucalyptol (0.81%), (+)-2-bornanone (0.85%), endo-borneol (1.24%), α -terpineol (1.11%), eugenol (0.46%), caryophyllene (2.28%), caryophyllene oxide (1.79%) and n-hexadecanoic acid (0.40%). These ingredients

TABLE 1 Effective active ingredients of AMS.

No.	PubChemCID	Compound name	Relative content/%	Chemical formula	Molecular weight/g·mol ⁻¹
1	1140	Toluene	0.35	C ₇ H ₈	92.14
2	7361	2-Furanmethanol	0.17	C ₅ H ₆ O ₂	98.10
3	7975	2-methyl-Pyridine	0.27	C ₆ H ₇ N	93.13
4	7929	1,3-dimethyl-Benzene	0.18	C ₈ H ₁₀	106.16
5	7501	Styrene	0.31	C ₈ H ₈	104.15
6	14266	2-methyl-2-Cyclopenten-1-one	0.07	C ₆ H ₈ O	96.13
7	14505	1-(2-furanyl)-Ethanone	0.09	C ₆ H ₆ O ₂	110.11
8	11565	3,5-dimethyl-Pyridine	0.21	C ₇ H ₉ N	107.15
9	7668	propyl-Benzene	0.06	C ₉ H ₁₂	120.19
10	240	Benzaldehyde	0.47	C ₇ H ₆ O	106.12
11	996	Phenol	1.39	C ₆ H ₆ O	94.11
12	13381	1-Decene	0.25	C ₁₀ H ₂₀	140.27
13	252324	(Z)-1-Phenylpropene	0.32	C ₉ H ₁₀	118.18
14	9223	Benzofuran	0.29	C ₈ H ₆ O	118.13
15	7936	2,4-dimethyl-Pyridine	0.30	C ₇ H ₉ N	107.15
16	14,287	2-ethyl-6-methyl-Pyridine	0.16	C ₈ H ₁₁ N	121.18
17	7,463	p-Cymene	1.09	C ₁₀ H ₁₄	134.22
18	440,917	D-Limonene	0.49	C ₁₀ H ₁₆	136.23
19	2,758	Eucalyptol	0.81	C ₁₀ H ₁₈ O	154.25
20	335	2-methyl-Phenol	0.51	C ₇ H ₈ O	108.14
21	342	3-methyl-Phenol	1.03	C ₇ H ₈ O	108.14
22	991698552	Succinic acid, tridec-2-yn-1-yl 3-methylpentyl ester	0.33	C ₂₃ H ₄₀ O ₄	380.60
23	62385	1-methyl-4-(1-methylethenyl)-Benzene	0.53	C ₁₀ H ₁₂	132.20
24	14257	Undecane	0.41	C ₁₁ H ₂₄	156.31
25	6616	Camphene	0.19	C ₁₀ H ₁₆	136.23
26	8500	1-(4-methylphenyl)-Ethanone	0.22	C ₉ H ₁₀ O	134.17
27	8794	Benzyl nitrile	0.43	C ₈ H ₇ N	117.15
28	159055	(+)-2-Bornanone	0.85	C ₁₀ H ₁₆ O	152.23
29	9231	Azulene	1.61	C ₁₀ H ₈	128.169
30	6552009	endo-Borneol	1.24	C ₁₀ H ₁₈ O	154.25
31	9268	Cyclododecane	1.97	C ₁₂ H ₂₄	168.32
32	17100	alpha-Terpineol	1.11	C ₁₀ H ₁₈ O	154.25
33	28453	2,6-dimethyl-Undecane	0.31	C ₁₃ H ₂₈	184.36
34	637759	4-phenyl-3-Buten-2-one	0.44	C ₁₀ H ₁₀ O	146.19
35	29025	Verbenone	0.50	C ₁₀ H ₁₄ O	150.22
36	5372813	2-methyl-3-phenyl-2-Propenal	0.95	C ₁₀ H ₁₀ O	146.19
37	12581	Benzenepropanenitrile	0.50	C ₉ H ₉ N	131.17
38	17355	4-phenyl-2-Butanone	0.72	C ₁₀ H ₁₂ O	148.20
39	1550846	(E)-3-phenyl-2-Propenenitrile	0.40	C ₉ H ₇ N	129.16
40	14109	hexyl-Benzene	0.88	C ₂₄ H ₃₈ O ₄	390.60
41	62465	4-ethyl-2-methoxy-Phenol	1.13	C ₉ H ₁₂ O ₂	152.19
42	17095	1-Tridecene	2.55	C ₁₃ H ₂₆	182.35
43	12388	Tridecane	2.11	C ₁₃ H ₂₈	184.36
44	7055	2-methyl-Naphthalene	0.93	C ₁₁ H ₁₀	142.20

(Continued)

TABLE 1 (Continued)

No.	PubChemCID	Compound name	Relative content/%	Chemical formula	Molecular weight/g·mol ⁻¹
45	332	2-Methoxy-4-vinyl phenol	0.83	C ₉ H ₁₀ O ₂	150.17
46	8817	5-ethenyl-2-methyl-Pyridine	0.54	C ₈ H ₉ N	119.16
47	7041	2,6-dimethoxy-Phenol	0.62	C ₈ H ₁₀ O ₃	154.16
48	5364455	Nonene	0.76	C ₉ H ₁₈	126.24
49	3314	Eugenol	0.46	C ₁₀ H ₁₂ O ₂	164.20
50	14115	heptyl-Benzene	0.61	C ₁₃ H ₂₀	176.30
51	519194	1-Methyl-4-n-hexylbenzene	0.38	C ₁₃ H ₂₀	176.30
52	19773	2,6,10-trimethyl-Dodecane	0.56	C ₁₅ H ₃₂	212.41
53	5352912	1-Tetradecene	3.62	C ₁₄ H ₂₈	196.37
54	12389	Tetradecane	1.58	C ₁₄ H ₃₀	198.39
55	11306	1,5-dimethyl-Naphthalene	1.27	C ₁₂ H ₁₂	156.22
56	11396	2,7-dimethyl-Naphthalene	1.11	C ₁₂ H ₁₂	156.22
57	5354499	Caryophyllene	2.28	C ₁₅ H ₂₄	204.35
58	11387	2,6-dimethyl-Naphthalene	1.29	C ₁₂ H ₁₂	156.22
59	11386	2,3-dimethyl-Naphthalene	0.72	C ₁₄ H ₁₂ O ₄	244.24
60	19774	2,6,10-Trimethyltridecane	1.07	C ₁₆ H ₃₄	226.44
61	6429347	1,4-Dimethylazulene	0.55	C ₁₂ H ₁₂	156.22
62	16607	octyl-Benzene	0.55	C ₁₄ H ₂₂	190.32
63	563197	Cycloisolongifolene	0.89	C ₁₃ H ₂₄	204.35
64	5364464	Z,Z-3,13-Octadecdien-1-ol	0.60	C ₁₈ H ₃₄ O	266.50
65	25913	1-Pentadecene	2.73	C ₁₅ H ₃₀	210.40
66	12391	Pentadecane	1.99	C ₁₅ H ₃₂	212.41
67	6432455	α-Selinene	0.51	C ₁₅ H ₂₄	204.35
68	13237	2,3,6-trimethyl-Naphthalene	0.98	C ₁₃ H ₁₄	170.25
69	16479	1,4,6-trimethyl-Naphthalene	0.14	C ₁₃ H ₁₈	174.28
70	6432640	1H-Cycloprop[e]azulen-7-ol,decahydro-1,1,7-trimethyl-4-methylene-,[1ar-(1a.alpha.,4a.alpha.,7a.beta.,7b.alpha.)]-	0.23	C ₁₅ H ₂₄ O	220.35
71	1742210	Caryophyllene oxide	1.79	C ₁₅ H ₂₄ O	220.35
72	11006	Hexadecane	0.40	C ₁₆ H ₃₄	226.44
73	11877394	Neointermedeol	0.47	C ₁₅ H ₂₆ O	222.37
74	23217	1-Heptadecene	0.63	C ₁₇ H ₃₄	238.50
75	12398	Heptadecane	0.29	C ₁₇ H ₃₆	240.50
76	10719	Chamazulene	0.39	C ₁₄ H ₁₆	184.28
77	5362709	(Z)-3-Tetradecene	0.59	C ₁₄ H ₂₈	196.37
78	10,390	Diphenylacetylene	0.31	C ₁₄ H ₁₀	178.23
79	95724	3-Phenanthrol	0.29	C ₁₄ H ₁₀ O	194.23
80	8217	1-Octadecene	0.44	C ₁₈ H ₃₆	252.48
81	11635	Octadecane	0.48	C ₁₈ H ₃₈	252.48
82	79362	Phthalic acid, monooctyl ester	0.35	C ₁₆ H ₂₂ O ₄	278.34
83	6423452	Phthalic acid, butyl tetradecyl ester	0.20	C ₂₆ H ₄₂ O ₄	418.60
84	6782	1,2-Benzenedicarboxylic acid, bis(2-methylpropyl) ester	0.24	C ₁₆ H ₂₂ O ₄	278.34
85	10446	Neophytadiene	0.34	C ₂₀ H ₃₈	278.50
86	6781	Diethyl Phthalate	0.10	C ₁₂ H ₁₄ O ₄	222.24

(Continued)

TABLE 1 (Continued)

No.	PubChemCID	Compound name	Relative content/%	Chemical formula	Molecular weight/g·mol ⁻¹
87	16221	Dimethyl palmitamine	0.77	C ₁₈ H ₃₉ N	269.50
88	29075	1-Nonadecene	0.19	C ₁₉ H ₃₈	266.50
89	8222	Eicosane	0.11	C ₂₀ H ₄₂	282.50
90	985	n-Hexadecanoic acid	0.40	C ₁₆ H ₃₂ O ₂	256.42
91	8907	Isopropyl palmitate	0.08	C ₁₉ H ₃₈ O ₂	298.50
92	3015374	Henicos-1-ene	0.08	C ₂₁ H ₄₂	294.60
93	7641	Hexanedioic acid, bis(2-ethylhexyl) ester	0.20	C ₂₂ H ₄₂ O ₄	370.60
94	8343	Bis(2-ethylhexyl) phthalate	1.23	C ₂₄ H ₃₈ O ₄	390.60

TABLE 2 17 Intersecting target genes.

Numberofnodes	17
Numberofedges	34
Averagenodedegree	4
Avg.Localclustering coefficient	0.47
Expected number of edges	7
PPEnrichmentp-value	<2.88e ⁻¹⁴

generally have antioxidant, anti-inflammatory, antibacterial, and antitumor effects. In addition, p-cymene has analgesic and neuroprotective effects (37); eucalyptol can reduce neural excitability and has soothing, neuroprotective, anti-anxiety, and antidepressant effects (38). The main chemical components of AMS have activities such as pain relief and sedation, which can be used to treat sleep disorders. However, many other components of AMS and their capabilities have not been investigated or verified, which will require a more thorough study.

This study showed that SLC2A1, SLC6A3, SLC6A4, MAOA, SCN2A, DRD1, DRD3, and CHRN2 are the core targets for AMS treatment of sleep disorders. The dopamine receptor family, including DRD1, DRD3, and SLC6A3, mainly acts on dopamine (DA), while MAOA and SLC6A4 primarily affect serotonin (5-HT). SCN2A is a subunit-encoding voltage-gated sodium ion channel in the central nervous system that is widely distributed at the beginning of the axons of glutamate neurons and is involved in the regulation of hippocampal replay within sharp wave ripples (SPW-Rs), which are essential for memory (39). SCN2A can control excitatory synaptic input (40), thus regulating neuronal excitability. DRD1 is a receptor for the excitatory neurotransmitter DA, and its activity is mediated by the G protein that activates adenylate cyclase. DRD1 can increase DA levels through the DRD1 MeCP2 BDNF TrkB signaling pathway, leading to insomnia (41). Sleep-related epilepsy can be caused by mutations in genes such as CHRN2, which encodes the nAChR subunit and is widely expressed in the forebrain (42). SLC6A4 is a 5-HT transporter that can help maintain 5-HT homeostasis in the central nervous system and affect sleep by regulating 5-HT transport. SLC6A4 transports 5-HT from the extracellular compartment to the cytoplasm through the exchange of Na⁺ during the electroneutral transport cycle, thereby limiting the intercellular signal transduction of 5-HT (43). In the raphe neurons of the brainstem, the uptake of 5-HT from the synaptic

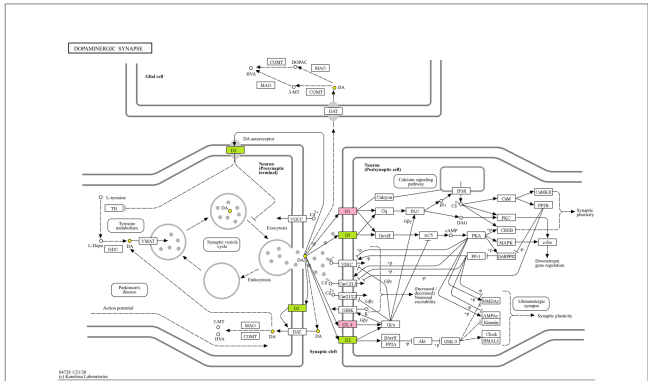


FIGURE 5
Dopaminergic synaptic pathway.

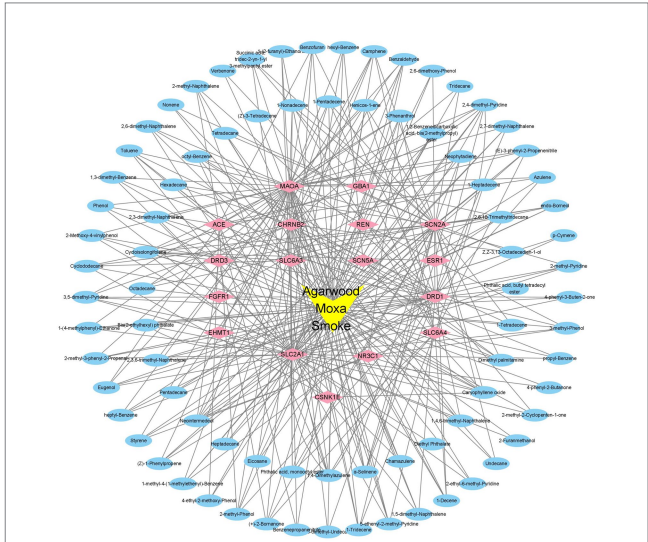


FIGURE 6
"Components-efficacy targets-action pathway network" of AMS.

gap to the presynaptic end is regulated, thereby terminating the transmission of 5-hydroxytryptamine signals at the synapse. In addition, mutations in the SLC6A4 and MAOA genes can induce structural and functional abnormalities in the dorsal raphe nucleus (DRN) and amygdala, thereby interfering with rapid eye movement

TABLE 3 Gene ontology BP, CC, MF top 8 entry information.

No.	Category	Type description	p-value
1	Biological processes	Behavior	2.34*10 ⁻¹⁵
2		Brain development	3.98*10 ⁻⁹
3		Circulatory system process	5.01*10 ⁻⁹
4		Cellular response to organic cyclic compound	2.10*10 ⁻⁷
5		Regulation of monoatomic ion transport	6.31*10 ⁻⁷
6		Import into cell	8.32*10 ⁻⁸
7		Response to hypoxia	3.80*10 ⁻⁷
8		Dopamine metabolic process	1.62*10 ⁻⁹
9	Cellular components	Presynaptic membrane	2.57*10 ⁻⁸
10		Plasma membrane raft	4.37*10 ⁻⁷
11		Intercalated disk	2.82*10 ⁻⁶
12		The apical part of the cell	2.04*10 ⁻³
13		Axon	5.01*10 ⁻³
14		GABA-ergic synapse	1.00*10 ⁻³
15		Postsynaptic specialization membrane	2.19*10 ⁻³
16		Serotonergic synapse	1.12*10 ⁻³
17	Molecular functions	Sodium ion transmembrane transporter activity	1.62*10 ⁻⁶
18		Calmodulin binding	4.27*10 ⁻⁶
19		Growth factor binding	5.37*10 ⁻⁵
20		Dopamine neurotransmitter receptor activity	2.95*10 ⁻⁶
21		Dopamine binding	6.17*10 ⁻⁶
22		Monoamine transmembrane transporter activity	1.95*10 ⁻⁵
23		Peptidyl-dipeptidase activity	1.12*10 ⁻³
24		Steryl-beta-glucosidase activity	1.12*10 ⁻³

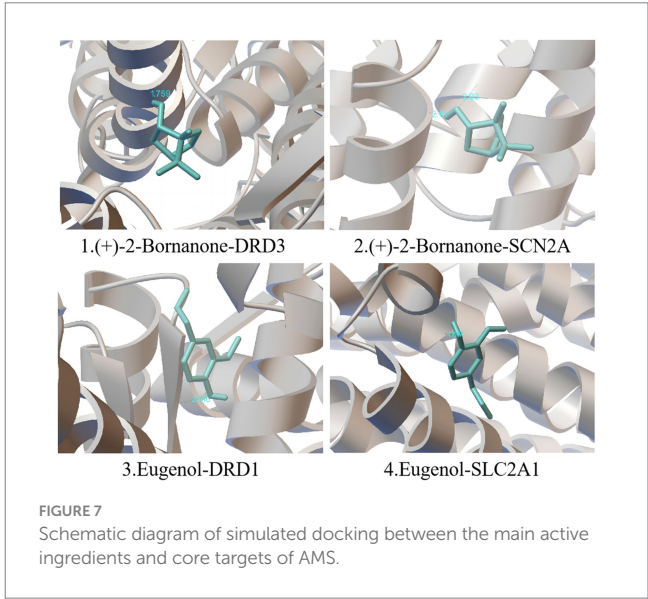
TABLE 4 Main active ingredients and core targets of AMS.

Number	Chemical compounds	Core targets	Binding energy (kcal/mol)
1	(+)-2-Bornanone	DRD3	-4.55
2	(+)-2-Bornanone	SCN2A	-3.97
3	Eugenol	DRD1	-4.76
4	Eugenol	SLC2A1	-4.82

TABLE 5 Comparison between existing and proposed study.

Reference number	Methods used	Results attained
(35)	Qing moxa smoke	89%
(36)	SAMP8 mice	91.2%
Proposed method	AMS	92.3%

(REM) sleep (44). In treating sleep disorders, AMS can mediate the expression of the genes above, thereby affecting emotions and sleep by controlling the transport and metabolism of related proteins and neurotransmitters, the permeability and exchange of Na⁺, and the transmission of synaptic signals.



The results of KEGG analysis showed that the mechanism of action of AMS in treating sleep disorders mainly involves the dopamine synaptic pathway, the neuroactive ligand-receptor

interaction, the renin-angiotensin system, the synaptic vesicle cycle pathway, and other signaling pathways. Among them, dopaminergic synapses are chemical synapses that play a crucial role in emotional disorders and can affect the connections of all members of the axonal protein superfamily of transmembrane molecules that play essential roles in neuropsychiatric disorders and excitatory cells. Excessive activation of the renin-angiotensin system pathway can lead to disturbances in the internal environment, increased reabsorption of Na^+ by the renal tubules, and elevated levels of renin and angiotensin, leading to elevated blood pressure, insomnia, anxiety, depression, and inflammation (45, 46). When insomnia occurs in the human body, the transmission of excitation signaling pathways is enhanced, and the content of molecules related to this pathway also increases. GO analysis of the active components of AMS showed that the core target genes were involved in processes such as behavior, brain development, circulatory system processes, and cellular responses to organic cyclic compounds. Therefore, based on the above results, the active ingredients of AMS can improve the behavior of individuals with sleep disorders and help maintain normal brain development and function by regulating the circulatory system; meanwhile, the active ingredients of AMS can communicate and activate various pathways through signal transduction and activating transcription factors, thereby controlling the levels of related neurotransmitters, hormones, signal molecules, and other substances.

5.1 Limitations and future scope

Using molecular docking and network pharmacology, the study explores the potential use of AMS in treating sleep disorders. However, *in vitro* and *in vivo* studies still need to be improved for better results. To comprehend the molecular processes behind the therapeutic benefits of AMS, future research should concentrate on thorough experimental investigations, clinical trials, and the integration of multi-omics data. Enhancing comprehension of AMS's therapeutic potential and creating novel treatment approaches may be accomplished by evaluating patient-centered outcomes and quality-of-life metrics.

6 Conclusion

The average amount of time people spend sleeping is steadily declining, the number of people who have insomnia is rising, and the causes of sleep disorders are becoming more complicated. Traditional Chinese medicine's diagnosis and treatment philosophy, based on syndrome differentiation and numerous potent ingredients, offers effective treatments for various insomnia-related conditions. Clinical experience has demonstrated that AMS is beneficial for qi circulation, pain relief, regulating meridians and kidneys, and calmness. In this study, we used GC-MS technology to identify the chemical components in AMS. The obtained compounds were subjected to network pharmacology analysis, and 94 active components of AMS were identified. Five hundred fourteen disease targets, 17 shared active component regulation targets, and sleep disorder-related targets were identified. A "components-targets-pathways" network

for AMS and GO was established. KEGG analyses were utilized to speculate that AMS may regulate sleep disorders through the following 7 pathways: dopaminergic synapse, the neuroactive ligand-receptor interaction, the renin-angiotensin system, central carbon metabolism in cancer, the synaptic vesicle cycle, chemical carcinogenesis-receptor activation, and calcium signaling pathways. Based on the specific roles of targets and components, 8 key targets were selected, including 9 potential active monomers. Molecular docking was carried out, and DRD3 and SCN2A showed good binding with (+)-2-Bornanone, and DRD1 and SLC2A1 showed good binding with eugenol. In this study, we explored the potential mechanisms underlying the calming and tranquilizing effects of AMS in sleep disorders from the following two aspects: material component analysis and network pharmacology, thereby providing a theoretical basis for further exploration and subsequent experimental research to evaluate the clinical application of AMS to improve sleep.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

NC: Writing – original draft, Data curation. YX: Writing – original draft, Formal analysis. WW: Writing – original draft, Funding acquisition. SC: Writing – original draft, Investigation. MZ: Writing – original draft, Methodology. YS: Writing – original draft, Project administration. YL: Writing – original draft, Conceptualization.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research study was supported by the International Science & Technology Cooperation Program of Hainan Province (GHYF2024005), the Hainan Province Major Science and Technology Plan Project (ZDZX2021034), the Research Project of the Chinese Ethnic Medicine Society (2022M10665-270301), and the Nanhai Xinxing Medical and Health Talent Platform Project of Hainan Province (NHXX-WJW-2023030).

Acknowledgments

The authors thank the members of their laboratory and their collaborators for their research work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Wu HH, Wen J, Han P, Xing HJ, Zhang XQ, Jia CS, et al. Using complex network to explore the characteristics of acupoint selection and the regularity of acupoint compatibility of acupuncture and moxibustion for insomnia. *World J Acupunct Moxibustion*. (2022) 32:61–9. doi: 10.1016/j.wjam.2021.08.006
- Vargas I, Nguyen AM, Muench A, Bastien CH, Ellis JG, Perlis ML. Acute and chronic insomnia: what has time and/or hyperarousal got to do with it. *Brain Sci*. (2020) 10:71. doi: 10.3390/brainsci10020071
- Dopheide JA. Insomnia overview: epidemiology, pathophysiology, diagnosis and monitoring, and nonpharmacological therapy. *Am J Manag Care*. (2020) 26:S76–S84. doi: 10.37765/ajmc.2020.42769
- Richey SM, Krystal AD. Pharmacological advances in the treatment of insomnia. *Curr Pharm Des*. (2011) 17:1471–5. doi: 10.2174/138161211796197052
- Sun XT. Effect of Chaihu Jialonggu oyster soup on insomnia patients with qi stagnation and anxiety and depression. *J Liaoning Univ Trad Chin Med*. (2020) 22:89–92.
- Zhang Y, Cong D, Liu P, Zhi XY, Shi C, Zhao J, et al. Study on the mechanism of regulating the hypothalamic cortical hormone-releasing hormone/corticotropin-releasing hormone type I receptor pathway by vibro-annular abdominal massage under the brain-intestine interaction in the treatment of insomnia. *Medicine*. (2021) 100:e25854. doi: 10.1097/MD.00000000000025854
- Zhuang J, Wu J, Fan L, Liang C. Observation on the clinical efficacy of traditional Chinese medicine non-drug therapy in the treatment of insomnia: a systematic review and Meta-analysis based on computer artificial intelligence system[J]. *Comput Intell Neurosci*. (2022) 2022:1–11. doi: 10.1155/2022/1081713
- Huang X, Li H, Ruan Y, Li Z, Yang H, Xie G, et al. An integrated approach utilizing raman spectroscopy and chemometrics for authentication and detection of adulteration of agarwood essential oils. *Front Chem*. (2022) 10:1036082. doi: 10.3389/fchem.2022.1036082
- Wang Z, Xu M, Shi Z, Bao C, Liu H, Zhou C, et al. Mild moxibustion for Irritable Bowel Syndrome with Diarrhea (IBS-D): A randomized controlled trial. *J Ethnopharmacol*. (2022) 289:115064. doi: 10.1016/j.jep.2022.115064
- Wang X X. *Comparative study on the curative effect of Chenxiang Moxibustion and Moxibustion on primary dysmenorrhea (cold coagulation and stasis type)*, (2023).
- Xu X, Shan S, Wang W, Liu H. Analysis of the components in moxa smoke by GC-MS and preliminary discussion on its toxicity and side effects. *Evid Based Complement Alternat Med*. (2020) 2020:2648758. doi: 10.1155/2020/2648758
- Yang J, Zheng X, Jin R, Ha L, Li J, et al. Effect of moxa smoke produced during combustion of Aiye (Folium Artemisiae Argyi) The missing references have been supplemented) on behavioral changes in mice inhaling the smoke. *J Tradit Chin Med*. (2016) 36:805–10. doi: 10.16/s0254-6272(17)30019-5
- Wang C, Wang Y, Gong B, Wu Y, Chen X, Liu Y, et al. Effective components and molecular mechanism of agarwood essential oil inhalation and the sedative and hypnotic effects based on GC-MS/MS and molecular docking. *Molecules*. (2022) 27:3483. doi: 10.3390/molecules27113483
- Wu HT, Ji CH, Dai RC, Hei PJ, Liang J, Wu XQ, et al. Traditional Chinese Medicine treatment for COVID-19: an overview of systematic reviews and meta-analyses. *J Integr Med*. (2022) 20:416–426. doi: 10.1016/j.joim.2022.06.006
- Capodice JL, Chubak BM. Traditional Chinese herbal medicine potential therapeutic application for the treatment of COVID-19. *Chin Med* (2021) 16:26. doi: 10.1186/s13020-020-00419-6
- Dai YJ, Wan SY, Gong SS, Liu JC, Li F, Kou JP, et al. Recent advances of traditional Chinese medicine on the prevention and treatment of COVID-19. *Chin J Nat Med*. (2020) 18:881–889. doi: 10.1016/S1875-5364(20)60031-0
- Ru J, Li P, Wang J, Zhou W, Li B, Huang C, et al. TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. *J Cheminform*. (2014) 6:13. doi: 10.1186/1758-2946-6-13
- Bolton EE, Wang Y, Thiessen PA, Bryant SH. Chapter 12-PubChem: integrated platform of small molecules and biological activities. *Ann Rep in Computer Chem*. (2008) 4:217–41. doi: 10.1016/S1574-1400(08)00012-1
- Gfeller D, Grosdidier A, Wirth M, Daina A, Michielin O, Zoete V. Swiss target prediction: a web server for target prediction of bioactive small molecules[J]. *Nucleic Acids Res*. (2014) 42:W32–8. doi: 10.1093/nar/gku293
- Davies M, Nowotka M, Papadatos G, Dedman N, Gaulton A, Atkinson F, et al. ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Res*. (2015) 43:W612–20. doi: 10.1093/nar/gkv352
- Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, et al. The universal protein resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res*. (2006) 34:D187–91. doi: 10.1093/nar/gkj161
- Stelzer G, Dalah I, Stein TI, Satanower Y, Rosen N, Nativ N, et al. In silico human genomics with GeneCards. *Hum Genomics*. (2011) 5:709–17. doi: 10.1186/1479-7364-5-6-709
- Bauer-Mehren A, Rautschka M, Sanz F, Furlong LI. DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. *Bioinformatics*. (2010) 26:2924–6. doi: 10.1093/bioinformatics/btq538
- Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, et al. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res*. (2007) 35:W169–75. doi: 10.1093/nar/gkm415
- Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. Jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics*[J]. (2014) 15:293. doi: 10.1186/1471-2105-15-293
- von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res*. (2003) 31:258–61. doi: 10.1093/nar/gkg034
- Prlic A, Bliven S, Rose PW, Bluhm WF, Bizon C, Godzik A, et al. Pre-calculated protein structure alignments at the RCSB PDB website[J]. *Bioinformatics*. (2010) 26:2983–5. doi: 10.1093/bioinformatics/btq572
- Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: new docking methods, expanded force field, and Python bindings[J]. *J Chem Inf Model*. (2021) 61:3891–8. doi: 10.1021/acs.jcim.1c00203
- Seeliger D, de Groot BL. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J Comput Aided Mol Des*. (2010) 24:417–22. doi: 10.1007/s10822-010-9352-6
- Deng H, Shen X. The mechanism of moxibustion: ancient theory and modern research. *Evid Based Complement Alternat Med*. (2013) 2013:379291. doi: 10.1155/2013/379291
- Tayebi N, Leon-Ricardo B, McCall K, Mehincovic E, Engelstad K, Huynh V, et al. Quantitative determination of SLC2A1 variant functional effects in GLUT1 deficiency syndrome. *ACTN*. (2023) 10:787–801. doi: 10.1002/acn3.51767
- Gaur S, Gross ME, Liao CP, Qian B, Shih JC. Effect of monoamine oxidase a (MAOA) inhibitors on androgen-sensitive and castration-resistant prostate cancer cells. *Prostate*. (2019) 79:667–77. doi: 10.1002/pros.23774
- Gould E, Kim JH. SCN2A contributes to oligodendroglia excitability and development in the mammalian brain. *Cell Rep*. (2021) 36:109653. doi: 10.1016/j.celrep.2021.109653
- Reith ME, Kortagere S, Wiers CE, Sun H, Kurian MA, Galli A, et al. The dopamine transporter gene SLC6A3: multidisease risksp. *Mol Psychiatry*. (2022) 27:1031–46. doi: 10.1038/s41380-021-01341-5
- Xu X, Shan S, Wang W, Liu H. Analysis of the components in moxa smoke by GC-MS and preliminary discussion on its toxicity and side effects. *Evid Based Complement Alternat Med*. (2020) 2020:1–13.
- Xu H, Zhao B, Cui Y, Lim MY, Liu P, Han L, et al. Effects of moxa smoke on monoamine neurotransmitters in SAMP8 mice. *Evid Based Complement Alternat Med*. (2013) 2013:1–6. doi: 10.1155/2013/178067
- Balahbib A, el Omari N, Hachlafi NE, Lakhdar F, el Menyiy N, Salhi N, et al. Health beneficial and pharmacological properties of p-cymene. *Food Chem Toxicol*. (2021) 153:112259. doi: 10.1016/j.fct.2021.112259
- Moreira P, Matos P, Figueirinha A, Salgueiro L, Batista MT, Branco PC, et al. Forest biomass as a promising source of bioactive essential oil and phenolic compounds for Alzheimer's disease therapy. *Int J Mol Sci*. (2022) 23:8812. doi: 10.3390/ijms23158812
- Takahashi S, Tanaka R, Takeguchi R, Kuroda M, Akaba Y, Ito Y. The role of molecular analysis of SLC2A1 in the diagnostic workup of glucose transporter 1 deficiency syndrome. *J Neurol Sci*. (2020) 416:117041. doi: 10.1016/j.jns.2020.117041
- Zandi PP, Jaffe AE, Goes FS, Burke EE, Collado-Torres L, Huuki-Myers L, et al. Amygdala and anterior cingulate transcriptomes from individuals with bipolar disorder reveal downregulated neuroimmune and synaptic pathways. *Nat Neurosci*. (2022) 25:381–9. doi: 10.1038/s41593-022-01024-6

41. Wang HG, Bavley CC, Li A, Jones RM, Hackett JE, Bayley Y, et al. Scn2a severe hypomorphic mutation decreases excitatory synaptic input and causes autism-associated behaviors. *JCI Insight*. (2021) 6:e150698. doi: 10.1172/jci.insight.150698
42. Shen B, Zhang D, Zeng X, Guan L, Yang G, Liu L, et al. Cannabidiol inhibits methamphetamine-induced dopamine release via modulation of the DRD1-MeCP2-BDNF-TrkB signaling pathway. *Psychopharmacology*. (2022) 239:1521–37. doi: 10.1007/s00213-021-06051-y
43. Becchetti A, Grandi LC, Cerina M, Amadeo A. Nicotinic acetylcholine receptors and epilepsy. *Pharmacol Res*. (2023) 189:106698. doi: 10.1016/j.phrs.2023.106698
44. Yang D, Gouaux E. Illumination of serotonin transporter mechanism and role of the allosteric site. *Sci Adv*. (2021) 7:l3857. doi: 10.1126/sciadv.abl3857
45. Ji Q, Li SJ, Zhao JB, Xiong Y, du XH, Wang CX, et al. Genetic and neural mechanisms of sleep disorders in children with autism spectrum disorder: a review. *Front Psych*. (2023) 14:1079683. doi: 10.3389/fpsy.2023.1079683
46. Jarrin DC, Alvaro PK, Bouchard MA, Jarrin SD, Drake CL, Morin CM, et al. Insomnia and hypertension: A systematic review. *Sleep Med Rev*. (2018) 41:3–38. doi: 10.1016/j.smrv.2018.02.003



OPEN ACCESS

EDITED BY

Udhaya Kumar,
Baylor College of Medicine, United States

REVIEWED BY

Hariharan Shanmugasundaram,
Vardhaman College of Engineering, India
Nagaraj M. Lutimath,
Dayananda Sagar Institutions, India

*CORRESPONDENCE

Jian Ping Li

✉ jpli2222@uestc.edu.cn

Amin Ul Haq

✉ khan.amin50@yahoo.com

RECEIVED 28 December 2023

ACCEPTED 03 April 2024

PUBLISHED 22 May 2024

CITATION

Jian W, Li JP, Haq AU, Khan S, Alotaibi RM,
Alajlan SA and Heyat MBB (2024) Feature
elimination and stacking framework for
accurate heart disease detection in IoT
healthcare systems using clinical data.
Front. Med. 11:1362397.
doi: 10.3389/fmed.2024.1362397

COPYRIGHT

© 2024 Jian, Li, Haq, Khan, Alotaibi, Alajlan
and Heyat. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Feature elimination and stacking framework for accurate heart disease detection in IoT healthcare systems using clinical data

Wang Jian¹, Jian Ping Li^{2*}, Amin Ul Haq^{2*}, Shakir Khan^{3,4},
Reemiah Muneer Alotaibi⁴, Saad Abdullah Alajlan⁴ and
Md Belal Bin Heyat⁵

¹School of Artificial Intelligence, Neijiang Normal University, Neijiang, Sichuan, China, ²School of Computer Science and Engineering, University of Electronic Science and Technology of China, Sichuan, Chengdu, China, ³University Centre for Research and Development, Department of Computer Science and Engineering, Chandigarh University, Mohali, India, ⁴College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia, ⁵CenBRAIN Neurotech Center of Excellence, School of Engineering, Westlake University, Hangzhou, Zhejiang, China

Introduction: Heart disease remains a complex and critical health issue, necessitating accurate and timely detection methods.

Methods: In this research, we present an advanced machine learning system designed for efficient and precise diagnosis of cardiac disease. Our approach integrates the power of Random Forest and Ada Boost classifiers, along with incorporating data pre-processing techniques such as standard scaling and Recursive Feature Elimination (RFE) for feature selection. By leveraging the ensemble learning technique of stacking, we enhance the model's predictive performance by combining the strengths of multiple classifiers.

Results: The evaluation metrics results demonstrate the superior accuracy and obtained the higher performance in terms of accuracy, 99.25%. The effectiveness of our proposed system compared to baseline models.

Discussion: Furthermore, the utilization of this system within IoT-enabled healthcare systems shows promising potential for improving heart disease diagnosis and ultimately enhancing patient outcomes.

KEYWORDS

heart disease, machine learning, classification, stacking, healthcare

1 Introduction

Heart disease (HD) is a serious public health problem that has affected millions of individuals worldwide according to the World Health Organization (WHO) (1, 2). Shortness of breath, muscle weakness, and swelling feet are prominent signs of HD (3). The diagnosis of HD is significantly important for patient treatment and recovery in the Medical Internet of Things system (MIoT) (4). Experts and medical specialists in MIoT systems have presented many non-invasive approaches for classifying and diagnosing cardiac disease (5). Machine learning (ML) and deep learning (DL) models are widely utilized in the design of computer-aided diagnosis systems (CAD) for the detection of heart disease (6).

Different heart disease diagnosis methods have been presented utilizing ML learning approaches in the literature. Detrano et al. (7) created an HD classification system utilizing ML algorithms. The Cleveland heart disease (CHD) dataset was used with global evolutionary and feature selection methods. Their proposed method recorded an accuracy of 77%. Humar et al. (8) proposed an HD detection method using a Neural Network (NN) and Fuzzy logic (FL). The classification accuracy of the said model was 87.4%. Palaniappan et al. (9) proposed a diagnosis method for HD diagnosis. The system was developed using ML models including Navies Bays (NB), Decision Trees (DT), and Artificial Neural Network (ANN). NB attained 86.12% accuracy, ANN achieved 88.12% accuracy, and 80.4% accuracy gained by the DT algorithm. Olaniyi et al. (10) proposed a three-phase model using the ANN for HD detection in angina that obtained an accuracy of 88.89%.

For the diagnosis of HD, Samuel et al. (11) designed an integrated model based on an ANN and Fuzzy AHP. In terms of accuracy, 91.10% was gained by the technique. Liu et al. (12) suggested a high-definition model based on Relief and rough set techniques. Their proposed method attained an accuracy of 92.32%. Mohan et al. (13) proposed an HD detection method using mixed ML algorithms. He also proposed a new strategy for selecting key features from data for effective machine learning classifier training and testing. They achieved 88.07% accuracy. Haq et al. (14) Proposed a machine learning-based diagnosis technique for identifying HD. ML models were used to detect HD. To choose the features, feature selection algorithms were utilized. For feature selection, they designed the Fast-Conditional-Mutual-Information (FCMIM) feature selection method. The proposed model (FCMIM-SVM) obtained a high accuracy of 92.37%. Tiwari et al. (15) proposed an ensemble approach for predicting cardiovascular illness. The framework (SE) employs a stacked ensemble classifier with machine learning algorithms such as ExtraTrees Classifier, Random Forest, and XGBoost. They have used different evaluation metrics for the proposed model (SE) evaluation. The proposed method obtained 92.34% accuracy.

The presented literature on the existing HD diagnosis models is shown in Table 1 in order to reach the problem gap in existing models in a systematic way. All of the prior treatments used a variety of methodologies to detect HD in its initial stages. However, all existing algorithms have low accuracy and are computationally complex to diagnose HD. The prediction accuracy of the HD detection approach, as shown in Table 1, requires significant enhancement for efficient and accurate detection of HD. Thus, the key concerns with the preceding methodologies are low accuracy and long computation times, which may be attributed to the usage of irrelevant features in the dataset. To solve these difficulties, new ways of identifying HD in IoT healthcare systems are necessary. Improving forecast accuracy is a major challenge and study area. Thus, the primary goal of this research is to develop an accurate and efficient HD diagnosis system.

In this research study, we have proposed an ML-based computer-aided diagnosis (CAD) approach for detecting HD early in the Medical Internet of Things (IoT) system. The objective is to develop a robust and efficient system that can assist healthcare professionals in accurately identifying HD in patients. In the designing of the CAD system, we applied data pre-processing

TABLE 1 Proposed models summary.

Model	FS	Data set	Acc (%)	Ref
ML algorithms	–	CHD	77	(7)
MLP + SVM	–	CHD	80.41	(16)
Hybrid MLmodel (HRFLM)	–	CHD	88.07	(13)
ANN + Fuzzy Logic (ANN-FL)	–	PID and CHD	87.4	(8)
ANN ensemble-based diagnosis system	–	CHD	89.01	(17)
IHDPS	–	–	88.12	(9)
3-phase technique using ANN	–	SCH.	88.89	(10)
XGBoost		CDHD	87.28	(18)
ANN-FUZZY-AHP	–	CHD	91.1	(11)
CART		HDD	87	(19)
RRS-HD	RFRS feature selection	SCH	92.32	(12)
HISFP	Relief, mRMR, LASSO	CHD	89	(2)
SVM		Cleveland Clinic dataset	96	(20)
FCMIM-SVM	Relief, mRMR, LASSO, and LLBFS	CHD	92.37	(14)
SE		Hungarian, Cleveland, Long Beach VA, Switzerland, and Statlog	92.34	(15)

techniques such as standard scalar and the removal of null values from the data set. To select related features from the data set, we incorporated the Recursive Feature Elimination (RFE) algorithm. This helps to balance the data for proper training of the algorithm and enhance the algorithm's predictive capability. The machine learning classifiers Random Forest (RF) and Ada Boost (AB) were used for the classification of affected and healthy control subjects. These models were trained and evaluated using the entire data set and selected feature data set. To further improve the predictive results of these models, we incorporated a stacking approach to select the best meta-classifier between the Random Forest and Ada Boost. We defined a parameter grid for grid search for both algorithms. Furthermore, a hold-out validation mechanism was utilized, and data were split for training and testing in portions of 80 and 20%, respectively. The Cleveland Heart Database was used to validate the proposed model. Different performance assessment

metrics were computed for model evaluation. The experimental results unequivocally demonstrated that our proposed model outperformed the baseline models in terms of predictive accuracy. Furthermore, its ease of use and compatibility with IoT healthcare systems make it an appealing and practical choice for heart disease prediction.

The innovative points of this research study are listed below:

- A CAD approach based on ML is designed to detect cardiac disease in its early stages in the MIoT systems.
- To normalize the dataset, we incorporated data preprocessing such as standard scalar and RFE algorithm for irrelevant feature elimination. The Random Forest and Ada Boost were trained and tested on entire selected feature datasets to classify heart disease and healthy control subjects.
- To further improve classification performance, the ensemble learning technique stacking was used to select the best meta-classifier between Random Forest and Ada Boost. The meta-classifier RF was used for the final classification.
- The proposed model performance was compared with baseline models, and our approach outperformed them. Hence, it is recommended for use in diagnosing heart disease in MIoT systems.

The structure of the remaining sections includes data collection and model methodology (Section 2), experiments (Section 3), discussion (Section 4), and conclusion (Section 5).

2 Research design

2.1 Data sets

The Cleveland heart disease dataset (CHD) (<https://www.kaggle.com/datasets/aavigan/cleveland-clinic-heart-disease-dataset>) is being examined for testing purposes in this study. Furthermore, for cross-validation of the models, we incorporated the data set Heart Statlog Cleveland Hungary (SCH) (<https://iee-dataport.org/open-access/heart-disease-dataset-comprehensive>).

2.2 Methodology

The proposed methodology is described in the following subsections:

1) *Recursive Feature Elimination (RFE) algorithm for feature selection*: feature selection is the process of selecting a subset of relevant features from a larger set of available features in a dataset. It is a critical step in machine learning and data analysis, as it helps improve model performance, reduce overfitting, and enhance interpretability. Feature selection also reduces the computation time of machine learning [Algorithm 1](#). RFE is a feature selection technique commonly used in machine learning to identify the most relevant features in a dataset. It aims to find the subset of features that are most relevant to a given machine learning task. It starts by taking a feature matrix X of shape $(n \text{ samples}, n \text{ features})$ and a target variable y of shape $(n \text{ samples})$ as input. Additionally, a

machine learning model is chosen to perform the feature selection process.

The RFE algorithm begins by initializing an empty list called “selected features” to store the indices of the selected features. It also creates another list of remaining features, which initially contains all the indices of the features in the “original feature” matrix.

The algorithm enters a while loop that continues until the number of selected features in selected features reaches the desired target number of features N . Inside the loop, the model is trained using the trained model and gets importance scores procedure.

This procedure fits the model on the subset of features given by X [: remaining features] and y . It then calculates the importance scores for each feature using a specific method provided by the chosen model. The importance scores represent the relevance or contribution of each feature to the model’s performance.

Next, the algorithm utilizes the least important feature procedure to identify the index of the least important feature based on the importance scores. This feature is then appended to the selected feature list and removed from the remaining feature list. The algorithm proceeds by selecting the subset of features from the original feature matrix X using the indices in the selected feature list, resulting in a new matrix called X selected. The model is then retrained using this reduced feature set by applying the train model procedure, which fits the model on selected X and y . The loop continues until the number of selected features reaches the target number N . At this point, the algorithm terminates, and the selected features list contains the indices of the optimal feature subset, according to the RFE algorithm. The RFE algorithm offers several advantages, including improved model interpretability, enhanced generalization capabilities, and reduced overfitting. By iteratively eliminating the least important features and retraining the model, RFE enables the identification of the most informative features for the given task, leading to more accurate and efficient models.

Pseudo-code for the Recursive Feature Elimination (RFE) algorithm is shown in [Algorithm 1](#).

2.3 Proposed classification algorithms

2.3.1 Random Forest ensemble learning algorithm

Random Forest (RF) ([21](#)) is an ensemble learning algorithm that combines multiple decision trees to make predictions. It is widely used for classification and regression tasks in machine learning. The algorithm creates subsets of the original dataset through bootstrapping and constructs decision trees by recursively partitioning the data based on feature splits. The final prediction is determined by aggregating the predictions of all the trees in the ensemble. Random Forest is known for its robustness against overfitting, ability to handle large datasets, and feature importance estimation. However, it can be computationally expensive and less interpretable compared with single decision trees. The hyperparameters with essential values of random forest are shown in [Table 2](#).

Data: Data set D , Original features matrix X , total instances in the data set n , target instance R_k , k is i^{th} instance, Target output classes label is y , $w[i]$, Target number of features " N "

Result: Reduced feature set S .

- Begin**
- Step 1:** The algorithm takes as input a feature matrix " X " of shape " $(n \text{ samples}, n \text{ features})$," a target variable " y " of shape " $(n \text{ samples})$," a chosen machine learning model, and a target number of features " N ";
- Step 2:** It initializes two lists: "selected features" to store the indices of the selected features and "remaining features" to keep track of the indices of the features that have not been selected yet;
- Step 3:** The algorithm enters a loop that continues until the desired number of features " N " is reached;
- Step 4:** Inside the loop, the algorithm trains the model on the remaining features, calculates the importance scores for each feature, and identifies the least important feature;
- Step 5:** The least important feature is appended to the "selected features" list and removed from the "remaining features" list;
- Step 6:** The model is retrained using the updated feature subset;
- Step 7:** The loop continues until the desired number of features " N " is obtained;
- Step 8:** Finally, the algorithm returns the "selected features" list containing the indices of the selected features;
- Finish;**

Algorithm 1. Recursive Feature Elimination (RFE) algorithm.

2.3.2 Ada Boost ensemble learning algorithm

AdaBoost (AB) (22) is an ensemble learning algorithm that puts together weak learners to form a strong classifier. It iteratively trains weak learners on weighted data, focusing on misclassified samples. The resulting prediction is a weighted combination of weak learners' predictions. AdaBoost handles complex decision boundaries and achieves high accuracy but can be sensitive to noise and outliers. The hyperparameters with essential values of Ada Boost algorithm are shown in Table 3.

2.4 Stacking model based on Random Forest and Ada Boost algorithms

The stacking approach is an ensemble technique for training several base classifiers on the same dataset. Instead of making individual predictions, the predictions of these base classifiers are

TABLE 2 Random Forest hyperparameters with essential values.

Parameters name	Description	Values
N estimators	Number of decision trees in the forest	1,000
Max depth	Maximum depth of each decision tree	20
Min samples split	Minimum number of samples required to split an internal node	10
Min samples leaf	Minimum number of samples required to be at a leaf node	5
Max features	Maximum features to use for splitting at each node	Randomness FS
Bootstrap	A boolean indicating whether to use bootstrap samples for training	True or False
Criterion	Function to measure the quality of a split (e.g., Gini impurity, entropy)	Entropy
Class weight	Weights associated with each class in classification tasks to handle class imbalance	Balance
Random state	Random seed for reproducibility	None

TABLE 3 Ada Boost algorithm hyperparameters with essential values.

Parameters name	Description	Values
N estimators	Parameters determine the number of weak learners to be included in the ensemble	200
Learning rate	Controls the contribution of each weak learner to the final prediction	0.001
Base estimator	Parameter specifies the weak learner used in the ensemble	–
Algorithm	Determines the algorithm used to update sample weights during training	"SAMME.R"

combined using a meta-classifier, which is typically a model such as logistic regression, random forest, or a neural network. The meta-classifier learns to make predictions based on the outputs of the base classifiers. By combining different types of classifiers, each with its strengths and weaknesses, the stacking approach aims to leverage the diverse perspectives and expertise of the individual classifiers to improve overall classification performance. This can lead to higher accuracy and better generalization compared with using a single classifier.

In this study, we trained two base classifiers (Random Forest and Ada Boost) using the entire training set. By using these two techniques, we aimed to introduce more diversity and variation into the ensemble. The predictions of each base model, Random Forest, and Ada Boost are then combined and used to train the meta-classifier, which in this case is also a Random Forest model.

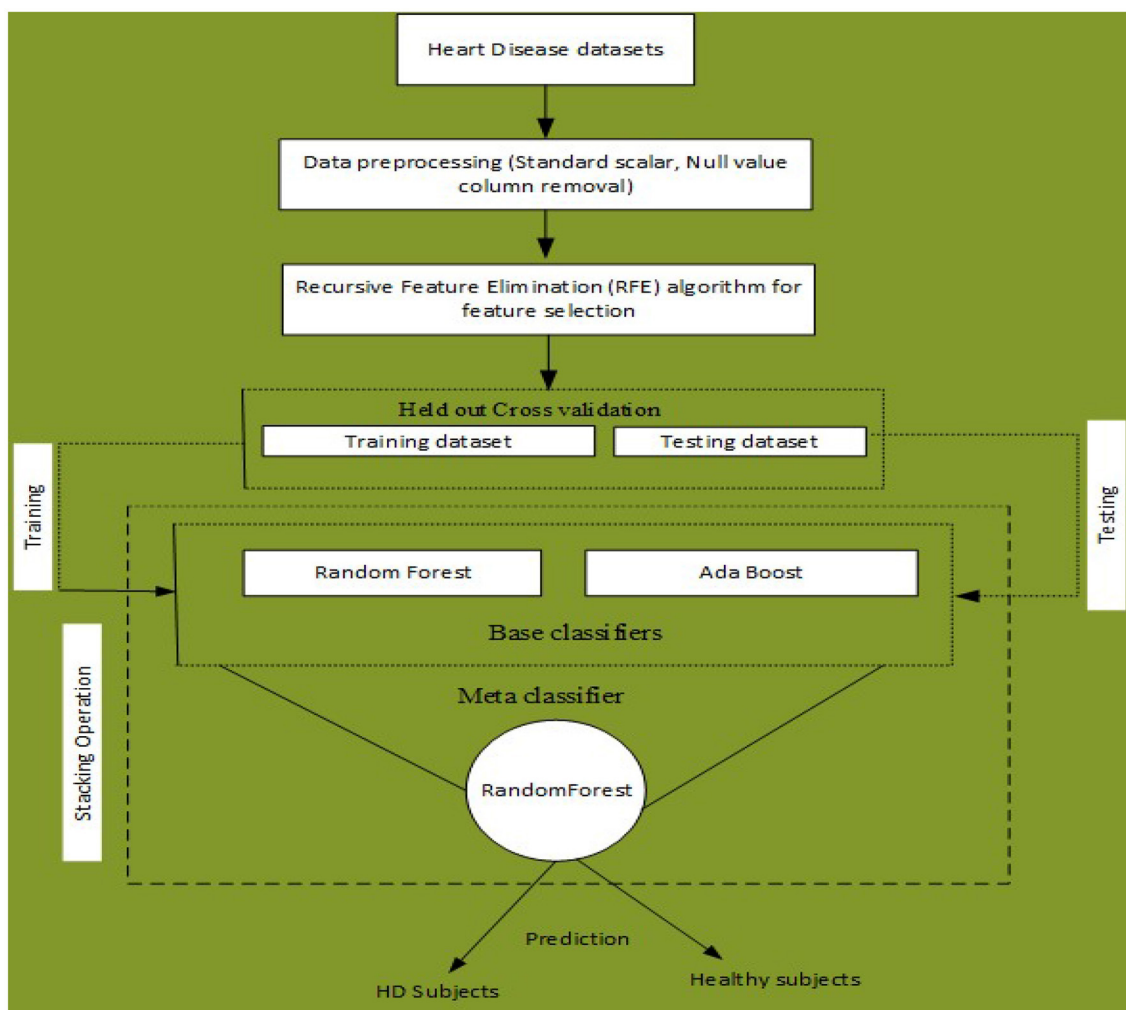


FIGURE 1

Proposed stacking-based (Stacking HD) model for Heart disease diagnosis in IoT healthcare systems.

2.5 Model cross validation

The model was trained and validated using the held-out cross-validation procedure (2). When the data set is large, the holdout CV is an appropriate validation approach. In this study, heart disease datasets such as CHD, CHDP, and SCH data sets were used and separated into 80% for training and 20% for model testing.

2.6 Performance evaluation criteria

The performance evaluation metrics (6) were used in this study to evaluate the proposed model performance. These evaluation metrics were expressed in equations mathematically Equations 1–6, respectively. TP denotes True Positive, TN denotes True Negative, FP denotes False Positive, and FN is False Negative.

$$\text{Accuracy (Acc)} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (1)$$

$$\text{Sensitivity (Sn)} = \frac{TP}{TP + FN} \times 100 \quad (2)$$

$$\text{Specificity (Sp)} = \frac{TN}{TN + FP} \times 100 \quad (3)$$

$$\text{Precision (Pr)} = \frac{TP}{TP + FP} \times 100 \quad (4)$$

$$\text{F1-Score (F1-S)} = 2 \times \frac{Pr \times Recall}{Pr + Recall} \times 100 \quad (5)$$

Matthews correlation coefficient (MCC):

$$\text{MCC} = \frac{T_1}{\sqrt{T_2 \times T_3 \times T_4 \times T_5}} \times 100 \quad (6)$$

where $T_1 = TP \times TN - FP \times FN$, $T_2 = TP + FP$, $T_3 = TP + FN$, $T_4 = TN + FP$, and $T_5 = TN + FN$.

```

Input : Cleveland Heart Disease dataset (CHD),
        Heart Statlog Cleveland Hungary dataset
        (HeartStatlogClevelandHungary)
Output: Evaluation metrics for all features and
        selected features
1 Step 1: Data preparation;
2 StandardizeData dataset return
  StandardScale(dataset);
3 CHD  $\leftarrow$  LoadDataset("Cleveland Heart Disease
  dataset");
4 HeartStatlogClevelandHungary  $\leftarrow$ 
  LoadDataset("Heart Statlog Cleveland Hungary
  dataset");
5 StandardizedCHD  $\leftarrow$  StandardizeData(CHD);
6 StandardizedHeartStatlog  $\leftarrow$ 
  StandardizeData(HeartStatlogClevelandHungary);
7 TrainData, TestData  $\leftarrow$ 
  HoldOutSplit(StandardizedCHD);
8 Step 2: Model construction;
9 TrainRandomForestdata, num_trees
  RandomForestModel  $\leftarrow$  RandomForest(num_trees);
10 RandomForestModel.fit(data);
11 return RandomForestModel;
12 TrainAdaBoostdata, num_trees AdaBoostModel  $\leftarrow$ 
  AdaBoost(num_trees);
13 AdaBoostModel.fit(data);
14 return AdaBoostModel;
15 RFEdat return RecursiveFeatureElimination(data);
16 RandomForestModel  $\leftarrow$  TrainRandomForest(TrainData,
  num_trees=1000);
17 AdaBoostModel  $\leftarrow$  TrainAdaBoost(TrainData,
  num_trees=200);
18 SelectedFeatures  $\leftarrow$  RFE(TrainData);
19 Stacking;
20 StackingModelPredictdata, base Models
  BaseModelOutputs  $\leftarrow$  [];
21 for model in baseModels do
22 | BaseModelOutputs.append(model.predict(data));
23 end
24 StackingModelInput  $\leftarrow$ 
  Concatenate(BaseModelOutputs);
25 StackingModelOutput  $\leftarrow$ 
  RandomForestPredict(StackingModel,
  StackingModelInput);
26 return StackingModelOutput;
27 BaseModelOutputsAllFeatures  $\leftarrow$ 
  StackingModelPredict(TestData,
  [RandomForestModel, AdaBoostModel]);
28 RandomForestModelSelectedFeatures  $\leftarrow$ 
  TrainRandomForest(TrainData[SelectedFeatures],
  num_trees=1000);
29 AdaBoostModelSelectedFeatures  $\leftarrow$ 
  TrainAdaBoost(TrainData[SelectedFeatures],
  num_trees=200);
30 BaseModelOutputsSelectedFeatures  $\leftarrow$ 
  StackingModelPredict(TestData[SelectedFeatures],
  [RandomForestModelSelectedFeatures,
  AdaBoostModelSelectedFeatures]);

```

```

31 Step 3: Model performance evaluation;
32 EvaluationMetricsAllFeatures  $\leftarrow$ 
  EvaluateModel(BaseModelOutputsAllFeatures,
  TestData);
33 EvaluationMetricsSelectedFeatures  $\leftarrow$ 
  EvaluateModel(BaseModelOutputsSelectedFeatures,
  TestData);

```

Algorithm 2. Stacking HD heart disease diagnosis.

Area under the ROC curve AUC:

The AUC represents the model's ROC, and a high AUC number indicates a high-performance model. These equations represent various performance metrics commonly used in binary classification tasks.

2.7 Proposed model (stacking HD)

An ML-based computer-aided diagnosis (CAD) model for detecting HD early stages in the Medical Internet of Things (IoT) system. In the designing of the CAD system, we applied data pre-processing techniques such as standard scalar and the removal of null values from the data set. To select related features from the data set, we incorporated the Recursive Feature Elimination (RFE) algorithm. This helps to balance the data for proper training of the algorithm and enhance the algorithm's predictive capability. The machine learning classifiers Random Forest (RF) and Ada Boost (AB) were used for the classification of affected and healthy control subjects. These models were trained and evaluated using the entire data set and selected feature data set. To further improve the predictive results of these models, we incorporated a stacking approach to select the best meta-classifier between the Random Forest and Ada Boost. We defined a parameter grid for grid search for both algorithms. Furthermore, a hold-out validation mechanism was utilized and data were split for training and testing in portions of 80 and 20%, respectively. The Cleveland Heart Database was used to validate the proposed model. Different performance assessment metrics were computed for model evaluation. The experimental results unequivocally demonstrated that our proposed model outperformed the baseline models in terms of predictive accuracy. The model flowchart is shown in Figure 1, and the model's method in Algorithm 2 is as follows.

3 Experiments

3.1 Experiments setup

For the implementation of the proposed model, we performed various experiments. First, we incorporated data preprocessing and feature selection techniques to balance the data set and remove the irrelevant features from the data set. The ML classifiers Random Forest and Ada Boost were trained on 80% the original feature data set and the selected feature data set and evaluated with 20% data.

Furthermore, as shown in Tables 2, 3, additional hyperparameters were adjusted in each model accordingly. The Cleveland Heart Disease and Heart Statlog Cleveland Hungary datasets were used for validation of the models. To further improve the predictive performance, a stacking mechanism was used.

The proposed model performance was evaluated by computing various evaluation metrics. The experiments were carried out on a laptop and run with a Google collaborator accelerator. All experiments required Python v3.7 and other machine-learning libraries. Consistent values are obtained after repeating the experiments several times. The results of all experiments were provided in tables and graphed.

3.2 Results and analysis

3.2.1 Results of data pre-processing

On the Cleveland Heart Disease dataset (CHD), the proposed model was tested. The original data set has 303 records and 75 columns; however, all published studies used only 14 columns. We did pre-processing on the data set, and 6 records were discarded due to empty values. Hence, the dataset has 297 records with 13 columns and 1 output column. As a result, a features matrix of 297×13 is created. We also employed a standard scalar to verify that each feature has a mean of 0 and a variance of 1; consequently, all features have the same coefficient. Furthermore, we duplicated 297 samples three times to increase the size of the data set. The number of samples in the new data set is $3 \times 297 = 891$. As a result, the new dataset, known as the Cleveland Heart Disease Proceeded (CHDP) data set, has a matrix size of 891×13 . The description of the CHD is shown in Table 4.

For cross-validation of the models, we incorporated the data set Heart Statlog Cleveland Hungary (SCH). This dataset has 1,190 samples with 11 columns. These datasets were collected and put in one place to enhance research on CAD-related machine learning and data mining methods and perhaps eventually advance clinical diagnosis and early treatment. The feature set Statlog Cleveland Hungary data set is shown in Table 5. The models were trained with Cleveland Heart Disease of feature matrix dataset 297×13 and $3 \times 297 = 891$ and tested with Heart Statlog Cleveland Hungary data set.

3.2.2 Results of REF algorithm and feature ranking and selected feature subsets from CHD and SCH data sets

To choose the optimal collection of features from the SCH and CHD data sets, the REF FS method was utilized. Table 6 shows the feature rating and selected feature sets. According to Table 6, these feature sets have a significant influence on the classification of HD and HC control subjects. From CHD data set, the subset of selected features included SEX, CTP, EIA, PES, and VCA. While from SCH data set, the selected subset of features are SEX, CTP, FBG, EIA, and PES. We have performed experiments on full and selected feature datasets of both data sets in the coming sections in order to check the models' results on full and selected feature sets.

TABLE 4 Description of cleveland heart disease (CHD) dataset (features matrix of 297×13).

Feature name	Feature code	Feature description
Age	AGE	Age in years
Sex	SEX	Male = 1 and Female = 0
Chest pain	CTP	Atypicalangina = 1, Typicalangina = 2, Asymptomatic = 3, Nonanginalpain = 4
Resting blood pressure	RBP	mm hg, hospitalized
Serum cholesterol	SCH	In mg/dl
Fasting blood sugar >120 mg/dl	FBS	fasting blood sugar >120 mg/dl ($T = 1, F = 0$)
Resting electrocardiographic	RES	Normal = 0, STT = 1, Hypertropy = 2
Maximum heart rate	MHR	–
Exercise included angina	EIA	Yes = 1, No = 0
Old peak = ST depression included by exercise relative to rest	OPK	–
Slope peak exercise St segment	PES	Up sloping = 1, Flat = 2, Down sloping = 3
Number of major vessels (0–3) colored by fluoroscopy	VCA	–
Thallium scan	THA	Normal = 3, Fixed defect = 6, Reversible defect = 7
Lable	LB	Heart disease = 1, healthy = 0

TABLE 5 Description of Statlog Cleveland Hungary (SCH) data set (features matrix of $1,190 \times 11$).

Feature name	Feature code	Feature description
Age	AGE	Age in years
Sex	SEX	Male = 1 and female = 0
Chest pain	CTP	Atypical-angina = 1, typical-angina = 2, Asymptomatic = 3, Non-anginal-pain = 4
Resting blood pressure	RBP	mm hg, hospitalized
Serum cholesterol	SCH	In mg/dl
Fasting blood sugar >120 mg/dl	FBS	fasting blood sugar >120 mg/dl ($T = 1, F = 0$)
Resting electrocardiographic	RES	Normal = 0, STT = 1, Hypertropy=2
Maximum heart rate	MHR	–
Exercise included angina	EIA	Yes = 1, No = 0
Old peak = ST depression included by exercise relative to rest	OPK	–
Slope peak exercise St segment	PES	Up sloping = 1, flat = 2, down sloping = 3
Targrt	TG	Heart disease = 1, healthy = 0

TABLE 6 Feature ranking and selected feature subsets from CHD and SCH data sets by REF algorithm, i.e., $297 * 5 \subset 297 * 13$ and $1,190 * 5 \subset 1,190 * 11$.

Dataset	Feature name	Feature code	Feature ranking	Selected feature
CHD	Age	AGE	7	
	Sex	SEX	1	SEX
	Chestpain	CTP	1	CTP
	Resting blood pressure	RBG	8	
	Serum cholesterol	SCH	9	
	Fasting blood sugar	FBG	2	
	Resting electrocardiographic	RES	5	
	Maximum heart rate	MHR	6	
	Exercise included angina	EIA	1	EIA
	OldPeak	OPK	3	
	SlopoST	PES	1	PES
	Flourosorpy	VCA	1	VCA
	Thal	THA	4	
SCH	Age	AGE	6	
	Sex	SEX	1	SEX
	Chestpain	CTP	1	CTP
	Resting blood pressure	RBG	5	
	Serum cholesterol	SCH	7	
	Fasting blood sugar	FBG	1	FBG
	Resting electrocardiographic	RES	3	
	Maximum heart rate	MHR	4	
	Exercise included angina	EIA	1	
	OldPeak	OPK	4	EIA
	SlopoST	PES	1	
				PES

3.2.3 Results of Random Forest and Ada Boost with full and selected feature data sets

The classification performance of Random Forest and Ada Boost was evaluated on whole and selected feature datasets of CHD, CHDP, and SCH datasets, respectively. The models were configured with basic hyperparameters, as shown in Tables 2, 3. The held-out cross-validation was incorporated, and data sets were divided into 80 and 20% ratios for training and validating of the models, respectively. The model’s performance was evaluated by computing different evaluation metrics, and the results were reported and discussed in detail.

Table 7 presented the results of classifiers Random forest and Ada boost trained and evaluated on full and selected feature sets on the CHD data set. On the full feature set, obtained results are 88.33% accuracy, 88.45% specificity, 89.23% sensitivity, 94.65% precision, 91.02% MCC, and 89.02% F1-score. While on selected features set the model 89.12%, 92.24%, 88.22%, 89.98%, 93.24%, and 90.00%, respectively. The model improved accuracy $89.12-88.33 = 0.79\%$ on the selected feature set. The performance of other metrics also

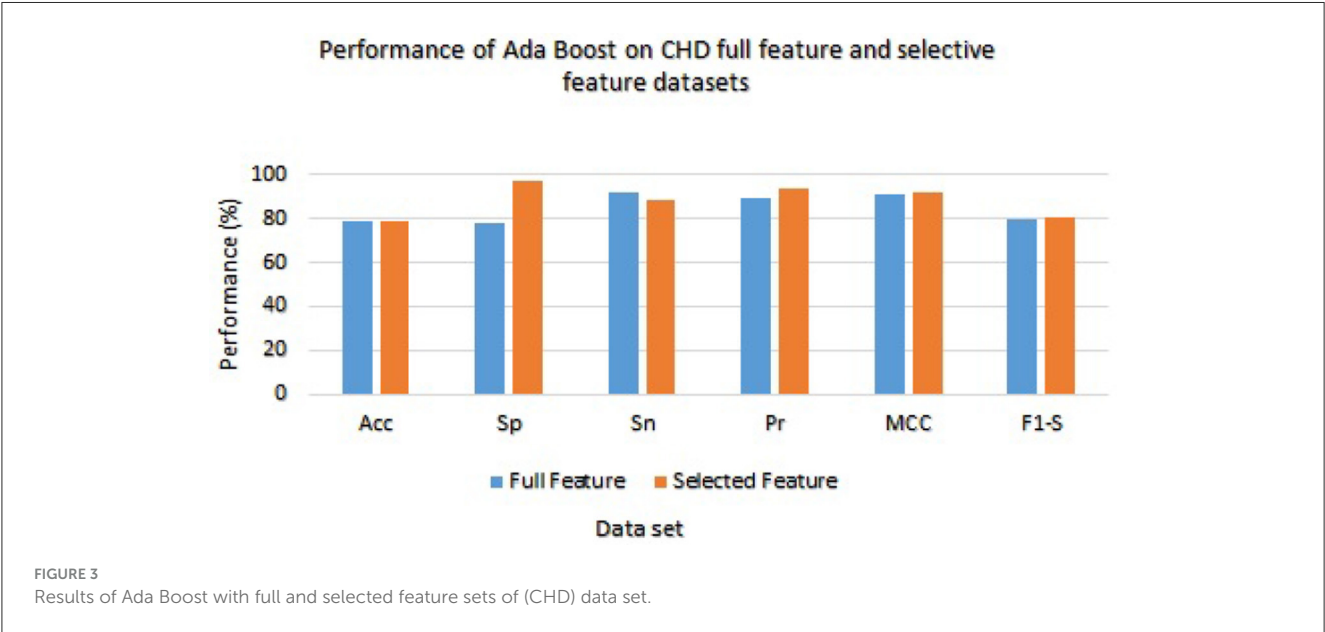
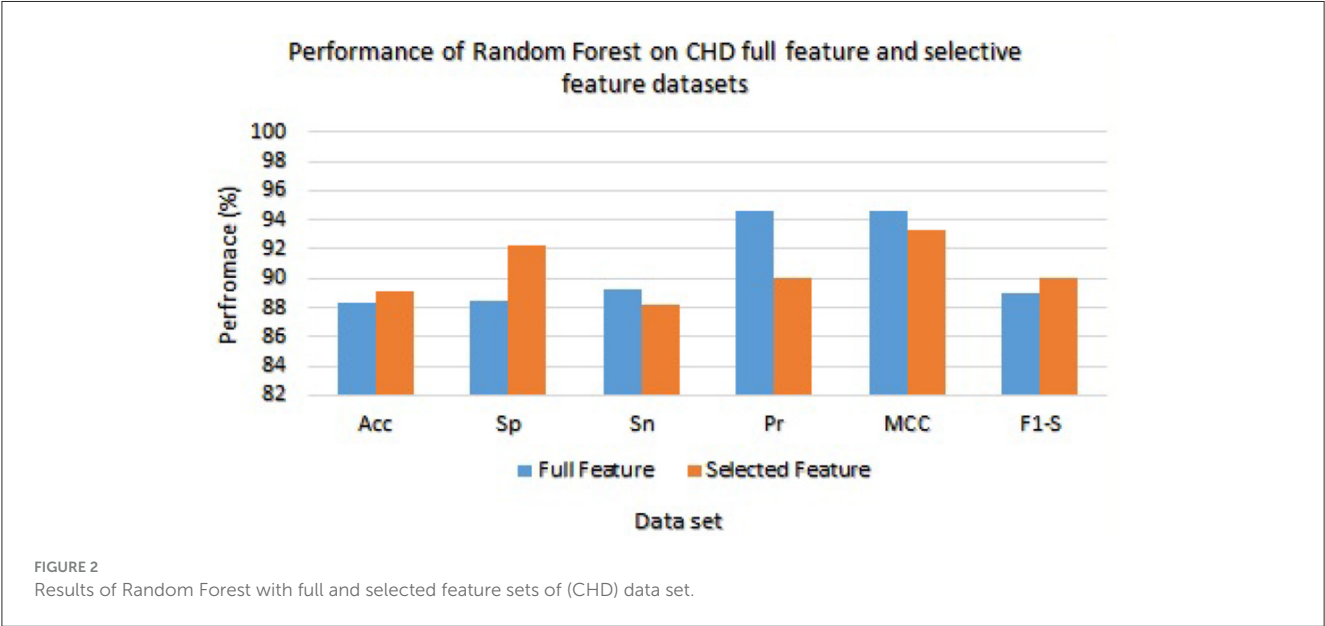
greatly improved. In Figure 2, Random Forest results are graphically presented.

The Ada Boost results are presented in Table 7 with the full feature set and obtained 78.33%, 78.21%, 92.11%, 89.34%, 91.00%, and 79.21% of accuracy, specificity, sensitivity, precision, MCC, and F1-score, respectively. On the selected feature set, the Ada Boost achieved 78.78%, 97.23%, 88.65%, 93.36%, 92.02%, and 80.58% of accuracy, sensitivity, specificity, precision, MCC, and F1-score values, respectively. Figure 3 graphically presents the model results of Ada boost on both selected and full feature data sets of CHD data set.

Table 8 presented the results of classifiers Random forest and Ada boost trained and evaluated on full and selected feature sets on the CHDP data set. The accuracy, specificity, sensitivity, precision, MCC, and F1-score values on the full feature set were 98.34%, 98.45%, 98.32%, 93.67%, 97.33%, and 98.32%, while those values on selected feature set were 98.89%, 99.00%, 98.77%, 98.67%, 96.00%, and 99.01%, respectively. The model improved accuracy $98.89-98.34 = 0.54\%$ on the selected feature set. The performance of other metrics also greatly improved. In Figure 4, Random Forest results are graphically presented.

TABLE 7 Results of Random Forest and Ada Boost with full and selected feature sets of CHD data set.

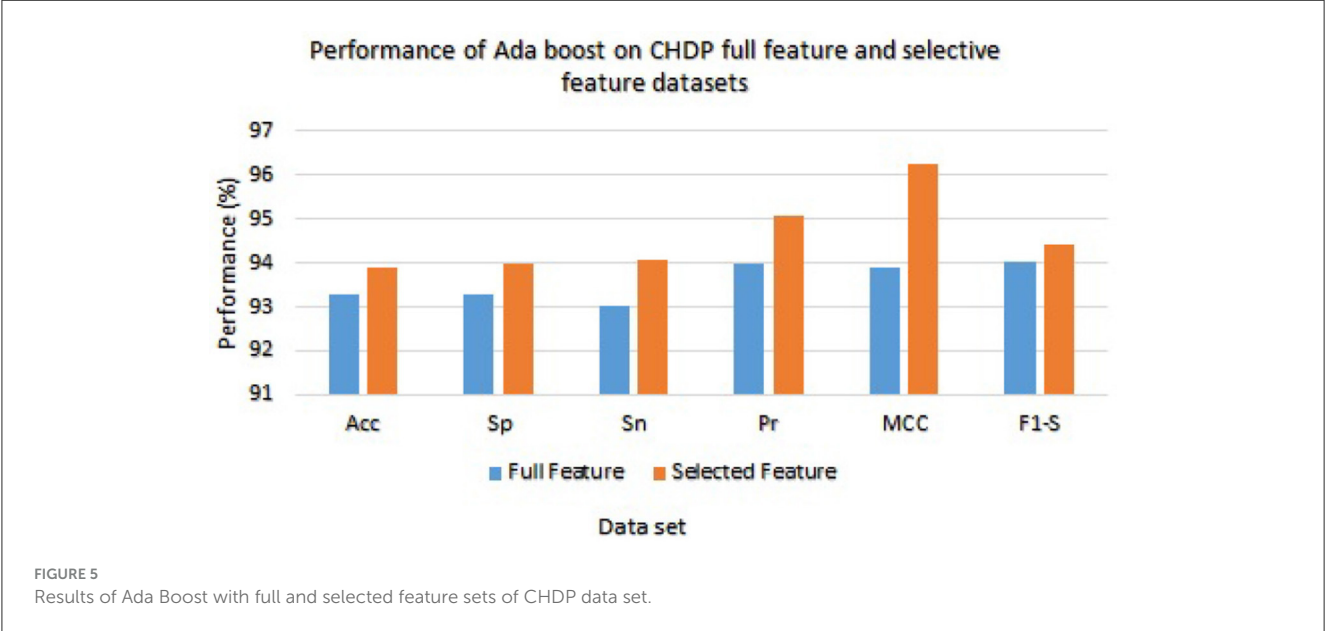
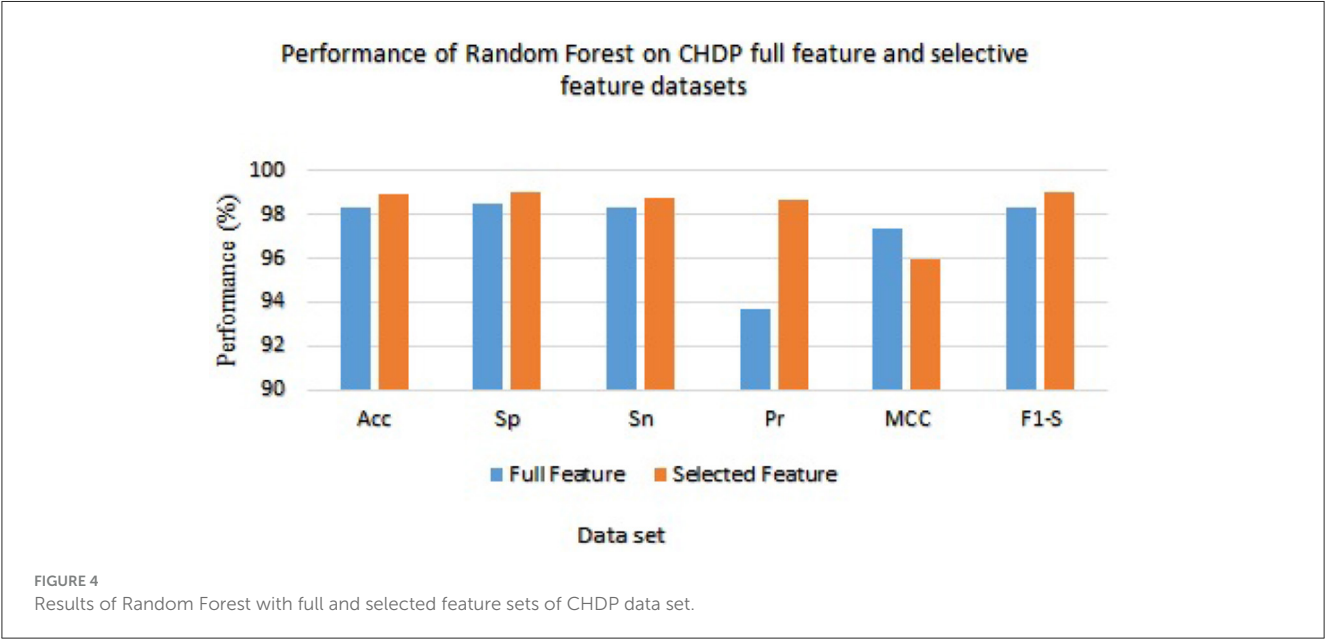
Model	Data set	Metrics					
		Acc (%)	Sp (%)	Sn (%)	Pr (%)	MCC (%)	F1-S (%)
Random forest	Full feature	88.33	88.45	89.23	94.65	91.02	89.02
–	Selected feature	89.12	92.24	88.22	89.98	93.24	90.00
Ada boost	Full feature	78.33	78.21	92.11	89.34	91.00	79.21
–	Selected feature	78.78	97.23	88.65	93.36	92.02	80.58



On the other hand, Ada Boost results with CHDP dataset are presented in Table 8 with the full feature set and obtained 93.29% accuracy, 93.28% specificity, 93.02% sensitivity, 94.00% precision, 93.89% MCC, and 94.02% F1-score. The Ada Boost achieved 93.89% accuracy, 93.89% specificity, 94.09% sensitivity, 95.09% precision, 94.23% MCC, and 94.43% F1-measure on the specified feature set. Figure 5 graphically presented the model results of Ada boost

TABLE 8 Results of Random Forest and Ada Boost with full and selected feature sets of CHDP data set.

Model	Dataset	Metrics					
		Acc (%)	Sp (%)	Sn (%)	Pr (%)	MCC (%)	F1-S (%)
Random forest	Full feature	98.34	98.45	98.32	93.67	97.33	98.32
–	Selected feature	98.89	99.00	98.77	98.67	96.00	99.01
Ada boost	Full feature	93.29	93.28	93.02	94.00	93.89	94.02
–	Selected feature	93.89	93.99	94.09	95.09	96.23	94.43



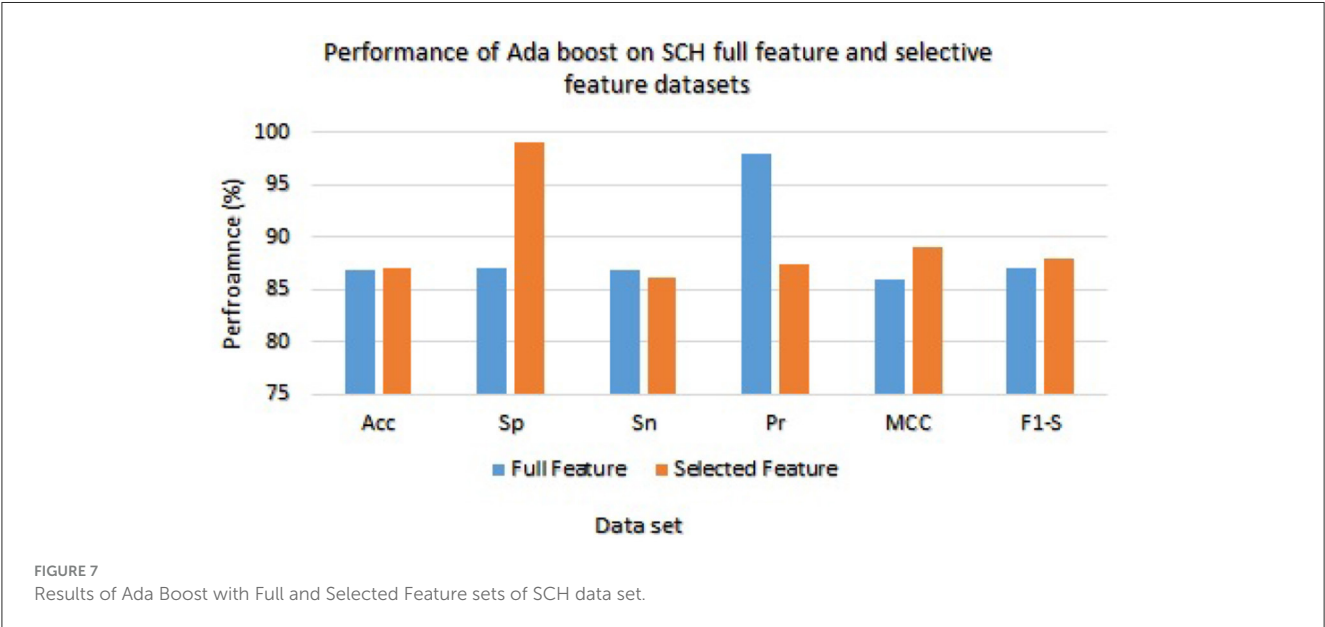
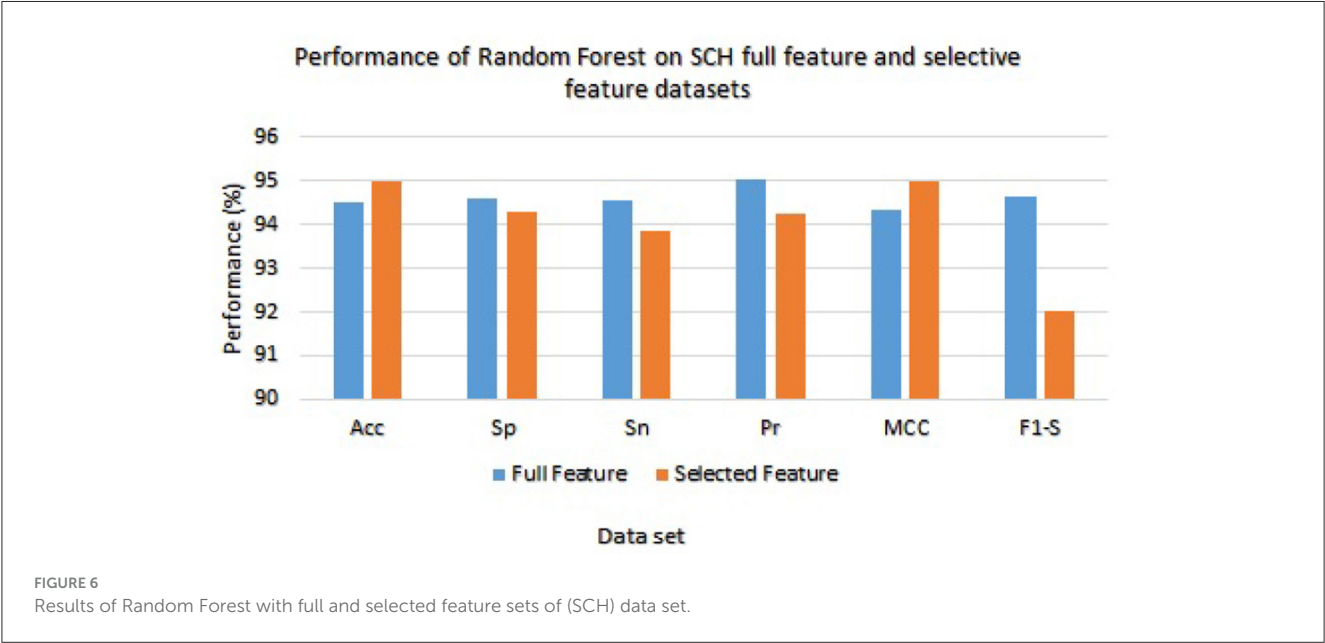
on both selected and full feature data sets of the CHDP data set.

We have checked the model’s performance on full and selected feature data sets (SCH) in order to evaluate these models.

Table 9 presented the Random Forest and Ada Boost classifier’s experimental results. With the full feature set, the Random Forest gained 94.53%, 94.59%, 94.56%, 95.02%, 94.33%, and 94.53% of accuracy, specificity, sensitivity, precision, MCC, and F1-score,

TABLE 9 Results of Random Forest and Ada Boost with full and selected feature sets of SCH data set.

Model	Dataset	Metrics					
		Acc (%)	Sp (%)	Sn (%)	Pr (%)	MCC (%)	F1-S (%)
Random forest	Full feature	94.53	94.59	94.53	95.02	94.33	94.53
–	Selected feature	95.00	94.30	93.87	94.23	95.01	92.04
Ada boost	Full feature	86.96	86.98	86.89	97.92	86.00	87.00
–	Selected feature	87.02	98.99	86.23	87.36	88.98	87.98



respectively. While accuracy, sensitivity, specificity, precision, MCC, and F1-score values on the selected feature set the Random Forest achieved 95.00%, 94.30%, 93.87%, 94.23%, 95.01%, and 92.04%, respectively. Figure 6 graphically presented the model results of Random Forest on both selected and full feature data sets of the SCH data set.

The Ada Boost results on full and selected feature data sets (SCH) are shown in Table 9. On the full feature set, the

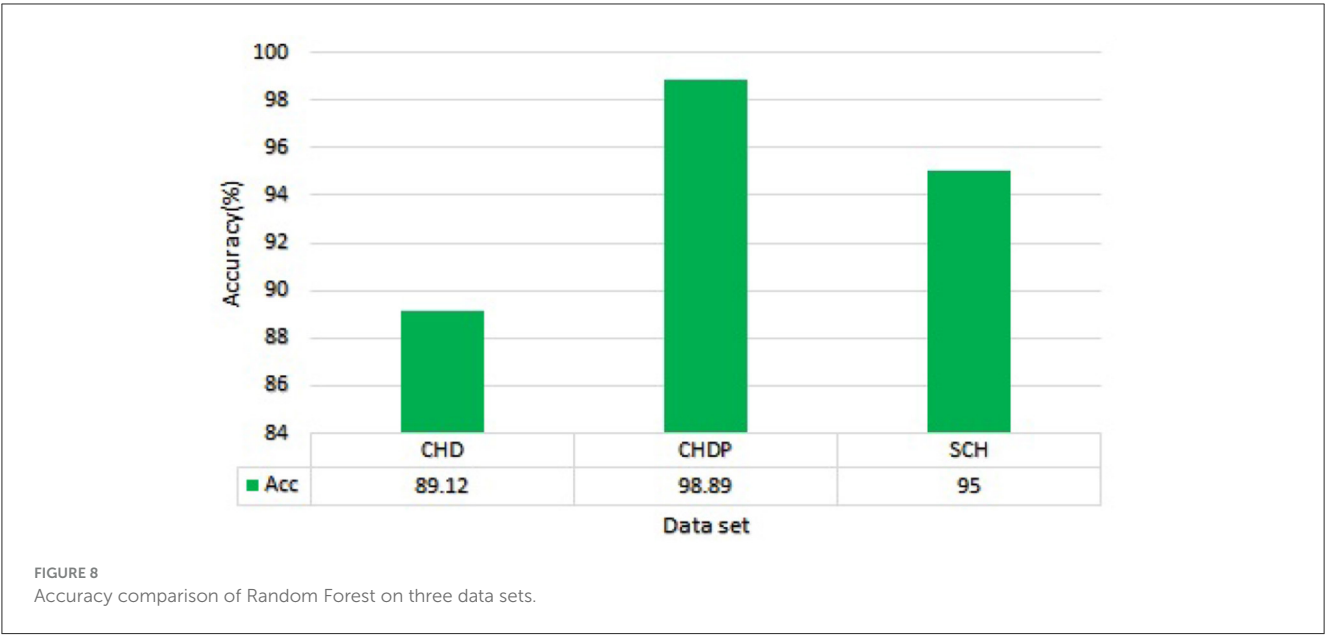
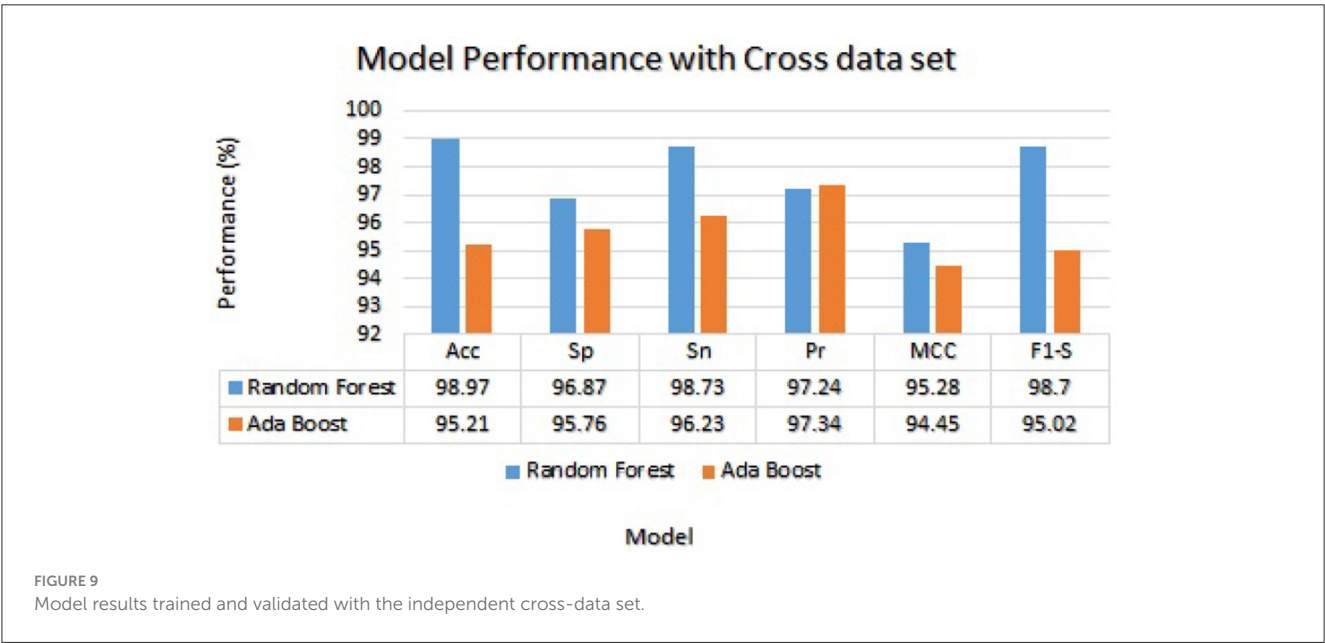


TABLE 10 Classifier evaluation with cross dataset.

Model	Metrics					
	Acc (%)	Sp (%)	Sn (%)	Pr (%)	MCC (%)	F1-S (%)
Random Forest	98.97	96.87	98.73	97.24	95.28	98.70
Ada Boost	95.21	95.76	96.23	97.34	94.45	95.02



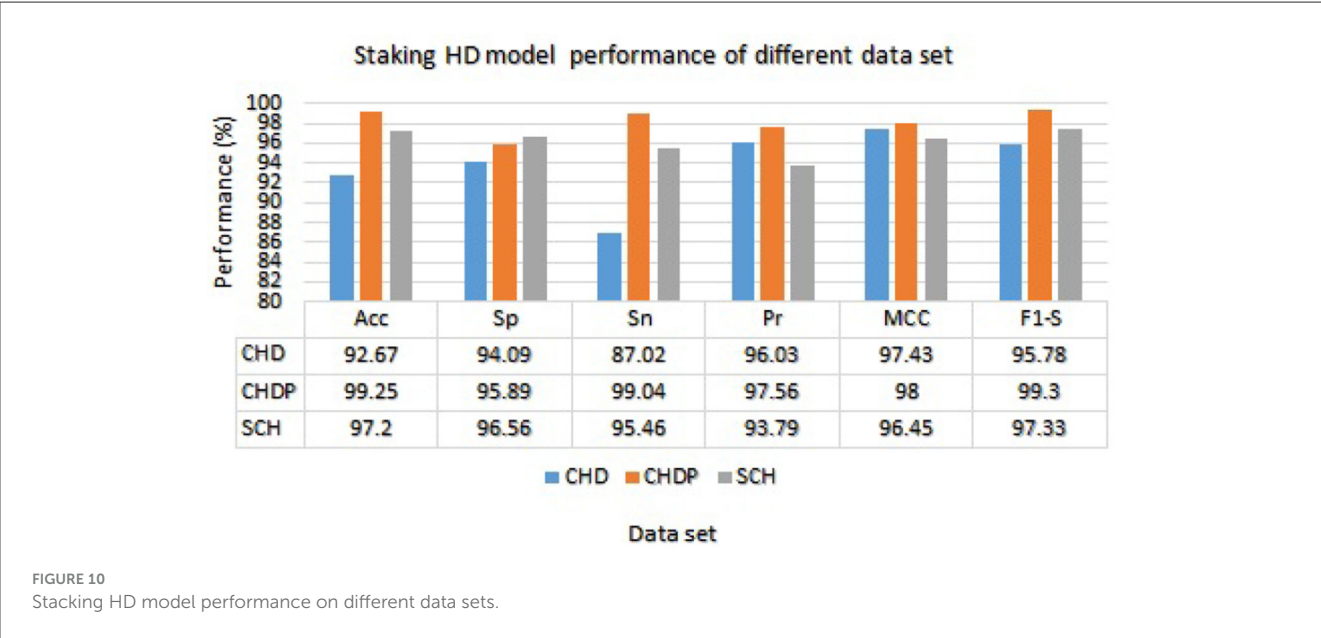
Ada boost achieved 86.96% accuracy, 86.98% specificity, 86.89% sensitivity, 97.92% precision, 86.00% MCC, and 87.00% F1-score. The Ada Boost improved predictive performance on selected feature dataset and obtained 87.02% accuracy, 98.99% specificity, 86.23% sensitivity, 87.36% precision, 88.98% MCC, and 87.98% F1-score. [Figure 7](#) graphically displayed the Ada Boost model

results on both the selected and full feature data sets of the SCH data set.

On the basis of the experimental results of Random Forest and Ada Boost classifiers on full and selected feature sets on three datasets including, CHD, CHDP, and SCH, as shown in [Tables 7–9](#), we concluded that the performance of Random Forest algorithm is

TABLE 11 Stacking HD model performance with CHD, CHDP, and SCH data sets.

Model	Datasets	Acc (%)	Sp (%)	Sn	Pr (%)	MCC (%)	F1-S (%)
Random Forest	CHD	92.67	94.09	87.02	96.03	97.43	95.78
Random Forest	CHDP	99.25	95.89	99.04	97.56	98.00	99.30
Random Forest	SCH	97.20	96.56	95.46	93.79	96.45	97.33



higher as compared with Ada Boost algorithm on CHDP data set. In terms of accuracy, Random forest with CHDP data set obtained 98.89% classification accuracy. On CHD data set, the accuracy of RF algorithm was 89.12% and the accuracy of SCH data set was 95.00%. Thus, on the basis of the data set, the Random forest classifier in CHDP data set is higher than in CHD and SCH data sets. Hence, Random Forest is a suitable classifier for the diagnosis of HD in IoT healthcare systems. The RF performance in terms of accuracy on three data sets is graphically presented in Figure 8 for better understanding.

3.2.4 Models performance evaluation with cross dataset

With separate cross-datasets, we examined the predictive outcomes of the Random Forest (RF) and Ada Boost (AB) classifiers. We trained the Random Forest and Ada Boost with CHD data set and tested with an independent SCH data set. The models were configured with basic hyperparameters as shown in Tables 2, 3. The model’s performance was evaluated by computing different evaluation metrics and experimental results, as shown in Table 10.

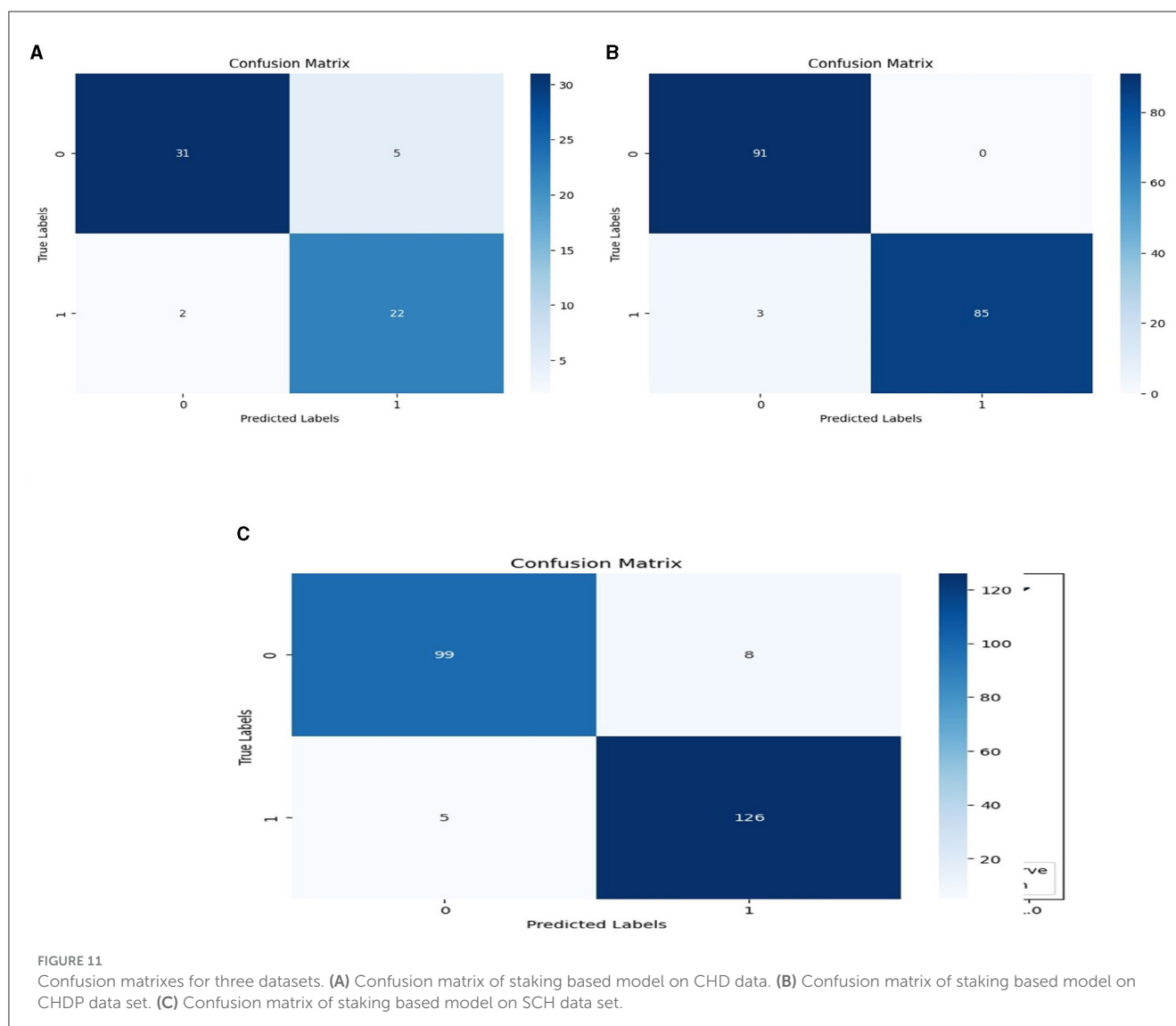
Table 10 reported performance metrics results for the random forest model including accuracy, sensitivity, specificity, precision, MCC, and F1-score which were 98.97%, 96.87%, 98.73%, 97.24%, 95.28%, and 98.70%, respectively. The test accuracy of the Random forest model is higher as compared to the test accuracy of the Ada Boost model on the same data. While the Ada Boost reached an accuracy of 95.21%, a specificity of 95.76%, a sensitivity of 96.23%,

a precision of 97.34%, MCC of 94.45%, and F1-score of 95.02%. The test accuracy is higher as compared to the test accuracy of the same data. The cross-data performance of Random Forest and Ada Boost is graphically shown in Figure 9.

3.2.5 Results of the stacking model (stacking HD)

We used the performance of all models (Random Forest and Ada Boost) as new training data to increase classification performance. The Random Forest model results were highest between Random Forest and Ada Boost models when the selected feature data sets of CHD, CHDP, and SCH were used. The outcomes of the stacking-based model (stacking HD) are shown in Table 11. The stacking-based model (stacking HD) performance of different data sets is presented graphically in Figure 10 for better understanding. The table presents that the results of the stacking-based model (stacking HD) are better and obtained 92.67% accuracy, 94.09% specificity, 87.02% sensitivity, 96.03% precision, 97.43% MCC, and 95.78% F1-score on the CHD selected feature data set. The performance of the stacking approach on CHD data is better than that of individuals models Random forest as reported in Table 7 such as 89.12% accuracy, 92.24% specificity, 88.22% sensitivity, 89.98% precision, 93.24% MCC and 90.00% F1-score. The Confusion Matrix (CM) and ROC curve of the stacking-based model on CHD data set are shown graphically in Figures 11A, 12A.

While on CHDP selected feature dataset, the stacking HD model meta classifier (Random Forest) obtained the higher performance in terms of 99.25% accuracy, 95.89% specificity,



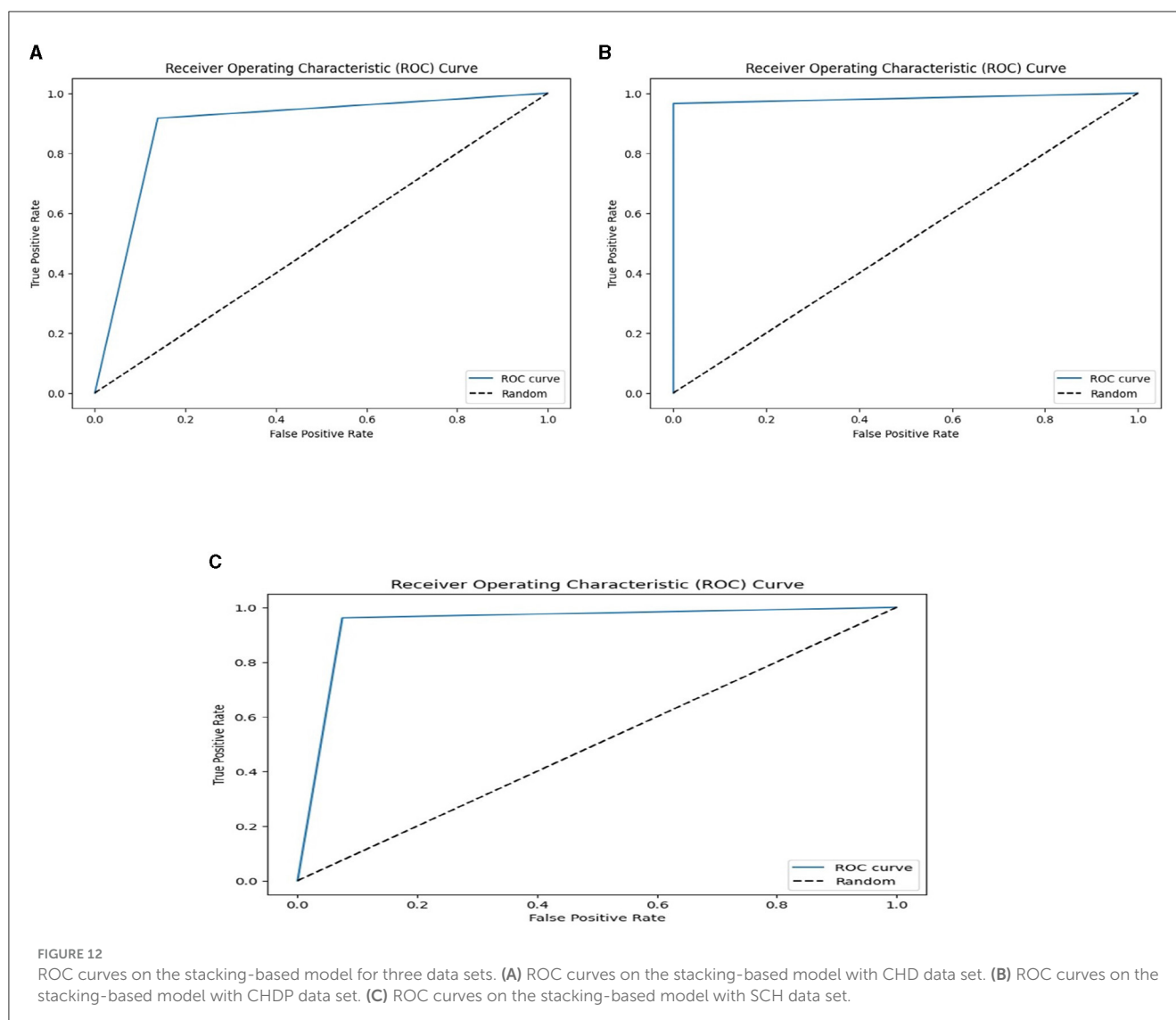
99.04% sensitivity, 97.56% precision, 98.00% MCC, and 99.30% F1-measure. The CM and ROC curves of the stacking-based model on CHDP data set are shown graphically in [Figures 11B, 12B](#). The stacking approach-based model on the SCH data set obtained 97.20% accuracy, 96.56% specificity, 95.46% sensitivity, 93.79% precision, 96.45% MCC, and 97.33% F1-score. The CM and ROC curves of the stacking-based model on SCH data set are shown graphically in [Figures 11C, 12C](#). The above stacking-based model (Stacking HD) results on different data sets presented that stacking-based models perform better than individual models. The result of the stacking-based model is the high performance of CHDP data set as compared with CHD and SCH data sets. Among the three stacking model, the stacking HD on the CHDP data set obtained a higher accuracy of 99.25%. Hence, the stacking HD model is an appropriate method to diagnose HD in its early stages. Random forest is considered as the meta classifier.

3.2.6 Comparison of stacking HD model with existing models

The proposed model (stacking HD) predictive accuracy is compared with baseline models, as shown in [Table 12](#). [Table 12](#) presented that the stacking HD model reached a higher 99.25% accuracy as compared with baseline models. The suggested method's great performance revealed that it correctly diagnoses HD and may be simply applied in IoT healthcare for the diagnosis of heart diseases.

4 Discussion

The diagnosis of heart disease (HD) is a critical task in the early stages of IoT healthcare systems. World Health Organization (WHO) reported that a large number of people are suffered from HD each year (1). To handle the initial stages of recognition of HD,



various diagnosis methods have been proposed by medical experts and researchers. Machine learning techniques based on Computer-Aided Diagnostic Systems (CAD) in an IoT healthcare system can accurately detect HD in its initial phases (30, 31). Machine learning techniques are widely used in CAD systems to diagnose critical diseases such as heart disease in IoT healthcare (32, 33). However, the existing HD diagnostic methods have the problem of lack of accuracy in the diagnosis HD correctly. The low prediction accuracy arises due to imbalanced data and irrelevant feature data for the ML model training. To address this issue, a new approach for properly and efficiently diagnosing heart disease is required for IoT healthcare systems.

The research study designed machine learning technique-based CAD systems for HD diagnosis in IoT-healthcare systems. In the designing of the CAD system, data pre-processing techniques such as standard scalar and removing null values attribute records from the data set. For related feature selection from the data set, we incorporated the Recursive Feature Elimination (RFE) algorithm to balance the data for good training of the model to enhance the model's predictive capability. The machine learning classifiers

Random Forest and Ada Boost were used for the classification of affected and healthy control subjects. These models were trained and evaluated using the entire data set and selected feature data set. To further improve the predictive performance of these models, we incorporated a stacking approach to select the best meta-classifier between the Random Forest and Ada Boost. We defined a parameter grid for grid search for both algorithms.

Furthermore, the held-out validation procedure was used, and data were split into sections of 80 and 20% for training and testing. The proposed model was validated using CHD, CHDP, and SCH databases. For model performance evaluation, various performance assessment metrics results were generated. The experimental results were compared with the existing state of the arts methods.

Here, the experimental results are briefly presented. The RFE algorithm from the CHD data set of the subset of selected features included SEX, CTP, EIA, PES, and VCA. While from the SCH data set, the selected subsets of features are SEX, CTP, FBG, EIA, and PES. The performance of the Random Forest algorithm on CHDP data was higher as compared with CHD and SCH data sets. Hence, Table 8 presented the results of the

TABLE 12 Proposed model performance comparison with baseline models.

Model	Acc (%)	Ref
ML algorithms	77	(7)
MLP + SVM	80.41	(16)
Random Forest model	88	(23)
GA-RBF	94.20	(24)
Hybrid MLmodel (HRFLM)	88.07	(13)
SGD	87.69	(25)
ANN + Fuzzy Logic (ANN-FL)	87.4	(8)
ANN ensemble-based diagnosis system	89.01	(17)
IHDPS	88.12	(9)
3p-ANN	88.89	(10)
RF	96.72	(26)
ANN-FUZZY-AHP	91.1	(11)
SVM	96.72	(27)
KNN	90.789	(28)
Random Forest	92.3	(29)
RRS-HD	92.32	(12)
HISFP	89	(2)
FCMIM-SVM	92.37	(14)
SE	92.34	(15)
Proposed stacking HD (CHD)	92.67	2024
Proposed stacking HD (CHDP)	99.25	2024
Proposed stacking HD (SCH)	97.20	2024

classifier Random forest trained and evaluated on full and selected feature sets on the CHDP data set. The values for the whole feature set's accuracy, sensitivity, specificity, precision, MCC, and F1-score were 98.34%, 98.45%, 98.32%, 93.67%, 97.33%, and 98.32%. While the values on selected feature set models were 98.89%, 99.00%, 98.77%, 98.67%, 96.00%, and 99.01%, respectively. The model improved accuracy $98.89-98.34 = 0.54\%$ on the selected feature set. The performance of other metrics also greatly improved. The Random Forest accuracy is also higher than the Ada Boost classifier. Similarly, when stacking techniques were incorporated, the Random Forest performance was higher than Ada Boost, and the Random Forest model was selected as the meta-model. According to Table 7, on the CHDP chosen feature dataset, the stacking technique selected the Random Forest meta classifier and produced the higher performance in terms of accuracy, sensitivity, specificity, precision, MCC, and F1-score, each with a score of 99.25%, 95.89%, 99.04%, 97.56%, 98.00%, and 99.30%.

The confusion matrix and ROC curves of the stacking approach with data sets CHD, CHDP, and SCH are shown in

Figures 11, 12. Hence, the ROC curve of the stacking model with the CHDP data set is higher, so it presents that the model accurately detected the HD as compared with CHD and SCH data sets.

Our analysis of the aforementioned results led us to the conclusion that the proposed model, stacking HD, provided better predictive outcomes and was easily implementable for HD detection in IoT-based healthcare systems.

5 Conclusion and future work direction

Machine learning-based Computer-Aided Diagnosis Systems are typically utilized to effectively identify heart disease. However, because current artificial diagnostic approaches are imprecise, medical practitioners are not adopting them into the heart diagnosis process efficiently. In the research study, we created an accurate technique for identifying HD using ML techniques. In the proposed approach, machine learning classifiers including Random-Forest (RF) and Ada-Boost are incorporated for the classification of heart disease and healthy control subjects. For data pre-processing and feature selection, we incorporated standard scalar and Recursive Feature Elimination (RFE) techniques to balance the data for proper training of the algorithm to enhance the model's predictive capability. We defined a parameter grid for grid search for both algorithms. To enhance algorithm accuracy, an ensemble learning technique was incorporated to select the best classification model. A held-out validation mechanism was utilized, and HD datasets were used to validate the proposed model.

The proposed model was evaluated using different evaluation metrics. According to experimental outcomes on the selected feature dataset (CHDP), the stacking technique selected meta classifier (Random Forest) and obtained the higher performance in terms of accuracy, 99.25%, and greater ROC curve. The proposed stacking HD model experimental outcomes presented that the model obtained higher results in terms of accuracy compared with existing models. Due to its excellent results, the proposed stacking HD model is recommended for HD detection in IoT healthcare systems. In the future, we will incorporate deep learning, transfer learning, and federated learning techniques to design a more advanced system for the diagnosis of heart disease in the IOT healthcare system.

Data availability statement

The data sets utilized in this work were collected from a public repository. The Cleveland Heart Disease dataset is available at the link: <https://www.kaggle.com/datasets/aavigan/cleveland-clinic-heart-disease-dataset>. Heart Statlog Cleveland Hungary data set is available at the link: <https://ieee-dataport.org/open-access/heart-disease-dataset-comprehensive>. All methods were performed in accordance with the relevant guidelines and regulations.

Ethics statement

Ethical review and approval was not required for the study on human participants, in accordance with the local legislation and institutional requirements.

Author contributions

WJ: Funding acquisition, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. JL: Funding acquisition, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision. AH: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. SK: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. RA: Writing – review & editing, Formal analysis, Visualization, Software. SA: Writing – review & editing, Methodology, Validation, Visualization, Software. MH: Data curation, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study

was supported by the National Natural Science Foundation of China (Grant No. 61370073), the National High Technology Research and Development Program of China, and the project of the Science and Technology Department of Sichuan Province (Grant No. 2021YFG0322).

Acknowledgments

The authors extend their appreciation to the Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia for funding this research through the project number IFP-IMSIU-2023119. The authors also appreciate the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) for supporting and supervising this project.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Mendis S. *Global Atlas on Cardiovascular Disease Prevention and Control*. Geneva: World Health Organization (WHO) (2011).
- Haq AU. A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mob Inf Syst*. (2018) 2018: 3860146. doi: 10.1155/2018/3860146
- Durairaj M. A comparison of the perceptive approaches for pre-processing the data set for predicting fertility success rate. *Int J Control Theory Appl*. (2016) 9:255–60.
- Kumar PM, Gandhi UD. A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases. *Comput Electr Eng*. (2018) 65:222–35. doi: 10.1016/j.compeleceng.2017.09.001
- Mpanya D. Predicting mortality and hospitalization in heart failure using machine learning: a systematic literature review. *IJC Heart Vasc*. (2021) 34:100773. doi: 10.1016/j.ijcha.2021.100773
- Haq AU. IIMFCBM: intelligent integrated model for feature extraction and classification of brain tumors using MRI clinical imaging data in IoT-healthcare. *IEEE J Biomed Health Inform*. (2022) 26:5004–12. doi: 10.1109/JBHI.2022.3171663
- Detrano R, Janosi A. International application of a new probability algorithm for the diagnosis of coronary artery disease. *Am J Cardiol*. (1989) 64:304–10. doi: 10.1016/0002-9149(89)90524-9
- Kahramanli H, Allahverdi N. Design of a hybrid system for diabetes and heart diseases. *Expert Syst Appl*. (2008) 35:82–9. doi: 10.1016/j.eswa.2007.06.004
- Palaniappan S. Intelligent heart disease prediction system using data mining techniques. In: *2008 IEEE/ACS International Conference on Computer Systems and Applications*. Doha: IEEE (2008), p. 108–15. doi: 10.1109/AICCSA.2008.4493524
- Olaniyi EO. Heart diseases diagnosis using neural networks arbitration. *Int J Intell Syst Appl*. (2015) 7:72. doi: 10.5815/ijisa.2015.12.08
- Samuel OW. An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. *Expert Syst Appl*. (2017) 68:163–72. doi: 10.1016/j.eswa.2016.10.020
- Liu X, Wang. A hybrid classification system for heart disease diagnosis based on the RFRS method. *Comput Math Methods Med*. (2017) 2017:8272091. doi: 10.1155/2017/8272091
- Mohan S, Thirumalai C. Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*. (2019) 7:81542–54. doi: 10.1109/ACCESS.2019.2923707
- Li JP, Haq AU. Heart disease identification method using machine learning classification in e-healthcare. *IEEE Access*. (2020) 8:107562–82. doi: 10.1109/ACCESS.2020.3001149
- Tiwari A, Chugh A, Sharma A. Ensemble framework for cardiovascular disease prediction. *Comput Biol Med*. (2022) 146:105624. doi: 10.1016/j.compbiomed.2022.105624
- Gudadhe M. Decision support system for heart disease based on support vector machine and artificial neural network. In: *2010 International Conference on Computer and Communication Technology (ICCCCT)*. Allahabad: IEEE (2010), p. 741–5. doi: 10.1109/ICCCCT.2010.5640377
- Das R, Turkoglu I. Effective diagnosis of heart disease through neural networks ensembles. *Expert Syst Appl*. (2009) 36:7675–80. doi: 10.1016/j.eswa.2008.09.013

18. Bhatt CM, Patel P, Ghetia T, Mazzeo PL. Effective heart disease prediction using machine learning techniques. *Algorithms*. (2023) 16:88. doi: 10.3390/a16020088
19. Ozcan M, Peker S. A classification and regression tree algorithm for heart disease modeling and prediction. *Healthc Anal*. (2023) 3:100130. doi: 10.1016/j.health.2022.100130
20. Shukur BS, Mijwil MM. Involving machine learning techniques in heart disease diagnosis: a performance analysis. *Int J Electr Comput Eng*. (2023) 13:2177. doi: 10.11591/ijece.v13i2
21. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324
22. Hastie T. Multi-class AdaBoost. *Stat Interface*. (2009) 2:349–60. doi: 10.4310/SII.2009.v2.n3.a8
23. Jansi Rani S, Chandran KS. Smart wearable model for predicting heart disease using machine learning: wearable to predict heart risk. *J Ambient Intell Humaniz Comput*. (2022) 13:4321–32. doi: 10.1007/s12652-022-03823-y
24. Doppala BP. A hybrid machine learning approach to identify coronary diseases using feature selection mechanism on heart disease dataset. *Distrib and arallel Databases*. (2021) 41:1–20. doi: 10.1007/s10619-021-07329-y
25. Pires IM. Machine learning for the evaluation of the presence of heart disease. *Procedia Comput Sci*. (2020) 177:432–7. doi: 10.1016/j.procs.2020.10.058
26. Al Ahdal A. Monitoring cardiovascular problems in heart patients using machine learning. *J Healthc Eng*. (2023) 2023:9738123. doi: 10.1155/2023/9738123
27. Saboor A. A method for improving prediction of human heart disease using machine learning algorithms. *Mob Inf Syst*. (2022) 2022:1410169. doi: 10.1155/2022/1410169
28. Shah D, Patel S. Heart disease prediction using machine learning techniques. *SN Comput Sci*. (2020) 1:1–6. doi: 10.1007/s42979-020-00365-y
29. Kishor A, Jeberson W. Diagnosis of heart disease using internet of things and machine learning algorithms. In: *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security: IC4S 2020*. Singapore: Springer (2021), p. 691–702. doi: 10.1007/978-981-16-0733-2_49
30. Ganesan M, Sivakumar N. IoT-based heart disease prediction and diagnosis model for healthcare using machine learning models. In: *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*. Pondicherry: IEEE (2019), p. 1–5. doi: 10.1109/ICSCAN.2019.8878850
31. Nancy AA, Ravindran D. IoT-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning. *Electronics*. (2022) 11:2292. doi: 10.3390/electronics11152292
32. Ahamed J, Manan Koli A, Ahmad K, Alam Jamal M, Gupta BB. CDPS-IoT: cardiovascular disease prediction system based on IoT using machine learning. *Int J Interact Multimedia Artif Intellig*. (2022) 7:78–86. doi: 10.9781/ijimai.2021.09.002
33. Ahmed F. An Internet of Things (IoT) application for predicting the quantity of future heart attack patients. *Int J Comput Appl*. (2017) 164:36–40. doi: 10.5120/ijca2017913773



OPEN ACCESS

EDITED BY

Amin Ul Haq,
University of Electronic Science and
Technology of China, China

REVIEWED BY

Ebrahim Elsayed,
Mansoura University, Egypt
Gurjot Singh Gaba,
Linköping University, Sweden
Misbah Abbas,
Nencki Institute of Experimental Biology
(PAS), Poland
Lakshmana Ramasamy,
Higher Colleges of Technology,
United Arab Emirates

*CORRESPONDENCE

Praveen Kumar Reddy Maddikunta
✉ praveenkumarreddy@vit.ac.in

RECEIVED 29 March 2024

ACCEPTED 15 May 2024

PUBLISHED 03 June 2024

CITATION

Ahmed R, Maddikunta PKR, Gadekallu TR,
Alshammari NK and Hendaoui FA (2024)
Efficient differential privacy enabled federated
learning model for detecting COVID-19
disease using chest X-ray images.
Front. Med. 11:1409314.
doi: 10.3389/fmed.2024.1409314

COPYRIGHT

© 2024 Ahmed, Maddikunta, Gadekallu,
Alshammari and Hendaoui. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Efficient differential privacy enabled federated learning model for detecting COVID-19 disease using chest X-ray images

Rawia Ahmed¹, Praveen Kumar Reddy Maddikunta^{2*},
Thippa Reddy Gadekallu^{3,4,5}, Naif Khalaf Alshammari⁶ and
Fatma Ali Hendaoui⁷

¹Computer Science Department, Applied College, University of Ha'il, Ha'il, Saudi Arabia, ²School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, Tamil Nadu, India, ³The College of Mathematics and Computer Science, Zhejiang A&F University, Hangzhou, China, ⁴Division of Research and Development, Lovely Professional University, Phagwara, India, ⁵Center of Research Impact and Outcome, Chitkara University, Rajpura, India, ⁶Mechanical Engineering Department, Engineering College, University of Ha'il, Ha'il, Saudi Arabia, ⁷Computer Science Department, Applied College, University of Ha'il, Ha'il, Saudi Arabia

The rapid spread of COVID-19 pandemic across the world has not only disturbed the global economy but also raised the demand for accurate disease detection models. Although many studies have proposed effective solutions for the early detection and prediction of COVID-19 with Machine Learning (ML) and Deep learning (DL) based techniques, but these models remain vulnerable to data privacy and security breaches. To overcome the challenges of existing systems, we introduced Adaptive Differential Privacy-based Federated Learning (DPFL) model for predicting COVID-19 disease from chest X-ray images which introduces an innovative adaptive mechanism that dynamically adjusts privacy levels based on real-time data sensitivity analysis, improving the practical applicability of Federated Learning (FL) in diverse healthcare environments. We compared and analyzed the performance of this distributed learning model with a traditional centralized model. Moreover, we enhance the model by integrating a FL approach with an early stopping mechanism to achieve efficient COVID-19 prediction with minimal communication overhead. To ensure privacy without compromising model utility and accuracy, we evaluated the proposed model under various noise scales. Finally, we discussed strategies for increasing the model's accuracy while maintaining robustness as well as privacy.

KEYWORDS

COVID-19 detection, decentralized training, adaptive differential privacy, federated learning, convolutional neural network, healthcare data privacy

1 Introduction

The global healthcare system faces an unprecedented challenge due to SARS-CoV-2. The COVID-19 pandemic has emerged as a significant global health crisis, impacting millions worldwide and causing widespread economic and societal disruption on a global scale. The rapid spread of the virus has led to the harnessing of cutting-edge technologies for patient data collection, disease prediction, surveillance, and management. COVID-19 disease-related data being generated or collected by the various Internet of Things (IoT) applications are being

managed and processed using efficient big data analytics and computational methods such as ML or DL algorithms (1). Diverse healthcare datasets are collected, encompassing epidemiological data (e.g., confirmed cases, deaths, recoveries), clinical records (e.g., symptoms, comorbidities), demographic information (e.g., gender, age), and socio-economic factors (e.g., population density, mobility patterns). However, this data inherently contains sensitive information related to specific patients, regions, or locations (2). Therefore, robust measures are crucial to safeguard data privacy and confidentiality during various activities such as sharing, exchanging, managing, and processing, which often involve multiple entities and tools. Healthcare data privacy standards guarantee that only authorized individuals or organizations have access to a patient's personal medical information. This protects sensitive information like a patient name, patient address, date of birth, and important medical status being shared without their consent (3). However, traditional centralized systems have major drawbacks, including significant processing time, increased network traffic, and a heightened risk of unauthorized data access.

Over the years, various methods have been developed for addressing the limitations of centralized architectures. While preserving data privacy and confidentiality through authorized access control. However, recent advances in applied AI technologies provide promising results with distributed learning techniques, resulting in increased data processing. FL is a distributed learning approach in which only model parameters are exchanged between the server and clients over several iterations, rather than actual data being transferred to the server. The clients perform training on their data using the model parameters provided by the server. Throughout this process, initial privacy is provided, and communication costs are reduced. Since the amount of data on clients is less compared to the central data pool, local learning is attained with minimal hardware requirements (4). Figure 1 illustrates the processing of medical data from various hospitals using FL architecture. Although FL achieves privacy through the physical isolation of data, it does not guarantee privacy for local data. During the model transmission process, the server can invert the client's local information using model gradients, leading to a potential inference attack. Even though FL fulfills the design principles necessary for achieving privacy, but still, the attacker can still steal the private information of a user through the intermediate results of the FL process (5). However, this can be addressed in two ways. First, we can consider encryption methods to protect the information flow of intermediate results such as Homomorphic Encryption (HE) (6) and Secure Multi-party Computation (MPC) (7). Secondly, we can consider the perturbation of the original private information, through techniques such as Differential Privacy (DP), which can prevent the revelation of intermediate results (8).

By introducing noise to the original dataset or learning parameters, the DP technique guarantees a high level of privacy protection in data analysis, thus making it impossible for attackers to access sensitive data. Although DP was proposed in 2006, its recent AI applications to improve data security, stabilize the learning process, develop unbiased models, and apply composition in specific AI domains have attracted significant interest from researchers and tech titans such as Google, Microsoft, and Apple (9). These organizations are interested in retrieving statistics from client devices, either by developing applications with Central Differential Privacy (CDP) or Local Differential Privacy (LDP)

techniques (10). CDP techniques involve the inclusion of random noise to the actual data after it has been acquired from all clients by a data curator in a central server. However, the LDP mechanism introduces noise before transmitting the data or learning parameter to the central server, guaranteeing privacy from the beginning of data transmission process. Besides applications in ML and DL, DP has also improved the convergence rate by guaranteeing privacy in distributed learning environments (11). An adaptive Differential Privacy Federated Learning Medical IoT (DPFL-MIoT) uses several techniques such as DP, FL, and deep neural networks with adaptive gradient descent to mask model parameters by infusing noise (12).

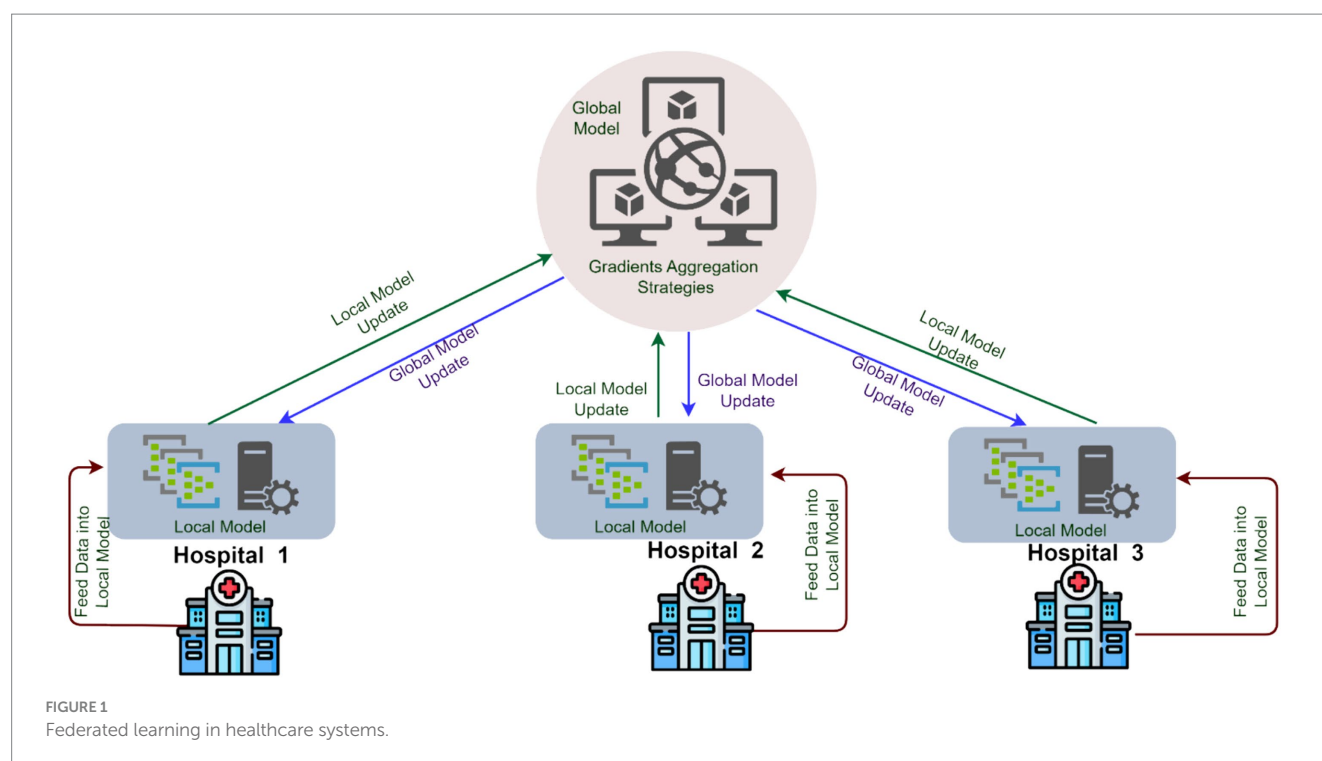
The main contributions of the work are as follows:

- 1 We have developed a distributed learning model to predict COVID-19 disease by considering the three different classes of Chest X-Ray images such as COVID, Normal, and Pneumonia.
- 2 We designed Adaptive Differential Privacy-Enhanced Federated Learning (DPFL) framework with an early-stopping technique to preserve patient data while maintaining utility.
- 3 We have conducted several experiments to analyze and evaluate the Utility and Privacy of the data, and the impact of the early stopping mechanism on the performance of the proposed DPFL model.

The rest of the paper is organized as follows: Section 2 discusses existing works on FL and AFL using DP. Section 3 presents the proposed FL models with a DP mechanism. A detailed discussion of the experimental setup, dataset, and obtained results are provided in Section 4. Finally, the conclusion and future research directions are discussed in Section 5.

2 Literature review

FL revolutionizes ML by decentralizing model training across devices, safeguarding local data privacy. This collaborative model involves a central server managing global parameters and clients with local datasets. Model updates from clients enhance the global model iteratively. FL offers advantages like privacy preservation, reduced communication overhead, and collaborative learning. Challenges include handling heterogeneous data and addressing communication and security concerns. This sets the stage for exploring privacy-preserving mechanisms like Differential Privacy within the FL framework. To reduce the prediction bias and to eradicate the overfitting problems caused by too small dataset, Chen et al. (13) have proposed a DP-based adaptive worker selection algorithm. The proposed framework generated a vulnerability prediction map considering COVID-19 data through various apps using distributed FL models to ensure privacy. Wu et al. (14) suggested an FL model with an adaptive gradient descent and differential privacy mechanism for a multiparty collaborative environment by ensuring efficient model training with minimal communication cost. Even though, the proposed technique enhances the accuracy and stability of the model but still lacks model convergence efficiency due to hyperparameter fluctuations. Ulhaq et al. (15) have developed a Differential privacy-enabled FL framework for COVID-19 disease diagnosis by ensuring data privacy. The authors have designed and



developed the theoretical model, hence the model needs to be implemented for further analysis.

Similarly, Wang et al. (16) have designed a privacy-enhanced disease diagnosis using FL. The proposed model incorporates Variational Autoencoder (VAE), differential privacy noise, and incentive mechanism during the disease diagnosis process in a distributed environment. Simulation results have shown that the accuracy of the global model decreases with an increase in the privacy budget. The privacy requirements of the individuals are not the same, hence the authors Liu et al. (17) have introduced a hybrid differential privacy technique to the existing privacy-friendly FL framework by dividing the user into groups as per their privacy requirements. The adaptive gradient clipping mechanism and improved composition methods of the model will improve the model accuracy by reducing the noise issues. To reduce the impact of noise on the accuracy of the model the authors Yang et al. (18) have proposed Kalman Filter-based Differential Privacy Federated Learning Method (KDP-FL). The Proposed algorithm was tested in a simulated environment; however, the Kalman filter noise reduction method results in better accuracy but increases the computational overhead.

To reduce and nullify the leakage of semantic information of the training data by the Generative Adversarial Networks (GAN), the author's Zhang et al. (19) have developed a "Federated Differentially Private Generative Adversarial Network (FedDPGAN)" model for the detection of COVID-19 pneumonia, which is aimed to improve the data privacy of the patients. DP-GAN of the proposed model protects the semantic information of the training dataset in a distributed learning environment. The model was tested and analyzed by considering both the IID and Non-IID settings of the COVID-19 dataset. The experimental results have shown 3% increase in the overall performance compared to the FL model by ensuring the privacy of data. Similarly, Ho et al. (20) introduced a privacy-focused FL system for COVID-19 detection, aiming to create a decentralized

learning framework among multiple hospitals that does not need the transfer of actual patient data. The proposed framework ensures the privacy of patient data by incorporating differential privacy techniques such as DP stochastic gradient descent (DP-SGD). The experimental results show that incorporating a spatial pyramid pooling layer into a 2D CNN, as well as specific design choices for handling Non-IID data, such as the number of total clients, the degree of client parallelism, and the computations per client, resulted in an increase in overall accuracy.

To achieve privacy with high utility in a distributed learning environment, the authors Li et al. (21) have proposed a secure Asynchronous Federated Learning (AFL) with DP algorithm for collaborative edge-cloud devices. The multi-stage adjustable private algorithm of the proposed model will dynamically adjust the noise and learning rates to improve the efficiency and convergence. The experimental findings show better results compared to the existing machine learning models with improved privacy. Lu et al. (22) has proposed a differentially private AFL approach for data sharing in vehicular networks. The authors have proposed local DP technique to nullify the attacks caused by the centralized curator during the weighted aggregation process. The experimental results have shown faster convergence with a few observations as the number of clients' increases such as increased training period required to learn from the server model with reduced accuracy. Nguyen et al. (23) has proposed a novel asynchronous federated optimization framework with buffered asynchronous aggregation and Differential privacy scheme. The model was aimed to achieve improved privacy and scalability. The simulation results of the model outperformed the traditional methods.

Li et al. (24) have proposed an optimized asynchronous federated model for a depression detection system. The model was designed to enhance both the communication efficiency and the convergence rate while maintaining users' privacy using the DP technique. The experimental results have shown 86.67% accuracy and minimal

communication cost. Even though the FL provides a privacy guarantee for the user's data, to strengthen the privacy safeguards the authors, Nampalle et al. (25) have proposed a novel FL with a DP technique for medical image classification. The proposed method consists of a novel noise calibration mechanism and adaptive privacy budget allocation strategy. Even though the simulation results have shown an improved efficiency in the classification of skin lesions and brain tumor images, the model requires further analysis and testing to improve the overall performance. Malik et al. (26) introduced DMFL_Net, a FL-based model for COVID-19 image classification. The study aims to improve COVID-19 classification, data privacy, and communication efficiency across medical institutions. The model incorporates DenseNet-169 into FL environment to enable collaborative training without sharing its contents to clients, thus guaranteeing privacy. The experiments were conducted on chest X-ray images to compare the performance of DMFL_Net with the conventional transfer learning approaches VGG-19 and VGG-16. The experimental results show that the proposed DMFL_Net model attains an accuracy of 98.45%, outperforming all other models and ensuring data privacy and optimal communication efficiency between participating hospitals. Dayan et al. (27) proposed a FL model named EXAM, that predicts the future oxygen requirements for COVID-19 patients based on chest X-rays, vital signs, and test results. The primary objective of the present study is to design a robust, generalizable model that can classify patients efficiently and effectively among different healthcare systems without the need for personal information sharing, thereby enhancing privacy and data security. The proposed model utilizes a 34-layer CNN (ResNet34) for extracting features from chest X-rays and a Deep & Cross network for integrating EMR features. The experiments were performed on data collected from 20 institutes around the world, and the results indicate that the proposed EXAM model enhanced accuracy and generalizability across trained models, with an AUC increase of 16 and 38% for generalizability.

Table 1 represents the summary of existing differential privacy-based Federated Learning models.

The literature review for Section 2 was carried out in accordance with the PRISMA guidelines shown in Figure 2.

3 Proposed model

In this section we present the preliminaries of Federated average algorithm and differential privacy mechanism. Following that, we present an overview of our proposed model, including the architecture and approaches used to classify Chest X-ray images to identify COVID-19 cases.

3.1 Differential privacy

Differential privacy (DP) enables the analysis of the features of an entire dataset or population without disclosing any personal information. A differentially private algorithm ensures that the inclusion or exclusion of a tuple from the dataset has no vital effect on the output. Dwork et al. defined DP as follows:

Definition 1: (ϵ, δ) —Differential Privacy—“A randomized algorithm $R: J \rightarrow K$ with input domain J and output range K is (ϵ, δ) -differentially private if for all pairs of neighboring datasets $J, J' \in J$,

and every measurable $L \subseteq K$, we have $\Pr(R(J) \in L) \leq e^\epsilon \cdot \Pr(R(J') \in L) + \delta$ where probabilities are with respect to the coin flips of R Equation.”

Where the privacy budget ϵ is used to determine the strengths of privacy protection and $\delta = 0$ result in ϵ -differential private mechanism. This type of DP is accomplished by introducing noise, which is identified through a sensitivity analysis of the dataset. Lower values of ϵ improve privacy but reduce effectiveness because of more noise, which lead to poor accuracy. Higher ϵ values improve data utility while compromising privacy. The chance of a further privacy violation after the ϵ guarantee is controlled by a measure called δ . When adjusting ϵ and δ , we must consider the desired prediction accuracy, acceptable privacy risk, and data sensitivity.

The following two probabilistic methods help to induce noise.

Laplace mechanism (10): The Laplace mechanism is a process of adding noise derived from the continuous Laplace distribution $\left(0, \frac{\Delta_p}{\epsilon}\right)$ where Δ_p is the sensitivity of function p , which measures the largest change in function p 's output generated by adding or removing a single individual's data from the dataset. A higher sensitivity indicates that the function is more responsive to changes in the input dataset. During the process of noise addition to the dataset, L1 sensitivity and the epsilon value (i.e., the privacy budget) are considered for effective results. Hence, the Laplace mechanism can be defined as below:

Definition 2: “Given a function $p: J^1 \rightarrow Y$, where Y is the set of all possible outputs, and $\epsilon > 0$.” The Laplace mechanism is represented in Eq. (1).

$$R(J) = p(J) + \text{Lap}\left(0, \frac{\Delta_p}{\epsilon}\right) \quad (1)$$

Gaussian mechanism (10): The Gaussian Mechanism is a substitution to the Laplace Mechanism, which adds Gaussian Noise and supports tractability of the privacy budget under composition. Unlike Laplace Mechanism, Gaussian Technique uses L2 sensitivity rather than the L1 sensitivity, providing better control over the privacy budget by ensuring reasonable privacy guarantees and smoother noise distribution of L2 sensitivity will also preserve the utility. It can be defined as below.

Definition 3: “Given two neighboring datasets J and J' in the dataset universe J^1 , a query function $p: J^1 \rightarrow G$, where G is the set of all possible outputs, and $\epsilon > 0$.” The ϵ -Gaussian DP (ϵ -GDP) mechanism is given in Eq. (2).

$$R(J) = p(J) + \mathcal{N}\left(0, \frac{\Delta_p^2}{\epsilon^2}\right) \quad (2)$$

Where, $\mathcal{N}\left(0, \frac{\Delta_p^2}{\epsilon^2}\right)$ is considered as the normal distribution.

3.2 Federated averaging process

In a FL system that includes one server and n clients, where each client maintains local database J_i where $i = \{1, 2, 3, \dots, n\}$. The server's objective is to continuously learn from the data stored on n clients through multiple iterations, employing the local weights sent by the n

TABLE 1 Summary of existing DP-based FL models.

References	Methodology	Advantages/salient feature	Disadvantages/future enhancement
Chen et al. (13)	“DP Based adaptive worker selection algorithm for FL with LSTM training model.”	Resolves the issues of inadequate amount of dataset, ensure users data privacy using DP mechanism	Requires further threat analysis.
Wu et al. (14)	Adaptive gradient descent mechanism with DP for collaborative learning	The model shows strong robustness and is less volatile.	The model suffers from convergence issues for a large set of data.
Ulhaq and Burmeister (15)	FL-based DP model for disease diagnosis.	Seven design principles are defined for effective implementation.	Only a theoretical model, hence it requires actual implementation for proper analysis
Wang et al. (16)	FL model with variational autoencoder (VAE) and DP preserve the patient's data privacy	The model guarantees high accuracy and low adversarial inference attacks	Lack of strategies to improve the accuracy of a global model.
Liu et al. (17)	Hybrid Differential Privacy Model for FL.	The model removes the adverse effect of noise addition by using the adaptive clip method	Lack of strategies to stabilize correctness, privacy, and communication in FL
Zhang et al. (19)	GAN-based DP mechanism for FL (FedDPGAN). GAN Based DP mechanism for FL (FedDPGAN).	High-quality training samples generation.	High-quality training samples generation.
Ho et al. (20)	FL-based DPSGD for disease analysis, CNN model incorporating a spatial pyramid pooling strategy.	Improved robustness of the Model and improved accuracy of Non-IID data.	The model requires further analysis by considering a large dataset.
Nampalle et al. (25)	Adaptive privacy budget allocation mechanism for FL.	Improved privacy of medical data.	The proposed model failed to harmonize privacy and model performance
Malik et al. (26)	DMFL_Net for the classification of COVID-19	High classification accuracy and robustness in privacy preservation.	The FL model's complexity limits its ability to scale to larger networks of organizations.
Dayan et al. (27)	FL for predicting clinical outcomes COVID-19 patients	The use of FL improved accuracy and privacy, making it appropriate for sensitive medical applications.	Due to the complexity of managing and synchronizing updates across the network, it does not scale smoothly as the number of participating sites increases.

clients to minimize loss. The optimization problem can be represented as shown in Eq. (3).

$$Wt^* = \arg \min_{Wt} \sum_{i=1}^n p_i F_i(Wt, \mathcal{J}_i) \quad (3)$$

Here, Wt^* denotes the server model parameter generated after aggregating the local models from n clients, Wt_i is denoted as the model parameter from the i th client, and F_i is considered as the loss function of the i th client. Overfitting to specific client datasets in a heterogeneous data environment is a challenge in FL. Regularization and model averaging methods are used to address this issue. Applying regularization to the loss functions F_i helps in minimize overfitting, and Federated Averaging engages averaging model updates from clients to reduce overfitting. p_i is proportional to the amount of data $\mathcal{J}_i$ contained by client i , affecting the client total model. The value of p_i impacts the convergence rate of the model. Managing these weights is essential for guaranteeing that the model performs well among all client data transfers. The training mechanism of FL systems consists of several steps: Initially, the FL model sets the server's weights. After that, it executes the following steps over multiple rounds:

Step 1: Forwarding the server weights: Server weights are forwarded to N clients in a network. Later, each client keeps a buffer to store the received weights in multiple iterations for future reference.

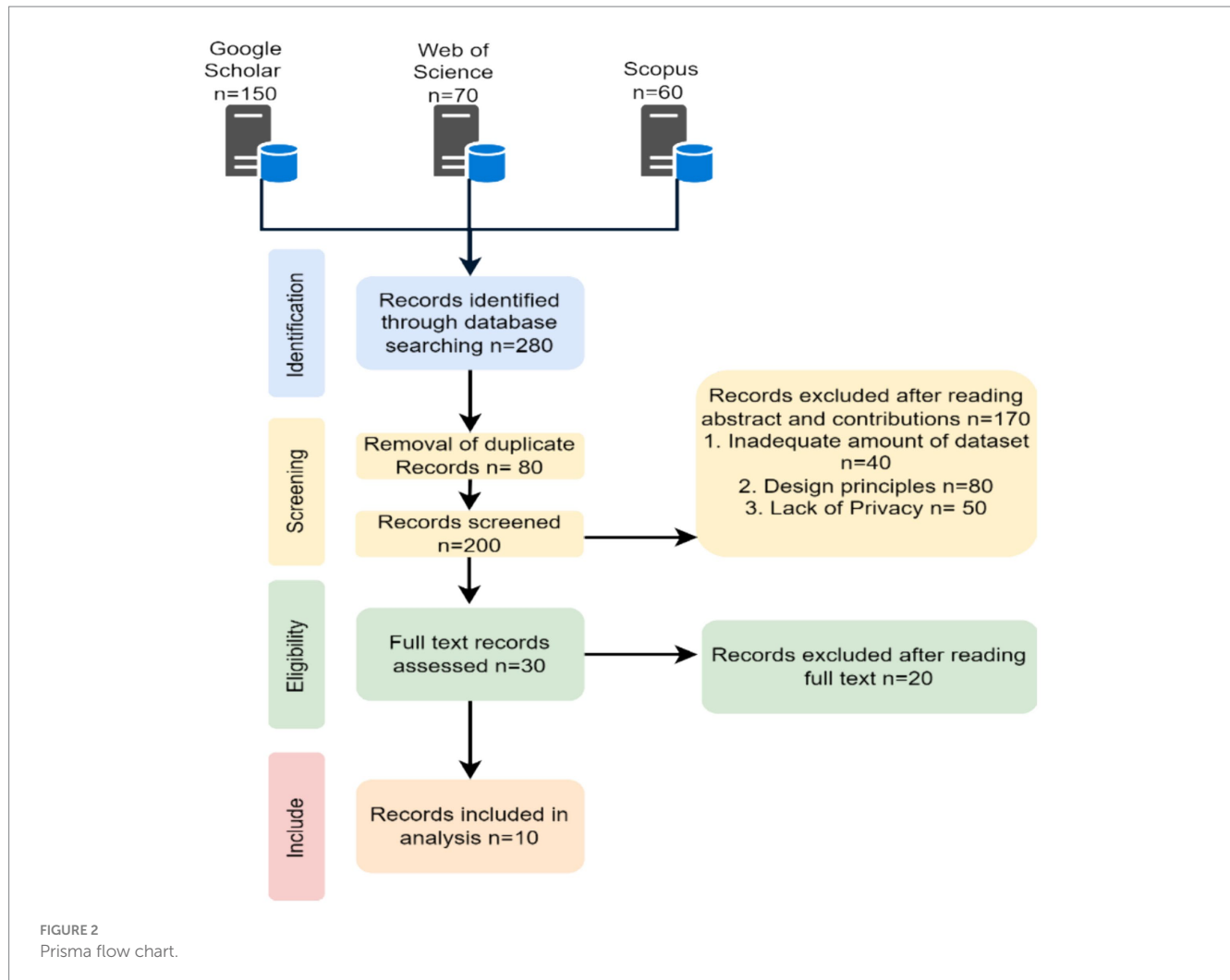
Step 2: Client Model Training: Using the latest model sent by the server, the clients will train their data on local machines. Soon after the training process, the updated models are returned to the server for further operations.

Step 3: Client Model Aggregation: The updated client model weights from n clients are transferred to the server. Later, the server will generate new weight by aggregating all client weight updates through mean computation, which is represented in Eq. (4).

$$Wt' = \frac{\sum_{i=0}^n Wt_i}{\sum_{i=0}^n i} \quad (4)$$

3.3 DP enabled federated averaging algorithm

In this section, we will discuss the architecture and steps involved in the proposed DPFL model and the pseudocode of the DPFL.



3.3.1 Model architecture

The proposed DP-based FL model is aimed at providing user-level privacy by modifying the basic Federated Average algorithms in two different ways:

1. Clip the Model Updates: Model clipping is performed using adaptive methods instead of predefined clipping norms. The adaptive approach updates the clipping threshold based on a specific quantile, ensuring that values are accurately estimated within that range. Also, enables the model to maintain stability and convergence while effectively controlling the magnitude of updates, aimed to improve training performance and model accuracy.

Let $A \in S$ be a random variable and $\beta \in [0, 1]$ be a quantile to be satisfied. Then, for any T is given in Eqs. (5, 6) results in Eq. (7).

$$\ell_{\beta}(T; A) = \begin{cases} (1-\beta)(T-A) & \text{if } A \leq T \\ \beta(A-T) & \text{otherwise} \end{cases} \quad (5)$$

So

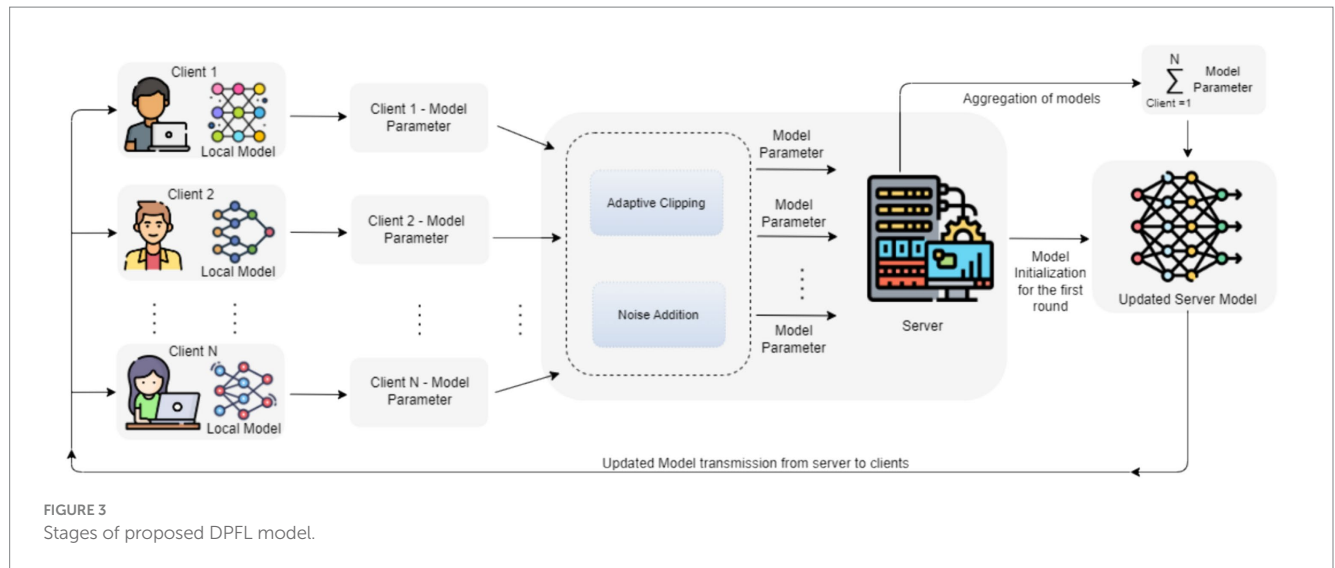
$$\ell'_{\beta}(T; A) = \begin{cases} (1-\beta) & \text{if } A \leq T \\ -\beta & \text{otherwise} \end{cases} \quad (6)$$

$$\mathbb{E}[\ell'_{\beta}(T; A)] = (1-\beta)\Pr[A \leq T] - \beta\Pr[A > T] = \Pr[A \leq T] - \beta \quad (7)$$

2. Addition of noise: In order to improve privacy without degrading the utility of data, the proposed model will be monitored using the standard deviation of the Gaussian noise and number of clients. Initially, we determine the noise tolerance of the model based on a varied amount of noise values by considering a small number of clients per round. Then we train the final model with increased noise on the sum and more clients per round. Reducing the number of clients at first eases the computational load and allows for effective noise level exploration. This methodology facilitates the assessment of the impact of varying noise levels on the usefulness of the information while offering valuable perspectives on the balance between privacy and usefulness. Figure 3 depicts the stages of the proposed DPFL model.

3.3.2 DPFL algorithm

Considering n as the number of users in a round and $\beta \in [0, 1]$ as the target quantile for the norm distribution where clipping is to be applied, for every iteration $m \in [M]$, let V^m represent the clipping



threshold, and η_V the learning rate. Let Y^m be the set of users sampled in round m . Each user $k \in Y^m$ will send the binary indicator a_k^m along with the usual model update Δ_k^m , where $a_k^m = \mathbb{I}(\Delta_{k,2}^m \leq V^m)$. Defining $\bar{a}^m = \frac{1}{n} \sum_{k \in Y^m} a_k^m$, we apply the update $V \leftarrow V \cdot \exp(-\eta_V (\bar{a} - \gamma))$. However, to prevent the leakage of private information through model updates, we add Gaussian noise to the sum $\tilde{a}^m = \frac{1}{n} \left(\sum_{k \in Y^m} a_k^m + \mathcal{N}(0, \sigma_a^2) \right)$.

The target quantile (β) for the normal distribution affects the clipping threshold (V^m) by selecting the value at which the distribution's tails are trimmed. Higher β values result in higher clipping thresholds, allowing for further removal of the distribution. The learning rate η_V in the update rule for V controls how quickly the clipping threshold adjusts to observed gradients. Higher η_V results in quicker V modifications, potentially speeding up convergence by allowing the model to react to changes in data distribution. Excessive η_V values disrupt training, leading to divergence. A lower η_V promotes stability but delay convergence rates. The regularization parameter γ maintains the clipping threshold within the intended bounds by modifying it in response to the discrepancy between the target value γ and the average clipping rate $\bar{a}$. Thus, the federated learning process's privacy-utility trade-off is adjusted by varying γ . Algorithm 1 depicts DPFL Algorithm.

ALGORITHM 1 : DPFL Algorithm

Function Train ($n, \beta, \eta_V, \eta_Z, \eta_V, x, \sigma_a, \beta$)	
Initialize model θ^0 , clipping bound V^0	
$x_\Delta \leftarrow (x^{-2} - (2\sigma_b)^{-2})^{-1/2}$	
For (each round $m = 0, 1, 2, \dots$) do	
$Y^m \leftarrow (\text{sample users uniformly})$	
For each user $k \in Y^m$ in parallel do	

	$(\Delta_k^m, a_k^m) \leftarrow \text{FedAvg}(k, \theta^m, \eta_V, V^m)$
End For	
	$\sigma_\Delta \leftarrow x_\Delta V^m$
	$\tilde{\Delta}^m = \frac{1}{n} \left(\sum_{k \in Y^m} \Delta_k^m + \mathcal{N}(0, I \sigma_\Delta^2) \right)$
	$\bar{\Delta}^m = \beta \bar{\Delta}^{m-1} + \tilde{\Delta}^m$
	$\theta^{m+1} \leftarrow \theta^m + \eta_Z \bar{\Delta}^m$
	$\tilde{a}^m = \frac{1}{n} \left(\sum_{k \in Y^m} a_k^m + \mathcal{N}(0, \sigma_a^2) \right)$
	$V^{m+1} \leftarrow V^m \cdot \exp(-\eta_V (\tilde{a}^m - \beta))$
End For	
function FedAvg(i, θ^0, η, V)	
$\theta \leftarrow \theta^0$	
$\mathcal{B} \leftarrow (\text{user } k\text{'s local data is divided into batches})$	
For batch $b \in \mathcal{B}$ do	
	$\theta \leftarrow \theta - \eta \nabla \ell(\theta; b)$
	$\Delta \leftarrow \theta - \theta^0$
	$a \leftarrow \mathbb{I}(\ \Delta\ \leq V)$
	$\Delta' \leftarrow \Delta \min\left(1, \frac{V}{\ \Delta\ }\right)$
End For	
return (Δ', a)	

3.4 Early stopping mechanism

The early stopping technique is a widely utilized method for regularization in DNN. It is an effective and simple technique that typically outperforms most of the general regularization approaches. During training, the model continually stores and updates the best parameters attained so far. If there's no further improvement in validation error after a set number of iterations, the training halts, retaining the last best parameters. When dealing with models that are prone to overfitting, it is common to recognize a gradual decrease in training error followed by an increase in validation error. Early stopping represents a balance between training duration and generalization error, minimizing communication overhead while still achieving optimal parameters. By reducing the need for communication and subsequently diminishing noise, early stopping enhances the utility of the data. The early stopping algorithm can be represented in Algorithm 2 as follows:

ALGORITHM 2 : General Early Stopping Mechanism

Input: s $\rightarrow$ represents the number steps during the evaluation period.	
e $\rightarrow$ represents the number of epochs, meaning it terminates after observing the worse performance.	
θ_0 $\rightarrow$ represents the initial parameter.	
$\theta \leftarrow \theta_0$	
$p \leftarrow 0$	
$q \leftarrow 0, r \leftarrow \infty, \theta^* \leftarrow \theta, p^* \leftarrow p$	
While $(q < e)$ do	
	Execute the training algorithm for s steps and update θ
	$p \leftarrow p + n$
	$r' \leftarrow \text{validation_set_error}(\theta)$
	If $r' < r$ then
	$q \leftarrow 0, \quad \theta^* \leftarrow \theta, \quad p^* \leftarrow p, \quad r \leftarrow r'$
	Else
	$q \leftarrow q + 1$
	End If
End while	
Output: The optimal parameter θ^* , the optimal number of training steps p^*	

4 Experimental results

This section discusses the experimental activities used to analyze and evaluate the effectiveness of the proposed algorithm. We discuss the dataset, experimental setup, model and training data, and performance analysis using various metrics.

4.1 Dataset description

The proposed model is evaluated considering the Covid19, Pneumonia, Normal Chest X-Ray Image dataset from Mendeley Data

TABLE 2 Distribution of the COVID-19 dataset into training and testing sets.

Data-split details	Normal	Covid-19	Pneumonia
Train data samples	1,442	1,300	1,440
Test data samples	360	326	360

(28). This dataset includes 5,228 chest X-ray images categorized into three categories: 1,626 COVID-19, 1,802 normal (asymptomatic), and 1,800 pneumonia (non-COVID-19). All images are resized to 256 * 256 pixels to reduce computational load, which is important in a FL environment where computations are distributed across devices of different capabilities. During the process we classify the image dataset into train and test sample datasets having 4,182 training samples and 1,046 testing samples, respectively. Table 2 describes the data distribution among each of the categories, and Figure 4 depicts sample images from each category.

4.2 Implementation and model

The proposed model is developed using the Python programming language and evaluated within a Tensorflow framework in a Colab environment. TensorFlow Federated and TensorFlow Privacy packages allow developers to simulate and test the functioning of distributed learning with privacy. TensorFlow Federated provides a wide range of FL-specific features. This allows for the modeling of FL processes on decentralized data, which is crucial for our research as data privacy and local computation are essential. The TensorFlow Privacy framework includes pre-built mechanisms, such as optimizers, to make it easier to integrate differential privacy into machine learning processes. The primary objective is to categorize the disease into three groups: normal, COVID-19, and pneumonia, through the use of CNN model. Our CNN model, depicted in Figure 5, contains two 3×3 convolutional layers with 32 and 64 channels, followed by a 2×2 max pooling layer. The two convolutional layers were used to achieve a balance between model complexity and computational efficiency, which is important in a FL environment where edge devices have limited computational resources. It includes a fully connected layer with 128 units and utilizes ReLU activation, a softmax output layer for classification. To prevent overfitting during the training process, two dropout layers with probabilities of 0.25 and 0.5 are positioned just before and after the fully connected layer.

4.3 Distributed and central architecture

The CNN model is trained in both distributed and traditional central learning environments considering the parameters as number_of_clients=100, client_ratio=0.3, local_epochs=2, and batch_size=16. With the increase in number of rounds, the accuracy in identifying COVID-19 diseases enhances more in FL-based environments. Therefore, the FL model shows superior learning capabilities compared to conventional learning systems. The FL-based model performs better after 50 rounds of execution. Therefore, the overall accuracy of the FL-based approach achieves 94.3%, while central learning is 93.5%. Figure 6 depicts an analysis of communication rounds between FL and central learning models, indicating that training on diverse datasets from various clients results

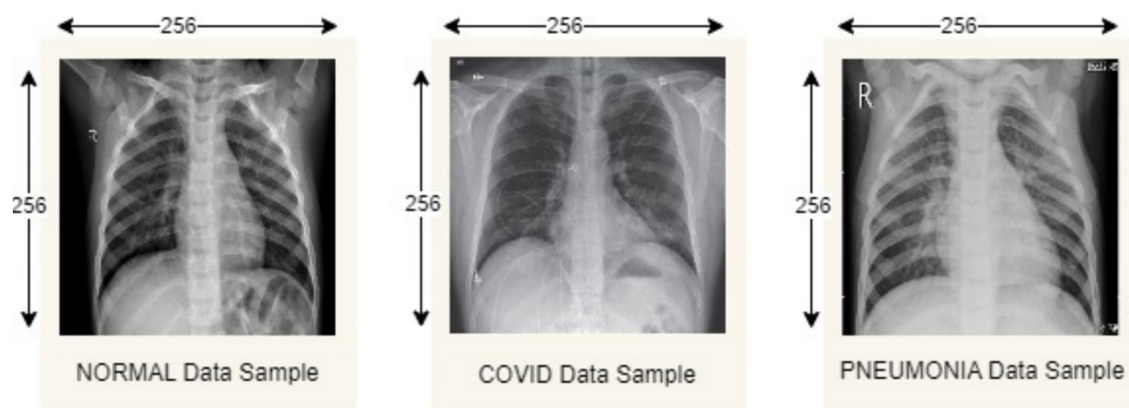


FIGURE 4
Normal, COVID19, pneumonia chest X-ray image samples.

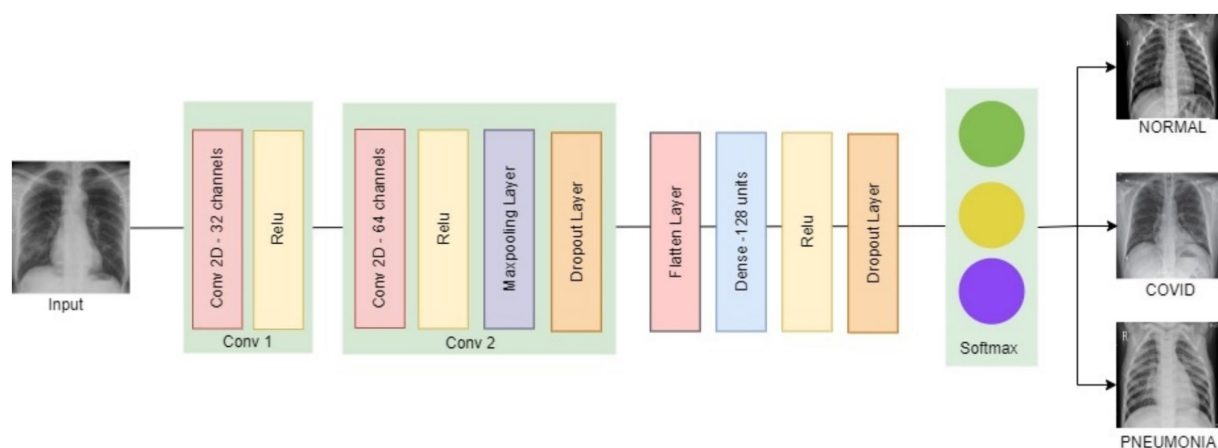


FIGURE 5
CNN model architecture.

in better model generalization. In FL, the client trains a model using local data and only shares model updates. This minimizes the risk of overfitting for COVID-19 patient data. Each round of FL training provides new updates from multiple client datasets, improving the model's ability to predict and achieve higher accuracy. This finding highlights distributed learning's advantage over traditional central learning methodologies in terms of improving model performance.

The proposed distributed learning techniques are further evaluated by comparing various existing CNN models such as Resnet18, Resnet50, and VGG18, with our model. The analysis uses `number_of_clients=100`, `client_ratio=0.3`, `local_epochs=2`, and `batch_size=16`. Our CNN has an optimal number of layers, and activation functions that handle the data's features more efficiently.

The model is designed to generalize better when trained on decentralized datasets and is highly parameter-efficient, resulting in higher accuracy with less parameters. This efficiency is important in FL, where models are updated throughout networks using minimal computational resources. Figure 7 depicts the accuracy analysis of the models where the CNN model outperforms the aforementioned models in terms of accuracy for different communication round. The

primary goal of FL is to manage communication rounds with the computational and communication overheads. Frequent updates result in faster convergence and higher accuracy. We noticed that as the number of rounds increased, the model's accuracy enhanced, implying that more frequent updates benefit model performance.

The proposed distributed FL model undergoes additional analysis by varying the batch size, which shows that the FL model's accuracy increases exponentially as the batch size increases across various rounds, as shown in Figure 8. Increasing the batch size leads to a larger volume of data processed during every round of training. Larger batch sizes help to smooth out noisy gradients and stabilize the training process, resulting in better convergence and accuracy. Therefore, this aids in enhancing the accuracy of the model's learning process.

4.4 FL with differential privacy mechanism

FL guarantees privacy by eliminating the need to share data between participants or servers. To improve the privacy mechanisms of FL-based learning, we proposed the Differential Privacy Federated

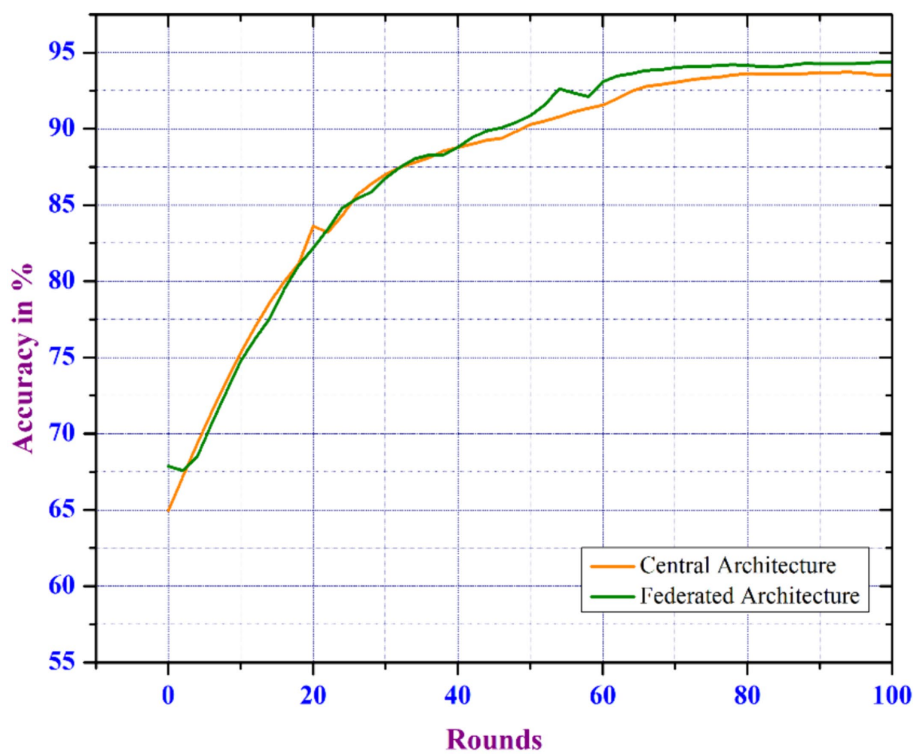


FIGURE 6
Comparison of model accuracy over communication rounds for central and federated learning architectures.

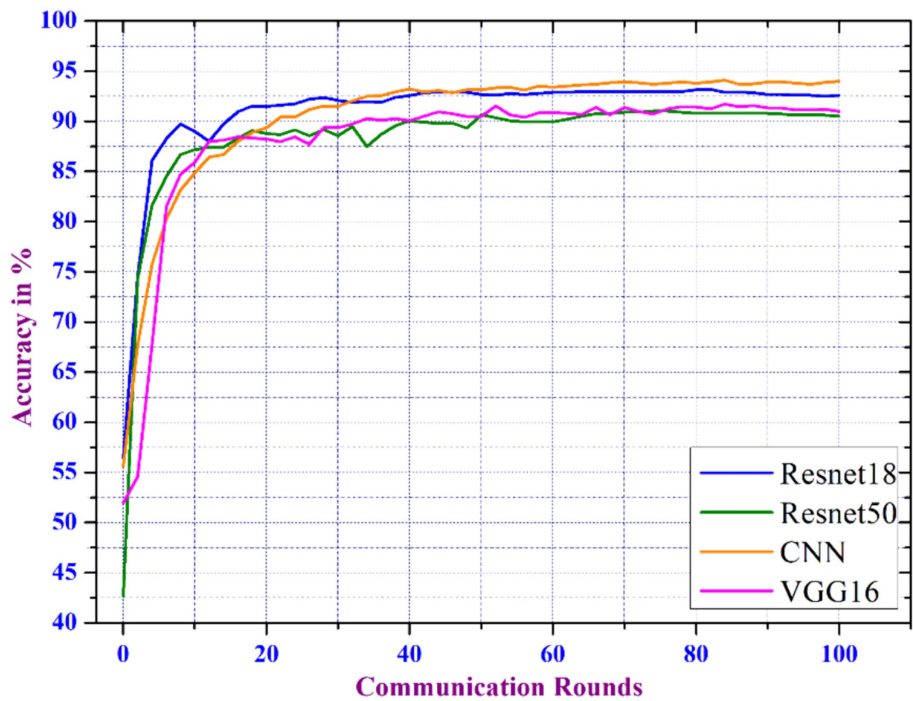


FIGURE 7
Comparative accuracy performance of CNN model against standard CNN architectures.

Learning model. The experiment is carried out in a distributed learning environment with a 0.2 noise_multiplier, 50 clients_per round, a learning_rate of 0.01, two epochs, and a client_ratio of 0.01. However, the introduction of noise reduces the accuracy of the DP-based FL when compared to the traditional FL. Figure 9 shows a 3% drop in accuracy for the DPFL-based model compared to FL. The

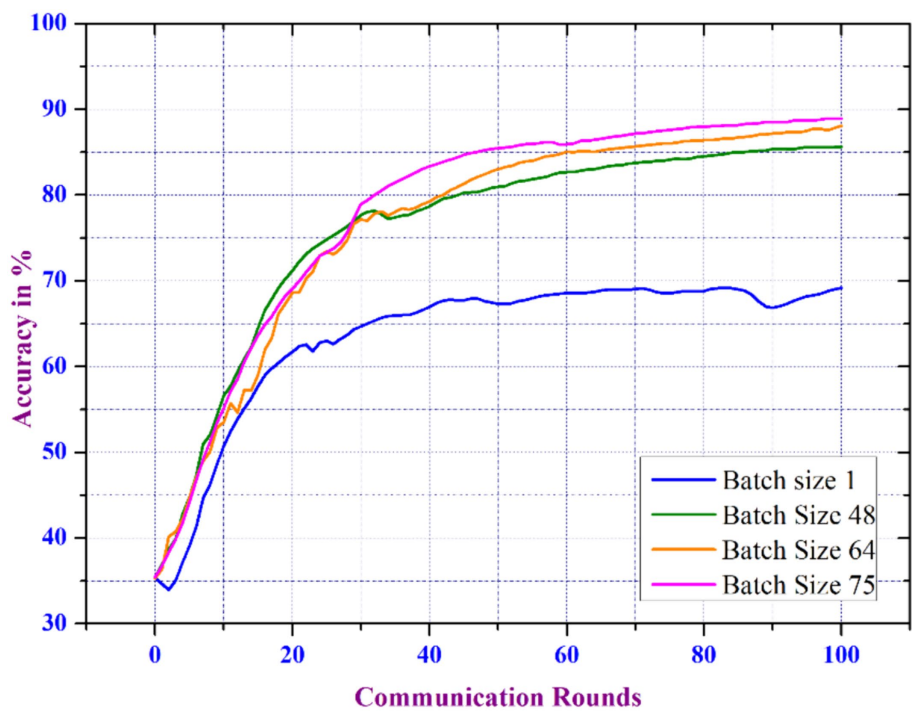


FIGURE 8
Accuracy analysis of FL model with respect to varied batch size.

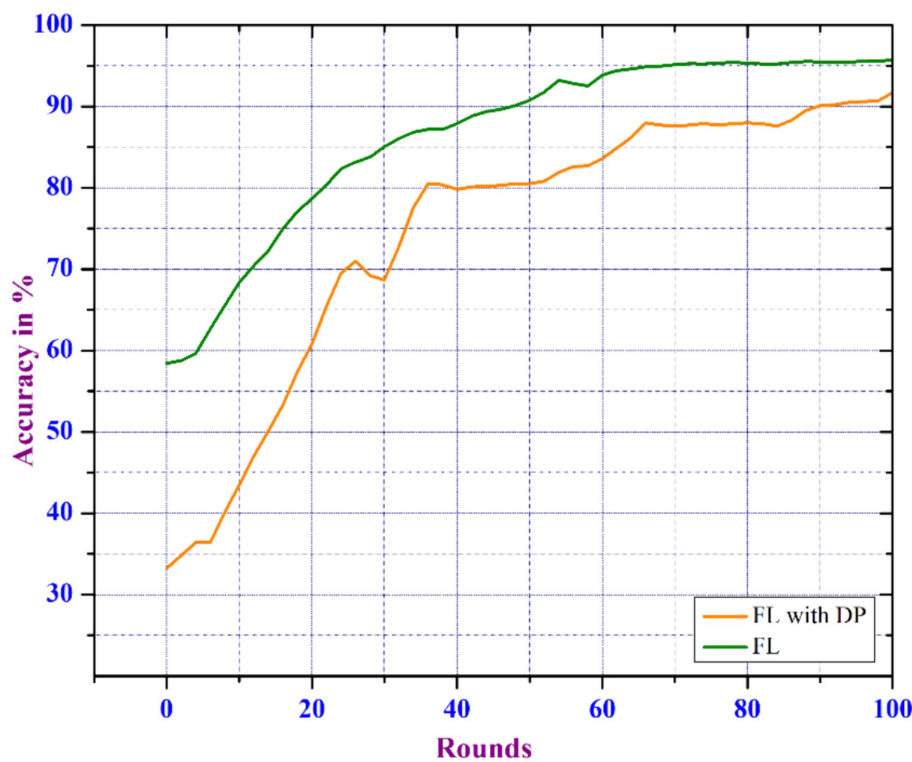


FIGURE 9
Comparison of FL vs. DP enabled FL.

noise disrupts the learning process, lowering the model's capability to accurately capture the underlying patterns in the data. As a result, the introduced noise necessitates a compromise between privacy and model accuracy.

4.5 Model noise sensitivity analysis

Model Noise Sensitivity Analysis in FL is important for deploying FL models in environments where data noise is unavoidable, as it helps to understand how noise in the data affects the performance and reliability of learning models trained on various decentralized devices or servers. In the healthcare domain, the main focus is the accuracy of diagnosis models, as inaccurate predictions can have an immediate effect on the health of patients (29). However, because medical records are so sensitive, patient data privacy is a major concern (30, 31). To meet these requirements, healthcare professionals can select a lower noise multiplier if the model's predictive accuracy is vital for critical diagnostic tasks. Yet, for less sensitive tasks, a higher noise multiplier may be sufficient to ensure more privacy. Our findings suggest a strategic approach in which noise levels are adjusted depending on the sensitivity of the data and the importance of the task. This enables health care professionals to keep patient trust by protecting their data while guaranteeing that the diagnostic models are as accurate as needed. Data scientists working in a variety of sectors particularly healthcare, are frequently challenged with creating models that balance usability and privacy standards. They could apply our findings to create adaptive privacy mechanisms that dynamically adjust the noise multiplier according to real-time assessments of data sensitivity and model performance. Understanding and minimizing the impact of noise can improve the reliability, accuracy, and effectiveness of FL models. To improve utility and maintaining privacy, our proposed model includes an adaptive clipping mechanism based on an increased noise addition mechanism. The adaptive clipping mechanism automatically adjusts the sensitivity between aggregated data as well

model updates, resulting in an optimal balance of data privacy and model utility. This mechanism helps in controlling the impact of noise introduced to ensure privacy, improving the model's learning efficiency, and protecting each data point. Initially, we train the model by considering 50 clients per round by considering noise multipliers in the range [0, 0.25, 0.5, 0.75, and 1.0].

Figures 10, 11 show that the model can tolerate noise multipliers up to 0.5, implying that noise multipliers of 0, 0.25, and 0.5 do not decrease the utility of the data. However, a noise multiplier of 0.75 reduces accuracy, while 1.0 causes the model to completely diverge. The adaptive clipping mechanism allows the model to withstand noise up to a certain level (0.5 in this case) while maintaining utility. This demonstrates the effectiveness of the proposed method, which balances privacy and accuracy. Additional simulations are carried out to determine the implications of changing the client count in each round while keeping a constant noise multiplier of 0.25 and client ratio of 0.01 throughout the process. As the client count increased from 10 to 40, the model's accuracy improved and the loss percentage decreased. However, based on the results of our previous experiments and with the goal of reducing data privacy risks while preserving data utility, we ran another simulation with a privacy budget of $1e-05$ and a total of 120 clients per round. In spite of the increased noise multiplier, the outcomes show enhanced precision in comparison to earlier tests, suggesting that the privacy-preserving mechanisms successfully discover a balance between privacy and utility. Figure 12 depicts the improved accuracy of the proposed model. Therefore, increasing the number of clients per round results in a more diverse and representative dataset, resulting in better generalization and model efficiency.

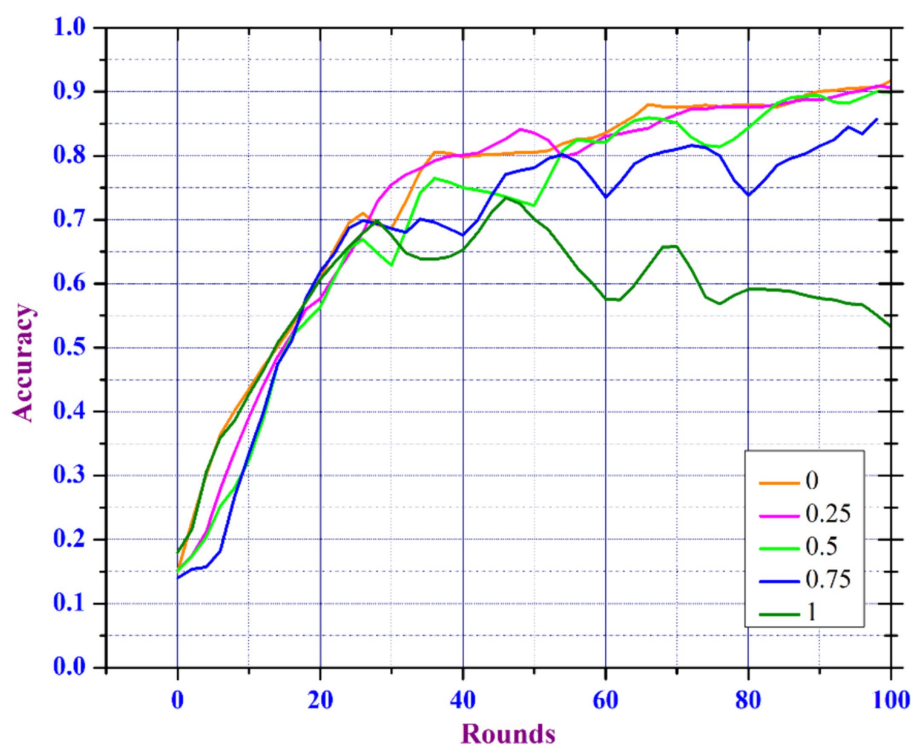


FIGURE 10
Accuracy analysis of DP enabled FL based on varied noise multiplier.

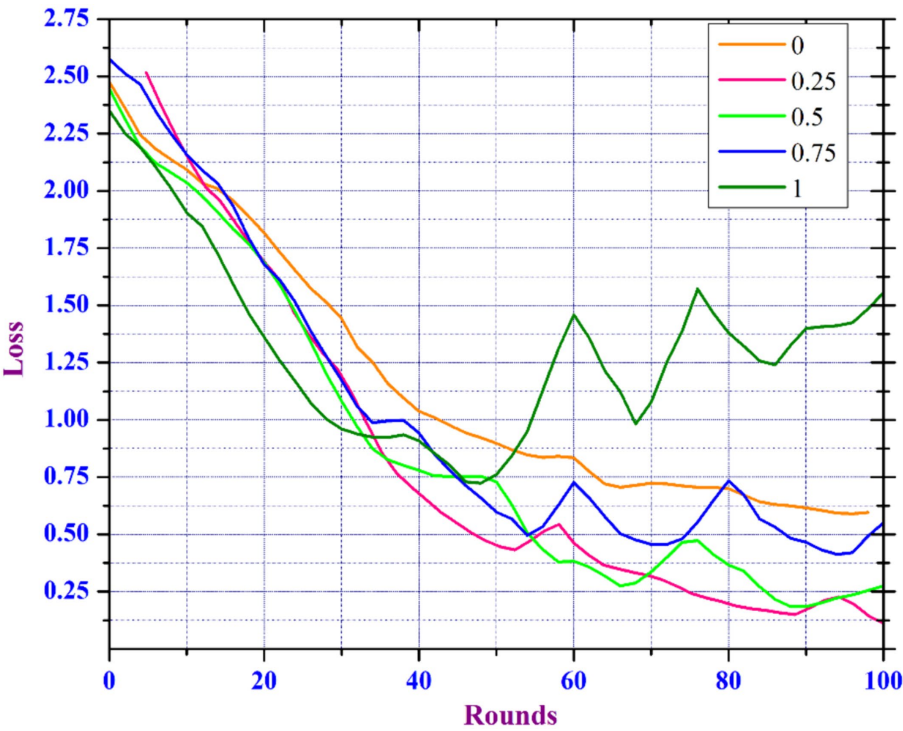


FIGURE 11
Loss analysis of DP enabled FL based on varied noise multiplier.

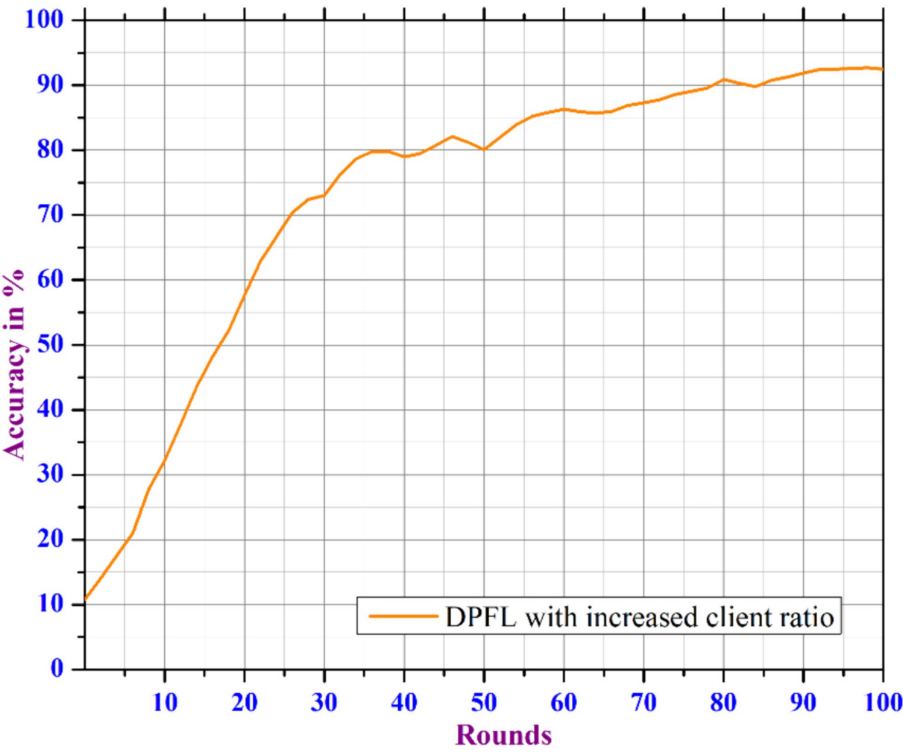


FIGURE 12
Accuracy analysis of DP enabled FL based on increased client ratio.

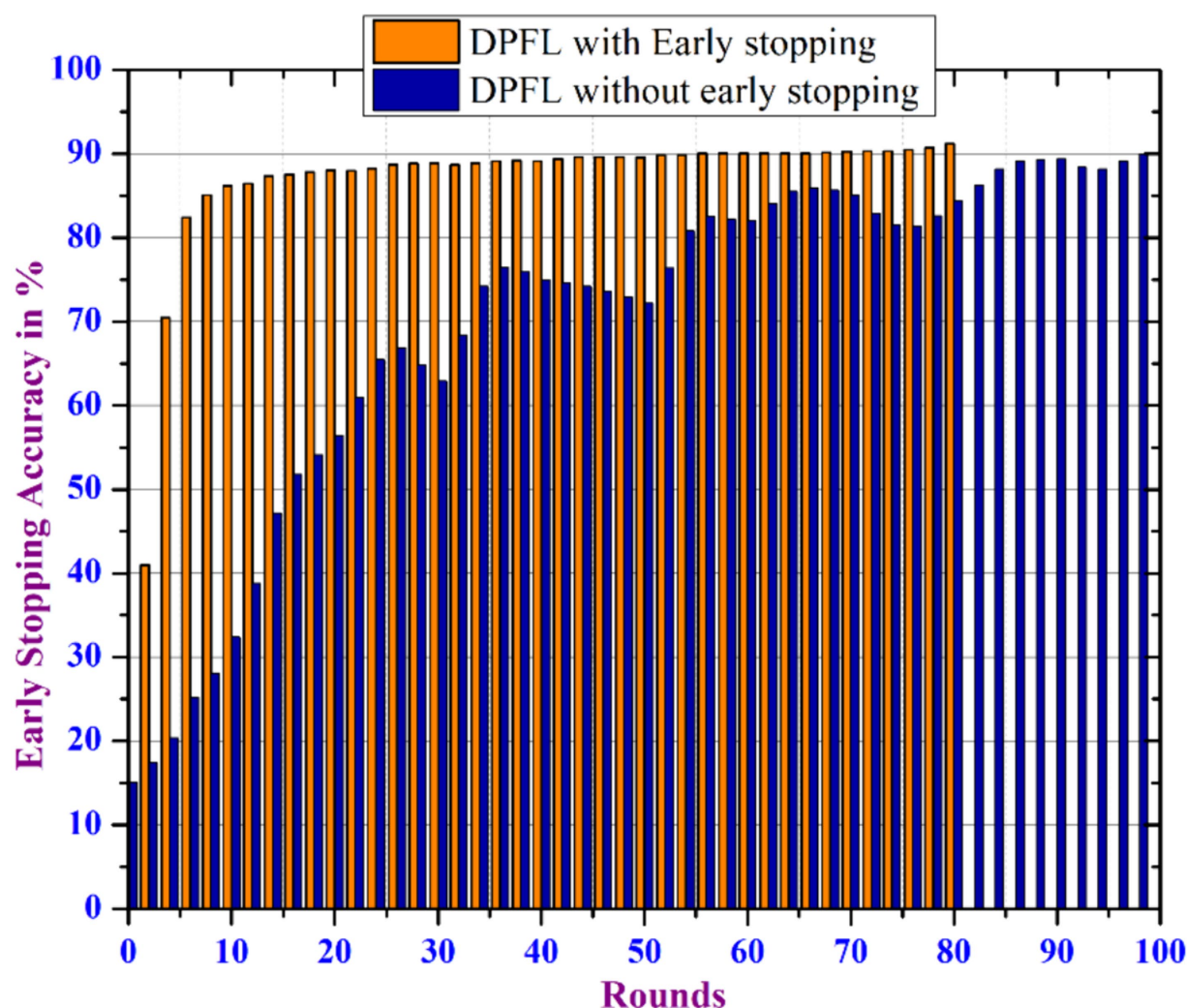


FIGURE 13
Accuracy analysis of early stopping mechanism.

4.6 Model performance for early stopping mechanism

Another experiment was carried out with a configuration of 50 clients_per round, a learning_rate of 0.01, and 100 epochs to investigate the impact of incorporating an early stopping mechanism into the proposed DPFL model, as shown in Figure 13. During the experiment, the proposed DPFL model's accuracy improved as the number of training epochs increased by dynamically adjusting the noise range within a specific privacy level. By evaluating the model's performance on a validation dataset during training, the early stopping mechanism terminate the training process when the model begins to overfit, thus improves the model's generalizability. As a result, the integration of the early stopping mechanism with DPFL model achieved an accuracy of 91.2% after 80 epochs, hence it ensures the consistent privacy level throughout the training process, without sacrificing accuracy and also minimizes overall communication costs.

Early termination of training may have a disproportionate impact on specific clients, resulting in biased model updates and imbalances.

This issue can be addressed by using the early stopping criterion based on client attributes or performance measures, ensuring that all clients contribute significantly to the training process and are treated equally.

5 Conclusion

In this work, we propose an enhanced Privacy-Preserving FL system with Differential Privacy techniques to predict COVID-19 using Chest X-Ray images. Initially, we trained Chest X-Ray image data using a CNN model, evaluating Federated and non-Federated training methods. The results show that FL-based training enhances performance by 0.8% over non-FL or traditional centralized learning. Secondly, we introduce an enhanced FL-based system that includes additional differential privacy and an adaptive noise inclusion mechanism. This system's adaptive clipping effectively identifies the model's noise tolerance level while preserving data utility across different noise scales. However, the proposed DPFL model's initial results show a 3% reduction in accuracy when

predicting COVID-19 due to the masking process. The integration of an efficient privacy-utility trade-off and an early stopping mechanism to DPFL has resulted in a 1% increase in accuracy and a decrease in communication rounds. As a result, the proposed early stopping-based DPFL model outperforms existing DP-based FL models in terms of COVID-19 predictions. The model can be further enhanced by considering the popular pre-trained models for a large dataset and also considering other aspects such as improving the scalability and robustness of the FL. Additionally the incorporation of various techniques for model personalization, model generalization, and fair client contribution evaluation will further strengthen the model.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

RA: Conceptualization, Data curation, Investigation, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing. PM: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. TG: Conceptualization, Data curation, Investigation, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing. NA: Conceptualization, Data curation, Funding acquisition, Methodology, Resources, Software, Visualization, Writing – original draft, Writing – review & editing. FH: Conceptualization, Formal analysis, Project administration,

Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research has been funded by Deputy for Research and Innovation, Ministry of Education through Initiative of Institutional Funding at University of Ha'il—Saudi Arabia through project number IFP-22 133.

Acknowledgments

The authors would like to thank the Scientific Research Deanship at University of Ha'il—Saudi Arabia through project number IFP-22 133.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Zikria YB, Ali R, Afzal MK, Kim SW. Next-generation internet of things (iot): opportunities challenges, and solutions. *Sensors*. (2021) 21:1174. doi: 10.3390/s21041174
- Alshehri F, Muhammad G. A comprehensive survey of the internet of things (IoT) and AI-based smart healthcare. *IEEE Access*. (2020) 9:3660–78. doi: 10.1109/ACCESS.2020.3047960
- Sinhasane S. Data privacy in healthcare: A necessity in protecting health information data. (2022). Available at: <https://mobisoftinfotech.com/resources/blog/data-privacy-in-healthcare/> (Accessed June 2, 2022).
- Zhang C, Xie Y, Bai H, Yu B, Li W, Gao Y. A survey on federated learning. *Knowl-Based Syst*. (2021) 216:106775. doi: 10.1016/j.knsys.2021.106775
- Li T, Sahu A.K., Talwalkar A., Smith V. Federated learning: challenges, methods, and future directions. In: *IEEE Signal Processing Magazine* (2020), 37, 50–60.
- Zhang L, Xu J, Vijayakumar P, Sharma P.K., Ghosh U. Homomorphic encryption-based privacy-preserving federated learning in iot-enabled healthcare system. In: *IEEE Transactions on Network Science and Engineering* (2022).
- Byrd D, Polychroniadou A. Differentially private secure multi-party computation for federated learning in financial applications. In: *Proceedings of the First ACM International Conference on AI in Finance*, (2020), pp. 1–9.
- Wei K, Li J, Ding M, Ma C, Yang HH, Farokhi F, et al. Federated learning with differential privacy: algorithms and performance analysis. *IEEE Trans Inf Forensics Secur*. (2020) 15:3454–69. doi: 10.1109/TIFS.2020.2988575
- Zhu T, Ye D, Wang W, Zhou W, Philip S.Y. More than privacy: applying differential privacy in key areas of artificial intelligence. In: *IEEE Transactions on Knowledge Data Engineering* (2020), 34, 2824–2843.
- El Oudrhiri A, Abdelhadi A. Differential privacy for deep and federated learning: a survey. *IEEE Access*. (2022) 10:22359–80. doi: 10.1109/ACCESS.2022.3151670
- Ni L, Huang P, Wei Y, Shu M, Zhang J. Federated learning model with adaptive differential privacy protection in medical IoT. *Wirel Commun Mob Comput*. (2021) 2021:1–14. doi: 10.1155/2021/8967819
- Khanna A., Schaffer V., Gürsoy G., Gerstein M. Privacy-preserving model training for disease prediction using federated learning with differential privacy. In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, (2022), pp. 1358–1361.
- Chen J.J., Chen R., Zhang X., Pan M. A privacy preserving federated learning framework for COVID-19 vulnerability map construction. ICC 2021-IEEE International Conference on Communications. IEEE, (2021) pp. 1–6
- Wu X, Zhang Y, Shi M, Li P, Li R, Xiong NN. An adaptive federated learning scheme with differential privacy preserving. *Futur Gener Comput Syst*. (2022) 127:362–72. doi: 10.1016/j.future.2021.09.015
- Ulhaq A., Burmeister O. Covid-19 imaging data privacy by federated learning design: A theoretical framework. (2020). arXiv [Preprint]. arXiv:2010.06177 2020.
- Wang X., Hu J., Lin H., Liu W., Moon H., Piran M.J. Federated learning-empowered disease diagnosis mechanism in the internet of medical things: from the privacy-preservation perspective. In: *IEEE Transactions on Industrial Informatics* (2022).
- Liu W, Cheng J, Wang X, Lu X, Yin J. Hybrid differential privacy based federated learning for internet of things. *J Syst Archit*. (2022) 124:102418. doi: 10.1016/j.sysarc.2022.102418
- Yang X, Dong Z. Kalman filter-based differential privacy federated learning method. *Appl Sci*. (2022) 12:7787. doi: 10.3390/app12157787
- Zhang L, Shen B, Barnawi A, Xi S, Kumar N, Wu Y. FedDPGAN: federated differentially private generative adversarial networks framework for the detection of COVID-19 pneumonia. *Inf Syst Front*. (2021) 23:1403–15. doi: 10.1007/s10796-021-10144-6

20. Ho TT, Tran KD, Huang Y. FedSGDCOVID: federated SGD COVID-19 detection under local differential privacy using chest X-ray images and symptom information. *Sensors*. (2022) 22:3728. doi: 10.3390/s22103728
21. Li Y, Yang S, Ren X, Zhao C. Asynchronous federated learning with differential privacy for edge intelligence. (2019) arXiv [Preprint]. arXiv:1912.07902 2019.
22. Lu Y, Huang X, Dai Y, Maharjan S, Zhang Y. Differentially private asynchronous federated learning for mobile edge computing in urban informatics. In: *IEEE Transactions on Industrial Informatics* (2019), 16, 2134–2143.
23. Nguyen J, Malik K., Zhan H., Yousefpour A., Rabbat M., Malek M., et al. Federated learning with buffered asynchronous aggregation. In: *International conference on artificial intelligence and statistics*. PMLR, (2022), pp. 3581–3607.
24. Li J, Jiang M, Qin Y, Zhang R, Ling SH. Intelligent depression detection with asynchronous federated optimization. *Complex Intell Syst.* (2023) 9:115–31. doi: 10.1007/s40747-022-00729-2
25. Nampalle K.B., Singh P, Narayan U.V., Raman B. Vision Through the Veil: Differential Privacy in Federated Learning for Medical Image Classification. (2023). arXiv [Preprint]. arXiv:2306.17794 2023.
26. Malik H, Naeem A, Naqvi RA, Loh WK. Dmfl_net: a federated learning-based framework for the classification of covid-19 from multiple chest diseases using x-rays. *Sensors*. (2023) 23:743. doi: 10.3390/s23020743
27. Dayan I, Roth HR, Zhong A, Harouni A, Gentili A, Abidin AZ, et al. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nat Med.* (2021) 27:1735–43. doi: 10.1038/s41591-021-01506-3
28. Kumar Sachin “Covid19-pneumonia-Normal chest X-ray images”, Mendeley Data, V1. (2022).
29. He B, Lu Q, Lang J, Yu H, Peng C, Bing P, et al. A new method for CTC images recognition based on machine learning. *Front Bioeng Biotechnol.* (2020) 8:897. doi: 10.3389/fbioe.2020.00897
30. Hu C, Xia T, Cui Y, Zou Q, Wang Y, Xiao W, et al. Trustworthy multi-phase liver tumor segmentation via evidence-based uncertainty. *Eng Appl Artif Intell.* (2024) 133:108289. doi: 10.1016/j.engappai.2024.108289
31. AbdulRahman S, Tout H, Ould-Slimane H, Mourad A, Talhi C, Guizani M. A survey on federated learning: the journey from centralized to distributed on-site learning and beyond. *IEEE Internet Things J.* (2020) 8:5476–97. doi: 10.1109/JIOT.2020.3030072



OPEN ACCESS

EDITED BY

Sultan Ahmad,
Prince Sattam Bin Abdulaziz University,
Saudi Arabia

REVIEWED BY

Adnen El Amraoui,
Université d'Artois, France
Furqan Aziz,
University of Birmingham, United Kingdom
Sourav Banerjee,
Kalyani Government Engineering College
(KGECE), India

*CORRESPONDENCE

Theyazn H. H. Aldhyani
✉ talldhyani@kfu.edu.sa

RECEIVED 16 April 2024

ACCEPTED 29 May 2024

PUBLISHED 20 June 2024

CITATION

Al-Nefaie AH, Aldhyani THH, Farhah N and
Koundal D (2024) Intelligent diagnosis system
based on artificial intelligence models for
predicting freezing of gait in Parkinson's
disease.

Front. Med. 11:1418684.

doi: 10.3389/fmed.2024.1418684

COPYRIGHT

© 2024 Al-Nefaie, Aldhyani, Farhah and
Koundal. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Intelligent diagnosis system based on artificial intelligence models for predicting freezing of gait in Parkinson's disease

Abdullah H. Al-Nefaie^{1,2}, Theyazn H. H. Aldhyani^{1,3*},
Nesren Farhah⁴ and Deepika Koundal^{1,5}

¹King Salman Center for Disability Research, Riyadh, Saudi Arabia, ²Department of Quantitative Methods, School of Business, King Faisal University, Al-Ahsa, Saudi Arabia, ³Applied College in Abqaiq, King Faisal University, Al-Ahsa, Saudi Arabia, ⁴Department of Health Informatics, College of Health Sciences, Saudi Electronic University, Riyadh, Saudi Arabia, ⁵School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India

Introduction: Freezing of gait (FoG) is a significant issue for those with Parkinson's disease (PD) since it is a primary contributor to falls and is linked to a poor superiority of life. The underlying apparatus is still not understood; however, it is postulated that it is associated with cognitive disorders, namely impairments in executive and visuospatial functions. During episodes of FoG, patients may experience the risk of falling, which significantly effects their quality of life.

Methods: This research aims to systematically evaluate the effectiveness of machine learning approaches in accurately predicting a FoG event before it occurs. The system was tested using a dataset collected from the Kaggle repository and comprises 3D accelerometer data collected from the lower backs of people who suffer from episodes of FoG, a severe indication frequently realized in persons with Parkinson's disease. Data were acquired by measuring acceleration from 65 patients and 20 healthy senior adults while they engaged in simulated daily life tasks. Of the total participants, 45 exhibited indications of FoG. This research utilizes seven machine learning methods, namely the decision tree, random forest, Knearest neighbors algorithm, LightGBM, and CatBoost models. The Gated Recurrent Unit (GRU)-Transformers and Longterm Recurrent Convolutional Networks (LRCN) models were applied to predict FoG. The construction and model parameters were planned to enhance performance by mitigating computational difficulty and evaluation duration.

Results: The decision tree exhibited exceptional performance, achieving sensitivity rates of 91% in terms of accuracy, precision, recall, and F1- score metrics for the FoG, transition, and normal activity classes, respectively. It has been noted that the system has the capacity to anticipate FoG objectively and precisely. This system will be instrumental in advancing consideration in furthering the comprehension and handling of FoG.

KEYWORDS

freezing, Parkinson's, gait, machine leaning, prediction, classification, transformers models

1 Introduction

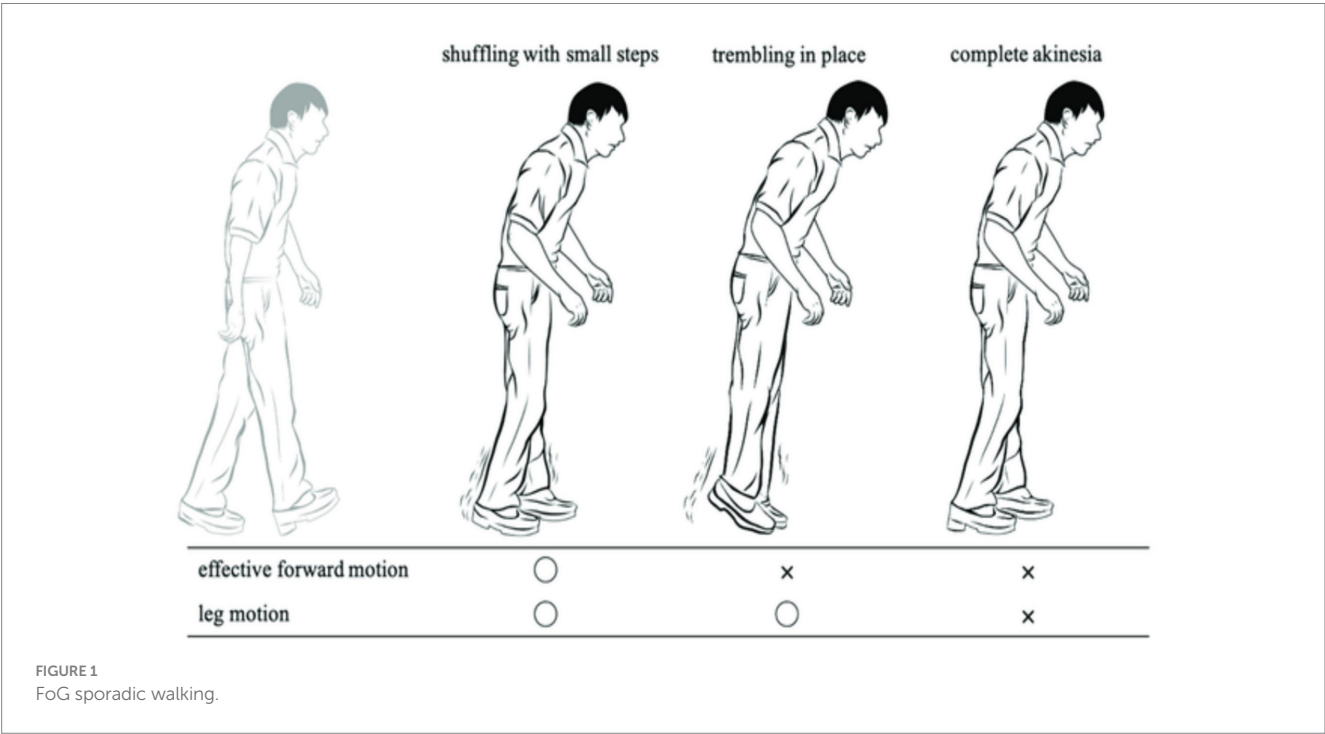
Parkinson’s disease (PD) is a degenerative neurological sickness that disturbs a large number of individuals (1). Freezing of gait (FoG) and the subsequent increased risk of falls are the primary disabling issues for a noteworthy figure of individuals with PD (2). There are presently few options for pharmacological therapies. Several tools and wearable devices that make available treatments, like rhythmical cueing and step-synchronized vibratory cueing, demonstrate good concert and results (3). Efficient treatment of FoG is now being investigated via examination on FoG recognition and prediction.

FoG is a sporadic walking problem characterized by sudden interruptions in stride or a significant decrease in forward movement of the feet (4). It greatly impacts quality of life and increases the likelihood of reductions and breakages in individuals with PD (2, 5). These symptoms may disrupt patients’ everyday activities, jeopardize their mental well-being, and lead to a weakening in their superiority of life. Approximately half of individuals with PD have encountered signs of FoG, which is the primary factor leading to falls (6–8). FoG is characterized as a temporary and intermittent inability or noteworthy reduction in the advancing motion of the feet, even when there is a desire to walk. In their study, Schaafsma et al. (9) categorized FoG into five distinct subtypes: start hesitation, turn hesitation, hesitation in confined spaces, hesitation toward a specific goal, and hesitation in wide spaces. Typically, FoG is linked to a particular sensation of “the feet being adhered to the ground” (10). FoG is influenced by surroundings, drugs, and anxiety, which might impact its frequency and duration (11). FoG is often considered to be a characteristic of akinesia, which is a severe type of bradykinesia (12). FoG is characterized by transient periods of immobility or the execution of very small steps while attempting to begin walking or change direction (2). The state of FoG is significantly influenced by ambient cues, cognitive input, medicines, and anxiety (11, 13). It is more common

to experience it at home rather than in a clinical environment, particularly in scenarios when there is full darkness or when there is a higher cognitive load, such as dual-tasking conditions (14–17). Figure 1 displays FoG sporadic walking.

FoG is a very debilitating condition often seen in individuals with PD. The symptoms often manifest in the later stages of the illness, with roughly 50% of all PD patients experiencing some indications and around 80% being significantly impacted (10, 18–20). Episodes of FoG often present as a sudden and temporary inability to initiate movement, often occurring while starting to walk, during making turns, or under stressful circumstances. During bouts of FoG, individuals with PD experience a phenomenon where they perceive their feet to be firmly stuck to the ground without any apparent cause (9). During episodes of FoG while walking, patients exhibit variations in their walking pattern and experience a significant decrease in the length of their steps. Additionally, they often display shaking in their legs (19, 20). The typical frequency range for normal gait steps, as measured by ankle sensors, is 0.5 to 3 Hz. However, FoG occurrences have a higher rate variety of 6 to 8 Hz (21–23).

Recent research has begun using machine leaning and deep learning for the resolution of automated categorization. Deep learning is a branch of artificial intelligence (AI) that utilizes algorithms having capability of mechanically extracting distinguishing features from information and data, such as signals acquired straight from sensors without any prior processing. Deep learning (DL) and machine learning (ML) have facilitated the creation of classifiers that cover the entire process and have demonstrated exceptional performance in various fields, including image processing, computer vision, medical information analysis, bioinformatics, natural language processing, logical reasoning, robotics, and control (24–27). Therefore, DL techniques have been used in human activity recognition (HAR) systems utilizing data collected from various light sensors (28, 29).



DL and ML methods have become more popular for detecting FoG in recent years, as seen by the employment of these techniques in several studies (30–34). The following are the most significant and noteworthy. Kim et al. (30) and Pepa et al. (32) introduced a novel sensing tool, namely a smartphone positioned in the pant pocket, as a more convenient method for monitoring patients with PD and detecting FoG. The researchers used a technique that relied on convolutional neural networks (CNN) to automatically extract distinctive characteristics from sensors integrated into an Android smartphone. The performance of the CNN classifier was compared to that of the random forest (RF) classifier, and the CNN classifier exhibited a sensitivity that was 20% greater than that of the RF classifier.

Approximately 7 to 10 million individuals worldwide are affected by PD, with a significant portion experiencing FoG. During an episode of FoG, a patient experiences a phenomenon where their feet get immobilized, making it impossible for them to go forward despite their efforts. FoG significantly impairs health-related quality of life, leading to depression, heightened fall risk, greater reliance on wheelchairs, and limited autonomy.

This study used a standardized dataset obtained from 65 participants, using a 3D accelerometer. The dataset has been categorized into four classes: Normal, Turn, Walking, and StartHesitation. Preprocessing methods were suggested to cleanse the dataset and address the issue of imbalanced classes. The output from the preprocessing approach was analyzed using several ML, deep learning and transformers modes to determine if the patients are experiencing FoG or are in a normal state. The primary contribution of this work is as follows:

- 1 The initial system employed for the classification of FoG used a new dataset.
- 2 In our research, we have categorized the dataset into four distinct classes namely Normal, Turn, Walking, and StartHesitation because the dataset did not have labels.
- 3 Employed various of ML, deep learning, and transformer approaches to predict the occurrence of FoG in patients with PD, the system achieved 91% with respect to accuracy.

2 Background of the study

FoG is an indication often seen in people with PD. However, the fundamental mechanisms of FoG are not well understood. Patients with PD often report this symptom as a sensation of their feet being firmly adhered to the ground (34–37). Handojoseno et al. (38) utilized the wavelet factors of electroencephalogram (EEG) data as the input for the multilayer perceptron neural network and KNN technique. This method achieved a sensitivity of 87% and an accuracy of 73% in predicting the transition from walking to FoG. Delval et al. (39) used a multi-camera setup to capture the gait kinematics gestures of patients. Deep pointers were affixed to the patients' bodies and recorded from various angles. Okuno et al. (40) utilized a plantar pressure measurement system of 1.92 m × 0.88 m for recording the walking patterns of patients by monitoring the weight exerted on their soles. While the sensors may all be used for FoG detection, the predominant method for FoG detection in community environments relies on inertial sensors.

Moore et al. (21) developed a portable monitoring apparatus and algorithm that used the occurrence features of vertical leg movement. This movement was measured using an accelerometer put on the left shank of 11 individuals with PD. The contributors' ages ranged from 45 to 72 years. The contributors were trained to go through a series of interior passages, including a tight entryway, and three obstacles. This research took into account the specific effects of the levodopa/carbidopa drug combination throughout both the “on” and “off” periods. The researchers used a threshold-based method to identify FoG, achieving a FoG detection rate of 78% and an accuracy rate of 89%. Delval et al. (39) conducted research in which they induced FoG in patients and used a series of measurable indicators to identify the presence of FoG. They used a 3D motion-analysis device to capture video footage of 10 sick and 10 healthy people while they were on a treadmill. Indicators were affixed to the heels, toes, ankles, shoulders, and on the T10 vertebra. Obstacles were encountered due to special situations, causing the patients to be in an inactive state. The identification of FoG in that particular investigation relied on a combination of threshold and frequency investigation. Bachlin et al. (41) devised a FoG recognition architecture using three accelerometers and implementing Moore's threshold-based algorithm (21). Upon detecting an episode of FoG, the device used a metronome to offer stimulation to the patient, aiding them in regaining their focus and stability. The system support resulted in improved gait for six out of eight individuals who had FoG. Azevedo et al. (41) Developed a FoG detector that included gait pattern analysis by using a solitary inertial sensor positioned on the lower extremity. Based on its findings, it determines that relying just on frequency-based analysis is insufficient for accurately identifying the occurrence of FoG. It is essential to not only detect but also forecast when a FoG event will take place. The authors used rhythm and tread data into their methodology to enhance the categorization process. In order to assess the walking patterns of individuals with PD, Jovicic et al. (42) developed a technique that utilizes inertial sensors placed on both lower legs to categorize different gait patterns. The system also distinguished between regular and pathological gait by utilizing an expert rule-based approach, based on data collected from 12 PD patients who walked over a convoluted course. A rule-based categorization approach was used for the identification and categorization of FoG. Pham et al. (43) introduced a FoG detection method that is not reliant on specific individuals. The uniqueness of this idea is in its ability to operate autonomously from the topic matter. An additional instance of a FoG recognition system that uses wearable accelerometers and video capture to categorize the occurrence is shown in the research conducted by Zach et al. (44). Their finding suggests that FoG may be detected with just one accelerometer placed in the lumbar area.

Pepa et al. (32) used soft computing approaches for FOG identification. A fuzzy method was created to integrate information pertaining to freeze index, energy, cadency fluctuation, and the derivative energy ratio. A building was constructed that relied on a smartphone as its foundation. Their findings demonstrated that, on average, the system exhibited a specificity of 92.33% and recall of 83.33% in classifying FoG events. Cole et al. (36) presented a method using dynamic neural networks (DNN) to accurately identify FoG. They gathered information from three accelerometers and an electromyographic shallow worn by patients and achieved favorable consequences in terms of detection. A noteworthy involvement of this study is the creation of a database documenting

unscripted and unimpeded everyday activities of PD patients, including instances of FoG. Ahlrichs et al. (22) introduced a FoG detector that utilizes a single accelerometer worn at the waist and a recognizer based on SVM. They documented the performance of 20 people with PD engaging in pre-planned everyday tasks. Patients were required to be documented both when taking medicine and while not taking medication. Their findings demonstrated a precision rate of 98.7%.

Rodriguez-Martin et al. (45) developed a ML method designed to identify episodes of FoG. Their preference for FoG detection was SVM. Their technique relies on a solitary 3D accelerometer positioned at the waist to identify FoG in real-world scenarios. A total of 21 individuals diagnosed with PD contributed in the research work. The patients were asked to execute two sets of pre-determined exercises during both their “off” and “on” times. These activities were associated with everyday existence. According to their research, the medicine had an impact on the patients’ motor reaction. Deep learning methods have been popular for detecting FoG in recent years, as seen by their frequent application in research (30, 34, 46–48). Kim et al. (30) used a novel sensing device, namely a smartphone positioned in the trouser pocket, to discover a more pragmatic approach for monitoring patients with PD and identifying FoG. The researchers used a technique that relied on CNN to automatically extract distinctive characteristics from sensors integrated into an Android smartphone. The performance of the CNN method was compared to that of the RF technique, and the CNN exhibited a sensitivity that was 20% better than that of the RF classifier. Xia et al. (49) suggested a FoG detection method based on CNN to accomplish automated feature learning and classification for FoG. Bachlin et al. (41) conducted experiments that relied on the patient’s input and studies that did not need the patient’s involvement. The most favorable outcomes were documented in the patient-dependent experiments. Same researchers used DL to predict FoG and PD (50–53).

3 Materials and methods

The proposed system aims to identify FoG, a distressing symptom that affects many individuals with PD. The proposed solution is built

upon a machine learning models that have been trained using data obtained from a wearable 3D sensor device positioned on the lower end. Figure 2 displays the framework of the FoG system based on a machine learning approach.

3.1 Dataset

The dataset was obtained from the Kaggle repository and consists of 3D accelerometer data from the lower back of individuals experiencing bouts of FoG, a debilitating condition often seen in individuals with PD. FoG has a detrimental effect on the ability to walk, hindering movement and independence. The goal is to identify the initiation and termination of each freezing episode, as well as the presence of three specific kinds of FoG events: start hesitation, turning, and walking. The data series consists of three unique datasets, each obtained under separate circumstances: (1) The tDCS FoG (tdcsfog) dataset consists of data series obtained in a laboratory setting, where individuals underwent a FoG-provoking procedure; (2) The DeFOG dataset consists of data series that were obtained in the subject’s home as they conducted a FoG-provoking regimen; and (3) The daily living dataset consists of 1 week of uninterrupted 24/7 recordings from 65 people. Out of the total number of participants, 45 display symptoms of FoG and also have series in the DeFOG dataset. In contrast, the other 20 patients do not show any symptoms of FoG and do not have series in any other part of the data. Table 1 displays meta data, whereas the training dataset is presented in Table 2.

3.2 Preprocessing approach

Data features engineering require the creation of new features or the transformation of existing features to enhance the effectiveness of a machine-learning model. Data preprocessing entails the extraction of pertinent information from unprocessed data and converting it into a format that is readily comprehensible by a model. The objective is to enhance the precision of the model by providing more significant and relevant data. The missing values in

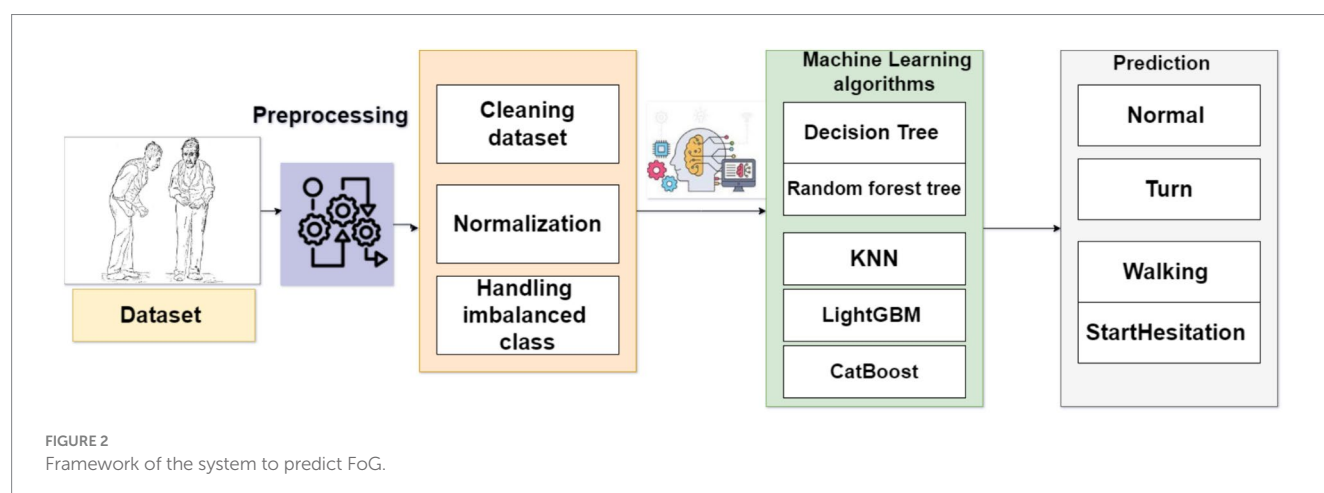
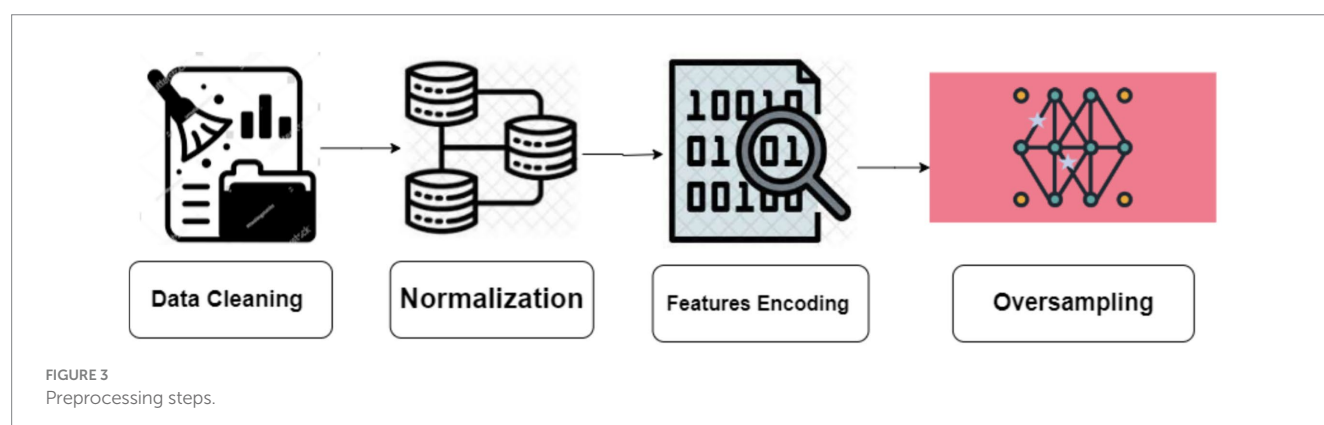


TABLE 1 Metadata of dcsfog and tdcsfog.

Features name	Description	Types of dataset
Visit		Int64
Medication		Int64
Time	A numerical value representing a discrete unit of time. The tdcsfog dataset records series at a frequency of 128 Hz, meaning there are 128 timesteps per second. On the other hand, the defog and daily series are recorded at a frequency of 100 Hz, resulting in 100 timesteps per second.	
AccV, AccML, AccAP	The lower-back sensor measures acceleration along three axes: vertical (V), mediolateral (ML), and anteroposterior (AP). The data is expressed in units.	Float64
Event	Class	Object

TABLE 2 features of dataset.

Features name	Description
Visit	Lab visits include an initial evaluation, two subsequent evaluations for distinct therapy phases, and a final evaluation for follow-up purposes.
Test	Test used
Medication	Subjects may have been either receiving or not receiving anti-parkinsonian medication throughout the recording.



the dataset were removed from all features. We have combined DeFOG features, namely Time, AccV, AccML, and AccA, with the DeFOG-metadata for Subject, Visit, and Medication Condition. Figure 3 shows the preprocessing steps for the classification of FoG of PD patients.

3.2.1 Normalization

Normalization is an essential preprocessing step for any machine-learning task. The process can be executed by either scaling or altering the initial data in order to equalize the influences of various characteristics in the data examples. In the present research work, we have standardized the input data to generate a representation among one and zero.

$$x_{normalize} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

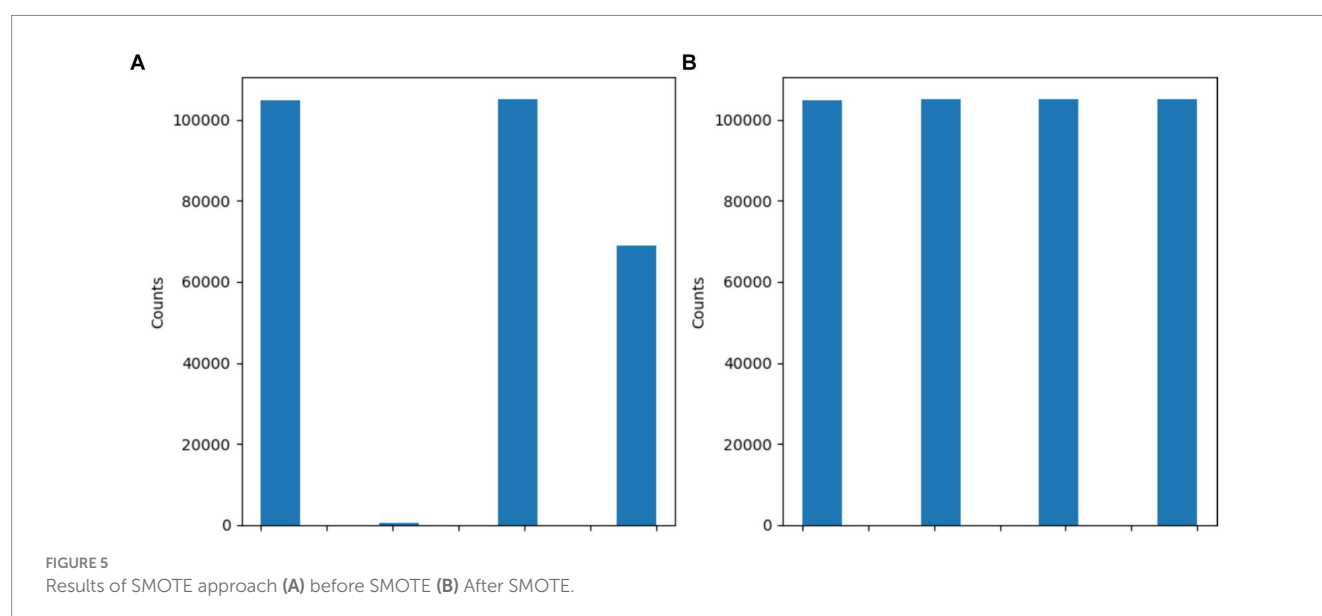
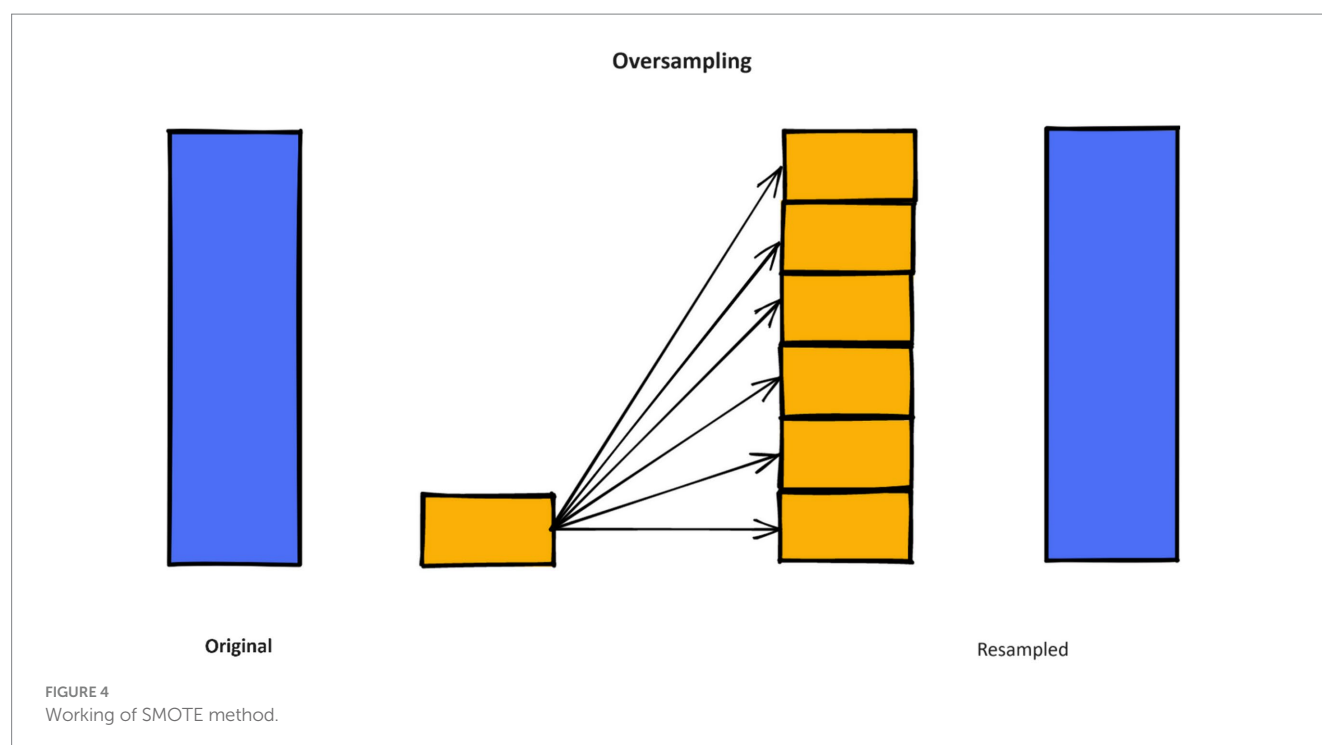
Where the x is training data, and $x_{\min}$ is maximum value [1] and $x_{\max}$ is minimum value [0].

3.2.2 Handling imbalance classes

Unbalanced data raises to a condition where the representation of observations and samples among dissimilar classes is unequal, with one class dominating the dataset and the other classes having insufficient representation.

The synthetic minority oversampling strategy (SMOTE) is a resampling strategy used to address extremely imbalanced datasets by creating synthetic samples in the minority class, hence increasing its representation. SMOTE is effective in increasing the figure of minority class examples and achieving class balance. To mitigate the problem of overfitting, the synthetic production of fresh samples deviated from the increase procedure.

The primary concept behindhand SMOTE technique is to create additional data samples in the minority class using interpolation between neighboring examples within this class (54). SMOTE enhances the amount of instances belonging to the minority class in an unbalanced dataset, thus improving the classifier's ability to generalize well. Figure 4 shows the SMOTE method in practice.



$$D_{new} = D_i + \left(\hat{D}_j - D_i \right) \times \delta \quad (2)$$

The dataset D_{new} represents the ADHD dataset. D_i consists of samples from the minority group, whereas $\hat{D}_j$ is a k-nearest neighbor of D_i . Let δ represent a uniformly distributed random number between 0 and 1. We used the SMOTE technique to enhance the categorizing process.

Figure 5 and Table 3 show the dataset before and after class distribution of the dataset using the SMOTE approach in the training dataset. The startHesitation class has less values (352); therefore, we have applied the SMOTE approach for handling this imbalance class to enhance the machine algorithms.

3.3 Algorithms

3.3.1 K-nearest neighbors

The KNN technique is a straightforward nonparametric approach that is often utilized for the purposes of regression and classification tasks. The KNN algorithm is a kind of instance-based learner, commonly referred to as idle learning. It does not build a categorization model-based approach till it is given samples to classify. The fundamental premise of KNN in categorization is to compare individual test samples with k nearby training samples in the variable space. The category of the test sample is determined based on the classification of its nearest k neighbors. Neighbors are often determined by calculating the

TABLE 3 Results of SMOTE approach.

Before the SMOTE	
Classes	Values
Normal	105,176
Turn	104,785
Walking	68,999
StartHesitation	352
After the SMOTE	
Normal	105,176
Turn	105,176
Walking	105,176
StartHesitation	104,785

TABLE 4 RF parameters.

Parameters name	Values
Estimators	500
Criterion	gini
Min_samples_leaf	1
Max_depth	10
Max_features	auto
Random_state	42

Euclidean distance between the data point being analyzed and its k nearest neighbors. The k parameter, denoting the quantity of nearest neighbors' number, is often kept minimal to avoid the inclusion of excessive data points that may distort the underlying characteristics of the data point under consideration. It is important to choose acceptable values for k in order to avoid overfitting and model instability, since large values of k might contribute to both issues. KNN utilizes the Euclidean distance metric. The underlying assumption is that each element in the dataset may be shown as a point in a space with N dimensions. KNN utilizes a parameter k to denote the number of examples to be considered, based on which the majority class is selected to categorize the new instance.

$$E_i = \sqrt{(x_1 - x_2)^2 + (x_3 - x_4)^2} \quad (3)$$

where x_1 , x_2 , x_3 , and x_4 calculate the Euclidean distance in a two-dimensional space.

3.3.2 Decision tree

A decision tree (DT) is a well-recognized nonparametric supervised learning technique. DT is one of the ML algorithms that can be applied for both regression and classification tasks. DT classifies the instances by traversing down the tree from the root to certain leaf nodes. Instances are categorized by evaluating the attribute specified by the node, beginning at the root node of the tree, and thereafter down the tree branch associated with the attribute value. The most often used criteria for splitting are "gini" for measuring Gini impurity and "entropy" for quantifying information gain, which may be mathematically represented.

$$Entropy(S) = - \sum_{i=1}^C p_i \log_2 p_i \quad (4)$$

$$Entropy(S|B) = \sum_{j=1}^J \frac{|S_j|}{|S|} entropy(S_j) \quad (5)$$

$$Gain(S|B) = entropy(S) - entropy(S|B) \quad (6)$$

The training dataset is indicated as S , while the freezing of gait dataset is represented by the class c , which encompasses both attack and normal data. The likelihood of seeing data that belongs to class S_i is represented as P_i . This probability is specifically related to the subsets of class S_i in the characteristics B .

3.3.3 Random Forest

A random forest (RF) classifier is a well-recognized collaborative classification technique used in machine learning and data science across several application domains. This approach employs "parallel ensembling," whereby several DT classifiers are concurrently trained on distinct sub-samples of the dataset. The ultimate result is decided via mainstream vote or averaging of the outcomes. Therefore, it reduces the issue of overfitting and enhances both the accuracy of predictions and control. Hence, the RF learning model, which utilizes many decision trees, often exhibits higher accuracy compared to a model based on a single decision tree. In order to construct a sequence of decision trees with regulated diversity, the method associates bootstrap combination (bagging) with arbitrary attributes selection. It is versatile for both classification and regression issues and is suitable for both categorical and continuous variables. Table 4 shows parameters of RF model.

3.3.4 LightGBM approach

LightGBM approach is a gradient boosting context that employs tree-based learning techniques. It is specifically engineered to be widely spread and highly effective, offering the following benefits: Enhanced training velocity and increased efficacy; Reduced memory consumption LightGBM provides support for parallel and GPU learning; Proficient at managing enormous volumes of data LightGBM is a rapid, circulated, and efficient gradient-boosting system that relies on decision tree methods. It is extensively used in a range of machine-learning tasks, including regression, ranking, and categorization (55). It is a furthering method that utilizes numerous weak machine-learning methods to create a powerful learning model. Boosting methods amplify the weights of incorrectly classified data while reducing the weightiness of successfully categorized data. Table 5 shows LightGBM parameter.

3.4 Gated recurrent unit–transformers

3.4.1 Gated recurrent unit

The GRU is a fundamental architecture of recurrent neural networks (RNNs) that has resemblance to Long Short-Term Memory (LSTM) models. GRU is specifically developed to represent

sequential data by enabling the selective retention or loss of information over time. Nevertheless, GRU possesses a more streamlined structure compared to LSTM, with a reduced number of parameters. This characteristic facilitates training and enhances computing efficiency.

The GRU is designed to handle sequential data by iteratively updating its hidden state in response to both the current input and the prior hidden state. During each iteration, the GRU calculates a “candidate activation vector” that integrates data from the input and the preceding hidden state. Subsequently, the candidate vector is employed to modify the concealed state for the subsequent time step. Two gates, namely the reset gate and the update gate, are used to calculate the candidate activation vector. The reset gate is responsible for determining the extent to which the previous hidden state is disregarded, whereas the update gate is responsible for determining the extent to which the candidate activation vector is integrated into the future hidden state.

$$\mu_t = \sigma(V_\mu x_t + W_\mu o_{t-1} + b_\mu) \tag{7}$$

$$r_t = \sigma(V_r x_t + W_r o_{t-1} + b_r) \tag{8}$$

$$i_t = \tanh(V_o x_t + W_o(r_t \odot o_{t-1}) + b_o) \tag{9}$$

$$o_t = \sigma(\mu_t \odot o_{t-1} (1 - \mu_t) \odot i_t) \tag{10}$$

TABLE 5 LightGBM parameters.

Parameters name	Values
Estimators	500
Learning_rate	0.01
Max_depth	10
Random_state	42

Input is i_t , output is o_t , update gate output is α_t , reset gate output is r_t , and Hadamard product is $\odot$. Weight matrices V , W , and b are parameters. The GRU encoder and Transformer path embeds input sequences using a recurrent GRU layer. Thirty-two GRU units encoded 200-dimensional vectors each timestep. Using multi-head self-attention with two heads, GRU embeddings may attend to each other based on learnt connections. Residual connections and layer normalization stabilize training. Flattening attention outputs to 1D vectors. Structure of GRU mode is presented in Figure 6.

3.4.2 Transformers

The self-attention mechanism-based sequence-to-sequence model Transformer is extensively used in natural language processing methods including machine translation, text summarization, language synthesis. Significant outcomes are achieved quickly. Transformers has a different architecture than RNN. The Transformer branch in the proposed GRU-Transformer model assumes a crucial function in capturing complex interdependencies and multidimensional characteristics present in the input sequence. The aforementioned objective is accomplished by utilizing the self-attention and multi-head attention processes of the Transformer, as seen in Figure 3. Its attention-based encoder-decoder structure enables the Transformer to effectively handle sequence-to-sequence tasks.

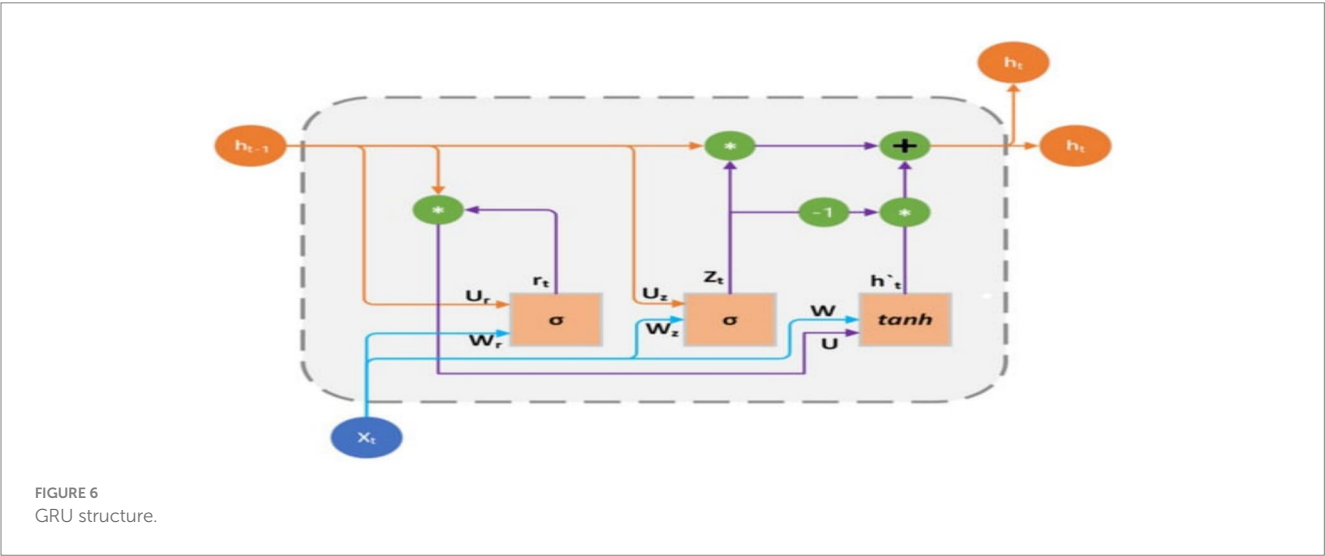
$$Q = XW_Q \tag{11}$$

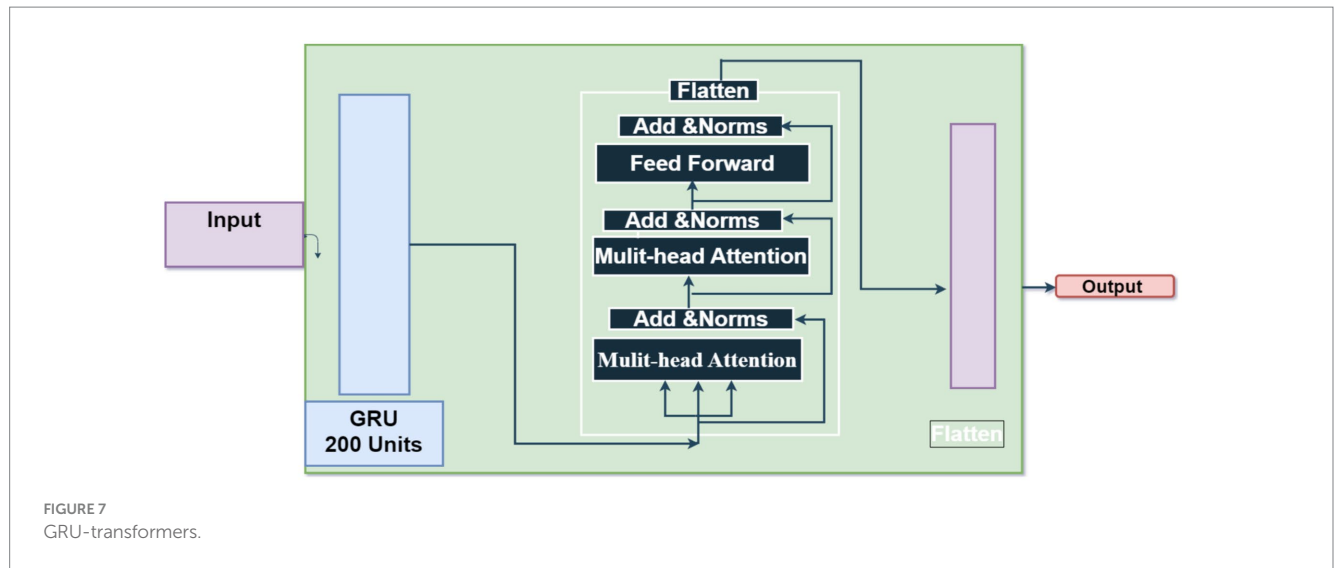
$$K = XW_K \tag{12}$$

$$V = XW_V \tag{13}$$

$$A = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \tag{14}$$

$$Y = AV \tag{15}$$





Where, X be the input and (K, Q, V) is query matrix, key matrix, value matrix, learnable weight matrix is A , attention matrix is Y , output matrix is $\sqrt{d_k}$, and attention header dimension, the scaling factor, reduces overly large or minuscule attention weights. To determine key value weight, softmax is used as a normalizer. The attention mechanism calculates the association between each input sequence item and the others to capture global dependencies.

The unit recurrent layer is 200 unit that stores sequence data and may capture dependencies. The parameter “return_sequences” sends the sequence of outputs for each time step to the next layer instead of just the final output. This Transformer component lets the model focus on different input sequence segments during prediction. Two 200-key dimension attention heads are used in the suggested method. This implementation helps the layer capture data relationships and connections. Attention boosts and accelerates learning. The residual link, or skip connection, solves the fading gradient problem by offering an alternate gradient movement path. Each time step of the sequence receives an individual 120-unit dense layer to extract unique characteristics. This strategy stochastically assigns input units to 0 during training after the TimeDistributed layer at 0.2 to reduce overfitting.

The output of the previous layers is turned into a unified vector to link with the final Dense layer for classification. The neural network generates probabilities for each of the four classes using a Dense layer with softmax activation. Figure 7 shows the structure of GRU-transformers. Parameters of GRU-transformers is presented in Table 6.

3.5 Long-term recurrent convolutional networks

LRCN neural networks combine the strengths of the CNN and RNN to handle sequential input with spatial and temporal dependency. The model's early layers use Convolutional Neural Networks (CNNs) to extract spatial properties from input data. These collected characteristics feed Recurrent Neural Networks (RNNs) to capture temporal relationships and long-term correlations. LRCN may

TABLE 6 paramters of GRU-transformers.

Parameters	Values
GRU	32 units
Multi-Head Attention	2 heads
Add	2 heads
Layer Norm	--
Flatten	---
Dropout	0.5
Dense	1,024

acquire spatial and temporal complex data representations by integrating CNN and LSTM components. This neural network design handles sequential data well. LRCN is an RNN developed to evaluate its performance on sequence input data.

$$C = \sum_{i=1}^i \sum_{j=1}^j I_{ij} F_{ij} \quad (16)$$

$$C_i^L = B_i^L + \sum_{j=1}^{x^{(L-1)}} F_{i,j}^L * C_j^{(L-1)} \quad (17)$$

Where, F represents a convolution kernel or filter, while i and j represent rows and columns of dataset. A unique two-dimensional output is obtained by convolving the input dataset.

With the kernel. B_i^L represents the bias matrix, whereas $F_{i,j}^L$ represents the filter connecting the j th feature map in the layer.

$$f_t = \sigma(W_{ef}X_t + W_{ef}h_{t-1} + W_{cf}C_{t-1} U_f +) \quad (18)$$

$$i_t = \sigma(W_{xi}X_t + W_{hi}h_{t-1} + W_{ci}C_{t-1} + U_i) \quad (19)$$

$$C_t = \sigma(f_t c_{t-1} + i_t \tanh(W_{xc}X_t + W_{hc}h_{t-1} + U)) \quad (20)$$

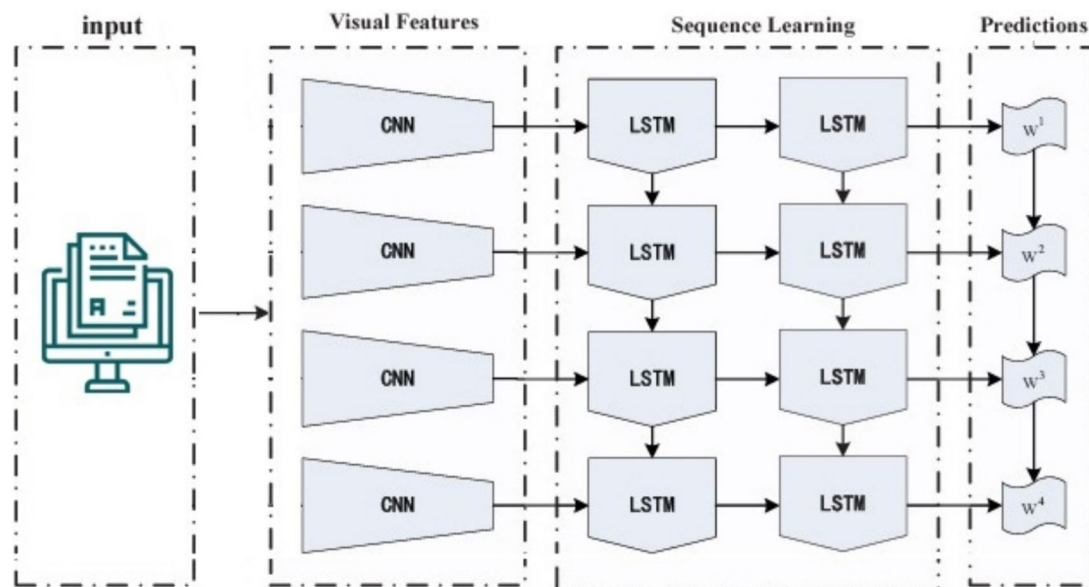


FIGURE 8
Structure of LRCN model.

$$o_t = \sigma(W_{xo}X_t + W_{ho}h_{t-1} + W_{co}C_{t-1} + U_o), \quad (21)$$

$$h_t = O_t \times \tanh(C_t) \quad (22)$$

Sequential forward and reverse methods apply the equations above. They represent the LSTM model equations. A gated cell in the LSTM network evaluates input data and retains it based on relevance or weight. The input gate, forget gate, and output gate make up the LSTM model. The forget gate f_t decides whether states to keep or discard. The input gate i_t modifies the value based on signals. The output gate o_t transmits cell status to neighboring neurons. The design has a logistic layer and a layer that generates a new vector to mix with the state. In a recurrent neural network (RNN), the hidden layer processes X_t using the weight matrix W to produce y_t . The LSTM model uses a memory cell called h_t , which is governed by three gates. The structure of LRCN is presented in Figure 8.

3.6 Evaluation metrics

Prior to further exploring our study, it is essential to elucidate the significance and computation techniques of several assessment measures. In this study, we have selected four primary assessment metrics: accuracy, precision, recall, f1-score, and rate of change (ROC).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \quad (23)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (24)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (25)$$

$$Fscore = \frac{2 * precision * Sensitivity}{precision + Sensitivity} \times 100\% \quad (26)$$

Algorithms of ML algorithms.

Let D be the dataset containing sensor data from FoG Parkinson's disease patients, where $D = \{(X_i, Y_i)\}_{i=1}^N$ where X_i represents the features and Y_i represents the corresponding FoG labels.

D is collected from wearable devices.

Data preprocessing.

clean the data $D' = \text{clean}(D)$.

Normalize the data $D'' = \text{normalize}(D')$

Resample the data $D''' = \text{resample}(D'')$

Feature extraction

Extract features: $X = \{X_i\}_{i=1}^N$

Model training

4.1 Select machine learning algorithms: $ML_Algorithms = \{DT, RF, KNN, LightGBM, CatBoost\}$

4.2 Split the data into training and testing sets: $D_{train}, D_{test} = \text{Split}(D''', 70\%)$

4.3 Train the models: $Model_j = \text{train}(ML_algorithm_j, D_{train})$; $j = 1, 2, 3, \dots, num_algorithms$

Model evaluation

$evaluate_models_j; Metrics_j = evaluate(Model_j, D_{test}), j = 1, 2, 3, \dots, num_algorithms$

FoG Detection:

Predict FoG instances: $\hat{Y} = \text{predict}(model_{best}, X)$

FoG_Events = detect($\hat{Y}$)

4 Experimental

This section presents the classification results and discoveries derived from a sequence of experiments carried out for predicting PD FoG by applying machine-learning algorithms. The main aim of these experiments was to evaluate the efficacy of several classification models

in accurately distinguishing various types of classes associated with gait behavior, specifically Normal, Turn, StartHesitation, and Walking. The evaluation primarily examined evaluation parameters such as accuracy, precision, recall, and f1-score for each class, offering valuable insights into the capabilities as well as limitations of the applied models. This part included simulation setup, split dataset, and machine-learning results.

4.1 Simulation setup

This module encompasses the specific steps and procedures involved in carrying out our suggested approaches. The instruments used in this document are enumerated in Table 7.

TABLE 7 Environmental requirements of the presented model.

Hardware	Software
RAM size 16 GB	Python
Intel(R) Core(TM) i7	Panda
CPU GHz	TensorFlow library
	Keras library
	Matplotlib
	NumPy library

TABLE 8 Testing results of the RF model.

Model	Class	Precision %	Recall %	F1-score %	Accuracy %
RF	Normal	87	85	90	90
	startHesitation	98	99	96	
	Turn	90	91	89	
	Walking	94	96	93	
	Weighted	90	90	90	

4.2 Split dataset

The dataset was divided into a 70% training dataset and a 30% testing dataset.

4.3 Results

4.3.1 Random forest testing results

Table 8 provides the testing results of the RF model for PD FoG. It had strong performance in accurately differentiating the “Turn class,” with a precision of 0.98, recall of 0.99, f1-score of 0.96, and a total accuracy of 90%. Though, there were complications in precisely detecting occurrences of the Turn class, as the recall rate was significantly lower despite a high precision score.

Figure 9 displays the confusion matrix of the RF model used for the classification of FoG of PD disease patients. The misclassification rate of the RF model in diagnosing FoG is less. The RF model exhibited a true negative rate of 25,586 for the classification of FoG. The number of true positive instances classified are 25,586 as Normal, 99 as Turn, 27,078 as startHesitation, and 18,999 as Walking.

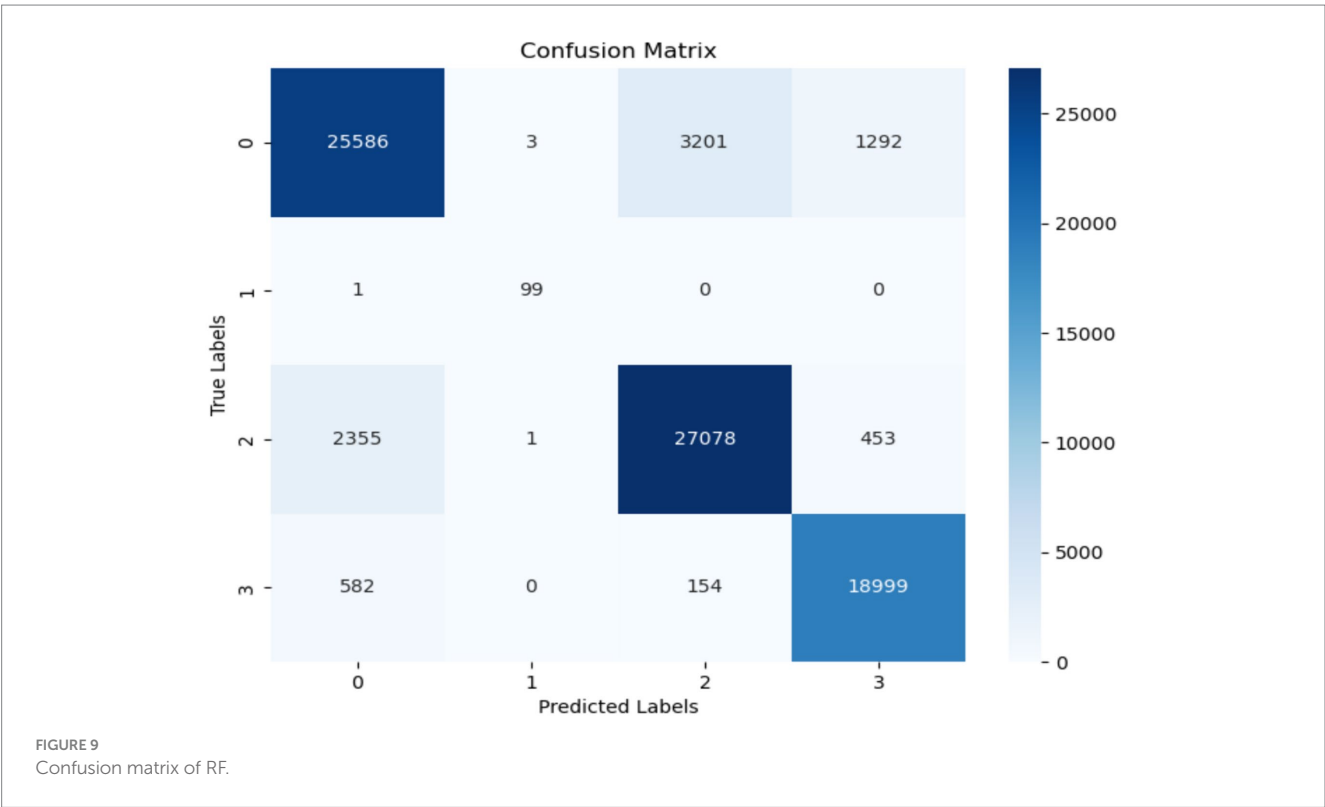
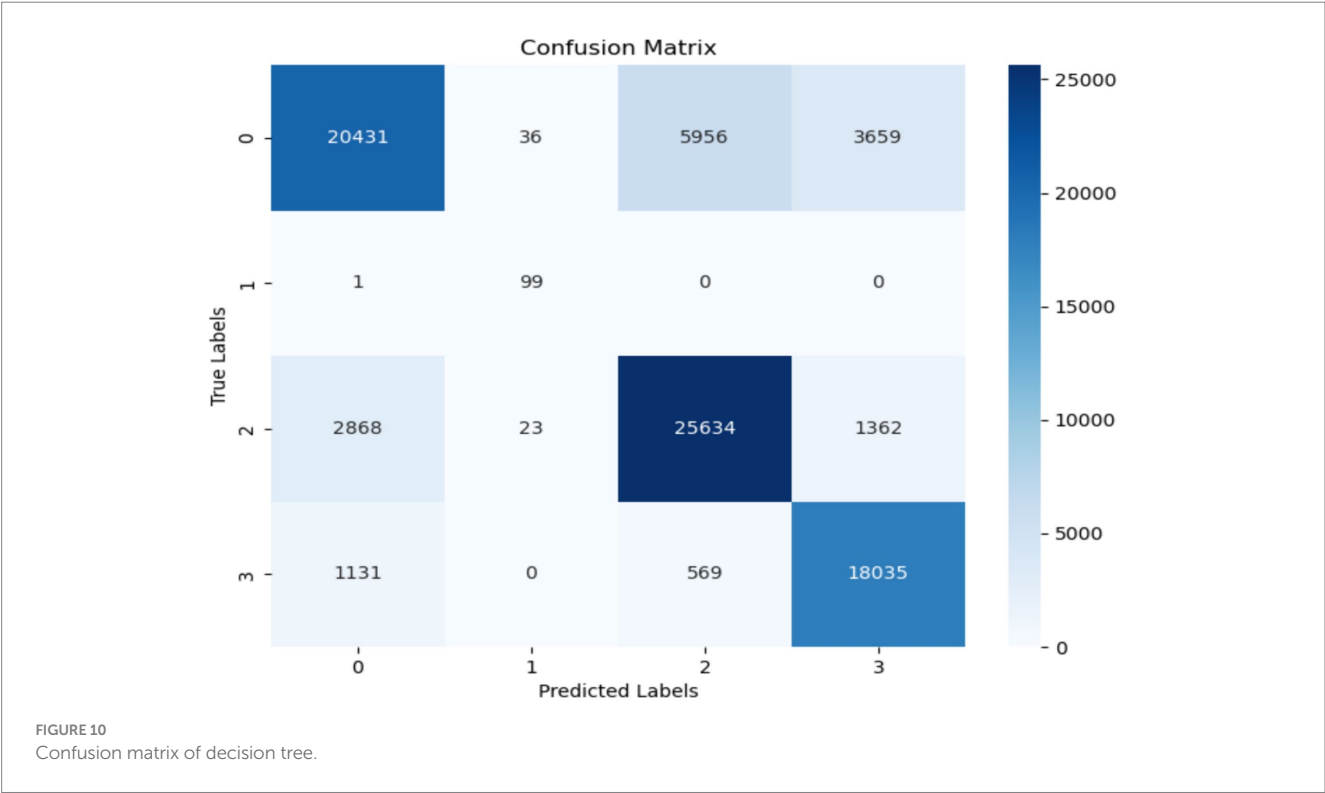


TABLE 9 Testing results of the DT model.

Model	Class	Precision %	Recall %	F1-score %	Accuracy %
RF	Normal	89	89	89	91
	startHesitation	97	99	94	
	Turn	90	90	90	
	Walking	96	97	95	
	Weighted	91	91	91	



4.3.2 Decision tree testing results

The experimental results while using the DT model demonstrated exceptional and excellent performance, notably in accurately categorizing the instances labeled as Turn. The model demonstrated exceptional precision (97%), recall (99%), F1-score (94%), and an overall accuracy of 91% for classes that existed in the clinical experimental dataset used. Although the model demonstrated strong accuracy and recall overall, it encountered difficulties in accurately detecting instances of the Turn class. This is evident from the poorer precision and recall scores specifically associated with this class. Table 9 summarizes the classification results based on the DT model.

The confusion matrix in Figure 10 displays the performance of the decision tree approach. The decision tree algorithm achieved a high accuracy of 91% throughout an evaluation stage. The program accurately classified 20,431 instances as normal. The misclassification of the class startHesitation is 2,868 instances more than that of the other classes, while the misclassification of the class Turn is only 1 instance.

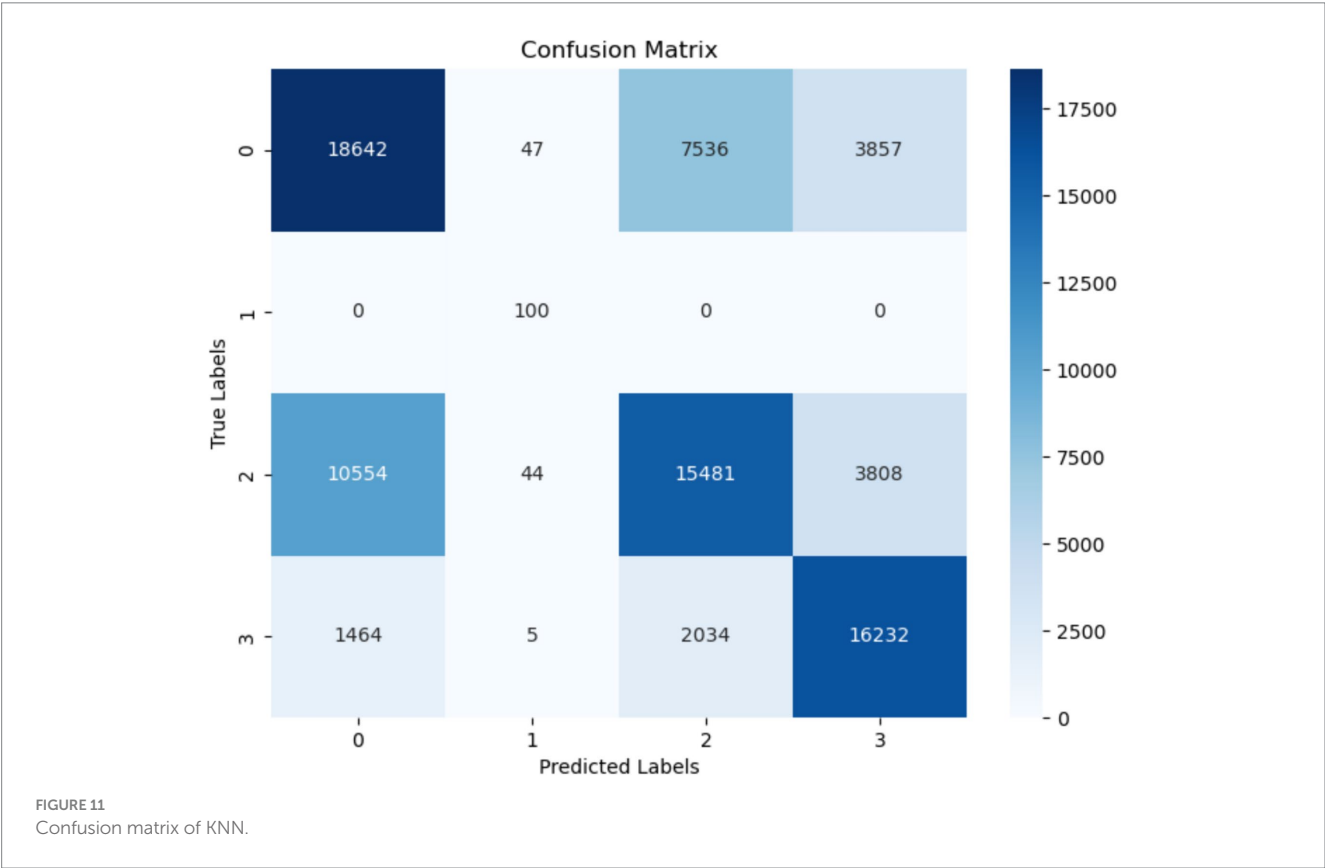
4.4 K-nearest neighbor’s classification results

The KNN model had excellent performance in accurately identifying instances belonging to the Walking class, achieving high precision (73%), recall (82%), f1-score (66%), and a total accuracy of 63%. Nevertheless, there were notable limitations in effectively classifying the Turn class samples, with both precision and recall scores being significantly noted in testing classification reports. Table 10 demonstrates the classification results based on the KNN model (Table 10).

The confusion matrix for the KNN model is displayed in Figure 11. The number of instances correctly predicted as “Normal” is 18,642, whereas there are no instances incorrectly predicted as “Turn.” However, the false positive rate is significantly high. The rate of false positives for the “startHesitation” class is particularly high, with a value of 10,554.

TABLE 10 Testing results of the KNN model.

Model	Class	Precision %	Recall %	F1-score %	Accuracy %
RF	Normal	58	53	64	63
	startHesitation	63	1.00	47	
	Turn	61	61	61	
	Walking	73	82	66	
	Weighted	63	63	63	



4.5 Classification results using the LightGBM model

This subsection presents the findings in detail of the classification results of the LightGBM model, which exhibited significant precision (84%), recall (91%), f1-score (78%), and overall accuracy (80%) in accurately categorizing the “Walking” cases. We faced complications in accurately identifying instances of the “Normal” category, leading to lower precision and recall scores. Table 11 displays the testing results of the LightGBM model.

Figure 12 displays the confusion matrix of the LightGBM model. It is worth noting that the misclassification (FP) rate for the “startHesitation” class is significantly high, with a total of 2,868 instances. The occurrence of false positives in the “Turn” class is extremely low, less than 1. The number of instances correctly classed as “Normal” and identified as negative is 20,431.

4.6 CatBoost model classification results

This section presents the results of the CatBoost model. The CatBoost algorithm exhibited remarkable precision (80%), recall (92%), f1-score (86%), and overall accuracy (82%) for the “Walking” class. Nevertheless, there were limitations in accurately categorizing cases that fell within the “startHesitation” class, leading to relatively low precision and recall ratings. Table 12 presents the testing and classification outcomes of the CatBoost model. The confusion matrix of CatBoost is presented Figure 13.

4.7 Results of GRU-transformers and LRCNN models

In this section GRU mode was combined with transformers model for classification FoG, we have used 200 hidden units for GRU model.

TABLE 11 Testing results of the LightGBM model.

Model	Class	Precision %	Recall %	F1-score %	Accuracy %
LightGBM	Normal	75	68	84	80
	startHesitation	77	99	63	
	Turn	83	86	80	
	Walking	84	91	78	
	Weighted	81	80	80	

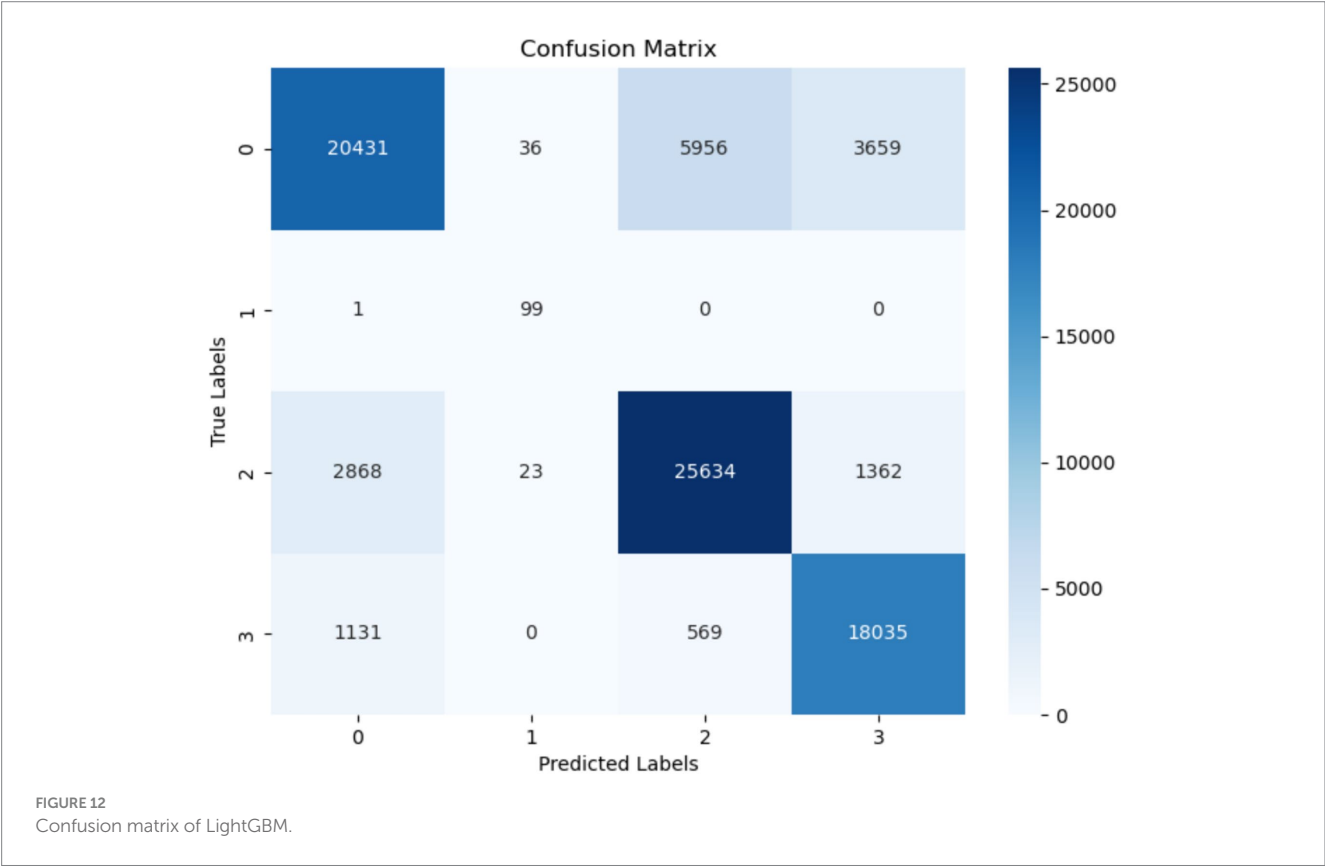


Table 13 shows the parameters of GRU-transformers and LRCNN models. It is noted that the accuracy of GRU-transformers and LRCN were achieved. It is investigated that the GRU-transformers and LRCN were better models for classification FoG.

The accuracy performance of the GRU-transformers is depicted in Figure 14. The GRU-transformers validation accuracy initially stood at 82% and then improved to 86% after 70 Epochs. The accuracy loss started from 0.43 and reached 0.32.

The performance and loss accuracy in the validation stages was calculated using the binary_crossentropy approach. The validation accuracy of the LRCN model is depicted in Figure 15. During the validation phase, the LRCN model exhibited started at 38% and reached to 86%. The accuracy loss is a decrease in accuracy loss from 0.42 to 0.35.

5 Results discussion

FoG is a motor disturbance categorized by an abrupt and fleeting inability to start or maintain walking, which poses difficulties for patients with PD. The timely identification and

predicting of FoG episodes are essential for efficient therapies and enhanced quality of life. The objective of this research was to evaluate the possibility of applying different machine-learning algorithms and GRU-transformers and LRCN models to predict FoG for a preventive strategy to mitigate the occurrence. In order to achieve this objective, random forest, k-nearest neighbor, LightGBM, and GRU-transformers and LRCN models algorithms were applied for detecting FoG.

The difficulties in classifying minority classes, specifically “startHesitation,” highlight the influence of imbalanced datasets on the effectiveness of models. Addressing these problems is essential in the context of FoG prediction to enable early detection of gait irregularities, facilitate prompt interventions, and enhance outcomes for individuals with PD. Therefore, we have applied an oversampling method for handling the imbalanced classes at the training phase. Figure 16 the relationship among features of the training dataset.

According to the experimental results obtained from using various machine-learning models, the decision tree model had a strong overall accuracy of 91% and an impressive f1-score of 0.96, particularly excelling in accurately categorizing “Normal” gait.

TABLE 12 Testing results of the CatBoost model.

Model	Class	Precision %	Recall %	F1-score %	Accuracy %
CatBoost	Normal	85	71	77	82
	startHesitation	27	1.00	42	
	Turn	81	86	84	
	Walking	80	92	86	
	Weighted	82	82	82	

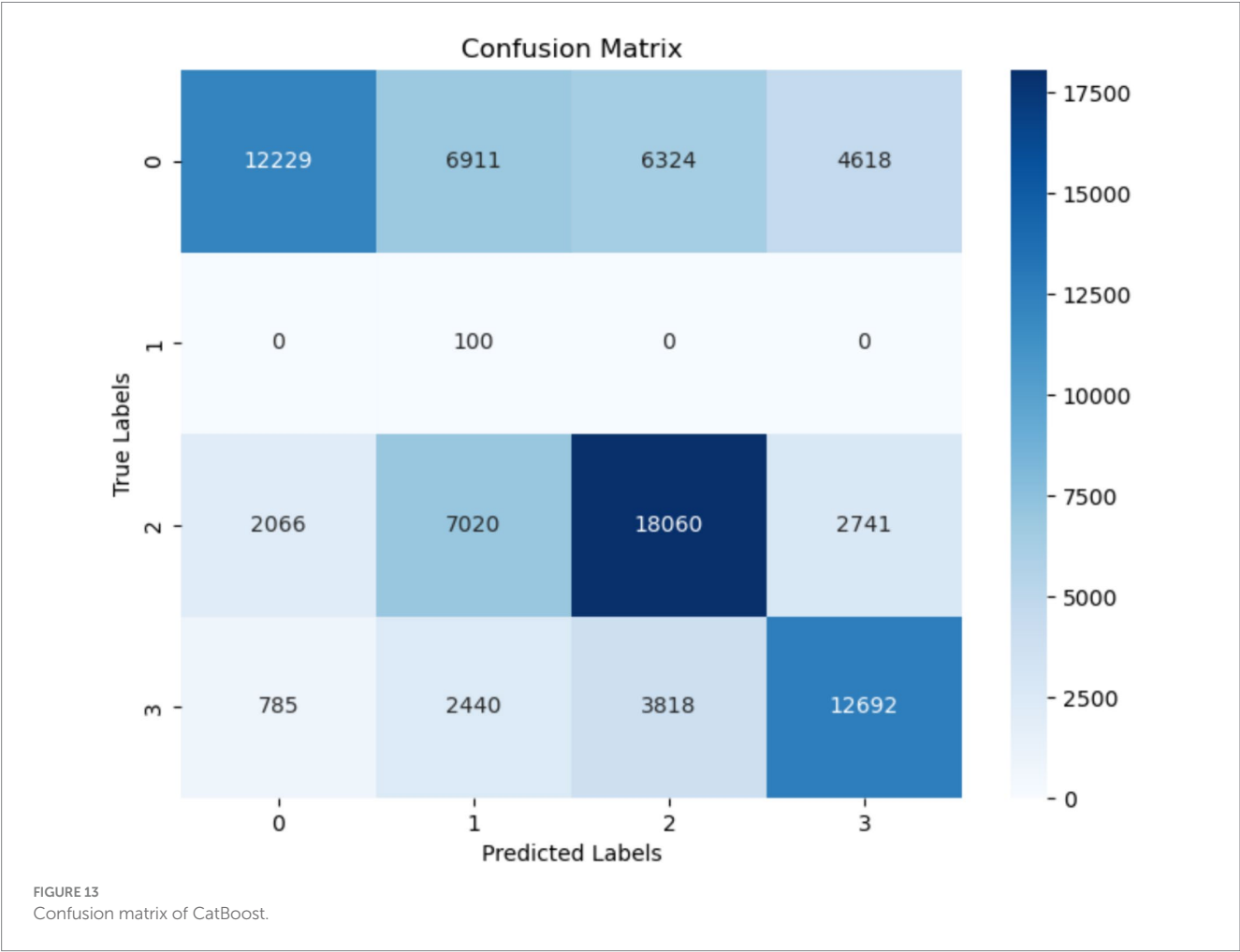


TABLE 13 Weight Avg. results of GRU-transformers and LRCNN model.

Models	Accuracy %	Precision%	Recall %	f1-score %
GRU-transformers	86	84	86	83
LRCNN	86	85	86	84

Nevertheless, the difficulties in precisely recognizing occurrences of “Turn” highlight the necessity of adjusting and optimizing the detection process to achieve a balance between accuracy and comprehensiveness. This is crucial for reliably identifying tiny irregularities in walking patterns that indicate FoG. The decision tree model demonstrated a notable weighted accuracy of 91% for all classes. The random forest approach scored a high accuracy (90%).

The KNN algorithm demonstrated a commendable level of accuracy (63%) and precision (63%). However, it is noted that the KNN achieved less accuracy compared with different existing ML approaches. Comprehending the influence of distance metrics and the quantity of neighbors is essential for enhancing its capacity to detect tiny variations linked to FoG. The LightGBM model showed potential, specifically in accuracy, attaining an accuracy of 80% and

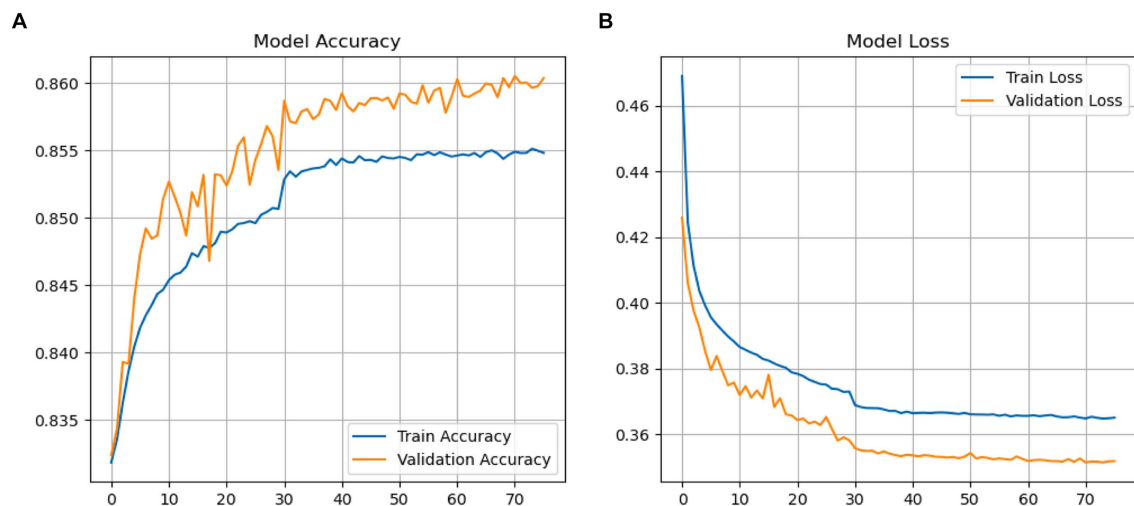


FIGURE 14
(A,B) Performance GRU-transformers.

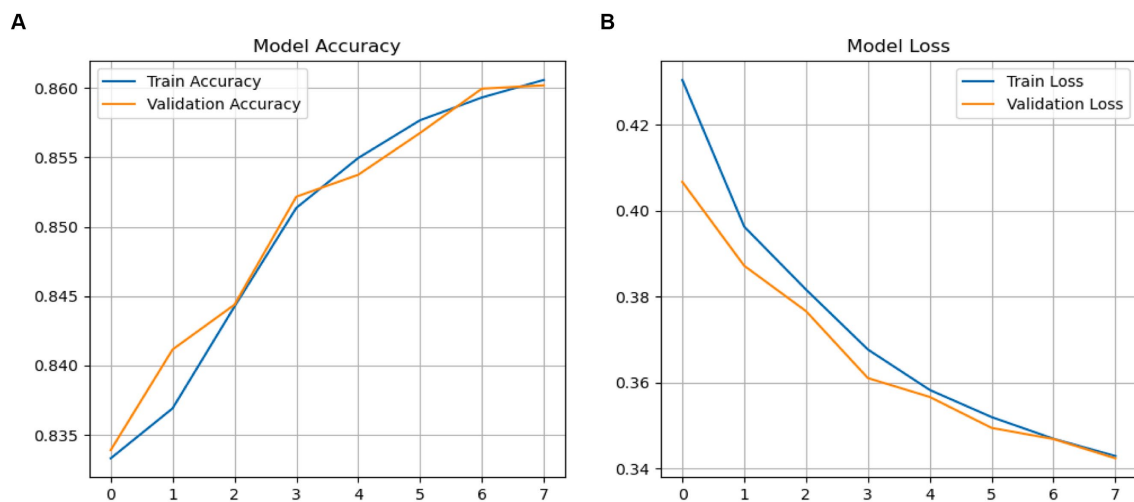


FIGURE 15
(A,B) Performance LRCN model.

an f1-score of 0.80. The CatBoost model demonstrated a strong precision of 0.82 and recall of 0.82, resulting in an accuracy of 82% and an f1-score of 82.

The ROC curve is a visual depiction that displays the performance of a classification algorithm at different levels of categorization. The graph depicts the relationship between two variables. The receiver operating characteristic (ROC) is computed using the following formula:

$$TRP = \frac{TP}{TP + FN} \quad (24)$$

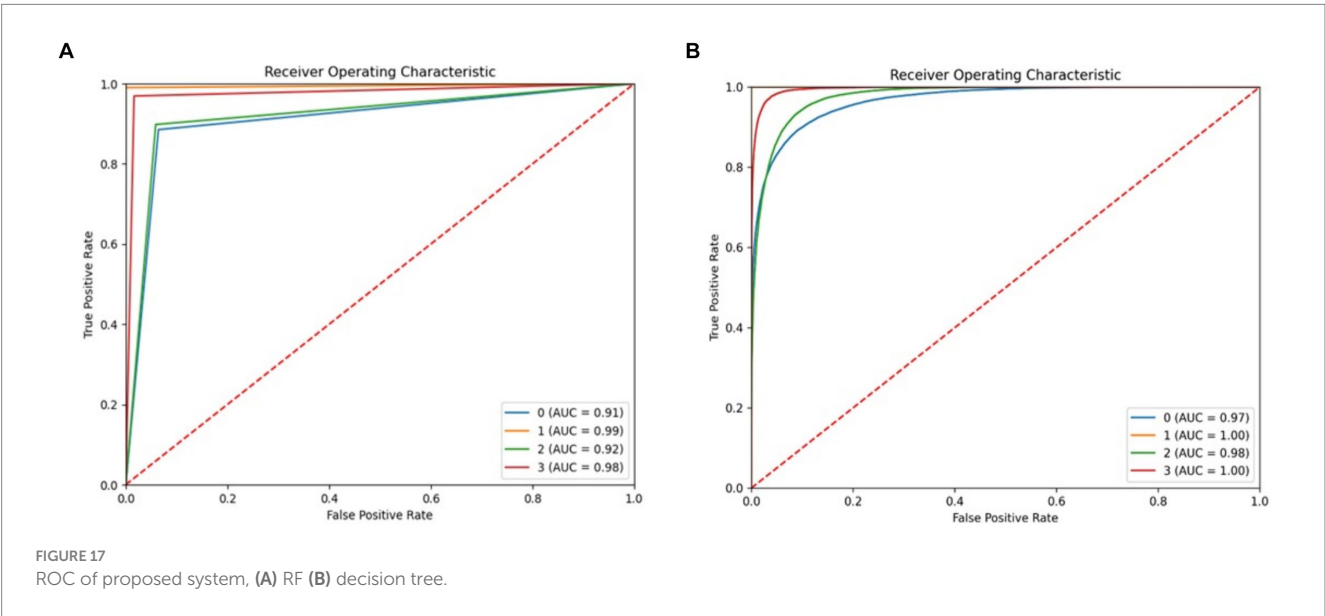
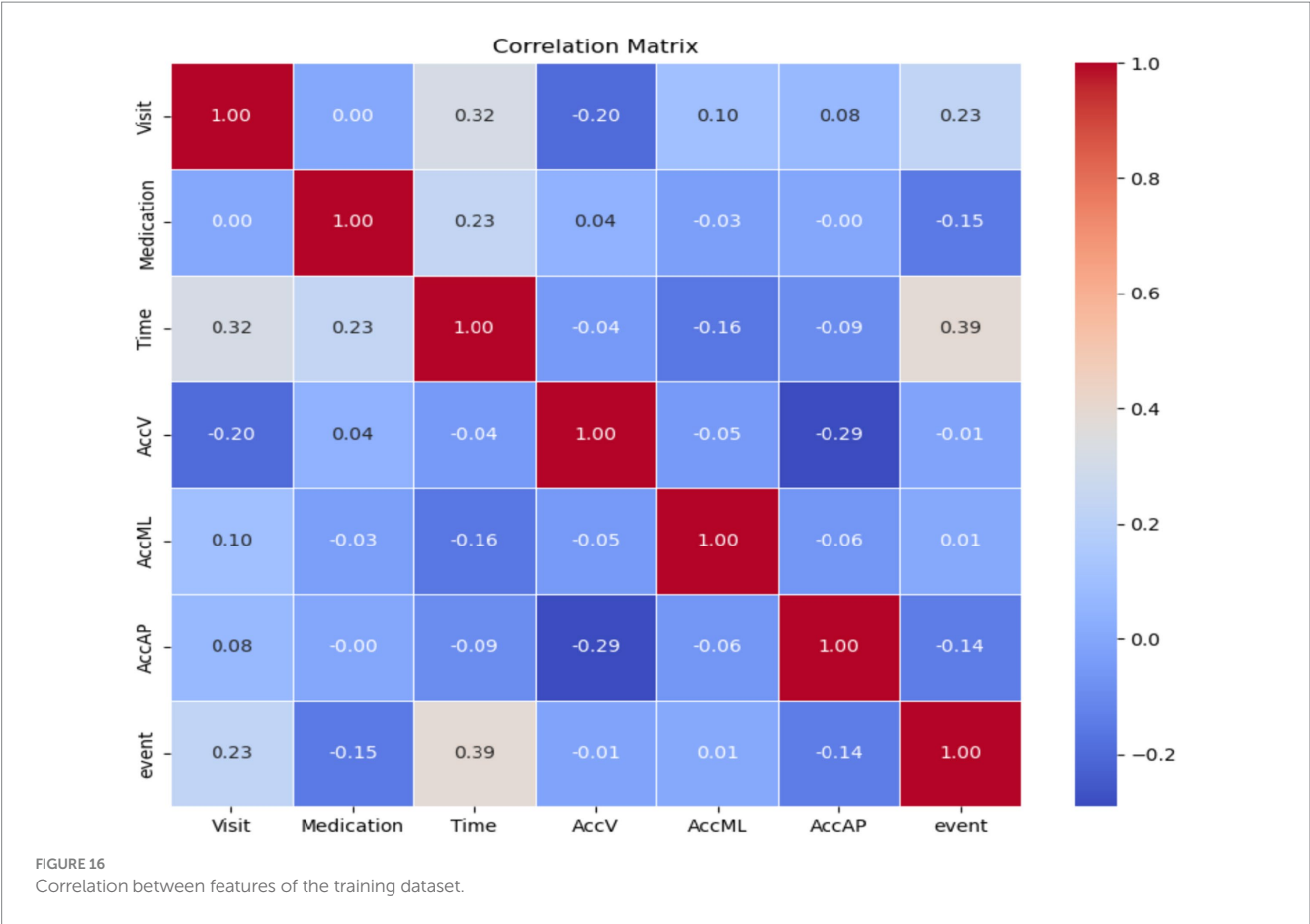
$$FPR = \frac{FP}{FP + TN} \quad (25)$$

Where TRP is the true positive rate and FPR is the false positive rate.

Figure 17 displays the ROC curve for both the DT and RF models. The DT model achieved a high ROC score of 99% for the “startHesitation” class and an ROC score of 98 for the “Walking” class. The receiver operating characteristic (ROC) analysis of the RF model yielded a high precision of 100% for the “startHesitation” class and an ROC value of 98% for the “Turn” class.

The LightGBM, and CatBoost algorithms scored less in accuracy. However, the ROC of the models are good, and the LightGBM, and CatBoost models scored ROC 100% in the “startHesitation” class. Figure 18 displays the ROC of LightGBM, and CatBoost models.

Figures 19, 20 displays ROC of GRU-transformers and LRCN models for predicting FoG. It is noted both models have achieved



similar performance, and GRU-transformers and LRCN were scored high percentage ROC=91 with class “Walking”.

Table 14 presents a comparison of the suggested algorithms, highlighting that the decision tree technique achieved a high accuracy rate of 91%.

6 Conclusion

FoG is a locomotive impairment observed in individuals with advanced PD, which has been linked to an elevated likelihood of falling and a worse overall quality of life. Freezing incidents can

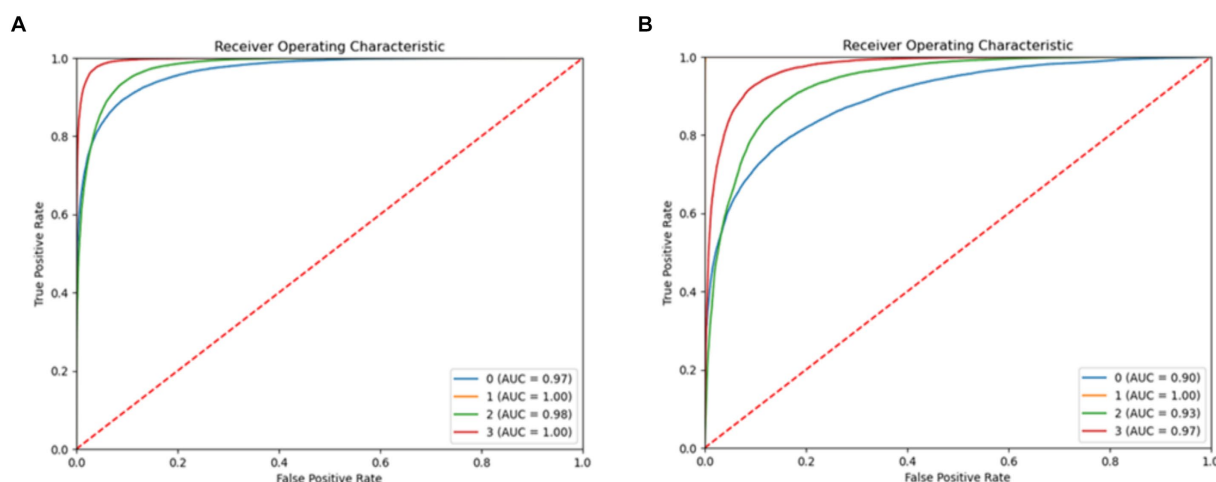


FIGURE 18
ROC of proposed system, (A) LightGBM (B) CatBoost.

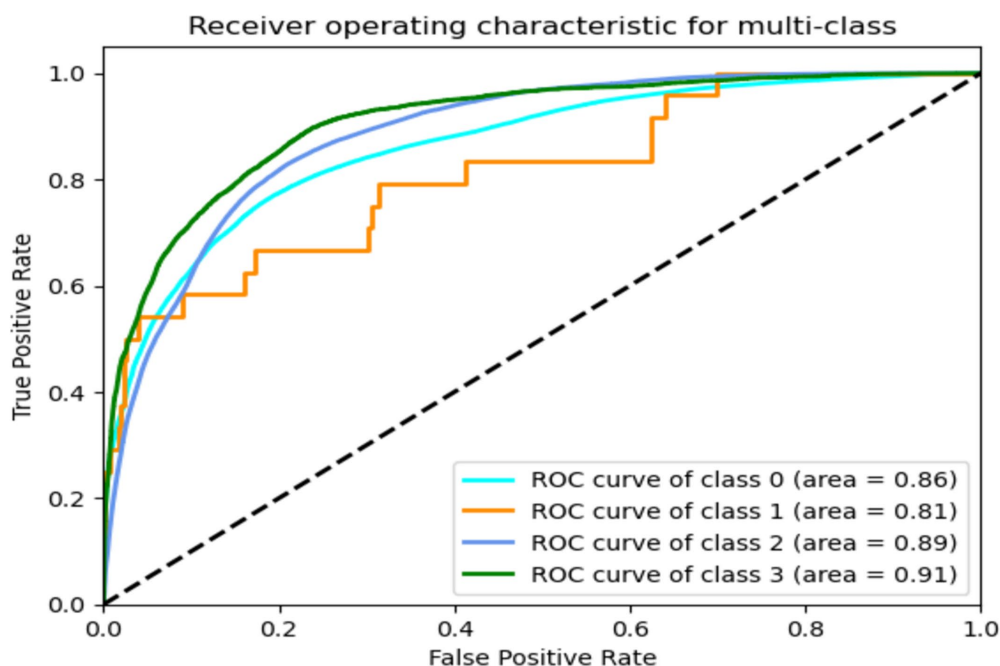


FIGURE 19
ROC of GRU-transformers.

be alleviated or averted through external intervention, such as the utilization of pictorial or auditory signals, which are triggered by FoG detection and prediction systems. The fundamental aim of this research work was predicting FoG using a standard dataset. This research concerted on the detection and prediction of FoG by analyzing 3D accelerometer data collected from the lower back of persons with PD, who frequently experience this terrible symptom. The dataset was obtained from a cohort of 65 participants. The dataset consists of four distinct classes: Normal, Turn, startHesitation, and Walking. Preprocessing techniques, such as cleaning the dataset and addressing imbalanced classes, were

implemented to enhance the performance of the ML methods. Various machine-learning algorithms, including decision tree, random forest, k-nearest neighbors, LightGBM, GRU-transformers and LRCN models, were employed for FoG detection and prediction. Of these, the decision tree algorithm exhibited a distinct advantage when working with datasets collected from sensors, achieving a high accuracy rate of 91%. This is the initial model employed for detecting FoG using this dataset. Furthermore, the main aim of this study also was to identify effective ML and DL algorithms that has capability of detecting and predicting FoG using a wearable system in real-time data.

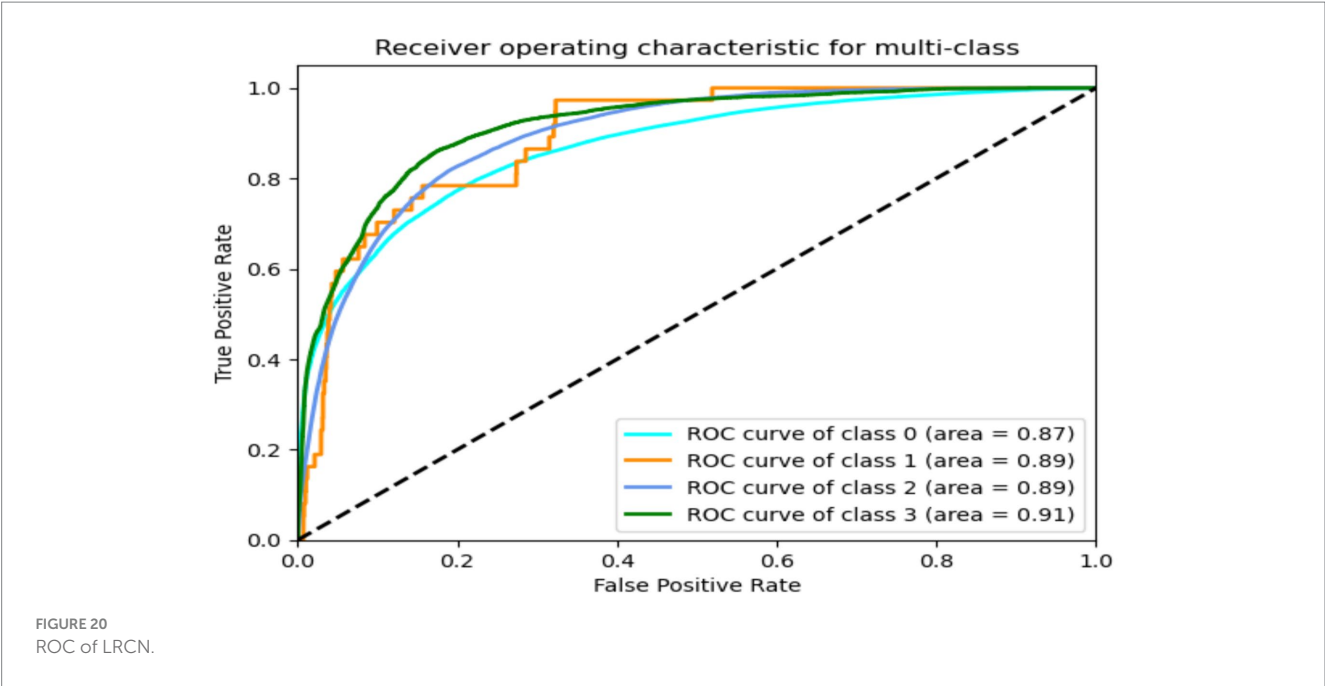


TABLE 14 Comparison results.

Models	Accuracy (%)
Random forest	90
Decision tree	91
KNN	63
LightGBM	83
CatBoos	80
GRU-transformers	86
GRU-transformers	86

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://kaggle.com/competitions/tlvmc-parkinsons-freezing-gait-prediction/data>.

Author contributions

AA-N: Writing – original draft, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. TA: Writing – review & editing, Software, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. NF: Writing – original draft, Visualization, Validation, Formal analysis, Data curation, Conceptualization. DK: Writing – review & editing, Visualization, Validation, Resources, Investigation, Funding acquisition, Formal analysis.

References

1. Harris JR. *Protein aggregation and Fibrillogenesis in cerebral and systemic amyloid disease*, vol. 65. Berlin, Heidelberg: Springer Science and Business Media (2012).

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. It was funded by King Salman center for disability through research group No: KSGR-2023-236.

Acknowledgments

The authors extend their appreciation to the king Salman center for disability research for funding this work through research group No: KSGR-2023-236.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

2. Nutt JG, Bloem BR, Giladi N, Hallett M, Horak FB, Nieuwboer A. Freezing of gait: moving forward on a mysterious clinical phenomenon. *Lancet Neurol.* (2011) 10:734–44. doi: 10.1016/S1474-4422(11)70143-0

3. Ehgoetz Martens KA, Pieruccini-Faria F, Almeida QJ. Could sensory mechanisms be a Core factor that underlies freezing of gait in Parkinson's disease? *PLoS One*. (2013) 8:e62602. doi: 10.1371/journal.pone.0062602
4. Tahafchi P, Judy JW. *Freezing-of-gait detection using wearable-sensor technology and neural-network classifier*. 2019 IEEE sensors, Montreal, QC, Canada. (2019), pp. 1–4.
5. Bloem BR, Hausdorff JM, Visser JE, Giladi N. Falls and freezing of gait in Parkinson's disease: a review of two interconnected, episodic phenomena. *Mov Disord*. (2004) 19:871–84. doi: 10.1002/mds.20115
6. Perez-Lloret S, Negre-Pages L, Damier P, Delval A, Derkinderen P, Destée A, et al. Prevalence, determinants, and effect on quality of life of freezing of gait in Parkinson disease. *JAMA Neurol*. (2014) 71:884–90. doi: 10.1001/jamaneurol.2014.753
7. Backer JH. The symptom experience of patients with Parkinson's disease. *J Neurosci Nurs*. (2006) 38:51–7. doi: 10.1097/01376517-200602000-00010
8. Giladi N, Nieuwboer A. Understanding and treating freezing of gait in parkinsonism, proposed working definition, and setting the stage. *Mov Disord*. (2008) 23:S423–5. doi: 10.1002/mds.21927
9. Schaafsma JD, Balash Y, Gurevich T, Bartels AL, Hausdorff JM, Giladi N. Characterization of freezing of gait subtypes and the response of each to levodopa in Parkinson's disease. *Eur J Neurol*. (2003) 10:391–8. doi: 10.1046/j.1468-1331.2003.00611.x
10. Giladi N, Treves T, Simon E, Shabtai H, Orlov Y, Kandinov B, et al. Freezing of gait in patients with advanced Parkinson's disease. *J Neural Transm*. (2001) 108:53–61. doi: 10.1007/s007020170096
11. Ehgoetz Martens KA, Ellard CG, Almeida QJ. Does anxiety cause freezing of gait in Parkinson's disease? *PLoS One*. (2014) 9:e106561. doi: 10.1371/journal.pone.0106561
12. Elkouzi A, Bit-Ivan EN, Elble RJ. Pure akinesia with gait freezing: a clinicopathologic study. *J Clin Mov Disord*. (2017) 4:15. doi: 10.1186/s40734-017-0063-1
13. Spildooren J, Vercruyse S, Desloovere K, Vandenbergh W, Kerckhofs E, Nieuwboer A. Freezing of gait in Parkinson's disease: the impact of dual-tasking and turning. *Mov Disord*. (2010) 25:2563–70. doi: 10.1002/mds.23327
14. Mancini M, Bloem BR, Horak FB, Lewis SJ, Nieuwboer A, Nonnekes J. Clinical and methodological challenges for assessing freezing of gait: future perspectives. *Mov Disord*. (2019) 34:783–90. doi: 10.1002/mds.27709
15. Peterson DS, King LA, Cohen RG, Horak FB. Cognitive contributions to freezing of gait in Parkinson disease: implications for physical rehabilitation. *Phys Ther*. (2016) 96:659–70. doi: 10.2522/ptj.20140603
16. Nieuwboer A, de Weert W, Dom R, Lesaffre E. A frequency and correlation analysis of motor deficits in Parkinson patients. *Disabil Rehabil*. (1998) 20:142–50. doi: 10.3109/09638289809166074
17. Kondo Y, Mizuno K, Bando K, Suzuki I, Nakamura T, Hashide S, et al. Measurement accuracy of freezing of gait scoring based on videos. *Front Hum Neurosci*. (2022) 16:828355. doi: 10.3389/fnhum.2022.828355
18. Okuma Y, Yanagisawa N. The clinical spectrum of freezing of gait in Parkinson's disease. *Mov Disord*. (2008) 23:426. doi: 10.1002/mds.21934
19. Giladi N. Freezing of gait. Clinical overview. *Adv Neurol*. (2001) 87:191–7.
20. Fahn S. The freezing phenomenon in parkinsonism. *Adv Neurol*. (1995) 67:53–63.
21. Moore ST, Macdougall HG, Ondo WG. Ambulatory monitoring of freezing of gait in Parkinson's disease. *J Neurosci Methods*. (2008) 167:340–8. doi: 10.1016/j.jneumeth.2007.08.023
22. Ahlrichs C, Samà A, Lawo M, Cabestany J, Rodríguez-Martín D, Pérez-López C, et al. Detecting freezing of gait with a tri-axial accelerometer in Parkinson's disease patients. *Med Biol Eng Comput*. (2016) 54:223–33. doi: 10.1007/s11517-015-1395-3
23. Kim H, Lee HJ, Lee W, Kwon S, Kim SK, Jeon HS, et al. *Unconstrained detection of freezing of gait in Parkinson's disease patients using smartphone*. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), Milano, Italy, 25–29 August 2015. (2015).
24. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature*. (2015) 521:436–44. doi: 10.1038/nature14539
25. Hassan M, Huda S, Uddin M, Almogren A, Alrubaian M. Human activity recognition from body sensor data using deep learning. *J Med Syst*. (2018) 42:1–8. doi: 10.1007/s10916-018-0948-z
26. Nweke HF, Teh YW, Al-Garadi M, Alo UR. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges. *Expert Syst Appl*. (2018) 105:233–61. doi: 10.1016/j.eswa.2018.03.056
27. Yang JB, Nguyen MN, San PP, Li XL, Krishnaswamy S. *Deep Convolutional Neural Networks On Multichannel Time Series For Human Activity Recognition*. In: Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI), Buenos Aires, Argentina, 25–31 July 2015; pp. 3995–4001. (2015).
28. Wang J, Chen Y, Hao S, Peng X, Hu L. Deep learning for sensor-based activity recognition: a survey. *Pattern Recog Lett*. (2019) 119:3–11. doi: 10.1016/j.patrec.2018.02.010
29. Ordóñez FJ, Roggen D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors*. (2016) 16:115. doi: 10.3390/s16010115
30. Kim HB, Lee HJ, Kim SK, Jeon HS, Park HY, Shin CW, et al. Validation of freezing-of-gait monitoring using smartphone. *Telemed E Health*. (2018) 24:899–907. doi: 10.1089/tmj.2017.0215
31. Masiala S, Huijbers W, Atzmueller M. *Feature-set-engineering for detecting freezing of gait in parkinson's disease using deep recurrent neural networks*. (2019)
32. Pepa L, Ciabattini L, Verdini F, Capecci M, Ceravolo M. *Smartphone based fuzzy logic freezing of gait detection in Parkinson's disease*. In: 2014 IEEE/ASME 10th international conference on mechatronic and embedded systems and Applications (MESA), IEEE, pp. 1–6. (2014).
33. Sigcha L, Costa N, Pavón I, Costa S, Arezes P, López JM, et al. Deep learning approaches for detecting freezing of gait in Parkinson's disease patients through on-body acceleration sensors. *Sensors*. (2020) 20:1895. doi: 10.3390/s20071895
34. Tahir A, Ahmad J, Shah SA, Morison G, Skelton DA, Larijani H, et al. WiFreeze: multiresolution scalograms for freezing of gait detection in Parkinson's leveraging 5G spectrum with deep learning. *Electronics*. (2019) 8:1433. doi: 10.3390/electronics8121433
35. Malik OA. Deep autoencoder for identification of abnormal gait patterns based on multimodal biosignals. *Int J Comput Digit Syst*. (2020) 10:2–8. doi: 10.12785/ijcds/100101
36. Cole BT, Roy SH, Nawab SH. *Detecting freezing-of-gait during unscripted and unconstrained activity*. In: Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 30 August–3 September 2011; pp. 5649–5652. (2011).
37. Nieuwboer A, Dom R, De Weert W, Desloovere K, Janssens L, Stijn V. Electromyographic profiles of gait prior to onset of freezing episodes in patients with Parkinson's disease. *Brain*. (2004) 127:1650–60. [doi: 10.1093/brain/awh189]
38. Handojoseno AA, Shine JM, Nguyen TN, Tran Y, Lewis SJ, Nguyen HT. *Using EEG spatial correlation, cross frequency energy, and wavelet coefficients for the prediction of Freezing of Gait in Parkinson's Disease patients*. In: Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013; pp. 4263–4266. (2013).
39. Delval A, Snijders AH, Weerdesteyn V, Duysens JE, Defebvre L, Giladi N, et al. Objective detection of subtle freezing of gait episodes in Parkinson's disease. *Mov Disord*. (2010) 25:1684–93. doi: 10.1002/mds.23159
40. Okuno R, Fujimoto S, Akazawa J, Yokoe M, Sakoda S, Akazawa K. Analysis of spatial correlation, plantar pressure pattern during gait in Parkinson's disease. In Proceedings of the 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, BC, Canada, 20–25 August 2008; pp. 1765–1768. (2008).
41. Bachlin M, Roggen D, Troste G, Plotnik M, Inbar N, Meidan I, et al. *Potentials of enhanced context awareness in wearable assistants for parkinson's disease patients with the freezing of gait syndrome*. In: 2009 international symposium on wearable computers, pp. IEEE, 123–130. (2009).
42. Djurić-Jović MD, Jović NS, Radovanović SM, Stanković ID, Popović MB, Kostić VS. Automatic identification and classification of freezing of gait episodes in Parkinson's disease patients. *IEEE Trans Neural Syst Rehabil Eng*. (2013) 22:685–94. doi: 10.1109/TNSRE.2013.2287241
43. Pham TT, Moore ST, Lewis SJG, Nguyen DN, Dutkiewicz E, Fuglevand AJ, et al. Freezing of gait detection in Parkinson's disease: a subject-independent detector using anomaly scores. *IEEE Trans Biomed Eng*. (2017) 64:2719–28. doi: 10.1109/TBME.2017.2665438
44. Zach H, Janssen AM, Snijders AH, Delval A, Ferraye MU, Auff E, et al. Identifying freezing of gait in parkinson's disease during freezing provoking tasks using waist-mounted accelerometry. *Parkinsonism Relat Disord*. (2015) 21:1362–6. doi: 10.1016/j.parkrel.2015.09.051
45. Rodríguez-Martín D, Samà A, Pérez-López C, Català A, Moreno Arosteigui JM, Cabestany J, et al. Home detection of freezing of gait using support vector machines through a single waist-worn triaxial accelerometer. *PLoS One*. (2017) 12:e0171764. doi: 10.1371/journal.pone.0171764
46. Hu K, Wang Z, Mei S, Ehgoetz Martens KA, Yao T, Lewis SJG, et al. Vision-based freezing of gait detection with anatomic directed graph representation. *IEEE J Biomed Health Inform*. (2019) 24:1215–25. doi: 10.1109/JBHI.2019.2923209
47. Masiala S, Huijbers W, Atzmueller M. Feature-set-engineering for detecting freezing of gait in Parkinson's disease using deep recurrent neural networks. *arXiv*. (2019) 2019:3428. doi: 10.48550/arXiv.1909.03428
48. Xia Y, Zhang J, Ye Q, Cheng N, Lu Y, Zhang D. Evaluation of deep convolutional neural networks for detection of freezing of gait in Parkinson's disease patients. *Biomed Signal Process Control*. (2018) 46:221–30. doi: 10.1016/j.bspc.2018.07.015
49. Shi B, Tay A, Au WL, Tan DML, Chia NSY, Yen SC. Detection of freezing of gait using convolutional neural networks and data from lower limb motion sensors. *IEEE Trans Biomed Eng*. (2022) 69:2256–67. doi: 10.1109/TBME.2022.3140258
50. Borzi L, Sigcha L, Rodríguez-Martín D, Olmo G. Real-time detection of freezing of gait in Parkinson's disease using multi-head convolutional neural networks and a single inertial sensor. *Artif Intell Med*. (2023) 135:102459. doi: 10.1016/j.artmed.2022.102459
51. Yang PK, Filtjens B, Ginis P, Goris M, Nieuwboer A, Gilat M, et al. Freezing of gait assessment with inertial measurement units and deep learning: effect of tasks, medication states, and stops. *J Neuroeng Rehabil*. (2024) 21:24. doi: 10.1186/s12984-024-01320-1

52. Aldhyani THH, Al-Nefaie AH, Koundal D. Modeling and diagnosis Parkinson disease by using hand drawing: deep learning model[J]. *AIMS Mathematics*. (2024) 9:6850–77. doi: 10.3934/math.2024334
53. Mukherjee M, Khushi M. SMOTE-ENC: a novel SMOTE-based method to generate synthetic data for nominal and continuous features. *Appl Syst Innov*. (2021) 4:18. doi: 10.3390/asi4010018
54. Joloudari JH, Marefat A, Nematollahi MA, Oyelere SS, Hussain S. Effective class-imbalance learning based on SMOTE and convolutional neural networks. *Appl Sci*. (2023) 13:4006. doi: 10.3390/app13064006
55. Zhou Y, Wang W, Wang K, Song J. Application of LightGBM algorithm in the initial Design of a Library in the cold area of China based on comprehensive performance. *Buildings*. (2022) 12:1309. doi: 10.3390/buildings12091309



OPEN ACCESS

EDITED BY

Surbhi Bhatia Khan,
University of Salford, United Kingdom

REVIEWED BY

Smitha Nayak,
University of Stirling, United Kingdom
Poonam Chaudhary,
The NorthCap University, India
Ramakrishna Kumar,
National University of Science and
Technology (Muscat), Oman
Abeer Saber,
Damietta University, Egypt

*CORRESPONDENCE

Surjeet Dalal
✉ profsurjeetdalal@gmail.com

RECEIVED 09 April 2024

ACCEPTED 27 May 2024

PUBLISHED 20 June 2024

CITATION

Darolia A, Chhillar RS, Alhussein M, Dalal S,
Aurangzeb K and Lilhore UK (2024) Enhanced
cardiovascular disease prediction through
self-improved Aquila optimized feature
selection in quantum neural network & LSTM
model.
Front. Med. 11:1414637.
doi: 10.3389/fmed.2024.1414637

COPYRIGHT

© 2024 Darolia, Chhillar, Alhussein, Dalal,
Aurangzeb and Lilhore. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Enhanced cardiovascular disease prediction through self-improved Aquila optimized feature selection in quantum neural network & LSTM model

Aman Darolia¹, Rajender Singh Chhillar¹, Musaed Alhussein²,
Surjeet Dalal^{3*}, Khursheed Aurangzeb² and
Umesh Kumar Lilhore⁴

¹Department of Computer Science and Applications, M.D. University, Rohtak, Haryana, India,

²Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia, ³Department of Computer Science and Engineering, Amity University, Gurgaon, Haryana, India, ⁴Department of Computer Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India

Introduction: Cardiovascular disease (CVD) stands as a pervasive catalyst for illness and mortality on a global scale, underscoring the imperative for sophisticated prediction methodologies within the ambit of healthcare data analysis. The vast volume of medical data available necessitates effective data mining techniques to extract valuable insights for decision-making and prediction. While machine learning algorithms are commonly employed for CVD diagnosis and prediction, the high dimensionality of datasets poses a performance challenge.

Methods: This research paper presents a novel hybrid model for predicting CVD, focusing on an optimal feature set. The proposed model encompasses four main stages namely: preprocessing, feature extraction, feature selection (FS), and classification. Initially, data preprocessing eliminates missing and duplicate values. Subsequently, feature extraction is performed to address dimensionality issues, utilizing measures such as central tendency, qualitative variation, degree of dispersion, and symmetrical uncertainty. FS is optimized using the self-improved Aquila optimization approach. Finally, a hybridized model combining long short-term memory and a quantum neural network is trained using the selected features. An algorithm is devised to optimize the LSTM model's weights. Performance evaluation of the proposed approach is conducted against existing models using specific performance measures.

Results: For dataset-1, accuracy-96.69%, sensitivity-96.62%, specificity-96.77%, precision-96.03%, recall-97.86%, F1-score-96.84%, MCC-96.37%, NPV-96.25%, FPR-3.2%, FNR-3.37% and for dataset-2, accuracy-95.54%, sensitivity-95.86%, specificity-94.51%, precision-96.03%, F1-score-96.94%, MCC-93.03%, NPV-94.66%, FPR-5.4%, FNR-4.1%. The findings of this study contribute to improved CVD prediction by utilizing an efficient hybrid model with an optimized feature set.

Discussion: We have proven that our method accurately predicts cardiovascular disease (CVD) with unmatched precision by conducting extensive experiments and validating our methodology on a large dataset of patient demographics and clinical factors. QNN and LSTM frameworks with Aquila feature tuning increase forecast accuracy and reveal cardiovascular risk-related physiological pathways. Our research shows how advanced computational tools may alter sickness

prediction and management, contributing to the emerging field of machine learning in healthcare. Our research used a revolutionary methodology and produced significant advances in cardiovascular disease prediction.

KEYWORDS

quantum neural network, cardiovascular disease prediction, feature extraction, self-improved Aquila optimization, machine learning

1 Introduction

CVD is a global health issue that kills many people. The WHO estimates a 37% mortality rate, affecting 17.9 million people (1). CVD deaths are mostly caused by stroke and heart disease. These frightening findings highlight the need to understand the complex causes of CVD. The complex nature of CVD, which is linked to risk factors like high blood pressure, insulin levels, smoking, and sedentary lifestyles, highlights the need for comprehensive prevention, early detection, and management strategies (2). Understanding these risk variables is essential for establishing targeted therapies and reducing the global effect of cardiovascular health issues as researchers study CVD (3). Studies show that up to 90% of CVD cases are avoidable, but early detection, treatment, and recovery are crucial. Early CVD detection is essential for timely interventions. However, CVD prediction is too sophisticated for the brain. Time dependency, erroneous results, and knowledge upgradation due to vast CVD datasets complicate identification (4). These datasets typically have irrelevant and redundant features that hamper classification. Noise from unwanted features affects system performance. Addressing this, our research focuses on FS to eliminate unwanted features before applying classification approaches. This process enhances model simplification, reduces the risk of overfitting, and improves computational efficiency (5).

Traditional diagnosis heavily relies on clinical signs and symptoms, making disease analysis challenging. Predicting CVD is particularly complex due to multiple contributing factors, leading to inconsistent outcomes and assumptions. In the medical domain, data mining (DM) methods, especially ML techniques (6), are employed to analyze diseases like cancer, stroke, diabetes (7), and CVD. This research specifically utilizes advanced DM approaches for studying CVD. Also, some more accurate DM approaches are being used to study heart disease. Researchers have applied various DM systems such as support vector machines (SVM), decision trees (DT), and artificial neural networks (ANN) to identify CVD (8). With all of the above methods, patient records are continuously categorized and predicted. It continuously checks the patient's movements and informs the patient and doctor of the risk of illness if there is a change. With the help of techniques like ML, doctors can easily detect CVD in the early stage itself. Amongst the traditional invasive-based method, angiography is represented as the well-known heart problem diagnosis method but, it has some limits. Conversely, a method such as intelligent learning-based computational approaches, non-invasive-based techniques is considered more effective for predicting CVD. Cardiovascular disease (CVD), one of the leading causes of death worldwide, causes much morbidity and death. Early detection and prediction are essential to prevent CVD and reduce its impact on individuals and healthcare systems. Medical advances in machine

learning and predictive analytics have created promising new opportunities for early cardiovascular disease risk factor diagnosis (9).

Predicting cardiovascular disease is crucial due to its incidence and damage. High-risk patients can be identified, advised on lifestyle changes, and prevented from developing cardiovascular disease (CVD). Genetic and risk factor-based predictive diagnostics provide individualized healthcare and tailored medicines. Traditional risk assessment and advanced machine learning algorithms predict cardiovascular disease. Traditional risk assessments like the Framingham Risk Score and Reynolds Risk Score use demographic, clinical, and biochemical data to estimate CVD risk across time. These techniques have directed primary preventive initiatives by identifying high-risk populations. Machine learning algorithms' ability to search massive data sets for detailed patterns has propelled their rise in cardiovascular disease prediction. More accurate and powerful predictive models have been constructed combining electronic health records, imaging data, genetic information, and lifestyle factors using supervised learning approaches such as logistic regression, support vector machines, random forests, and neural networks. Before predictive analytics can fully forecast cardiovascular illness, many challenges must be overcome. Multiple data sources, such as genetic data, wearable sensor data, and socioeconomic characteristics, make cohesive prediction models difficult. Integrating all these data types while maintaining privacy, interoperability, and quality is still difficult. When clinical decision-making is crucial, machine learning model interpretability is a concern. Black-box algorithms can produce accurate predictions, but healthcare practitioners are wary of them since they do not expose their inner workings. Because cardiovascular disease risk changes, models must be developed and validated for varied populations and healthcare systems (10).

Future multidisciplinary teams of medics, data scientists, and AI professionals will improve cardiovascular disease prediction. Integrating data from microbiomics, proteomics, metabolomics, and genomes may lead to new cardiovascular risk biomarkers and better risk prediction models. Wearables, smartphone health apps, and remote monitoring systems enable real-time risk assessment and personalized treatments based on lifestyle and physiological parameters. Here, a Hybrid Intelligent Model with an Optimal Feature Set is introduced for the prediction of CVD. The main contributions are summarized below:

1. The proposed research addresses the issue of dimensionality reduction by implementing FS techniques to reduce the number of features.
2. To introduce the SIAO method for optimal FS, overcoming challenges in extensive CVD datasets.
3. Proposing a hybrid model that combines LSTM and QNN to enhance the prediction performance of CVD.

The subsequent sections follow a structured framework: Section 2 reviews conventional CVD prediction models. In Section 3, the proposed model architecture is presented, and discussions on feature extraction, central tendency, dispersion, qualitative variation, and symmetrical uncertainty are provided. Section 4 introduces SIAO for optimal FS. The hybrid LSTM-QNN classification method is covered in Section 5. Experimental results and discussions are presented in Section 6. Section 7 contains the conclusion, summarizing contributions, and suggesting future research.

2 Literature review

This section critically analyses CVD prediction approaches, highlighting significant research and their contributions to the discipline. Using an Improved Quantum CNN (IQCNN) for accuracy, Pitchal et al. (11) developed an automated model for heart disease prediction that includes preprocessing, feature extraction, and prediction. This technique, which surpassed Bi-LSTM and CNN with 0.91 accuracy, shows promise for using IoT technologies for health diagnosis. Innovative computer methods improve cardiac disease prediction in their work.

Li et al. (12) used a hybrid deep learning (DL) model to predict CVD. The hybrid model, which uses 7,291 patient data and two deep neural network (DNN) models and one RNN for training, outperformed standard methods in prediction accuracy. Secondary training with a kNN model improved predicted accuracy. Prediction accuracy of 82.8%, precision of 87.08%, recall of 88.57%, and F1-score of 87.82% in the test set outperform single-model ML predictions. The hybrid model reduced overfitting, improving CAD prediction and clinical diagnosis. Singh et al. (13) examined how IoMT devices transformed continuous CVD patient monitoring. Their study proposed an advanced DL framework for the IoMT ecosystem that could improve patient care by predicting CVD. They effectively extract spatial and sequential characteristics from diverse IoMT data sources, such as pulse oximeters and electrocardiograms, using their innovative hybrid CNN-RNN architecture. With the utilization of transfer learning (TL) and real-world data, the proposed model surpasses previous methods in terms of precision and resilience. Their research assists medical professionals in gaining insights into predictive factors, enhancing the model's ability to be understood and its impact on therapy.

In their study, Oyewola et al. (14) utilized an ensemble optimization DL method to diagnose early CVD. They employed the Kaggle Cardiovascular Dataset for both training and testing purposes. The ensemble model achieves superior performance compared to neural network architectures, boasting an impressive accuracy rate of 98.45%. The research examined and provided a practical solution to streamline CVD diagnosis for doctors. It showcased the model's impressive speed and precision in identifying patients and interpreting CVD test results, leading to advancements in healthcare practices. Incorporating wearable systems, exploring advanced ensemble techniques, and utilizing diverse data sources have been found to enhance predictive capabilities and improve model performance in real-world healthcare settings, according to recent research. In 2023, a team of researchers developed a cutting-edge model for assessing the risk of cardiovascular disease (CVD). They utilized advanced algorithms and optimization strategies to create the SOLSSA-CatBoost

model, which shows great promise in this field. Their approach proved to be highly effective, surpassing the performance of multiple machine learning models and optimization techniques on Kaggle CVD datasets. They achieved impressive F1-scores of 90 and 81.51%. This work contributes to the field of predictive healthcare by offering a more precise tool for assessing the risk of cardiovascular disease. However, further research is required to evaluate its practicality and effectiveness in diverse populations.

In their study, Palanivel et al. (15) discussed the global health concern of cardiovascular disease (CVD) and emphasized the importance of early prediction. They presented a compelling approach that combines FS and an innovative Multi-Layer Perceptron (MLP) for Enhanced Brownian Motion based on Dragonfly Algorithm (MLP-EBMDA) classification using DM methods. This contribution encompasses an optimized unsupervised feature selection technique, a distinctive classification model with an accuracy of 94.28%, and a methodical approach to predicting early cardiovascular disease. The methodology is meticulously organized and precise, but it requires validation and real-world implementation.

In their study, Yewale et al. (16) devised a comprehensive framework for predicting cardiovascular disease. They made a deliberate choice to exclude FS and instead focused on data balance and outlier identification. Their work involved utilizing the Cleveland dataset to investigate various performance factors and achieve an impressive accuracy rate of 98.73% and sensitivity rate of 98%, surpassing previous research findings. The methodology demonstrates an impressive level of precision, with a specificity of 100%, positive prediction value of 100%, and negative prediction value of 97%. It also implemented OD by using a separate forest for a thorough analysis. Their work is notable for its meticulous evaluation metrics.

In their study, Behera et al. (17) devised a novel approach combining machine learning techniques to predict heart and liver diseases. They utilized a modified particle swarm optimization (PSO) algorithm in conjunction with support vector machines (SVM). The study focused on the rising occurrence of heart and liver disorders and the importance of promptly detecting them for better patient outcomes. By integrating SVM with modified PSO, the hybrid model achieved significant improvements in classification accuracy, error reduction, recall, and F1-score. The research's empirical foundation is strengthened by the data from the UCI ML collection. In their study, Sudha and Kumar (18) proposed a hybrid CNN and LSTM network for predicting cardiovascular disease, aiming to tackle the pressing issue of timely and accurate detection on a global scale. Utilizing cutting-edge DL advancements, the suggested model seamlessly combined CNN and LSTM to surpass the accuracy limitations of traditional ML methods. The hybrid system achieved an accuracy of 89% on a CVD dataset following 10 k-fold cross-validation. The suggested analysis outperformed SVM, Naïve Bayes (NB), and DT models in terms of performance. Their approach stands out with its distinctive technique, impressive precision, and practicality as an alternative to ML models in predicting CVD.

Elavarasi et al. (19), provided a summary of the recent challenges in predicting cardiovascular disease (CVD), focusing on the issues faced by traditional systems and the complexity of deep learning (DL). They utilized the elephant search algorithm (ESA) to explore innovative interpretability solutions during their investigation. ESA is seamlessly integrated with SVM to enhance the accuracy of CVD prediction, even though it faces challenges

when dealing with large datasets and computational complexity. They strive to enhance FS by enhancing the accuracy and interpretability of CVD dataset. Their research enhanced clinical decision support systems (DSSs), shedding light on the ongoing debate surrounding CVD prediction methodologies.

Table 1 summarizes standard CVD prediction models' features and drawbacks. An Automated IQCNN Model improved heart disease prediction and IoT diagnostics (11), however dataset specificity and scalability were issues. Wei et al. (20)'s SOLSSA-CatBoost Model improved CVD risk assessment accuracy through algorithmic fusion, however real-world applicability was questioned. In Palanivel et al. (15), the MLP-EBMDA classification model showcased optimized unsupervised FS, a novel classification model with higher accuracy, and a systematic approach to early CVD

prediction. Li et al. (12), proposed a hybrid DL model with features like the utilization of two DNN models and an RNN, achieving average accuracy and effectively addressing overfitting challenges. Singh et al. (13), introduced an IoMT-Enhanced DL framework, incorporating a hybrid architecture combining CNNs and RNNs, extracting spatial and sequential features from heterogeneous IoMT data sources, and emphasizing interpretability and impact on treatment processes. Oyewola et al. (14) proposed an ensemble optimization DL technique that stands out for outperforming various NN architectures with high accuracy and simplifying CVD diagnosis for medical professionals. Elavarasi et al. (19) presented an ESA-integrated SVM for CVD prediction, focusing on interpretability through FS and optimizing FS using ESA and SVM while addressing challenges associated with traditional systems. Yewale et al. (16),

TABLE 1 Review of features and challenges of conventional models based on a prediction of CVD.

References	Deployed model	Features	Challenges
Pitchal et al. (11)	Automated IQCNN Model	Incorporates preprocessing, feature extraction, and prediction with IQCNN; Notable high accuracy level; Advances IoT use in health diagnostics	Reliance on specific datasets; Scalability in diverse healthcare settings
Li et al. (12)	Hybrid DL model	Utilizes two DNN models and an RNN; Achieved Average accuracy, precision, recall, and F1-score	Effectively addresses overfitting challenges
Singh et al. (13)	IoMT-Enhanced DL framework	Hybrid architecture combining CNNs and RNNs; Extracts spatial and sequential features from heterogeneous IoMT data sources; Incorporates TL and real-world data	Ensuring interpretability and impact on treatment processes
Oyewola et al. (14)	Ensemble Optimization DL technique	Outperforms various NN architectures with higher accuracy; Simplifies CVD diagnosis for medical professionals	Use of sophisticated ensemble techniques
Palanivel et al. (15)	MLP-EBMDA classification	Optimized unsupervised FS; Novel classification model with high accuracy; Systematic approach to early CVD prediction	Use more datasets to get accurate results.
Yewale et al. (16)	Ensemble techniques with data balancing and OD	Achieves High accuracy and sensitivity; Demonstrates High specificity and positive prediction value	Need to use a diverse composition of metrics.
Behera et al. (17)	Hybrid ML algorithm with PSO and SVM	Utilizes modified particle swarm optimization and SVM; Showcases Average classification accuracy and error reduction	Need to investigate the runtime complexity.
Sudha and Kumar (18)	Hybrid CNN and LSTM Network	Combines CNNs with LSTM networks for CVD prediction; Achieves High accuracy validated through k-fold cross-validation	Apply the hybrid approach to real-world applications
Wei et al. (20)	SOLSSA-CatBoost Model	Integrates improved SSA with CatBoost; Enhanced by salp swarm algorithm, OBL, and lateral mutation; Superior F1-scores	Real-world applicability and diverse population performance
Elavarasi et al. (19)	ESA-integrated SVM	Addresses challenges with traditional systems; Focuses on interpretability through FS; Optimizes FS using ESA and SVM	Handling large datasets and computational complexity

ensemble techniques with data balancing and OD achieved higher accuracy and sensitivity, demonstrating high specificity and positive prediction value, although a need for a diverse composition of metrics was identified. Behera et al. (17) proposed a hybrid ML algorithm incorporating PSO and SVM showcased the utilization of modified PSO-SVM, resulting in average classification accuracy and error reduction, with a call to investigate runtime complexity. Finally, Sudha and Kumar (18) proposed a hybrid CNN and LSTM CVD prediction approach with great accuracy proven by 10 k-fold cross-validation and recommended for real-world applications. These systematic reviews shed light on these models' strengths and weaknesses, leading to CVD prediction methodology development between paragraphs belonging to the same section.

In essence, our proposed model, as outlined, integrates the strengths of DL, and bio-inspired algorithms techniques while systematically addressing the limitations identified in the existing approaches. The innovative features of our model, including optimized FS through SIAO and the hybridization of LSTM and QNN, contribute to its potential to provide enhanced accuracy, efficiency, and practical applicability in real-world CVD prediction scenarios.

3 Methodology

The hybrid model averages LSTM and QNN classifier outputs to predict. The SIAO method optimizes LSTM weight adjustment, improving prediction model accuracy. Thus, CVD prediction works. As shown in Figure 1, CVD prediction involves four key steps: preprocessing, feature extraction, FS, and prediction.

- **Step 1: Preprocessing** – The initial stage removes duplicates and missing data to ensure data quality and dependability for analysis.
- **Step 2: Feature Extraction** – This phase involves detailed feature extraction. Central tendency, qualitative variation, dispersion, and symmetrical uncertainty are identified. These attributes help solve the dataset's high dimensionality problem.
- **Step 3: Feature Selection** – The Symmetrical Uncertainty-based Iterative Algorithm Optimization (SIAO) technique is used to choose features optimally. This smart selection procedure improves model efficiency and accuracy by using only the most important features.
- **Step 4: CVD Prediction** – A hybrid model combining LSTM and QNN technology is trained using ideally selected features. This stage optimizes LSTM model weights using the SIAO algorithm. This optimization technique improves the model's predictive power.

The proposed CVD prediction approach is shown in Figure 1.

3.1 Optimal selection of features via self-improved Aquila optimization

The extracted features pose challenges related to dimensionality reduction, prompting the utilization of an SIAO Algorithm for optimal FS in this research endeavor.

3.1.1 SIAO algorithm

In 2021, Abualigah et al. (1) proposed an Aquila Optimization (AO), which is a modern swarm intelligence (SI) algorithm. Aquila consists of 4 types of hunting behaviors for specific sorts of prey. Aquila adeptly adapts hunting strategies for specific prey, utilizing its rapid velocity and robust talons; correspondingly, the AO Algorithm comprises four intricately designed stages as follows:

Expanded Exploration (X_1): Excessive ascend with a vertical bend. Eq. 1 and Eq. 2 define the mathematical expression of expended exploration of AO, in which Aquila flies excessively over the floor and explores the quest area widely, and then a vertical dive can be taken as soon as the Aquila identifies the prey's location.

$$X_1(t+1) = X_{best}(t) \cdot \left(1 - \frac{1}{t}\right) + (X_{M(t)} - X_{best}(t)) \cdot rand \quad (1)$$

$$X_M(t) = \frac{1}{N} \sum_{i=1}^N X_i(t) \quad (2)$$

Where better position attained was represented as $X_{best}(t)$ and $X_{M(t)}$ represents the mean position of Aquila in the present iteration. t denoted as the current iteration and the T represents the maximum iteration. The size of the population is mentioned as N and a random number (between 0 and 1) is indicated as $rand$.

Narrowed Exploration (X_2): Outline flight with the brief skim attack. Narrowed exploration is one of the frequently used hunting approaches for Aquila Employing brief gliding maneuvers for targeted prey attacks, the AO Algorithm elegantly combines sliding within the selected area and precise aerial navigation around the prey, with the refined exploration process succinctly defined by Eq. 3.

$$X_2(t+1) = X_{best}(t) \cdot LF(D) + X_{R(t)} + (y - x) \cdot rand \quad (3)$$

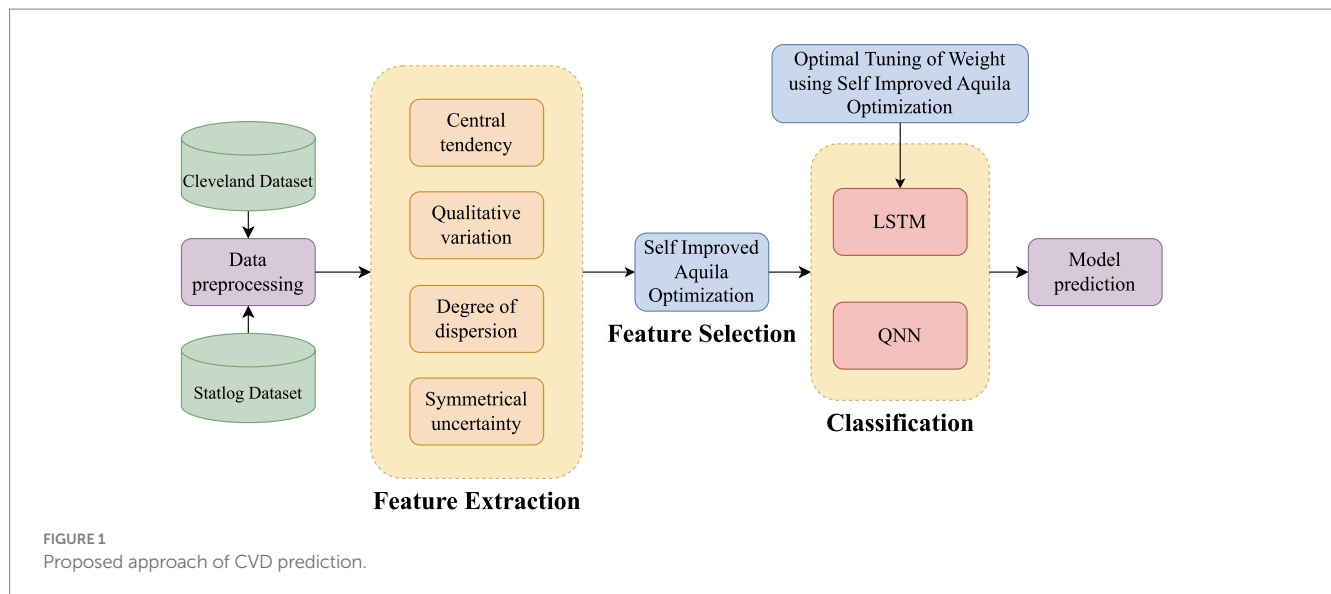
Where Hawks' random position is indicated as $X_{R(t)}$, and the size of a dimension is denoted as D . Function of Levy flight $LF(D)$, is expressed in below Eq. 4a and Eq. 4b.

$$LF(D) = s \times \frac{u \times \sigma}{|v|^{\frac{1}{K}}} \quad (4a)$$

$$\sigma = \left[\frac{\Gamma(1+K) \times \sin\left(\frac{\pi K}{2}\right)}{\Gamma\left(\frac{1+K}{2}\right) \cdot K \cdot 2^{\left(\frac{K-1}{2}\right)}} \right] \quad (4b)$$

Where Γ and K means stable values equivalent to 0.01 & 1.5; u and v denote random values between 0 & 1; y and x represent the spiral shape in the search. These values are mathematically calculated as follows (See Eq. 5):

$$x = r \times \sin \varnothing$$



$$y = r \times \cos \varnothing$$

$$r = r1 + 0.00565 \times D_1$$

$$\varnothing = -\omega \times D_1 + \frac{3 \times \pi}{2} \quad (5)$$

Where, the search cycle number is represented as $r1$, which exists between the range of 1 and 20, the value of ω is equal to 0.005. Also, D_1 is mentioned as the integer values and D indicates the size of the dimensions.

Extended Exploitation (X_3): Executing the minimal flight strategy with a calibrated descent attack, the Aquila adopts a nuanced approach. In this tactical maneuver, the prey's location is approximately ascertained, prompting the Aquila to initiate a vertical assault. The AO algorithm strategically capitalizes on the identified region, meticulously navigating closer to the prey before launching the attack. This intricate behavior is mathematically articulated in Eq. 6.

$$X_3(t+1) = X_{best}(t) - X_M(t) \cdot \alpha - r \text{ and } ((UB - LB) \cdot rand + LB) \cdot \varnothing \quad (6)$$

The parameters of the exploitation adjustment are assigned a value of 0.1 in this context. UB and LB are boundary values. In this, we have proposed Eq. 7 for choosing a random number between 0 and 1, which is calculated using a logistics map. The mathematical expression for the random value is given in Eq. 7.

$$rand = LB + rand(0,1) \times (UB - LB) \quad (7)$$

Subsequently, the arithmetic crossover is performed, in which two regions are randomly selected, and by performing linear combination 2 offspring are produced.

Narrowed Exploitation (X_4): Executing a strategy involving pursuit and ground-based assault, the Aquila pursues prey, following the trajectory of its escape, culminating in an attack on the ground, as mathematically articulated in Eq. 8–11.

$$X_4(t+1) = QF \times X_{best}(t) - (G_1 \cdot X(t) \cdot rand - G_2 \cdot LF(D) + rand \times G_1) \quad (8)$$

$$QF(t) = t^{\frac{2 \times rand - 1}{(1-T)^2}} \quad (9)$$

$$G_1 = 2 \times rand - 1 \quad (10)$$

$$G_2 = 2 \times \left(1 - \frac{t}{T}\right) \quad (11)$$

Where a current position is denoted as $X(t)$, for search strategy balancing quality function value and is indicated as $QF(t)$. During the tracking of prey, Aquila's movement parameter is denoted by G_1 . When chasing the prey, the slope of flight is termed as G_2 , which is minimized linearly from 2 to 0.

Algorithm 1 describes the steps of proposed SIAO algorithms.

Algorithm 1: Proposed SIAO

Step 1: Initialization Phase.

Commence by initializing the population of the AO.

Initialize the relevant parameters associated with AO.

WHILE (termination condition) do.

Calculate the values of the fitness function.

$X_{best}(t)$ finds the best solution.

for ($i = 1, 2, \dots, N$) do.

Improve the mean value of the present solution.

Improve the x , y , $LF(D)$, G_1 , G_2

If $t \leq (2/3) \times T$ then

If $r \text{ and } \leq 0.5$ then

Step 2: Expanded exploration (X_1).

Improve the present solution using Eq. 1.

If $\text{Fitness } X_1(t+1) < \text{Fitness } (X(t))$ then

$$X(t) = (X_1(t+1))$$

If $\text{Fitness } (X_1(t+1)) < \text{Fitness } (X_{best}(t))$ then

$$X_{best}(t) = (X_1(t+1))$$

Step 3: Narrowed exploration (X_2).

Improve the present solution using Eq. 3.

If $\text{Fitness } X_2(t+1) < \text{Fitness } (X(t))$ then

$$X(t) = (X_2(t+1))$$

If $\text{Fitness } (X_2(t+1)) < \text{Fitness } (X_{best}(t))$ then

$$X_{best}(t) = (X_2(t+1))$$

If then $r \text{ and } \leq 0.5$

The rand is calculated using the proposed Eq.

$$\text{"rand"} = LB + \text{rand}(0,1) \times (UB - LB).$$

Step 4: Expanded Exploitation (X_3).

Improve the present solution detailed in Eq. 6.

If $\text{Fitness } X_3(t+1) < \text{Fitness } (X(t))$ then

$$X(t) = (X_3(t+1))$$

If $\text{Fitness } (X_3(t+1)) < \text{Fitness } (X_{best}(t))$ then

$$X_{best}(t) = (X_3(t+1))$$

Step 4: Narrowed Exploitation (X_4).

Improve the present solution detailed in Eq. 8.

If $\text{Fitness } (X_4(t+1)) < \text{Fitness } (X(t))$ then

$$X(t) = (X_4(t+1))$$

If $\text{Fitness } (X_4(t+1)) < \text{Fitness } (X_{best}(t))$ then

$$X_{best}(t) = (X_4(t+1))$$

return the best solution (X_{best}).

3.1.2 Solution encoding

In this work, the optimization strategy is applied in two phases. For selecting the optimal FS from the extracted feature set F , the selected features are termed as F_s . Second, the weight of LSTM indicated as W_f is tuned optimally, and the tuned weights are denoted as W_f^* . The graphical representation in Figure 2 illustrates the input solution for the envisaged SIAO methodology.

3.2 Classification process via hybrid LSTM-QNN classifier

As delineated earlier, the optimal features chosen undergo integration into a hybrid classifier for disease presence prediction. To augment the classifier's performance, the fine-tuning of LSTM weights is intricately executed through the application of the proposed SIAO methodology (Figure 3).

3.2.1 LSTM model

The learning outcome of RNN influences the base theory of LSTM. LSTM can study the lengthy dependencies among variables (21). The long-period series is evaluated using LSTM pseudocode. Activation functions like tanh and sigmoid are essential for NNs, as they introduce non-linearity, allowing the network to tackle complex data patterns and decisions. The resultant outcome enhances the explosion gradient disappearance of the NN algorithm. For controlling the process of memorizing LSTM uses the mechanism called Gating. The unit of LSTM comprises three gates namely input, output, and forget gates.

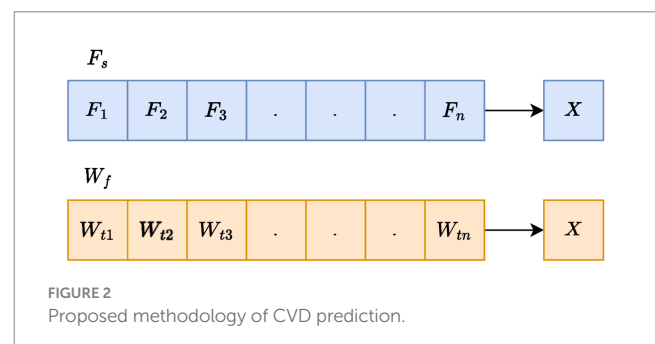
1. *Forget Gate*: Here, the attention and ignorance of information are decided. Through the function of the sigmoid, the information from the current input and hidden state is passed where the current input is denoted as $X(t)$ and the hidden state is indicated as $h(t-1)$. 0 and 1 are the range of values generated by the sigmoid function. For the point-by-point multiplication, the value of $f(t)$ is used in Eq. 12.

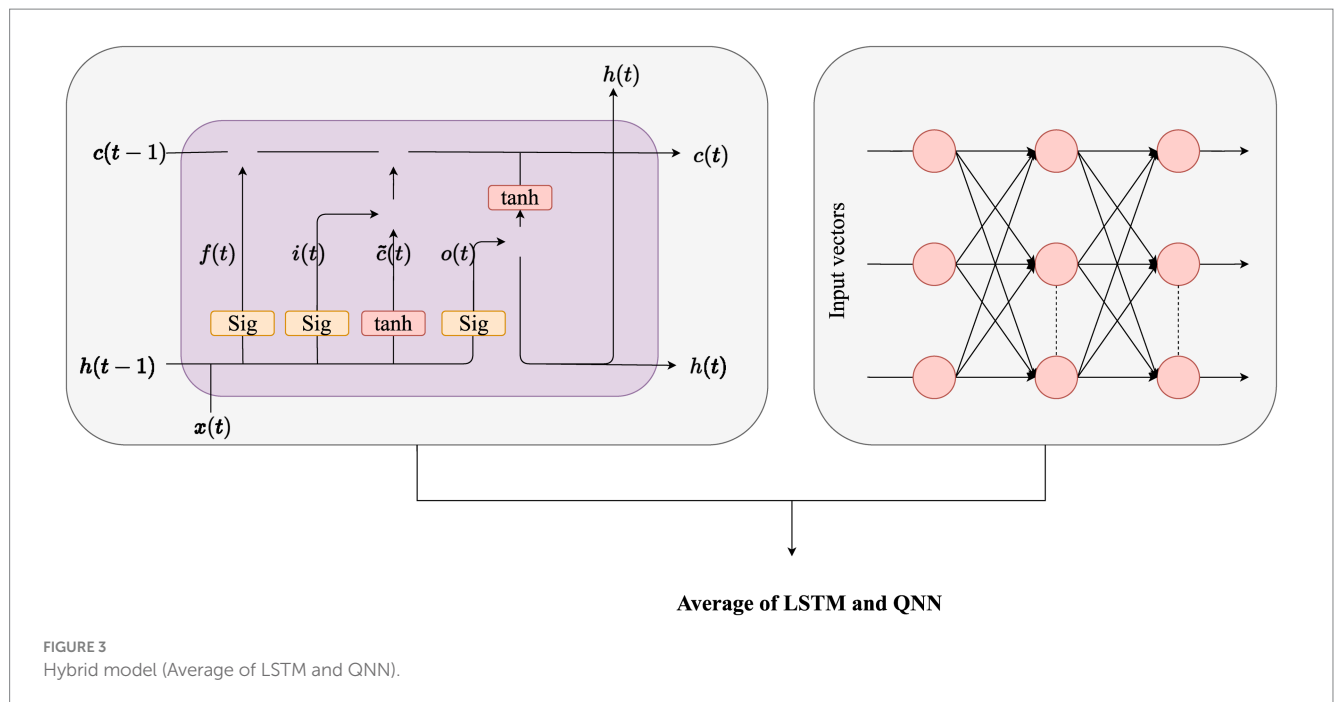
$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_t) \quad (12)$$

where timestamp id is denoted as t , f_t denotes the forget gate of t , input is determined as x_t , h_{t-1} is the previous hidden state, W_f signifies weight matrix, and b_t denotes the connection bias at timestamp t .

2. *Input Gate*: Here, the operations were performed to update the status of cells. The current position state and hidden position state are projected into the function of the sigmoid. The transformation of values takes place between 0 and 1. Then the same information will get passed to the function of the. For performing network regulation, the tanh operator generates a value range between 0 and 1. The generated values are ready for point-by-point multiplication in Eq. 13–17.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (13)$$





$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (14)$$

$$h_t = o_t \times \tanh(C_t) \quad (17)$$

Where, W_i denotes the weight matrix, b_i indicated the bias vector at t , the value generated by $\tanh$ is denoted as $\tilde{C}_t$, weight matrix of the $\tanh$ operator between cell state information and network output is indicated as W_c , and the bias vector is represented as b_c .

3. *Cell state*: The subsequent step is to select and save the information in the cell state. The multiplication is performed for the previous cell state and forgets the vector. If the value of the resultant outcome is 0, then in the cell state the value will drop. Then the point-by-point addition is performed by the output value of the vector in the input.

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (15)$$

Here, the cell state of information is denoted as C_t , the previous timestamp is indicated by C_{t-1} , and the value generated by $\tanh$ is expressed as $\tilde{C}_t$.

4. *Output Gate*: To determine the value of the hidden state, the output gate is utilized. In this state, the information on the inputs that came before it is stored. Within the beginning, the sigmoid function will be given both the value of the current state as well as the value of the hidden state that came before it. A new cell state will be generated as a result of this, and it will be sent to the function that is responsible for calculating $\tanh$. After that, a multiplication operation will be carried out on those outputs on a point-by-point basis. The network decides the information that is carried out for the hidden state based on the results that it has obtained. The hidden state that is produced as a result is then utilized for prediction.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (16)$$

Where the output gate at t is denoted by o_t , out gates' weight matrix is indicated by W_o , a vector is represented as b_o , and the output of LSTM is mentioned as h_t .

3.2.2 QNN model

A QNN (22), as elucidated in reference, constitutes a multi-layered feedforward NN renowned for its efficacy in classifying uncertain data. The QNN state shift function embodies a linear composition of multiple sigmoid functions, commonly referred to as a multi-level switch function. Unlike the binary expression of traditional sigmoid functions with two states, the QNN's hidden layer neural cells exhibit a richer spectrum of states. Introducing a discrete quantum interval for the sigmoid function allows for the mapping of diverse data onto distinct levels, affording enhanced classification flexibility. The quantum interval within a QNN is acquired through a training process. Structurally, a QNN comprises input, hidden, and output layers, with the output function of the hidden layer mathematically articulated in Eq. 18.

$$b_r = \frac{1}{ns} \sum_{s=1}^{ns} f \left[\beta (W^T X - \theta_s) \right] \quad (18)$$

Where, $\gamma = 1, 2, 3, \dots, u$ and $f(x) = 1 / (1 + \exp(-x))$ is an excitation function in which W is expressed as the weight of the network, X is the input vector, the slope factor is indicated as β , the input of the quantum cell is represented as $W^T X$, and the quantum interval is termed as θ_s .

3.3 Preprocessing phase

The preprocessing phase is conducted as an initial step to assess the data quality. Data cleaning is performed to eliminate incorrect and incomplete data. Additionally, null values and duplicate entries are removed during this preprocessing phase.

3.3.1 Central tendency

Toward a central point the size of the sample inclined toward infinity. This data property is termed a central tendency and the point toward the data gets inclined is termed a central tendency measure (23). A central propensity can be suitable for both a constrained association of features and for a theoretical transference. Moreover, some of the measures of central tendency for n data points with value $I_{data(i)}$ extracted in our proposed model are given as follows:

1. *Arithmetic Mean (AM, $\overline{I_{data(i)}}$)*: The arithmetic mean, a fundamental measure of central tendency, is denoted as the sum of all data annotations divided by the total number of data points. Eq. 19 expresses the mean of the data.

$$\overline{I_{data(i)}} = \frac{1}{n} \sum_{i=1}^n I_{data(i)} \quad (19)$$

2. *Median*: A statistical metric denoting the central value within a dataset, effectuates a division of the dataset into two equidistant halves. This partition is achieved through the meticulous arrangement of data points in ascending order, facilitating the identification of a singular data point characterized by an equitable distribution of values both superior and inferior to it. The methodology for ascertaining the median subtly diverges contingent on whether the dataset harbors an odd or even count of values. Eq. 20 elucidates the mathematical formulation encapsulating the concept of the median.

$$Median = \left\lceil \frac{I_{data(i)} + 1}{2} \right\rceil \quad (20)$$

3. *Mode*: In the dataset, one of the frequently occurring values is the mode. The mode is also a degree of central tendency that identifies the group or rating that happens the maximum often inside the distribution of data.
4. *Standard deviation (SD, σ)*: In statistics, standard deviation measures the dataset dispersion relative to the mean. Also, the SD is calculated as the variance square root. Eq. 21 denotes the mathematical expression for SD.

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (I_{data(i)} - \overline{I_{data(i)}})^2} \quad (21)$$

The minimum value obtained was considered as the initial order statistics and the maximum value is the last order statistics.

5. *Geometric mean (GM)*: A sophisticated measure of central tendency, that computes the product of specified values in a numerical series. Importantly, it is undefined if any element in the series is negative or zero, as succinctly expressed in Eq. 22.

$$GM = \left(\prod_{i=1}^n I_{data(i)} \right)^{\frac{1}{n}} \quad (22)$$

6. *Harmonic Mean (HM)*: Delineated as the reciprocal of the AM, computed from the reciprocals of individual annotations. Its evaluation is confined to a comprehensive "positive scale," ensuring meticulous consideration of positive values exclusively. Eq. 23 elegantly captures the intricate mathematical formulation underpinning the HM.

$$HM = \frac{n}{\sum_{i=1}^n \left(\frac{1}{I_{data(i)}} \right)} \quad (23)$$

7. *Trimmed Mean (TM)*: It encompasses the determination of the mean following the selective omission of specific elements from the extremes of a probability distribution or pattern. This procedure uniformly excludes an equal quantity from both the high and low ends.
8. *Interquartile range (IQR)*: Within statistical analysis, IQR assumes a pivotal role as a metric to gauge the dispersion of data and observations. The precise mathematical notation for IQR is succinctly expressed in Eq. 24, providing an exact quantification of this statistical characteristic.

$$IQR = \frac{2}{n} \sum_{i=\frac{n}{4}+1}^{\frac{3n}{4}} I_{data(i)} \quad (24)$$

9. *Midrange*: The midrange is defined as the mean of the maximum and minimum number in the dataset. It is expressed mathematically in Eq. 25.

$$M_{rg} = \frac{low(I_{data(i)}) + high(I_{data(i)})}{2} \quad (25)$$

10. *Midhinge*: The midhinge is considered as the estimation of central tendency (C) shown in Eq. 29.

$$M_{hg} = \frac{C_1(I_{data(i)}) + C_3(I_{data(i)})}{2} \quad (26)$$

11. *Trimean*: A trimean is represented as the general tendency of a data set and its mathematical notation is given in Eq. 27 where C_1 , C_2 , C_3 are central tendencies for quartiles.

$$Trim = \frac{C_1 + 2.C_2 + C_3}{4} \quad (27)$$

12. *Winsorized means*: This method pertains to an averaging technique that initially substitutes the smallest and largest values with the annotations nearest to them. This strategic replacement is executed to mitigate the influence of anomalous extreme values during the computation process.

3.3.2 Degree of dispersion

In statistical analysis, dispersion, also interchangeably referred to as variability, spread, or scatter, characterizes the degree of deviation or spreading inherent within a distribution (24). This metric delineates the extent to which data points diverge or converge from a central tendency, offering valuable insights into the distribution's inherent dynamics.

1. *IQR*: Serves as a sophisticated metric embodying statistical dispersion, elucidating the nuanced spread encapsulated between the 75 and 25 percentiles. This measure offers a granular depiction of variability by meticulously assessing the interquartile span.
2. *Range*: In the domain of statistical analysis, the Range assumes the role of a fundamental measurement, meticulously quantifying the explicit divergence existing between the uppermost and lowermost values within a dataset. This metric provides an unambiguous reflection of the dataset's inherent variability.
3. *Mean absolute difference (MAD)*: It is a quantitative facet of dispersion, that delineates the dissonance between two independently drawn values from a probability distribution. This metric affords a nuanced insight into the distributional nuances characterizing the dataset.
4. *Average absolute deviation (AAD)*: It assumes the mantle of quantifying the normative divergence of data points from the pivotal central tendency within an informational index, thereby encapsulating the comprehensive variability inherent in the dataset.
5. *Distance standard deviation*: In the insight's hypothesis, the departure distance relationship is a fraction of dependence between two mutually uneven vectors of unrestricted measurement. A diverse fraction of divergence is "dimensionless."
6. *Coefficient of Variation (CV)*: It ensconced within the domain of probability statistics, and surfaces as a comprehensive barometer of dispersal within a probability or recurrence distribution. Articulated as a ratio, the CV serves as a standardized gauge, representing the fraction of SD to the mean.
7. *Quartile coefficient of dispersion (QCD)*: A nuanced statistical metric, that assumes symbolic relevance in evaluating divergence within a dataset. Its precise calculation leverages the first (P_1) and third (P_3) quartiles for each dataset, culminating

in the articulation of the scattering coefficient, as rigorously expressed in Eq. 28.

$$QCD = \frac{P_3 - P_1}{P_3 + P_1} \quad (28)$$

8. *Replicating the coefficient of Gini & relative mean difference*: MAD, which is a precise measure of accurate divergence equivalent to the AAD of 2 independent attributes drawn from a probability distribution. A noteworthy metric associated with MAD is the AAD, representing the MAD divided by the AM and twice the Gini coefficient.
9. *Entropy (H)*: The entropy of a discrete variable displays invariance in both location and scale, signifying inherent scale independence. In contrast to traditional dispersion measures, the entropy of a continuous variable remains constant across regions and seamlessly accommodates new information, exhibiting a unique scalability. The entropy function $H(y)$ for continuous variable x , c can be arithmetically expressed in Eq. 29.

$$H(y) = H(x) + \log(c) \quad (29)$$

3.3.3 Qualitative variation (QV)

This index is the *measure of arithmetical dispersion in the ostensible distribution* (25). Between the 0 and 1 bounds, the data normalization exists and then changes to level 4. The data level changes are expressed in Table 2.

Twenty-three features are there in the QV index. Also, indices of Wilcox's and its characteristics include RanVR, MNDif, R packages, ModVR, B index, HRel, StDev, MNDif, and AvDev. Gibbs' indices include M1, M2, M4, and M6, while single-order sample indices encompass Menhinick's, Lloyd & Ghelardi's, Shannon–Wiener, Average taxonomic distinctness, Hill's diversity numbers, Theil's H, Brillouin, McIntosh's D and E, Cotgreave's, Bulla's E, Berger–Parker, Index of qualitative variation, Margalef's, Caswell's V, Rarefaction, Smith and Wilson's B, Q statistic, Harvey, Camargo's, E, Smith & Wilson's, Simpson's, Heip's, Rényi entropy, Strong's, Horn's, and Fisher's alpha. F_{QV} determined the characteristics of extracted qualitative variation.

3.3.4 Symmetric uncertainty

The characteristics and class of symmetric uncertainty are evaluated based on the estimated SU relationship metric (26). The communal information is calculated using Eq. 30.

$$CM(Q,P) = \sum PO(Q,B) \log_2 \frac{PO(Q,P)}{PO(Q).PO(P)} \quad (30)$$

In Eq. 31, communal information is indicated by CM , the feature is represented as Q , the class is denoted as P , and the function of

TABLE 2 Transformed data levels.

Datapoint transferred	Data range normalized
1	If 0–0.25
2	If 0.25–0.5
3	If 0.5–0.75
4	If 0.75–1

probability is represented as PO . Also, Eq. 31 indicates symmetrical uncertainty.

$$SU(Q,P) = 2(CM(Q,P)) / (H(Q).H(P)) \quad (31)$$

In Eq. 32, the entropy function is indicated as H . F_{SU} denotes the symmetric uncertainty feature. So, the entire feature F combines the features that are extracted like central tendency F_{CT} , degree of dispersion F_D , qualitative variation F_{QV} , and symmetrical uncertainty F_{SU} were termed in Eq. 32.

$$F = F_{CT} + F_D + F_{QV} + F_{SU} \quad (32)$$

4 Results and discussions

4.1 Simulation details

The execution of the CVD prediction model within the Python 3.11 environment involves a systematic evaluation, methodically assessing a plethora of Type I metrics and Type II metrics. This comprehensive scrutiny unfolds across two distinct datasets: Dataset 1, sourced from the Cleveland dataset (UCI Machine Learning Repository, n.d.-a) featuring 76 attributes, with a focused exploration of a refined subset of 14 attributes, notably emphasizing the Cleveland dataset. Meanwhile, Dataset 2, attained from the (UCI Machine Learning Repository, n.d.-b) comprises 13 attributes and an intricately defined cost matrix denoted as ‘abse’ and ‘pres.’ The orchestrated training and testing processes systematically unfold across varied proportions (60, 70, 80, and 90%), providing a structured lens for a nuanced examination of the predictive model’s performance.

$$\begin{matrix} \text{absence} & 0 & 1 \\ \text{presence} & 5 & 0 \end{matrix}$$

In the above matrix, the row indicates the true values and the columns predicted.

4.2 Performance analysis of the adopted and traditional model for Dataset-1

The performance of the proposed model is evaluated over the existing models like SVM (21), DBN (Deep Belief Network) (22), RNN (27), DCNN (Deep CNN) (28), 7 classifiers (DT, NB, LR, SVM, k-NN, ANN and Vote (a hybrid technique with NB and LR)) (4), 4

ML classifiers (DT, LR, XGBoost, SVM) (29), BiGRU (Bidirectional Gated Recurrent Unit) (30), SMO (Sequential Minimal Optimization)+HC (Hybrid Classifiers) (26), SSA (Salp Swarm Algorithm)+HC (31), DHOA (Dear Hunting Optimization Algorithm)+HC (32), AO+HC (7), SI+AO+LSTM+QNN+HC, accordingly. The predictive model’s performance is rigorously evaluated through key metrics, including accuracy, precision, and sensitivity, across various learning percentages (60, 70, 80, and 90%). Figure 4 illustrates the exceptional accuracy of the compositional model, achieving a remarkable 95.54% during the 90% training phase. The projected approach consistently surpasses the performance of other existing models at all learning percentages, such as SVM, DBN, RNN, DCNN, 7 classifiers, 4 ML classifiers, BiGRU, SMO+HC, SSA+HC, DHOA+HC, AO+HC, SI+AO+LSTM+QNN+HC. Figure 5 sheds light on the superior sensitivity of the proposed SI-AO-LSTM-QNN approach, particularly evident with a peak sensitivity of 95.86% at the 90% training percentage. This notable performance outshines other existing approaches. Sensitivity rates for the 60, 70, and 80% training percentages are also substantial, standing at 91.6, 92.95, and 94.39%, respectively. Precision analysis, as depicted in Figure 6, further emphasizes the prowess of the proposed model. Achieving the highest precision rate of 96.03% during the 90% training phase, the SI-AO-LSTM-QNN approach outperforms the already existing models. Precision rates for the other training percentages are commendable, measuring at 92.76, 94.33, and 95.47%.

Table 3 provides a comprehensive performance analysis for the prediction of CVD on Dataset 1, focusing on a Training percentage (TP) of 90%. Various metrics, including accuracy, sensitivity, specificity, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), Negative Predictive Value (NPV), False Positive Rate (FPR), and False Negative Rate (FNR), are reported for a range of existing models, as well as the proposed model, SI+AO+LSTM+QNN+HC. Notably, the proposed model achieves outstanding results with an accuracy of 96.69%, sensitivity of 96.62%, specificity of 96.77%, precision of 96.03%, recall of 97.86%, F1-score of 96.85%, MCC of 96.37%, NPV of 96.25%, FPR of 3.23%, and FNR of 3.38%. These metrics collectively indicate the superior predictive capabilities of the proposed SI+AO+LSTM+QNN+HC model, showcasing its robust performance compared to other existing models across a diverse set of evaluation criteria.

4.3 Performance analysis of the adopted and traditional model for Dataset-2

In dataset 2, the proposed model is compared to SVM, DBN, RNN, DCNN, 7 classifiers, 4 ML classifiers, BiGRU, SMO+HC, SSA+HC, DHOA+HC, AO+HC, SI+AO+LSTM+QNN+HC, and others. Notably, the SI-AO-LSTM-QNN approach consistently outperforms the existing models, achieving higher values across critical metrics. Specifically, for accuracy, sensitivity, and precision, the proposed model attains impressive rates of 96.69, 96.62, and 96.03%, respectively. These superior metrics are observed consistently across various learning percentages, with the highest values achieved at the 90th learning percentage. Figure 7 visually represents the accuracy comparison, illustrating that the SI-AO-LSTM-QNN model excels, achieving the highest

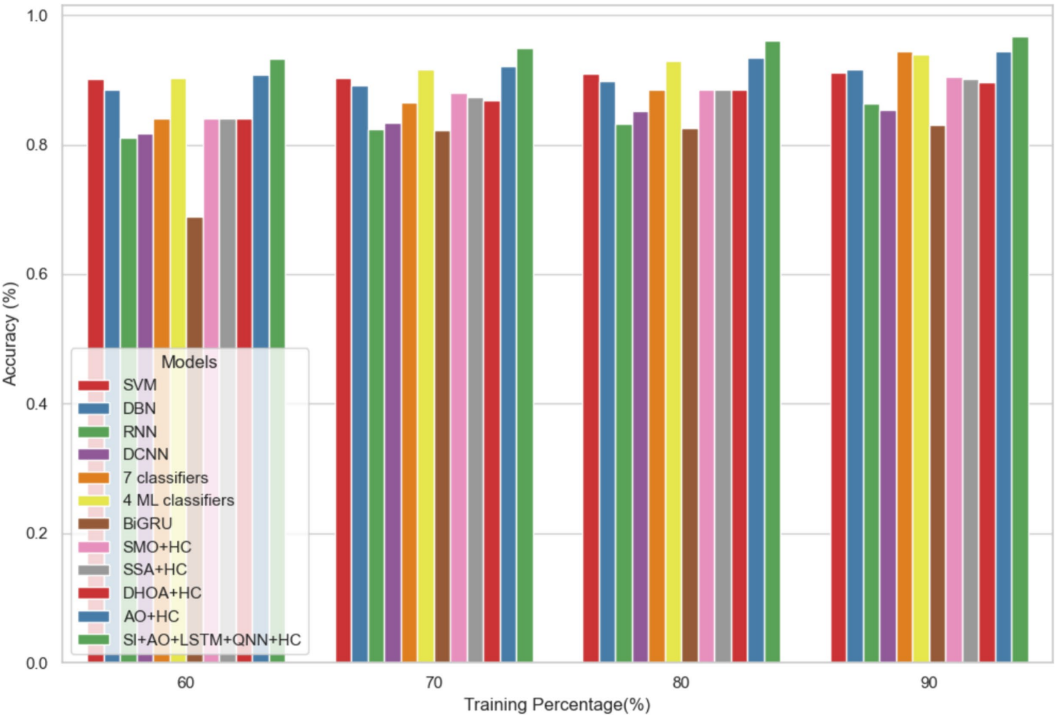


FIGURE 4
Comparative analysis of the accuracy rates in predicting CVD on Dataset-1.

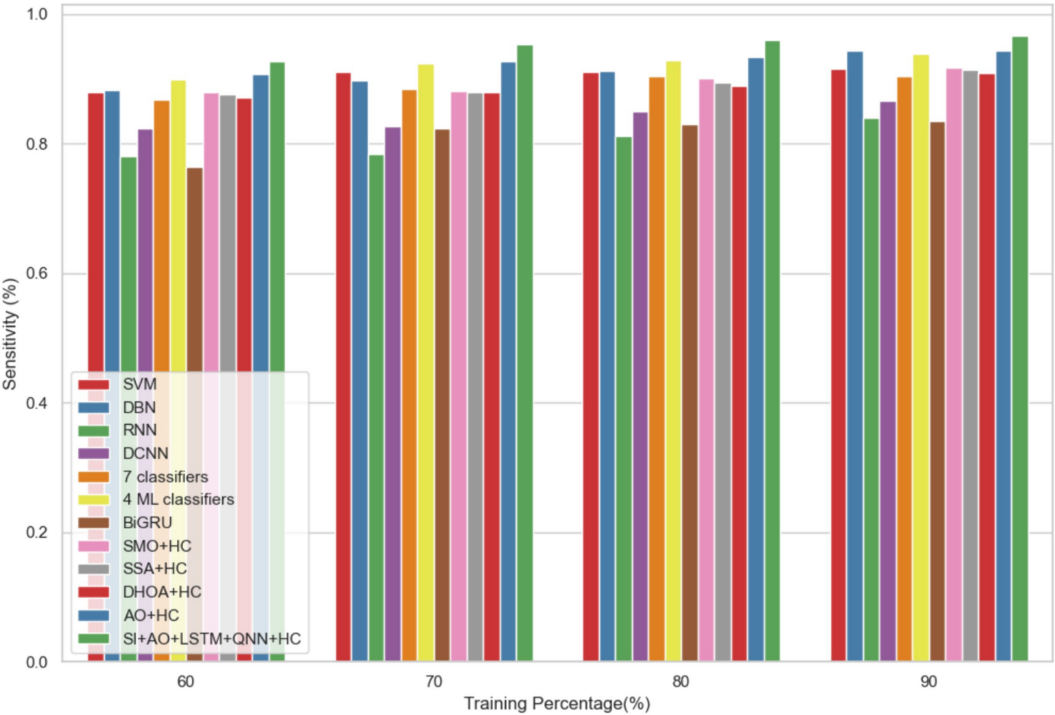


FIGURE 5
Comparative analysis of the sensitivity rates in predicting CVD on Dataset-1.

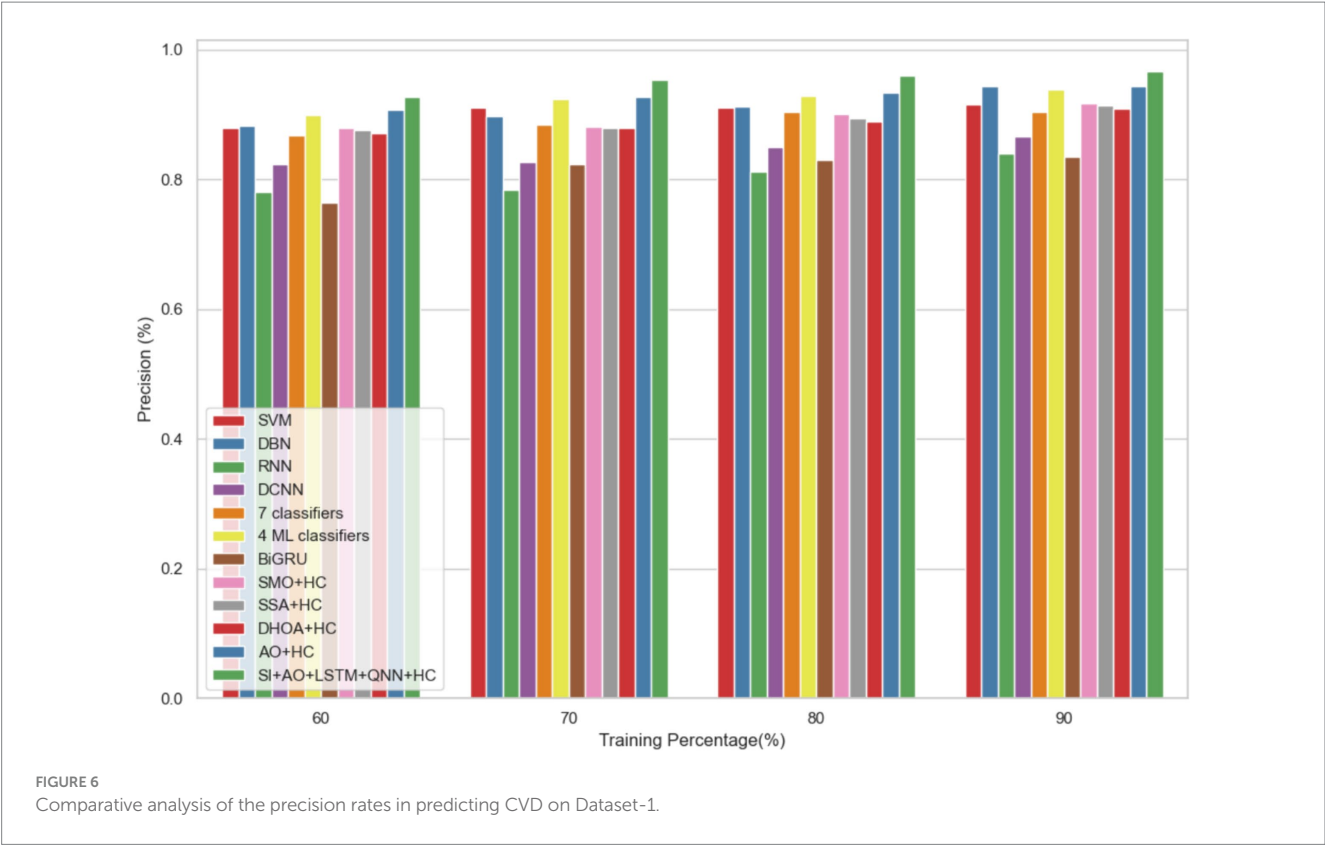


TABLE 3 Performance analysis for prediction of CVD of dataset 1 for TP = 90%.

Metrics	Accuracy	Sensitivity	Specificity	Precision	F1-score	MCC	NPV	FPR	FNR
SVM (33)	0.91079	0.915957	0.91079	0.91079	0.91079	0.91079	0.914113	0.08921	0.084043
DBN (34)	0.916754	0.943232	0.921099	0.923449	0.918028	0.91879	0.939337	0.078901	0.056768
RNN (29)	0.863295	0.83929	0.884078	0.864878	0.851892	0.780951	0.861998	0.115922	0.16071
DCNN (6)	0.854458	0.865934	0.852423	0.840533	0.853045	0.724256	0.876004	0.147577	0.134066
7 classifiers (4)	0.944006	0.904322	0.944193	0.94492	0.944902	0.877832	0.90148	0.055807	0.095678
4 ML classifier (9)	0.939135	0.938981	0.939321	0.940045	0.940026	0.928548	0.936801	0.060679	0.061019
BiGRU (25)	0.831094	0.834353	0.836954	0.803626	0.818701	0.704706	0.862893	0.163046	0.165647
SMO + HC (35)	0.905631	0.917888	0.891647	0.899332	0.908415	0.871494	0.913507	0.108353	0.082112
SSA + HC (30)	0.901581	0.91371	0.887748	0.895251	0.904286	0.868333	0.909493	0.112252	0.08629
DHOA + HC (28)	0.896831	0.908896	0.883071	0.890535	0.899522	0.863758	0.904701	0.116929	0.091104
AO + HC (7)	0.94413	0.943851	0.945169	0.948321	0.94577	0.933364	0.94166	0.054831	0.056149
LSTM (32)	0.828358	0.735632	0.961921	0.965309	0.834964	0.689585	0.7164	0.038079	0.264368
QNN (21)	0.9273	0.9461	0.9075	0.9136	0.9295	0.8740	0.9428	0.0924	0.0538
Proposed model	0.966922	0.966244	0.967714	0.9603	0.968473	0.963715	0.962534	0.032286	0.033756

accuracy among the compared models. Figure 8 showcases the precision performance, indicating higher values, especially at the 80th and 90th learning percentages. Lastly, Figure 9 presents the sensitivity analysis, highlighting the consistently superior

sensitivity of the proposed model across different training percentages. Table 4 provides a comprehensive performance analysis for the prediction of CVD on Dataset 2, with a focus on the TP rate of 90%.

Various metrics, including accuracy, sensitivity, specificity, precision, F1-score, MCC, NPV, FPR, and FNR, are reported for a range of existing models, as well as the proposed model SI + AO + LSTM + QNN + HC. The SI + AO + LSTM + QNN + HC model outshines the other models consistently across all metrics, achieving an accuracy of 95.55%, sensitivity of 95.87%, specificity of 94.52%, precision of 96.03%, F1-score of 96.94%, MCC of 93.09%, NPV of 94.67%, FPR of 5.48%, and FNR of 4.13%. These superior metrics signify the robust predictive capabilities of the proposed model, showcasing its effectiveness in comparison to a diverse set of existing models across a spectrum of evaluation criteria on Dataset-2.

4.4 Convergence analysis

Convergence analysis of the proposed SI-AO-LSTM-QNN, in comparison to conventional methods like SMO, SSA, DHOA, AO, and SI-AO, is visually presented in Figures 10, 11. The primary objective of the adopted methodology revolves around convergence analysis, with a specific focus on maximizing accuracy. The analysis reveals that heightened convergence is achieved with an increase in the iteration count. Given the inverse relationship between accuracy and errors, the overarching goal of this research is to attain the highest possible detection accuracy, thereby minimizing error rates. In Figure 10, which pertains to Dataset-1, the graphical representation illustrates superior convergence of the proposed work over existing counterparts, with maximal convergence evident at the 20th iteration. Likewise, in Figure 11, corresponding to Dataset-2, the presented work demonstrates robust convergence,

surpassing other methods and reinforcing its effectiveness in the classification.

4.5 Statistical analysis

Tables 5, 6 provide a comparative statistical analysis of accuracy for the proposed SI-AO-LSTM-QNN model against traditional schemes on Dataset-1 and Dataset-2, respectively. The stochastic nature of the optimization algorithm led to five independent runs, and statistical measures such as mean, SD, median, worst, and best were recorded for accuracy. In Table 6 for Dataset-1, the proposed SI-AO-LSTM-QNN model showcases a superior mean performance of 95.23%, outperforming traditional methods. Notably, the method exhibits a narrow SD of 1.279, indicating consistency across runs. The worst-case scenario is observed at 93.31%, and the best-case scenario attains an impressive 96.69%. In comparison, other traditional methods show varying performance levels, with SI-AO-LSTM-QNN consistently demonstrating higher accuracy.

4.6 Analysis on features

Tables 7, 8 provide an in-depth analysis of feature performance in predicting CVD for Dataset-1 and Dataset-2, respectively. In Dataset-1, the proposed feature exhibits superior predictive capabilities with an accuracy of 95.59%, outperforming scenarios without FS (94.58%) and optimization (94.62%). The proposed feature also excels in key metrics such as F1-score, precision, sensitivity, specificity, MCC, NPV, FPR,

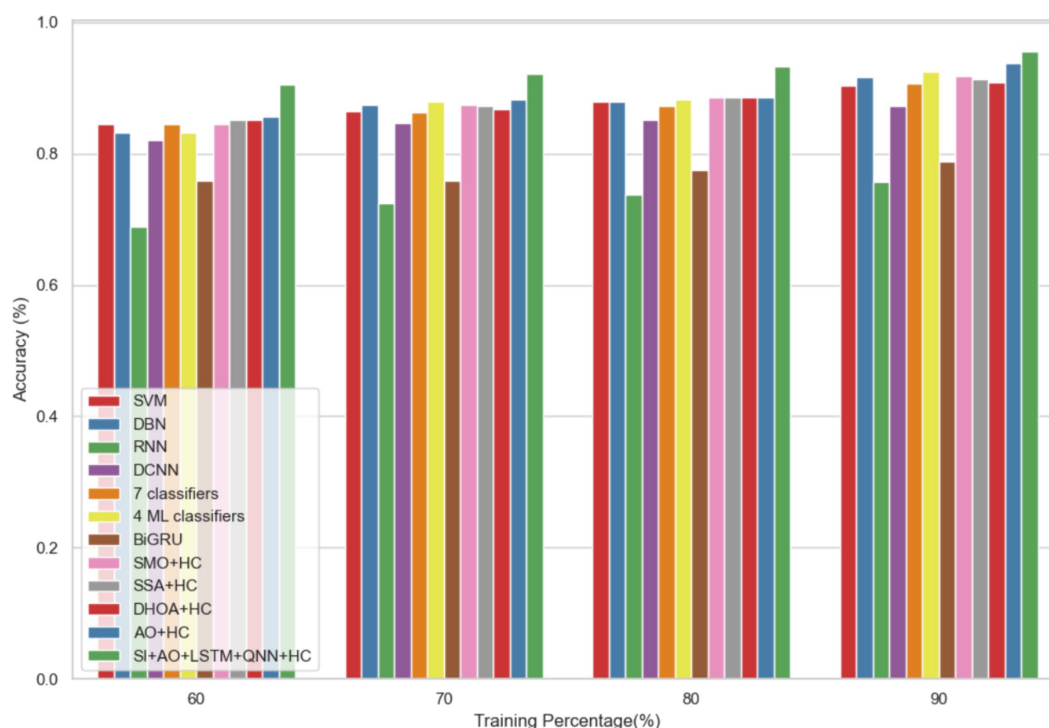


FIGURE 7
Comparative analysis of the accuracy rates in predicting CVD on Dataset-2.

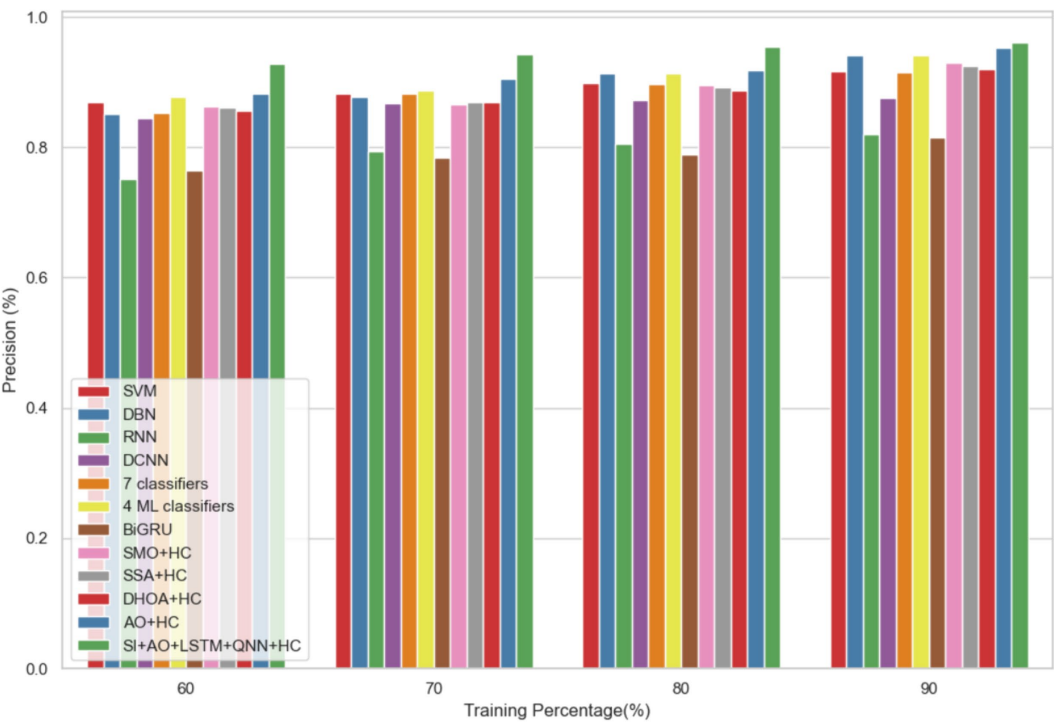


FIGURE 8
Comparative analysis of the precision rates in predicting CVD on Dataset-2.

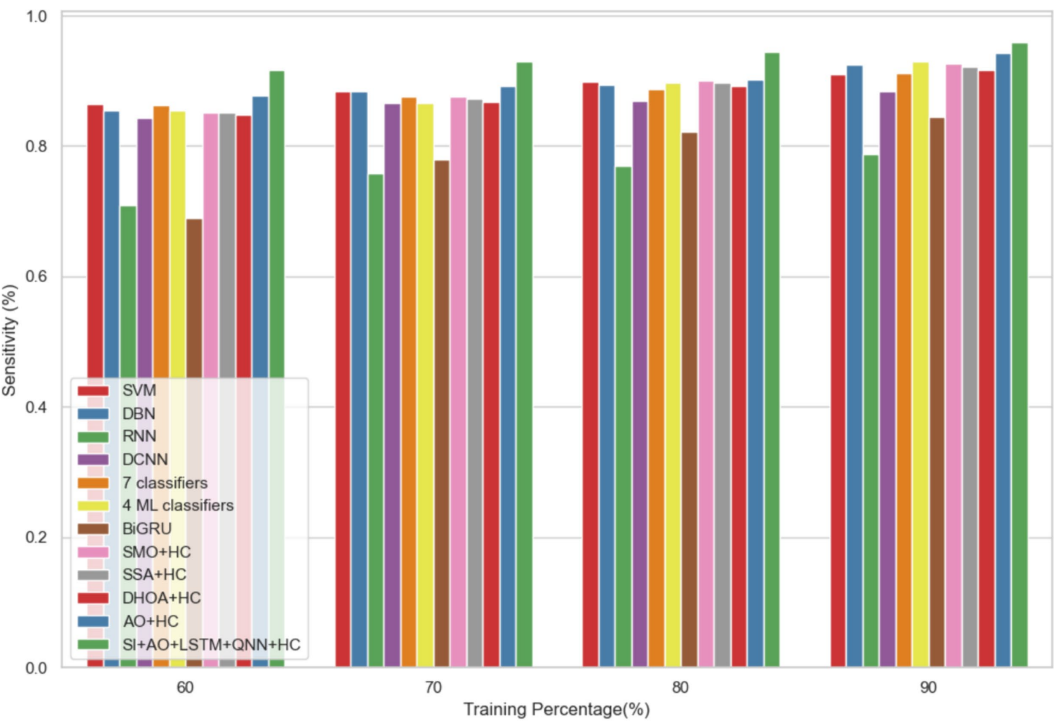
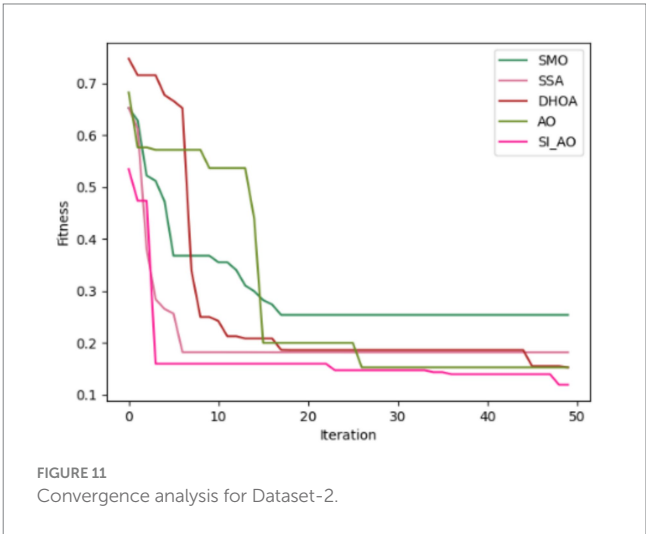
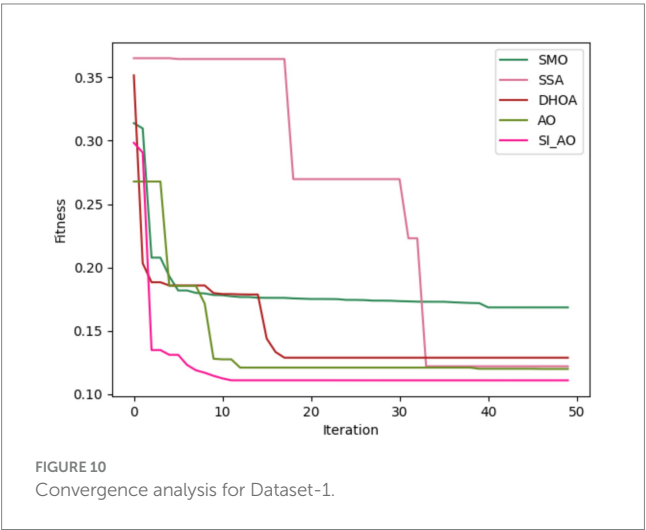


FIGURE 9
Comparative analysis of the sensitivity rates in predicting CVD on Dataset-2.

TABLE 4 Performance analysis for prediction of CVD of Dataset 2 for TP = 90%.

Metrics	Accuracy	Sensitivity	Specificity	Precision	F1-score	MCC	NPV	FPR	FNR
SVM (33)	0.903943	0.91079	0.884254	0.916544	0.918485	0.887562	0.898065	0.115746	0.08921
DBN (34)	0.91717	0.925456	0.89554	0.941486	0.937214	0.867554	0.885457	0.10446	0.074544
RNN (29)	0.757692	0.788	0.774623	0.820833	0.804082	0.754632	0.779862	0.225377	0.212
DCNN (6)	0.873491	0.884144	0.812193	0.875556	0.880953	0.782849	0.841727	0.187807	0.115856
7 classifiers (4)	0.907628	0.911215	0.892074	0.914643	0.9168	0.884667	0.905379	0.107926	0.088785
4 ML classifier (9)	0.925602	0.929508	0.912176	0.941193	0.939203	0.870667	0.910303	0.087824	0.070492
BiGRU (25)	0.788	0.844286	0.745249	0.815172	0.829474	0.780115	0.758131	0.254751	0.155714
SMO + HC (35)	0.918383	0.925771	0.89847	0.92913	0.931322	0.884695	0.912411	0.10153	0.074229
SSA + HC (30)	0.913569	0.920919	0.893732	0.924301	0.926481	0.880019	0.907629	0.106268	0.079081
DHOA + HC (28)	0.908756	0.916067	0.888993	0.919472	0.921641	0.875343	0.902847	0.111007	0.083933
AO + HC (7)	0.938558	0.943341	0.921635	0.952722	0.952239	0.888595	0.916232	0.078365	0.056659
LSTM (32)	0.8703	0.8703	0.8703	0.9306	0.8995	0.7206	0.7704	0.1296	0.1296
QNN (21)	0.9079	0.9328	0.862	0.9258	0.9293	0.7973	0.874	0.1379	0.0671
Proposed model	0.955479	0.958691	0.945167	0.9603	0.969417	0.930939	0.946673	0.054833	0.041309



and FNR, underscoring its effectiveness in enhancing the overall predictive accuracy for CVD in Dataset-1. Specifically, the proposed feature demonstrates improved sensitivity and NPV, suggesting its robust ability to correctly identify positive cases and avoid false negatives.

Turning attention to Dataset-2 in Table 8, the proposed feature showcases exceptional predictive performance, achieving an accuracy of 96.65% compared to scenarios without FS (93.49%) and optimization (94.34%). The proposed feature consistently outperforms across various metrics, emphasizing its importance in accurate CVD prediction. Particularly noteworthy are the high values for precision, sensitivity, and F1-score, indicating the ability of the proposed feature

to correctly classify positive cases and minimize false positives. Overall, both tables affirm that the inclusion of the proposed feature, with careful selection and optimization, significantly improves the predictive accuracy of CVD across different datasets.

5 Conclusion and future work

The conclusion of the paper underscores the significant advancements made in the prediction of CVD through the development and application of a *Hybrid Model that integrates LSTM*

TABLE 5 Comparative statistical analysis of accuracy for proposed and traditional schemes in Dataset-1.

Method	Mean	Median	SD	Worst	Best
SVM (33)	90.64	90.65	0.367	90.18	91.07
DBN (34)	89.81	89.50	1.156	88.58	91.67
RNN (29)	83.27	82.81	1.914	81.11	86.32
DCNN (6)	83.92	84.28	1.505	81.67	85.44
7 classifiers (4)	88.35	87.49	3.830	84.03	94.40
4 ML classifier (9)	92.25	92.35	1.337	90.37	93.91
BiGRU (25)	79.21	82.39	5.930	68.95	83.10
SMO + HC (35)	87.79	88.25	2.334	84.10	90.56
SSA + HC (30)	87.52	87.95	2.243	84.03	90.15
DHOA + HC (28)	87.29	87.72	2.119	84.03	89.68
AO + HC (7)	92.73	92.83	1.348	90.83	94.41
LSTM (32)	95.23	95.45	1.279	93.31	96.69

TABLE 6 Comparative statistical analysis of accuracy for proposed and traditional schemes in Dataset-2.

Method	Mean	Median	SD	Worst	Best
SVM (33)	87.33	87.24	2.15	84.45	90.39
DBN (34)	87.62	87.76	3.00	83.25	91.71
RNN (29)	72.73	73.10	2.48	68.95	75.76
DCNN (6)	84.82	84.91	1.85	82.13	87.34
7 classifiers (4)	87.21	86.81	2.28	84.48	90.76
4 ML classifier (9)	88.00	88.11	3.29	83.25	92.56
BiGRU (25)	77.02	76.69	1.19	75.92	78.80
SMO + HC (35)	88.09	88.02	2.62	84.48	91.83
SSA + HC (30)	88.07	87.91	2.25	85.13	91.35
DHOA + HC (28)	87.84	87.68	2.13	85.13	90.87
AO + HC (7)	89.10	88.45	2.97	85.65	93.85
LSTM (32)	92.90	92.74	1.81	90.56	95.54

and QNN. This model, optimized by a novel algorithm, demonstrates exceptional efficacy in handling complex healthcare data, as evidenced by its superior performance metrics over existing models. Notably, the model achieves a remarkable 14.05% improvement in accuracy on Dataset-1 and a 20.7% enhancement on Dataset-2, with sensitivity metrics that outperform a broad spectrum of current models including SVM, DBN, RNN, DCNN, BiGRU, SMO, SSA, DHOA, and AO. These results not only validate the model’s capability in accurately predicting CVD but also highlight its potential to significantly impact future

TABLE 7 Feature analysis for Dataset-1.

Metrics	Without FS	Without optimization	Proposed feature
Accuracy	0.945848375	0.946277097	0.955888608
F1-score	0.947162427	0.959159303	0.963933803
Precision	0.952755906	0.96353167	0.958619611
Sensitivity	0.941634241	0.954826438	0.969368891
Specificity	0.950331126	0.92962963	0.930392593
MCC	0.891698291	0.880765057	0.915424716
NPV	0.938675388	0.91355778	0.950517634
FPR	0.049668874	0.07037037	0.06960741
FNR	0.058365759	0.045173562	0.03063111

TABLE 8 Feature analysis for Dataset-2.

Metrics	Without FS	Without optimization	Proposed feature
Accuracy	0.934916	0.943396	0.96652
F1-score	0.939274	0.956938	0.974655
Precision	0.922979	0.961538	0.969282
Specificity	0.911591	0.925926	0.940741
Sensitivity	0.956864	0.952381	0.980151
MCC	0.870635	0.87446	0.925606
NPV	0.949849	0.909091	0.96109
FPR	0.088409	0.074074	0.059259
FNR	0.043136	0.047619	0.019849

healthcare practices by providing more precise and reliable diagnoses. Looking forward, the research identifies several areas for potential improvement and expansion, such as refining the optimization algorithm, further tuning the hybrid model, broader evaluation across diverse datasets, exploration of real-time implementation possibilities, and incorporation of additional data sources. These directions aim to further enhance the model’s accuracy and applicability, contributing to the ongoing evolution of predictive healthcare models and ultimately, to the advancement of patient care in cardiovascular diseases.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: <https://archive.ics.uci.edu/dataset/45/heart+disease>.

Author contributions

AD: Conceptualization, Investigation, Methodology, Writing – review & editing. RC: Methodology, Supervision, Writing – review &

editing. MA: Formal analysis, Project administration, Writing – review & editing. SD: Conceptualization, Investigation, Writing – original draft. KA: Funding acquisition, Resources, Supervision, Writing – review & editing. UL: Methodology, Project administration, Validation, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was funded by Research Supporting Project Number (RSPD2024R553), King Saud University, Riyadh, Saudi Arabia.

References

- Abualigah L, Yousri D, Abd M, Elaziz AA, Ewees MAA, Al-Qaness AHG. Aquila optimizer: a novel meta-heuristic optimization algorithm. *Comput Ind Eng*. (2017) 157:107250. doi: 10.1016/j.cie.2021.107250
- Aman, Chhillar RS. Optimized stacking ensemble for early-stage diabetes mellitus prediction. *IJECE*. (2023) 13:7048–140. doi: 10.11591/ijece.v13i6.pp7048-7055
- Aman, Chhillar RS. Disease predictive models for healthcare by using data mining techniques: state of the art. *SSRG Int J Eng Trends Technol*. (2020) 68:52–7. doi: 10.14445/22315381/IJETT-V68I10P209
- Amin MS, Chiam YK, Varathan KD. Identification of significant features and data mining techniques in predicting heart disease. *Telemat Inform*. (2014) 36:82–93. doi: 10.1016/j.tele.2018.11.007
- Arabasadi Z, Alizadehsani R, Roshanzamir M. Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm. *Comput Methods Prog Biomed*. (2017) 141:19–26. doi: 10.1016/j.cmpb.2017.01.004
- Arooj S, Rehman S, Imran A, Almuhaimeed A, Alzahrani AK, Alzahrani A. A deep convolutional neural network for the early detection of heart disease. *Biomedicine*. (2022) 10:2796. doi: 10.3390/biomed10112796
- Barfungpa SP, Sarma HKD, Samantaray L. An intelligent heart disease prediction system using hybrid deep dense Aquila network. *Biomed Signal Process Control*. (2023) 84:104742. doi: 10.1016/j.bspc.2023.104742
- Dalal S, Goel P, Onyema EM, Alharbi A, Mahmoud A, Algarni MA, et al. Application of machine learning for cardiovascular disease risk prediction. *Comput Intell Neurosci*. (2023) 2023:1–12. doi: 10.1155/2023/9418666
- Bharti R, Khamparia A, Shabaz M, Dhiman G, Pande S, Singh P. Prediction of heart disease using a combination of machine learning and deep learning. *Comput Intell Neurosci*. (2021) 2021:1–11. doi: 10.1155/2021/8387680
- Cai Y, Cai YQ, Tang LY, Wang YH, Gong M, Jing TC, et al. Artificial intelligence in the risk prediction of cardiovascular disease and development of an independent validation screening tool: a systematic review. *BMC Med*. (2024) 22:56. doi: 10.1186/s12916-024-03273-7
- Pitchal P, Ponnusamy S, Soundararajan V. Heart disease prediction: improved quantum convolutional neural network and enhanced features. *Expert Syst Appl*. (2024) 249:123534. doi: 10.1016/j.eswa.2024.123534
- Li F, Chen Y, Hongzeng X. Coronary heart disease prediction based on hybrid deep learning. *Rev Sci Instrum*. (2024) 95:118–26. doi: 10.1063/5.0172368
- Singh S, Singh A, Limkar S. Prediction of heart disease using deep learning and internet of medical things. *Int J Intell Syst Appl Eng*. (2024) 12:512–25.
- Oyewola DO, Dada EG, Misra S. Diagnosis of cardiovascular diseases by ensemble optimization deep learning techniques. *IJHISL*. (2024) 19:1–21. doi: 10.4018/IJHISL.334021
- Palanivel N, Indumathi R, Monisha N, Arikandhan K, Hariharan M. Novel implementation of heart disease classification model using RNN classification. *ICSCAN*. (2023) 2023:25–38. doi: 10.1109/ICSCAN58655.2023.10395314
- Yewale D, Vijayaragavan SP, Bairagi VK. An effective heart disease prediction framework based on ensemble techniques in machine learning. *IJACSA*. (2023) 14:218–25. doi: 10.14569/IJACSA.2023.0140223
- Behera MP, Sarangi A, Mishra D, Sarangi SK. A hybrid machine learning algorithm for heart and liver disease prediction using modified particle swarm optimization with support vector machine. *Procedia Comput Sci*. (2023) 218:818–27. doi: 10.1016/j.procs.2023.01.062
- Sudha VK, Kumar D. Hybrid CNN and LSTM network for heart disease prediction. *SN Comput Sci*. (2023) 4:172. doi: 10.1007/s42979-022-01598-9
- Elavarasi D., Kavitha R., Aanankumar S.. "Navigating heart health with an elephantine approach in clinical decision support systems." In Proceedings of the 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS). (2023), pp. 847–902.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Wei X, Rao C, Xiao X, Chen L, Goh M. Risk assessment of cardiovascular disease based on SOLSSA-CatBoost model. *Expert Syst Appl*. (2023) 219:119648. doi: 10.1016/j.eswa.2023.119648
- Jeswal SK, Chakraverty S. Recent developments and applications in quantum neural network: a review. *Arch Comput Meth Eng*. (2019) 26:793–807. doi: 10.1007/s11831-018-9269-0
- Kumar PR, Ravichandran S, Narayana S. Ensemble classification technique for heart disease prediction with meta-heuristic-enabled training system. *Bio Algorithm Med Syst*. (2021) 17:119–36. doi: 10.1515/bams-2020-0033
- Darolia A, Chhillar RS. Analyzing three predictive algorithms for diabetes mellitus against the pima Indians dataset. *ECS Trans*. (2022) 107:2697–704. doi: 10.1149/10701.2697ecst
- Aman, Chhillar R.S., "The upsurge of deep learning for disease prediction in healthcare." In Proceedings of the international conference on innovations in data analytics, Singapore: Springer Nature Singapore, (2022), pp. 102–119.
- Han Q, Liu H, Huang M, Wu Y, Zhang Y, "Heart disease prediction based on MWMOTE and res-BiGRU models." In Proceedings 2023 IEEE 6th International Conference on Pattern Recognition and Artificial Intelligence (PRAI), (2023), pp. 991–999.
- Pegolotti L, Pfaller MR, Rubio NL, Ding K, Brufau RB, Darve E, et al. Learning reduced-order models for cardiovascular simulations with graph neural networks. *Comput Biol Med*. (2024) 168:107676. doi: 10.1016/j.compbmed.2023.107676
- Lee HH, Lee H, Bhatt DL, Lee GB, Han J, Shin DW, et al. Smoking habit change after cancer diagnosis: effect on cardiovascular risk. *Eur Heart J*. (2024) 45:132–5. doi: 10.1093/eurheartj/ehad199
- Moon J, Posada-Quintero HF, Chon KH. A literature embedding model for cardiovascular disease prediction using risk factors, symptoms, and genotype information. *Expert Syst Appl*. (2023) 213:118930. doi: 10.1016/j.eswa.2022.118930
- Omankwu OC, Ubah VI. Hybrid deep learning model for heart disease prediction using recurrent neural network (RNN). *J Sci Technol Res*. (2023) 5:1259–81. doi: 10.1109/PRAI59366.2023.10332042
- Patro S.P., Padhy N., Sah R.D. "Anticipation of heart disease using improved optimization techniques." In Proc. ICCCL. Cham: Springer Nature, (2022), pp.105–113
- Rani P, Kumar R, Ahmed NMOS, Jain A. A decision support system for heart disease prediction based upon machine learning. *J Reliab Intell Environ*. (2021) 7:263–75. doi: 10.1007/s40860-021-00133-6
- Revathi TK, Balasubramaniam S, Sureshkumar V, Dhanasekaran S. An improved long short-term memory algorithm for cardiovascular disease prediction. *Diagnostics*. (2024) 14:239. doi: 10.3390/diagnostics14030239
- Sandhya Y. Prediction of heart diseases using support vector machine. *IJRASET*. (2020) 8:126–35. doi: 10.22214/ijraset.2020.2021
- Irene D, Shiny TS, Vadivelan N. Heart disease prediction using hybrid fuzzy K-medoids attribute weighting method with DBN-KELM based regression model. *Med Hypotheses*. (2020) 143:110072. doi: 10.1016/j.mehy.2020.110072
- Vadicherla D, Sonawane S. Decision support system for heart disease based on sequential minimal optimization in support vector machine. *IJESSET*. (2013) 4:19–26.
- Jain A, Rao ACS, Jain PK, Hu YC. Optimized levy flight model for heart disease prediction using CNN framework in big data application. *Expert Syst Appl*. (2023) 223:119859. doi: 10.1016/j.eswa.2023.119859
- Sarkar BK. Hybrid model for prediction of heart disease. *Soft Comput*. (2020) 24:1903–25. doi: 10.1007/s00500-019-04022-2
- Sumit RS, Chhillar S, Dalal S, Dalal UK, Lilhore S, Samiya . A dynamic and optimized routing approach for VANET communication in smart cities to secure intelligent transportation system via a chaotic multi-verse optimization algorithm. *Clust Comput*. (2024) 2024:1–24. doi: 10.1007/s10586-024-04322-9



OPEN ACCESS

EDITED BY

Hikmat A. M. Abdeljaber,
Applied Science Private University, Jordan

REVIEWED BY

Aminreza Karamoozian,
Shaoxing University, China
Saeed Ahmed,
Lund University, Sweden
Hira Munir,
Coventry University, United Kingdom

*CORRESPONDENCE

Jiancheng (Charles) Ji
✉ jcji20@szpu.edu.cn

RECEIVED 13 February 2024

ACCEPTED 03 April 2024

PUBLISHED 26 June 2024

CITATION

Butt MHF, Li JP, Ji J, Riaz W, Anwar N, Butt FF,
Ahmad M, Saboor A, Ali A and Uddin MY
(2024) Intelligent tumor tissue classification
for Hybrid Health Care Units.
Front. Med. 11:1385524.
doi: 10.3389/fmed.2024.1385524

COPYRIGHT

© 2024 Butt, Li, Ji, Riaz, Anwar, Butt, Ahmad,
Saboor, Ali and Uddin. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Intelligent tumor tissue classification for Hybrid Health Care Units

Muhammad Hassaan Farooq Butt¹, Jian Ping Li¹,
Jiancheng (Charles) Ji^{2*}, Waqar Riaz^{2,3}, Noreen Anwar⁴,
Faryal Farooq Butt⁵, Muhammad Ahmad⁶, Abdus Saboor¹,
Amjad Ali⁷ and Mohammed Yousuf Uddin⁸

¹School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China, ²Institute of Intelligent Manufacturing, Shenzhen Polytechnic University, Shenzhen, China, ³Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, ⁴Computer Engineering and Software Engineering Department, Polytechnique Montreal, Montreal, QC, Canada, ⁵Islamabad Medical Complex, NESCOM Hospital, Islamabad, Pakistan, ⁶Department of Computer Science, National University of Computer and Emerging Sciences, Chiniot-Faisalabad Campus, Chiniot, Pakistan, ⁷Department of Computer Science and Software Technology, University of Swat, Saidu Sharif, Pakistan, ⁸Department of Information Systems, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Alkhair, Saudi Arabia

Introduction: In the evolving healthcare landscape, we aim to integrate hyperspectral imaging into Hybrid Health Care Units to advance the diagnosis of medical diseases through the effective fusion of cutting-edge technology. The scarcity of medical hyperspectral data limits the use of hyperspectral imaging in disease classification.

Methods: Our study innovatively integrates hyperspectral imaging to characterize tumor tissues across diverse body locations, employing the Sharpened Cosine Similarity framework for tumor classification and subsequent healthcare recommendation. The efficiency of the proposed model is evaluated using Cohen's kappa, overall accuracy, and f1-score metrics.

Results: The proposed model demonstrates remarkable efficiency, with kappa of 91.76%, an overall accuracy of 95.60%, and an f1-score of 96%. These metrics indicate superior performance of our proposed model over existing state-of-the-art methods, even in limited training data.

Conclusion: This study marks a milestone in hybrid healthcare informatics, improving personalized care and advancing disease classification and recommendations.

KEYWORDS

hyperspectral imaging classification, Sharpened Cosine Similarity, deep learning, tumor tissues, Hybrid Health Care

1 Introduction

Our research explores the utilization of hyperspectral imaging (HI) to revolutionize tumor tissue classification in various body regions, aiming to impact the medical field significantly. This approach promises to refine diagnostic accuracy and pave the path for more personalized treatment plans. Taking a step toward the era of highly personalized, adequate healthcare, our study aims to enhance patient care. The reason HI is utilized for disease diagnosis is grounded in the understanding that changes in tissue's optical properties, stemming from morphological and biochemical alterations during disease

progression, can be detected (1). For instance, rapid cell division in malignant cells leads to increased metabolic enzyme levels and the formation of new vessels through angiogenesis to meet the demand for nutrients and oxygen (2).

HI capitalizes on these changes to identify lesions and abnormal tissue without needing histological examination, saving time and improving treatment efficacy. Biopsy samples, which are stable and easily obtained from patients, implement scanning HI feasible. Recent studies have explored correlations between HI and histological examination results to validate HI as an accurate disease diagnostic tool. Various tissues, including the breast (3), liver (4), brain (5), kidney (6), stomach (7), head and neck (8), and thyroid gland (9), have been investigated, demonstrating HI's capability for disease diagnosis. The complexity of HI is addressed by employing artificial intelligence, which exhibits comparable diagnostic accuracy compared to histology.

One notable advantage of HI-based disease diagnosis is its ability to directly examine biopsy tissue during surgery. Unlike histology, which typically takes hours, HI can analyze tissue within minutes. This rapid analysis enables real-time assessment of resection margins to check for residual tumor tissues. In a study, HI successfully identified breast cancer from excised breast tissue during surgery with an accuracy exceeding 84% (10). Additionally, HI has found application in identifying blood cells, showcasing its potential to delineate abnormal tissue without relying on biochemical techniques (11). These applications underscore the capacity of HI to support swift and accurate decision-making in clinical settings. Our research contributes significantly to the field in addressing the pressing need for more adaptable and precise tumor classification in healthcare diagnostics. The following points outline the key contributions made in this study:

- Versatile tumor classification: introduces a hyperspectral imaging-based classifier offering location-independent and adaptable tumor classification, surpassing the limitations of existing methods.
- Sharpened Cosine Similarity (SCS): SCS is proposed as an innovative technique within the hyperspectral imaging classification framework, demonstrating superior precision and efficiency for tumor classification, especially under limited training data.
- Empirical evaluation: provides a rigorous empirical evaluation of the proposed model, substantiating its superior performance through metrics like Cohen's kappa, overall accuracy, and f1-score.
- Hybrid Health Care (HHC) integration: applies hyperspectral imaging classification within HHC Units, contributing to personalized and effective medical care solutions with broader implications for healthcare informatics.

2 Literature review

Traditional imaging techniques like Magnetic Resonance Imaging (MRI) (12), Computed Tomography (CT) Scans (13), Positron Emission Tomography (PET) Scans (14), Functional MRI

(fMRI) (15), and Magnetic Resonance Spectroscopy (MRS) (16) have their own set of challenges in tumor detection (17). While these methods are indispensable, their specificity to specific tumor types hinders widespread application. Furthermore, implementing advanced deep learning algorithms presents scalability and real-time processing issues in clinical environments (18). Addressing these limitations, our approach offers a more versatile and computationally efficient alternative, enhancing its potential for clinical integration.

Elaborating on existing imaging modalities, MRI stands out for its high sensitivity (90%–95%) in brain tumor detection but grapples with the risk of false results and limitations in pinpointing specific tumor types or smaller lesions (19). CT Scans, utilizing X-rays, exhibit a sensitivity range of 60%–90% and a specificity of ~90%. Still, the method is constrained by radiation risks and less detailed soft tissue imaging (20). PET Scans employing ionizing radiation show varying sensitivity (70%–90%) and reasonable specificity (80%–90%), yet are subject to sensitivity limitations due to tumor characteristics and tracer use (21). fMRI, indicating brain activity through blood flow, offers high sensitivity (80%–90%) and specificity in identifying key brain areas but is susceptible to motion artifacts and variable interpretation (22). MRS provides a window into the biochemical makeup of tissues, yielding crucial data on tumor metabolism and types (23). Each modality contributes uniquely to tumor diagnosis, balancing specific advantages and inherent challenges.

Tumors, formed when cells behave abnormally, exhibit a range of sizes and can emerge anywhere in the body. Genes mutation, whether inherited, acquired gradually, or induced by substances like alcohol and tobacco, transform cells into cancerous ones (24). Growing tumors can invade neighboring tissues, displace normal cells, and produce enzymes breaking down surrounding tissues. Local invasion occurs when tumors grow larger, and metastasis happens when cancer cells spread to other body parts through blood or lymphatics (25). Classification involves categorizing tumors broadly by tissue, organ, or system, specifically by type, grading based on cellular and structural features using the World Health Organization (WHO) system, and staging using the Tumor Node Metastasis (TNM) system (26). Solid neoplasms, including carcinomas, sarcomas, and lymphomas, are classified based on type. The WHO Classification of Tumors provides detailed insights into tumor histotypes across various organ systems (27). According to the WHO system, tumor grading assigns a numerical grade (1–3) based on cellular differentiation. Staging relies on the TNM system, considering the presence of distant metastases (M), lymph node involvement (N), and the size or extension of the primary tumor (T) (28).

The most common cause of cancer death among children under the age of 15 and the second fastest-growing cause of cancer death among those over the age of 65 are brain tumors, which originate in brain cells and may be benign or malignant (29). Gene defects, exposure to certain chemicals, and radiation therapy to the head increase the risk of these tumors (30). Gliomas, the most common type, form from neural cells, including astrocytomas and ependymomas (31). Other types, such as brain stem gliomas, optic nerve gliomas, primitive neuroectodermal tumors (PNET), medulloblastomas, craniopharyngiomas,

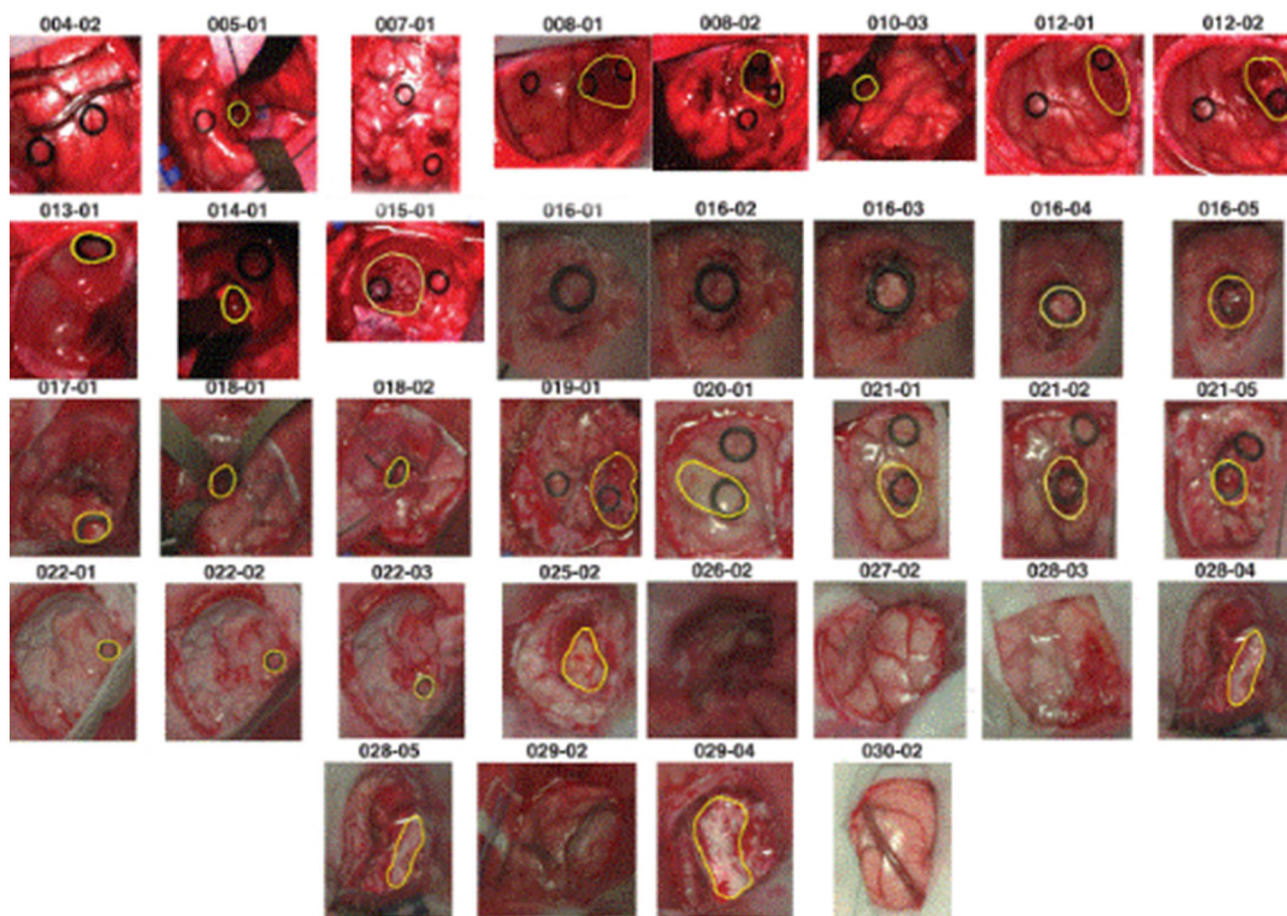


FIGURE 1

RGB representations of dataset images with PatientID-ImageID codes, delineating approximate tumor areas guided by neurosurgeon expertise and the IGS system.

and pineal region tumors, pose distinct challenges in terms of location and characteristics (32). Understanding these variations is crucial for tailored treatment approaches and underscores the complexity of brain tumor classification and detection.

Moreover, Lung carcinoma, or lung cancer (33), results from genetic mutations in airway cells triggered by factors like smoking (34). It manifests as non-small-cell lung cancer (85%) and small-cell lung cancer (15%) (35). Breast cancer originates from mutated breast cells, often spreading invasively, with common types being lobular, ductal carcinoma *in situ* (DCIS), and invasive ductal carcinoma (IDC) (36, 37). Meningiomas, arising from brain membranes, may compress nearby tissues, and their slow growth lacks a defined cause (38). HI stands poised to revolutionize tumor classification and identification by capturing unique optical properties associated with different tumor types (39). HI offers a non-invasive and potentially rapid method for precise diagnosis, contributing to improved treatment strategies and patient outcomes in lung, breast, and meningiomas.

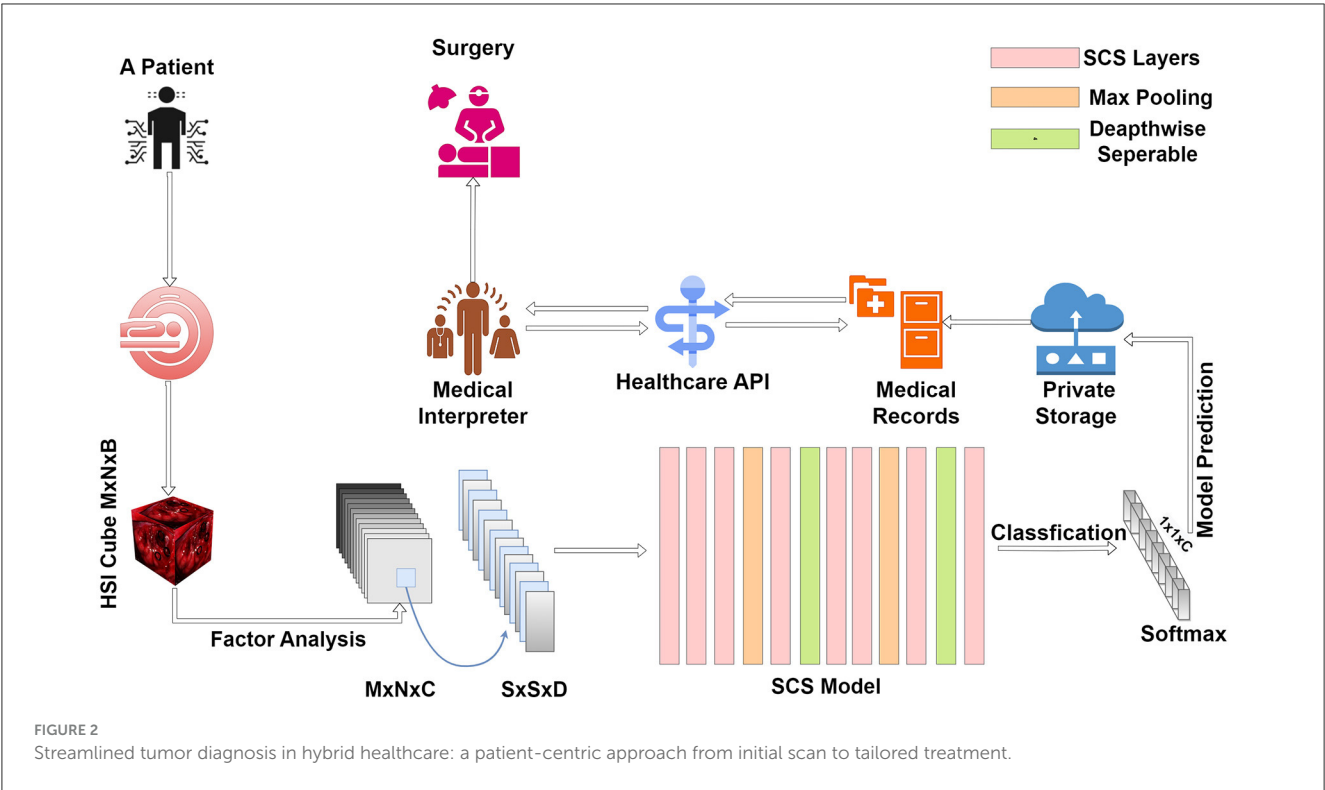
3 Materials and methods

3.1 Dataset

The dataset we used to conduct experiments was initially collected and published by the *In-vivo* HS Human Brain database (40) comprising 36 *in-vivo* brain surface images from 22 unique patients. This labeled dataset includes tumor and normal tissue, blood vessels, and other irrelevant materials within the surgical scene (referred to as background). Tumor types are differentiated in the dataset, encompassing primary (grade IV glioblastoma and grade III and II anaplastic oligodendrogliomas) and secondary tumors (lung and breast). Additionally, RGB representations of hyperspectral cubes within the *in-vivo* hyperspectral human brain image database are presented in Figure 1. The dataset designates the approximate tumor area using a yellow line, aiding in identifying the rubber ring marker corresponding to pathological analyses of the tumor tissue. Patient ID and Image ID details in Table 1 offer a comprehensive overview, including image characteristics and the pathological diagnosis of each image. The total number of labeled pixels for each class and image is specified, addressing cases where

TABLE 1 Patient-specific image data and label distribution where “N” refers to “normal” tissue, “T” signifies “tumor tissue,” “BV” represents “blood vessels,” and “B” denotes the “background.”

Patient ID	Image ID	Size	True labels				Diagnosis
			N	T	BV	B	
004	02	389 × 345 × 826	✓	✗	✓	✓	Normal brain
005	01	483 × 488 × 826	✓	✗	✓	✓	Renal carcinoma (S)
007	01	582 × 400 × 826	✓	✗	✓	✗	Normal brain
008	01	460 × 549 × 826	✓	✓	✓	✓	Grade IV glioblastoma (P)
008	02	480 × 553 × 826	✓	✓	✓	✓	Grade IV glioblastoma (P)
010	03	460 × 549 × 826	✓	✗	✓	✓	Grade IV glioblastoma (P)
012	01	443 × 497 × 826	✓	✓	✓	✓	Grade IV glioblastoma (P)
012	02	445 × 498 × 826	✓	✓	✓	✓	Grade IV glioblastoma (P)
013	01	298 × 253 × 826	✓	✗	✓	✓	Lung carcinoma (S)
014	01	317 × 244 × 826	✗	✓	✓	✓	Grade IV glioblastoma (P)
015	01	376 × 494 × 826	✓	✓	✓	✓	Grade IV glioblastoma (P)
016	01	376 × 494 × 826	✓	✗	✓	✓	Normal brain
016	02	335 × 326 × 826	✓	✗	✗	✓	Normal brain
016	03	376 × 494 × 826	✓	✗	✓	✓	Normal brain
016	04	383 × 297 × 826	✓	✗	✓	✓	Grade IV glioblastoma (P)
016	05	414 × 292 × 826	✓	✗	✓	✓	Grade IV glioblastoma (P)
017	01	441 × 399 × 826	✓	✗	✓	✓	Grade IV glioblastoma (P)
018	01	479 × 462 × 826	✓	✗	✓	✓	Grade I glioblastoma (P)
018	02	510 × 434 × 826	✓	✗	✓	✓	Grade I glioblastoma (P)
019	01	601 × 535 × 826	✓	✗	✓	✓	Meningioma
020	01	378 × 330 × 826	✓	✓	✓	✓	Grade IV glioblastoma (P)
021	01	452 × 334 × 826	✓	✓	✓	✓	Breast carcinoma (S)
021	02	448 × 324 × 826	✓	✓	✓	✓	Breast carcinoma (S)
021	05	378 × 330 × 826	✓	✗	✓	✓	Breast carcinoma (S)
022	01	597 × 527 × 826	✓	✗	✓	✓	Grade III anaplastic oligodendroglioma (P)
022	02	611 × 527 × 826	✓	✗	✓	✓	Grade III anaplastic oligodendroglioma (P)
022	03	592 × 471 × 826	✗	✓	✗	✗	Grade III anaplastic oligodendroglioma (P)
025	02	473 × 403 × 826	✓	✓	✓	✓	Grade IV glioblastoma (P)
026	02	340 × 324 × 826	✓	✗	✓	✗	Normal brain
027	02	493 × 476 × 826	✓	✗	✓	✓	Normal brain
028	03	422 × 398 × 826	✓	✗	✓	✓	Normal brain
028	04	482 × 408 × 826	✗	✗	✗	✓	Lung adenocarcinoma (S)
028	05	482 × 390 × 826	✗	✓	✗	✗	Lung adenocarcinoma (S)
029	02	365 × 371 × 826	✓	✗	✓	✓	Normal brain
029	04	399 × 342 × 826	✗	✓	✗	✓	Grade II anaplastic oligodendroglioma (P)
030	02	382 × 285 × 826	✓	✗	✓	✓	Normal brain



certain images were diagnosed as specific tumor types without labeled tumor samples due to procedural challenges.

The authors (40) mention the inherent challenges in acquiring *in-vivo* HI during neurosurgical procedures; the dataset primarily captures common tumor types over two years. The customized hyperspectral acquisition system, a preliminary demonstrator, is designed to capture tumor images on the surface or in easily focused deeper layers. The authors utilize a push broom camera for spatial scanning; the system's limitations include increased acquisition time and potential spatial coherence issues due to patient brain movement and procedural artifacts. As snapshot cameras offer real-time image acquisition but have fewer spectral bands than push-broom cameras, future investigations using high spectral resolution push-broom cameras are warranted. The dataset creation process by authors (40) addresses challenges from limited patient availability, presenting a preliminary database for exploring HI applications in tissue and tumor identification, tumor boundary delineation, and providing pertinent information for neurosurgeons. Their methodology leverages spectral characteristics guided by intraoperative MRI, surgeon expertise, and pathological analysis results. Subsequent data acquisition efforts are anticipated to broaden the database, encompassing more tumor types with detailed pathological descriptions.

3.2 HHC: AI tumor diagnostics

Our innovative methodology for tumor tissue classification within an HHC Unit unfolds with the patient's arrival at the facility. The initial phase involves a hyperspectral sensor scan,

capturing intricate details of the patient's internal composition. This technology provides a comprehensive overview, laying the foundation for precise diagnosis. Following the hyperspectral scan, the acquired data undergoes processing through Factor Analysis. This step is crucial for dimension reduction, ensuring that the hyperspectral cube retains only relevant features essential for accurate classification. The processed data then traverses through the layers of our SCS model. As a breakthrough in tumor classification, the SCS model enhances precision, even when trained with limited data. This stage is pivotal for predicting and classifying tumor tissues, contributing to superior performance compared to existing models.

Once the classification is complete, the results are securely stored within the hospital's private records, ensuring data confidentiality. This stored information becomes a valuable resource for future reference and analysis. Integrated into the HHC Unit is a seamless access mechanism through Healthcare APIs. Healthcare professionals can leverage these APIs to access detailed reports and results related to tumor tissues. This integration streamlines the diagnostic process, providing a user-friendly interface for medical interpretation.

The final phases of our methodology involve the medical interpreter within the Healthcare API, aiding healthcare professionals in interpreting results and making informed recommendations. These recommendations extend to surgical interventions and ongoing medical care, all tailored to the specific classification of tumor tissues and their respective locations. Figure 2 presents a comprehensive and patient-centric approach to tumor tissue classification within the HHC Unit. By seamlessly integrating hyperspectral imaging, Factor Analysis, and the

innovative SCS model, we aim to revolutionize healthcare diagnostics and enhance the overall patient experience.

3.3 Proposed Sharpened Cosine Similarity method

HI represented as $\mathbf{X} \in \mathbb{R}^{(M \times N) \times B}$, where the dimensions $(M \times N)$ correspond to a specific area on the tissue surface and B denotes the total number of spectral bands in the HI. Each pixel within $\mathbf{X}$, indicated as x_{ij} where $i = 1, 2, \dots, M$ and $j = 1, 2, \dots, N$, is grouped into C unique tissue types, collectively expressed as $\mathbf{Y} = (y_1, y_2, \dots, y_n)$. Moreover, every $x_{ij} \in \mathbf{X}$ describes a tissue pixel through a spectral vector $x_{ij} = [x_{ij,1}, x_{ij,2}, \dots, x_{ij,B}] \in \mathbf{X}$, containing a series of B spectral data points.

In the initial processing phase, spatial characteristics are emphasized by implementing a patch extraction method. This preliminary step involves the creation of a hyperspectral cube, $\mathbf{x}_{ij} \in \mathbb{R}^{(s \times s) \times D}$, encapsulating the area surrounding the focal pixel (i, j) over a region of dimensions $s \times s$. This approach is instrumental in enhancing the model's ability to distinguish between different features by integrating spectral and spatial attributes. As such, the spectral-spatial cubes $\mathbf{x}_{ij}$, drawn from the primary data and conforming to the dimensionality $\mathbb{R}^{(s \times s) \times D}$, are consolidated into the dataset $\mathbf{X}$ in preparation for subsequent feature extraction processes. The concluding step involves the selection of training and testing samples across each distinct class.

In neural networks, the convolution operation involves a sliding dot product operation, symbolized as $\mathbf{w} \cdot \mathbf{x}_{ij}$, between an image patch $\mathbf{x}_{ij}$ and a filter $\mathbf{w}$, which might miss crucial information due to its basic similarity measure. Enhancing this with normalization transforms the operation into cosine similarity, defined as $\frac{\mathbf{w} \cdot \mathbf{x}_{ij}}{\|\mathbf{w}\| \|\mathbf{x}_{ij}\|}$. This is similar to calculating the cosine of the angle between vectors, utilizing Euclidean distance.

To address these limitations, Strided Cosine Similarity (SCS) was developed as expressed in Equation (1). It operates similarly to convolution but includes key differences. In standard convolution, the operation is a dot product $\mathbf{w} \cdot \mathbf{x}_{ij}$, while SCS involves normalizing the vectors. The normalization in SCS ensures the magnitude of vectors is unity before the dot product, leading to an expression like $\frac{\mathbf{w} \cdot \mathbf{x}_{ij}}{\|\mathbf{w} + \mathbf{q}\| \|\mathbf{x}_{ij} + \mathbf{q}\|}$, where $\mathbf{q}$ is a small value to avoid numerical instability.

The similarity measure in SCS ranges between -1 and 1 , indicating complete opposition or perfect alignment of the kernel and image patch, respectively. To mitigate the issue of small magnitudes, which can lead to noise inclusion, additional parameters are introduced in SCS, formulated as;

$$\text{SCS}(\mathbf{w}, \mathbf{x}_{ij}) = \frac{\mathbf{w} \cdot \mathbf{x}_{ij}}{\|\mathbf{w} + \mathbf{q}\| \|\mathbf{x}_{ij} + \mathbf{q}\|} \quad (1)$$

Similar to conventional convolution in deep learning, SCS is a striding operation that extracts features from an image patch. However, it includes an additional step of magnitude normalization before the dot product, leading to what some literature refers to as Sharpened Cosine normalization. The effectiveness of SCS surpasses traditional convolutional processes in terms of speed due to fewer required parameters and the absence of normalization or activation functions.

In contrast to standard pooling, absolute max-pooling is employed in SCS for backpropagation filter updates, selecting the highest magnitude irrespective of the sign. The overall model with SCS is trained over 50 epochs, a batch size of 256, and a learning rate of 0.001. The learning rate significantly influences the model's learning rate, while momentum aids accuracy and speed. A root mean square prop and momentum-based optimizer, specifically the Adam optimizer, is utilized for its efficiency and computational advantages.

4 Experiment analysis

This section presents an overview of the evaluation metrics, baselines SOTA and implementation details.

4.1 Evaluation metrics

The results presented in this study are evaluated using the following metrics:

Kappa statistic: This statistical measure assesses the level of agreement between predicted classifications and ground-truth maps, as defined by Equation (2). In this equation, A_o represents the observed agreement, calculated using Equation (3), while A_e denotes the expected agreement, computed using Equation (4).

$$\kappa = \frac{A_o - A_e}{1 - A_e} \quad (2)$$

where,

$$A_o = \frac{TP + TN}{TP + FN + FP + TN} \quad (3)$$

and,

$$A_e = \left(\frac{FN + TN}{TP + FN + FP + TN} \times \frac{FP + TN}{TP + FN + FP + TN} \right) + \frac{TP + FN}{TP + FN + FP + TN} \quad (4)$$

Here, TP and FP denote true positives and false positives, respectively, while TN and FN represent true negatives and false negatives.

Average accuracy (AA): AA signifies the average classification performance across different classes, as depicted in Equation (5).

$$AA = \frac{TP + TN}{TP + TN + FN} \quad (5)$$

Overall accuracy (OA): OA is computed as the ratio of correctly classified examples to the total number of test examples, as defined by Equation (6).

$$OA = \frac{1}{N} \sum_{i=1}^N TP_i \quad (6)$$

In the equations above, TP represents true positives, FP represents false positives, TN represents true negatives, and FN represents false negatives.

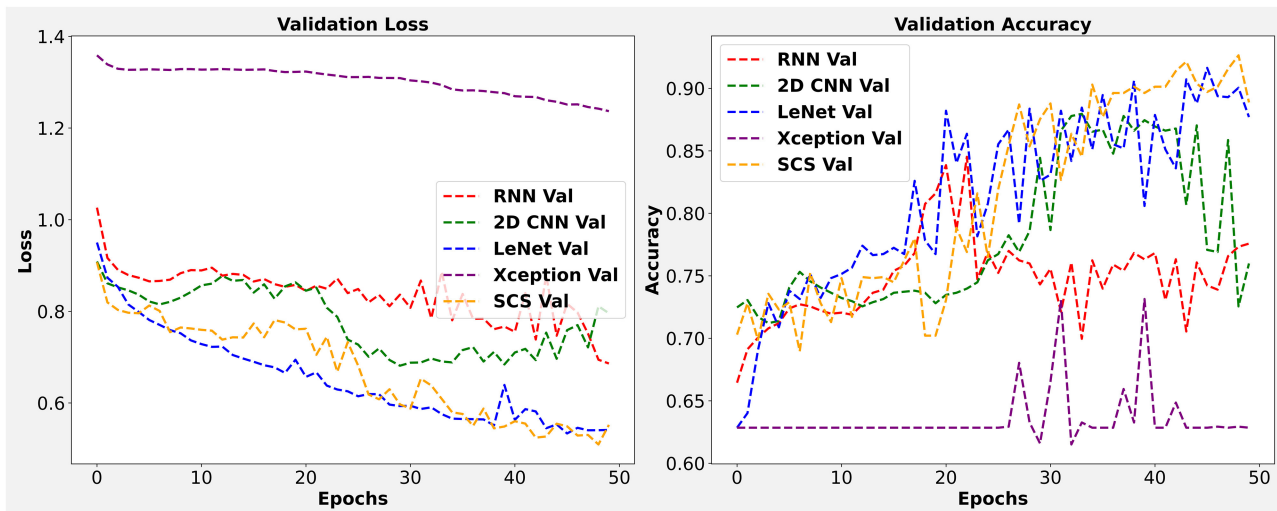


FIGURE 3
Visualization of validation loss and accuracy for 2D CNN, RNN, LeNet, Xception and SCS.

4.2 Baseline models

4.2.1 Recurrent Neural Networks

The Recurrent Neural Networks (RNN) architecture (41) presents a blend of convolutional and fully connected layers within a Sequential model. Beginning with a Conv2D layer employing a 3×3 kernel and ReLU activation, the subsequent MaxPooling2D layer downsamples the spatial dimensions. Flattening the output precede a fully connected layer of 100 neurons, integrated with Batch Normalization and ReLU activation for regularization. With softmax activation, the final layer tailors the output to fit the specified number of classes. This design reflects a hybrid approach, incorporating convolutional operations followed by dense layers, offering flexibility for various applications in classification tasks.

4.2.2 2-Dimensional Convolution Neural Network

The 2-Dimensional Convolution Neural Network (2D CNN) architecture (42) is structured within a Sequential model, featuring a Conv2D layer with a 3×3 kernel and ReLU activation, applied to input data of shape (window size, window size, kernel size). Subsequently, a MaxPooling2D layer down-samples spatial dimensions with a pooling size adjustment option. The flattened output leads to a fully connected layer with 100 neurons, supplemented by Batch Normalization and ReLU activation for regularization. The final layer, employing softmax activation, tailors the output to match the specified number of classes. This design reflects a standard 2D convolutional neural network suitable for diverse classification tasks with image data. Adjustments to the pooling size provide adaptability based on specific requirements.

4.2.3 LeNet

The LeNet architecture, a seminal convolutional neural network devised by Yann LeCun in the 1990s, marked

a breakthrough in computer vision (43). Comprising two convolutional layers with 5×5 filters and ReLU activation, each succeeded by average pooling; the network captures hierarchical features in the input. The subsequent dense layers, with 120 and 84 neurons, distill high-level representations. The final layer, employing softmax activation, tailors the output to the number of classes. LeNet's simplicity and efficacy laid the foundation for modern CNNs, influencing subsequent developments in image classification.

4.2.4 Xception

The Xception architecture (44) is a 2D variant of the Xception neural network, known for its depth-wise separable convolutions and exceptional performance in image classification tasks. The model begins with an entry flow featuring a series of convolutional layers with batch normalization and ReLU activation. The residual block 1 introduces separable convolutions, preserving spatial information efficiently. The middle flow comprises eight repeated blocks, each containing three separable convolutional layers, facilitating feature extraction. The exit flow further refines features with a combination of separable convolutions and residual connections. The model concludes with a global average pooling layer and a dense layer with softmax activation, tailoring the output to the specified number of classes. The Xception architecture is designed to capture complex hierarchical features in image data, making it suitable for various image classification tasks. Adjustments to the number of filters and other parameters can be made based on specific requirements.

4.3 Implementation details

For our empirical assessment, we utilized *in-vivo* HS Human Brain database which is already discussed in Section 3.1 and accessible on request on this

TABLE 2 Performance analysis of the SOTA models on each predicted class for different patient IDs and image IDs.

Patient ID	Image ID	Classes	2D CNN	RNN	LeeNet	Xception	SCS model
004	02	Normal tissue	0.88	0.86	0.97	0.77	0.96
		Tumor tissue	–	–	–	–	–
		Hypervascularized tissue	0.44	0.18	0.26	0.00	0.82
		Background	0.54	0.68	0.75	0.00	0.77
005	01	Normal tissue	0.85	0.87	0.95	0.72	0.98
		Tumor tissue	–	–	–	–	–
		Hypervascularized tissue	0.50	0.22	0.28	0.10	0.98
		Background	0.50	0.63	0.80	0.15	0.86
008	01	Normal tissue	0.82	0.85	0.89	0.82	0.99
		Tumor tissue	0.95	0.92	0.98	0.15	1.00
		Hypervascularized tissue	0.84	0.81	0.89	0.16	0.99
		Background	0.89	0.92	0.96	0.19	1.00
013	01	Normal tissue	0.99	0.89	0.99	0.95	0.99
		Tumor tissue	–	–	–	–	–
		Hypervascularized tissue	0.92	0.88	0.95	0.11	1.00
		Background	0.94	0.93	0.99	0.15	0.99
018	01	Normal tissue	1.00	1.00	1.00	0.98	1.00
		Tumor tissue	–	–	–	–	–
		Hypervascularized tissue	0.98	0.98	0.99	0.18	1.00
		Background	1.00	1.00	0.25	0.15	1.00
019	01	Normal tissue	0.97	0.87	1.00	0.91	1.00
		Tumor tissue	–	–	–	–	–
		Hypervascularized tissue	0.94	0.31	0.99	0.25	0.99
		Background	0.91	0.69	0.99	0.18	1.00
021	01	Normal tissue	0.93	0.26	0.98	0.95	1.00
		Tumor tissue	0.71	0.47	0.97	0.35	0.99
		Hypervascularized tissue	0.95	0.42	1.00	0.13	1.00
		Background	0.98	0.73	0.95	0.23	0.98
022	01	Normal tissue	1.00	0.80	1.00	0.89	1.00
		Tumor tissue	–	–	–	–	–
		Hypervascularized tissue	0.95	0.69	1.00	0.19	1.00
		Background	0.97	0.81	1.00	0.23	1.00
028	05	Normal tissue	–	–	–	–	–
		Tumor tissue	1.00	1.00	1.00	0.95	1.00
		Hypervascularized tissue	–	–	–	–	–
		Background	–	–	–	–	–
029	04	Normal tissue	–	–	–	–	–
		Tumor tissue	0.98	0.76	0.98	0.86	1.00
		Hypervascularized tissue	–	–	–	–	–
		Background	0.99	0.75	0.86	0.12	1.00

The bold values represent the Class-Wise Performance Analysis Based on F1-Score for Different Patient IDs and Image IDs.

TABLE 3 Comparative performance analysis of SOTA at patient and image level for each predicted class.

Patient ID	Image ID	Classes	2D CNN	RNN	LeeNet	Xception	SCS model
004	02	Kappa accuracy	49.93	47.66	71.67	30.02	80.53
		Overall accuracy	77.00	77.09	85.61	62.86	89.91
		Average accuracy	59.23	54.37	67.09	33.33	86.25
		F1-score	77	77	86	63	90
		Training time (seconds)	10.15	10.35	10.23	226.65	68.92
		Testing time (seconds)	0.75	0.78	0.70	4.68	2.04
005	01	Kappa accuracy	58.46	60.51	68.85	29.45	91.98
		Overall accuracy	66.77	63.92	70.35	35.25	97.6
		Average accuracy	61.66	57.33	67.66	32.33	94.0
		F1-score	68	59	70	34	97
		Training time (seconds)	39.63	41.60	42.93	254.75	83.45
		Testing time (seconds)	1.71	1.95	2.10	5.65	3.12
008	01	Kappa accuracy	84.60	84.0	89.84	31.47	98.45
		Overall accuracy	90.10	89.91	95.89	37.35	100.0
		Average accuracy	87.50	87.25	93.90	33.45	99.50
		F1-score	91	90	95	35	100
		Training time (seconds)	8.94	8.42	7.30	154.89	75.2
		Testing time (seconds)	1.38	2.91	1.63	2.84	1.73
013	01	Kappa accuracy	92.59	86.6	96.47	39.25	99.45
		Overall accuracy	93.97	92.9	98.6	41.79	99.76
		Average accuracy	95.16	90.4	97.66	40.33	99.89
		F1-score	95	90	98	42	100
		Training time (seconds)	5.73	7.94	6.76	107.21	80.72
		Testing time (seconds)	0.23	1.71	1.60	2.40	1.65
018	01	Kappa accuracy	99.23	99.51	72.12	41.62	99.71
		Overall accuracy	99.61	99.75	76.59	44.82	99.85
		Average accuracy	99.11	99.40	74.66	43.66	99.38
		F1-score	100	100	76	45	100
		Training time (seconds)	36.57	107.21	86.04	352.47	239.37
		Testing time (seconds)	1.36	1.60	1.40	4.37	1.83
019	01	Kappa accuracy	93.38	25.45	98.58	41.68	98.95
		Overall accuracy	95.88	61.83	99.11	45.23	99.50
		Average accuracy	86.93	43.01	98.45	44.66	99.33
		F1-score	96	62	99	46	100
		Training time (seconds)	17.17	95.1	38.46	537.42	354.81
		Testing Time (seconds)	0.59	2.89	1.43	6.85	4.29
021	01	Kappa accuracy	85.12	45.74	97.83	38.61	98.55
		Overall accuracy	90.99	48.39	98.91	42.35	99.75
		Average accuracy	89.25	45.19	97.56	41.50	99.25
		F1-score	90	47	98	41	100
		Training time (seconds)	3.47	5.41	38.46	37.42	34.18
		Testing time (seconds)	1.54	2.89	1.43	3.85	3.19

(Continued)

TABLE 3 (Continued)

Patient ID	Image ID	Classes	2D CNN	RNN	LeeNet	Xception	SCS model
022	01	Kappa accuracy	97.86	65.68	99.82	43.12	99.80
		Overall accuracy	98.92	79.33	99.89	47.98	99.90
		Average accuracy	97.33	74.64	99.90	46.66	99.85
		F1-score	98	79	100	48	100
		Training time (seconds)	11.80	67.98	23.49	335.96	139.54
		Testing time (seconds)	0.48	1.54	0.85	3.38	1.24
028	05	Kappa accuracy	–	–	–	–	–
		Overall accuracy	100	100	100	97.42	100
		Average accuracy	100	100	100	95.36	100
		F1-score	100	100	100	95	100
		Training time (seconds)	4.71	6.45	9.08	96.41	43.18
		Testing time (seconds)	0.20	0.37	0.53	1.76	0.93
029	04	Kappa accuracy	–	–	–	–	–
		Overall accuracy	98.9	75.8	92.69	50.81	100
		Average accuracy	98.3	75.64	91.45	49.05	100
		F1-score	99	78	92	50	100
		Training time (seconds)	3.91	5.59	5.05	35.55	27.91
		Testing time (seconds)	0.40	0.68	0.43	1.86	1.12

The bold values represent the Comparative Performance Analysis of SOTA and SCS Models Across Various Evaluation Metrics.

<https://hsibraindatabase.iuma.ulpgc.es/>. This experiment used a Jupyter notebook running on an Intel 11th Gen processor and 32 GB of RAM. For all experiments, the training, validation, and test samples distribution was set at 15%, 15%, and 70%, respectively. To ensure an equitable comparison, all models, including the RNN, 2D CNN, LeeNet, Xception, and proposed SCS models, were executed simultaneously with a single, randomly chosen set of samples. The reported results were achieved using a patch size of 3×3 , and the three most informative bands were identified through Factor Analysis (FA). Regarding training parameters, the models began with randomly initialized weights, which were subsequently optimized via backpropagation using the Adam optimizer and a softmax loss function. Figure 3 present a detailed analysis of the validation loss and accuracy for all models under consideration. In this study, we adhere to this principle by keeping these parameters uniform across all compared methods, including our SCS pipeline, within a single execution run.

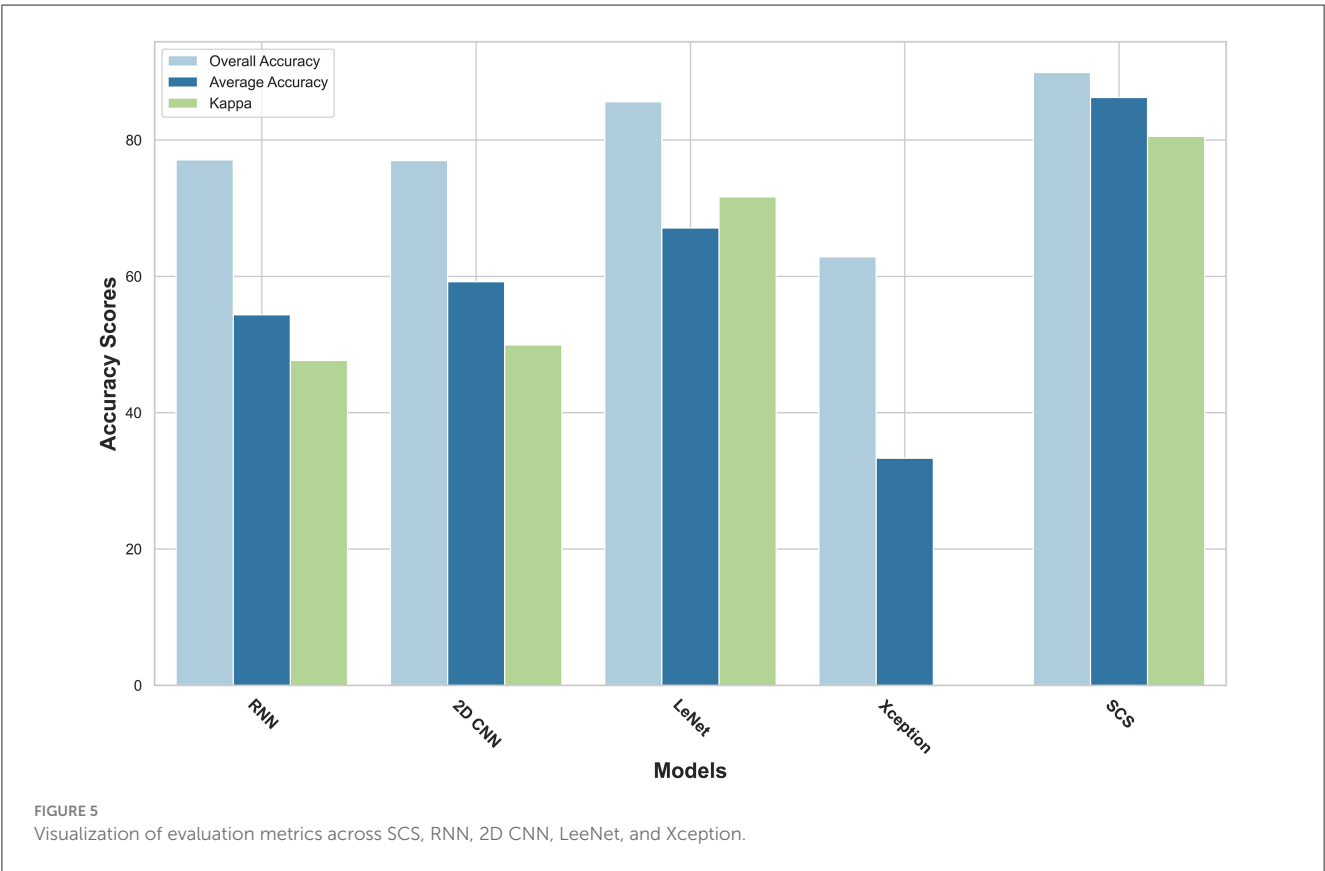
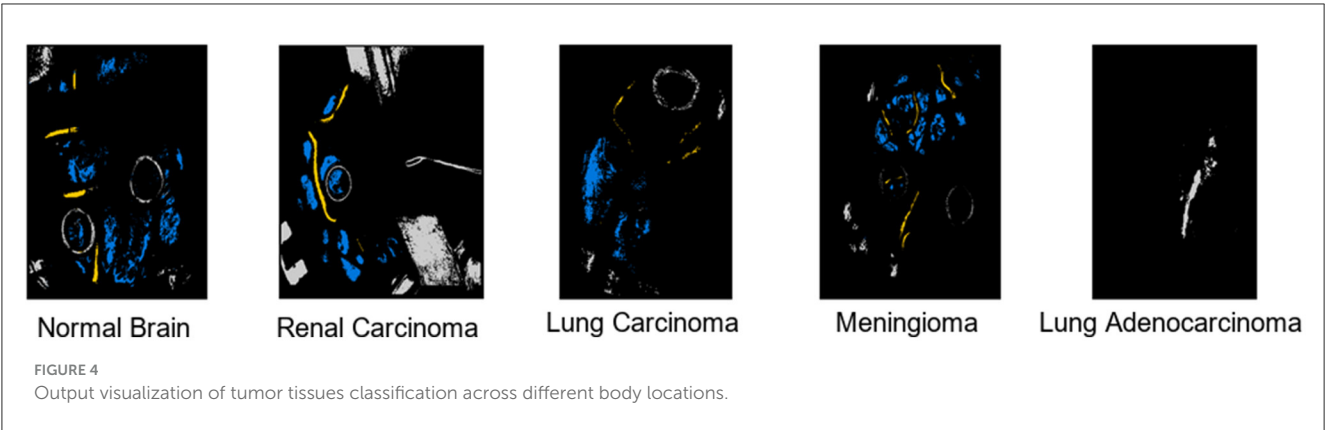
5 Discussion

In this section, we conduct a twofold comparative analysis to evaluate the performance of our SCS pipeline for the Hybrid Healthcare Unit. Firstly, at the patient level, we assess the system's efficacy in providing personalized tumor tissue classifications and treatment recommendations. Subsequently, at the same tumor class level, we analyze the system's precision in distinguishing minute variations within specific classes. These comparative experiments

aim to comprehensively understand the Hybrid Healthcare Unit's capabilities, addressing individual patient needs and the challenges within distinct tumor classes.

5.1 Comparative experiment—Class level

We present a comprehensive performance analysis based on the F1-Score, comparing SOTA, including 2D CNN, RNN, LeeNet, Xception, and our proposed SCS across different patients. The objective of this comparative experiment class level is to evaluate and compare the performance of these models in accurately classifying different tissue types in HI as results discussed in Table 2. Across different patients, our SCS consistently achieves high accuracy in predicting tissue classes, as presented in Table 3. Notably, for Patient IDs 005, 008, 022, 028, and 029, SCS achieves exceptional accuracy close to or at 100% in classifying normal tissue, hypervascularized tissue, and background classes. This demonstrates the model's robustness in handling diverse cases. In cases where tumor tissue is present, the SCS model also successfully achieves accurate predictions comparison with other models (2D CNN, RNN, LeeNet, and Xception 2D). The model's effectiveness in leveraging SCS-enhanced features for accurate tissue classification irrespective of body location are shown in classification map outputs (Figure 4). These sub-figures correspond to different tumor tissue types: Normal Brain, Renal Carcinoma, Lung Carcinoma, Meningioma, and Lung Adenocarcinoma. The model's ability to handle various tissue classes and consistent



accuracy across different patients and images highlight its potential as a valuable tool in medical diagnostics, particularly for tumor tissue classification.

5.2 Comparative experiment—Patient level

A detailed analysis of the performance of SOTA models on key metrics, including Kappa Accuracy, Overall Accuracy, Average Accuracy, F1-Score, Training Time, Testing Time, and Memory Consumption. Table 3 summarizes the performance metrics for each patient and their corresponding image IDs across various tissue classes. From the patient-level experiment, the SCS model consistently outperformed both models across multiple performance metrics. For instance, in Patient ID 004,

the SCS model achieved a Kappa Accuracy of 80.53, surpassing 2D CNN (49.93), RNN (47.66), LeeNet (71.67) and Xception (30.02). Similar trends were observed regarding Overall Accuracy, Average Accuracy, and F1 score, where the SCS model consistently demonstrated superior performance across all patient IDs. Notably, in Patient ID 021, the SCS model achieved a Kappa Accuracy of 98.55, significantly surpassing 2D CNN (85.12), RNN (45.74), LeeNet (97.83), and Xception (38.61). SCS model's ability to consistently attain high accuracy, coupled with efficient training times and memory consumption, underscores its potential for accurate tissue classification in HI data, highlighting its value in practical medical applications. Although other models such as 2D CNN, RNN, and LeeNet have less training time, their accuracy is low compared to the SCS model; as we know, in deep learning, there is a trade-off between speed and accuracy. Figure 5 shows results underscore the superior performance of the SCS model across

various metrics, indicating its efficacy in accurately classifying tissue types in HI data. The consistent out performance of the SCS model reaffirms its potential to enhance medical diagnostics and contribute to real-world applications.

6 Conclusion

Our research highlights the pivotal role of HI integrated with AI in advancing tumor tissue classification with the new Hybrid Health Care Units landscape. The innovative application of the Sharpened Cosine Similarity framework has proven highly effective, achieving remarkable performance metrics of 91.76% Cohen's kappa, 95.60% overall accuracy, and 94.29% f1-score. These results, surpassing current SOTA research even under limited training data, affirm our proposed model's robustness and potential clinical impact. The scarcity of specific hyperspectral medical data has been acknowledged as a challenge, emphasizing the need for ongoing efforts to expand and diversify datasets for further validation and generalization of our approach. However, the demonstrated superiority of our model in tumor classification positions it as a valuable tool for enhancing diagnostic capabilities in medical imaging. Future research could extend the proposed model by diversifying and expanding hyperspectral medical datasets for broader validation. Exploring real-time implementation in clinical settings and investigating additional AI techniques could enhance predictive capabilities. Furthermore, exploring broader applications beyond tumor classification, such as skin conditions, could maximize the model's utility. These efforts would advance healthcare informatics, improving diagnostic accuracy within Hybrid Health Care Units.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://hsibraindatabase.iuma.ulpgc.es>.

Author contributions

MB: Visualization, Methodology, Conceptualization, Writing—review & editing, Writing—original draft. JL: Writing—review & editing, Supervision. JJ: Writing—review & editing, Supervision. WR: Writing—original draft, Funding acquisition. NA: Writing—original draft, Methodology, Formal analysis. FB: Writing—original draft, Validation, Data curation. MA: Writing—review & editing, Methodology, Conceptualization. AS: Writing—original draft, Visualization. AA: Writing—review

& editing, Formal analysis, Validation. MU: Writing—review & editing, Supervision, Validation.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The research conducted for this paper has been made possible through the generous support of several funding sources. Specifically, we acknowledge the stable support project of Shenzhen colleges and universities (20220811224429001), the Post-doctoral Later-stage Foundation Project of Shenzhen Polytechnic (6021271014K), and the Shenzhen Polytechnic Scientific Research Start-up Project (6022312028K). Furthermore, our work has been supported by the National Science Foundation of Guangdong Province (2022A1515011724) and the Featured Innovation Project of Guangdong Provincial Department of Education (2021KTSCX275). These funding sources have played a crucial role in enabling our research endeavors and advancing our understanding in the field of healthcare informatics.

Acknowledgments

The authors would like to express their sincere appreciation to FF Butt, a dedicated medical professional at NESCOM Hospital in Islamabad, Pakistan. FF Butt played a pivotal role in providing invaluable insights into the domain knowledge concerning various types of tumor tissues within the human body. Her expertise extended to a comprehensive understanding of the existing medical imaging modalities utilized for tumor tissue classification, along with a thorough grasp of the historical and background aspects. The authors acknowledge FF Butt's significant contributions, which greatly enhanced the depth and quality of this research.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Bashkatov AN, Genina EA, Tuchin VV. Optical properties of skin, subcutaneous, and muscle tissues: a review. *J Innov Opt Health Sci*. (2011) 4:9–38. doi: 10.1142/S1793545811001319
2. Kroemer G, Pouyssegur J. Tumor cell metabolism: cancer's Achilles' heel. *Cancer Cell*. (2008) 13:472–82. doi: 10.1016/j.ccr.2008.05.005

3. Aboughaleb IH, Aref MH, El-Sharkawy YH. Hyperspectral imaging for diagnosis and detection of *ex-vivo* breast cancer. *Photodiagnosis Photodyn Ther.* (2020) 31:101922. doi: 10.1016/j.pdpdt.2020.101922
4. Zhang Y, Yu S, Zhu X, Ning X, Liu W, Wang C, et al. Explainable liver tumor delineation in surgical specimens using hyperspectral imaging and deep learning. *Biomed Opt Express.* (2021) 12:4510–29. doi: 10.1364/BOE.432654
5. Ortega S, Halicek M, Fabelo H, Camacho R, Plaza MdL, Godtliebsen F, et al. Hyperspectral imaging for the detection of glioblastoma tumor cells in H&E slides using convolutional neural networks. *Sensors.* (2020) 20:1911. doi: 10.3390/s20071911
6. Lv M, Chen T, Yang Y, Tu T, Zhang N, Li W, et al. Membranous nephropathy classification using microscopic hyperspectral imaging and tensor patch-based discriminative linear regression. *Biomed Opt Express.* (2021) 12:2968–78. doi: 10.1364/BOE.421345
7. Liu N, Guo Y, Jiang H, Yi W. Gastric cancer diagnosis using hyperspectral imaging with principal component analysis and spectral angle mapper. *J Biomed Opt.* (2020) 25:066005. doi: 10.1117/1.JBO.25.6.066005
8. Halicek M, Dormer JD, Little JV, Chen AY, Myers L, Sumer BD, et al. Hyperspectral imaging of head and neck squamous cell carcinoma for cancer margin detection in surgical specimens from 102 patients using deep learning. *Cancers.* (2019) 11:1367. doi: 10.3390/cancers11091367
9. Halicek M, Dormer JD, Little JV, Chen AY, Fei B. Tumor detection of the thyroid and salivary glands using hyperspectral imaging and deep learning. *Biomed Opt Express.* (2020) 11:1383–400. doi: 10.1364/BOE.381257
10. Kho E. Hyperspectral Imaging for Resection Margin Assessment during Breast Cancer Surgery [PhD Thesis - Research external, graduation UT]. University of Twente, Enschede (2020).
11. Huang Q, Li W, Zhang B, Li Q, Tao R, Lovell NH. Blood cell classification based on hyperspectral imaging with modulated Gabor and CNN. *IEEE J Biomed Health Inform.* (2019) 24:160–70. doi: 10.1109/JBHI.2019.2905623
12. Van der Leest M, Israel B, Cornel EB, Zamecnik P, Schoots IG, van der Lelij H, et al. High diagnostic performance of short magnetic resonance imaging protocols for prostate cancer detection in biopsy-naïve men: the next step in magnetic resonance imaging accessibility. *Eur Urol.* (2019) 76:574–81. doi: 10.1016/j.eururo.2019.05.029
13. Huang S, Yang J, Shen N, Xu Q, Zhao Q. Artificial intelligence in lung cancer diagnosis and prognosis: current application and future perspective. *Semin Cancer Biol.* (2023) 89:30–7. doi: 10.1016/j.semcancer.2023.01.006
14. Pulumati A, Pulumati A, Dwarakanath BS, Verma A, Papineni RV. Technological advancements in cancer diagnostics: improvements and limitations. *Cancer Rep.* (2023) 6:e1764. doi: 10.1002/cnr.2.1764
15. King AD, Thoeny HC. Functional MRI for the prediction of treatment response in head and neck squamous cell carcinoma: potential and limitations. *Cancer Imaging.* (2016) 16:1–8. doi: 10.1186/s40644-016-0080-6
16. Faghihi R, Zeinali-Rafsanjani B, Mosleh-Shirazi MA, Saeedi-Moghadam M, Lotfi M, Jalli R, et al. Magnetic resonance spectroscopy and its clinical applications: a review. *J Med Imaging Radiat Sci.* (2017) 48:233–53. doi: 10.1016/j.jmir.2017.06.004
17. Karim S, Qadir A, Farooq U, Shakir M, Laghari AA. Hyperspectral imaging: a review and trends towards medical imaging. *Curr Med Imaging.* (2023) 19:417–27. doi: 10.2174/1573405618666220519144358
18. Tobore I, Li J, Yuhang L, Al-Handarish Y, Kandwal A, Nie Z, et al. Deep learning intervention for health care challenges: some biomedical domain considerations. *JMIR mHealth uHealth.* (2019) 7:e11966. doi: 10.2196/11966
19. Gull S, Akbar S. Artificial intelligence in brain tumor detection through MRI scans: advancements and challenges. In: Mohan Goyal L, Saba T, Rehman A, Larabi-Marie-Sainte S, editors. *Artificial Intelligence and Internet of Things*. Boca Raton, FL: CRC Press (2021), p. 241–76. doi: 10.1201/9781003097204-10
20. Demehri S, Baffour FI, Klein JG, Ghotbi E, Ibad HA, Moradi K, et al. Musculoskeletal CT imaging: state-of-the-art advancements and future directions. *Radiology.* (2023) 308:e230344. doi: 10.1148/radiol.230344
21. Islam MZ, Naqvi RA, Haider A, Kim HS. Deep learning for automatic tumor lesions delineation and prognostic assessment in multi-modality PET/CT: a prospective survey. *Eng Appl Artif Intell.* (2023) 123:106276. doi: 10.1016/j.engappai.2023.106276
22. Pasquini L, Peck KK, Jenabi M, Holodny A. Functional MRI in neuro-oncology: state of the art and future directions. *Radiology.* (2023) 308:e222028. doi: 10.1148/radiol.222028
23. Seelen LW, van den Wildenberg L, van der Kemp WJ, Mohamed Hoessein FA, Mohammad NH, Molenaar IQ, et al. Prospective of 31P MR spectroscopy in hepatopancreatobiliary cancer: a systematic review of the literature. *J Magn Reson Imaging.* (2023) 57:1144–55. doi: 10.1002/jmri.28372
24. Pecorino L. *Molecular Biology of Cancer: Mechanisms, Targets, and Therapeutics*. Oxford: Oxford University Press. (2021). doi: 10.1093/hesc/9780198833024.001.0001
25. Pérez-González A, Bévant K, Blanpain C. Cancer cell plasticity during tumor progression, metastasis and response to therapy. *Nat Cancer.* (2023) 4:1063–82. doi: 10.1038/s43018-023-00595-y
26. Carbone A. Cancer classification at the crossroads. *Cancers.* (2020) 12:980. doi: 10.3390/cancers12040980
27. Choi JH, Ro JY. The 2020 WHO classification of tumors of bone: an updated review. *Adv Anat Pathol.* (2021) 28:119–38. doi: 10.1097/PAP.0000000000000293
28. Gress DM, Edge SB, Greene FL, Washington MK, Asare EA, Brierley JD, et al. Principles of cancer staging. In: Amin MB, Edge S, Greene F, Byrd DR, Brook-land RK, Washington MK, et al., editors. *AJCC Cancer Staging Manual*, 8th edition. New York, NY: Springer (2017), p. 3–30.
29. Thierheimer M, Cioffi G, Waite KA, Kruchko C, Ostrom QT, Barnholtz-Sloan JS. Mortality trends in primary malignant brain and central nervous system tumors vary by histopathology, age, race, and sex. *J Neurooncol.* (2023) 162:167–77. doi: 10.1007/s11060-023-04279-6
30. Tariq MR, Ali SW, Fatima N, Jabeen A, Qazi AS, Hameed A, et al. Radiation therapies in cancer. In: Saleem Qazi A, Tariq K, editors. *Therapeutic Approaches in Cancer Treatment*. New York, NY: Springer (2023), 59–77. doi: 10.1007/978-3-031-27156-4_4
31. Triulzi FM, editor. Gliomas, glioneuronal tumors and neuronal tumors: ependymaltumors. choroid plexus tumors. In: *Neuroradiology of Brain Tumors: Practical Guide based on the 5th Edition of WHO Classification*. New York, NY: Springer; 2023. p. 125–137. doi: 10.1007/978-3-031-38153-9_5
32. Gillespie C, Pettorini B. Central nervous system tumors. In: Losty PD, La Quaglia M, Sarnacki S, Fuchs J, Taguchi T, editors. *Pediatric Surgical Oncology*. Boca Raton, FL: CRC Press (2022), p. 272–312. doi: 10.1201/9781351166126-28
33. Saboor A, Li JP, Ul Haq A, Haq A, Rehman SU. Lung cancer accurate detection employing machine learning stacking approach and relief algorithm for feature selection using clinical data. In: *2023 20th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. Chengdu (2023), p. 1–10. doi: 10.1109/ICCWAMTIP60502.2023.10387125
34. Schuller HM. The impact of smoking and the influence of other factors on lung cancer. *Expert Rev Respir Med.* (2019) 13:761–9. doi: 10.1080/17476348.2019.1645010
35. Siringo M, Baena J, Bote de Cabo H, Torres-Jiménez J, Zurera M, Zugazagoitia J, et al. Future perspectives in the second line therapeutic setting for non-oncogene addicted non-small cell lung cancer. *Cancers.* (2023) 15:5505. doi: 10.3390/cancers15235505
36. Yeong J, Thike AA, Tan PH, Iqbal J. Identifying progression predictors of breast ductal carcinoma in situ. *J Clin Pathol.* (2017) 70:102–8. doi: 10.1136/jclinpath-2016-204154
37. Agbley BLY, Li JP, Haq AU, Bankas EK, Mawuli CB, Ahmad S, et al. Federated fusion of magnified histopathological images for breast tumor classification in the internet of medical things. *IEEE J Biomed Health Inform.* (2023) 1–12. doi: 10.1109/JBHI.2023.3256974
38. Solomon DA, Pekmezci M. Pathology of meningiomas. In: McDermott MW, editor. *Handbook of Clinical Neurology*, Vol. 169. Amsterdam: Elsevier (2020), p. 87–99. doi: 10.1016/B978-0-12-804280-9.00005-6
39. Zhang Y, Wu X, He L, Meng C, Du S, Bao J, et al. Applications of hyperspectral imaging in the detection and diagnosis of solid tumors. *Transl Cancer Res.* (2020) 9:1265. doi: 10.21037/tcr.2019.12.53
40. Fabelo H, Ortega S, Szolna A, Bulters D, Piñeiro JF, Kabwama S, et al. *In-vivo* hyperspectral human brain image database for brain cancer detection. *IEEE Access.* (2019) 7:39098–116. doi: 10.1109/ACCESS.2019.2904788
41. Mou L, Ghamisi P, Zhu XX. Deep recurrent neural networks for hyperspectral image classification. *IEEE Trans Geosci Remote Sens.* (2017) 55:3639–55. doi: 10.1109/TGRS.2016.2636241
42. Cao X, Yao J, Xu Z, Meng D. Hyperspectral image classification with convolutional neural network and active learning. *IEEE Trans Geosci Remote Sens.* (2020) 58:4604–16. doi: 10.1109/TGRS.2020.2964627
43. Verdhann V, Verdhann V. Image classification using LeNet. In: *Computer vision Using Deep learning: Neural Network Architectures with Python and Keras*. Berkeley, CA: Apress (2021), p. 67–101. doi: 10.1007/978-1-4842-6616-8_3
44. Banumathi J, Muthumari A, Dhanasekaran S, Rajasekaran S, Pustokhina IV, Pustokhin DA, et al. An Intelligent deep learning based Xception model for hyperspectral image analysis and classification. *Comput Mater Contin.* (2021) 67:2393–407. doi: 10.32604/cmc.2021.015605



OPEN ACCESS

EDITED BY

Amin Ul Haq,
University of Electronic Science and
Technology of China, China

REVIEWED BY

Lakshmeeswari Gondi,
Gandhi Institute of Technology and
Management (GITAM), India
Saeed Mian Qaisar,
Effat University, Saudi Arabia
Kishore Konda Chidella,
University of Nevada, Las Vegas, United States
J. Mohana Lakshmi,
Malnad College of Engineering, India

*CORRESPONDENCE

Shayla Islam
✉ shayla@ucsiuniversity.edu.my
Chan Yeob Yeun
✉ chan.yeun@ku.ac.ae

RECEIVED 26 March 2024

ACCEPTED 20 May 2024

PUBLISHED 05 July 2024

CITATION

Praveen SP, Hasan MK, Abdullah SNHS,
Sirisha U, Tirumanadham NSMKM, Islam S,
Ahmed FRA, Ahmed TE, Noboni AA,
Sampedro GA, Yeun CY and Ghazal TM (2024)
Enhanced feature selection and ensemble
learning for cardiovascular disease prediction:
hybrid GOL2-2T and adaptive boosted
decision fusion with babysitting refinement.
Front. Med. 11:1407376.
doi: 10.3389/fmed.2024.1407376

COPYRIGHT

© 2024 Praveen, Hasan, Abdullah, Sirisha,
Tirumanadham, Islam, Ahmed, Ahmed,
Noboni, Sampedro, Yeun and Ghazal. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Enhanced feature selection and ensemble learning for cardiovascular disease prediction: hybrid GOL2-2T and adaptive boosted decision fusion with babysitting refinement

S. Phani Praveen¹, Mohammad Kamrul Hasan²,
Siti Norul Huda Sheikh Abdullah², Uddagiri Sirisha¹,
N. S. Koti Mani Kumar Tirumanadham³, Shayla Islam^{4*},
Fatima Rayan Awad Ahmed⁵, Thowiba E. Ahmed⁶,
Ayman Afrin Noboni⁷, Gabriel Avelino Sampedro^{8,9},
Chan Yeob Yeun^{10*} and Taher M. Ghazal²

¹Department of Computer Science and Engineering, Prasad V Potluri Siddhartha Institute of Technology, Vijayawada, India, ²Faculty of Information Science and Technology, University Kebangsaan Malaysia, Bangi, Selangor, Malaysia, ³Department of Computer Science and Engineering, Sir C R Reddy College of Engineering, Eluru, India, ⁴Institute of Computer Science and Digital innovation, UCSI University, Kuala Lumpur, Malaysia, ⁵Computer Science Department, College of Computer Engineering and Science, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia, ⁶Computer Science Department, College of Science and Humanities-Jubail, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia, ⁷Department of Surgery, Medical College For Women and Hospital, Dhaka, Bangladesh, ⁸Faculty of Information and Communication Studies, University of the Philippines Open University, Los Baños, Philippines, ⁹Center for Computational Imaging and Visual Innovations, De La Salle University, Manila, Philippines, ¹⁰Centre for Cyber Physical Systems, Computer Science Department, Khalifa University, Abu Dhabi, United Arab Emirates

Introduction: Global Cardiovascular disease (CVD) is still one of the leading causes of death and requires the enhancement of diagnostic methods for the effective detection of early signs and prediction of the disease outcomes. The current diagnostic tools are cumbersome and imprecise especially with complex diseases, thus emphasizing the incorporation of new machine learning applications in differential diagnosis.

Methods: This paper presents a new machine learning approach that uses MICE for mitigating missing data, the IQR for handling outliers and SMOTE to address first imbalance distance. Additionally, to select optimal features, we introduce the Hybrid 2-Tier Grasshopper Optimization with L2 regularization methodology which we call GOL2-2T. One of the promising methods to improve the predictive modelling is an Adaboost decision fusion (ABDF) ensemble learning algorithm with babysitting technique implemented for the hyperparameters tuning. The accuracy, recall, and AUC score will be considered as the measures for assessing the model.

Results: On the results, our heart disease prediction model yielded an accuracy of 83.0%, and a balanced F1 score of 84.0%. The integration of SMOTE, IQR outlier detection, MICE, and GOL2-2T feature selection enhances robustness while improving the predictive performance. ABDF removed the impurities in the model and elaborated its effectiveness, which proved to be high on predicting the heart disease.

Discussion: These findings demonstrate the effectiveness of additional machine learning methodologies in medical diagnostics, including early recognition improvements and trustworthy tools for clinicians. But yes, the model's use and

extent of work depends on the dataset used for it really. Further work is needed to replicate the model across different datasets and samples: as for most models, it will be important to see if the results are generalizable to populations that are not representative of the patient population that was used for the current study.

KEYWORDS

multivariate imputation by chained equations, synthetic minority over-sampling technique, interquartile range, adaptive boosted decision fusion, cardiovascular disease and adaboost decision fusion (ABDF)

1 Introduction

Many communities are affected by heart disease, a major global health problem that is responsible for many cases of sickness and death. There is an increasing need to understand the complexity of heart diseases as our understanding of cardiovascular health expands. Think about this: someone dies of cardiovascular issues every 37 s in America, which highlights the urgency to quell this unseen epidemic (American Heart Association, 2022). This mind-boggling figure shows how huge numbers of people, families, and societies are affected by cardiac diseases (1).

The human heart is one fantastic example of biologically engineered machinery that coordinates life's intricate workings by driving vital energy through a network of complex vessels. However, repercussions can be disastrous when this symphony gets disrupted. Heart problems include conditions like coronary artery disease, heart failure, arrhythmias and congenital malformations. Their etiology is multifactorial involving genetic predispositions, behavioral factors and countless sophisticated biochemical pathways (2). Beyond the confines of medical practice, heart diseases contain a rich assortment of stories—chronicles of courage, sadness and hope. Every heartbeat affects those whose lives are touched by it and every diagnosis carries along its own path for each of them which are distinct and personal.

A major global health issue, cardiovascular disease, and cardiovascular disorders. Coronary artery disease (CAD), the most common, causes narrowing or blockage of the coronary arteries, leading to angina or myocardial infarction. Heart failure reduces oxygen delivery because the heart cannot pump blood properly. Mild exercise causes an abnormal heart rate that can impair circulation. Valvular heart disease damages the heart muscles and limits blood flow. Cardiomyopathy occurs when the heart muscle contracts or stiffens, reducing its ability to carry blood (3).

Poor diet, lack of physical activity, tobacco use, alcohol abuse and obesity are major risk factors. Heart disease prevention includes healthy eating, exercise, weight control, and smoking cessation. Treatment options range from medical to surgical, depending on the severity. Routine inspections detect and address them quickly (4). Knowing the risk factors and prioritizing cardiovascular health helps reduce the impact of cardiovascular disease.

Risk factors for cardiovascular disease include smoking and alcohol misuse. Coronary artery disease (5), hypertension, decreased oxygen saturation, and accelerated blood clotting are all consequences of smoking. Consuming alcohol raises the risk of hypertension, heart disease (6, 7), and cholesterol. When smoked and drunk at the same time, oxidative stress rises, the immune system is weakened, and

blood arteries and cholesterol are damaged. Heart disease, particularly myocardial, cerebral, and cardiac insufficiency, is greatly increased by this lethal combination. It is vital to quit smoking, restrict alcohol intake, and maintain cardiovascular health since these habits add up to a lot of harm. Although beating addiction could be difficult, the rewards in terms of heart health are substantial.

Adaptive enhanced decision fusion is crucial for disease prediction, especially in cardiovascular health. Combining numerous models and adjusting to changing data patterns enhances early disease detection and prediction. The ABDF educates doctors on cardiac illnesses to help them choose the best treatments and improve patient outcomes. In the complex realm of cardiovascular diseases, its versatility allows quick risk assessment and appropriate intervention. ABDF is a cutting-edge ensemble learning approach that enhances cardiovascular health patient care and predictive analytics.

As data reveals, the cardiovascular problem percentage among people in India as diagnosed in the year 2020 is shown in Figure 1, using the breakdown by age group. In cardiovascular matters, most often, the older age group was seen having more frequent problems than the younger age group. The rate of cardiovascular disease found among the teenagers of the age group below 19 is about 2.98%, which is comparably lower compared to that of the young people of the age group 20–29, which registers about 5%. Investigators have been able to ascertain that the 45- to 59-year-old population group had an illness rate of cardiovascular problems of about 11.9%, while that of the 30- to 44-year-old group was about 6.28%. At a rate of 18.7%, the above-60-year-old succession group accounts for the highest prevalence of cardiovascular diseases. Given the existence of age disparities, policymakers should focus on the development of auxiliary policies, early detection, effective healthcare delivery, and educational campaigns that will help in the ongoing battle against the rising frequency of cardiovascular diseases among the aging population (8–12).

2 Literature review

In 2020, Shah et al. (13) examine data mining and machine learning for heart disease prediction. The study stresses the need of precise and timely identification of heart disease, a top worldwide mortality. Using the enormous Cleveland database of UCI repository, 303 cases and 76 characteristics are rigorously condensed to 14 important elements. The study compares popular algorithms including Naïve Bayes, decision tree, K-nearest neighbor (KNN), and random forest for heart disease prediction. KNN was the most accurate algorithm, demonstrating predictive modeling potential. The finding agrees with earlier research

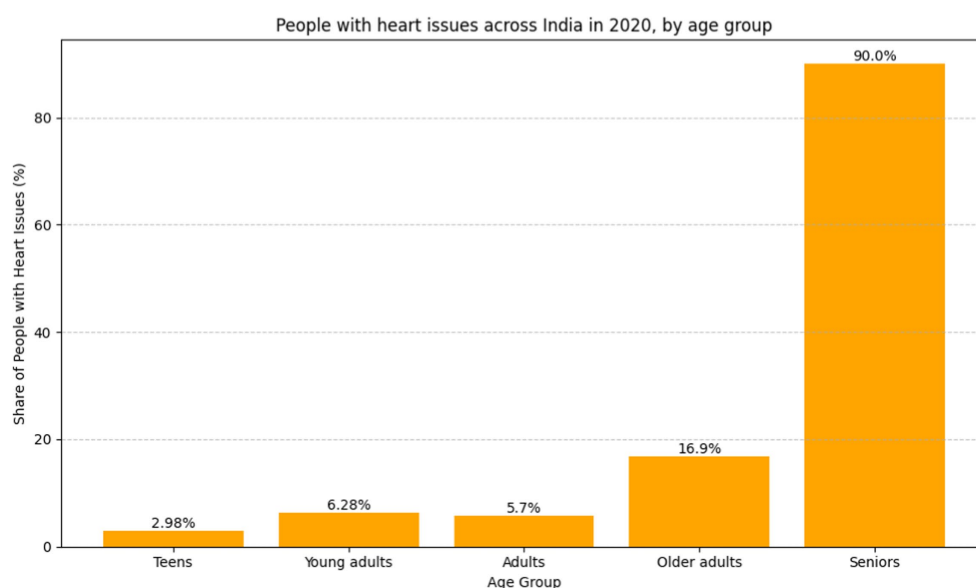


FIGURE 1

Bar graph people with heart issues across India in 2020, by age group (8) (<https://www.statista.com>).

that many algorithms are needed for complete findings. Future data mining approaches such time series analysis, clustering, association rules, support vector machines, and evolutionary algorithms are suggested to improve predicted accuracy. While insightful, the paper admits its limits and advocates for further research to improve early and accurate heart disease prediction algorithms.

In 2020, Katarya et al. (14) conducted a survey saying that heart disease is a global issue with rising treatment expenses, therefore early detection is essential. Alcohol, tobacco, and inactivity are essential heart disease indicators. The paper recommends using machine learning, particularly supervised methods, for healthcare decision-making and prediction to address this essential issue. Several algorithms, including as ANN, DT, RF, SVM, NB, and KNN, being investigated for heart disease prediction. The research summarizes these algorithms' performance to reveal their efficacy. In conclusion, automated technologies to anticipate cardiac disease early on help healthcare professionals diagnose and empower patients to monitor their health. Feature selection is critical, and hybrid grid search and random search are suggested for optimization. Search algorithms for feature selection and machine learning will improve cardiac disease prediction, leading to better healthcare treatments, according to the report.

In 2021, Jindal et al. (15) highlights the increasing number of heart diseases and the need for prediction models. The declaration acknowledges the challenge of correct diagnosis and promotes machine learning techniques for accurate projections. Logistic regression and KNN are compared to naive Bayes in the research. The proposed heart disease prediction system reduces costs and improves medical care. The research also includes a Logistic Regression, Random Forest Classifier, and KNN cardiovascular disease detection model. The model's accuracy is 87.5%, up from 85% for previous models. The literature shows that the KNN method outperforms other algorithms with an accuracy rate of 88.52%. The article claims that machine learning can predict cardiac issues more accurately than conventional techniques, improving patient care and lowering costs.

In 2019, Gonsalves et al. (16) uses Machine Learning (ML) approaches such as Naïve Bayes (NB), Support Vector Machine (SVM), and Decision Tree (DT) to predict Coronary Heart Disease (CHD). Coronary heart disease (CHD) is a major cause of death around the world, highlighting the need of early detection. The work uses historical medical data and three supervised learning approaches to discover CHD data correlations to improve prediction precision. The summary of the literature acknowledges the complexity of medical data and CHD prediction linkages, stressing the challenges of existing techniques. The study's focus on NB, SVM, and DT matches existing research techniques, highlighting the availability of disease prediction machine learning algorithms. Early screening and identification are crucial for patient well-being, resource allocation, and preventative interventions, according to the research. The discussion of ML model performance, including accuracy, sensitivity, specificity, and other characteristics, sheds light on Naive Bayes, Support Vector Machines, and Decision Trees. Despite not meeting threshold rates, the Naive Bayes (NB) classifier looks to be the best option for the dataset. According to the literature review, unsupervised learning and data imbalance should be studied in the future. This will enhance prediction algorithms and may lead to mobile CHD diagnosis apps.

In 2018, Nashif et al. (17), addresses cardiovascular problems across the globe and highlights the necessity to detect and monitor them early. The cloud-based heart disease prediction system uses powerful machine learning. Interestingly, the Support Vector Machine (SVM) method has 97.53% accuracy. Real-time patient monitoring using Arduino for data collection is presented in the study, focusing on remote healthcare. Comparative evaluations show SVM outperforms other models. The abstract concludes with potential issues including photoplethysmography-based blood pressure monitoring. The literature analysis highlights cloud-based prediction and real-time patient monitoring as a solution to PPG-based system constraints.

In 2023, Bhatt et al. (18) used Machine learning to create a cardiovascular disease prediction model. The study employed 70,000 Kaggle-downloaded real-world samples. Huang initialization improves k-modes clustering classification accuracy. GridSearchCV optimizes random forest, decision tree, multilayer perceptron, and XGBoost models. With 86.37 to 87.28% accuracy, the models are great. Multiple layer perceptron outperforms other models. The study adjusts age, blood pressure, and gender to account for heart disease progression. Despite promising outcomes, the study had limitations. These include employing a single dataset, only considering particular clinical and demographic features, and not comparing results to other test datasets. More research is needed to overcome these restrictions, compare clustering algorithms, test the model on new data, and improves interpretability. Machine learning—particularly clustering algorithms—can effectively predict cardiac illness and guide focused treatment and diagnostic measures.

In 2023, Abood Kadhim et al. (19) examines the growing use of artificial intelligence—specifically machine learning—in cardiac disease diagnosis and prediction. Support vector machines, random forests, and logistic regression are tested on Cleveland Clinic data. Research on artificial intelligence in cardiac care is also examined. The study found that support vector machines are the most accurate heart disease diagnosis tools at 96%. It also presents a 95.4% accurate random forest model for cardiac attacks. The findings demonstrate the importance of AI in healthcare decision-making and early cardiac problem intervention.

Recent researches have stressed the need for global cardiovascular disease diagnosis and identification. Several papers in 2020 and 2021 studied Naïve Bayes, decision tree, K-nearest neighbor (KNN), and random forest algorithms using data mining and machine learning methods. The primary findings are that K-Nearest Neighbors (KNN) may predict heart disease, that supervised machine learning may make healthcare decisions, and that logistic regression, KNN, and naïve Bayes are comparable. These findings show the usefulness of

predictive models in addressing the rising number of cardiac ailments, leading to healthcare technology advances for early identification and better patient treatment (Figure 2).

2.1 Motivation

Due to the global the amount of cardiovascular diseases, data mining and machine learning research on heart disease prediction is escalating. Heart disease is the most common cause of mortality worldwide. To reduce mortality rates, these medical conditions must be accurately and quickly detected. Researchers are studying machine learning to improve diagnostic skills since conventional methods frequently make inaccurate predictions. These studies aim to enhance early diagnosis and treatment. Medical data is complex and risk variables change, making machine learning an intriguing method for finding meaningful patterns and improving heart disease prediction.

2.2 Research gap

Despite the wealth of knowledge in machine learning approaches to heart disease prediction, additional research is needed. Shah et al. (13), Katarya et al. (14), Jindal et al. (15), Gonsalves et al. (16), Nashif et al. (17), Bhatt et al. (18), and Abood Kadhim et al. (19) all emphasize the importance of accurate and early heart disease detection. These researches have examined how K-nearest neighbor (KNN), Support Vector Machine (SVM), Random Forest, and logistic regression can increase predicted accuracy. These attempts are intriguing, but they also highlight limits like dataset dependence, feature selection optimization issues, and the need for more unsupervised learning research. Address data imbalance and real-time patient monitoring equipment concerns. Thus, even though machine learning could change cardiac illness prediction, more research is needed to improve algorithms, overcome

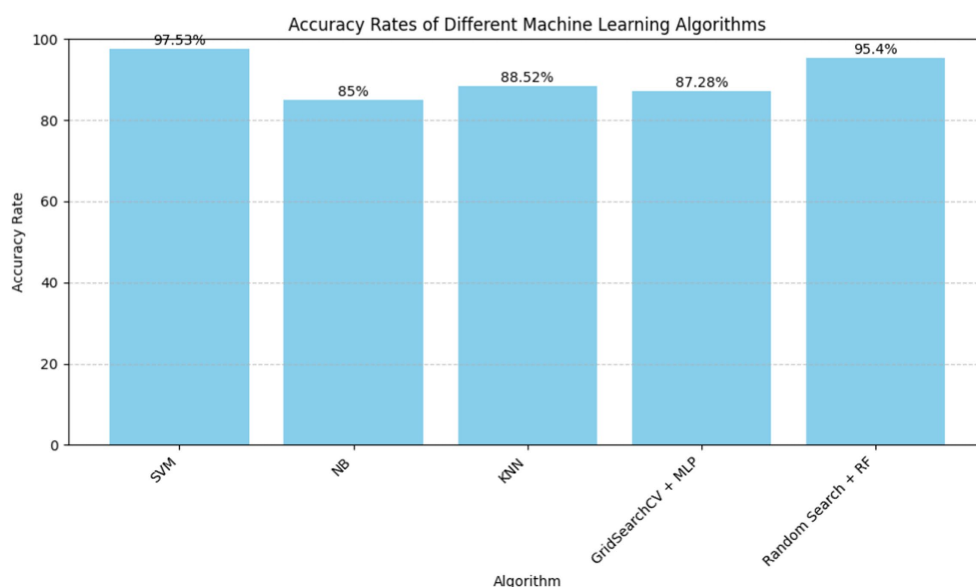


FIGURE 2
Machine learning algorithms for heart disease prediction.

data constraints, and improve cardiovascular health care outcomes. The current study lacks detailed algorithm assessments, leaving the best technique for exact predictions unknown. There is also insufficient research into using advanced data mining methods like time series analysis and evolutionary algorithms to better forecast heart illness. Overcome these gaps to increase prediction model robustness and precision in this critical healthcare sector.

The research's scope is to create trustworthy and effective cardiovascular disease diagnostic tools. Our goal is to reduce heart disease deaths and improve heart disease predictions using powerful machine learning.

- SMOTE, IQR outlier identification, and MICE are used to solve data difficulties in this work. We also introduce Hybrid GOL2-2T, a hybrid feature selection approach.
- It uses L2 regularization and the Grasshopper Optimization Algorithm.
- A babysitter algorithm and Adaptive Boosted Decision Fusion (ABDF) ensemble learning increase predictive modeling accuracy.
- Our model will be assessed by accuracy, recall, and AUC score.

The main goal of this project is to develop reliable diagnostic tools for early diagnosis and treatment of cardiovascular diseases. This can help doctors improve patient outcomes and reduce illness.

In the subsequent sections, Section 2 provides a comprehensive literature analysis of the corpus of recent publications. The suggested methodology is then presented in Section 3. Section 4 offers a thorough summary of the results and the discussion that follows. In Section 5, prospective avenues for further research are explored and the article is summarized with a conclusion.

3 Proposed methodology

For the two-tier Feature Selection Hybrid GOL2-2T, starting from the data pre-processing stage among the partitions, 70% of the data partition is allotted for the training set and 30% for the testing set. An objective under this category makes it easy to evaluate the performance of the models in question based on it deeply. The second to the last step is the missing data estimate, which makes use of the Multivariate Imputation by Chained Equations (MICE) approach. This, in return, ensures the completeness of information from one or many variables. In this case, the following techniques were corrected with a deficiency of training the model and have high interoperability with the techniques of machine learning; Imputation, Data scaling, and Label encoding. Inside the method, it has the Inter Quartile Range (IQR) to identify and deal with an outlier in an effort to enhance the resilience of the model through a reduction in influence that emanates from abnormal data points. The major maxim is SMOTE, which a synthetic minority is over-sampling technique aimed at the problem of class imbalance. The technique established a fair representation through the development of synthetic minorities, toward the reduction of biases that may associate with the general over-representation of the dominant class.

2-tier Feature Selection is based on the L2 Regularization (Ridge) (20) along with the Grasshopper Optimization (GOA) method; therefore, the proposed Hybrid GOL2-2T model is going to form a 2-level model for Feature Selection. It also employs ABDF hyperparameters, which have been babysitting algorithm to be fine-tuned after proper pre-processing of the dataset. Therefore, AdaBoost Decision Fusion (ABDF) maximizes the predictive modeling tasks' accuracies by pulling the performance measures out with respect to other models for comparison (Figure 3).

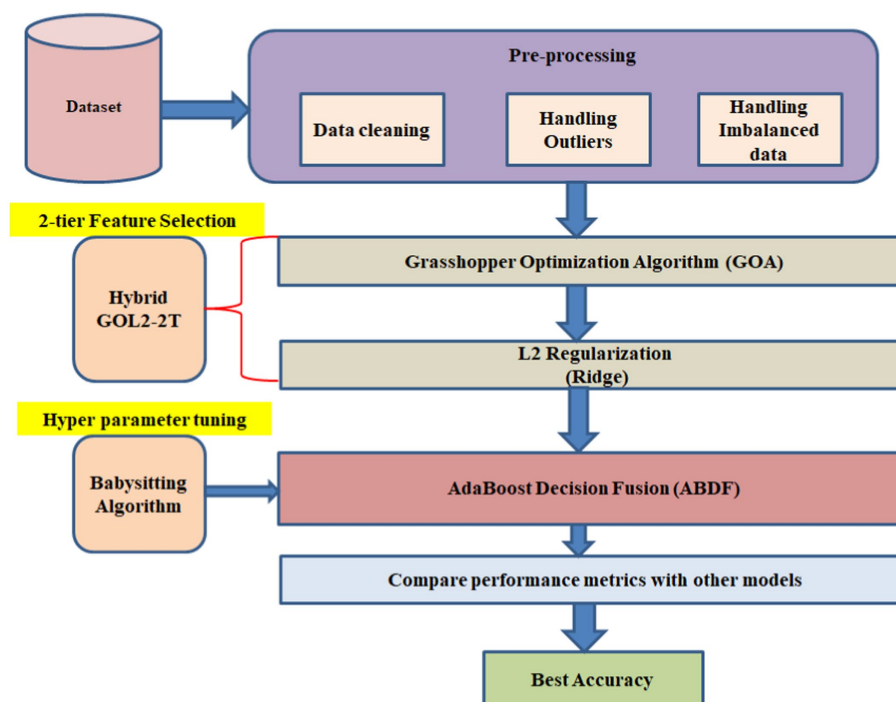
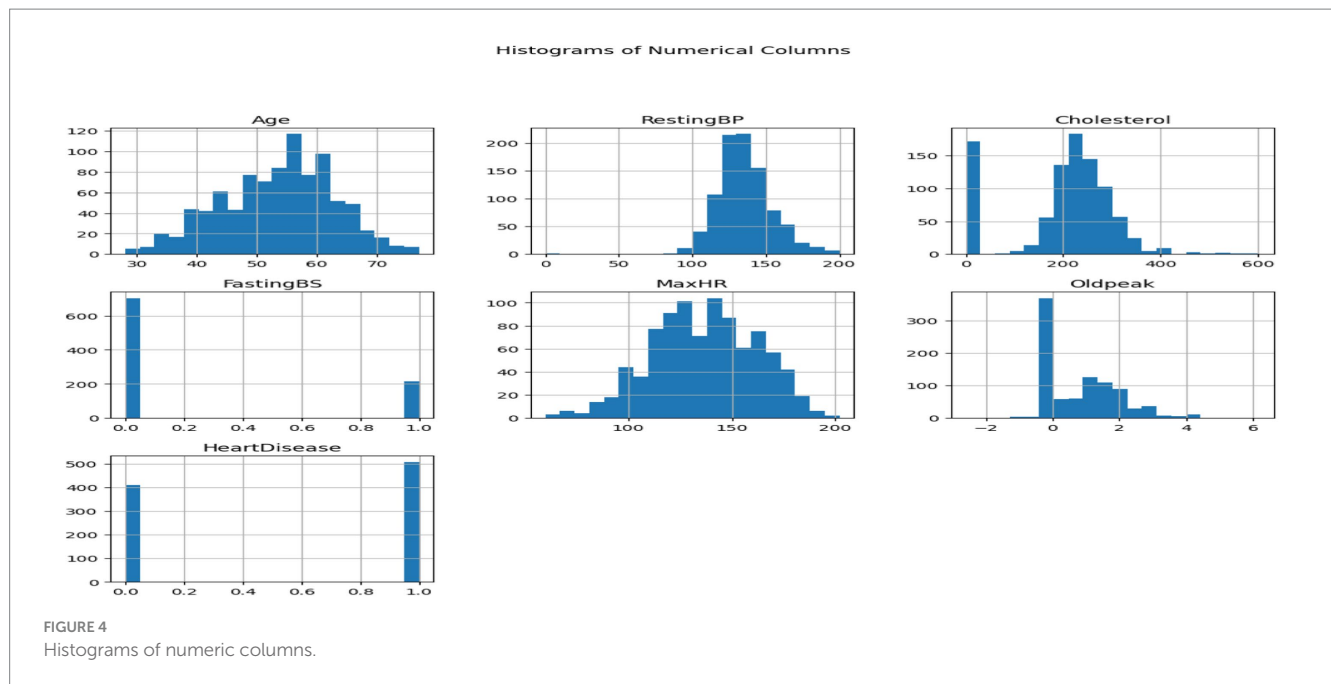


FIGURE 3
Heart disease forecasting workflow.



3.1 Data collection

The 1988 heart disease dataset (21) is an excellent resource for studying and forecasting cardiovascular disease prevalence. Age, gender, type of chest pain, blood pressure, cholesterol levels, and the presence of numerous cardiovascular diseases are among the 14 important factors. There is a large variety of ages represented in the dataset, with the majority falling between 40 and 60. Of those, 207 are male and 96 are female. With a value of 1 for males and 0 for females, the variable “sex” is included in the data for each issue as an essential health indicator. While we display resting blood pressure (trestbps) and serum cholesterol levels (chol) as whole numbers, we categorize chest discomfort as 1, 2, 3, or 0. Exang, exercise-induced angina, exercise-induced ST depression compared to rest, the slope of the peak exercise ST segment, the number of major vessels colored by fluoroscopy, and thalassemia type are some other factors that improve the dataset. In order to promote a thorough study of cardiovascular health and facilitate the development of reliable prediction systems, the “target” property shows whether heart disease is present (1) or absent (0) (Figure 4).

3.1.1 Visualizing the attributes of heart disease dataset using pair plot

This dataset encompasses six numerical variables: RestingBP, Cholesterol, FastingBS, MaxHR, Oldpeak. Two variables are distributed in each grid subplot. Variable correlations in the Heart Disease dataset are shown in the pair plot. The correlation between two variables is displayed in every matrix scatterplot. The level of heart disease dictates the color of the dots. Early detection of data patterns and trends can be aided by this. It can reveal whether there are commonalities between those who have cardiac disease and those who do not.

Histograms show variable distribution, while scatter plots show the connection between paired variables. In the upper left subplot, RestingBP distribution is presented. The y-axis shows data point frequency, and the x-axis shows RestingBP levels. The bottom right

subplot displays the association between MaxHR and Oldpeak, an off-diagonal plot. This subplot shows Oldpeak on the y-axis and MaxHR on the x-axis. Examining the pair plot can reveal patterns and linkages, such as cholesterol-resting blood pressure correlations. This graphical tool simplifies dataset analysis, especially for outliers and linear correlations. We consider non-diagonal scatter plots while examining linear relationships. Straight lines between scatter plot dots indicate the variables’ direction and strength. Outliers are scatter plot data points far from the main cluster. If we want to use machine learning to forecast cardiac disease from patient data, we need to understand these tendencies. It might be necessary to make adjustments and do further research on visual representations in order to have a better understanding of the dataset (Figure 5).

3.2 Pre-processing

3.2.1 Data cleaning with MICE

Data pretreatment requirements include cleaning the data to ensure dataset correctness and completeness and that it is analysis or model training ready. Absent data often hurts machine learning models. MICE (22) handle missing data thoroughly and statistically through Multiple Imputation by Chained Equations shown in equation (1). In an iterative process, MICE calculate conditional distributions for all variables with missing data using observed data and other variable imputations. As iterations continue until convergence, the process creates various entire datasets. To accommodate for missing value uncertainty, each dataset has its own imputations. Multiple Imputation by Chained Equations (MICE) works well for non-random missing data patterns in real-world datasets where observed values may affect missing. It evaluates variables and predicts data distributions. The MICE technique provides imputations, updates models, and combines findings to provide credible imputed datasets. Finalized datasets can be used to train machine-learning models. MICE address

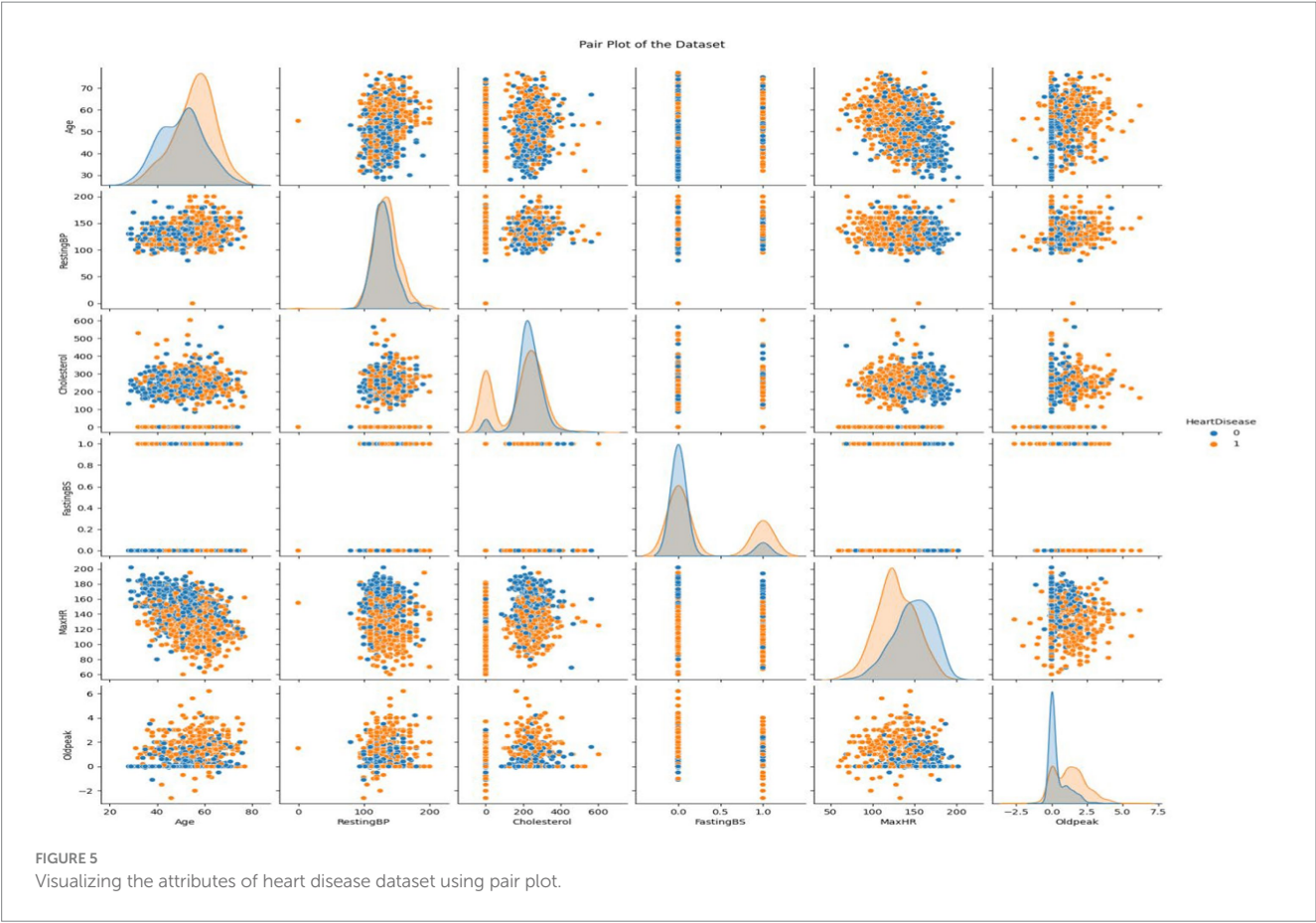


TABLE 1 Machine learning algorithms for heart disease prediction.

Algorithm	Accuracy rate	Citation
SVM	97.53	Nashif et al. (17)
NB	85	Gonsalves et al. (16)
KNN	88.52	Jindal et al. (15)
KNN	90.78	Shah et al. (13)
GridSearchCV + MLP	87.28	Bhatt et al. (18)
Random Search + RF	95.4	Abood Kadhim et al. (19)

missing data to improve model performance and assure unbiased parameter estimates.

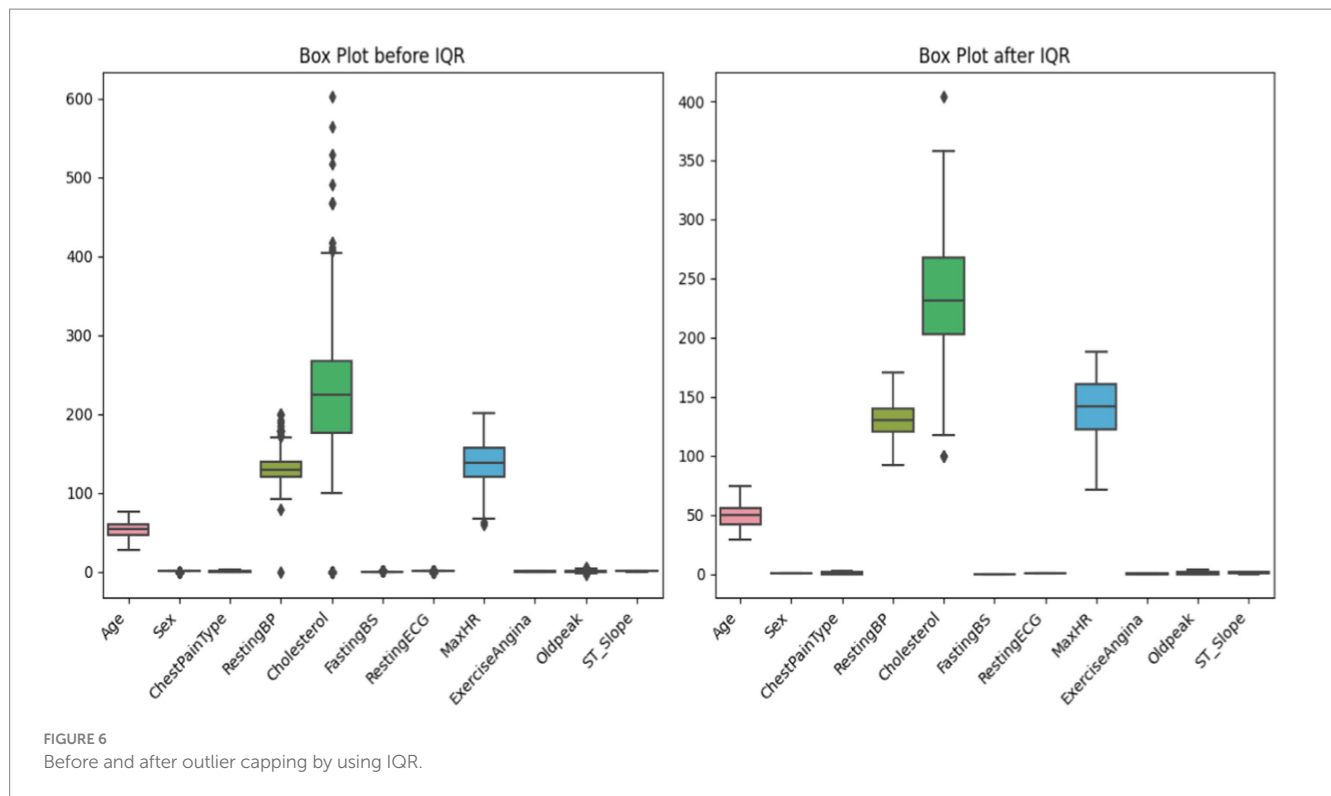
$$y_{ji}^{\wedge imputed} = f(x_{i,-j}) + \epsilon_i \tag{1}$$

- $y_{ji}^{\wedge imputed}$ shows the value that has been ascribed to the absent item.
- f : The missing value is estimated by the function. The data type of variable j might affect this function.
- $x_{i,-j}$: With the exception of variable j , all observed values of the variables are represented by the vector in the i th observation.
- ϵ_i : Error term

The observed values of all the variables in this context, with the exception of variable j in observation i , are stored in the vector $x_{i,-j}$. By using these observed values, the function f is used to estimate the missing value. The assumed value's error word ϵ_i denotes any inexplicable volatility or unpredictability.

3.2.2 Scaling with label encoder

There are two essential methods for preparing machine learning data: label encoding and scaling. To transform categorical data into a numerical form, Label Encoding assigns unique integer labels to each category. One method for giving numerical values to categorical variables is Label Encoding (23). With Label Encoding, "Male" and "Female" would be represented as 0 and 1, respectively, in a "Gender" column. For algorithms that can only take numerical input, this simplifies the usage of categorical variables. On the flip side, numerical features can be scaled to be uniform in size so that no one characteristic can have an outsized impact due to size disparities. Model convergence and performance are both enhanced by methods Standard Scaling, [shown in equation (2)] which ensure that all features contribute equally. A typical preprocessing step involves converting categorical characteristics using Label Encoding and then scaling numerical features to make their magnitudes consistent. Label Encoding and Scaling, when used together; make it easy to get datasets ready to be used in machine learning algorithms.



$$X_{scaled} = \frac{X - \mu}{\sigma} \quad (2)$$

- The initial feature value was X .
- The feature values mean is represented by μ .
- The feature values' standard deviation is represented by σ .

3.2.3 Handling outliers with IQR

Careful data preparation, including outlier removal, improves machine learning model durability. Interquartile Range (IQR) is a prominent method for finding and treating dataset outliers. Interquartile range (IQR) is the difference between a distribution's third and first quartiles, or 75th and 25th percentiles [shown in equation (3)]. Abnormal data points fall below or above the lower and higher limits ($Q1 - 1.5 * IQR$ and $Q3 + 1.5 * IQR$, respectively) [shown in equations (4, 5)]. Outliers might hurt the model's performance, but the IQR-based technique would find and fix them. To minimize outliers' impact on learning, alter them. This reduces model sensitivity to unexpected data sets. This is crucial for algorithms that respond fast to data distribution changes (Table 1).

The initial stage in IQR-based outlier treatment is splitting the sample into quartiles and determining the IQR (24). Outliers can be deleted or altered by comparing them against boundaries. This technique emphasizes creating more extensive and reliable datasets to improve ML model generalizability and prediction accuracy. The IQR outlier control approach must be used to prepare data for future machine learning experiments to ensure reliability and efficiency (Figure 6).

$$IQR_{outlier} = Q_3 - Q_1 \quad (3)$$

TABLE 2 Before and after applying SMOTE.

Before applying SMOTE		After applying SMOTE	
Class	Count	Class	Count
0	410	0	508
1	508	1	508

$$\text{LowerBound} = Q_1 - 1.5 * IQR \quad (4)$$

$$\text{UpperBound} = Q_3 + 1.5 * IQR \quad (5)$$

3.2.4 Handling imbalanced dataset with SMOTE

To ensure that machine learning algorithms are not biased toward the dominant class and hence reduce prediction accuracy, imbalanced datasets must be handled. In order to rectify class imbalance, particularly in cases when minority occurrences are underrepresented, this system applies the Synthetic Minority Over-sampling Technique (SMOTE) (25) [shown in equation (6)]. Class distribution has an imbalance with 508 class 1 instances and 410 class 0 instances (shown in Table 2 and Figure 7). It would indicate that the 0.8071 imbalance ratio is less than the 1 - imbalance_threshold threshold. SMOTE manipulates the underrepresented class's dataset presence by creating false instances of it. This is accomplished by building artificial instances along line segments that connect instances of minority classes. With a more evenly distributed dataset, the model may learn from more examples and, perhaps, make better predictions with new data.

Model prediction is improved with SMOTE (26) because it decreases class imbalance. When data from minority groups is limited, this strategy really shines in terms of model performance. To aid in

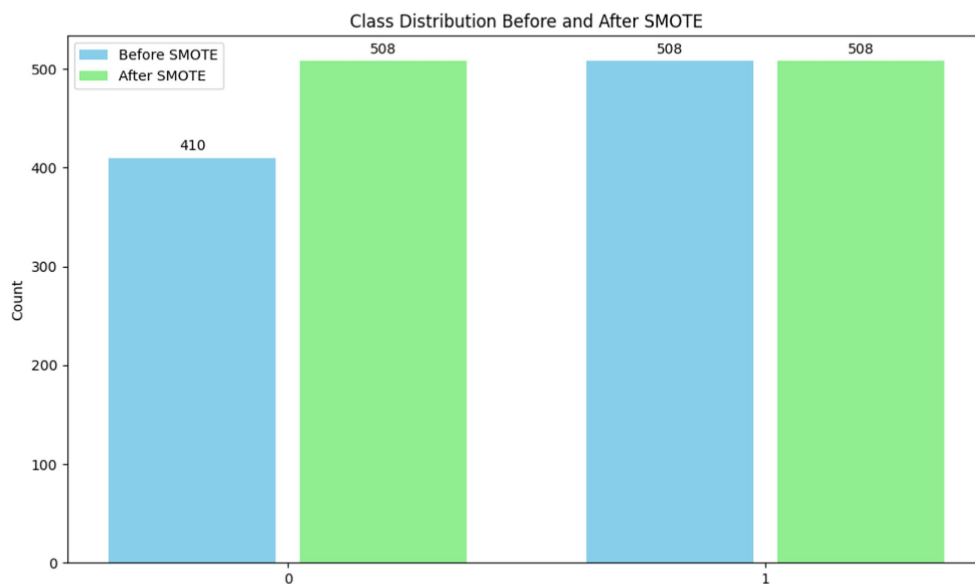


FIGURE 7
Before and after applying SMOTE.

the management of unbalanced datasets, SMOTE encourages correct and equitable predictions across all classes.

$$\text{Imbalanced Ratio (IR)} = \frac{\text{The count of occurrences in the majority class}}{\text{The count of occurrences in the minority class}} \quad (6)$$

3.2.5 Feature selection using hybrid GOL2-2 T

A new hybrid feature selection approach called the Hybrid GOL2-2 T, in which L2 regularization is fused with the Grasshopper Optimization Algorithm (GOA) (27), is discussed. This solution of the metaheuristic attracts a promising subset of the feature set through the application of an objective function and global search. We then applied L2 regularization to the selected feature set. Majorly, the objective of L2 regularization is to penalize too many coefficients, promote sparsity, and preserve only the most useful features. Hybrid GOL-2 T combining fine tuning powers from L2 regularizations with the muscular strength of GOA combined gives a dependable feature selection technique. In this respect, models that provide predictive classification via two-level approaches should have higher classification accuracy and dependability since they help in selecting the most relevant characteristics and reducing overfitting. As has been correctly pointed out, for these reasons, this approach has gained significant acceptance and has become an indispensable tool for many machine learning applications, like regression and classification tasks.

3.2.6 Grasshopper optimization algorithm

Developed in 2017 by Saremi et al. (32), the Grasshopper Optimization Technique (GOA) is a metaheuristic optimization technique inspired by nature. The idea originated from the way

grasshoppers behaved in unison. GOA has been used to solve a variety of optimization problems, including feature selection in the context of machine learning. Here is a brief description of how GOA works shown in Algorithm 1, complete with formulas and the algorithm itself:

Algorithm 1: Grasshopper Optimization Algorithm (GOA)

Initialize population of grasshoppers (solutions)

Initialize best solution (best_solution)

Initialize number of iterations (iterations)

While (termination criterion is not met)

For each grasshopper ii

 Calculate social interaction component S_{ii} shown in equation (7)

 Calculate gravity component G_{ii} shown in equation (8)

 Calculate wind component A_{ii} shown in equation (9)

 Calculate movement of grasshopper ii (x_{ii})

 Update position of grasshopper ii (x_{ii})

 Evaluate objective function for new position ($fitness_{ii}$)

If ($(fitness_{ii}) > fitness\ of\ best_solution$)

 Update best solution (best_solution)

End If

End For

 Update number of iterations (iterations)

End While

Return best solution

$$S_{ii} = C * \left(\frac{\sum (x_{jj} - x_{ii})}{N} \right) \quad (7)$$

$$G_{ii} = -g * (x_{ii} - o) \quad (8)$$

$$A_{ij} = U * (x_e - x_{ij}) \quad (9)$$

Where,

- c is a decreasing coefficient that balances the processes of exploration and exploitation.
- g is a constant that determines the strength of the gravity component is the center of the search space.
- U is a constant that determines the strength of the wind component.
- x_e is the position of the best solution found so far.
- N is the number of grasshoppers.
- x_{ij} and x_{jj} are the positions of the grasshoppers.

The algorithm generates grasshoppers, each representing a possible solution. The first grasshopper in the population gets the best answer. The algorithm then loops through each grasshopper in the population. The application calculates grasshopper social interaction, gravity, and wind components. These components steer the grasshopper toward the best alternative.

The components calculated in the previous stage are used to modify the grasshopper's movement. The objective function measures grasshopper positioning and solution efficacy. A new site becomes the ideal option if it outperforms the old one. After reaching grasshopper population termination criteria, the technique continues iteratively. A maximum number of iterations, a minimum

fitness value, or any other suitable stopping condition may be used for the job. After optimization, the technique returns the ideal answer (Figure 8).

3.2.7 L2 regularization

L2, sometimes called ridge regression (28), is a machine learning technique used to reduce a model's complexity by adding a penalty term to the loss function. The penalty term is directly correlated with the square of the magnitudes of the coefficients, encouraging the model to have smaller coefficients and reducing the likelihood of overfitting shown in Algorithm 2.

The L2 regularization term is added to the loss function as shown in equation (10).

$$\text{Loss} = \text{MSE} + \left(\alpha * \sum (\text{coefficient}^2) \right) \quad (10)$$

Where:

MSE is the mean squared error between the predicted and actual values shown in equation (11).

α is the regularization parameter (a hyperparameter).

Coefficient is the coefficient of the feature in the model.

The algorithm for L2 regularization can be described as follows:

Algorithm 2: L2 regularization

Initialize coefficients to small random values

While (termination criterion is not met)

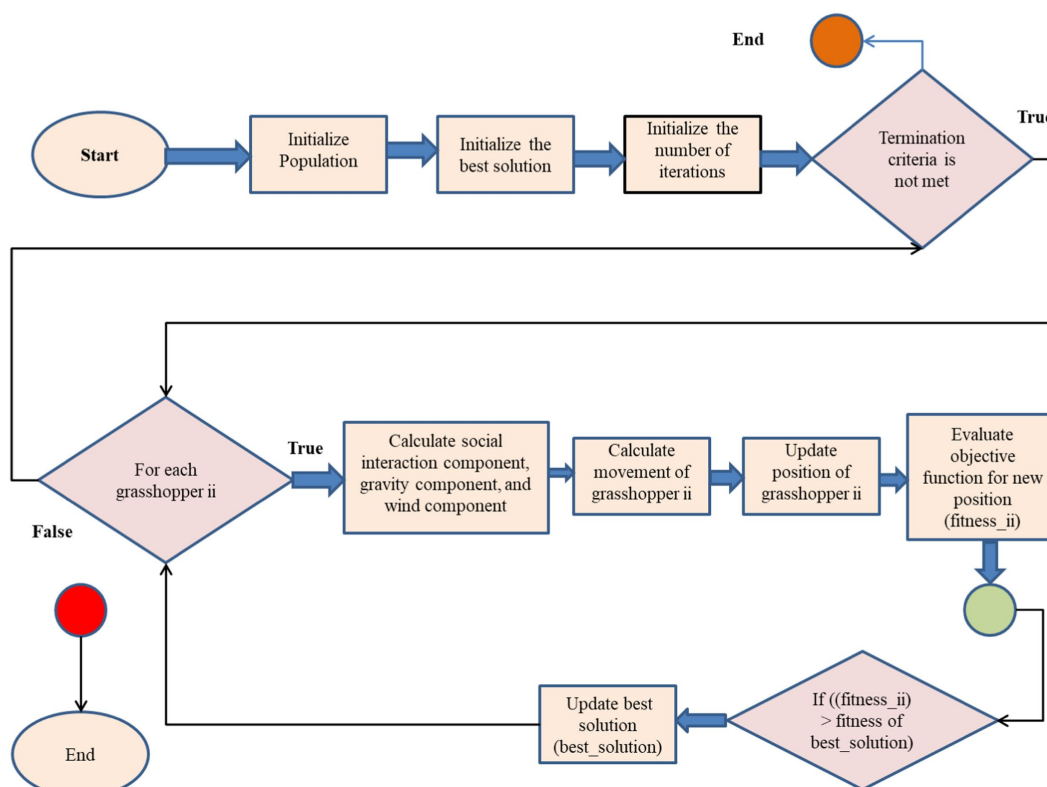


FIGURE 8
Feature selection flow chart for Grasshopper optimization algorithm.

Calculate MSE using the current values of the coefficients by using [equation \(11\)](#)

Calculate sum of squared coefficients by using [equation \(12\)](#)

Calculate regularized loss function as the sum of the MSE and the regularization term (alpha * sum of squared coefficients)

Update coefficients to minimize the regularized loss function

End While

Return the optimized coefficients

$$\text{Mean Squared Error (MSE)} = \frac{1}{n} \sum_{ii=1}^n (y_{ii} - y_{ii}^t)^2 \quad (11)$$

$$\text{Sum of squared coefficients (SSC)} = \sum_{jj=1}^p \theta_{jj}^2 \quad (12)$$

Where,

- p is the number of coefficients.
- y_{ii} as the data point's observed value ii
- y_{ii}^t as the anticipated value for data point ii .

The L2 regularization approach may be used to a wide range of models due to its computational efficiency. To achieve the optimal balance between bias and variance, the regularization hyperparameter alpha has to be changed. Features that are more effective at lowering the Mean Squared Error (MSE) are chosen when L2 regularization reduces the size of the model's coefficients. L2 regularization may be used as a feature selection method by selecting only those features in the model that have coefficients greater than zero ([Figure 9](#)).

3.3 Hyperparameter tuning using babysitting algorithm

The babysitting Algorithm (BA) ([29](#)) in AdaBoost ([30](#)) decision fusion manually evaluates the model's performance after iteratively modifying the hyperparameters. Setting hyperparameters,

constructing a table, separating the dataset into training, validation, and testing sets, and progressively experimenting with different combinations are the steps. For each combination, an AdaBoost classifier is trained on the training set and assessed on the validation set using a performance metric. The hyperparameter table is updated when the trial number, hyperparameters, and performance measure change. Select the hyperparameters with the best validation set outcomes after all trials. The training and validation sets are utilized to train a new AdaBoost classifier using the optimum hyperparameters. For an impartial evaluation, the finished model is tested on the testing set shown in [Algorithm 3](#).

Algorithm 3: Hyperparameter Tuning Babysitting on AdaBoost Decision Fusion

```
// Initialize hyperparameters and performance metric
InitializeHyperparameters()
// Initialize the hyperparameter table
InitializeHyperparameterTable()
// Main loop for hyperparameter tuning
while (stopping criterion not met) do
    // Iterate through hyperparameter combinations
    for each hyperparameter combination do
        // Train AdaBoost classifier with current hyperparameters
        model = TrainAdaBoostClassifier(current_hyperparameters)
        // Evaluate the model's performance on the validation set
        performance_metric = EvaluateModelPerformance(model, validation_set)
        // Update hyperparameter table with current hyperparameters and performance metric
        UpdateHyperparameterTable(current_hyperparameters, performance_metric)
    end for
    // Select best hyperparameters based on the highest performance metric
    best_hyperparameters = SelectBestHyperparameters()
    // Train AdaBoost classifier with the best hyperparameters on the combined training and validation sets
    best_model = TrainAdaBoostClassifier(best_hyperparameters, combined_training_validation_set)
```

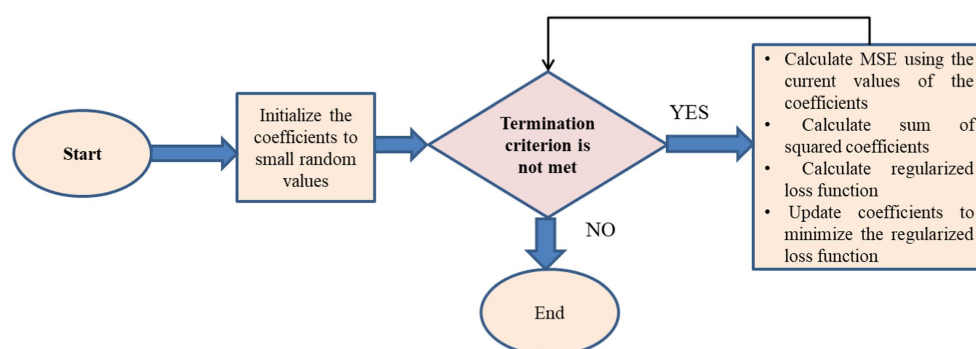


FIGURE 9
Feature selection flow chart for L2 regularization.

```

// Evaluate the final model on the testing set
final_performance_metric = EvaluateModelPerformance(best_
model, testing_set)
// Update stopping criterion based on convergence or
maximum iterations
UpdateStoppingCriterion()
end while

```

3.4 Model building for heart failure prediction

3.4.1 Ensemble technique with adaptive boosted decision fusion

“Adaptive Boosted Decision Fusion (31) is an advanced ensemble learning algorithm that effectively combines the principles of Adaptive Boosting (AdaBoost) and Decision Fusion.” To prioritize instances that are harder to classify, this innovative approach has the algorithm adaptively changing the weights [shown in equation (13)] given to less effective learners. When combined with decision fusion, ABDF sequential training method for weak models allows for the efficient integration of results from many decision-makers [shown in equations (14–18)]. The ultimate result is a very accurate and reliable prediction model that is both adaptable and resilient. One way to make the ensemble better is via adaptive boosted decision fusion, which uses iterative refinement and smartly gives different learners different weights depending on how well they do. When it's critical to combine multiple decision-making viewpoints to get superior predicted outcomes, this method shines.

Input:

Training dataset: $D = \{(ux_1, uy_1), (ux_2, uy_2), \dots, (ux_n, uy_n)\}$.

Where ux_i the feature is vector and uy_i is the corresponding label.

Number of weak learners: UT

Initialization:

$$1. \text{ Initialize instance weights : } uw_i = \frac{1}{n} \text{ for } i = 1, 2, 3, \dots, n \quad (13)$$

2. Initialize an empty ensemble of weak learners.

$$\text{For each iteration : } ut = 1, 2, 3, \dots, UT : \quad (14)$$

3. Train a weak learner uh_t using the current instance weights.

i. Compute the error of the weak learner :

$$\epsilon_t = \sum_{i=1}^n uw_i \cdot | (uh_t(ux_i) \neq uy_i) | \quad (15)$$

where $|(\cdot)|$ is the indicator function.

$$\text{ii. Compute the learner weight : } \alpha_t = \frac{1}{2 \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right)} \quad (16)$$

iii. Update instance weights :

$$\text{for } i = 1, 2, 3, \dots, n, uw_i \leftarrow uw_i \cdot \exp(-\alpha_t \cdot uy_i \cdot uh_t(ux_i)) \quad (17)$$

TABLE 3 Selected features with scores using GOA.

Features	Score
cp	0.047363
trestbps	0.002171
chol	0.000873
thalach	0.007371
oldpeak	0.047905
ca	0.03526

TABLE 4 Selected Features with Scores using L2 regularization.

Feature	Score
cp	0.051832
oldpeak	0.047056
ca	0.037252
thalach	0.007355
trestbps	0.002556

$$\text{Normalize weights : } uw_i \leftarrow \frac{uw_i}{\sum_{i=1}^n uw_i} \quad (18)$$

iv Add the weak learner uh_t to the ensemble with weight α_t .

Output:

$$\text{Ensemble of weak learners : } \{(\alpha_1, uh_1), (\alpha_2, uh_2), \dots, (\alpha_T, uh_{UT})\} \quad (19)$$

Predictions:

For a new instance ux , the final prediction is given by :

$$H(ux) = \text{sign} \left(\sum_{ut=1}^{UT} \alpha_t (ux) \right) \quad (20)$$

This method combines the best features of AdaBoost and Decision Fusion in a way that strengthens the ensemble (26), making it better at handling misclassifications and making accurate predictions. A long-lasting ensemble model that frequently outperforms individual models is produced by ABDF iterative method of correcting errors of weak models [shown in equation (19)]. Classification problems, such as the prediction [shown in equation (20)] of cardiac illness, frequently use ABDF. It finds usage in a variety of domains due to its flexibility in accommodating varied poor learners (Tables 3, 4).

4 Result and discussion

4.1 Performance assessments

4.1.1 Feature selection outcome using GOL2-2T

The Grasshopper Optimization Algorithm (GOA) (32) identified heart disease predictors. This method found critical characteristics like chest pain type (cp), resting blood pressure (trestbps), serum cholesterol (chol), maximum heart rate (thalach),

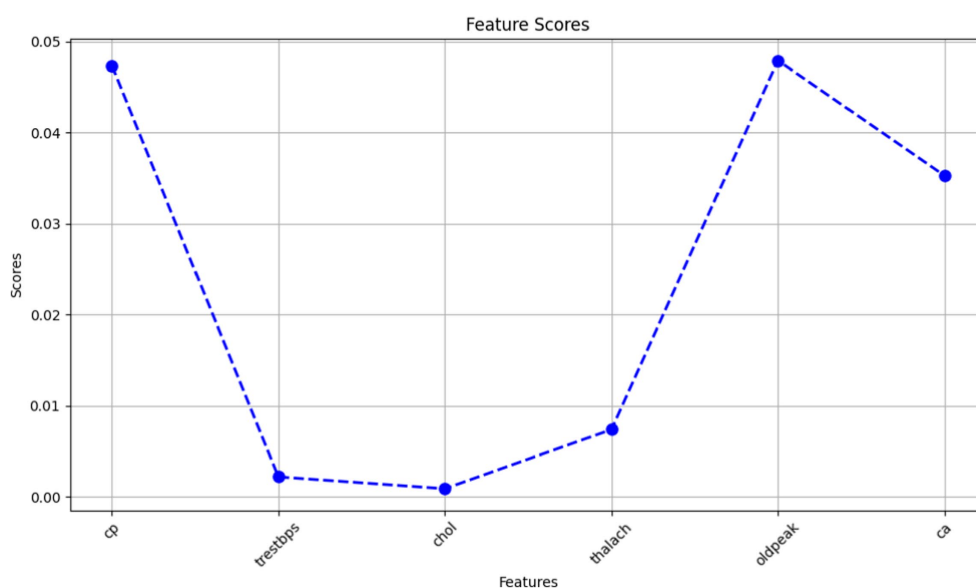


FIGURE 10
A line graph denoting selected features with scores using GOA.

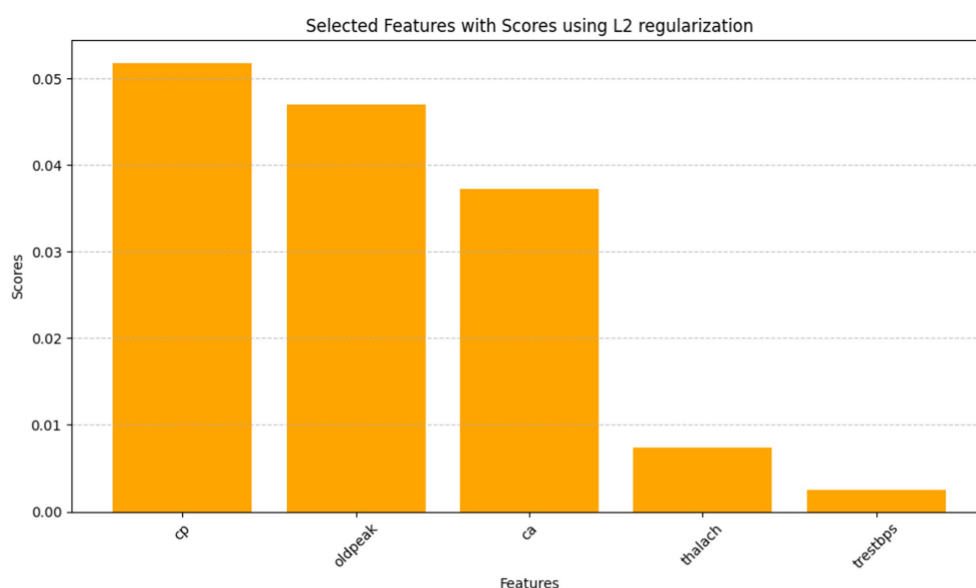


FIGURE 11
A bar graph denoting selected features with scores using L2 regularization.

ST depression caused by exercise compared to rest (oldpeak), and the number of main vessels colored by fluoroscopy (ca). High scores showed relevancy. The prediction model ranked attributes by score. Next, we used ridge regression, also known as L2 regularization, to enhance feature selection. Revised features included oldpeak, thalach, ca, trestbps, and cp. Reevaluating characteristics using L2 regularization yielded scores that accurately represent their value in heart disease prediction. Comparing the two feature selection approaches shows convergence in the selected qualities, suggesting they may

be essential for heart disease identification. However, slight discrepancies in feature significance showed that GOA and L2 regularization use different techniques and criteria. We need more study to evaluate the predictive modeling of the upgraded features and the implications for heart disease diagnostics (Figures 10, 11).

4.1.2 Hyperparameter tuning outcome using babysitting algorithm on ABDF

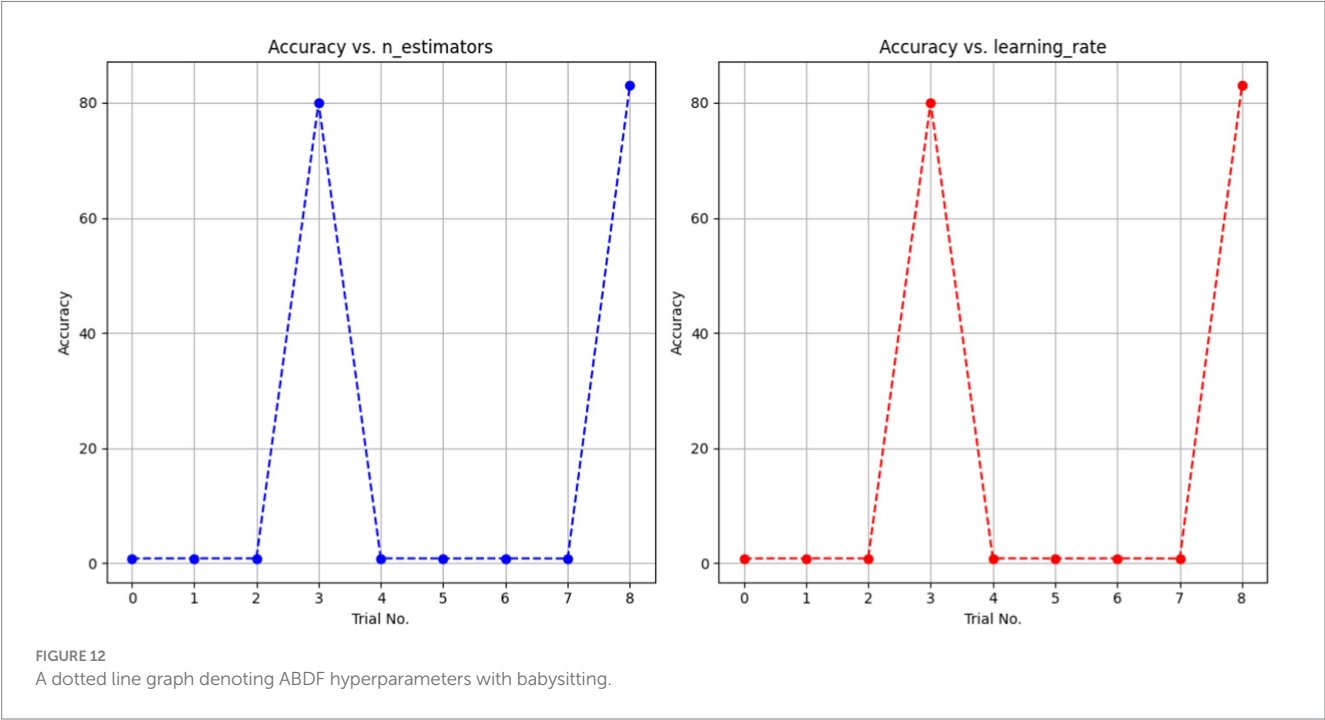
The AdaBoost Decision Fusion model's hyperparameters were optimized by a two-pronged approach involving tuning the

TABLE 5 AdaBoost decision fusion model hyperparameters tuning summary.

Models used	Hyperparameters tuning algorithm	Hyperparameters	Search Space
AdaBoost decision Fusion	Babysitting	n_estimators	50–200
		learning_rate	0.5–1

TABLE 6 AdaBoost decision fusion hyperparameters with babysitting.

Trial no.	Accuracy	n_estimators	learning_rate
0	0.802	50	0.1
1	0.82	50	0.5
2	0.812	50	1
3	80.00	100	0.1
4	0.822	100	0.5
5	0.804	100	1
6	0.819	200	0.1
7	0.79	200	0.5
8	83.00	200	1



n_estimators and learning rate with the help of the Babysitting Algorithm (see in Table 5). A narrow range of the search space for n_estimators, which was from 50 to 200, and a more broad range of the learning rate, which was from 0.5 to 1, was seen. The hyperparameter optimization was made through a number of runs by substituting various combinations of parameters for n_estimators and learning_rate (see in Table 6 and Figure 12). The data obtained from the ABDF model showed deviation across the many attempts conducted in the experiment; Trial No. 8 gave 8 as the most accurate results, their accuracy being 83.00%. The

TABLE 7 IQR outlier detection ABDF performance metrics.

IQR outlier detection with ABDF results	
Metrics	Values
Accuracy	0.83
Precision	0.84
Recall	0.85
f1_score	0.84
AUC Score	0.89

crucial aspiration of this process was the attainment of an optimal accuracy and robustness model for the ABDF model, specifically as it concerned the given task.

4.2 IQR outlier detection with ABDF

Heart disease may be reliably predicted using the ABDF method and the IQR outlier preprocessing strategy. The model

achieves an 83% accuracy rate in instance categorization and an 84% success rate in accurately anticipating predicted positives (see in Table 7 and Figure 13). The model correctly identifies a large number of positive examples, as evidenced by its impressive recall score of 85%. An F1 Score of 84% (a measure of both recall and accuracy) indicates that the model is performing well. With an Area Under the Curve (AUC) score of 89% (see in Figure 14), the model clearly can differentiate between positive and negative occurrences. Based on these metrics, it appears that preprocessing

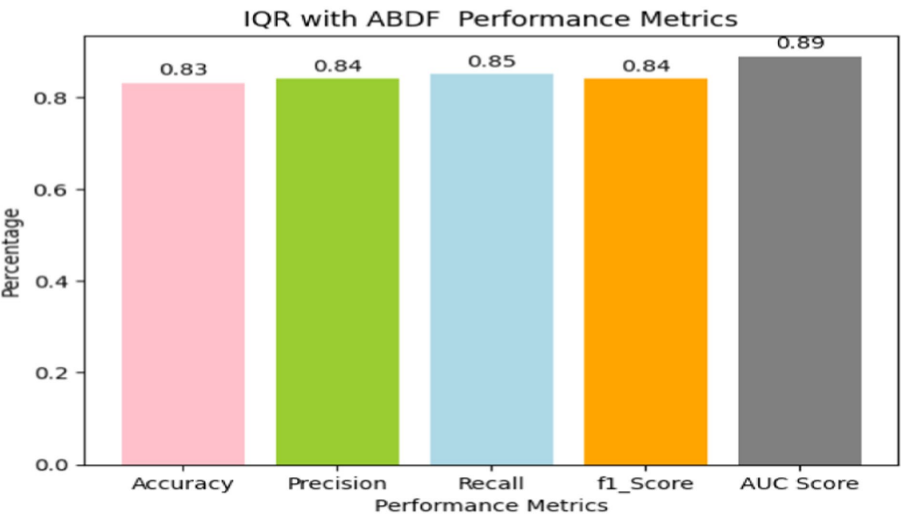


FIGURE 13
Bar graph shows IQR outlier detection with ABDF performance metrics.

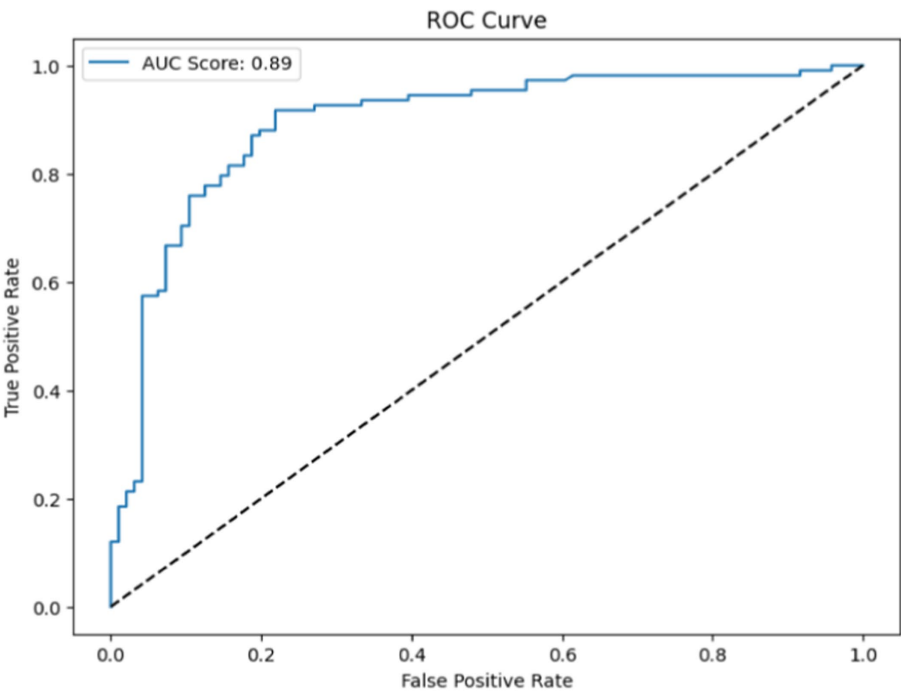
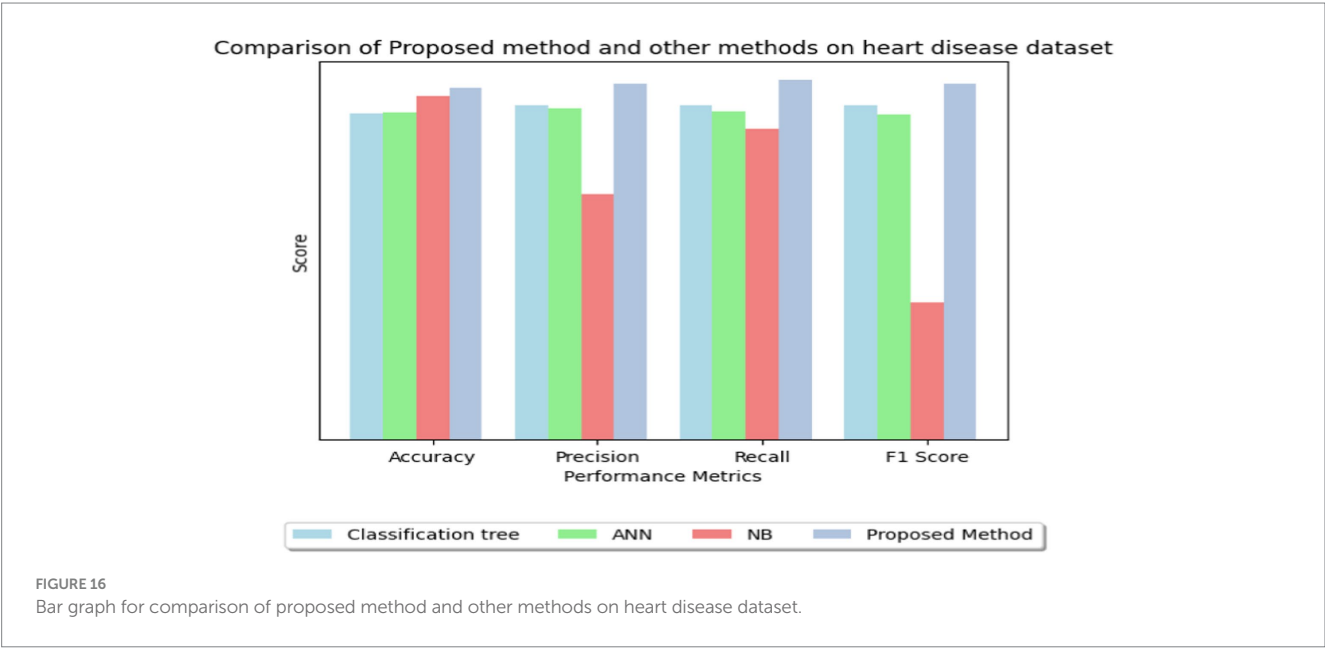
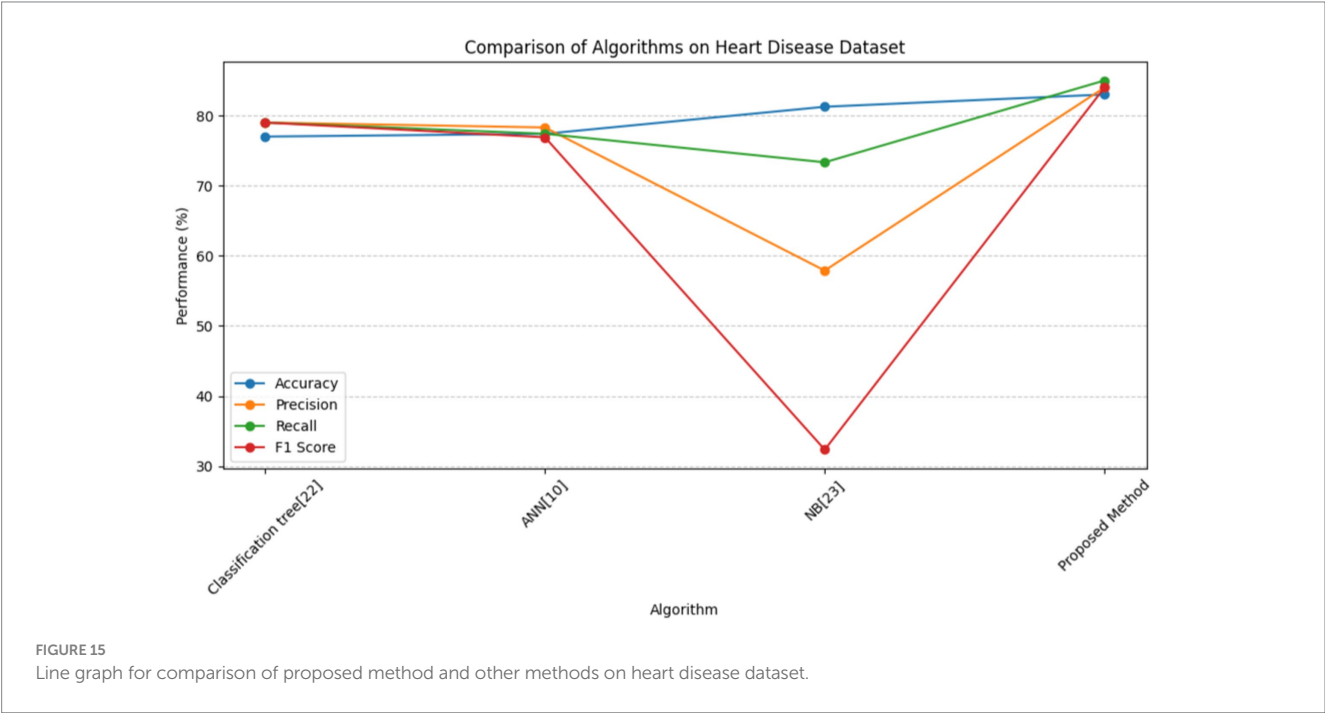


FIGURE 14
ROC for IQR outlier detection with ABDF.

TABLE 8 Comparison of proposed method and other methods on heart disease dataset.

Algorithm	Accuracy	Precision	Recall	f1_score
Classification tree (33)	77.0	79.0	79.0	79.0
ANN (17)	77.39	78.30	77.40	76.90
NB (34)	81.25	57.89	73.33	32.35
Proposed method	83.0	84.0	85.0	84.0



using ABDF and IQR improves the accuracy, precision, recall, and overall predictive performance of models used to forecast cardiac diseases. According to its reliable performance, the model may be relied on by healthcare providers to aid in the rapid identification and treatment for people at risk of heart disease.

4.2.1 Comparison of proposed method and other methods on heart disease dataset

In Table 8, multiple approaches are used to a heart disease dataset to assess accuracy, precision, recall, and F1-score. The suggested technique outperforms the others with 83.0% accuracy. This shows that it locates dataset instances properly. This method outperforms the Classification Tree and Artificial Neural Network (ANN) methods in classification testing. The new approach outperforms previous methods in accuracy, recall, and F1-score. Its great overall performance is due to its balanced trade-off between precisely recognizing positive examples (precision) and capturing all positive occurrences (recall).

The Naive Bayes (NB) technique exceeds the suggested method in accuracy (81.25%) but much worse in precision, recall, and F1-score. More particular, the NB technique has poorer precision and F1-score than the suggested strategy, suggesting more false positives and a worse accuracy-recall trade-off. The findings suggest that the proposed technique balances accuracy and precision-recall, making it suitable for heart illness classification (see in Figures 15, 16). The comparison research also emphasizes the need of choosing the right technique for favorable performance indicators. This scenario shows that the recommended strategy is better than the present options.

5 Discussions

Our work presents an 83% reliable machine learning heart disease prediction approach. We used cutting-edge methods like SMOTE, IQR outlier detection, MICE, and GOL2-2T, a hybrid feature selection technique, to improve predictive accuracy and robustness. Combining these techniques improved feature selection and model performance, according to our findings. Our heart disease patient identification approach is very accurate. These results demonstrate the need of using cutting-edge machine learning algorithms in medicine to identify and cure diseases early.

Our findings may help doctors predict cardiac disease, improving patient care and intervention. Our accurate diagnostic equipment may enhance patient outcomes and minimize cardiovascular disease mortality. However, our research has some drawbacks. Our hopeful results are limited to a dataset and may not apply to other patient populations or healthcare situations. Data quality and feature selection criteria may also affect our model's performance.

We urge additional research to corroborate our results across a variety of datasets and populations. Using additional machine learning methods (35–40) and domain-specific information may improve the model's interpretability and prediction accuracy. To evaluate the long-term effects of early cardiac disease identification on patient outcomes, longitudinal studies are needed. In conclusion, our results emphasize the necessity for ongoing study to develop cardiovascular prediction analytics.

6 Conclusion and future scope

In conclusion, our study met the urgent demand for precise and effective cardiovascular disease prognostic diagnostic tools. MICE, IQR outlier detection, SMOTE, and Adaptive Boosted Decision Fusion (ABDF) were used to improve heart disease prediction models' precision and reliability. The Hybrid GOL2-2T feature selection technique has enhanced our process by discovering important features and decreasing overfitting.

We solved class imbalance, missing data, and outlier identification to create a model that outperforms previous methods. The accuracy rate of 83.0% and balanced F1 score of 84.0% of our heart disease prediction method were impressive. The accuracy, recall, and AUC score demonstrate the validity and applicability of our methods. Our findings show that powerful machine learning techniques must be used in healthcare to produce reliable cardiovascular disease diagnosis tools. The study gives doctors tools for early diagnosis and effective treatment of cardiovascular disease risk.

Future study may improve prediction models and examine additional factors to improve diagnostic precision.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

SP: Visualization, Data curation, Investigation, Software, Validation, Writing – original draft. MH: Data curation, Investigation, Funding acquisition, Project administration, Supervision, Writing – review & editing. SA: Writing – review & editing. US: Conceptualization, Software, Validation, Writing – original draft. NT: Conceptualization, Formal analysis, Investigation, Writing – review & editing. SI: Investigation, Methodology, Resources, Visualization, Writing – review & editing. FA: Resources, Software, Validation, Writing – review & editing. TA: Conceptualization, Data curation, Formal analysis, Writing – review & editing. AN: Data curation, Investigation, Writing – review & editing. GS: Writing – review & editing. CY: Funding acquisition, Writing – review & editing, Conceptualization, Formal analysis, Resources, Visualization. TG: Formal analysis, Methodology, Validation, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Barik S, Mohanty S, Rout D, Mohanty S, Patra AK, Mishra AK. Heart disease prediction using machine learning techniques. *Adv Electr Control Signal Syst.* (2020) 665:879–88. doi: 10.1007/978-981-15-5262-5_67
- Riyaz L, Butt MA, Zaman M, Ayob O. Heart disease prediction using machine learning techniques: a quantitative review. *Adv Intell Syst Comput.* (2021) 1394:81–94. doi: 10.1007/978-981-16-3071-2_8
- Fu Q, Chen R, Ding Y, Xu S, Huang C, He B, et al. Sodium intake and the risk of various types of cardiovascular diseases: a Mendelian randomization study. *Front Nutr.* (2023) 10:509. doi: 10.3389/fnut.2023.1250509
- Huang L, Wu J, Lian B, Zhang D, Zhai Y, Cao L. Successful robot-assisted laparoscopic resection of pheochromocytoma in a patient with dilated cardiomyopathy: a case report on extremely high-risk anesthesia management. *Medicine.* (2023) 102:e35467. doi: 10.1097/md.00000000000035467
- Wang Y, Zhai W, Zhang H, Cheng S, Li J. Injectable Polyzwitterionic lubricant for complete prevention of cardiac adhesion. *Macromol Biosci.* (2023) 23:e2200554. doi: 10.1002/mabi.202200554
- Zhou Y, Sun X, Yang G, Ding N, Pan X, Zhong A, et al. Sex-specific differences in the association between steps per day and all-cause mortality among a cohort of adult patients from the United States with congestive heart failure. *Heart Lung.* (2023) 62:175–9. doi: 10.1016/j.hrtlng.2023.07.009
- Liu Z, Fan Y, Zhang Z, Fang Y, Cheng X, Yang Q, et al. mTOR in the mechanisms of atherosclerosis and cardiovascular disease. *Discov Med.* (2021) 31:129–40.
- Statista. Share of People with Heart Problems India 2020, by Age Group. (2023). Available at: <https://www.statista.com/statistics/1123509/india-share-of-respondents-with-heart-issues-by-age-group/>.
- Manikandan S. "Heart attack prediction system." (2017). International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS).
- Sirisha U, Praveen SP, Srinivasu PN, Barsocchi P, Bhoi AK. Statistical analysis of design aspects of various YOLO-based deep learning models for object detection. *Int J Comput Intell Syst.* (2023) 16:126. doi: 10.1007/s44196-023-00302-w
- Srinivasu PN, JayaLakshmi G, Jhaveri RH, Praveen SP. Ambient assistive living for monitoring the physical activity of diabetic adults through body area networks. *Mob Inf Syst.* (2022) 2022:1–18. doi: 10.1155/2022/3169927
- Krishna T, Praveen SP, Ahmed S, Srinivasu PN. Software-driven secure framework for mobile healthcare applications in IoMT. *Intell Decis Technol.* (2023) 17:377–93. doi: 10.3233/IDT-220132
- Shah D, Patel S, Bharti SK. Heart disease prediction using machine learning techniques. *SN Comp Sci.* (2020) 1:1–6. doi: 10.1007/s42979-020-00365-y
- Katarya R, Srinivas P. "Predicting heart disease at early stages using machine learning: a survey." (2020). International Conference on Electronics and Sustainable Communication Systems (ICESC).
- Jindal H, Agrawal S, Khera R, Jain R, Nagrath P. Heart disease prediction using machine learning algorithms. *IOP Conf Ser Mater Sci Eng.* (2021) 1022:012072. doi: 10.1088/1757-899x/1022/1/012072
- Gonsalves AH, Thabtah F, Mohammad RMA, Singh G. "Prediction of coronary heart disease using machine learning." Proceedings of the 2019 3rd International Conference on Deep Learning Technologies (2019).
- Nashif S, Rakib Raihan M, Rasedul Islam M, Imam MH. Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system. *World J Eng Technol.* (2018) 6:854–73. doi: 10.4236/wjet.2018.64057
- Bhatt CM, Patel P, Ghetia T, Mazzeo PL. Effective heart disease prediction using machine learning techniques. *Algorithms.* (2023) 16:88. doi: 10.3390/a16020088
- Abood Kadhim M, Radhi AM. Heart disease classification using optimized machine learning algorithms. *Iraqi J. Comp. Sci. Math.* (2023) 4:31–42. doi: 10.52866/ijcsm.2023.02.02.004
- Huang H, Wu N, Liang Y, Peng X, Shu J. SLNL: a novel method for gene selection and phenotype classification. *Int J Intell Syst.* (2022) 37:6283–04. doi: 10.1002/int.22844
- Heart Disease Dataset. Kaggle. (2019). Available at: <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset/data>.
- Samad MD, Abrar S, Diawara N. Missing value estimation using clustering and deep learning within multiple imputation framework. *Knowl-Based Syst.* (2022) 249:108968. doi: 10.1016/j.knosys.2022.108968
- Gandla VR, Mallela DV, Chaurasiya R. "Heart failure prediction using machine learning." International Conference on Applied Computational Intelligence and Analytics (Acia-2022) (2023).
- Mamun A, Paul BK, Ahmed K, Bui FM, Quinn JM, Moni MA. Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison. *Comput Biol Med.* (2021) 136:104672. doi: 10.1016/j.combiomed.2021.104672
- Ishaq A, Sadiq S, Umer M, Ullah S, Mirjalili S, Rupapara V, et al. Improving the prediction of heart failure patients' survival using SMOTE and effective.
- Data Mining Techniques. IEEE Access 9 (2021): 39707–716. doi: 10.1109/access.2021.3064084
- Rani P, Kumar R, Ahmed NS, Jain A. A decision support system for heart disease prediction based upon machine learning. *J Rel Intell Environ.* (2021) 7:263–75. doi: 10.1007/s40860-021-00133-6
- Hasan MK, Habib AA, Islam S, Safie N, Ghazal TM, Khan MA, et al. Federated learning enables 6 G communication technology: requirements, applications, and integrated with intelligence framework. *Alex Eng J.* (2024) 91:658–68. doi: 10.1016/j.aej.2024.02.044
- Hasan MK, Hosain M, Ahmed MK, Islam S, Aledaily AN, Yasmeen S, et al. Encrypted images in a V-BLAST assisted SC-FDMA wireless communication system. *Trans Emerg Telecommun Technol.* (2024) 35:e4882. doi: 10.1002/ett.4882
- Ahmed AA, Hasan MK, Nafi NS, Aman AH, Islam S, Nahi MS Optimization technique for deep learning methodology on power Side Channel attacks. 2023 33rd International Telecommunication Networks and Applications Conference IEEE (2023).
- Tirumanadham K, Kumar M, Thailaynayaki S, Sriram M. "Evaluating boosting algorithms for academic performance prediction in E-learning environments." (2024) International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE).
- Dey C, Bose R, Ghosh KK, Malakar S, Sarkar R. LAGOA: learning automata based grasshopper optimization algorithm for feature selection in disease datasets. *J Ambient Intell Human Comp.* (2021) 13:3175–94. doi: 10.1007/s12652-021-03155-3
- Ponti J, Moacir P. Combining classifiers: from the creation of ensembles to the decision fusion. (2011). 24th SIBGRAPI Conference on graphics, patterns, and images tutorials.
- Dwivedi AK. Performance evaluation of different machine learning techniques for prediction of heart disease. *Neural Comp Appl.* (2016) 29:685–93. doi: 10.1007/s00521-016-2604-1
- Liao H, Murah MZ, Hasan MK, Aman AHM, Fang J, Hu X, et al. A survey of deep learning Technologies for Intrusion Detection in internet of things. *IEEE Access.* (2024) 12:4745–61. doi: 10.1109/ACCESS.2023.3349287
- Islam MM, Hasan MK, Islam S, Balfaqih M, Alzahrani AI, Alalwan N, et al. Enabling pandemic-resilient healthcare: narrowband internet of things and edge intelligence for real-time monitoring. *CAAI Trans Intell Technol.* (2024). doi: 10.1049/cit.12314
- Lu S, Yang J, Yang B, Li X, Yin Z, Yin L, et al. Surgical instrument posture estimation and tracking based on LSTM. *ICT Express.* (2024). doi: 10.1016/j.icte.2024.01.002
- Kim S, Kim M, Kim J, Jeon JS, Park J, Yi HG. Bioprinting methods for fabricating in vitro tubular blood vessel models. *Cyborg Bionic Syst.* (2023) 4:43. doi: 10.34133/cbsystems.0043
- Bing P, Liu Y, Liu W, Zhou J, Zhu L. Electrocardiogram classification using TSST-based spectrogram and ConViT. *Front Cardiovasc Med.* (2022) 9:543. doi: 10.3389/fcvm.2022.983543
- Gao X, Huang D, Hu Y, Chen Y, Zhang H, Liu F, et al. Direct Oral anticoagulants vs. vitamin K antagonists in atrial fibrillation patients at risk of falling: a Meta-analysis. *Front Cardiovasc Med.* 9:329. doi: 10.3389/fcvm.2022.833329
- Srinivasu PN, Sirisha U, Sandeep K, Praveen SP, Maguluri LP, Bikku T. An interpretable approach with explainable AI for heart stroke prediction. *Diagnostics.* (2024) 14:128. doi: 10.3390/diagnostics14020128



OPEN ACCESS

EDITED BY

Sultan Ahmad,
Prince Sattam Bin Abdulaziz University,
Saudi Arabia

REVIEWED BY

Mohammed Almaiah,
The University of Jordan, Jordan
Ghaida Muttashar Abdulsahib,
University of Technology, Iraq
Inayat Khan,
University of Engineering and Technology,
Mardan, Pakistan

*CORRESPONDENCE

Mohammed Al-Yaari
✉ malyaari@kfu.edu.sa

RECEIVED 22 May 2024

ACCEPTED 05 July 2024

PUBLISHED 19 July 2024

CITATION

Alsharif N, Al-Adhaileh MH, Al-Yaari M,
Farhah N and Khan ZI (2024) Utilizing deep
learning models in an intelligent eye-tracking
system for autism spectrum disorder
diagnosis.

Front. Med. 11:1436646.

doi: 10.3389/fmed.2024.1436646

COPYRIGHT

© 2024 Alsharif, Al-Adhaileh, Al-Yaari, Farhah
and Khan. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Utilizing deep learning models in an intelligent eye-tracking system for autism spectrum disorder diagnosis

Nizar Alsharif^{1,2}, Mosleh Hmoud Al-Adhaileh^{1,3},
Mohammed Al-Yaari^{1,4*}, Nesren Farhah⁵ and Zafar Iqbal Khan⁶

¹King Salman Center for Disability Research, Riyadh, Saudi Arabia, ²Department of Computer Engineering and Science, Albaha University, Al Bahah, Saudi Arabia, ³Deanship of E-learning and Information Technology, King Faisal University, Al-Ahsa, Saudi Arabia, ⁴Department of Chemical Engineering, College of Engineering, King Faisal University, Al-Ahsa, Saudi Arabia, ⁵Department of Health Informatics, College of Health Sciences, Saudi Electronic University, Riyadh, Saudi Arabia, ⁶Department of Computer Science, College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia

Timely and unbiased evaluation of Autism Spectrum Disorder (ASD) is essential for providing lasting benefits to affected individuals. However, conventional ASD assessment heavily relies on subjective criteria, lacking objectivity. Recent advancements propose the integration of modern processes, including artificial intelligence-based eye-tracking technology, for early ASD assessment. Nonetheless, the current diagnostic procedures for ASD often involve specialized investigations that are both time-consuming and costly, heavily reliant on the proficiency of specialists and employed techniques. To address the pressing need for prompt, efficient, and precise ASD diagnosis, an exploration of sophisticated intelligent techniques capable of automating disease categorization was presented. This study has utilized a freely accessible dataset comprising 547 eye-tracking systems that can be used to scan pathways obtained from 328 characteristically emerging children and 219 children with autism. To counter overfitting, state-of-the-art image resampling approaches to expand the training dataset were employed. Leveraging deep learning algorithms, specifically MobileNet, VGG19, DenseNet169, and a hybrid of MobileNet-VGG19, automated classifiers, that hold promise for enhancing diagnostic precision and effectiveness, was developed. The MobileNet model demonstrated superior performance compared to existing systems, achieving an impressive accuracy of 100%, while the VGG19 model achieved 92% accuracy. These findings demonstrate the potential of eye-tracking data to aid physicians in efficiently and accurately screening for autism. Moreover, the reported results suggest that deep learning approaches outperform existing event detection algorithms, achieving a similar level of accuracy as manual coding. Users and healthcare professionals can utilize these classifiers to enhance the accuracy rate of ASD diagnosis. The development of these automated classifiers based on deep learning algorithms holds promise for enhancing the diagnostic precision and effectiveness of ASD assessment, addressing the pressing need for prompt, efficient, and precise ASD diagnosis.

KEYWORDS

autism spectrum disorder, eye tracking, deep learning, VGG19, MobileNet, DenseNet169, hybrid model

1 Introduction

Autism Spectrum Disorder (ASD) is a neurological condition that involves complications in both spoken and non-spoken communication, as well as challenges in social interaction. It is also marked by monotonous and stereotyped behaviors (1). The intensity of indicators and the impact of ASD differ from one circumstance to another. As to the Centers for Disease Control and Prevention (CDC), the commonness of ASD is assessed to be 1 in 54 children. This condition affects individuals from diverse racial, ethnic, and socioeconomic backgrounds. Furthermore, the prevalence of ASD in boys is four times higher than in girls. Additionally, girls with ASD often have fewer observable symptoms compared to boys (2). Autism is a persistent and enduring condition that remains present throughout a person's whole life (3). Hence, it is of utmost importance to identify ASD at an early stage, since individuals who are identified with ASD during early infancy can greatly benefit from suitable therapies, leading to a favorable long-term result (4).

Facial expressions communicate a wealth of personal, emotional, and social information from early infancy. Even in a short interaction, people may effortlessly focus on and rapidly comprehend the intricate details of a person's face, accurately identifying their emotional state and social situation, and frequently recalling their face later (5). Neuroimaging research has indicated that eye interaction can stimulate brain movement in parts of the brain associated with social interactions. Additionally, studies on human development have provided evidence that infants and young children have a natural inclination to pay attention to and comprehend faces that make direct eye contact. Increasing evidence suggests that ASD is related with an aberrant design pattern of eye tracking conduct (6, 7). Therefore, it is widely accepted that autism is characterized by impairments in facial handling. Nevertheless, the precise attributes of these discrepancies and the correlations among atypical face processing and deviant socio-emotional function in ASD remain inadequately comprehended.

Eye tracking, a non-invasive and straightforward measurement technique, has garnered the attention of scientists in recent years (8–11). The use of eye tracking in ASD research is justified by the correlation between ASD and different attention patterns, which differ from those seen in typical development (12–15). Hence, the use of eye tracking based system to quantify eye activities and gaze designs should assist in understanding the aberrant behavior associated with persons diagnosed with ASD, as well as distinguishing individuals with ASD from typically developing (TD) individuals. Eye tracking is a method used by certain computational systems to aid in the identification of mental problems (16, 17). Eye tracking technology is beneficial in addressing ASD, a neurodevelopmental disease marked by challenges in social communication and repetitive activities. An early indication of ASD is the absence of visual engagement, namely the lack of eye contact. This trait is seen in infants as early as six months old, irrespective of the cultural context in which they are raised. Eye-tracking technology is essential in diagnosing ASD through the analysis of visual patterns (18). A device based on eye-tracking framework classically comprises a high-determination digital camera device and a sophisticated technique based machine learning algorithm that accurately determines the coordinates of eye gaze when persons watch films or pictures. This technology's eye gaze data may help customize therapy to ASD patients' social issues (19). To further understand how eye-tracking biomarkers might

discriminate ASD subgroups, we should explore the effects of closely related mental illnesses such as attention deficit hyperactivity disorder (ADHD), nervousness, and attitude complaint. We may better understand how these variables may affect our ability to distinguish different groups in a medical setting by doing this. Research indicates that children who having the cases of Autism ASD and ADHD tend to have shorter periods of focused attention on faces while looking at static social cues that are not very complex, compared to children who simply have ASD and those with TD (20).

Research has shown that eye-tracking data can be utilized as medical indicators that can be applied in medical health domain to identify ASD in children at an initial state (18). Biomarkers, sometimes referred to as biological markers, are quantifiable and impartial signs that offer insights around a patient's apparent organic state. Bodily fluids or soft tissue biopsies are frequently employed to assess the efficacy of handling for a disease or medicinal disorder.

A crucial element of social interaction is maintaining eye contact, a skill that individuals with ASD often find challenging. Eye tracking technology may be applied to measure the length of time someone maintains eye interaction and the occurrence and track of their eye movements. This provides measurable signs of difficulties in social interactions. Individuals with ASD may also exhibit other irregularities in pictorial processing, including heightened focus on specific details, sensory hypersensitivity, and difficulties with complex visual tasks. Hence, the sophisticated deep learning algorithms, namely MobileNet, VGG19, DenseNet169, and the hybrid of MobileNet-VGG19, were applied for the early-stage recognition of ASD. The primary contributions of this research article are as follows:

- This work introduces a new method for creating eye-tracking event detectors using a deep learning methodology.
- The research asserts that it has attained accuracy (100%) in identifying ASD by employing the MobileNet algorithm. This indicates that the DenseNet169 and hybrid of MobileNet-VGG19 model that was created has demonstrated encouraging outcomes in accurately differentiating persons with ASD from those who do not have ASD, using eye tracking data.
- The proposed methodology was compared with different existing systems that used the same dataset; it is observed that our model achieved high accuracy because we have used a different preprocessing approach from improving dataset.
- This work presents an innovative artificial intelligence (AI) technique for the diagnosis of ASD. Its objective is to differentiate persons with autism from those without utilizing deep learning models, relying on publicly accessible eye-tracking datasets. The suggested approach was evaluated against other existing systems that utilized the same dataset. It was found that the proposed system achieved a high accuracy rate of 100% when compared to one of the deep learning models.

2 Background

ASD can be detected by early screening techniques utilizing DL algorithms. These approaches have become more prominent because of their accuracy rate and capability to grip large volumes of data. It assists experts in automating the diagnostic procedure and reducing

the time spent on tests (21, 22). AI techniques are used in the rehabilitation process to lessen symptoms of ASD. This research analyzes the utilization of DL approaches in the past five years for diagnosing ASD through the application of eye tracking techniques.

Fang et al. (23) introduced a novel method for identifying children with ASD based on stimuli that include the ability to follow someone's gaze. Individuals with ASD exhibited typical patterns of visual attention, especially while observing social settings. The scientists developed a novel deep neural network (DNN) method to abstract distinctive characteristics and categorize children with ASD and healthy controls based on individual images.

Elbattah et al. (24) developed a machine learning (ML)-based approach to aid in the diagnosing process. This approach relies on acquiring knowledge of sequence-oriented patterns in action eye motions. The primary philosophy was to represent eye-tracking data as written documents that analyze a sequence of rapid eye movements (saccades) and periods of gaze fixation. Therefore, the study utilized the natural language processing (NLP) technique to transform the unorganized eye-tracking information.

Li et al. (25) introduced an automated evaluation framework for detecting typical intonation patterns and predictable unique phrases that are important to ASD. Their focus was on the linguistic and communication difficulties experienced by young children with ASD. At first, the scientists utilized the Open SMILE toolkit to extract high-dimensional auditory characteristics at the sound level. They also employed a support vector machine (SVM) backend as the standard baseline. Furthermore, the researchers suggested many DNN arrangements and structures for representing a shared prosody label derived directly from the audio spectrogram after the constant Q transform.

Identification and intervention for ASD have enduring effects on both ASD children as well as their families, necessitating informative, medical, social, and economic assistance to enhance their overall well-being. Professionals have problems in conducting ASD assessments due to the absence of recognized biophysiological diagnostic techniques (25, 26). Therefore, the diagnosis is often determined by a thorough evaluation of behavior, using reliable and valid standardized techniques such as the Autism Diagnostic Observation Schedule (ADOS) (27) and the Autism Diagnostic Interview-Revised (ADI-R) (28). These tools, widely approved in investigation and research domains, are considered the most reliable method for diagnosing ASD in medical situations (29, 30). However, using them involves the use of many materials, a significant amount of time, and is somewhat expensive (25, 26). Furthermore, the diagnostic technique necessitates the involvement of skilled and knowledgeable interviewers, who have the potential to influence the process. This is accompanied by the inclusion of intricate clinical procedures (25, 31). Collectively, these difficulties frequently contribute to a postponed identification, leading to a delay in the initiation of early intervention (26). Research indicates that early treatments for children with ASD before the age of five result in a much higher success rate of 67%, compared to a success rate of just 11% when interventions begin after the age of 5 (32).

Eye-tracking technology is regarded as a beneficial method for doing research on ASD since it allows for the early detection of autism and its characteristics (33, 34) in a more objective and dependable manner compared to traditional assessments (35). There has been a significant rise in the amount of eye-tracking research focused on autism in the past period. This increase can be attributed to improved

accessibility to eye-tracking technology and the development of specialized devices and software that make recording eye-tracking data easier and more cost-effective.

Machine learning and eye-tracking devices are often used together. Data-driven machine learning uses sophisticated mathematics learning, statistical estimates, and information theories (36, 37). This method trains a computer program to examine data and find statistical trends (36–39). Machine learning may improve autism investigation studies by giving an unbiased and reproducible second evaluation (18), including initial autism detection (40), analysis (41), behavior (16), and brain activity (17). Machine learning may also be a viable biomarker-based tool for objective ASD diagnosis (42). ASD is diagnosed via machine learning in IoT systems (43). By helping ASD youngsters learn, assistive technology may improve their lives. This method is backed by studies (44).

Various studies have utilized artificial neural network (ANN) to classify cases of ASD. For example, in ref. (18), the authors investigated the integration of eye-tracking technologies with ANN to assist in the detection of ASD. Initially, other approaches that did not use neural networks were used. The precision achieved by this ensemble of models was adequate. Subsequently, the model underwent testing using several ANN structures. According to the results, the model with a single layer of 200 neurons achieves the maximum level of accuracy. In ref. (45), researchers examined ASD children's visual attention when observing human faces. They extract semantic characteristics using DNN. When viewing human faces, ASD feature maps differ from those without ASD. These feature maps are combined with CASNet features. They contrasted CASNet to six different deep learning based techniques. CASNet has outdone all other models in every situation. The scientists used eye movement patterns to classify children with TD and ASD (46). They combined CNNs with LSTMs. CNN-LSTM extracted features from saliency maps and scan route fixation points. SalGAN pretrained prediction model preprocessed and input network data. The validation dataset accuracy of the proposed model is 74.22%.

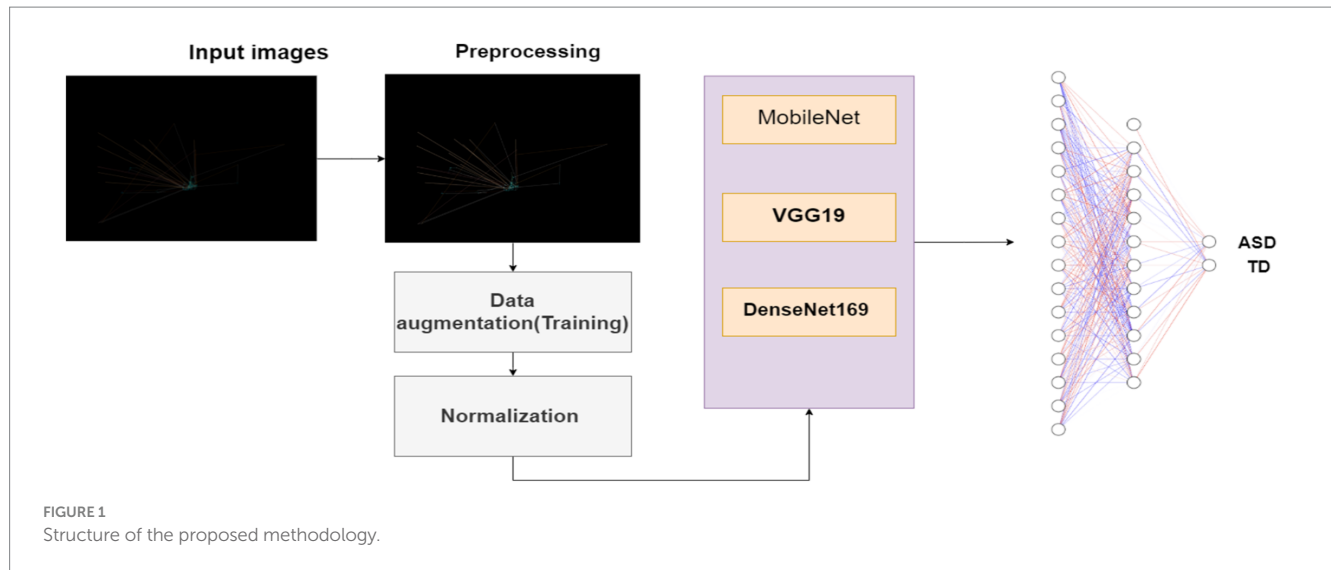
Akter et al. (47) proposed a method that uses transfer learning to identify ASD by analyzing face features. They developed an improved facial recognition system using transfer learning, which can accurately identify individuals with ASD.

Raj and Masood (48) utilized several machine and deep learning techniques with the aim of identifying ASD in youngsters. They utilized three publicly available datasets obtained from the UCI Repository.

Xie et al. (49) proposed a two-stream deep learning network for the detection of visual attention in individuals with ASD. The suggested framework was built using two VGGNets that were derived from the VGG16 architecture and were similar to each other.

3 Methods

This section presents in depth the planned methodology applied to develop ASD detection system using deep learning techniques capable to detect ASD from eye tracking images based features. This methodology includes dataset collection, data preprocessing, deep learning classification model, evaluation metrics and results analysis. The framework of this methodology is shown in Figure 1.



3.1 Dataset

The dataset was obtained from a public repository that contains eye-tracking images. The collection presently comprises 547 images. The default images dimensions were established at 640×480 . More precisely, there were 328 images for the people without ASD, and 219 images for the persons diagnosed with ASD. Figure 2 shows samples of eye-tracking images that were used for examining the proposed methodology.

3.2 Data preprocessing

It is an important step in making the images dataset for training machine learning models. We applied various data preprocessing methods to make certain the dataset is suitable for model training which are discussed as follows.

- **Image Resize:** The first step in data preprocessing encompasses resizing all images in the dataset to a standard size of 640×480 pixels. This ensures uniformity in image measurements and facilitates effective processing during model training.
- **Image Enhancement:** For all images in the dataset used, we applied a specific preprocessing step by improving their resolution by 20% using the Image Enhance module. This enhancement aims to enhance the quality and clarity of the images data, particularly for those where it's considered necessary.
- **Vectorization:** After resizing and enhancing the images, we converted them into numerical arrays using vectorization techniques. This step includes transforming each image into a multi-dimensional array of pixel values, making it compatible with computational operations and deep learning algorithms.
- **Normalization:** after transformation to numerical arrays, we normalized the pixel values to fall within the range of $[0, 1]$. Normalization ensures that the pixel values are scaled appropriately, facilitating more stable and efficient model training by preventing issues related to large variations in input images data.
- **Splitting Data:** Once the images are preprocessed and converted into numerical arrays, we divide the dataset into

three sets namely training, validation, and testing. This step is essential for evaluating model results, as it allows us to train the model on one subset of data, validate its performance on another subset, and finally test its generalization ability on a separate unseen subset.

- **Data Augmentation:** To increase the diversity and robustness of the training dataset, data augmentation techniques, using the Image Data Generator module, was applied. This method involve rotation, shifting, and flipping of images, introducing variations that help avoid overfitting and enhance the model's capability to be generalized to new, unseen images data.

3.3 Improving the deep learning algorithms

3.3.1 The VGG19 model

The VGG19 model (50) is a sequential model architecture constructed in this study for the purpose of detecting ASD based on eye-tracking features. Initially, the model incorporates the pre-trained VGG19 architecture, with the weights initialized from the ImageNet dataset, excluding the fully connected layers, and specifying the input shape to match the dimensions of the input images with size of $(640, 480)$. Subsequently, a GlobalAveragePooling2D layer is added to obtain a condensed representation of the features extracted by VGG19. Following this, several dense layers are appended to the model, comprising 1,024, 128, and 64 neurons, each activated by the rectified linear unit (ReLU) function, to facilitate the learning of intricate patterns within the data. Lastly, a Dense layer with 2 units and a softmax activation function are employed for binary classification, enabling the model to predict the probability of ASD presence. Figure 3 shows the VGG1 model structure.

Upon compiling the model, utilizing the sparse categorical cross-entropy loss function and RMSprop optimizer with a learning rate of 0.0001, data augmentation approach is adopted throughout training process to improve the model's generality competences. Through this architecture, the model aims to effectively discern the presence of ASD based on the provided eye-tracking features, leveraging the robustness of the VGG19

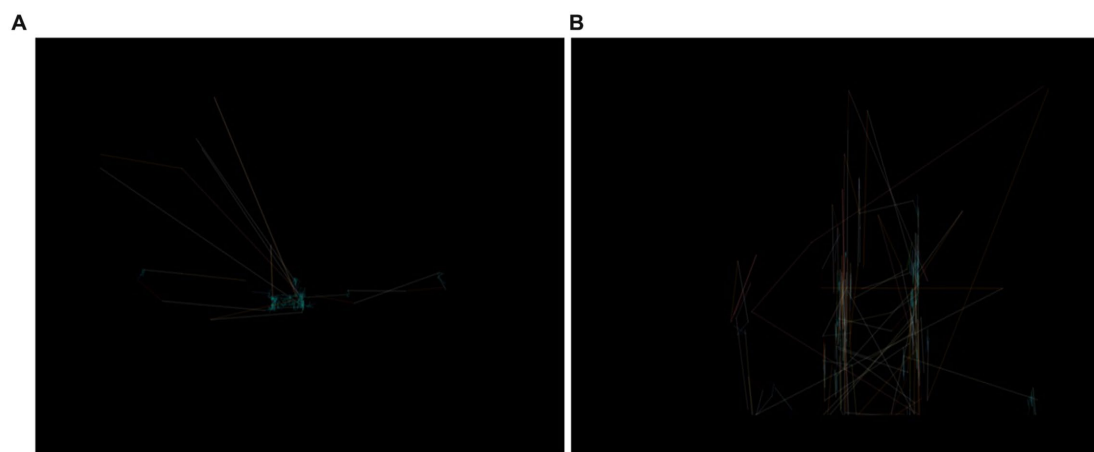


FIGURE 2
Sample of images: (A) ASD (B) TD.

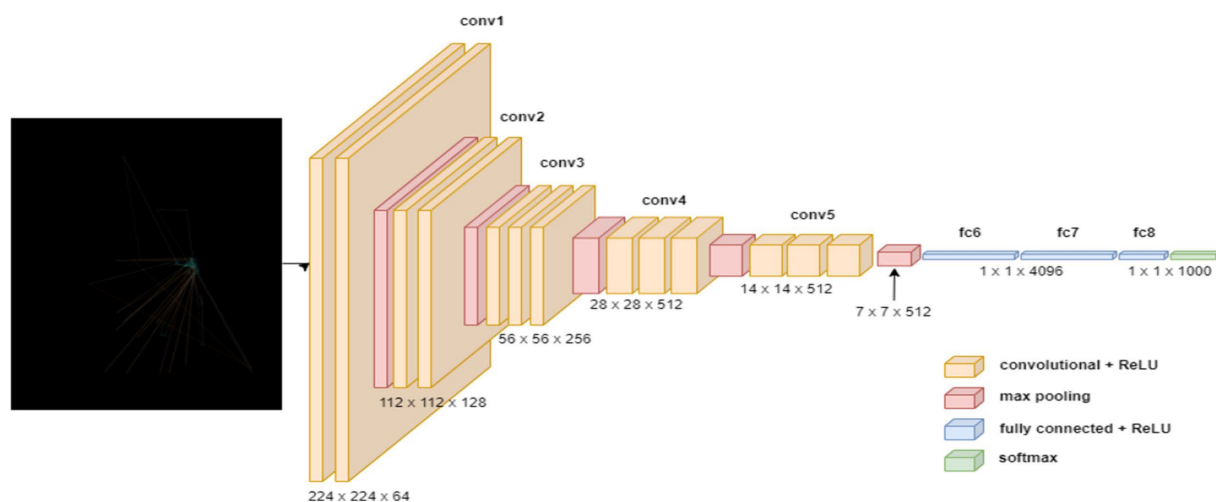


FIGURE 3
Structure of the VGG19 model.

convolutional neural network. Table 1 outlines the parameters of VGG19 model.

3.3.2 The MobileNet model

The MobileNet (51) model architecture has a sequential model structure, which allows for the systematic building of a neural network layer by layer. The MobileNet pre-trained convolutional neural network (CNN) is used as the basis model in this methodology, which is prepared with learnt representations from the ImageNet dataset. However, the fully connected layers of the MobileNet are excluded to facilitate transfer learning. Following integration of the MobileNet base model, a Global Average Pooling 2D layer is used to compress the three-dimensional spaces of the feature maps formed by the convolutional layers. The pooling layer calculates the mean value of each feature map over all spatial locations, resulting in a fixed-size vector representation of the input image, regardless of its size.

Successively, many dense (completely linked) layers are added to capture more complex characteristics and perform classification tasks. The dense layers are composed of 1,024, 128, and 64 neurons, respectively, each of which is activated using the ReLU activation function. The ReLU activation function is selected for its capacity to introduce non-linearity, hence improving the complexity of the model and the efficiency of training.

The classification layer of the model that is named as output layer consists of a dense layer with 2 units, representing the two classes for binary classification (ASD or TD). These units are activated using the softmax function. This function generates probability for every class. This model architecture seeks to utilize the data obtained by MobileNet and conduct classification based on these features. It then proceeds to fine-tune the dense layers to suit the particular purpose of ASD detection using eye-tracking features. The MobileNet architecture is presented in Figure 4 and model's parameters are listed in Table 2.

3.3.3 The DenseNet169 model

We also applied the DenseNet169 (52) model as the base, which is tailored for ASD detection based on eye-tracking features. Utilizing pre-trained weights from the ImageNet dataset, the model excludes the fully connected layers for transferring learning tasks. After integrating a Global Average Pooling 2D layer to condense feature maps, dense layers capture higher-level features. Dropout layers mitigate overfitting, and the output layer, activated by softmax, produces class probabilities. With frozen base model layers, the model is compiled with appropriate functions and benefits from learning rate scheduling. Data augmentation enhances training, aligning with the ASD detection task's needs. Figure 5 displays the structure of DenseNet169 model, and Table 3 outlines the parameters used in DenseNet169 model.

3.3.4 The hybrid model

The framework of this a combination model employs the capacities of two solidified convolutional neural network (CNN)

structures, VGG19 (46) and MobileNet (51) models, to enhance its efficacy in recognizing ASD using eye-tracking features. At first, the model provides in the pre-trained VGG19 and MobileNet structures, although without their completely connected layers. It then freezes all layers to maintain their learnt representations. Global Average Pooling 2D layers are subsequently employed to acquire feature representations from the output of each model. These representations are merged to develop a united feature vector, which is then handled through numerous robust layers to capture complicated data patterns. Following that, the model is collected utilizing acceptable loss and optimization functions, while data augmentation approaches are employed during training to improve its generalization capability. This hybrid model aims to improve classification accuracy in the ASD detection task by combining the features learned by VGG19 and MobileNet. By using the capabilities of both architectures, it seeks to attain heightened accuracy. Table 4 summarizes the parameters used in the hybrid VGG19-MobileNet model, and Figure 6 displays the structure of hybrid model.

TABLE 1 Parameters of the VGG19 model.

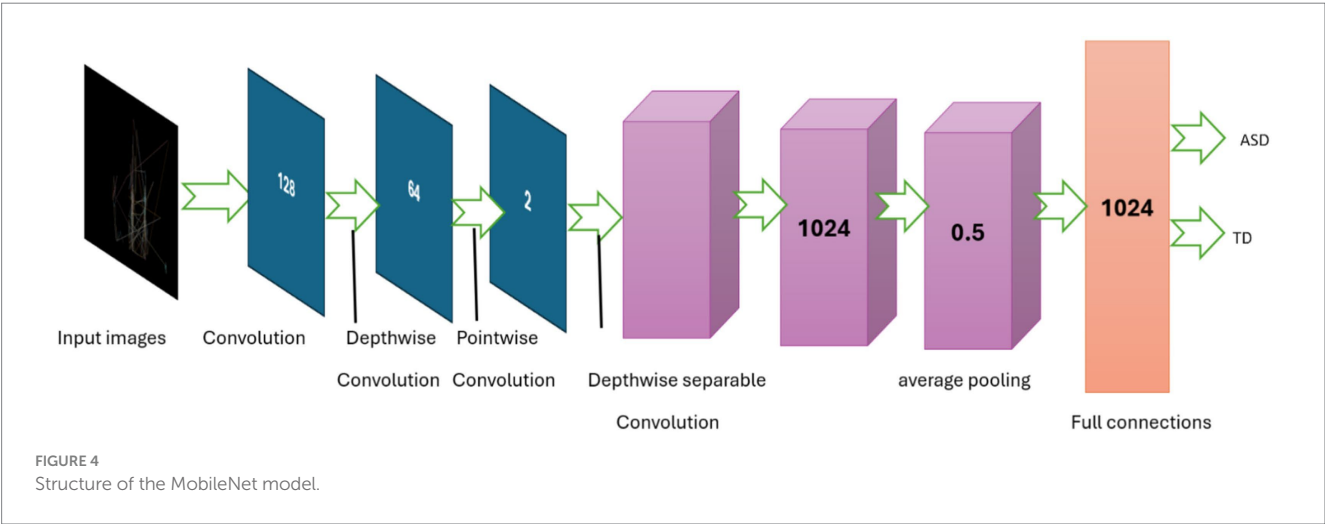
Parameter	Description
Architecture	Sequential
Base model	VGG19 (pre-trained on ImageNet)
Input shape	(640, 480, 3)
Global pooling layer	Global Average Pooling 2D
Dense layers	1,024, 128, 64 neurons with ReLU activation
Output layer	Dense layer with 2 units, softmax activation (binary classification)
Loss function	Sparse categorical cross-entropy
Optimizer	RMSprop with learning rate of 0.0001
Metrics	Accuracy
Data augmentation	Applied during training using Image Data Generator
Training batch size	16
Validation batch size	32
Number of epochs	100

3.4 Evaluation metrics

Assessing the performance and testing results obtained by the proposed deep learning models namely MobileNet, VGG19, DenseNet169 and hybrid of MobileNet-VGG19 are crucial for gauging the effectiveness of the models. The evaluation measures provide an alternative perspective on the model's advantages and disadvantages. There are several matrices used to quantify performance, including accuracy, recall (sensitivity), specificity, and F1-score. These evaluation matrices, expressed by Equations (1–4), can be calculated from the confusion matrix.

$$Accuracy = \frac{TP + TN}{FP + FN + TP + TN} \times 100 \tag{1}$$

$$Recall = Sensitivity = \frac{TP}{TP + FP} \times 100\% \tag{2}$$



$$Specificity = \frac{TN}{TN + FN} \times 100\%$$

(3)

$$F1 - score = 2 * \frac{precision \times Recall}{precision + Recall} \times 100\%$$

(4)

where *TP*, *TN*, *FP*, and *FN* stand for true positives, true negatives, false positives, and false negatives, respectively.

4 Results

This section focuses on the gained testing results of each model for spotting ASD using eye-tracking characteristics. The testing

TABLE 2 Parameters of the MobileNet model.

Parameter	Description
Architecture	Sequential
Weights	Image net
Input shape	(640, 480, 3)
Pooling layer	0.5
Dense	256
Output layer	Dense layer with 2 units, softmax activation (binary classification)
Loss function	Sparse categorical cross-entropy
Optimizer	adam with learning rate of 0.0001
Metrics	Accuracy
Data augmentation	Applied during training using Image Data Generator
Training batch size	16
Validation batch size	32
Number of epochs	100

process included evaluation of four separate deep learning models: MobileNet, VGG19, DenseNet169, and a combination of VGG19 and MobileNet called the hybrid model.

4.1 Models' configuration

The efficacy of the advanced deep learning algorithm was evaluated in a specific environment to identify ASD using an eye-tracking method. Table 5 presents the environment of the DL models.

4.2 Splitting dataset

The dataset was segregated into three subsets: training, testing, and validation. Table 6 displays the specific division that was employed in the proposed method for diagnosing ASD.

4.3 The test classification results of the MobileNet model

The MobileNet model demonstrated outstanding performance in all parameters, attaining perfect precision, recall, and F1-score for both ASD and non-ASD classes. This indicates that the model accurately categorized all cases of ASD and non-ASD without any incorrect positive or negative predictions, resulting in a remarkable overall accuracy of 100%. Table 7 presents the testing classification results of MobileNet.

The impressive performance of MobileNet underscores its efficacy in accurately recognizing instances of ASD through the utilization of eye-tracking characteristics. Figure 7 depicts the confusion matrix, which reveals that 33 images were correctly identified as true negatives (TN), 22 images were correctly classified as true positives (TP), and there were no instances of false positives (FP) or false negatives (FN). Based on the empirical data, it has been determined that the MobileNet model obtained a high level of accuracy.

Figure 8 displays the performance of the MobileNet model. The model's accuracy exhibited a progressive increase in validation

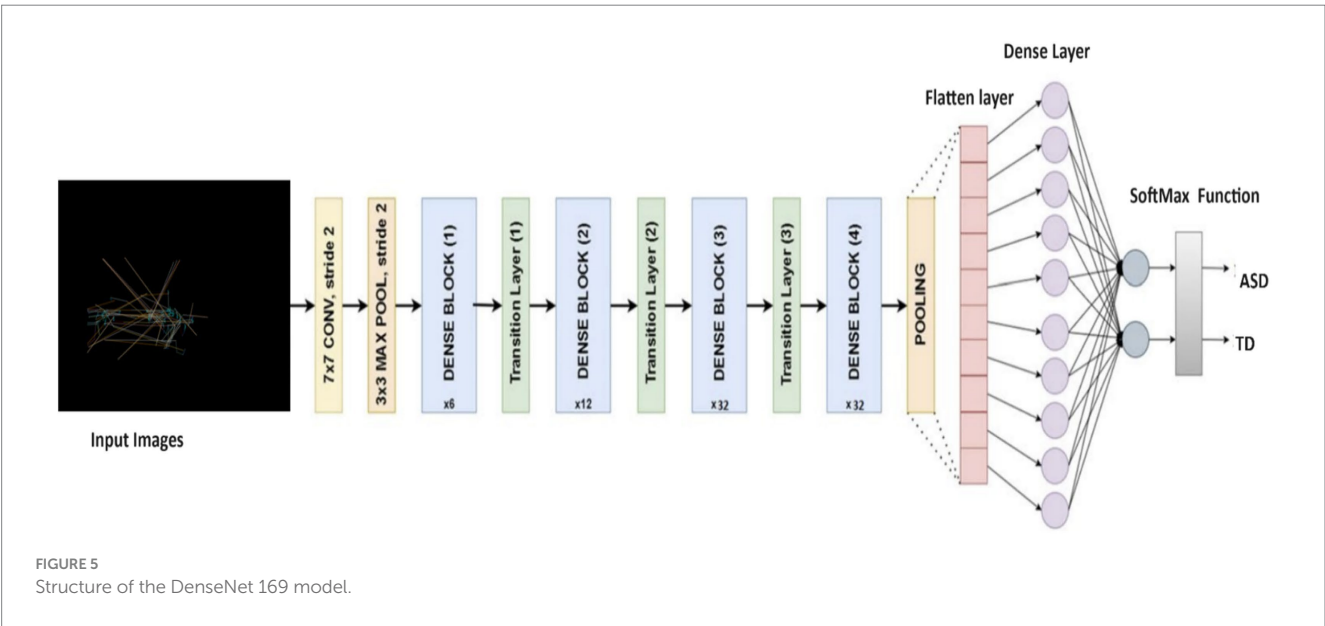
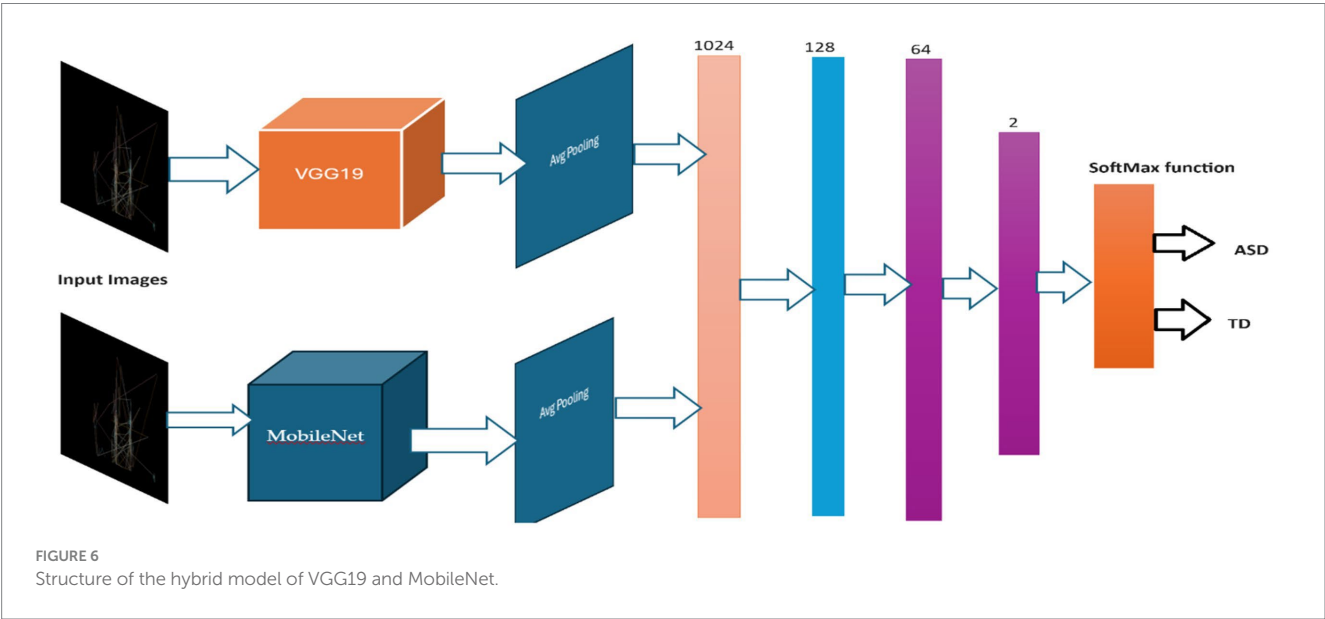


TABLE 3 Parameters of the DenseNet169 model.

Parameter	Description
Base model	DenseNet169 pre-trained CNN initialized with ImageNet weights, excluding fully connected layers
Global Average Pooling 2D	Condenses spatial dimensions of feature maps
Dense layers	512 and 256 neurons with ReLU activation, capturing higher-level features
Dropout layers	Dropout rate of 0.5 for regularization, preventing overfitting
Output layer	Dense layer with 2 units for binary classification, activated by softmax
Frozen base model layers	Retains learned features during training
Loss function	Sparse categorical cross-entropy
Optimizer	RMSprop with learning rate of 0.0001
Learning rate scheduler	Reduces learning rate based on validation loss
Data augmentation	Applied during training to improve generalization

TABLE 4 Parameters of the hybrid model.

Parameter	Description
Pre-trained models	VGG19 and MobileNet are used as pre-trained CNN architectures.
Trainable layers	All layers in both VGG19 and MobileNet models are frozen
Output layers	Global Average Pooling 2D layers are added to the output of each model
Concatenated output	The outputs of both models are concatenated to create a fused feature vector
Dense layers	Several dense layers with ReLU activation functions: 1024, 128, and 64 units
Output activation	Softmax activation function is used for the output layer
Loss function	Sparse categorical cross entropy loss function is used
Optimizer	RMSprop optimizer with a learning rate of 0.0001 is employed
Data augmentation	Image data augmentation techniques are applied during training
Training epochs	The model is trained for 100 epochs
Batch size	Batch size is set to 16 for training and 32 for validation



performance, starting at 50% and reaching 100%. In contrast, the accuracy in training performance had a smoothing effect, starting at 65% and also reaching 100%. The decline in the MobileNet starting and validation performance has resulted in a fall of 1.6% to reach 0.0. This confirms that the MobileNet model has achieved a high percentage score.

4.4 Testing results of the VGG19 model

This subsection introduces the testing classification results gained by the VGG19 model which achieved an accuracy of 87%, its recall,

TABLE 5 Environment of the proposed DL.

GPU	GPU T4 X 2 Kaggle
Memory	4GB
Language	Python
TensorFlow	
Keras	
Panda	

TABLE 6 Dataset.

Training set	77.78%
Validation set	22.22%
Testing set	10%

TABLE 7 Testing classification results of the MobileNet model.

Class	Precision (%)	Recall (%)	F1-score (%)	Support (%)	Accuracy (%)
Non ASD	100	100	100	33	100
ASD	100	100	100	22	
Macro average	100	100	100	55	

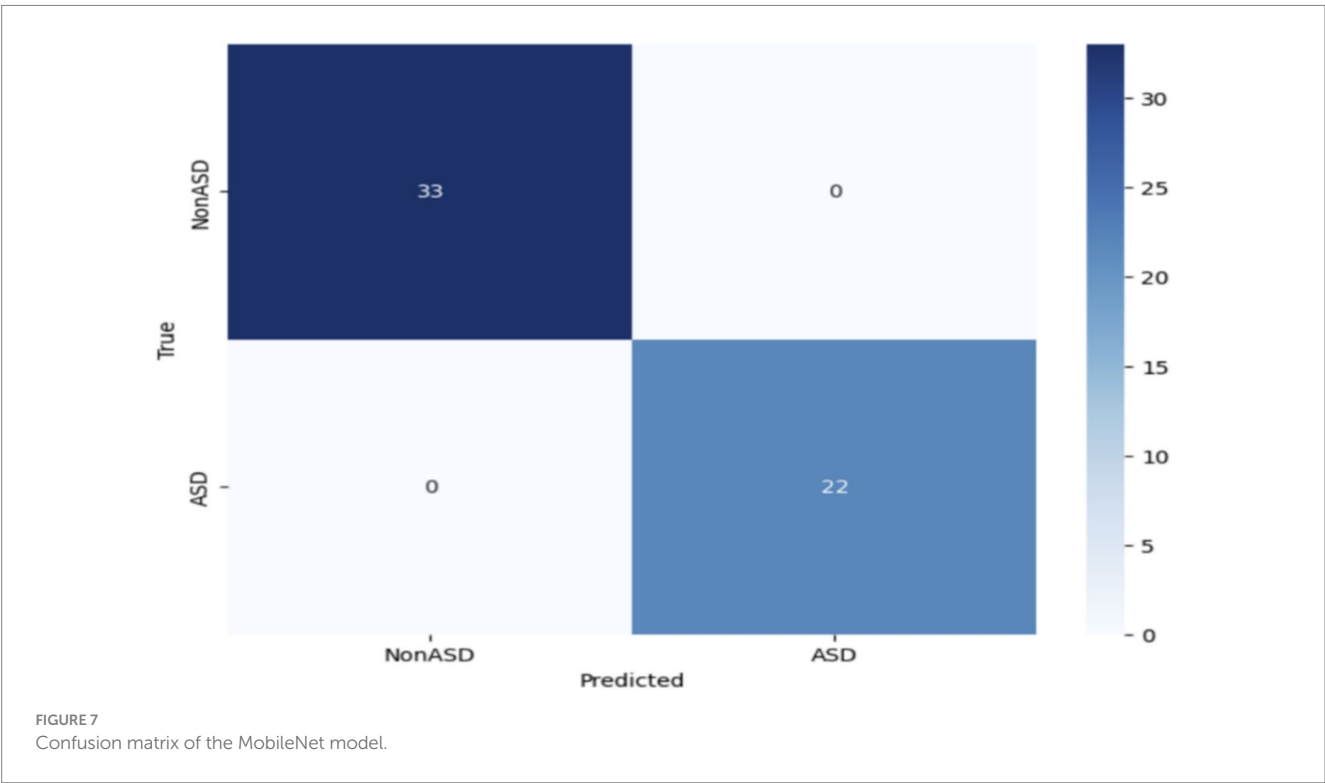
precision and F1-score for the ASD class were pointedly lower than those for the non-ASD class. This suggests that although the model demonstrated good performance in appropriately categorizing individuals without ASD, it encountered difficulties in correctly identifying individuals with ASD, resulting in a greater incidence of false negatives. Table 8 summarizes and presents the testing results of VGG19 model.

Further modification or improvement of the VGG19 design may be required to enhance its effectiveness in diagnosing ASD. Figure 9 depicts the confusion matrix of the VGG19 model used to categorize Autism Spectrum Disorder (ASD) using an eye-tracking method. The VGG19 model correctly identified 31 images as true negatives (TN) and 19 images as true positives (TP). However, it misclassified 3 images and incorrectly classified 2 images as false negatives (FN).

Figure 10 illustrates the process of validating and training the VGG19 model. The VGG19 model achieved a validation accuracy of 87%. The VGG19 model attained an accuracy rate of 89% in diagnosing Autism Spectrum Disorder (ASD) using the eye-tracking dataset during training. The loss of the VGG19 model decreased to 0.3.

4.5 Testing classification results of the hybrid VGG19-MobileNet model

The hybrid VGG19-MobileNet model exhibited strong performance, with a 91% accuracy with well-balanced precision, recall, and F1-score for both ASD and non-ASD categories. The hybrid model successfully utilized the advantageous qualities of both VGG19 and MobileNet architectures, leading to enhanced classification performance. Table 9 presents the testing classification results obtained by the hybrid VGG19-MobileNet model.



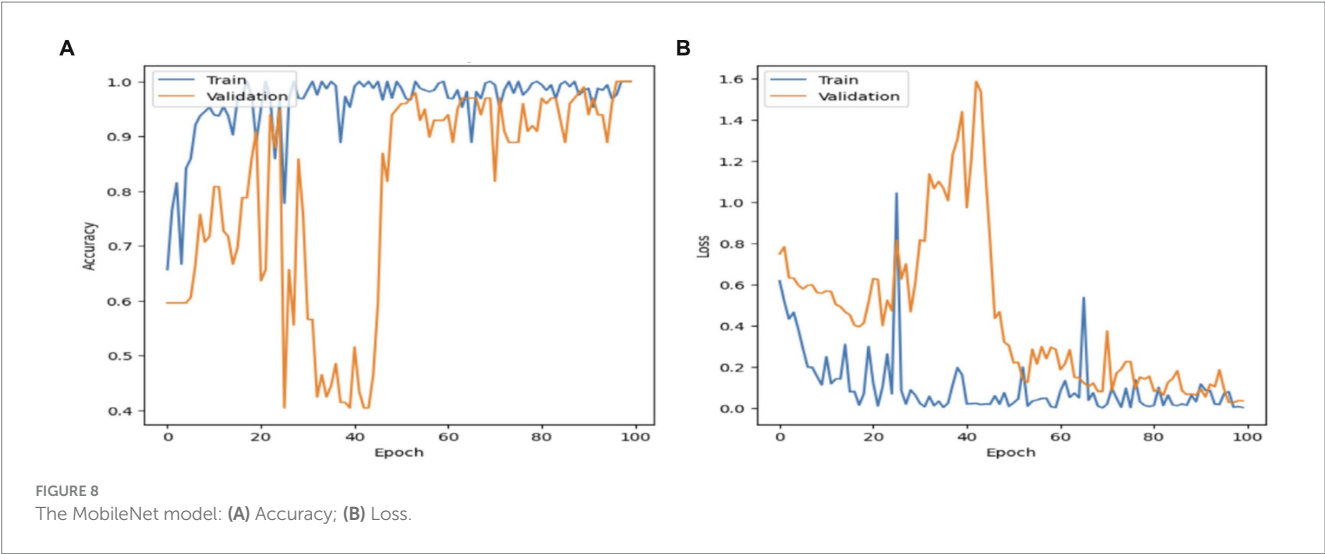
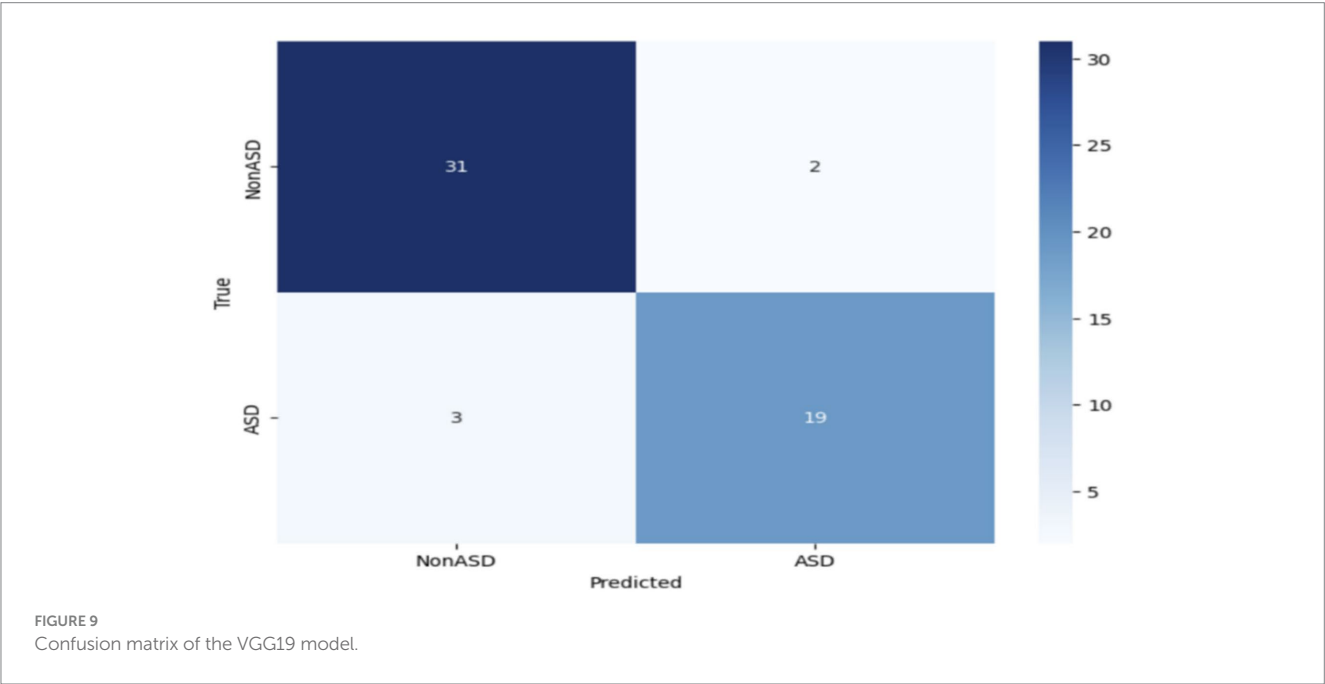


TABLE 8 Testing results of the VGG19 model.

Class	Precision (%)	Recall (%)	F1-score (%)	Support (%)	Accuracy (%)
Non ASD	82	100	90	33	87
ASD	100	68	81	22	
Macro average	91	84	86	55	



The model’s ability to accurately differentiate between cases of ASD and non-ASD highlights its potential utility in clinical settings for diagnosing ASD based on eye-tracking features. Figure 11 presents the confusion matrix of the hybrid VGG19-MobileNet model. In this hybrid model, 31 images were accurately labeled as TD and 19 images were accurately classified as ASD (autism spectrum disorder). The hybrid model correctly classifies 3 images as FP and incorrectly classifies 2 images as FN.

The results performance of the VGG19-MobileNet model is depicted in Figure 12. The VGG19-MobileNet model obtained a

validation accuracy of 91% and a training accuracy of 92%. The hybrid model had a reduction from 0.6 to 0.4.

4.6 Testing results of the DenseNet169 model

The DenseNet169 model attained an accuracy of 78%, exhibiting superior precision, recall, and F1-score for the non-ASD class in comparison to the ASD class. This indicates that although the model

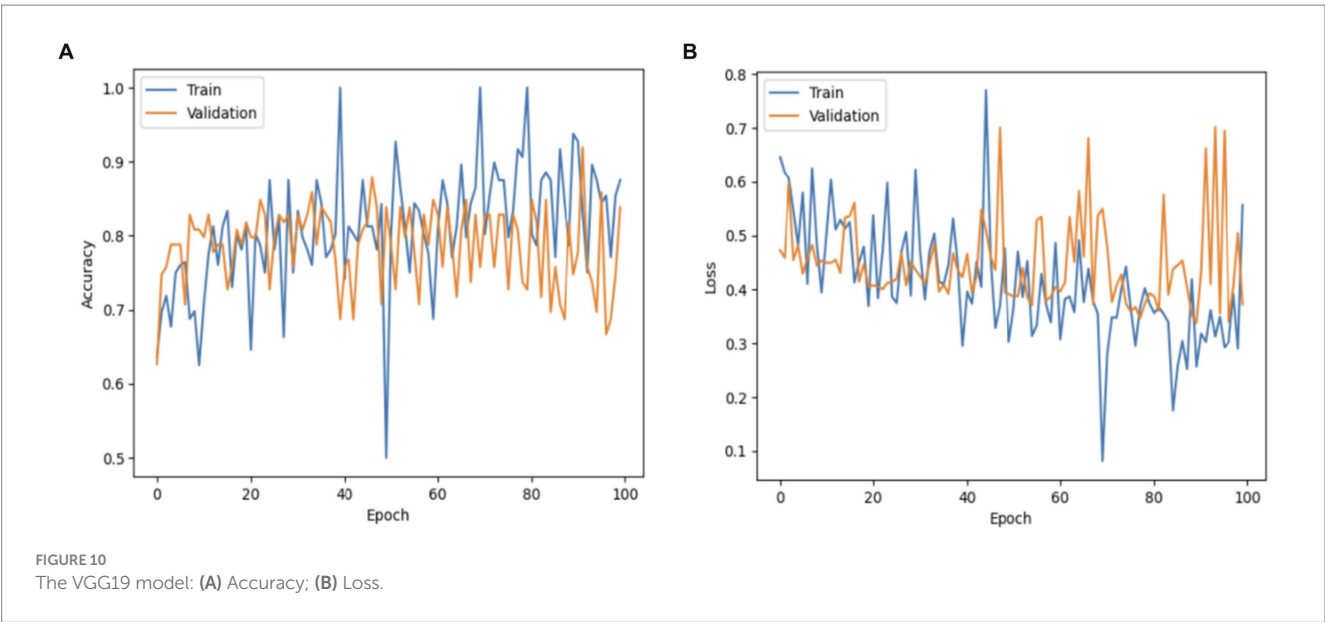
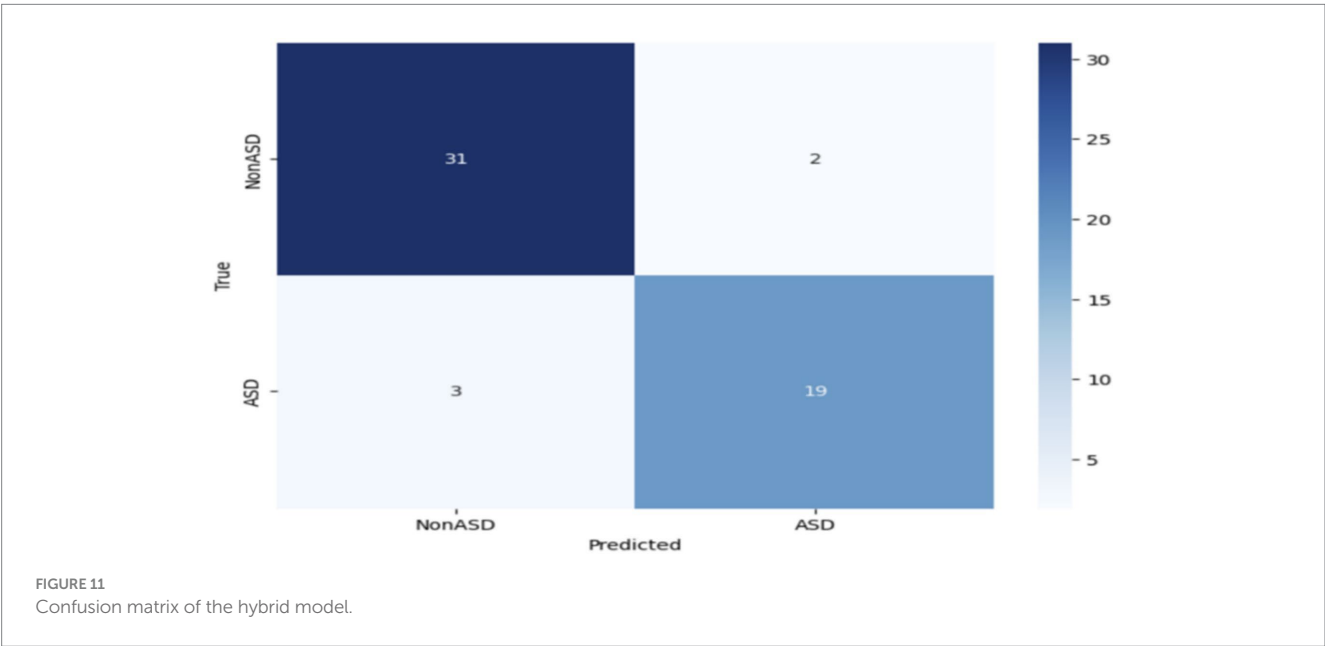


TABLE 9 Testing results of the hybrid model.

Class	Precision (%)	Recall (%)	F1-score (%)	Support (%)	Accuracy (%)
Non ASD	91	94	93	33	91
ASD	90	86	88	22	
Macro average	91	90	90	55	



performed well in accurately categorizing those without ASD, its ability to identify individuals with ASD was comparatively less effective. Table 10 summarizes the testing classification results of the DenseNet169 model.

The elevated rate of false negatives in ASD cases highlights possible opportunities for enhancing the model's ability to detect ASD-related characteristics. In general, although all models demonstrated potential in detecting ASD, there is a need for more

improvement and optimization of model structures to boost the accuracy and precision of ASD diagnosis using eye-tracking data.

5 Discussion

ASD is a neurodevelopmental condition marked by enduring difficulties in social interaction, communication, and restricted or

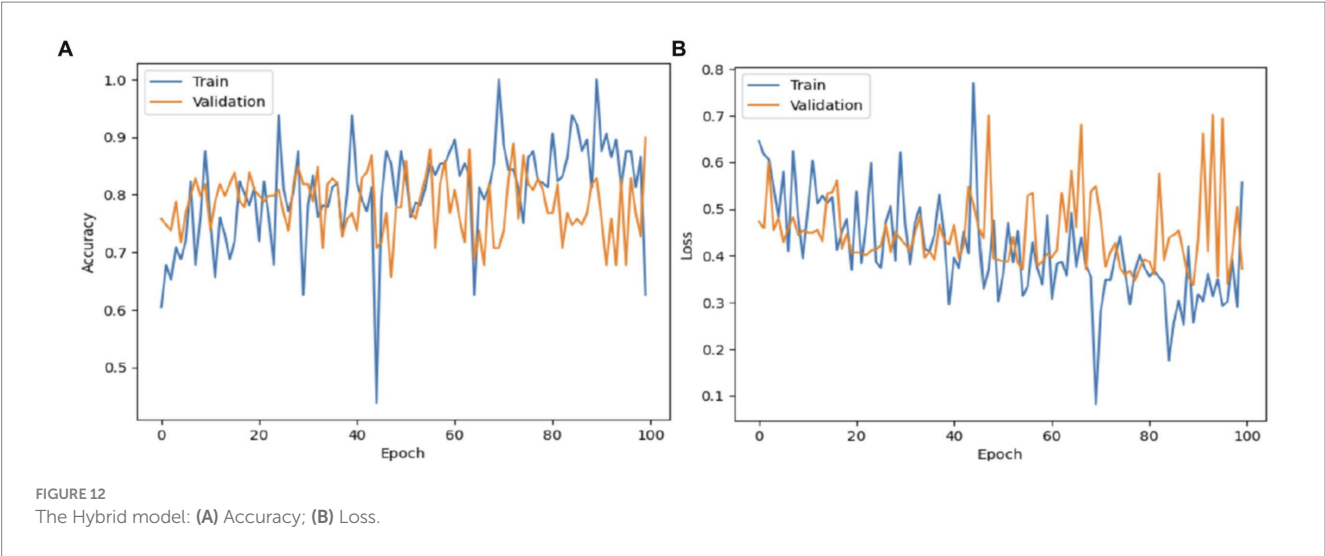


TABLE 10 Testing results of the DenseNet169 model.

Class	Precision (%)	Recall (%)	F1-score (%)	Support (%)	Accuracy (%)
Non ASD	74	97	84	33	78
ASD	92	50	65	22	
Macro average	83	73	74	55	

repetitive behaviors. People with Autism Spectrum Disorder (ASD) can display a diverse array of symptoms and levels of functioning, resulting in significant variation within the spectrum. Eye-tracking technology is the technique of observing and documenting the movement of a person's eyes in order to examine different aspects of visual attention, perception, and cognitive processing. Eye-tracking studies in individuals with ASD commonly examine gaze fixation patterns, saccades (quick eye movements), and pupil dilation to explore disparities in visual processing and social attention between individuals with ASD and those who are typically developing.

The experimental results presented in this study demonstrate the efficacy of several convolutional neural network (CNN) models in detecting and predicting Autism Spectrum Disorder (ASD) by utilizing eye-tracking features. The classification accuracy, precision, recall, and F1-score of each model offer valuable insights into their efficacy in detecting ASD cases using eye movement patterns.

The MobileNet model exhibited outstanding performance, attaining flawless precision, recall, and F1-score for both ASD and non-ASD categories. This indicates that MobileNet successfully diagnosed all cases of ASD and non-ASD, demonstrating its potential usefulness in diagnosing ASD using eye-tracking data.

Although the VGG19 model achieved an accuracy of 87%, its precision, recall, and F1-score for the ASD class were somewhat lower, suggesting a higher occurrence of false negatives. This implies that VGG19 might have difficulties in reliably detecting cases of ASD solely based on eye movement patterns.

The DenseNet169 model attained an accuracy of 78%, exhibiting superior precision, recall, and F1-score for the non-ASD class in comparison to the ASD class. This disparity suggests possible

constraints in the model's ability to detect ASD-related eye movement characteristics, resulting in an increased occurrence of incorrect negative diagnoses for individuals with ASD.

The hybrid VGG19-MobileNet model exhibited strong performance, with a 91% accuracy with well-balanced precision, recall, and F1-score for both ASD and non-ASD categories. This suggests that the hybrid model successfully utilized the advantages of both VGG19 and MobileNet architectures to enhance ASD identification using eye-tracking features.

Figure 13 displays the receiver operating characteristics (ROC) findings of the proposed deep learning (DL) model. The MobileNet model earned a high accuracy score of 100%, while both the VGG19 and hybrid models achieved the same accuracy score of 96%.

In summary, the experimental results highlight the capability of CNN models, specifically MobileNet and the hybrid VGG19-MobileNet model, to accurately detect ASD cases using eye-tracking data. However, additional study is required to optimize the design of models and increase their ability to detect patterns in eye movements associated to ASD. This will ultimately lead to better accuracy in diagnosing and treating ASD. The proposed system was compared to several current eye-tracking systems (46–48), as seen in Table 11 and Figure 14. Our enhanced MobileNet model achieved a perfect score of 100%, surpassing all other current systems.

6 Conclusion

Eye tracking is a commonly used method for detecting ASD in both young children and adults. Research including eye

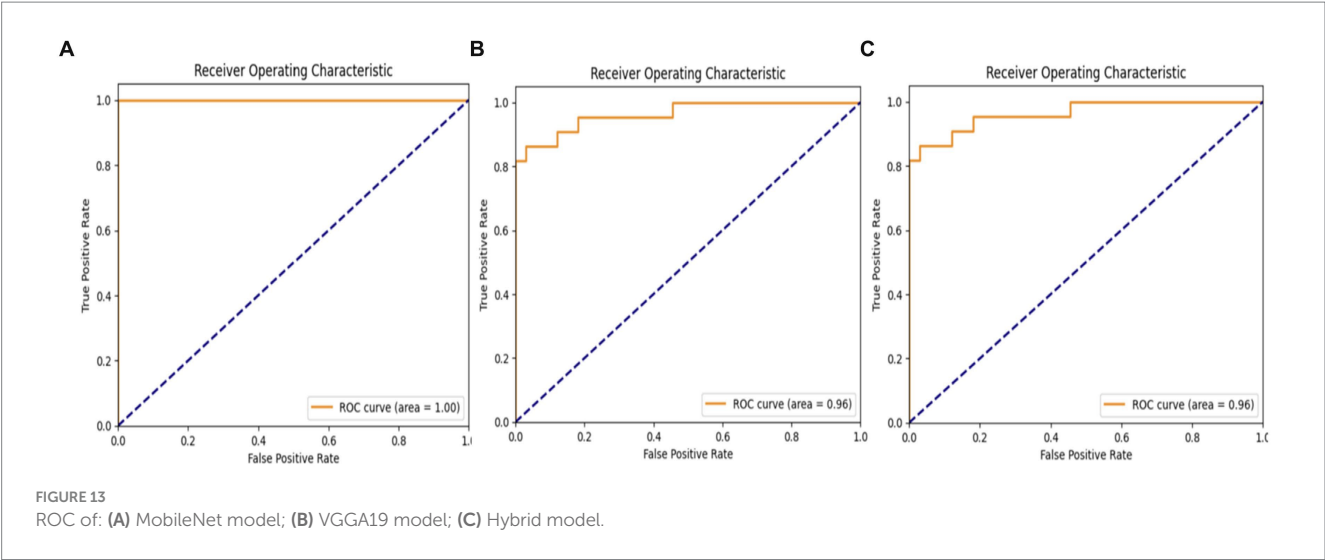
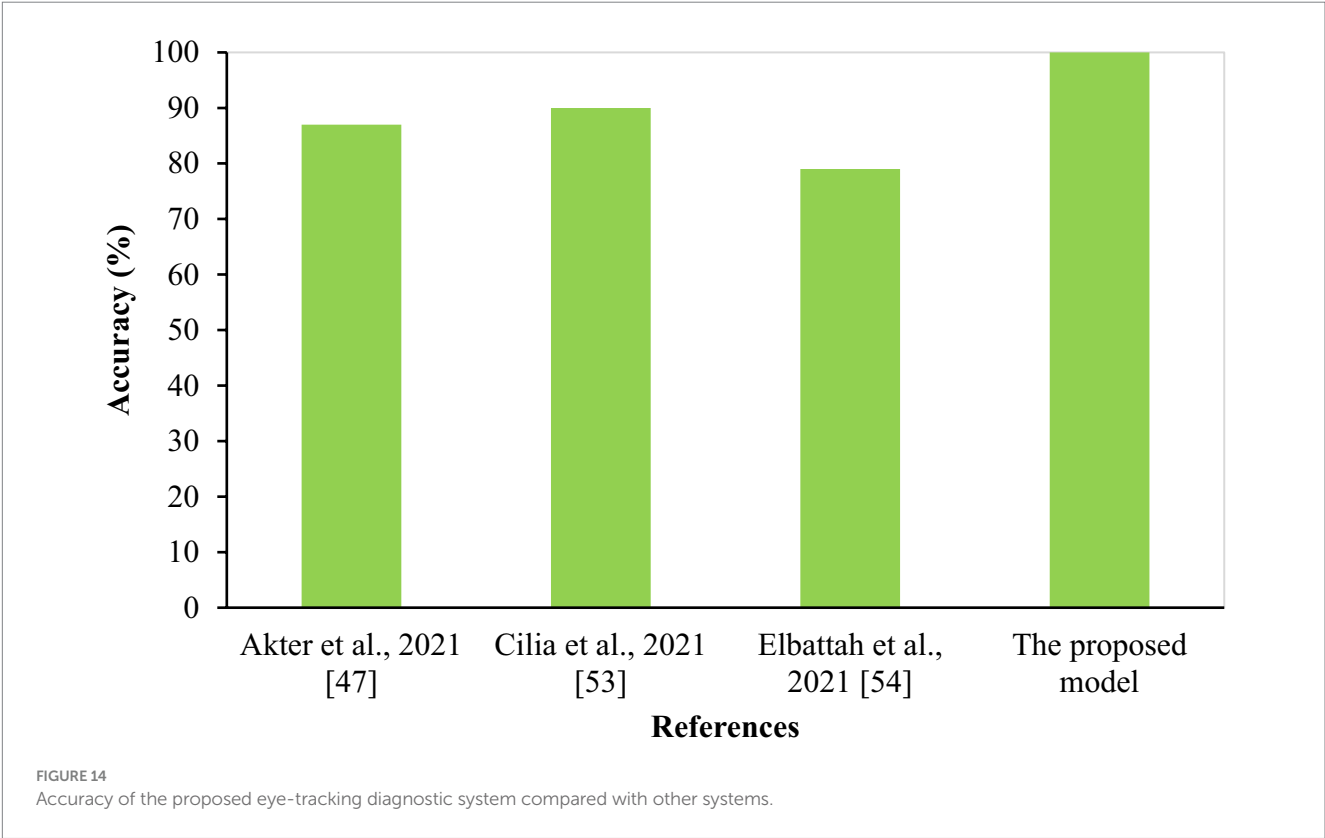


TABLE 11 Results of the proposed eye-tracking diagnosis system compared with other systems.

Authors, years	Dataset	Approach	Accuracy %
Akter et al., 2021 (47)	Same	DT, SVM, LR, KNN, and MLP	Accuracy (87%), and AUC (79%)
Cilia et al., 2021 (53)	Same	CNN	90%
Elbattah et al., 2021 (54)	Same	Variational Autoencoder (VAE)	79%
This proposed model	Same	MobileNet	Accuracy (100%), and ROC (100%)



tracking has revealed that individuals with autism have distinct gaze patterns compared to normally developing individuals. Various diagnostic procedures have been considered for the diagnosis of ASD, such as parent interviews, homogenous behavioral appraisals, and neurological examinations. Eye-tracking technology has gained significance for supporting the study and analysis of autism. This research presents a methodology that utilizes advanced deep learning algorithms, including MobileNet, VGG19, DenseNet169, and a hybrid of MobileNet-VGG19, to analyze and display the eye-tracking patterns of persons diagnosed with ASD. The study specifically focuses on children and adults in the initial phases of growth. The primary concept is to convert the movement patterns of the eye into a visual depiction, allowing for the use of image-based methods in activities connected to diagnosis. The visualizations generated are freely accessible as an image collection for use by other studies seeking to explore the capabilities of eye-tracking in the setting of Autism ASD. The collection consists of 547 images, with 328 images representing persons without ASD and 219 images representing those diagnosed with ASD. The MobileNet model scored high accuracy 100%, the proposed methodology was compared with different with existing ASD model, it is investigated that our model out performance.

An important avenue for future study is to expand the sample size by include a wider range of participants, including a greater number of persons with ASD and TD individuals. By increasing the size of the sample, researchers might potentially uncover additional patterns and subtleties in the data.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://figshare.com/articles/dataset/Visualization_of_Eye-Tracking_Scanpaths_in_Autism_Spectrum_Disorder_Image_Dataset/7073087.

References

- Bölte S, Hallmayer J. FAQs on autism, Asperger syndrome, and atypical autism answered by international experts. Cambridge, MA, USA: Hogrefe Publishing (2010).
- CDC. (2018). Screening and diagnosis of autism Spectrum disorder for healthcare providers. Available at: <https://www.cdc.gov/ncbddd/autism/hcp-screening.html> (Accessed April 13, 2024).
- Maenner MJ, Shaw KA, Baio J. Prevalence of autism Spectrum disorder among children aged 8 years. Atlanta, GA, USA: CDC (2020).
- Dawson G. Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Dev Psychopathol.* (2008) 20:775–803. doi: 10.1017/S0954579408000370
- Sasson NJ. The development of face processing in autism. *J Autism Dev Disord.* (2006) 36:381–94. doi: 10.1007/s10803-006-0076-3
- Senju A, Johnson MH. The eye contact effect: mechanisms and development. *Trends Cogn Sci.* (2009) 13:127–34. doi: 10.1016/j.tics.2008.11.009
- Senju A, Johnson MH. Atypical eye contact in autism: models, mechanisms and development. *Neurosci Biobehav Rev.* (2009) 33:1204–14. doi: 10.1016/j.neubiorev.2009.06.001
- Klin A, Jones W, Schultz R, Volkmar F, Cohen D. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch Gen Psychiatry.* (2002) 59:809–16. doi: 10.1001/archpsyc.59.9.809
- Falck-Ytter T, Bolte S, Gredeback G. Eye tracking in early autism research. *J Neurodev Disord.* (2013) 5:28. doi: 10.1186/1866-1955-5-28
- Boraston Z, Blakemore SJ. The application of eye-tracking technology in the study of autism. *J Physiol.* (2007) 581:893–8. doi: 10.1113/jphysiol.2007.133587
- Black MH, Chen NTM, Iyer KK, Lipp OV, Bölte S, Falkmer M, et al. Mechanisms of facial emotion recognition in autism spectrum disorders: insights from eye tracking and electroencephalography. *Neurosci Biobehav Rev.* (2017) 80:488–515. doi: 10.1016/j.neubiorev.2017.06.016
- Fujioka T, Inohara K, Okamoto Y, Masuya Y, Ishitobi M, Saito DN, et al. Gazefinder as a clinical supplementary tool for discriminating between autism spectrum disorder and typical development in male adolescents and adults. *Mol Autism.* (2016) 7:19. doi: 10.1186/s13229-016-0083-y
- Klin A, Lin DJ, Gorrindo P, Ramsay G, Jones W. Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature.* (2009) 459:257–61. doi: 10.1038/nature07868
- Frazier TW, Klingemier EW, Parikh S, Speer L, Strauss MS, Eng C, et al. Development and validation of objective and quantitative eye tracking-based measures of autism risk and symptom levels. *J Am Acad Child Adolesc Psychiatry.* (2018) 57:858–66. doi: 10.1016/j.jaac.2018.06.023
- Papagiannopoulou EA, Chitty KM, Hermens DF, Hickie IB, Lagopoulos J. A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders. *Soc Neurosci.* (2014) 9:610–32. doi: 10.1080/17470919.2014.934966
- Zhou Y, Yu F, Duong T. Multiparametric MRI characterization and prediction in autism spectrum disorder using graph theory and machine learning. *PLoS One.* (2014) 9:e90405. doi: 10.1371/journal.pone.0090405

Author contributions

NA: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft. MA-A: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Writing – original draft. MA-Y: Visualization, Writing – original draft, Writing – review & editing. NF: Conceptualization, Formal analysis, Writing – review & editing. ZK: Conceptualization, Formal analysis, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors extend their appreciation to the King Salman Center for Disability Research for funding this work through Research Group no. KSRG-2023-500.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

17. Minissi ME, Giglioli IAC, Mantovani F, Raya MA. Assessment of the autism spectrum disorder based on machine learning and social visual attention: a systematic review. *J Autism Dev Disord.* (2021) 52:2187–202. doi: 10.1007/s10803-021-05106-5
18. Carette R., Elbattah M., Cilia F., Dequen G., Guérin J.L., Bosche J. (2019). Learning to predict autism Spectrum disorder based on the visual patterns of eye-tracking Scanpaths. In Proceedings of the HEALTHINF, Prague, Czech Republic, 22–24 February 2019, pp. 103–112. doi: 10.5220/0007402601030112
19. Kanhirakadavath MR, Chandran MSM. Investigation of eye-tracking scan path as a biomarker for autism screening using machine learning algorithms. *Diagnostics.* (2022) 12:518. doi: 10.3390/diagnostics12020518
20. Shic F, Naples AJ, Barney EC, Chang SA, Li B, McAllister T, et al. The autism biomarkers consortium for clinical trials: evaluation of a battery of candidate eye-tracking biomarkers for use in autism clinical trials. *Mol Autism.* (2022) 13:15. doi: 10.1186/s13229-021-00482-2
21. Haq AU, Li JP, Ahmad S, Khan S, Alshara MA, Alotaibi RM. Diagnostic approach for accurate diagnosis of COVID-19 employing deep learning and transfer learning techniques through chest X-ray images clinical data in E-healthcare. *Sensors.* (2021) 21:8219. doi: 10.3390/s21248219
22. Ruchi S, Singh D, Singla J, Rahmani MKI, Ahmad S, Rehman MU, et al. Lumbar spine disease detection: enhanced CNN model with improved classification accuracy. *IEEE Access.* (2023) 11:141889–901. doi: 10.1109/ACCESS.2023.3342064
23. Fang Y., Duan H., Shi F., Min X., Zhai G. (2020). Identifying children with autism spectrum disorder based on gaze-following. In Proc. in 2020 IEEE Int. Conf. on Image Processing (ICIP). Abu Dhabi, United Arab Emirates, pp. 423–427. doi: 10.1109/ICIP40778.2020.9190831
24. Elbattah M., Guérin J.L., Carette R., Cilia F., Dequen G. (2020). NLP-based approach to detect autism spectrum disorder in saccadic eye movement. In Proc. in 2020 IEEE Symp. Series on Computational Intelligence (SSCI), Canberra, Australia, pp. 1581–1587.
25. Li M, Tang D, Zeng J, Zhou T, Zhu H, Chen B, et al. An automated assessment framework for atypical prosody and stereotyped idiosyncratic phrases related to autism spectrum disorder. *Comput Speech Lang.* (2019) 56:80–94. doi: 10.1016/j.csl.2018.11.002
26. Tsuchiya KJ, Hakoshima S, Hara T, Ninomiya M, Saito M, Fujioka T, et al. Diagnosing autism spectrum disorder without expertise: a pilot study of 5- to 17-year-old individuals using Gazefinder. *Front Neurol.* (2021) 11:1963. doi: 10.3389/fneur.2020.603085
27. Vu T., Tran H., Cho K.W., Song C., Lin F., Chen C.W., et al. (2017). Effective and efficient visual stimuli design for quantitative autism screening: an exploratory study. In Proceedings of the 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Orlando, FL, USA, 16–19 February, pp. 297–300. doi: 10.1109/BHI.2017.7897264
28. Lord C, Rutter M, Le Couteur A. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord.* (1994) 24:659–85. doi: 10.1007/BF02172145
29. Goldstein S, Ozonoff S. Assessment of autism Spectrum disorder. New York, NY, USA: Guilford Publications (2018).
30. Kamp-Becker I, Albertowski K, Becker J, Ghahreman M, Langmann A, Mingeback T, et al. Diagnostic accuracy of the ADOS and ADOS-2 in clinical practice. *Eur Child Adolesc Psychiatry.* (2018) 27:1193–207. doi: 10.1007/s00787-018-1143-y
31. He Q, Wang Q, Wu Y, Yi L, Wei K. Automatic classification of children with autism Spectrum disorder by using a computerized visual-orienting task. *PsyCh J.* (2021) 10:550–65. doi: 10.1002/pchj.447
32. Fenske EC, Zalsenski S, Krantz PJ, McClannahan LE. Age at intervention and treatment outcome for autistic children in a comprehensive intervention program. *Anal Interv Dev Disabil.* (1985) 5:49–58. doi: 10.1016/S0270-4684(85)80005-7
33. Frank MC, Vul E, Saxe R. Measuring the development of social attention using free-viewing. *Infancy.* (2012) 17:355–75. doi: 10.1111/j.1532-7078.2011.00086.x
34. Yaneva V, Eraslan S, Yesilada Y, Mitkov R. Detecting high-functioning autism in adults using eye tracking and machine learning. *IEEE Trans Neural Syst Rehabil Eng.* (2020) 28:1254–61. doi: 10.1109/TNSRE.2020.2991675
35. Sasson NJ, Ellison JT. Eye tracking young children with autism. *J Vis Exp.* (2012) 61:3675. doi: 10.3791/3675
36. Bone D, Goodwin MS, Black MP, Lee C-C, Audhkhasi K, Narayanan S. Applying machine learning to facilitate autism diagnostics: pitfalls and promises. *J Autism Dev Disord.* (2015) 45:1121–36. doi: 10.1007/s10803-014-2268-6
37. Poonguzhali R, Ahmad S, Thiruvannamalai Sivasankar P, Anantha Babu S, Joshi P, Prasad Joshi G, et al. Automated brain tumor diagnosis using deep residual U-net segmentation model. *Comput Mater Contin.* (2023) 74:2179–94. doi: 10.32604/cmc.2023.032816
38. Liu W, Li M, Yi L. Identifying children with autism Spectrum disorder based on their face processing abnormality: a machine learning framework. *Autism Res.* (2016) 9:888–98. doi: 10.1002/aur.1615
39. Ahmad S, Khan S, Fahad AlAjmi M, Kumar Dutta A, Minh Dang L, Prasad Joshi G, et al. Deep learning enabled disease diagnosis for secure internet of medical things. *Comput Mater Contin.* (2022) 73:965–79. doi: 10.32604/cmc.2022.025760
40. Peral J, Gil D, Rotbei S, Amador S, Guerrero M, Moradi H. A machine learning and integration based architecture for cognitive disorder detection used for early autism screening. *Electronics.* (2020) 9:516. doi: 10.3390/electronics9030516
41. Crippa A, Salvatore C, Perego P, Forti S, Nobile M, Molteni M, et al. Use of machine learning to identify children with autism and their motor abnormalities. *J Autism Dev Disord.* (2015) 45:2146–56. doi: 10.1007/s10803-015-2379-8
42. Alam M.E., Kaiser M.S., Hossain M.S., Andersson K. An IoT-belief Rule Base smart system to assess autism. In Proceedings of the 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT), Dhaka, Bangladesh, 13–15 September 2018, pp. 672–676.
43. Hosseinzadeh M, Koohpayehzadeh J, Bali AO, Rad FA, Souri A, Mazaherinezhad A, et al. A review on diagnostic autism Spectrum disorder approaches based on the internet of things and machine learning. *J Supercomput.* (2020) 77:2590–608. doi: 10.1007/s11227-020-03357-0
44. Syriopoulou-Delli CK, Gkiolnta E. Review of assistive technology in the training of children with autism spectrum disorders. *Int J Dev Disabil.* (2020) 68:73–85. doi: 10.1080/20473869.2019.1706333
45. Duan H, Min X, Fang Y, Fan L, Yang X, Zhai G. Visual attention analysis and prediction on human faces for children with autism spectrum disorder. *ACM Trans Multimed Comput Commun Appl (TOMM).* (2019) 15:1–23. doi: 10.1145/3337066
46. Tao Y., Shyu M.L. (2019). SP-ASDNet: CNN-LSTM based ASD classification model using observer scanpaths. In Proceedings of the 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Shanghai, China, 8–12 July, pp. 641–646.
47. Akter T, Ali MH, Khan M, Satu M, Uddin M, Alyami SA, et al. Improved transfer-learning-based facial recognition framework to detect autistic children at an early stage. *Brain Sci.* (2021) 11:734. doi: 10.3390/brainsci11060734
48. Raj S, Masood S. Analysis and detection of autism spectrum disorder using machine learning techniques. *Procedia Comput Sci.* (2020) 167:994–1004. doi: 10.1016/j.procs.2020.03.399
49. Xie J., Wang L., Webster P., Yao Y., Sun J., Wang S., et al. (2019). A two-stream end-to-end deep learning network for recognizing atypical visual attention in autism spectrum disorder. arXiv [preprint]. Available at: <https://doi.org/10.48550/arXiv.1911.11393>.
50. Simonyan K., Zisserman A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv [preprint]. Available at: <https://doi.org/10.48550/arXiv.1409.1556>
51. Howard A.G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., et al. (2017). MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv [preprint]. Available at: <https://doi.org/10.48550/arXiv.1704.04861>
52. Huang G., Liu Z., Van Der Maaten L., Weinberger K.Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708.
53. Cilia F, Carette R, Elbattah M, Dequen G, Guérin J-L, Bosche J, et al. Computer-aided screening of autism spectrum disorder: eye-tracking study using data visualization and deep learning. *JMIR Hum Factors.* (2021) 8:e27706. doi: 10.2196/27706
54. Elbattah M, Loughnane C, Guérin J-L, Carette R, Cilia F, Dequen G. Variational autoencoder for image-based augmentation of eye-tracking data. *J Imaging.* (2021) 7:83. doi: 10.3390/jimaging7050083



OPEN ACCESS

EDITED BY

Udhaya Kumar,
Baylor College of Medicine, United States

REVIEWED BY

Manoj Kumar,
University of Wollongong in Dubai,
United Arab Emirates
Suci Aulia,
Telkom University, Indonesia

*CORRESPONDENCE

Sultan Ahmad
✉ s.alisher@psau.edu.sa
Eid Albalawi
✉ elbalawi@kfu.edu.sa

RECEIVED 07 May 2024

ACCEPTED 26 June 2024

PUBLISHED 19 July 2024

CITATION

Albalawi E, Neal Joshua ES, Joys NM, Bhatia Khan S, Shaiba H, Ahmad S and Nazeer J (2024) Hybrid healthcare unit recommendation system using computational techniques with lung cancer segmentation. *Front. Med.* 11:1429291. doi: 10.3389/fmed.2024.1429291

COPYRIGHT

© 2024 Albalawi, Neal Joshua, Joys, Bhatia Khan, Shaiba, Ahmad and Nazeer. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Hybrid healthcare unit recommendation system using computational techniques with lung cancer segmentation

Eid Albalawi^{1*}, Eali Stephen Neal Joshua², N. M. Joys³,
Surbhi Bhatia Khan^{4,5}, Hadil Shaiba⁶, Sultan Ahmad^{7,8*} and
Jabeen Nazeer⁷

¹Department of Computer science, College of Computer Science and Information Technology, King Faisal University, Hofuf, Saudi Arabia, ²Department of Computer Science & Engineering, Gandhi Institute of Technology and Management (GITAM) (Deemed to be University), Visakhapatnam, Andhra Pradesh, India, ³Department of Computer Science & Engineering, Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam, Andhra Pradesh, India, ⁴School of Science, Engineering and Environment, University of Salford, Manchester, United Kingdom, ⁵Department of Electrical and Computer Engineering, Lebanese American University, Byblos, Lebanon, ⁶Department of Computer Science, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia, ⁷Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Alkharij, Saudi Arabia, ⁸School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India

Introduction: Our research addresses the critical need for accurate segmentation in medical healthcare applications, particularly in lung nodule detection using Computed Tomography (CT). Our investigation focuses on determining the particle composition of lung nodules, a vital aspect of diagnosis and treatment planning.

Methods: Our model was trained and evaluated using several deep learning classifiers on the LUNA-16 dataset, achieving superior performance in terms of the Probabilistic Rand Index (PRI), Variation of Information (VOI), Region of Interest (ROI), Dice Coefficient, and Global Consistency Error (GCE).

Results: The evaluation demonstrated a high accuracy of 91.76% for parameter estimation, confirming the effectiveness of the proposed approach.

Discussion: Our investigation focuses on determining the particle composition of lung nodules, a vital aspect of diagnosis and treatment planning. We proposed a novel segmentation model to identify lung disease from CT scans to achieve this. We proposed a learning architecture that combines U-Net with a Two-parameter logistic distribution for accurate image segmentation; this hybrid model is called U-Net++, leveraging Contrast Limited Adaptive Histogram Equalization (CLAHE) on a 5,000 set of CT scan images.

KEYWORDS

image segmentation, two-parameter logistic type distribution, performance evaluation, CLAHE, ROI segmentation, lung cancer detection

1 Introduction

Lung cancer begins in the lungs and spreads throughout the rest of the body (1), including the brain. Lung cancer is the most common cause of cancer-related mortality worldwide (2). Although lung cancer is more frequent in smokers, it may also occur in nonsmokers (3). The incidence of lung cancer is often and excessively increased with smoking. Lung cancer risk may be lowered even if you have smoked for a long period. Segmentation, a type of image compression, is necessary to infer information from photos. Imaging modalities (4), including Magnetic Resonance Imaging (MRI) and Computed Tomography (CT), can be utilized to create Computer-Aided Diagnostic (CAD) (5) models

that can be used to diagnose and treat patients in precision medicine. Using a limited quantity of medical image data, we demonstrated the efficacy of our proposed model, which we refer to as U-NET++. A method known as the dice coefficient loss was used to compute the findings of the investigations. An approach to labeling preprocessing that is in line with the approaches that are already in use is presented in this paper.

The main novelty of this study is as follows.

- To propose the segmentation model for identifying lung disease made on CT scans with the limited set of CT scan images using the CLAHE.
- To develop the learning architecture combining U-Net with a two-parameter logistic distribution for image segmentation, was used for segmentation.
- To train the models using several deep learning classifiers and evaluate the performance of the models using benchmarks on the LUNA16 dataset using different information retrieval metrics.

The following section describes the organization of the subsequent sections of this study.

A considerable amount of important research is presented in Section 2. Deep learning architectures are used in segmenting medical images by U-NET++, which is created by combining the two-parameter model recommended with distribution learning of the U-Net type. Section 3 provides a comprehensive explanation of the topic. At this point, the criteria for evaluating the model's performance discussed in the fourth part of the section are presented.

2 Related works

A meta-analysis of the literature was performed. Table 1 clearly shows the literature matrix representation of their meta-analysis and the strong relationships between the authors and their respective works. CT scans were assessed based on the image brightness. Different areas of the same region should have the same intensity; hence, segmentation is an effective method to separate objects. Various segmentation procedures were found to be useful in this study. Three-step segmentation-based strategy for distinguishing lung regions.

First, the lung was segmented using gray-level thresholding. Dynamic programming then divides the lung lobes. Finally, morphology-based smoothing approaches were employed. Region-based segmentation includes enlarging, dividing, and combining the areas (17).

A novel convolutional network type known as U-NET++ was developed to analyze CT images used in the biological sciences. U-NET++ was used in this study to extract lung fields from CT images. In healthcare, U-NET++ is nothing more than a variation of ConvNet, combined with various *ad hoc* data augmentation methods.

The robustness of the model was compromised because the authors of (6–8) carried out their research using the same data potential. The traditional U-Net network (9–16) is a semantic segmentation network built using a fully convolutional neural

network. Although it has a relatively small number of layers, the network is nevertheless capable of functioning well, although less complex than its predecessors. The UNET network consists of two main components: down-sampling and up-sampling algorithms. The process of feature extraction, also known as down sampling, involves using convolutional, and pooling layers. This stage is accountable for obtaining characteristics from the original image. A deconvolution technique is employed to enhance the feature map's intricacy. The alternative term for the structure that involves down-sampling and up-sampling is the decoder-encoder structure. The original picture undergoes convolutional and pooling layers during the down-sampling process. This leads to the generation of feature maps that include different levels of information. Regarding visual characteristics, the feature maps exhibit diverse abstraction levels. Combining the down-sampled feature map makes it possible to retrieve a larger portion of the abstract detail information lost during training. As a consequence, the network becomes more successful at segmentation. During the up-sampling process, the deconvolution layer systematically increases the feature image's dimensions. Consequently, the lung's three-dimensional nature results in a substantial loss of spatial information. Consequently, a substantial quantity of relevant information is lost when down-sampling occurs. As retrieving all data is impractical, up-sampling yields imprecise outcomes and disregards visual nuances. Moreover, in addition to the aforementioned concerns, implementing a deep neural network is necessary for future advancement. According to the results of applying U-NET++ to a new dataset, the precision of the IOU and Dice coefficients improved. The test results demonstrate that the U-NET++ architecture improves the efficiency of multiscale conversion and fully connected systems. The authors in (18) propose a novel approach for lung CT scan classification. They combined handcrafted features were extracted using Q-deformed entropy (QDE), which captured image texture based on intensity variations, with features automatically learned by a Convolutional Neural Network (CNN). This fusion strategy aimed to improve the identification of healthy lungs from those affected by conditions like COVID-19 or pneumonia (18). This proposed approach demonstrated the benefits of combining handcrafted and automatically learned features. Segmentation focused the model on relevant lung regions, and the LSTM network effectively utilized the fused features for accurate classification.

3 Materials and methods

3.1 U-NET++ architectural design

This study introduces the U-NET++ hybrid model, which utilizes a two-parameter logistic function to identify lung nodules from CT scans accurately. Lung CT scans were classified as “benign” or “malignant” when used as an input for a binary classification system. A unique hybrid model that combines U-Net (19) and two-parameter logistic distribution was developed to segment and diagnose lung cancer. The model was generated using the dataset of LUNA-16 lung CT images. The U-NET++ model is highly esteemed as a leading architecture in computer vision, primarily because it is built on established computer vision

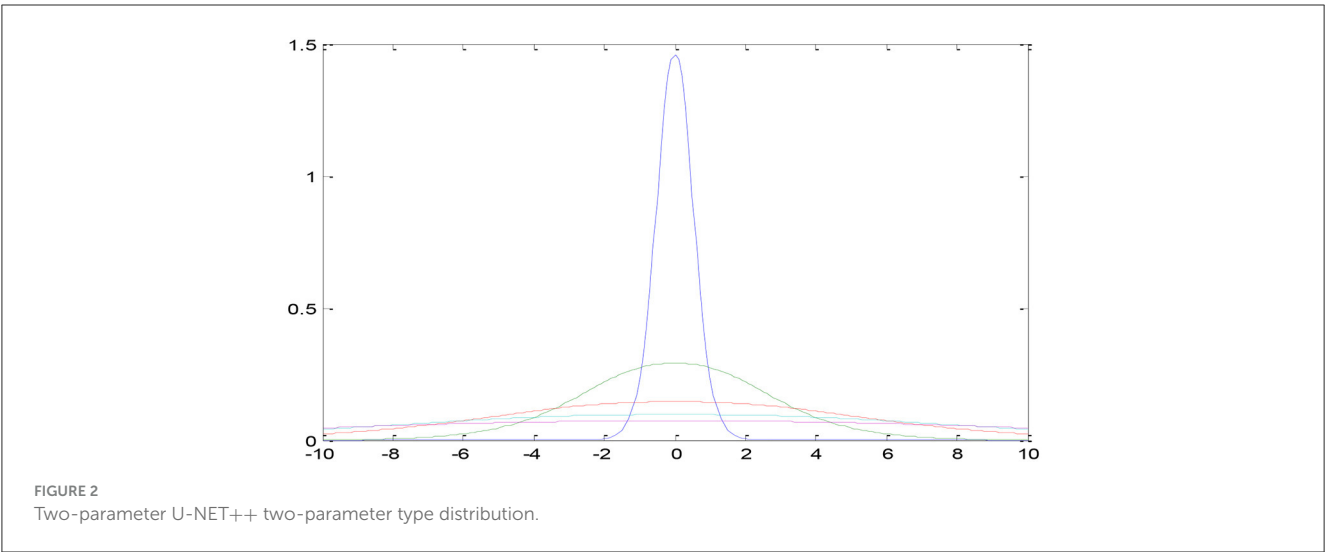
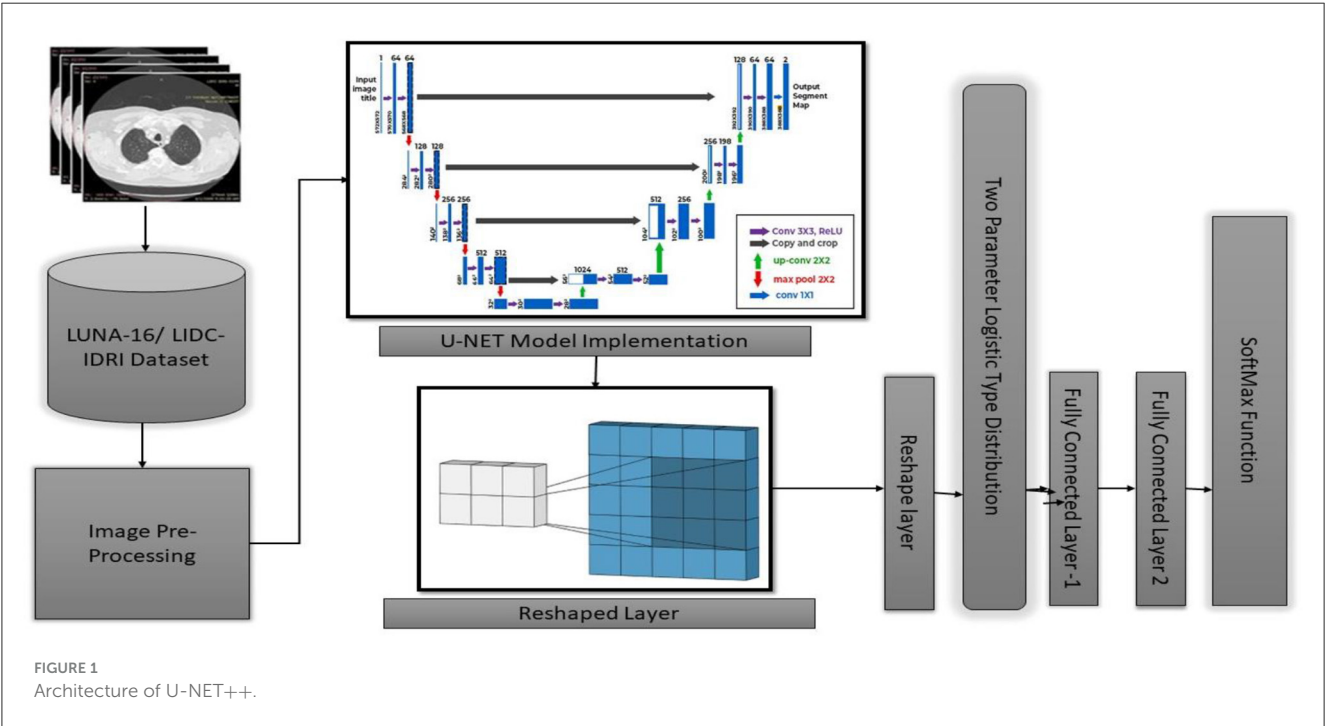
TABLE 1 Presents the related study and limitations in the works.

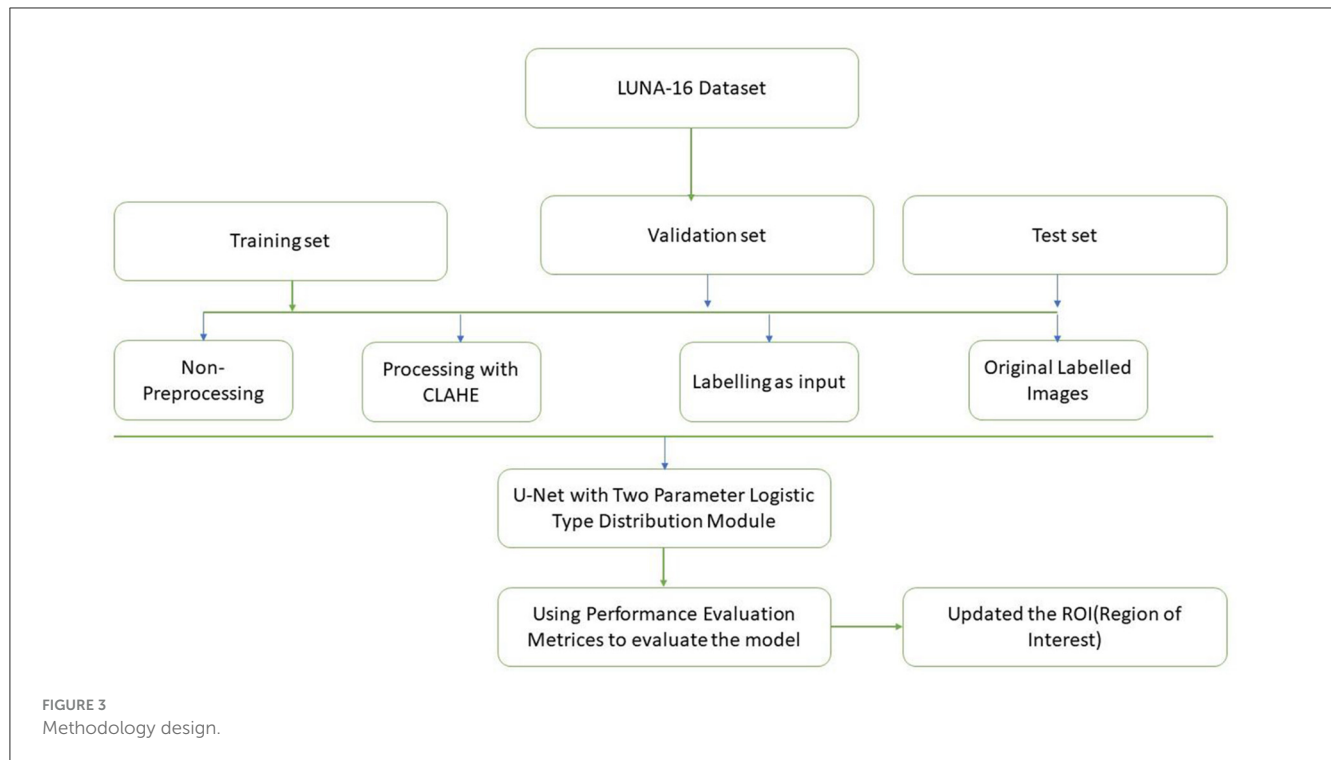
References	Dataset	Split	Key arguments	Drawbacks
Huang and Hu (6)	Lung Nodule Analysis 2016 dataset and Alibaba Tianchi Lung Cancer Detection Competition dataset.	60:40	The Noisy U-Net (NU-Net) increases the diagnosis of early lung cancer nodules by increasing the sensitivity to tiny nodules measuring between 3 and 5 mm in diameter. This is achieved by adding distinct noise to hidden layers during training.	Insufficient validation across a number of clinical situations or datasets has been done to evaluate NU-Net's applicability and robustness. The practical application is restricted since it ignores false positives and the algorithm's inconsistent performance with diverse nodule properties. The lack of advanced method comparison studies limits NU-Net's effectiveness compared to U-Net.
Zhao et al. (7)	LUNA-16	70:30	The proposed approach for accurately detecting cancerous lung lesions from CT scans involves using a patch-based 3D U-Net and a contextual convolutional neural network.	The article lacks a thorough validation or explanation of the model's performance variability across various datasets or in real-world clinical situations. Furthermore, the lack of a comparison to current approaches hinders the ability to assess the superiority or applicability of the proposed strategy.
Chiu et al. (8)	LUNA-16	70:30	The 2D U-Net approach effectively identifies lung nodules in medical pictures. The detection performance may be improved by utilizing ROI segmentation models and further labeling.	The use of the ROI segmentation technique enhances the accuracy of lung nodule identification. The U-Net-based network architecture demonstrates high proficiency in segmenting lung nodules. Additionally, complementary labeling appears to be helpful in situations when there is a scarcity of data.
Gao et al. (9)	LUNA-16	70:30	The U-Net model, which incorporates an attention mechanism and residual structure, effectively segments lung cancer bone metastases in SPECT images, improving early identification and treatment outcomes.	The research will likely neglect practical factors, such as variations in SPECT imaging circumstances or anomalies that may undermine the model's robustness in real clinical settings.
Cai et al. (10)	LUNA-16	60:40	The U-Net deep learning network consistently enables the identification of lung cancer nodules larger than 3 mm in diameter, hence facilitating the progress of early detection and therapy methods for this disease.	The research work fails to describe the AI model's clinical validation and integration in real-world healthcare settings, obscuring its practicality. It prioritizes model accuracy above false positives and negatives, which are essential for successful practical diagnosis. Due to its architecture and lack of testing against more adaptable modern methods, the U-Net and PSP Net AI models' effectiveness is unknown. Due to its dataset dependence, the model may not work for all patient groups or imaging situations (Luna16).
Banu et al. (11)	LUNA-16	70:30	The use of WEU-Net, also known as weight excitation U-Net, enhances the early identification of lung cancer by precisely segmenting lung nodules in CT images.	The work does not explain how the model shows nodule variety, size, and consistency across datasets. The therapeutic adoption of this technology depends on time efficiency and computational needs, which are being disregarded. The lack of a comparison with other cutting-edge segmentation methods hinders our comprehension of WEU-Net's efficacy. To conclude, the model's interpretability and therapeutic potential in diagnostic and treatment planning are undisputed.
Xia (12)	LUNA-16	60:40	When it comes to detecting supplemental lung cancer, RUNet image segmentation outperforms 3D U-Net. Pro-CRP, CEA, and NSE serve as diagnostic markers for malignant lung cancers.	The research lacks a thorough examination of any biases or confounding variables that may impact the accuracy of diagnoses and the performance of the model when selecting patients. The research did not assess the generalizability of the findings to larger groups of patients or other imaging techniques other than MRI. The absence of a comparative analysis with other verified segmentation approaches impedes the understanding of the specific benefits that RUNet offers in contrast to other methods. Moreover, there is insufficient information about using the model in clinical settings to verify its effectiveness in real-world situations or with external datasets.
Chhabra et al. (13)	IIITD-CLF	8:2	The study discusses how regularization and patch size affect how well the model works. segmentation with different network designs and patch sizes to make it more accurate.	Factors like scalability, external validity, possible bias, and limited generalizability should be considered.
Venkatesh et al. (14)	LIDC-IDRI	70:30	It aims to revolutionize the detection of lung cancer by offering a more accurate and efficient approach compared to existing approaches.	The evaluation of the effectiveness of the suggested technique in relation to existing methods is limited due to the lack of a comparative study with state-of-the-art systems. Further investigation is required to enable the idea's implementation in real-world clinical environments, considering ethical concerns, regulatory challenges, and the potential to scale up.

(Continued)

TABLE 1 (Continued)

References	Dataset	Split	Key arguments	Drawbacks
Madhu et al. (15)	POCUS	70:30	This paper presents XCovNet, an improved Xception neural network, which outperforms existing deep learning models for point-of-care lung ultrasound data analysis, enabling accurate identification of COVID-19.	The study enhances medical imaging technology for the detection of infectious diseases by developing XCovNet and showcasing its improved performance in comparison to current models. This is essential to fulfill the need for accurate and expedient diagnostic tools in contexts with limited resources.
Lamba et al. (16)	GSCE25066	70:30	The aim of the project is to use machine learning techniques to find crucial genes for cancer subtyping. These genes will then be validated using the Kaplan-Meier Survival Model.	The study paper does not explicitly discuss any recognized research constraints in the categorization of breast cancer subtypes based on gene expression data. Subsequent studies in this domain might examine the impact of different feature selection methods on the effectiveness of models and the reliability of findings across different datasets.





approaches. When assessed using the ImageNet test dataset, this model achieved a precision rate of 91%. The main architectural improvement in the model is the filter size, an improved version of the U-NET. Figure 1 illustrates the architecture of the proposed model.

In the current section, a detailed presentation of the combination of two- and three-parameter logistic distribution models is presented. Figure 2 shows a two-parameter U-NET++ logistic-type distribution. In general, the pixel intensities are the content through which the quantification of the image details performed on several regions of the images. The brightness of a picture or image can be measured by using several performance metrics such as the moisture in the surroundings, lightening of the images, vision, and the surrounding environmental conditions. This measurement can be performed using the pixel values and pixel intensities. For instance, pixel (a, b) intensity measurement was performed using the function $z = f(a, b)$ and considered as a random variable. To better analyze and understand the performance of the currently considered model and the intensities of pixels for various images, the model was designed for both parametric and parametric models. The pdf of the pixel intensity is given by

$$f(y, u, \Omega^2) = \frac{\left[\frac{3}{(12+\pi^2)} \right] \left[4 + \left(\frac{y-u}{\Omega} \right)^2 \right] e^{-\left(\frac{y-u}{\Omega} \right)^2}}{k \left[1 + e^{-\left(\frac{y-u}{\Omega} \right)^2} \right]}, \quad -\infty < y < \infty, -\infty < u < \infty, \Omega > 0 \quad (1)$$

Where y is the pixel Intensity, u is the mean of pixels s , and Ω is the variance of the entered image's pixels.

3.2 U-NET++ algorithm

3.2.1 U-Net with two parameter type distribution Ω_i^2

For updating Ω_i^2 we differentiate $R(Q, Q)$ with respect to Ω_i^2 and equate it to zero. That is $\frac{\partial}{\partial \Omega_i^2} (R(Q, Q)) = 0$. This implies $E \left[\frac{\partial}{\partial \Omega} (\log L(Q, Q^{(l)})) \right] = 0$. The derivative was applied and implemented for both parameter models with σ_i^2 for the two-parameter model, with estimation error of 0.001 and it was with the biased estimation. From the Equations 1–6 segmentation algorithm used in the proposed algorithm.

$$\left[\sum_{s=1}^N \sum_{i=1}^K P_i(y_s, Q^l) \log \beta_i \frac{\left[\frac{3}{12+\pi^2} \right] \left[4 + \left(\frac{y_s - U_i}{\Omega_i} \right)^2 \right] e^{-\left(\frac{y_s - U_i}{\Omega_i} \right)^2}}{\Omega_i \left[1 + e^{-\left(\frac{y_s - U_i}{\Omega_i} \right)^2} \right]} \right] = 0. \quad (2)$$

The updated equations of σ_i^2 at $(l+1)^{th}$ iteration is

$$\Omega_i^{2(l+1)} = \frac{\sum_{s=1}^N \left[\left[\frac{(y_s - U_i^{(l+1)})}{\Omega_i^{3(l)}} \right] - \left[\frac{1}{\Omega_i^{2(l)}} \right] - \left[\frac{(y_s - U_i^{(l+1)})^2}{\Omega_i^4 \left(1 + e^{-\left(\frac{y_s - U_i^{(l+1)}}{\Omega_i^{(l)}} \right)^2} \right)} \right] \right] p_i(y_s, Q^{(l)})}{\sum_{s=1}^N \frac{(x_s - \mu_i^{(l+1)}) p_i(x_s, \theta^{(l)})}{\sigma_i^4(l) (4\sigma_i^2(l) + (x_s - \mu_i^{(l+1)})^2)}}. \quad (3)$$

For three-parameter logistic type distribution: -

$$\left[\frac{\frac{\partial}{\partial \Omega_i^2}}{\sum_{s=1}^N \sum_{i=1}^K P_i(y_s, Q^l) \log \beta_i \left[\frac{\frac{3}{3p+\pi^2}}{\Omega_i} \left[p + \left(\frac{y_s - U_i}{\Omega_i} \right)^2 \right] e^{-\left(\frac{y_s - U_i}{\Omega_i} \right)^2} \right]} \right] = 0. \tag{4}$$

$$\Upsilon_i^{(l+1)} = \frac{\sum_{s=1}^N \frac{P_i(y_s, Q^{(l)}) (y_s - U_i^{(l+1)})}{2\Omega_i^3(l)} - \sum_{s=1}^N \frac{P_i(y_s, Q^{(l)}) (y_s - U_i^{(l+1)})}{\sigma_i^2(l) \left[\frac{(y_s - U_i^{(l+1)})}{1+e} \right]} - \sum_{s=1}^N \frac{P_i(y_s, Q^{(l)})}{2\Omega_i^2(l)}}{\sum_{s=1}^N \frac{P_i(y_s, Q^{(l)}) (y_s - U_i^{(l+1)})^2}{\sigma_i^4(l) \left[p\Omega_i^2(l) + (y_s - U_i^{(l+1)})^2 \right]}}. \tag{5}$$

Were

$$p_i(y_s, Q^{(l)}) = \left[\frac{\beta_i^{(l+1)} f_i(y_s, \Upsilon_i^{(l+1)}, \Omega_i^{2(l)})}{\sum_{i=1}^k \beta_i^{(l+1)} f_i(y_s, \Upsilon_i^{(l+1)}, \Omega_i^{(l)})} \right]. \tag{6}$$

3.3 Module design

Figure 3 discuss about the methodology design followed in our proposed work. A typical image processing method is contrast-limited adaptive histogram (CLAHE) equalization. Smooth regions become noisier with adaptive histogram equalization. CLAHE may enhance noise in hectic circumstances. Histogram size may be limited by CLAHE. Understand that deep learning variation is a major issue. Use two tag techniques for variety. Match the center to the background to reduce variation. This study employed the dice coefficient loss function used by picture segmentation pros. The experiment suggests labeling may be better than initial marking in cases with insufficient data. Medical images are hard to classify and find. Everyone agrees transferring less data is hard. Semi-supervised learning overcomes auto-labeling naming issues. Proposed study successfully locates the lung using ROI segmentation from CT scans. Process attention model. The ROI segmentation model during data processing may find lung tumors, study suggests.

4 Model parameters and discussions

4.1 LUNA-16 dataset

A total of 5,000 CT scans were obtained from LUNA-16. Four expert radiologists annotated the images in the LIDC/IDRI database for 2 years (20–22). Each radiologist diagnosed the nodules as non-nodules, nodules with a diameter of ≤ 3 mm, or nodules with a diameter of ≥ 3 mm (23). This article examines the annotation process in detail. Three of every four nodules larger than 3 mm in diameter must be identified by radiologists (24). Non-standard findings have not been noted before (non-nodules, nodules <3 mm, and nodules annotated by only one or two radiologists). Table 2 shows various illustrations of nodules in the LUNA-16 dataset.

TABLE 2 Various benign and malignant nodules present in the LUNA-16 dataset.

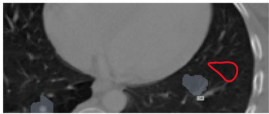
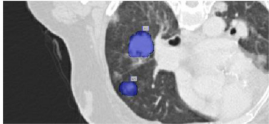
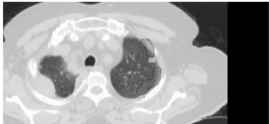
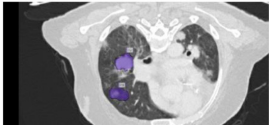
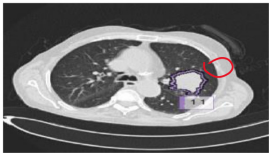
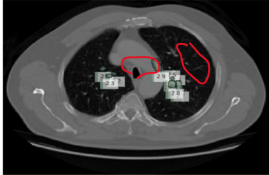
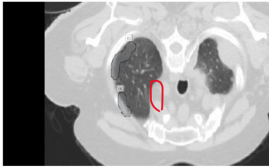
S. No.	Nodule name	Nodule image
1	Small nodule	
2	Ground glass opacity nodule	
3	Rough edged nodule	
4	Thick walled nodule	
5	Granular nodule	
6	Pleural surface nodule	
7	Pulmanory region nodule	

TABLE 3 Presents the standard deviation of various features in LUNA-16/LIDC-IDRI dataset.

Features in LUNA-16	Testing	Training
Malevolence	1.98 ± 0.95	1.65 ± 1.03
Conjecture	2.60 ± 0.70	2.65 ± 0.77
Subtlety	1.89 ± 0.74	3.65 ± 0.69
Lobulation	2.73 ± 0.67	2.36 ± 0.71
Diameter in mm	9.17 ± 3.51	8.56 ± 0.56
Margin	3.03 ± 1.56	3.68 ± 0.58

Table 3 presents various feature extraction values obtained from the LUNA16 database. A node, which refers to a specific structure, has a wide range of characteristics, with malignancy being used as an example to illustrate this. The estimation of the node's outline coordinates is utilized, whereas the surrounding area of the nodule

TABLE 4 Dataset details.

Dataset name	Description
LUNA-16	Comprising annotated lung CT scans collected from partnering medical institutions, Includes data from LUNA 16 and LIDC-IDRI
Number of samples	5,000 annotated CT scans slices
Image resolution	128 × 128
Annotation methods	Expert radiologists using semi-automated tools
Preprocessing steps	-Slice normalization -Rescaling to uniform dimensions -Augmentation for training set

is often underestimated. Lobulation refers to the configuration and attributes of a nodule. The measurement of a nodule in millimeters determines its diameter, which in turn determines its length. The border of the nodule indicates a transparent region.

Table 4 describes the dataset used in our study. We compiled a custom LUNA-16 dataset by combining annotated lung CT scans from various sources, including LIDC-IDRI datasets. This dataset comprises 5,000 annotated CT scans slices, each with a resolution of 512×512 pixels. The images were annotated by expert radiologists using semi-automated tools, ensuring high quality labels for training and evaluation.

4.2 Study design

Three categories of data were created, namely training, validation, and testing. We built a model, trained it using validation data, and tested it. This method is repeated until a firm understands how our model reacts in real-world scenarios. Allow average pooling and expand the size of the final output by using layers in the filter. We examined our test data to determine what we could learn from it in order to enhance the model. Because we are neither testing nor training a model on a test dataset, we can utilize it only once per session. Two-parameter and three-parameter mixtures generate a model using a single test dataset, which significantly reduces the time and effort required. Figure 4 illustrates the study design.

4.3 Split and pre-process data

Jpeg serves as the data transport format in our architecture in the same way as DICOM. The Neuroimaging Informatics Technology Initiative (NIFIT) (25) is a 501(c)(3) not-for-profit organization committed to the advancement of neuroimaging informatics (NITI) (26). Despite its origins in neuroimaging, it is now commonly used in brain and other medical imaging. By memorizing the coordinates, it is possible to relate pixel values (i, j, k) to the position space (x, y, z) (x, y, z). Each data scan may provide three-dimensional medical images comprising 128×128 slices of varying thicknesses. Additional RAM is required to store the data in the DICOM format.

CLAHE12 contribute to the enhancement of CT scan quality (Contrast Limited Adaptive Histogram Equalization). The artwork places a premium on contrast and visual detail. The Hounsfield center values for the lung window and soft tissue were 600, 1,500, and 50,400. As a result, the lung window is the most frequently used Hounsfield range for lung image diagnosis. As shown in Table 1, the Hounsfield values of various body components were dispersed. Following sampling, the objective was to compress a snapshot to preserve the memory. Standardization is the next step in reducing computing costs. Subsequently, CLAHE was used to enhance nodule contrast and visibility.

Contrast-limited adaptive histogram equalization (CLAHE) has been used in image processing for a long time. Instead of adaptive Histogram Equalization (AHE13) (27), it cannot be used. Standard adaptive equalization may amplify noise in ordinarily homogeneous areas of the image. Consequently, the histogram tends to focus on this region. The CLAHE has the potential to enhance noise in locations where it is almost continuous. In Figure 5, the LUNA-16 dataset is preprocessed using the Wiener filter and CLAHE.

The CLAHE approach can be used to decrease the histogram concentration. When utilizing CLAHE, the concentrated histogram component was maintained. On the other hand, the exceeding histogram was maintained and equally distributed throughout all histogram bins. The Wiener filter is an extremely successful technique for visual noise reduction. PET/CT scans were afflicted with an additive noise of constant intensity. Figure 6 shows an example of the original CT scan image, second image is with CLAHE and third one is with CLAHE and weiner.

4.4 Architecture and implementation

The lung segmentation method utilized in this study used 5,000 lung CT scan images and masks. Each CT scan image has a resolution of 128×128 pixels. Images s black and white the final consequence is a split lung. The technique begins with the data being saved in memory and each image being resized to 32×32 pixels. Image processing was accelerated by shrinking the photographs. The images were corrected after rescaling. Subsequently, the dataset was partitioned into 70 percent training set and 30 percent test set. Rotation was performed to increase the number of training samples. There were eight rotating copies for each training sample. In Table 5, U-NET++ is composed of layer blocks that compress and stretch clockwise. The augmented dataset was initially used to define the input layer. The following are the layers of convolution, non-linearity, and down sampling. Non-linearity is first applied to decrease the final image size, followed by convolution to apply a filter, and finally max-pooling. The image is concatenated by applying similar layers in contracting and expanding patterns, and then up-sampled to make it larger. The output layer provides a lung segmentation image. After all layers have been trained, the U-Net ConvNet is created (28). For example, using Adam as the optimizer, the dropout was set to 0.5, epochs were set to 10, and steps per epoch were set to 200 (29). Each layer, similar to the model architecture, has its own set of filters. We examined the performance of U-Net ConvNet using test data.

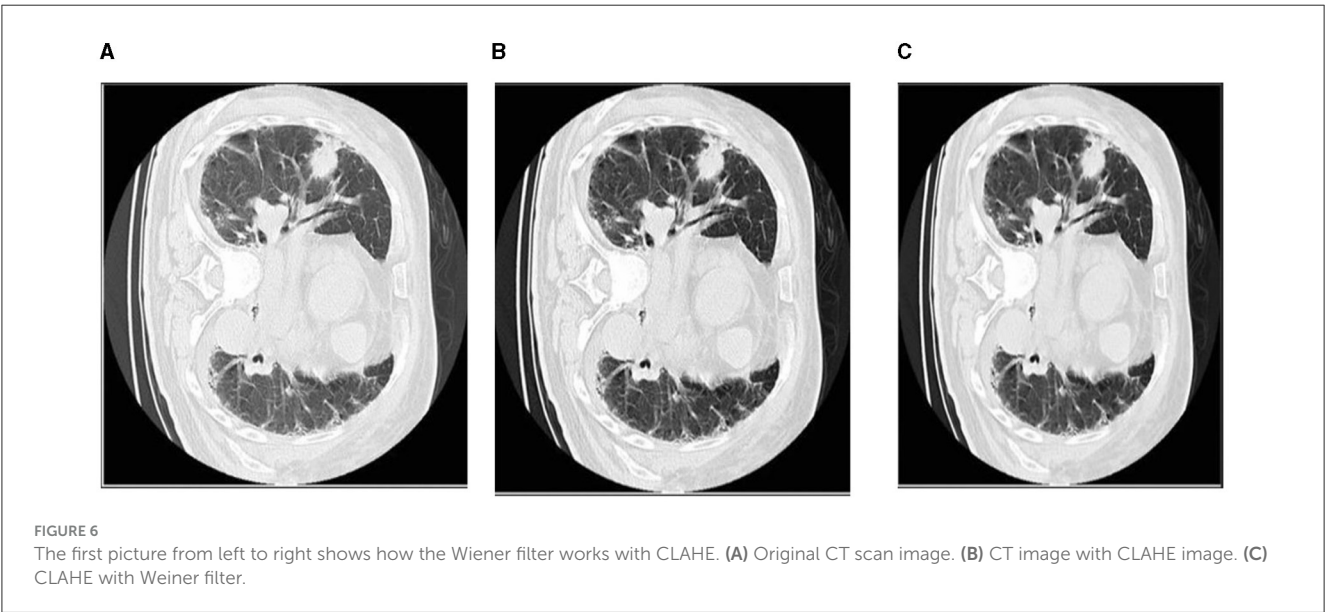
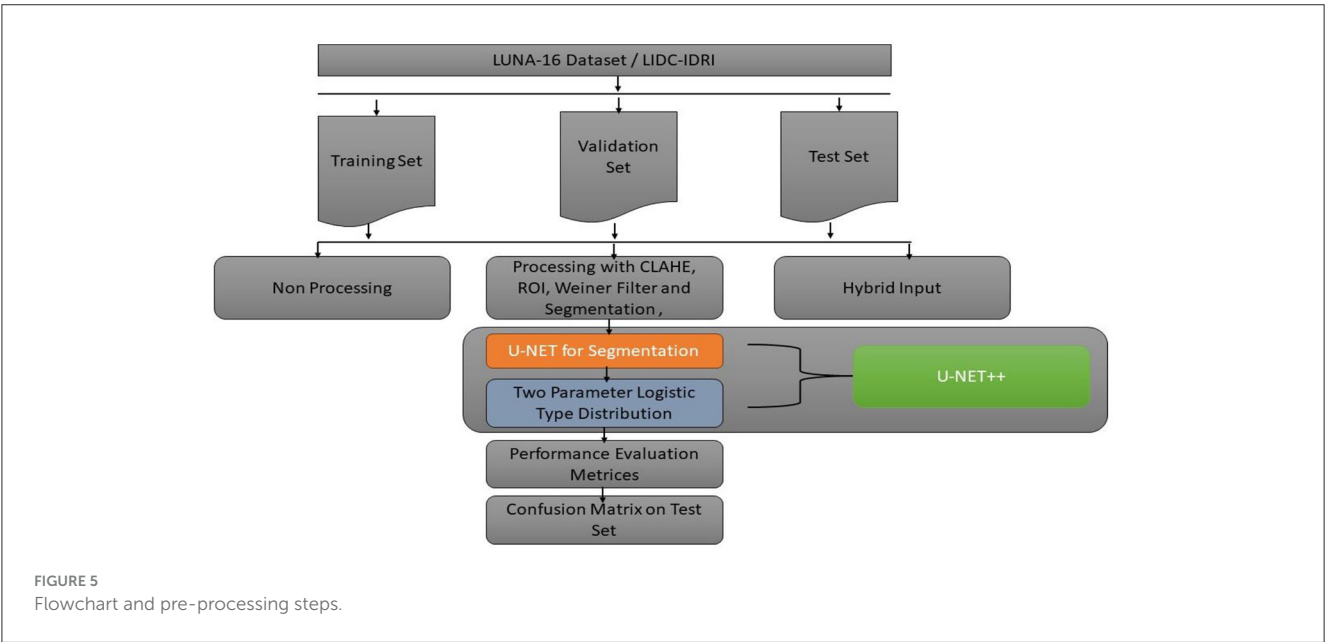
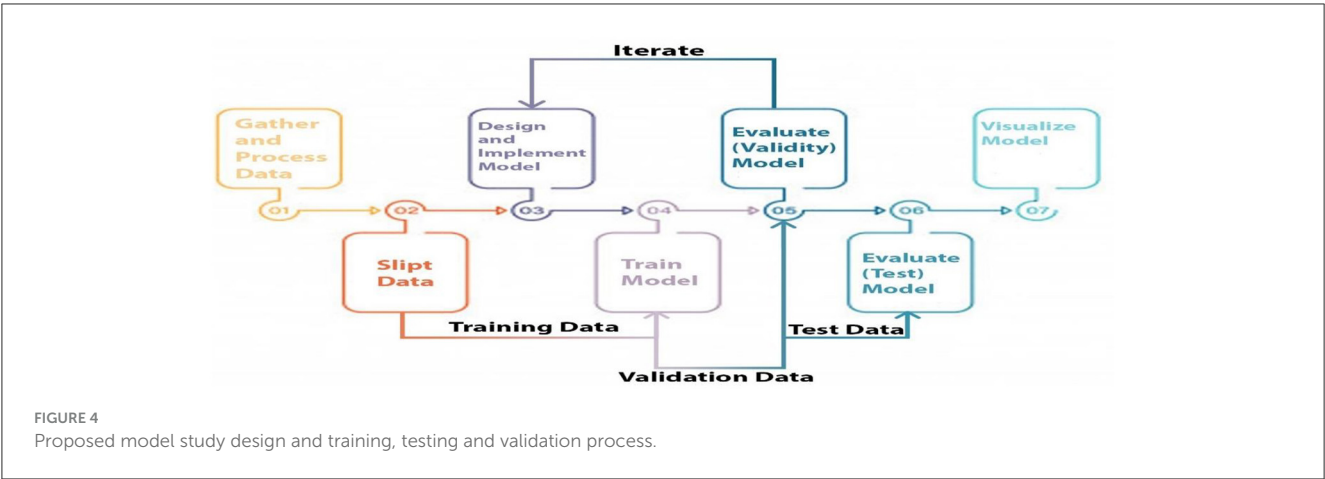


TABLE 5 Proposed network architecture with two parameters distribution.

Layer	Type	Input size	Output size	Kernal size	Stride	Padding
U-NET++ down sampling encoder process						
Layer 1	Conv+ReLU	128 × 128	128 × 128	3 × 3	1	1
Layer 2	Conv+ReLU	128 × 128	128 × 128	3 × 3	1	1
Layer 3	Max Pooling	128 × 128	64 × 64	2 × 2	2	0
Layer 4	Conv+ReLU	128 × 128	64 × 64	3 × 3	1	1
Layer 5	Conv+ReLU	64 × 64	64 × 64	3 × 3	1	1
Layer 6	Max Pooling	64 × 64	32 × 32	2 × 2	2	0
Layer 7	Conv+ReLU	32 × 32	32 × 32	3 × 3	1	1
Layer 8	Conv+ReLU	32 × 32	32 × 32	3 × 3	1	1
Layer 9	Max Pooling	32 × 32	16 × 16	2 × 2	2	0
Layer 10	Conv+ReLU	16 × 16	16 × 16	3 × 3	1	1
Layer 11	Conv+ReLU	16 × 16	16 × 16	3 × 3	1	1
Layer 12	Max Pooling	16 × 16	8 × 8	2 × 2	2	0
Layer 13	Conv+ReLU	8 × 8	8 × 8	3 × 3	1	1
Layer 14	Conv+ReLU	8 × 8	8 × 8	3 × 3	1	1
Layer 15	Max Pooling	8 × 8	4 × 4	2 × 2	2	0
U-Net++ up-sampling decoder process						
Layer 16	Up sample Transposed Conv	4 × 4	8 × 8	2 × 2	2	0
Layer 17	Conv+ReLU	8 × 8	8 × 8	3 × 3	1	1
Layer 18	Conv+ReLU	8 × 8	8 × 8	3 × 3	1	1
Layer 19	Up sample Transposed Conv	8 × 8	16 × 16	2 × 2	2	0
Layer 20	Conv+ReLU	16 × 16	16 × 16	3 × 3	1	1
Layer 21	Conv+ReLU	16 × 16	16 × 16	3 × 3	1	1
Layer 22	Up sample Transposed Conv	16 × 16	32 × 32	2 × 2	2	0
Layer 23	Conv+ReLU	32 × 32	32 × 32	3 × 3	1	1
Layer 24	Conv+ReLU	32 × 32	32 × 32	3 × 3	1	1
SoftMax function	Convolutional Layer_8	32 × 32	32 × 32	3 × 3	1	1
Benign or malignant	SoftMax Function	32 × 32	32 × 32	3 × 3	1	0

There were five columns in total. The first column provides the layer name, followed by the number of filters, filter type/size, dimension, and concatenated layers. Eleven convolutional layers were used. The input layer is the first layer. A 32×32 -pixel input layer is displayed in this picture. For the Con1 layer, eight 3×3 filters are needed. The size of the images remained unchanged. Con1 was closely related to other con1. After the con layers, there were ReLU layers.

4.4.1 Simulation settings

To facilitate the replication of our work, we provide a detailed description of the simulation settings and the dataset used. This

information includes hardware and software configurations, data preprocessing steps, and hyperparameter settings.

The simulation settings outlined in Table 6 provides comprehensive details on the hardware software environment used for our requirements. Our setup included an Intel core i9-10900k CPPU and an NVIDIA GEFORCE RTX 3090 GPU, ensuring sufficient computational power for training deep learning models. We utilized Ubuntu 20.04 LTS as our operating system, with python 3.8 and TensorFlow 2.4 for model development and training.

Table 7 details the hyperparameters and model configuration. We implemented a U-NET++ with 20 layers, utilizing a kernal size of 3×3 and max pooling layer of 2×2 . The ReLU activation function was used throughout the network, with a sigmoid

TABLE 6 Simulation setting used in our proposed work.

Component	Description
Hardware	
CPU	Intel Core i9-10900K
GPU	NVIDIA GeForce RTX 3090
RAM	64 GB DDR4
Software	
Operating System (OS)	Ubuntu 20.04 LTS
Programming language	Python 3.8
Deep learning framework	TensorFlow 2.4, Keras
Data preprocessing	
Normalization	Rescale pixel values to range [0,1]
Augmentation techniques	Rotation, translation, flipping, scaling
Data split	70% training, 15% validation, 15% testing
Model training	
Optimizer	Adam
Learning	0.001
Batch size	32
Epochs	100
Loss function	Dice loss
Metrics	Dice co-efficient, IoU, Sensitivity, Specificity

TABLE 7 Hyperparameters and model configurations.

Parameter	Values
Network architecture	U-NET++
Number of layers	20
Kernal size	3 × 3
Pooling	Max pooling (2 × 2)
Activation function	ReLU (Rectified Linear Unit)
Output layer activation	Sigmoid
Dropout rate	0.5
Regularization	L2 regularization with delta = 0.001

activation function in the output layers for binary segmentation (30). A dropout rate of 0.5 and L2 regularization were applied to prevent overfitting.

These settings and configurations provide a robust framework for replicating our lung cancer segmentation model and can serve as a foundation for further research and development in this domain.

4.5 Training process

The loss function expresses the loss of the die coefficients. Frequently, the dice coefficient is used to segment medical images,

as shown in Figure 7. It is often used to compare two samples. This experiment generated sufficient compelling evidence to be deemed to be conclusive.

This research is mostly concerned with two-dimensional pictures. It might end up saving a lot of money in the long term. Another example is graphics processing unit (GPU) throttling. Owing to memory limitations, the majority of GPUs have difficulty in training 3D models. 2D and 3D models are available for downloading in various formats. We break down our findings into different segmentation strategies with an emphasis on unbalanced and tiny datasets. In addition, the model training process converged in 200 epochs. The confusion matrix can be used to evaluate real-world data and calculate metrics such as accuracy, sensitivity, and specificity. The testing loss is approximately 0.4 in Figure 6, whereas CLAHE and Wiener may be as low as 0.1 without pre-processing.

5 Results discussion and comparison with other models

The results were enhanced by using the ROI segmentation method. It seems that it has the capacity to address the problem of the model’s inaccurate positioning of labels. As a consequence, following the recommended methodology may lead to decreased losses. Furthermore, it was shown that the training session continued to slow down. The lesson is enhanced in its effectiveness as shown in Figure 8. It is advisable to apply the same treatment to both one-dimensional and two-dimensional data. The objective of this strategy is to eliminate any errors in labeling in both directions. Over time, there was a gradual reduction in the size of each point. Engaging in conversations with individuals helps achieve both objectives.

If the dataset is insufficient, it may be necessary to round up more labels. Overall, there were 159 cancerous tumors, and the standard deviation of the Dice coefficient was 0.2. Although its model had a low mFPI, the DL-based model was successful in detecting lung tumors from chest X-rays, the results are shown in Figure 8. The evaluations of the proposed models are presented in Table 8.

TensorFlow was used to evaluate the effectiveness of the U-NET++ approach for the segmentation of lung tumors. The evaluation was performed with the assistance of an image segmentation examiner. Images from LUNA-16 were used to complete the segmentation process. The results of the logistic distributions with the two parameters are shown in the following table. Based on the information shown in Table 9, it is presumed that the intensities of the image pixels adhere to a combination of logistic-type distributions with two parameters.

The pixel intensities in each of the k sectors of the image were assumed to follow a two-parameter logistic distribution, with unique parameters. This assumption was based on the fact that a picture. The histogram of pixel intensities was analyzed to estimate the segment count for each CT scan image used in the experiment. The histograms that indicate the pixel intensities that may be observed in the CT scan images are shown in Figure 9.

Typically, malignant tumors have higher average radius values compared to benign tumors, as seen by histograms and bar graphs. The average radius of malignant

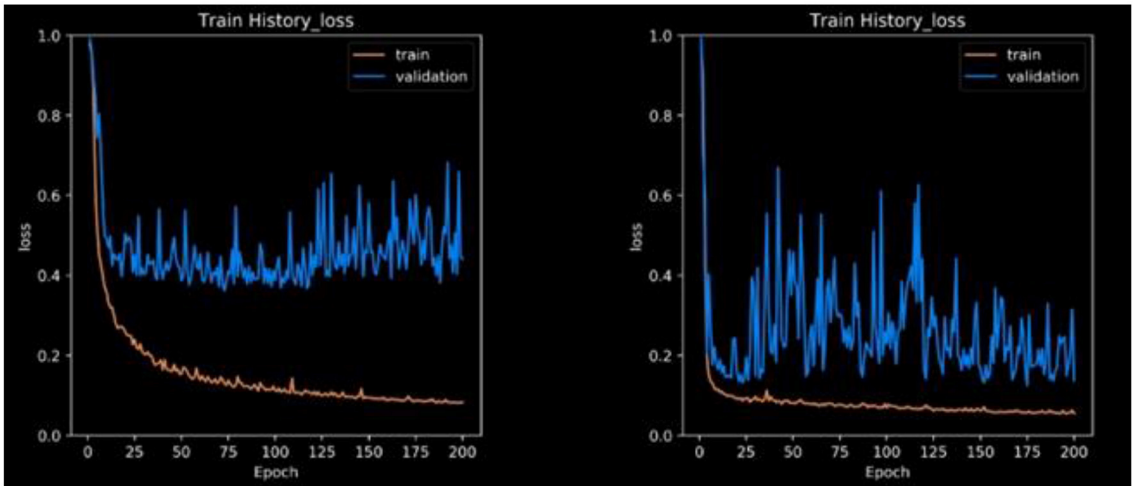


FIGURE 7
The proposed framework with respect to both training and validation accuracy.

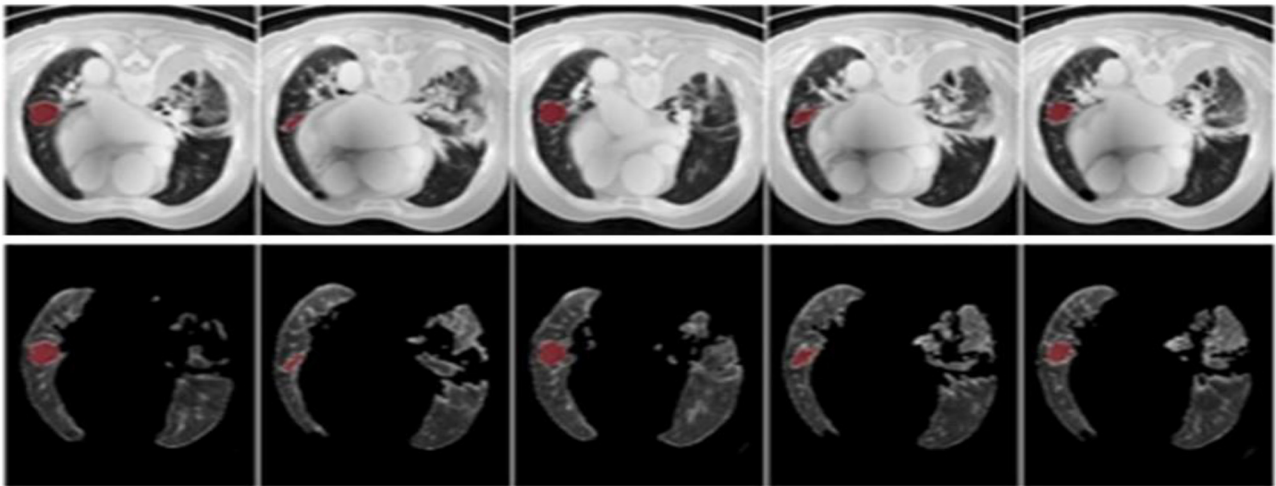


FIGURE 8
Prior to and during the segmentation procedure, the ground-truth forecast was used in each of these instances.

TABLE 8 The evaluation report of the different lung nodule semantic segmentation with comparison to our proposed algorithm.

Evaluation	Cai et al. (10)	Banu et al. (11)	Proposed model
Dice similarity index	87.22%	90.24%	91.76%
Error matrix	Accuracy 90%	Accuracy 89%	Accuracy: 90%
	Sensitivity 90%	Sensitivity 90%	Sensitivity: 89%
	Specificity 89%	Specificity 86%	Specificity: 90%

tumors is 20.1020, whereas benign tumors normally have a radius of 11.3286. These data indicate the differences in average radius values between benign and malignant tumor types.

5.1 Visualization of the model

After examining the data, they found a connection, as shown in Table 10, between how well the suggested method worked and

TABLE 9 The refined value of k with two-parameter U-NET architecture.

Constraints	First parameters		Revised calculations									
	Image region		Image area									
	1	2	1	2	1	2	1	2	1	2	1	2
α_i	0.500	0.500	0.2588	0.7428	0.500	0.500	0.2588	0.7428	0.500	0.500	0.2588	0.7428
μ_i	60.54	121.98	19.48	136.18	60.54	121.98	19.48	136.18	60.54	121.98	19.48	136.18
σ_i^2	94.2568	128.784	124.281	117.251	94.2568	128.784	124.281	117.251	94.2568	128.784	124.281	117.251

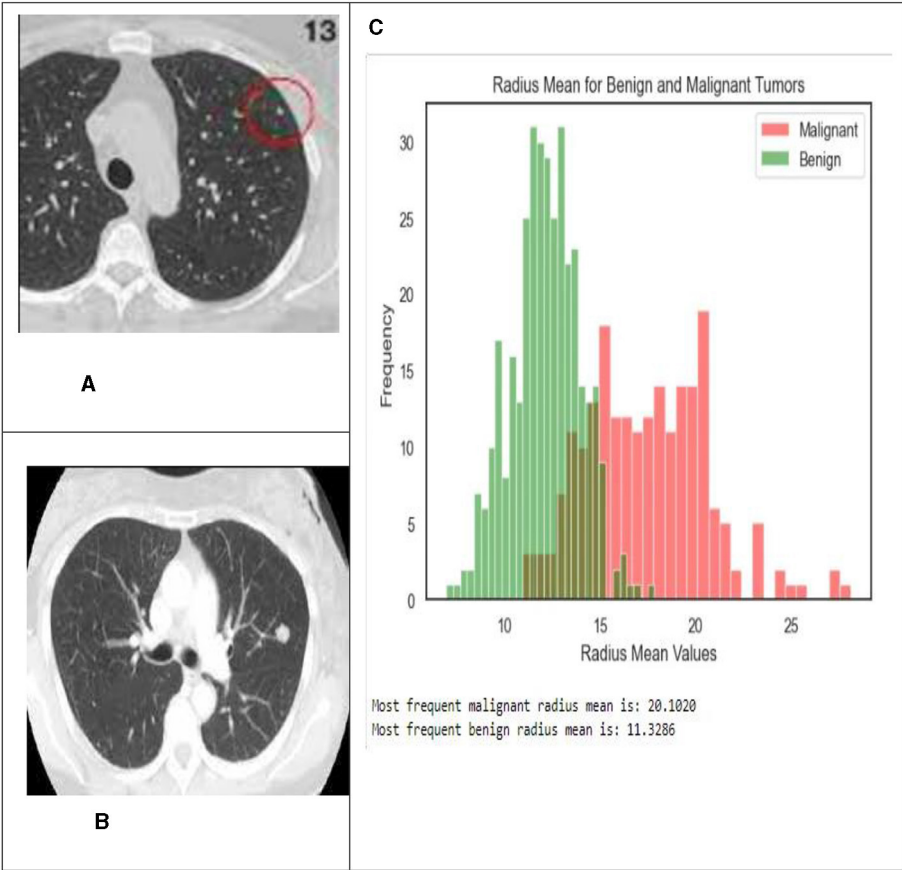


FIGURE 9 In this illustration, the pixel intensities generated from CT scan images of lung nodules that were either benign or malignant were included. (A) Malignant tumor. (B) Benign tumor. (C) Shows the radius mean for benign and malignant tumors.

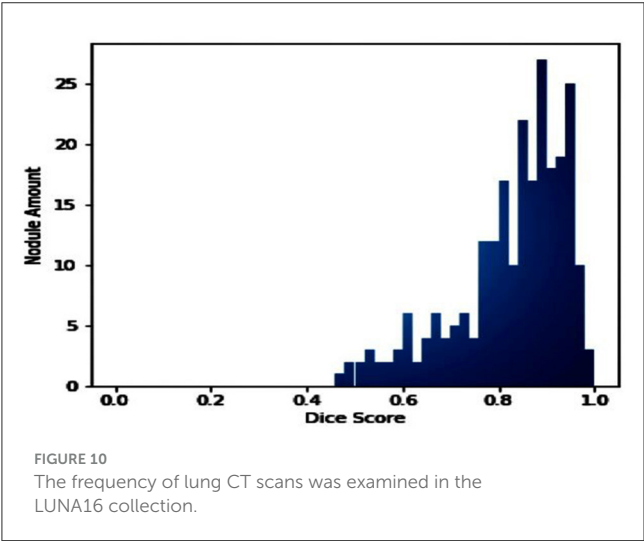
TABLE 10 Comparing the proposed model’s quantitative segmentation results to well-established benchmark models.

References	Classifier models	Dice coefficient (%)	Sensitivity (%)	Specificity (%)
Huang and Hu (6)	NU-NET	89.26 ± 12.45	89.63 ± 23.56	89.21 ± 14.25
Zhao et al. (7)	U-NET	76.24 ± 17.89	85.45 ± 12.54	88.24 ± 15.45
Chiu et al. (8)	2D U-NET	81.89 ± 14.56	91.25 ± 12.89	78.26 ± 15.45
Gao et al. (9)	U-NET	86.45 ± 56.78	78.56 ± 23.57	87.65 ± 23.90
Cai et al. (10)	U-NET	87.22 ± 56.45	75.67 ± 23.74	56.24 ± 22.56
Banu et al. (11)	3D U-NET	90.24 ± 24.45	80.26 ± 23.77	79.23 ± 22.74
Xia (12)	WU-NET	83.12 ± 25.06	88.96 ± 26.32	80.24 ± 23.56
Proposed work	U-NET++	91.76 ± 26.67	89.54 ± 3.65	85.98 ± 25.98

Bold values indicate highest compared with other classifiers.

other ways of showing the same thing. To determine how well the U-NET++ model segmented the LUNA16 trial dataset, five radiotherapists were used for comparison with real experts. Of the three radiologists, 81.26% were good at segmenting patients. The U-NET++ model was also tested by comparing it with the U-NET model and many other benchmark models, such as the newest ResNet152V2.

The number of nodes, Dice coefficient value index, and distribution are presented in Figure 10. This allowed the U-NET++ model to be tested on a test set. Giving each node a number and placing it in the midst of a test set trial is standard.



Duan et al. (23) employed a U-NET architecture with advanced deep learning techniques, resulting in a dice co-efficient of 0.88. Similarly, Duan et al. (23) utilized V-NET incorporating 3D convolutional layers, achieving a dice co-efficient of 0.90. The method by Petit et al. (25) leveraged transformer networks, while Ali et al. (26) utilized efficient net for a more parameter efficient approach.

Table 11 shows a numeric comparison of how well the new method U-NET++ works with three other deep learning models, U-Net (7), NU-Net (6), and WU-Net (12), using CT images of lung nodules from a dataset that was already made public, the suggested method is better than the average method for segmenting images of lung nodules.

We used Fisher’s least significant difference (LSD) method in SPSS software to look at the numeric results and see if the suggested way in Table 12 worked. By using the LSD test, we can see that the suggested method does better than standard methods in terms of IoU, recall, precision, and F1-score ($p < 0.001$).

After preprocessing the image, shown in the Figure 11A the grouped picture, Figure 11B what was found when Lung tumors were identified. Figure 11C results of cutting lung tumors into whole pieces. Figure 11D the findings of the lung tumor search. Figure 11E picture showing the effects on a specific area of lung tumors when they are cut into pieces. Figure 11F a picture of a lung tumor that was accurately cut into pieces.

Our Model, built on a U-NET++ architecture, demonstrated a baseline performance with a dice-coefficient of 91.76% and an IoU of 89.78%. Recent methods, such as the swin Transformer by Ronneberger et al. (27), achieved higher performance metrics through the use of advanced architectures and techniques.

TABLE 11 Quantitative evaluation of lung cancer segmentation methods based on key performance metrics, model architectures and unique features.

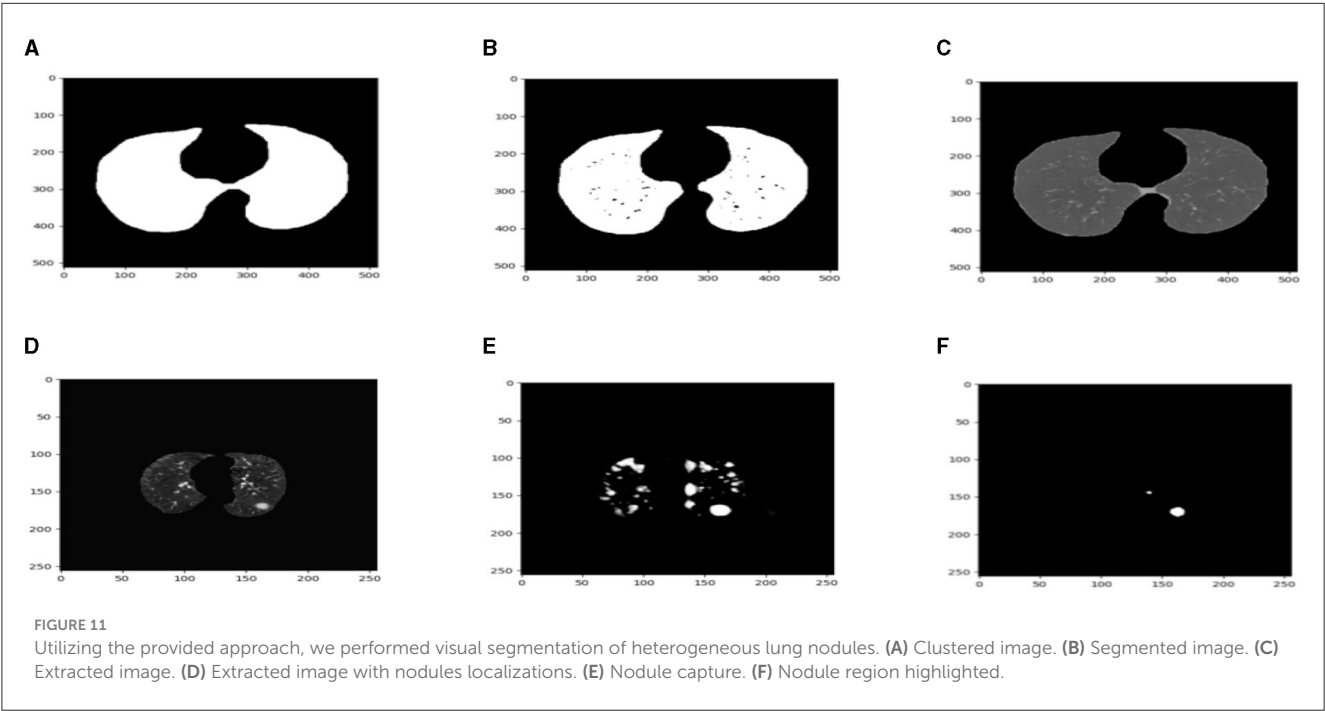
Method		Score			96% C.I for mean	
		Mean	St. deviation	St. error	Lower bound	Upper bound
IoU	Proposed methods	0.879	0.105	0.049	0.748	0.925
	U-Net (7)	0.805	0.104	0.073	0.677	0.916
	NU-NET (6)	0.780	0.117	0.062	0.680	0.824
	WU-NET (12)	0.731	0.144	0.088	0.613	0.865
Recall	Proposed method	0.933	0.035	0.043	0.814	0.963
	U-Net (7)	0.870	0.022	0.038	0.759	0.955
	NU-NET (6)	0.850	0.027	0.030	0.755	0.967
	WU-NET (12)	0.805	0.154	0.83	0.702	0.954
Precision	Proposed method	0.950	0.130	0.040	0.834	0.991
	U-Net (7)	0.890	0.106	0.029	0.780	0.992
	NU-NET (6)	0.880	0.086	0.024	0.770	0.996
	WU-NET (12)	0.831	0.154	0.062	0.704	0.946
F1-Score	Proposed method	0.940	0.120	0.040	0.826	0.993
	U-Net (7)	0.886	0.086	0.023	0.754	0.975
	NU-NET (6)	0.851	0.117	0.035	0.789	0.923
	WU-NET (12)	0.813	0.128	0.057	0.721	0.948

Bold values indicates is highest compared with other classifiers.

TABLE 12 Statistical analysis.

Multiple comparisons					96% C.I for mean	
Dependent	Model(a)	Methods(b)	Mean difference	Sig	Lower bound	Upper bound
IoU	Proposed model	U-Net (7)	0.072 [#]	<=0.001	0.061	0.094
		NU-NET (6)	0.096 [#]	<=0.001	0.084	0.125
		WU-NET (12)	0.145 [#]	<=0.001	0.125	0.165
Recall	Proposed model	U-Net (7)	0.055 [#]	<=0.001	0.038	0.074
		NU-NET (6)	0.061 [#]	<=0.001	0.050	0.075
		WU-NET (12)	0.115 [#]	<=0.001	0.103	0.135
Precision	Proposed model	U-Net (7)	0.056 [#]	<=0.001	0.045	0.078
		NU-NET (6)	0.065 [#]	<=0.001	0.055	0.085
		WU-NET (12)	0.122 [#]	<=0.001	0.104	0.145
F1-Score	Proposed model	U-Net (7)	0.048 [#]	<=0.001	0.038	0.065
		NU-NET (6)	0.072 [#]	<=0.001	0.065	0.089
		WU-NET (12)	0.121 [#]	<=0.001	0.112	0.137

[#]Indicates proposed Method is better than the existing classifiers.



The images in Figure 12 show a DSC value of at least 0.8 can be trusted for most tumors. The dice index results were compared with the U-NET++ architecture's specific performance to ensure that the model's results were correct. The Dice similarity score (DSC) for the U-NET++ model was 90.84%, which is an unusually high level of success. Because it has fewer parameters than the original U-NET design, the U-NET++ model can effectively separate features and divide them into groups.

The ROC curves in Figure 11 demonstrate that radiologists have the capacity to obtain much greater levels of specificity (i.e., decreased false positive rates) with a low impact on

sensitivity (31). By narrowing down the requirement for a positive screen for individuals who are recommended to undergo repeat computed tomography (CT) scans, it is possible to achieve a specificity of 92.4%, while slightly decreasing the sensitivity to 86.9%.

6 Conclusions and future work

Lung segmentation is necessary for the effective diagnosis and identification of lung disorders. There has been a frenzy of

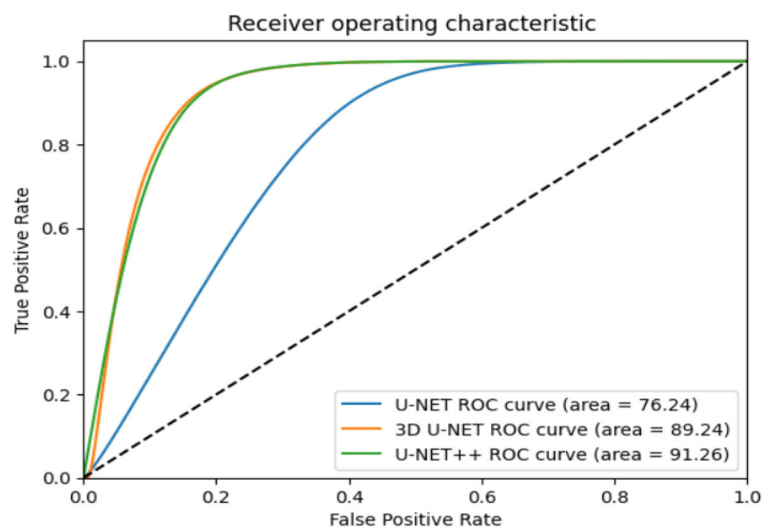


FIGURE 12
AUC curve for the proposed classifier with respect to other classifiers.

lung segmentation research over the past few years, all aimed at improving the accuracy. To identify and categorize lung illnesses, automated analysis of a CT scan must first “segment” the lung. The precision at which lung segmentation can be performed has been the subject of several studies. Deep learning algorithms and basic thresholding approaches have been applied to lung segmentation. U-NET++ is particularly effective in separating cells and neurons from images acquired using a PET Scan. In this study, U-NET++ was used for lung segmentation. The accuracy of the lung segmentation using U-NET++ was 91%. The original purpose of U-NET++ was to separate tiny images. The lungs were effectively divided using CT images. By shrinking the images, they were reduced from 128×128 to 32×32 pixels. There were 25 convolutional layers in total in this network. It is much more accurate to train U-NET++ using an original image size of 128×128 . The convolutional layers may be increased in size to enhance the accuracy of the filter.

Data availability statement

The dataset used for the findings will be shared by the corresponding authors upon reasonable request.

Author contributions

EA: Investigation, Software, Writing – original draft, Writing – review & editing. EN: Methodology, Software, Supervision, Writing – original draft, Writing – review & editing. NJ: Formal analysis, Project administration, Supervision, Writing – original draft, Writing – review & editing. SB: Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing. HS: Resources, Validation, Writing – original draft, Writing – review & editing. SA: Conceptualization, Methodology, Visualization, Writing – original draft, Writing – review & editing.

JN: Resources, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R135), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Acknowledgments

This work was supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R135), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. This work was supported by the Deanship of Scientific Research, Vice President for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [GrantA013].

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or

claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Drokin I, Ericheva E. End-to-end lung nodule detection framework with model-based feature projection block. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2021). p. 91–100. doi: 10.1007/978-3-030-87589-3_10
2. Shirokikh B, Shevtsov A, Dalechina A, Krivov E, Kostjuchenko V, Golanov A, et al. Accelerating 3D medical image segmentation by adaptive small-scale target localization. *J Imag.* (2021) 7:35. doi: 10.3390/jimaging7020035
3. Chen W, Yang F, Zhang X, Xu X, Qiao X. MAU-Net: multiple attention 3D U-Net for lung cancer segmentation on CT images. *Procedia Comput Sci.* (2021) 192:543–52. doi: 10.1016/j.procs.2021.08.056
4. Blanc D, Racine V, Khalil A, Deloche M, Broyelle J-A, Hammouamri I, et al. Artificial intelligence solution to classify pulmonary nodules on CT. *Diagn Interv Imaging.* (2020) 101:803–10. doi: 10.1016/j.diii.2020.10.004
5. Zhao Y, Chen S, Shen F, Long D, Yu T, Wu M, et al. In vitro neutralization of autocrine IL-10 affects Op18/stathmin signaling in non-small cell lung cancer cells. *Oncol Rep.* (2019) 41:501–11. doi: 10.3892/or.2018.6795
6. Huang W, Hu L. Using a noisy U-Net for detecting lung nodule candidates. *IEEE Access.* (2019) 7:67905–15. doi: 10.1109/ACCESS.2019.2918224
7. Zhao C, Han J, Jia Y, Gou F. Lung nodule detection via 3D U-Net and contextual convolutional neural network. In: *2018 International Conference on Networking and Network Applications (NaNA)*, IEEE (2018). p. 356–361. doi: 10.1109/NANA.2018.8648753
8. Chiu T-W, Tsai Y-L, Su S-F. Automatic detection of lung nodes with deep learning in segmentation and imbalanced data labeling. *Sci Rep.* (2021) 11:11174. doi: 10.1038/s41598-021-90599-4
9. Gao R, Lin Q, Man Z, Cao Y. Automatic lesion segmentation of metastases in SPECT images using U-Net-based model. In: *2nd International Conference on Signal Image Processing and Communication (ICSIPC 2022)* (2022). doi: 10.1117/12.2643560
10. Cai Y, Chen R, Liu H, Peng Z, Yuan Y, Lu Y, et al. Lung nodule CT medical image analysis based on deep learning. In: *AOPC 2022: AI in Optics and Photonics* (2023). doi: 10.1117/12.2652067
11. Banu SF, Sarker MM, Abdel-Nasser M, Rashwan HA, Puig D. WEU-Net: weight excitation U-Net for lung nodule segmentation. In: *Artificial Intelligence Research and Development* (2021). doi: 10.3233/FAIA210154
12. Xia L. Auxiliary diagnosis of lung cancer with magnetic resonance imaging data under deep learning. *Comput Math Methods Med.* (2022) 2022:1994082. doi: 10.1155/2022/1994082
13. Chhabra M, Ravulakollu KK, Kumar M, Sharma A, Nayyar A. Improving automated latent fingerprint detection and segmentation using deep convolutional neural network. *Neural Comput Applic.* (2023) 35:6471–97. doi: 10.1007/s00521-022-07894-y
14. Venkatesh C, Chinna Babu J, Kiran A, Nagaraju CH, Kumar M, A. hybrid model for lung cancer prediction using patch processing and deeplearning on CT images. *Multimed Tools Appl.* (2024) 83:43931–52. doi: 10.1007/s11042-023-17349-8
15. Madhu G, Kautish S, Gupta Y, Nagachandrika G, Biju SM, Kumar M. XCovNet: An optimized xception convolutional neural network for classification of COVID-19 from point-of-care lung ultrasound images. *Multimed Tools Appl.* (2024) 83:33653–74. doi: 10.1007/s11042-023-16944-z
16. Lamba M, Munjal G, Gigras Y, Kumar M. Breast cancer prediction and categorization in the molecular era of histologic grade. *Multimed Tools Appl.* (2023) 82:29629–48. doi: 10.1007/s11042-023-14918-9
17. Hesamian MH, Jia W, He X, Kennedy P. Deep learning techniques for medical image segmentation: achievements and challenges. *J Digit Imaging.* (2019) 32:582–96. doi: 10.1007/s10278-019-00227-x
18. Hasan AM, Al-Jawad MM, Jalab HA, Shaiba H, Ibrahim RW, Al-Shamasneh AR. Classification of Covid-19 coronavirus, pneumonia and healthy lungs in CT Scans using Q-deformed entropy and deep learning features. *Entropy.* (2020) 22:517. doi: 10.3390/e22050517
19. Seo H, Badii Khuzani M, Vasudevan V, Huang C, Ren H, Xiao R, et al. Machine learning techniques for biomedical image segmentation: an overview of technical aspects and introduction to state-of-art applications. *Med Phys.* (2020) 47:e148–e167. doi: 10.1002/mp.13649
20. Huang H, Lu RF, Tao JL Li Y, Zhang JQ. Segmentation of lung nodules in CT images using improved U-Net++. *Acta Photonica Sinica.* (2021) 50:73–83. doi: 10.3788/gzxb20215002.0210001
21. Jeong EJ, Choi D, Lee DW. U-Net deep-learning-based 3D cell counter for the quality control of 3D cell-based assays through seed cell measurement. *SLAS Technol.* (2021) 26:468–476. doi: 10.1177/24726303211017532
22. Ji S, Xiao Z. Integrated context and multi-scale features in thoracic organs segmentation. *J Image Graph.* (2021) 26:2135–2145. doi: 10.32604/cmc.2022.020561
23. Duan X, Xie D, Zhang R, Li X, Sun J, Qian C, et al. A novel robotic bronchoscope system for navigation and biopsy of pulmonary lesions. *Cyborg Bionic Syst.* (2023) 4:0013. doi: 10.34133/cbsystems.0013
24. Cui Y, Wang X, Lin F, Li W, Zhao Y, Zhu F, et al. MiR-29a-3p improves acute lung injury by reducing alveolar epithelial cell PANoptosis. *Aging Dis.* (2022) 13:899–909. doi: 10.14336/AD.2021.1023
25. Petit O, Thome N, Rambour C, Themyr L, Collins T, Soler L. U-net transformer: Self and cross attention for medical image segmentation. In: *Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12*, Springer International Publishing (2021). p. 267–276. doi: 10.1007/978-3-030-87589-3_28
26. Ali Z, Irtaza A, Maqsood M. An efficient U-net framework for lung nodule detection using densely connected dilated convolutions. *J Supercomput.* (2022) 78:1602–23. doi: 10.1007/s11227-021-03845-x
27. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A. editors *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. MICCAI 2015. *Notes in Computer Science*, Cham: Springer (2015). doi: 10.1007/978-3-319-24574-4_28
28. Zheng W, Lu S, Yang Y, Yin Z, Yin L, Ali H. Lightweight transformer image feature extraction network. *PeerJ Comput Sci.* (2024) 10:e1755. doi: 10.7717/peerj-cs.1755
29. Yang C, Sheng D, Yang B, Zheng W, Liu C. A dual-domain diffusion model for sparse-view CT reconstruction. In: *IEEE Signal Processing Letters* (2024). doi: 10.36227/techrxiv.23726703.v2
30. Zheng J, Yue R, Yang R, Wu Q, Wu Y, Huang M, et al. Visualization of zika virus infection via a light-initiated bio-orthogonal cycloaddition labeling strategy. *Front Bioeng Biotechnol.* (2022) 10:1051. doi: 10.3389/fbioe.2022.940511
31. Rocha J, Cunha A, Mendonça AM. Conventional filtering versus U-Net based models for pulmonary nodule segmentation in CT images. *J Med Syst.* (2020) 44:81. doi: 10.1007/s10916-020-1541-9



OPEN ACCESS

EDITED BY

Sultan Ahmad,
Prince Sattam bin Abdulaziz University,
Saudi Arabia

REVIEWED BY

Prabhishek Singh,
Bennett University, India
Mohammad Tabrez Quasim,
University of Bisha, Saudi Arabia
Ayesha Mukhtar,
King Saud bin Abdulaziz University for Health
Sciences, Saudi Arabia
Piyush Kumar Shukla,
Rajeev Gandhi Technical University, India

*CORRESPONDENCE

Badar Almarri
✉ baalmarri@kfu.edu.sa

RECEIVED 05 April 2024

ACCEPTED 06 June 2024

PUBLISHED 19 July 2024

CITATION

Alarjani M and Almarri B (2024) Multivariate
pattern analysis of medical imaging-based
Alzheimer's disease. *Front. Med.* 11:1412592.
doi: 10.3389/fmed.2024.1412592

COPYRIGHT

© 2024 Alarjani and Almarri. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Multivariate pattern analysis of medical imaging-based Alzheimer's disease

Maitha Alarjani and Badar Almarri*

Department of Computer Science, College of Computer Sciences and Information Technology, King Faisal University, Al-Hofuf, Saudi Arabia

Alzheimer's disease (AD) is a devastating brain disorder that steadily worsens over time. It is marked by a relentless decline in memory and cognitive abilities. As the disease progresses, it leads to a significant loss of mental function. Early detection of AD is essential to starting treatments that can mitigate the progression of this disease and enhance patients' quality of life. This study aims to observe AD's brain functional connectivity pattern to extract essential patterns through multivariate pattern analysis (MVPA) and analyze activity patterns across multiple brain voxels. The optimized feature extraction techniques are used to obtain the important features for performing the training on the models using several hybrid machine learning classifiers for performing binary classification and multi-class classification. The proposed approach using hybrid machine learning classification has been applied to two public datasets named the Open Access Series of Imaging Studies (OASIS) and the AD Neuroimaging Initiative (ADNI). The results are evaluated using performance metrics, and comparisons have been made to differentiate between different stages of AD using visualization tools.

KEYWORDS

Alzheimer's disease, blood flow, multivariate pattern analysis, fMRI, neuroimaging, biomarker

1 Introduction

The human brain is a highly complex organ regulating the human neurological system. The human neocortex has up to 100 billion neurons connecting throughout the brain (1). They constitute a vast, interconnected network linked to human activities and emotions. Various neuroimaging techniques can acquire a wide range of brain signals. The term "neuroimage" is based on the representation of brain functionality or architecture (2). AD is among the most common types of memory loss in the twenty-first century and is a significant healthcare problem. As per statistics, there are ~5.5 Americans aged 65 years and older affected by AD (3). AD is a progressive brain disease. It is marked by a loss of executive function that treatment cannot resolve. Thus, studies have been conducted to develop ways to predict the disease, especially before symptoms appear, to slow or prevent them from worsening (4).

Traditionally, AD was detected through an invasive technique. Recently, multiple neuroimaging modalities have been developed to identify AD: positron emission tomography (PET) uses specific radiotracers to visualize and quantify amyloid plaques in the brain; electroencephalography (EEG) is utilized to obtain the electrical activity; and functional magnetic resonance imaging (fMRI) is utilized to measure the functionality of the brain with the help of oxygen level change detection in various parts of the brain, such as voxels (5). Moreover, the anatomical brain features are studied using magnetic resonance imaging (MRI), having high spatial determination, and can compare soft tissues (6).

Because neuroimaging techniques are rapidly changing, combining large amounts of high-dimensional, multimodal neuroimaging data is challenging. Thus, computer-aided machine learning methods for consolidative study have rapidly become extremely popular, and multiple neuroimaging modalities have recently been developed to identify AD. A popular neuroimaging process for examining brain activity in neurodegenerative illnesses is resting-state fMRI (7).

Based on recent research, brain changes associated with AD begin up to two decades before symptoms appear. Due to the high cost and side effects of current medicines, it is essential to focus on enhancing the quality of life or reducing the impact of the disease. To this end, a computer learning model showing significant performance in predicting the disease earlier can help minimize losses (8).

The structure of the study is outlined as follows: Section II provides background on the phases of AD and BOLD data. Section III reviews previous study on fMRI data. Section IV introduces the framework, while Section V shows the results. Section VI discusses the evaluation metrics used, and Section VII compares the findings with previous studies. Finally, Section VIII concludes the study and outlines future research directions.

1.1 Motivation and contribution

In recent years, computer-aided design systems have become increasingly important in diagnosing and grading AD, a severe disease affecting many people, particularly the older population. AD causes memory loss and an inability to function in one's environment. The biology of the disease is not yet fully understood, and no cure or medication is currently available to prevent its progress. Early detection is essential for minimizing the impact of the illness and enhancing patients' quality of life. However, classifying AD is challenging due to various constraints involved in fMRI scans, such as low spatial resolution, image artifacts, and motion aftereffects. Despite the low spatial resolution, the abstract and high-level shapes can still provide valuable information for our analysis. With a large amount of data, we have the potential to capture a wide range of variations, which can help improve the robustness and generalization of the model, based on inter subjects. This diversity can also help identify and characterize AD patterns and various sub-types or stages. Addressing these problems at different stages is necessary to develop a robust detection and classification framework for AD.

The primary contributions of this study include:

- To apply techniques using MVPA to consider patterns across multiple variables simultaneously.
- To identify relevant features in order to mitigate the impact of irrelevant or redundant ones by using the LASSO method.
- To propose a framework for detecting AD based on brain signals using hybrid machine learning classifiers.
- To evaluate the results using performance metrics on the public datasets of OASIS and ADNI for improved accuracy rates.

1.2 Early diagnosis benefits

Early detection of AD is crucial for several reasons (9):

- **Early intervention:** It is referred to as the strategies implemented for the early detection of AD. As there is no treatment for AD, the progression of the disease can help to manage the symptoms (10).
- **Treatment planning:** Early detection of AD allows for the timely implementation of comprehensive treatment plans, including medications, lifestyle changes, and cognitive interventions.
- **Clinical trials:** Early detection enables individuals to participate in clinical trials for new treatments, which are crucial for advancing our understanding of AD and developing new therapies.
- **Learn about the management of AD symptoms.**
- **Develop a community for assistance.**
- **Conduct clinical studies to test any recent possible medication (9).**

2 Background

2.1 Phases of AD

AD has been classified into four stages (11), as shown in Figure 1:

2.1.1 Normal control

Normal control is also known as cognitive normal, which is the natural process of cognitive aging. Individuals of 66 years of age healthily, retaining their ability to think, respond and communicate. This is related to the natural aging process (12). They show no symptoms of AD.

2.1.2 Mild Cognitive Impairment or prodromal stage

The intermediate stage between healthy control and AD is referred to as MCI. During this stage, an individual experiences short-term memory loss and difficulty remembering the names of familiar people or objects as a symptom. According to studies, 80% of MCI patients advance to AD after a certain time period of ~5–6 years (12).

Individuals may experience minor abnormalities in cognitive function, but they are insufficiently severe to meet the criteria for the diagnosis of Alzheimer's disease in Early Mild Cognitive Impairment (EMCI stage) (13). Therefore, this stage is generally considered harmless. Not everyone with MCI will develop AD, and some people may even show improvement in their mental abilities. This stage damages the medial temporal lobe of the hippocampus and causes symptoms of short-term memory loss.

More progression is toward another alarming stage, which is Late Mild Cognitive Impairment (LMCI) (13), affecting the lateral and parietal lobes of the brain. Reading difficulties, poor object

recognition, difficulty knowing the names of people, and a lack of sense of direction are all symptoms of this stage.

2.1.3 Alzheimer's disease

AD is the final stage of the disease, characterized by severe memory loss, including the names of people and things. This stage is incurable (14). The stage of AD begins in the hippocampus and entorhinal cortex and gradually spreads to other brain sections, affecting the frontal, temporal, and occipital lobes of the brain. Poor judgment, impulsivity, a short attention span, and vision issues are all symptoms of this period. Advancing age, hereditary variables, brain traumas, vascular illnesses, pathogens, and external conditions are among the risk factors contributing to AD development, as shown in Figure 2. What leads to the pathological changes observed in AD remains unclear. While several theories exist, two of the most prominent ones suggest that cholinergic dysfunction and amyloid protein abnormalities may be significant risk factors. However, no widely accepted explanation exists for the underlying mechanisms of AD (15).

2.2 Blood Oxygenation Level-Dependent signal

Several important factors influence the BOLD signal, as shown in Figure 3.

The complex interaction between neural action and causing a hemodynamic reply, and how an MRI scanner can detect this response. The magnetic field intensity, echo duration, and type of imaging technology used are only a few of the experimental factors in the scanning of fMRI that influence the number of BOLD signals detected by each scanner. For instance, although the hemodynamic response is the same, a 1% BOLD signal throughout an echo of 30 ms is comparable to 2% over an echo period of 60 ms, and the reaction is continuous. Additionally, BOLD imaging is prone to several aberrations, including field inhomogeneities, ghosting, and head motion (17). Determining how accurately the BOLD reply imitates a specific hemodynamic response is challenging due to the number of interacting variables.

The balloon method by Buxton et al. (18) has been developed through extensive research on the type of hemodynamic reply, particularly by Friston et al. (19). As previously mentioned by Buxton et al. (18), the BOLD signal vascular basis is primarily thought to be a relative inequality between rises in blood flow of local cerebral and concomitant (albeit smaller) rises in oxygen digestion, resulting in a brief drop in the deoxyhemoglobin to oxyhemoglobin ratio.

The blood volume, hematocrit, vascular geometry, and oxygenation levels of basal are other physiological variables affecting changes in the deoxyhemoglobin concentration (20, 21). Despite these crucial starting conditions, the hemodynamic response can differ significantly between species and cortical areas. Different facets of the hemodynamic response may alter on various timescales and have various neuronal underpinnings and effects on the signal of BOLD. It is now widely acknowledged that the signal of BOLD also occurs at prominent draining veins, possibly

a few centimeters below the neuronally active part, in addition to capillaries. Inferentially, such changes in the signal would be located spatially apart from the stimulated brain tissue.

Consequently, regarding the "brain vs. vein" debate (22), suggest that the density is based on microvascular, which will consistently be less than that of neurons (23), is impeded by massive contributions of vessels, and is more likely to be the aspect restraining the BOLD-based fMRI spatial resolution. The spin-echo fMRI method reduces these vein contributions, making them potentially useful for more precisely tracing the neuronal sources of fMRI BOLD (24, 25). Capillaries have a more significant impact on image intensity (26), with stronger field strengths. Therefore, these two might become more beneficial when used together.

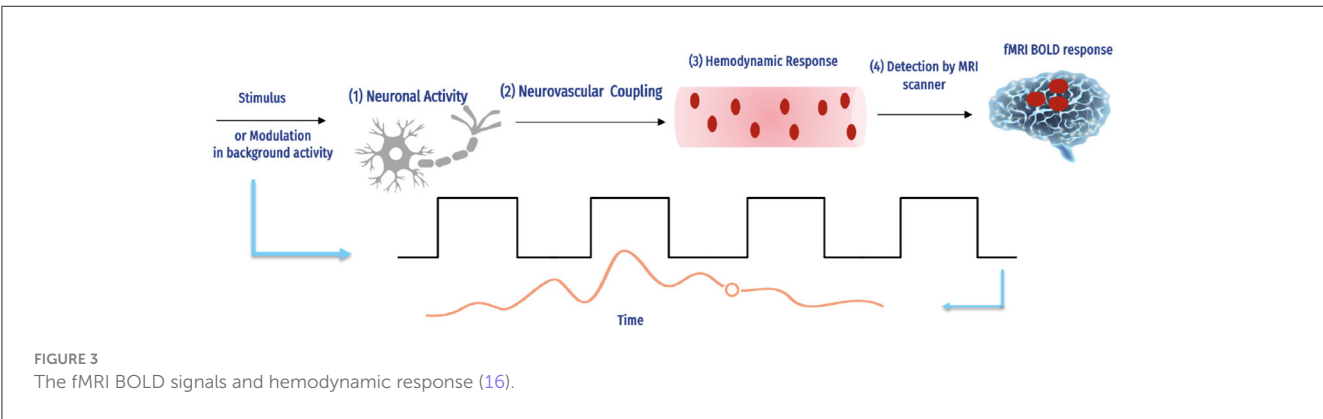
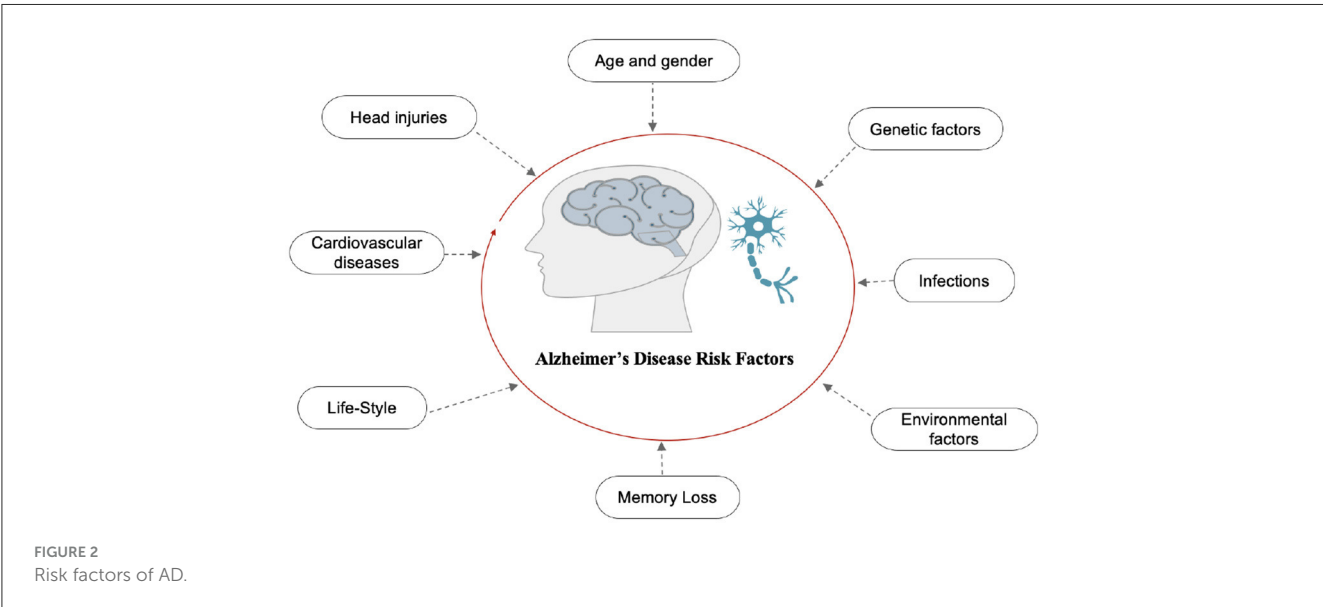
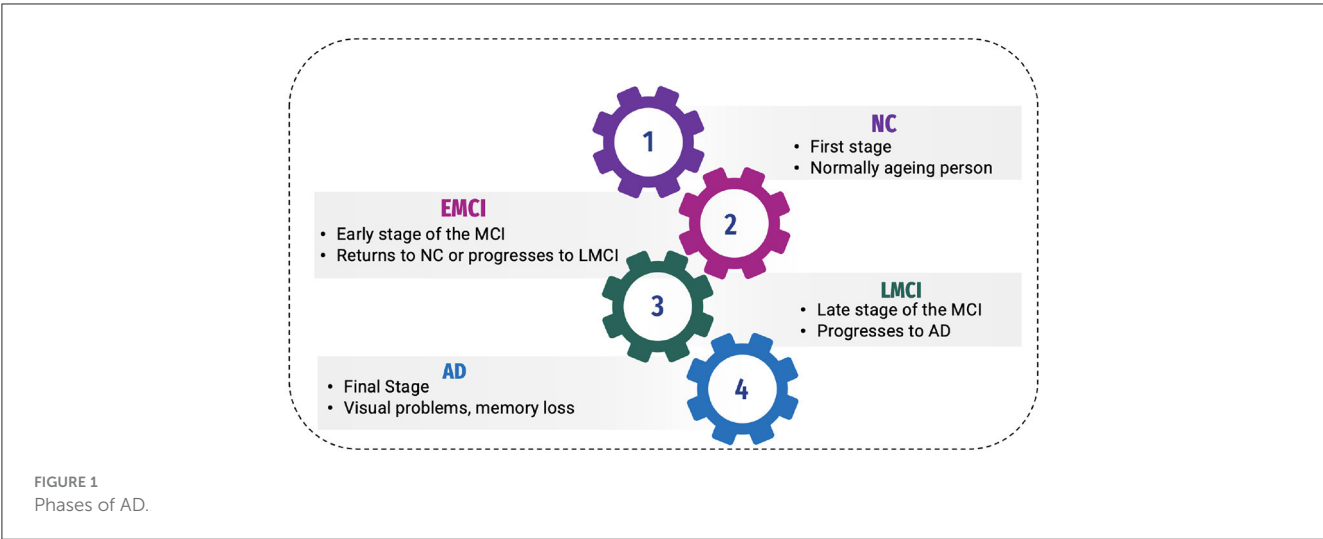
3 Literature review

Several studies have developed ideas for systems that could be used to classify AD. This section examines current studies using deep learning (DL) and machine learning (ML) models in systems for diagnosing and detecting AD. Some previous studies on detecting this disease have used standard ML methods (27). Additionally, many neuroimaging studies feature extraction strategies for fMRI signals; for example, Lama and Kwon (28) implemented graph theory to help predict AD at three stages: AD, MC, and NC, with classifications based on the linear support vector machine (SVM) and the regularized extreme learning machine. The Node2vec graph embedding approach converts graph features into feature vectors.

Parmar et al. (29) developed a 3D-CNN that uses rs-fMRI data to predict AD development. By employing unconventional techniques, they extracted patterns from neuroimaging data and found that a simple deep-learning model works well in categorizing AD. The findings of the study suggest a promising future, where fMRI-based biomarkers could assist in the early diagnosis and classification of AD. The study achieved 96.67% accuracy.

Guo and Zhang (30) introduced a distinct network using an autoencoder(AE) to detect natural aging and progression disorders. The network is based on biased neural networks and can easily diagnose AD. The researchers evaluated the system using the fMRI AD dataset and observed that it provides 25% better accuracy than other methods. The study achieved a remarkable 94.6% accuracy. Another study by Alarjani et al. (31) compared machine learning (ML) and deep learning (DL) models for early detection of AD using fMRI data. A 3D convolutional neural network (3D-CNN) extracted features from support vector machine (SVM) for classification. The 3D-CNN achieved 98.3% accuracy, while the SVM achieved 97.5%.

Shahparian et al. (32) developed an ML-based system that detected AD using fMRI images. The system is used to calculate time series for specific anatomical regions using the individual's fMRI data, and the latent low-rank representation method is utilized to extract pertinent features. Based on the acquired characteristics, the SVM classifier determines whether the person is healthy at the onset of the disease or has AD. The proposed method has an accuracy exceeding 97.5%. The problem with vascular dementia (VD) and AD is that both are more frequent. These may cause controversial diagnoses while using classical MRI and



clinical methods. Castellazzi et al. (33) different ML algorithms alongside combinations of MRI data are analyzed. AD and VD are two of the most common. Concerning AD and VD, they may demonstrate multiple neurological symptoms that may lead to

ambiguous diagnoses when using MRI criteria and conventional clinical. To overcome this problem, a method to classify AD and VD is presented. The system is assessed by three algorithms, such as ANN, SVM, and neuro-fuzzy inference.

Wang and Lim (34) conducted a new assessment approach introduced for individuals with AD and MCI compared with NC individuals, which utilized the zoom-in neural network DL algorithm. By extracting features from the resting-state fMRI dataset obtained from the ADNI, the algorithm could detect the implicated regions during AD by utilizing the automated anatomical labeling (AAL) Atlas. The study found that the ZNN obtained good results of 97.7, 84.8, and 72.7% accuracy for distinguishing AD from NC and MCI, NC from MCI and AD, and MCI from NC and AD, respectively. This was achieved using seven discriminative ROIs in the AAL-90.

Data optimization is indeed a complex task in the field of neuroimaging. However, Zamani et al. (35) proposed an interesting approach integrating artificial neural network (ANN) with evolutionary algorithms to optimize the neuroimaging data with multiple parameters. Using the rs-fMRI data based on the resting state, they measured the FC and computed 1,155 parameters. They tested the system using the ADNI dataset and achieved 94% accuracy.

To achieve AD discrimination at various stages, Nguyen et al. (36) suggested a voxel-wise discriminative system for multi-measuring rs-fMRI and combining hybrid MVPA and extreme learning machine (ELMs) and applied it to two different datasets. Jiao et al. (37) proposed a method focusing on the multi-scale combination of features. This approach utilizes global static features, moment features, and more refined features extracted from networks that are static, dynamic, and high-order functional. Subsequently, SVM was used to classify EMCI versus NC. Lu et al. (38) developed a system categorizing AD, MCI, and CN of fMRI data using FC throughout the brain rather than feature selection. They then used an ELM to classify binary stages. Unfortunately, this framework is only appropriate for a small dataset.

Yang et al. (39) extensively applied the brain function network to classify AD biomarkers 240 in the MCI stage. They used multiple time points of rs-fMRI data by combining the fused sparse network model based on centralized learning that is parameter-free. The essential features selected by the similarity network fusion method were then used to classify them using SVM. In addition, Chan et al. (40) proposed approach for AD uses a graph neural network (GNN) on MRI and fMRI scans. It encodes scans into brain graphs, clusters representations learned by the GNN to identify disease subtypes, and constructs population graphs for final decision-making. This approach outperforms existing methods, identifying three AD subtypes and revealing unique biomarkers, such as left cuneus and left isthmus cingulate cortex degeneration.

Lama et al. (41) constructed the brain network using Pearson's correlation-based FC of fMRI data. The brain network's graph features were transformed into feature vectors using the Node2vec graph embedding technique. Furthermore, they selected features using various approaches, which they then applied to classifiers: single-layered extreme learning and multi-layered ELM. Koluragi et al. (42) combined SVM and EfficientNetB0 to improve the performance. The integrated approach outperformed individuals, leveraging EfficientNetB0's efficient resource utilization and balance.

In earlier research, rs-fMRI used a mono-band frequency range and focused on low-order neurodynamics. Thus, high-order

neurodynamics were deliberately excluded. To address these issues, Sethuraman et al. (43) proposed an automated system to detect AD using rs-fMRI. The system constructs a high-order neurodynamic functional network using different levels of rs-fMRI time-series data, such as slow4 and slow5, and the full-band ranges from 0.027 to 0.08 Hz, 0.01 to 0.027 Hz, and 0.01 to 0.08 Hz. SVM and k-nearest neighbor (KNN) were used for ML, and AlexNet and Inception were implemented to classify various stages of AD. The system achieved 96.61% accuracy in differentiating between AD and NC. Begum and Selvaraj (44) used deep CNN (DCNN) and 3D densely connected convolutional neural network algorithms to diagnose AD and perform feature analysis on fMRI data.

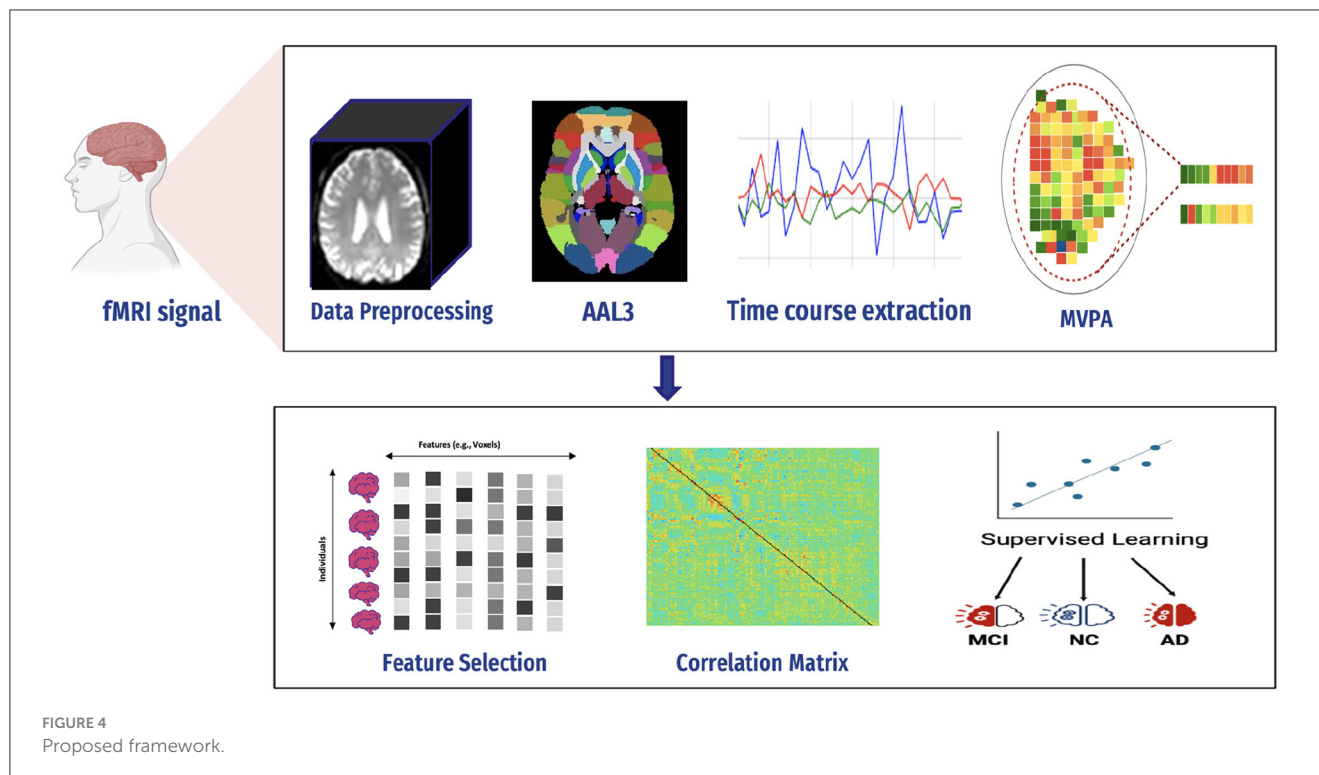
To enhance early detection (45), the effectiveness of Extreme Learning Machines (ELMs) was assessed alongside fMRI-based FC metrics. The non-linear methods such as MIC and eMIC were applied as classification features leads to robust outcomes. The study achieved a 95% accuracy rate in distinguishing between AD and NC using these methods. The study conducted by Penalba-Sánchez et al. (46) investigated the dynamic and static FC of resting-state fMRI using various methods across 116 ROIs for four participant groups. Additionally, they utilized graph theory metrics to investigate network segregation and integration. The results showed that the EMCI group had a longer typical path length and lower degree compared with the healthy control (HC) group.

3.1 Important of gap

MVPA techniques can enhance the ability to detect significant changes in the activity of the brain that may not be noticeable with traditional univariate methods. This is particularly important in AD, where early detection of subtle changes can be crucial for timely intervention. Additionally, MVPA allows a more detailed understanding of how different brain regions interact and contribute to cognitive processes. This can provide valuable insights into the underlying mechanisms of AD and other neurological disorders.

4 Proposed framework

AD is a serious health condition affecting many people, particularly the older population worldwide. It is a debilitating illness causing memory loss and impairing one's ability to interact with their surroundings. Early detection is crucial in mitigating the effects of Alzheimer's disease and improving the quality of life. Recognizing the disease at its onset enables the reduction of its impact on patients. We constructed a predictive framework to detect AD at an early stage based on human brain imaging techniques: fMRI. Figure 4 presents a summary of the proposed framework. It includes the following steps: (1) data collection (i.e., fMRI), (2) preprocessing of fMRI data to avoid artifacts (i.e., noisy), (3) computing FC through MVPA, (4) extracting time series of fMRI data, (5) computing correlation matrices for each stage, (6) feature selection to select relevant features (i.e., voxel), (7) supervised learning, and (8) evaluation and analysis.



4.1 fMRI signal preprocessing

Since medical images are complex and difficult to extract features, various techniques must be used to process images in the dataset.

A flexible preprocessing pipeline is used to prepare functional and structured data, including realignment, slice timing correction (STC), normalization to MNI space, and smoothing (47). For realignment, we utilized the SPM realignment unwarping procedure suggested by Andersson et al. (48). Then, scans are co-registered based on a reference image, such as the first scan of the first session. For this, a least square technique and a transformation of a 6-parameter (rigid body) are utilized, as presented in the study by Friston et al. (49). After that, the interpolation of the B-spline was resampled to reduce the effects of motion and magnetic artifacts.

Temporal misalignment and methods were applied to identify scans. A reference BOLD image was developed by applying the mean to the scans, and the outliers were excluded. The SPM unified normalization algorithm is used to perform the normalization and obtain the standard MNI space (50, 51), with the probability map template based on default Ixi-549 tissue, as resampled to 2 mm isotropic voxels. Finally, the spatial convolution of the data was performed with the help of a Gaussian kernel of 6 mm full-width at half-maximum (FWHM) for smoothing (see Figure 5 for an illustration).

4.2 Functional connectivity

An essential application of fMRI studies is brain network mapping in AD patients and between the brain network

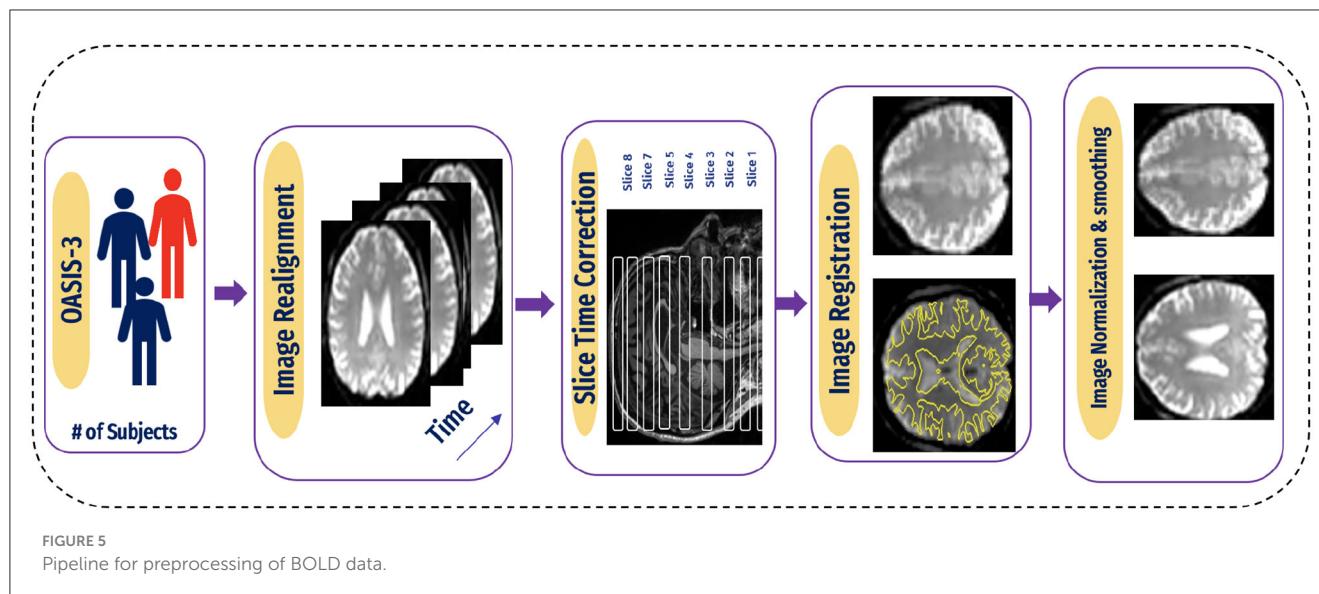
mapping routes. At rest, the default mode network is among the most exciting networks (52). DMN relates to knowing previous events, imagining future events, self-relevant mental processing, and checking external information (53). Alterations in DMN functional activity have been linked to neurological disorders (54–56). Most studies show decreased FC in the DMN. In a study by Koch et al. (57), the power of the DMN in rs-fMRI was examined to differentiate between three groups: CN individuals, MCI, and patients with AD. Moreover, this can be constructed using numerous imaging technologies [for example, EEG/magnetoencephalography [MEG] and structural, diffusion, and functional MRI]. Ways to analyze FC include UNIVAR and MVPA.

4.2.1 Univariate analysis

UNIVAR is a method used to analyze fMRI data. UNIVAR assesses the individual voxel neural activation or the average voxel activation of the brain. Thus, it is used for the localization of brain regions participating in processing specific stimuli such as face versus object. The conclusion about the brain regions participating in cognitive processes is also drawn from the study by Haynes and Rees (58). A general linear model is employed on each voxel, which is why it is called univariate (59). FCA characterizes communication between various brain regions during a task or rest. It also measures the relationship strength between the BOLD signal of the time series (60), as shown in Figure 6.

$$\forall x, y \quad r_n(x, y) = g_n * b(x, y) + \epsilon_n(x, y) \cdot \sigma(x, y) \quad (1)$$

Null hypothesis $C.b(x, y) = 0$



Here in Equation 1, n refers to the number of subjects in a study, and x and y are the voxel pair. The characterizing FC of these two voxels can be considered as $r_n(x, y)$, where g_n is referred to as the vector of a predictor of each subject n . The unknown regression coefficient of an unknown vector is $b(x, y)$, while $\epsilon_n(x, y)$ and $\sigma(x, y)$ are error term and inter-subject variance, respectively. A null hypothesis can be formed using $C.b(x, y) = 0$.

Many studies used UNIVAR, such as in the study by Moeller et al. (61), to identify the region's dynamic activity close to the expected waveform. In another method Bu et al. (62), the authors examined the UNIVAR and MVPA overlap.

4.2.2 MultiVoxel (or Multivariate) Pattern Analysis

Multivariate Pattern Analysis MVPA is the most used technique for analyzing functional data. In this study, the spatial pattern of neural activation across various voxels is considered (e.g., voxels in fMRI or channels in MEG/EEG). It also assesses whether it has information related to the task (63). It is called multivariate because it is based on analyzing a set of voxels rather than single voxel modeling (64). The similarities of such patterns can also be investigated by the activation of these patterns, such as by viewing a scene vs. a face, Norman et al. (65), as shown in Figure 7. The MVPA can be mathematically defined as follows (Equation 2).

$$\forall x \quad r_n(x) = g_n * B(x) + \epsilon_n(x) * \sum(x) \quad (2)$$

Null hypothesis $C.B(x).P(x) = 0$

While $r_n(x)$ refers to connectivity value whole map, unknown predictor of regression coefficients is denoted by $B(x)$. $\epsilon_n(x)$ refers to residual error. $\sum(x)$ is denoted as voxel-by-voxel matrix of positive definite. While C denotes between subject, $P(x)$ represents contrast matrix of between-voxels. There are many studies that used MVPA such as in the study by Yoon et al. (66), it used validate impairment hypothesis in schizophrenia-distributed representations. In another method, Lee et al. (67) conducted hypothesis by using MVPA to check that based on the brain

prediction, the efficiency of models has variations across the stimuli types.

4.3 Region of interest

After preprocessing BOLD fMRI data, we can extract features from the fMRI data depending on the atlas. Automated Anatomical Labeling (AAL) atlas is a tool used in neuroimaging that provides a pre-defined anatomical division of the human brain. This tool is widely used in neuroscience research, particularly in functional and structural brain imaging studies, such as fMRI and PET. The AAL atlas helps researchers to identify and label specific brain regions in their neuroimaging data. The human brain is divided into anatomical regions, each with a specific label in the AAL atlas. AAL atlas provides standardized three-dimensional coordinates for each region, which researchers can use to locate and precisely label brain imaging data areas. The AAL atlas performs various analyzes, including region-of-interest (ROI) studies in functional brain imaging, to map brain activity during specific tasks or resting-state conditions (68). Few types of the AAL atlas are as follows: AAL1 (69), AAL2 (70), Chinese AAL (71), AAL3 (68). Dealing with high-dimensional and small sample datasets such as fMRI data is challenging when it comes to classification and modeling. To address this issue, the AAL template is utilized in this study to calculate the functional link matrix after processing the original image. In Figure 8, the AAL3 used to perform feature extraction to identify relevant brain regions or patterns for the fMRI. AAL3 includes 170 regions, masking objects with an atlas to extract time series within each ROI (see Figure 9).

4.4 Compute connectivity

Multiple techniques are available to calculate the FC of fMRI. These techniques include connectivity maps of seed-to-voxel, ROI-to-ROI connectivity matrices, independent component analysis,

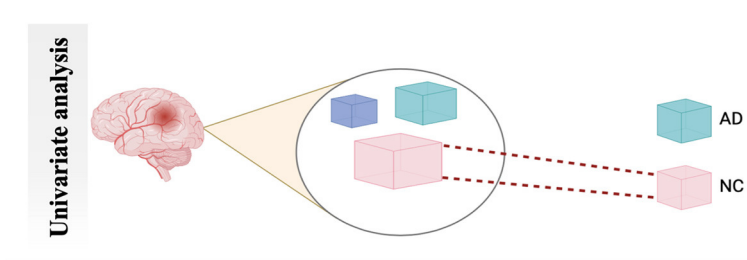


FIGURE 6
Schematic diagram of univariate analysis.

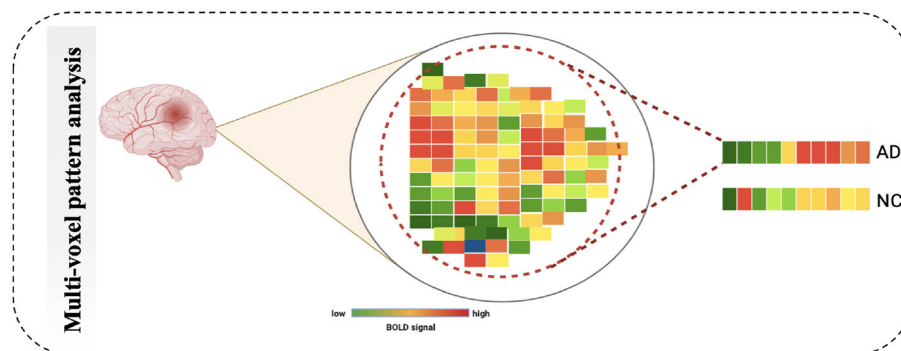


FIGURE 7
Schematic diagram of Multi-Voxel Pattern Analysis.

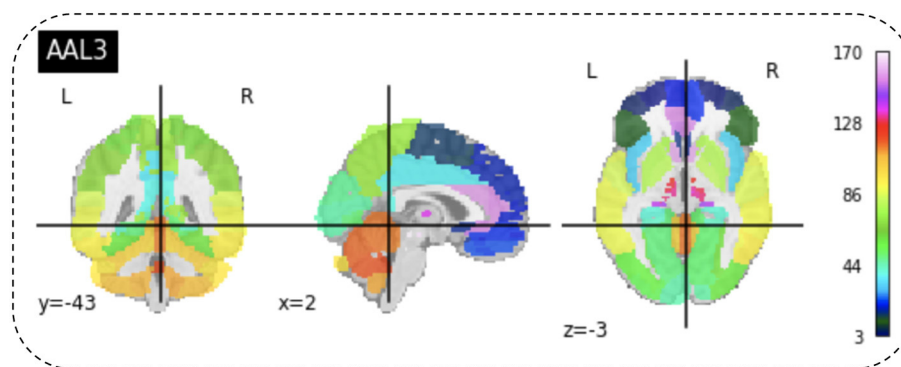


FIGURE 8
View for the AAL3 template.

and multivariate pattern analysis (MVPA). This study proposes FC using MVPA to analyze individual voxel resolution in the brain-wide connectome. This approach uses the MVPA methods to overcome the challenges of brain-wide connectome analysis. MVPA was applied to a 4D BOLD dataset to compute the correlation matrix between voxel time series within each ROI and remove relevant voxels based on their correlation with other voxels. These analyzes calculate a series of associated connectivity patterns and spatial maps that illustrate the voxel connectivity to the rest of the brain. Based on the provided fMRI time-series data, the calculated correlation matrix will then contain correlation values between ROI pairs. The FC matrix is displayed using the

AAL3 template, which includes 166 brain regions, resulting in a connectivity matrix of 166 X 166. The correlation matrix ranges from 0 to 1, with 0 indicating no correlation and 1 indicating a high degree of correlation. The matrix is shown in [Figures 10, 11](#).

4.5 Feature selection

4.5.1 LASSO

Suppose we have a data (x^i, Y_i) , $i = 1, 2, \dots, N$, where $x^i = (X_{i1}, Y_{ip})^T$ refers to the variables used for prediction, y_i refers to the response. In the usual setup of regression, we suppose that either all

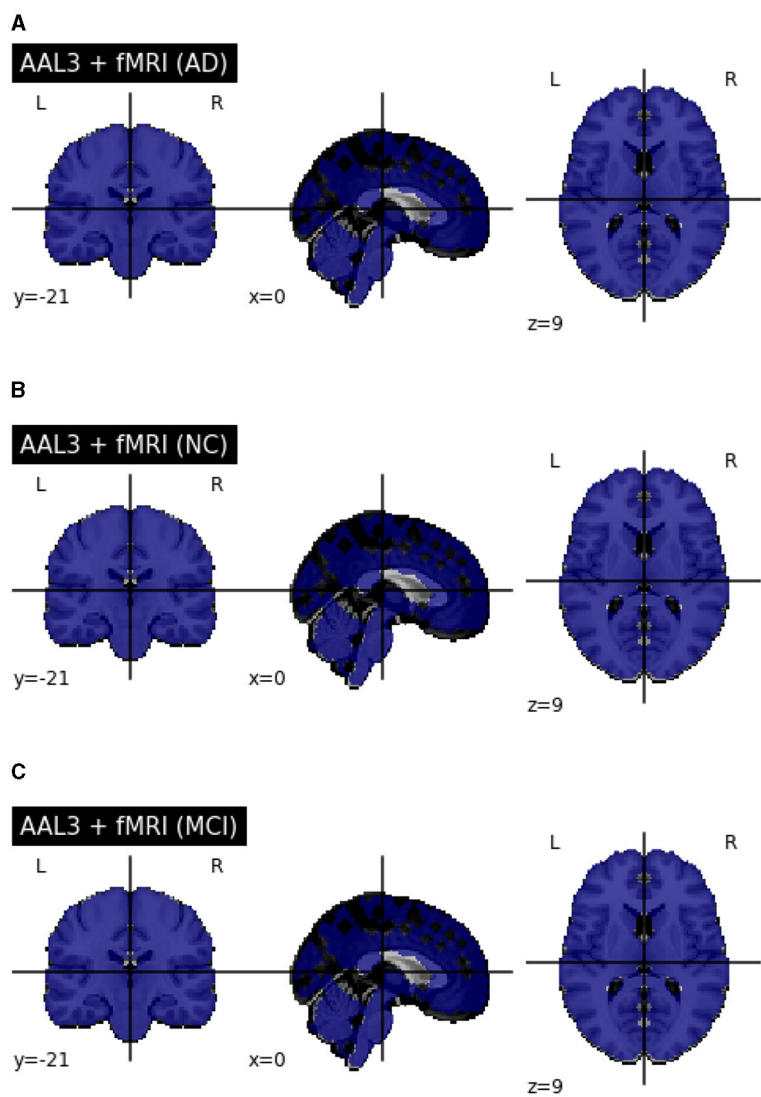


FIGURE 9
(A) Performed mask and functional of AD. (B) Performed mask and functional of MCI. (C) Performed mask and functional of NC.

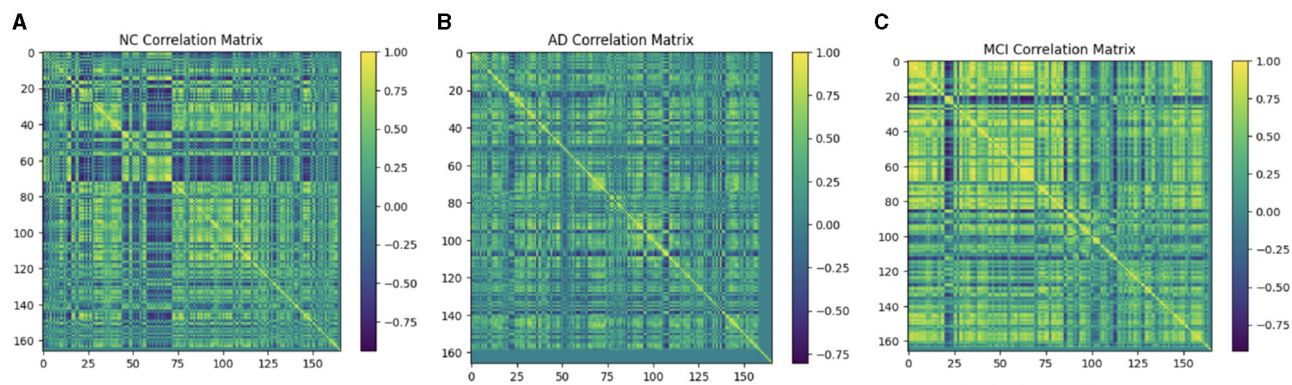


FIGURE 10
Functional connection matrix and brain network visualization for each stage (AD, MCI, and NC). (A) Functional connectivity NC. (B) Functional connectivity AD. (C) Functional connectivity MCI.

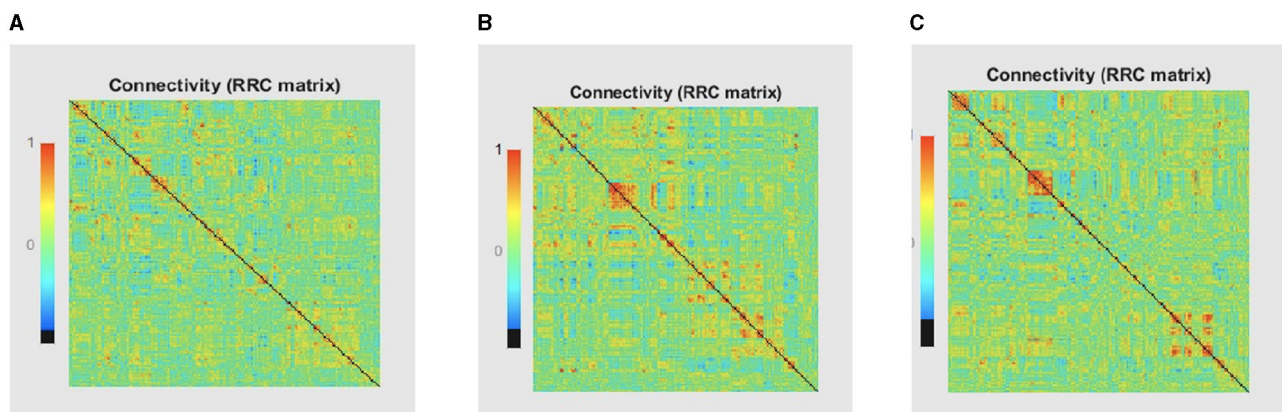


FIGURE 11

Functional connection matrix and brain network visualization for each stage (AD, MCI, and NC) in ADNI dataset. (A) Functional connectivity NC. (B) Functional connectivity AD. (C) Functional connectivity MCI.

observation is independent or y_i s are independently conditionally of the given y_{ij} s we can suppose that X_{ij} referred to as standardized $\sum_i x_{ij}/N = 0$, $\sum_i x_{ij}^2/N = 1$. Suppose $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)^T$, and the lasso estimate $(\hat{\alpha}, \hat{\beta})$

Here, $t \geq 0$ is referred to as a parameter for tuning. For all t , the solution for α is $\hat{\alpha} = \hat{y}$. We can consider without losing the generality that $\hat{\alpha} = 0$ which omit α . The solution of the above equation is a problem of quadratic programming having linear constraints of inequality.

The amount of shrinkage is controlled by the parameter $t \geq 0$. It is applied for estimation. Suppose $\hat{\beta}_j$ refers to the estimates of full least squares. Let $\sum |\hat{\beta}_j|$, then the shrinkage will occur due to $t < 0$. This shrinkage will occur in the solutions toward 0. There are some coefficients and value of these coefficients will be 0. If $t = t_0/2$, then the affect will be same as searching the best subset having a size of $p/2$. It is not necessary that the matrix of design will be of full rank.

The motivation behind the Lasso is from a proposal by Breiman, and it can be defined as Equations 3 and 4.

$$(\hat{\alpha}, \hat{\beta}) = \arg \min \left\{ \sum_{i=1}^N \left(y_i - \alpha - \sum_j \beta_j x_{ij} \right)^2 \right\} \quad \text{subject to } \sum_j |\beta_j| \leq t. \quad (3)$$

$$\sum_{i=1}^N \left(y_i - \alpha - \sum_j c_j \hat{\beta}_j^0 x_{ij} \right)^2 \quad \text{subject to } c_j \geq 0, \sum c_j \leq t \quad (4)$$

As previously mentioned, fMRI data are high-dimensional, with many voxels (3D pixels) representing regions of the brain. In this context, LASSO helps select a subset of these most relevant voxels for a particular analysis. Lasso is used as a regularization technique in linear regression methods. It adds a penalty term to the sum of squared errors, encouraging sparsity in the resulting model. It promotes the selection of a subset of features (voxels or ROI) while setting others to zero. Identifying relevant voxels or ROIs:

Lasso regression is employed to identify relevant features (relevant brain regions or voxels). The Lasso coefficients provide information about the importance of each feature by setting a suitable penalty parameter (alpha = 0.01). Features with non-zero coefficients are considered relevant, and those with coefficients set to zero are effectively excluded from the model (72). We selected the λ value that minimized the cross-validated mean squared error (MSE), as shown in Figures 12, 13.

4.6 Machine learning

Machine learning (ML) is among the most efficient and robust tools that have entered the medical imaging domain in the last few years. The recent advances in this field have enabled intelligent algorithms capable of assisting human experts in making wise decisions. Data are prepared in various directions, such as single and hybrid models, to classify the disease by organizing the time series of relevant voxel(s) into a matrix and labeling samples as AD, MCI, or NC based on their task condition or behavioral response. Table 1 presents the hyper-parameters of ML.

4.6.1 Single model

4.6.1.1 Support vector machine

Support vector machine (SVM) is among the most common classification and regression analysis algorithms. They use patterns found through data analysis and pattern recognition to predict newly collected data. The SVM classifies data into different classes by creating a hyperplane. The nearest points from each class are kept as far apart as feasible by the hyperplane, which is selected to optimize the margin between the two classes (73).

4.6.1.2 K-Nearest Neighbor

K-Nearest Neighbor (KNN) is an important ML model based on supervised learning. The approach assumes similarity between new and existing subjects. Subsequently, it places the latest subject in the group, which is mostly similar to the existing categories,

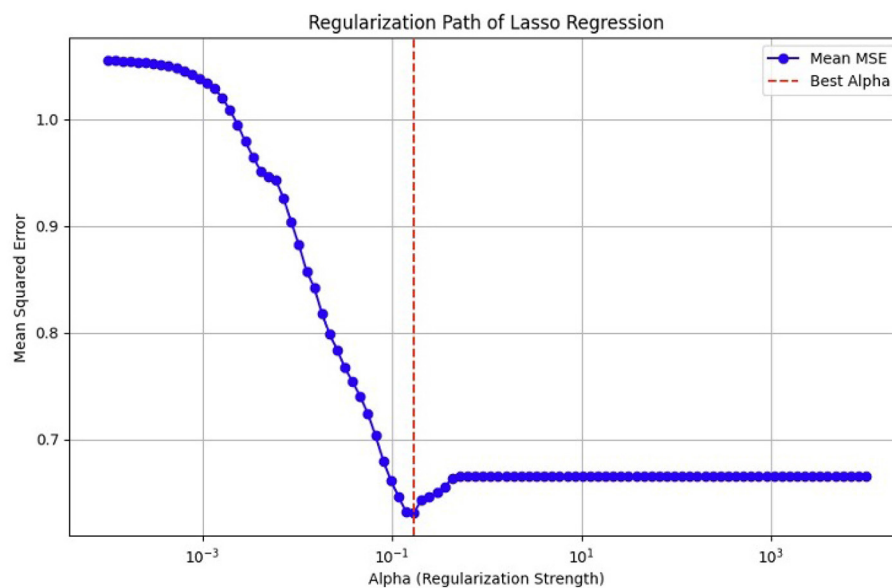


FIGURE 12
MSE of the LASSO fit, cross-validated with a parameter lambda (λ), for the OASIS dataset.

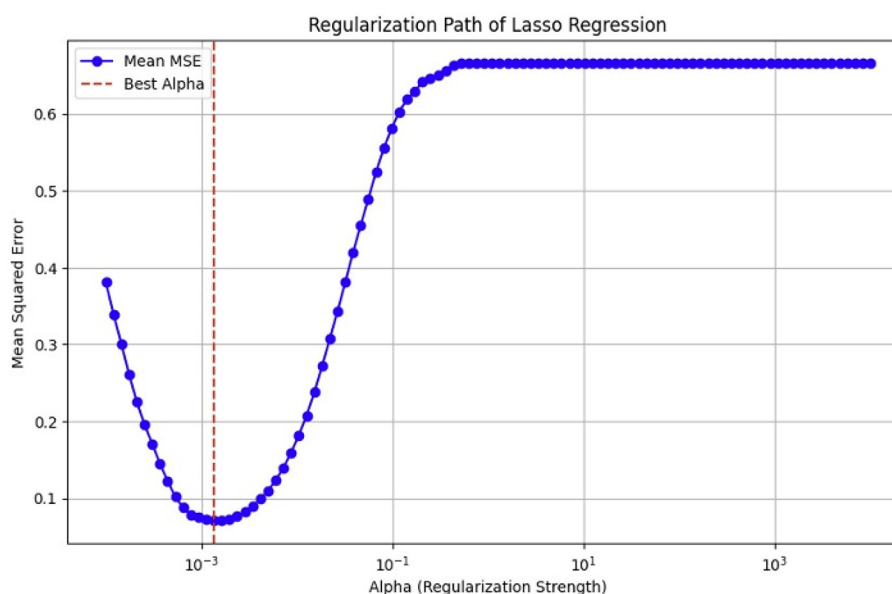


FIGURE 13
Mean squared error (MSE) of the LASSO fit, cross-validated with a parameter lambda (λ), for the ADNI dataset.

such as AD, MCI, and NC. The KNN algorithm saves all existing data and generates new subjects based on similarity. Once a new subject is developed, the KNN method instantly categorizes it into a suitable category. Notably, KNN is a non-parametric technique, so no assumptions about original data are made. During the training phase, the KNN algorithm stores the dataset and classifies new subjects into a category similar to the old data (74).

4.6.1.3 AdaBoost

AdaBoost classification involves training of numerous weak classifiers on the same training set to create a robust classifier. The

weak classifier is a stump of the tree. The models then decide which prediction is the best; however, this technique depends on the weak classifier. It would increase the accuracy if used along with another algorithm (75).

4.6.2 Hybrid approach

In our case, the ensemble classifier combined the predictions of three base classifiers, namely, SVM, AdaBoost, and KNN classifiers by voting (Figure 14). In soft voting, each data point in fMRI

collects the probability estimates (class probabilities) from each of the individual classifiers in the ensemble. The average or weighted average of these probabilities is computed for each class to make a final decision for that data point. The predicted class label for each data point is determined by selecting the class with the highest average probability.

TABLE 1 Tuning for machine learning models.

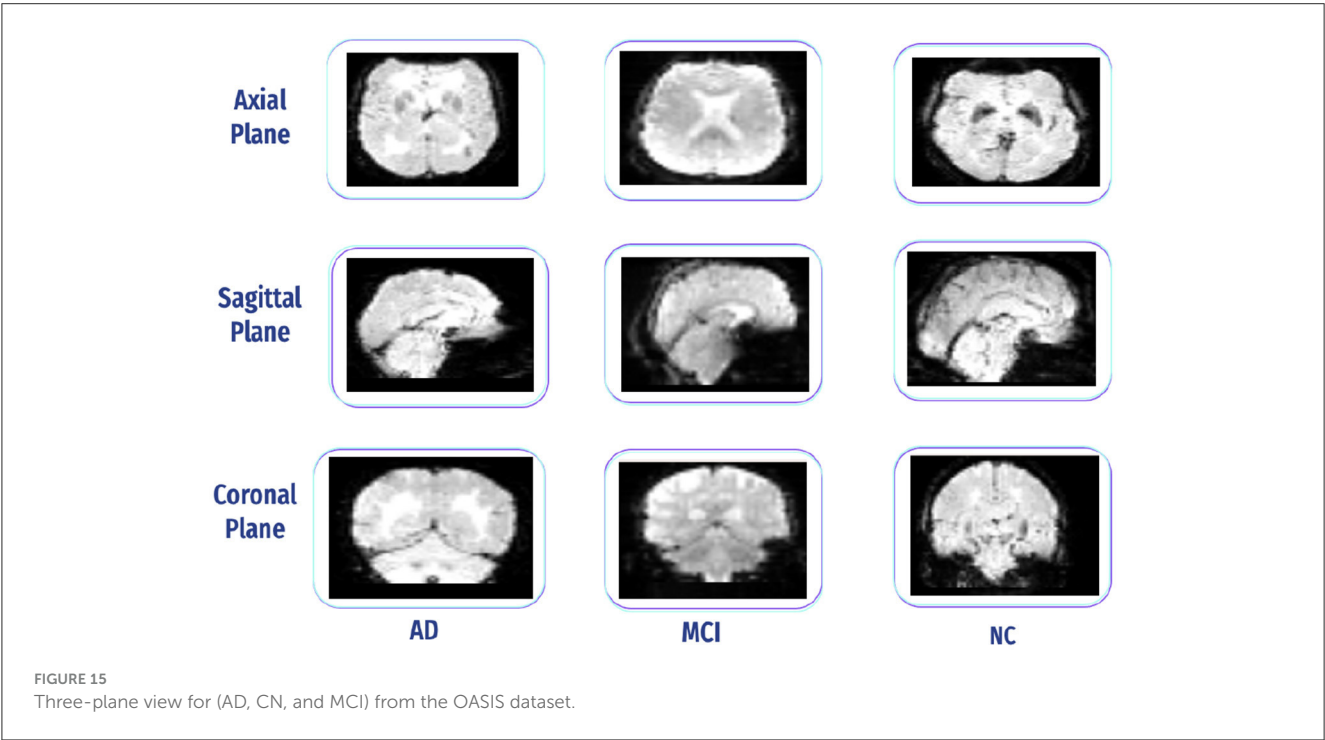
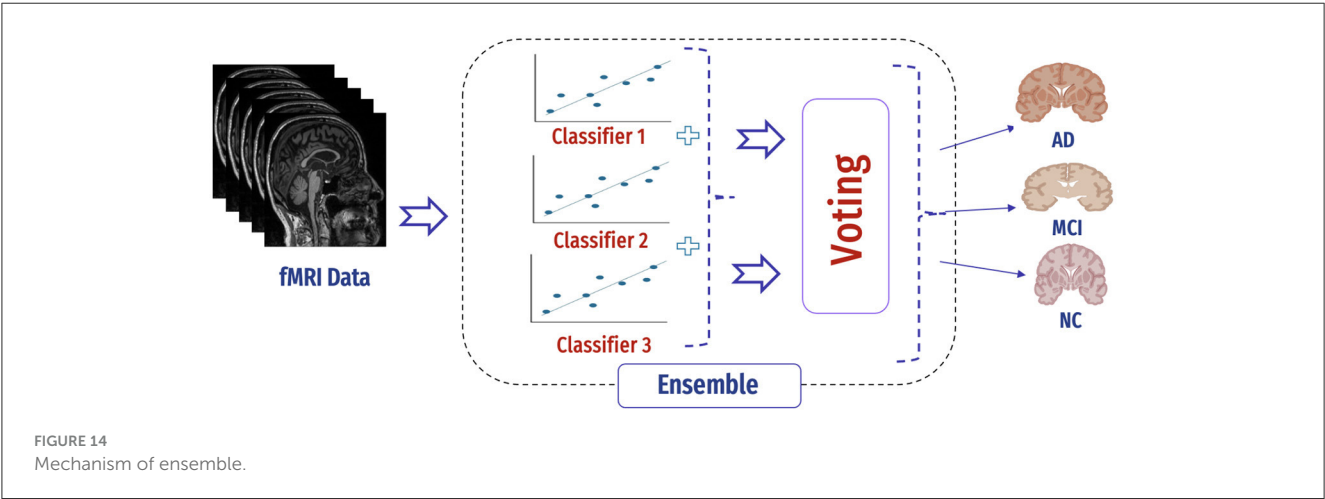
Model	Hyperparameters
SVM	kernel = "sigmoid", C = 0.2, random state = 300
AdaBoost	Default
KNN	n_neighbors = 300, weights = "uniform", p = 2
HML	voting = "hard"

5 Results and analysis

5.1 Datasets

5.1.1 Open Access Series of Imaging Studies

A free-access data set was retrieved from the OASIS dataset. It has three versions: OASIS-I, OASIS-II, and OASIS-III. OASIS-III is a longitudinal neuroimaging, biomarker, cognitive, and clinical dataset for normal aging and AD with ages varying between 42 and 95 years, including 1,379 subjects (male/female) and 2,842 MRI sessions, which include T1w, T2w, and resting-state BOLD (rs-BOLD). In our case, We used rs-BOLD data, typically acquired as a sequence of 3D brain volumes, with each volume representing a snapshot of brain activity, including all data for the mild cognitive impairment (MCI) stage, with all slices. We balanced the selection by choosing approximately the same number of samples for AD



and normal cognition (NC) classes to prevent bias. Figure 15 presents all three plane views of fMRI data. It has the following requirements: Each stage has several subjects (males and females), including functional and structural data (i.e., T1 W), and each subject has an array of size $64 * 64 * 36 * 164$, representing height * width * number of horizontal slices * number of data points. The data are captured using 3.0 T (Tesla) scanners with a slice thickness of 2.4 mm, and the flip angle is 80 degrees (76). Table 2 displays the subjects' demographic information.

5.1.2 AD Neuroimaging Initiative

Its free-access dataset is retrieved from the ADNI. The ADNI was initiated in 2004 under the leadership of Dr. Michael

W. Weiner. ADNI is a collaborative effort involving multiple institutions and researchers in the United States and Canada. It is a longitudinal study that was carried out in stages at several centers in North America (ADNI1, ADNIGO, ADNI2, and ADNI3). ADNI aims to develop biomarkers as clinical trial outcome measures. The ADNI includes MRI, PET, fMRI, and DTI and genetic data sessions at various stages for males and females (77). Additionally, we can select a sagittal, coronal, and axial plane, adding them to data collections and downloading them as NIFTI files, as shown in Figure 16. In this study, we downloaded fMRI data comprising 95 normal, 35 MCI, and 55 AD subjects.

Table 3 shows that the ADNI dataset is not imbalanced. However, we still need to address the issue of imbalanced data,

TABLE 2 Subject cohort of fMRI (OASIS dataset).

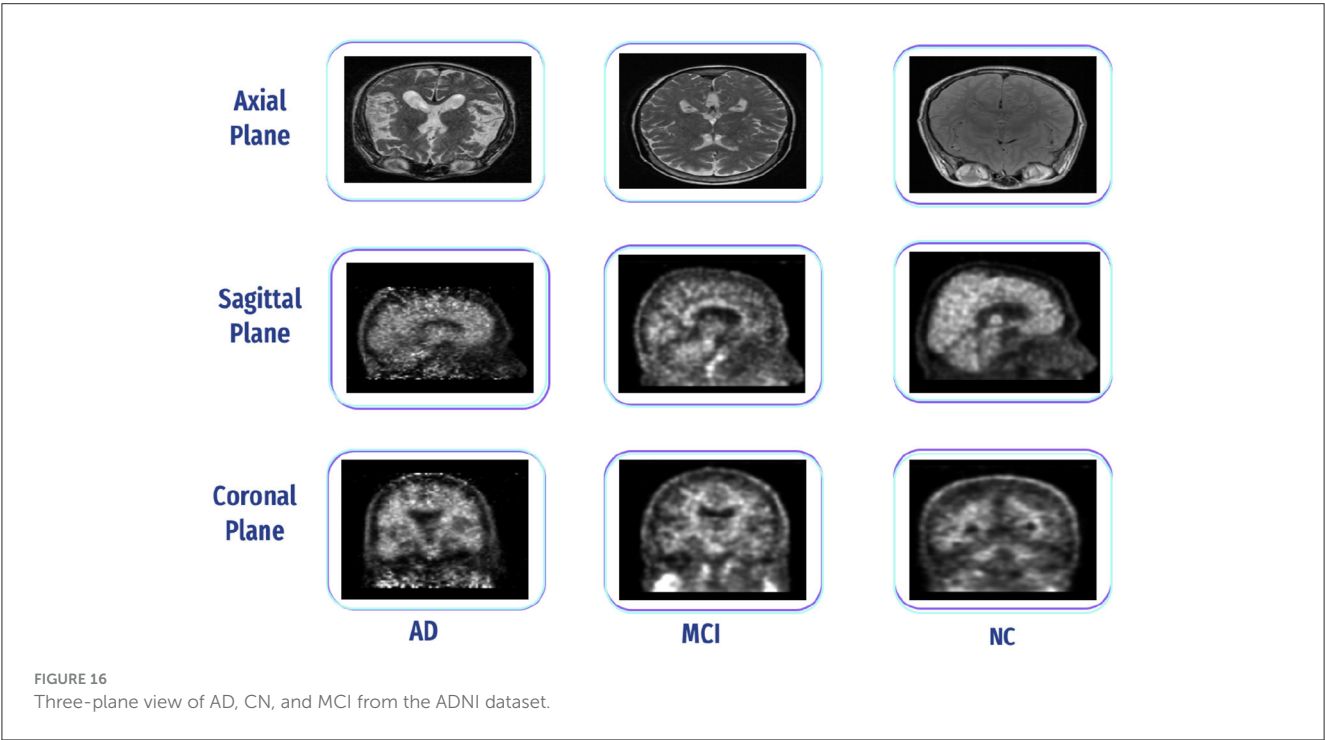
	AD	MCI	NC
Type of image	DICOM ^a		
Number of subjects	101	95	102
Total of slice	16,564	15,580	16,728
Sex (M/F)	72/29	40/55	60/42
Clinical dementia rating (CDR)	0.5	0	1
Flip angle	77°		
Voxel size in fMRI	3x3x3mm ³		
TR/TE	2 s/25 ms		
Width	64	64	64
Height	64	64	64
Acquisition scanner	3.0 T (Tesla)		

^a Digital Imaging and Communications in Medicine.

TABLE 3 Subject cohort of fMRI (ADNI dataset).

	AD	MCI	NC
Type of image	NIFTI*		
Number of subjects	55	35	95
Total of slices	1096	701	1900
Male/female	30/25	20/15	50/45
Range of age	65–75		
Acquisition plane	Axial rsfMRI (eyes open)		
Voxel size in axial rs-fMRI	3x3x3mm ³		
TR/TE	TR = 3 s; TE = 30		
Thickness	3.312999963760376 mm		
Acquisition scanner	Philips medical systems		

*Neuroimaging informatics technology initiative.



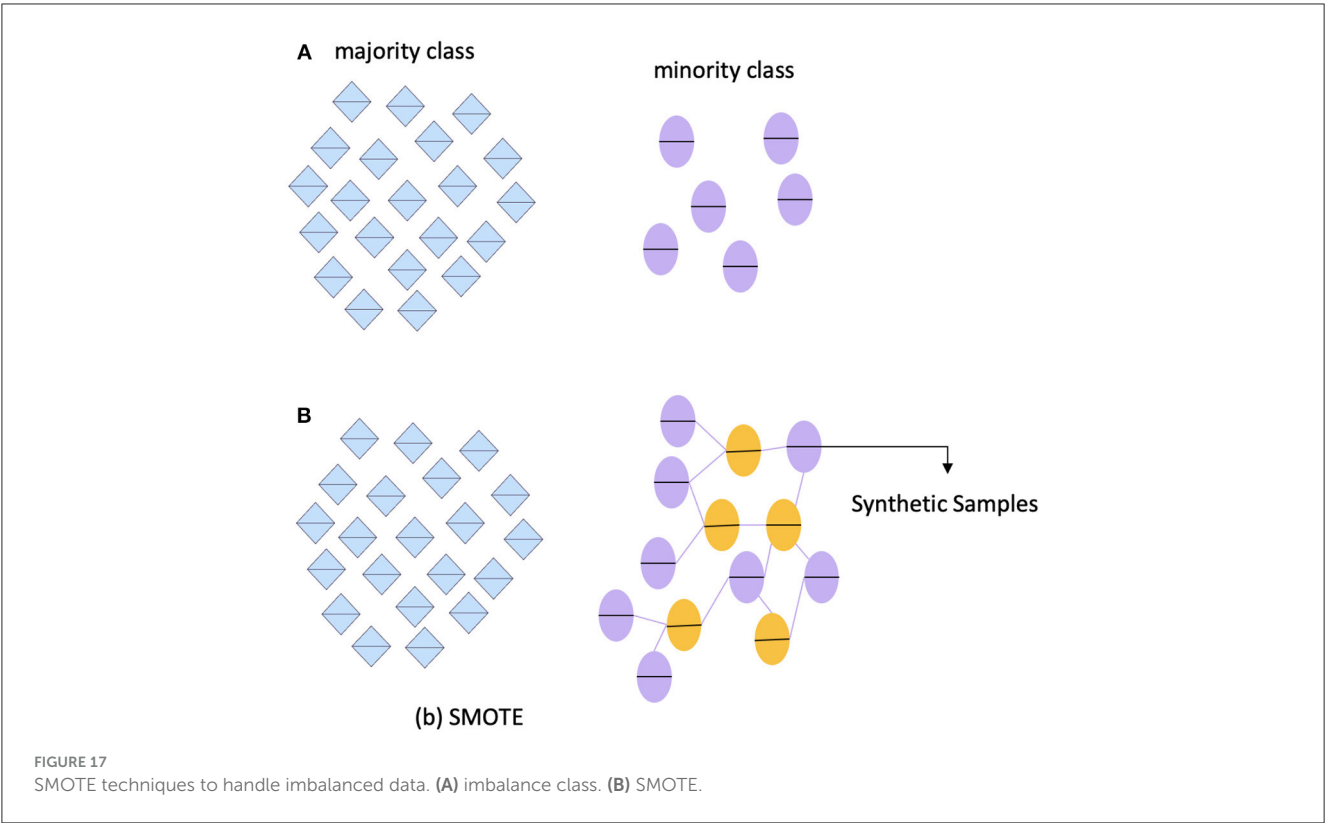


TABLE 4 Different between OASIS and ADNI datasets.

	OASIS	ADNI
Availability	Openly available to the scientific community	available to researchers, but access is subject to data use agreements and restrictions.
Size and Scope	Relatively smaller dataset with fewer subjects and a narrower focus multiple	larger with data from sites and a broader range of assessments.
Cost and resources	Require less computational resources.	Require additional resources expertise due to its comprehensive nature and complex data structure.

as it can affect accuracy. To do this, we utilized the Synthetic Minority Oversampling Technique (SMOTE) (78). One of the most common techniques used to tackle imbalanced data is SMOTE. This technique involves several steps, including identifying the minority classes, selecting their instances, finding the nearest neighbors, and creating synthetic samples. To ensure an equitable representation of participants, the minority class, known as "MCI," was oversampled in this framework. It is important to note that the dataset has no missing or null values, eliminating the need for data imputation or removal. Figure 17 outlines the step-by-step process for predicting AD.

Table 4 shows the main differences between the OASIS and ADNI datasets are as follows: OASIS provides open access to a diverse population but with less comprehensive data, while ADNI provides extensive data

and standardized protocols but with restricted access and a more homogeneous population.

5.2 Evaluation analysis

The trained model's performance is measured using evaluation metrics, with each implementation having a different preprocessing and classifier training (79).

5.2.1 K-fold cross-validation

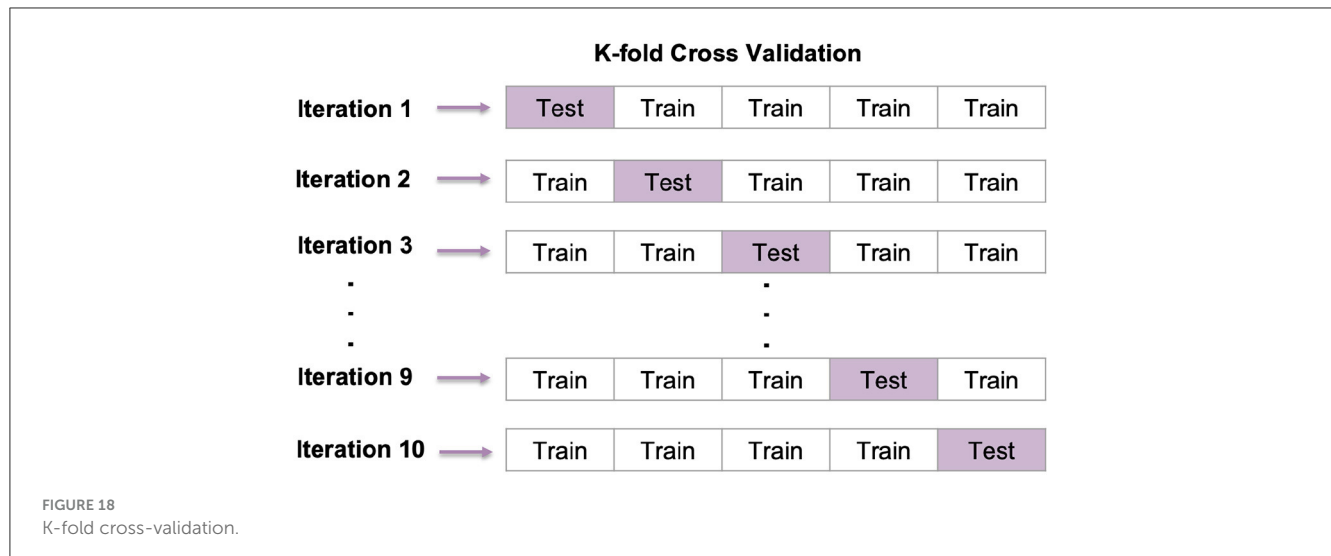
Cross-validation is a widely used method in ML for evaluating how well a model can make predictions. This method is easy to understand and helps reduce bias during evaluation. We have used 10-fold cross-validation for each configuration created by combining available values (80), as shown in Figure 18.

5.2.2 Classification metrics

This section presents the experimental results, focusing on precision, recall, F1-score, and accuracy metrics used for disease classification. These metrics are calculated based on confusion metrics, and accordingly, the performance metrics were identified in Equation (5) (81):

$$Precision = TP / (TP + FP)$$

(5)



Recall, also referred to as sensitivity (SN), is defined as the capability to identify AD patients. The formula is given in Equation (6):

$$\text{Recall} = TP / (TP + FN) \quad (6)$$

F1- score is a degree of the accuracy of the test, which reflects both the recall and precision of the test to calculate the score. The given formula is for the F1- score in Equation (7):

$$\text{F1-score} = 2TP / (2TP + FP + FN) \quad (7)$$

Accuracy (ACC) is the likelihood of correct positive and negative forecasts, as shown in Equation (8).

$$\text{ACC} = TP + TN / TP + TN + FP + FN \quad (8)$$

Where the parameters TP, FP, TN, and FN are defined as follows:

- True positive (TP): The subject has AD, and categorization outcome is positive (AD).
- False positive (FP): A subject has NC, and categorization outcome is positive.
- True negative (TN): A subject has NC, and categorization outcome is negative (Normal).
- False negative (FN): The person has AD, but the test is negative.

6 Comparison with previous studies

Based on the related study presented in Table 5, it was observed that most studies depend on a single model for ML and DL to classify AD. Moreover, most of the studies used a small dataset and the AAL-90 atlas to define the nodes (regions) of the brain. Thus, it is necessary to improve a model in various ways to extract and select essential features. Our study used MVPA for fMRI scans from the

OASIS-3 and ADNI datasets to extract activation and connectivity patterns.

7 Conclusion

AD is referred to as a neurodegenerative disease that worsens gradually and irreversibly over time. In this article, we proposed a framework to compute FC through MVPA. The fMRI data are relatively complex, with numerous voxels representing different brain regions in 3D space. We used LASSO to select a subset of relevant voxels for a specific analysis to simplify the analysis and focus on the most critical voxels.

We defined the ROIs or brain areas to analyze FC. These ROIs are often selected based on previous knowledge or hypotheses. Moreover, time-series data were extracted from these ROIs. For MVPA, the activity across multiple voxel patterns is crucial. Each data point represents the activity pattern in a specific ROI for a given task, and a correlation matrix of fMRI data is then computed. We applied our framework to single and HML algorithms to classify AD stages based on the activity patterns within ROIs. Our method surpasses state-of-the-art techniques in identifying AD, MCI, and NC in the experimental results.

Medical image classification is a crucial issue in computer science that has been extensively studied over recent decades. While significant improvement has been made in the reliability of various methods, they may need to provide accurate results due to their limitations in terms of universality, susceptibility to illumination effects, and the inadequacy of data quality, resulting in poor accuracy. We have many dimensions, few data points for each scan, and the training sample in fMRI. Additionally, trades between having enough non-redundant features to capture and not having too many noise features lead to overfitting on our data; so it is hard to distinguish between a noise and a signal accurately. In addition, we applied AAL3 to extract the ROI that includes 170 regions, but in preprocessed and defined again, it only used 166 regions, and some region skipped. Finally, the variety of public datasets is not that wide. Additionally, we posed the problem as an fMRI scan in all

TABLE 5 Comparison between our study and similar studies in the fMRI literature.

References	Perf	Data	Model	Type of class			
				AD/MCI	AD/NC	MCI/NC	AD/MCI/NC
Lama and Kwon (28)	Acc		LSVM	-	-	-	96.11%
	f1-Sc		-	-	-	-	97.3%
	Rec	ADNI	-		-	-	95.03%
	Prec			-	-	-	97.18%
Parmar et al. (29)	Acc			-	94.58%	-	-
	f1-Sc				94.82%	-	-
	Rec	ADNI	3D-CNN	-	95.2%	-	-
	Prec			-	94.44%	-	-
Guo and Zhang (30)	Acc			-	-	-	-
	f1-sco			-	-	-	-
	Rec	ADNI-2	AE	-	94.6%	-	-
	Pre			-	96.7%	-	-
Shahparian et al. (32)	Acc			98.26%	97.51%	-	-
	f1-sco			98.9%	98.28%	-	-
	Rec	ADNI	SVM	97.83%	100%	-	-
	Pre			100%	96.63%	-	-
Wang and Lim (34)	Acc			72.7%	84.8 %	97.7%	-
	f1-sco			-	-	-	-
	Rec	ADNI	ZNN	-	-	-	-
	Pre			-	-	-	-
Jiao et al. (37)	Acc			-	-	91.13%	-
	f1-sco			-	-	-	-
	Rec	ADNI-2	SVM	-	-	93.17%	-
	Pre			-	-	87.92%	-
Yang et al. (39)	Acc			-	-	91.13%	-
	f1-sco			-	-	-	-
	Rec	ADNI-2	SVM	-	-	93.17%	-
	Pre			-	-	87.92%	-
Begum and Selvaraj (44)	Acc			97.52%	97.53%	-	-
	f1-sco			97.14%	98.46%	-	-
	Rec	ADNI	3D-DCNN	90.48%	95.42%	-	-
	Pre			94.98%	97.98%	-	-
Our work	Acc			83.3%	93.18%	92.15%	87.79%
	f1-sco			83.44%	93.28%	92.44%	88.64%
	Rec	ADNI	SVM	83.30%	93.28%	92.15%	87.79%
	Pre			83.3%	94.29%	93.99%	92.60%
	Acc			84%	88.71%	89.57%	88.28%
	f1-sco			84.21%	88.83%	90%	88.41%
	Rec	ADNI	KNN	84%	88.71%	89.57%	88.28%
	Pre			88%	89.96%	89.96%	90.34%

(Continued)

TABLE 5 (Continued)

References	Perf	Data	Model	Type of class			
				AD/MCI	AD/NC	MCI/NC	AD/MCI/NC
	Acc			87.92%	89.98%	80%	91%
	f1-sco			87.94%	89.52%	81.19%	91%
	Rec	ADNI	Adaboost	87.92%	89.98%	80%	89%
	Pre			88.53%	91.39%	88.38%	92%
	Acc			94.93%	90.94%	96.15%	93.96%
	f1-sco			94.98%	91%	96.23%	94%
	Rec	ADNI	*HML	94.93%	90.94%	96.15%	93.96%
	Pre			95.61%	93%	96.65%	94.78%
	Acc			86.97%	81.3%	90%	88%
	f1-sco			86.49%	80.49%	90.51%	88%
	Rec	OASIS	SVM	86.97%	81.33%	90%	85%
	Pre			90.79%	89 %	90%	91%
	Acc			86.4%	86%	89%	84%
	f1-sco			86.26%	86.8%	90.5%	84%
	Rec	OASIS	KNN	86.47%	86.7%	89%	77%
	Pre			87.58%	89.8%	89.8%	92%
	Acc			82.94%	86.16%	89.94%	87%
	f1-sco			82.9%	86%	82.9%	87%
	Rec	OASIS	Adaboost	82.94%	86.1%	89.94%	86%
	Pre			82.93%	89.75%	89.30%	88%
	Acc			95.47%	95.11%	93.5%	92%
	f1-sco			95.50%	95.13%	93.54%	92%
	Rec	OASIS	HML	95.47%	95.11%	93.49%	93%
	Pre			96.22%	95.83%	93.34%	92%

* Hybrid machine learning. Bold values indicate highest score.

of our experiments. The main obstacle remains the intricate nature of the data and the restricted sample size within the existing dataset.

In the future, we intend to improve early detection performance by employing advanced AI methods such as explainable AI (XAI), to provide explainable results, in addition to label predictions. Moreover, we will extend the framework to track different disease modalities, such as PET and MRI. Moreover, we aim to increase the number of stages to include all the stages of AD, such as EMCI and LMCI.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

BA: Conceptualization, Formal analysis, Project administration, Validation, Writing - review &

editing. MA: Conceptualization, Data curation, Methodology, Software, Visualization, Writing - original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors extend their appreciation to the Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number: INTS065.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Schaefer M, Michelin L, Kepner J. Naming schema for a human brain-scale neural network. *arXiv Preprint*. (2021). doi: 10.48550/arXiv.2109.10951
- Herculano-Houzel S. The human brain in numbers: a linearly scaled-up primate brain. *Front Hum Neurosci*. (2009) 2009:31. doi: 10.3389/neuro.09.031.2009
- Bird TD. *Alzheimer Disease Overview*. GeneReviews (2018).
- Apostolova LG, Thompson PM. Mapping progressive brain structural changes in early Alzheimer's disease and mild cognitive impairment. *Neuropsychologia*. (2008) 46:1597–612. doi: 10.1016/j.neuropsychologia.2007.10.026
- Weiner MW, Aisen PS, Jack Jr CR, Jagust WJ, Trojanowski JQ, Shaw L, et al. The Alzheimer's disease neuroimaging initiative: progress report and future plans. *Alzheimer's Dement*. (2010) 6:202–11. doi: 10.1016/j.jalz.2010.03.007
- Carr MW, Grey ML. Magnetic Resonance Imaging: overview, risks, and safety measures. *Am J Nurs*. (2002) 102:26–33. doi: 10.1097/0000446-200212000-00012
- Goedert M, Spillantini MG. A century of Alzheimer's disease. *Science*. (2006) 314:777–81. doi: 10.1126/science.1132814
- Richens JG, Lee CM, Johri S. Improving the accuracy of medical diagnosis with causal machine learning. *Nat Commun*. (2020) 11:3923. doi: 10.1038/s41467-020-17419-7
- Allsop D. Introduction to Alzheimer's disease. *Alzheimer's Dis*. (2000) 2000:1–21. doi: 10.1097/00002093-200000001-00001
- Mosconi L, Brys M, Glodzik-Sobanska L, De Santi S, Rusinek H, De Leon MJ. Early detection of Alzheimer's disease using neuroimaging. *Exp Gerontol*. (2007) 42:129–38. doi: 10.1016/j.exger.2006.05.016
- Medical Care Corporation. *Education Progression of Alzheimer's* (2011). Available online at: <https://www.mccare.com/education/alzprogression.html> (accessed June 4, 2023).
- Kim SM, Song J, Kim S, Han C, Park MH, Koh Y, et al. Identification of peripheral inflammatory markers between normal control and Alzheimer's disease. *BMC Neurol*. (2011) 11:1–6. doi: 10.1186/1471-2377-11-51
- Guo M, Li Y, Zheng W, Huang K, Zhou L, Hu X, et al. A novel conversion prediction method of MCI to AD based on longitudinal dynamic morphological features using ADNI structural MRIs. *J Neurol*. (2020) 267:2983–97. doi: 10.1007/s00415-020-09890-5
- Wang SSS, Good TA. An overview of Alzheimer's disease. *J Chin Inst Chem Eng*. (2005) 36:533–59.
- Korolev IO. Alzheimer's disease: a clinical and basic science review. *Med Stud Res J*. (2014) 4:24–33.
- Arthurs OJ, Boniface S. How well do we understand the neural origins of the fMRI BOLD signal? *Trends Neurosci*. (2002) 25:27–31. doi: 10.1016/S0166-2236(00)01995-0
- Turner R, Ordidge RJ, Haacke E, Liang ZP. Technical challenges of functional magnetic resonance imaging. *IEEE Eng Med Biol Mag*. (2000) 19:42–54. doi: 10.1109/51.870231
- Buxton RB, Wong EC, Frank LR. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magnet Reson Med*. (1998) 39:855–64.
- Friston KJ, Mechelli A, Turner R, Price CJ. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. *NeuroImage*. (2000) 12:466–77. doi: 10.1006/nimg.2000.0630
- Kennan RP, Zhong J, Gore JC. Intravascular susceptibility contrast mechanisms in tissues. *Magnet Reson Med*. (1994) 31:9–21.
- Ogawa S, Menon R, Tank DW, Kim S, Merkle H, Ellermann J, et al. Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model. *Biophys J*. (1993) 64:803–12.
- Frahm J, Merboldt KD, Hänicke W, Kleinschmidt A, Boecker H. Brain or veins oxygenation or flow? On signal physiology in functional MRI of human brain activation. *NMR Biomed*. (1994) 7:45–53.
- Menon RS, Kim SG. Spatial and temporal limits in cognitive neuroimaging with fMRI. *Trends Cogn Sci*. (1999) 3:207–16.
- Boxerman JL, Bandettini PA, Kwong KK, Baker JR, Davis TL, Rosen BR, et al. The intravascular contribution to fMRI signal change: Monte Carlo modeling and diffusion-weighted studies in vivo. *Magnet Reson Med*. (1995) 34:4–10.
- Lee SP, Silva AC, Ugurbil K, Kim SG. Diffusion-weighted spin-echo fMRI at 9.4 T: microvascular/tissue contribution to BOLD signal changes. *Magnet Reson Med*. (1999) 42:919–28.
- Menon RS, Thomas CG, Gati JS. Investigation of BOLD contrast in fMRI using multi-shot EPI. *NMR Biomed*. (1997) 10:179–82.
- Warren SL, Moustafa AA. Functional magnetic resonance imaging, deep learning, and Alzheimer's disease: a systematic review. *J Neuroimag*. (2023) 33:5–18. doi: 10.1111/jon.13063
- Lama RK, Kwon GR. Diagnosis of Alzheimer's disease using brain network. *Front Neurosci*. (2021) 15:605115. doi: 10.3389/fnins.2021.605115
- Parmer HS, Nutter B, Long R, Antani S, Mitra S. Deep learning of volumetric 3D CNN for fMRI in Alzheimer's disease classification. *Med Imag 2020 Biomed Appl Mol Struct Funct Imag*. (2020) 11317:66–71. doi: 10.1117/12.2549038
- Guo H, Zhang Y. Resting state fMRI and improved deep learning algorithm for earlier detection of Alzheimer's disease. *IEEE Access*. (2020) 8:115383–92. doi: 10.1109/ACCESS.2020.3003424
- Alarjani M. "Detect Alzheimer's disease from neuroimaging using artificial intelligence techniques," in *2023 3rd International Conference on Computing and Information Technology (ICCIIT)*. Tabuk: IEEE. (2023). p. 192–200.
- Shahparian N, Yazdi M, Khosravi MR. Alzheimer disease diagnosis from fMRI images based on latent low rank features and support vector machine (SVM). *Curr Sign Transduct Ther*. (2021) 16:171–7. doi: 10.2174/1574362414666191202144116
- Castellazzi G, Cuzzoni MG, Cotta Ramusino M, Martinelli D, Denaro F, Ricciardi A, et al. A machine learning approach for the differential diagnosis of Alzheimer and vascular dementia fed by MRI selected features. *Front Neuroinformat*. (2020) 14:25. doi: 10.3389/fninf.2020.00025
- Wang B, Lim JS. Zoom-in neural network deep-learning model for Alzheimer's disease assessments. *Sensors*. (2022) 22:8887. doi: 10.3390/s22228887
- Zamani J, Sadr A, Javadi AH. Classification of early-MCI patients from healthy controls using evolutionary optimization of graph measures of resting-state fMRI, for the Alzheimer's disease neuroimaging initiative. *PLoS ONE*. (2022) 17:e0267608. doi: 10.1371/journal.pone.0267608
- Nguyen DT, Ryu S, Qureshi MNI, Choi M, Lee KH, Lee B. Hybrid multivariate pattern analysis combined with extreme learning machine for Alzheimer's dementia diagnosis using multi-measure rs-fMRI spatial patterns. *PLoS ONE*. (2019) 14:e0212582. doi: 10.1371/journal.pone.0212582
- Jiao Z, Xia Z, Ming X, Cheng C, Wang SH. Multi-scale feature combination of brain functional network for eMCI classification. *IEEE Access*. (2019) 7:74263–73. doi: 10.1109/ACCESS.2019.2920978
- Lu J, Zeng W, Zhang L, Shi Y, et al. Diagnosis of Alzheimer's disease with extreme learning machine on whole-brain functional connectivity. *Concept Magnet Reson B Magnet Reson Eng*. (2022) 2022:1047616. doi: 10.1155/2022/1047616
- Yang P, Zhou F, Ni D, Xu Y, Chen S, Wang T, et al. Fused sparse network learning for longitudinal analysis of mild cognitive impairment. *IEEE Trans Cybernet*. (2019) 51:233–46. doi: 10.1109/TCYB.2019.2940526
- Chan YH, Ang JL, Gupta S, He Y, Rajapakse JC. Subtype-specific biomarkers of Alzheimer's disease from anatomical and functional connectomes via Graph Neural Networks. In: *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Seoul: IEEE (2024). p. 2195–9.
- Lama RK, Kim JI, Kwon GR. Classification of Alzheimer's disease based on core-large scale brain network using multilayer extreme learning machine. *Mathematics*. (2022) 10:1967. doi: 10.3390/math10121967
- Koluragi P, Deshpande Y, Dhumale G, Chikkamath S, Nirmala S, Budihal SV. Multi-model approach for Alzheimer's disease detection and classification. In: *2024 3rd International Conference for Innovation in Technology (INOCON)*. Indai: IEEE (2024). p. 1–5.
- Sethuraman SK, Malaiyappan N, Ramalingam R, Basheer S, Rashid M, Ahmad N. Predicting Alzheimer's disease using deep neuro-functional networks with resting-state fMRI. *Electronics*. (2023) 12:1031. doi: 10.3390/electronics12041031

44. Begum AP, Selvaraj P. Alzheimer's disease classification and detection by using AD-3D DCNN model. *Bullet Electr Eng Informat.* (2023) 12:882–90. doi: 10.11591/eei.v12i2.4446
45. Chauhan N, Choi BJ. Classification of Alzheimer's disease using maximal information coefficient-based functional connectivity with an extreme learning machine. *Brain Sci.* (2023) 13:1046. doi: 10.3390/brainsci13071046
46. Penalba-Sánchez L, Oliveira-Silva P, Sumich AL, Cifre I. Increased functional connectivity patterns in mild Alzheimer's disease: a rsfMRI study. *Front Aging Neurosci.* (2023) 14:1037347. doi: 10.3389/fnagi.2022.1037347
47. Nieto-Castanon A. *Handbook of Functional Connectivity Magnetic Resonance Imaging Methods in CONN.* Boston, MA: Hilbert Press (2020).
48. Andersson JL, Hutton C, Ashburner J, Turner R, Friston K. Modeling geometric deformations in EPI time series. *Neuroimage.* (2001) 13:903–19. doi: 10.1006/nimg.2001.0746
49. Friston KJ, Ashburner J, Frith CD, Poline JB, Heather JD, Frackowiak RS. Spatial registration and normalization of images. *Hum Brain Map.* (1995) 3:165–89.
50. Ashburner J. A fast diffeomorphic image registration algorithm. *Neuroimage.* (2007) 38:95–113. doi: 10.1016/j.neuroimage.2007.07.007
51. Ashburner J, Friston KJ. Unified segmentation. *Neuroimage.* (2005) 26:839–51. doi: 10.1016/j.neuroimage.2005.02.018
52. Margulies DS, Böttger J, Long X, Lv Y, Kelly C, Schäfer A, et al. Resting developments: a review of fMRI post-processing methodologies for spontaneous brain activity. *Magnet Reson Mater Phys Biol Med.* (2010) 23:289–307. doi: 10.1007/s10334-010-0228-5
53. Liu Y, Yu C, Zhang X, Liu J, Duan Y, Alexander-Bloch AF, et al. Impaired long distance functional connectivity and weighted network architecture in Alzheimer's disease. *Cerebr Cortex.* (2014) 24:1422–35. doi: 10.1093/cercor/bhs410
54. Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, et al. Dissociable intrinsic connectivity networks for salience processing and executive control. *J Neurosci.* (2007) 27:2349–56. doi: 10.1523/JNEUROSCI.5587-06.2007
55. Chaovalitwongse WA, Won D, Seref O, Borghesani P, Askren MK, Willis S, et al. Network optimization of functional connectivity within default mode network regions to detect cognitive decline. *IEEE Trans Neural Syst Rehabil Eng.* (2017) 25:1079–89. doi: 10.1109/TNSRE.2017.2679056
56. Sheline YI, Raichle ME. Resting state functional connectivity in preclinical Alzheimer's disease. *Biol Psychiat.* (2013) 74:340–7. doi: 10.1016/j.biopsych.2012.11.028
57. Koch W, Teipel S, Mueller S, Benninghoff J, Wagner M, Bokde AL, et al. Diagnostic power of default mode network resting state fMRI in the detection of Alzheimer's disease. *Neurobiol Aging.* (2012) 33:466–78. doi: 10.1016/j.neurobiolaging.2010.04.013
58. Haynes JD, Rees G. Decoding mental states from brain activity in humans. *Nat Rev Neurosci.* (2006) 7:523–34. doi: 10.1038/nrn1931
59. Raizada RD, Kriegeskorte N. Pattern-information fMRI: new questions which it opens up and challenges which face it. *Int J Imag Syst Technol.* (2010) 20:31–41. doi: 10.1002/ima.20225
60. Yang Z, Fang F, Weng X. Recent developments in multivariate pattern analysis for functional MRI. *Neurosci Bullet.* (2012) 28:399–408. doi: 10.1007/s12264-012-1253-3
61. Moeller JR, Habeck CG, et al. Reciprocal benefits of mass-univariate and multivariate modeling in brain mapping: applications to event-related functional MRI, H(2)(15)O-, and fdg-pet. *Int J Biomed Imag.* (2006) 2006:79862. doi: 10.1155/IJBI/2006/79862
62. Bu X, Hu X, Zhang L, Li B, Zhou M, Lu L, et al. Investigating the predictive value of different resting-state functional MRI parameters in obsessive-compulsive disorder. *Transl Psychiat.* (2019) 9:17. doi: 10.1038/s41398-018-0362-9
63. Mur M, Bandettini PA, Kriegeskorte N. Revealing representational content with pattern-information fMRI—an introductory guide. *Soc Cogn Affect Neurosci.* (2009) 4:101–9. doi: 10.1093/scan/nsn044
64. Lewis-Peacock JA, Norman KA. Multi-voxel pattern analysis of fMRI data. *Cogn Neurosci.* (2014) 512:911–20. doi: 10.7551/mitpress/9504.003.0099
65. Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci.* (2006) 10:424–30. doi: 10.1016/j.tics.2006.07.005
66. Yoon JH, Tamir D, Minzenberg MJ, Ragland JD, Ursu S, Carter CS. Multivariate pattern analysis of functional magnetic resonance imaging data reveals deficits in distributed representations in schizophrenia. *Biol Psychiat.* (2008) 64:1035–41. doi: 10.1016/j.biopsych.2008.07.025
67. Lee IS, Jung Wm, Park HJ, Chae Y. Spatial information of somatosensory stimuli in the brain: multivariate pattern analysis of functional magnetic resonance imaging data. *Neural Plast.* (2020) 2020:8307580. doi: 10.1155/2020/8307580
68. Rolls ET, Huang CC, Lin CP, Feng J, Joliot M. Automated anatomical labelling atlas 3. *Neuroimage.* (2020) 206:116189. doi: 10.1016/j.neuroimage.2019.116189
69. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage.* (2002) 15:273–89. doi: 10.1006/nimg.2001.0978
70. Rolls ET, Joliot M, Tzourio-Mazoyer N. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *Neuroimage.* (2015) 122:1–5. doi: 10.1016/j.neuroimage.2015.07.075
71. Liang P, Shi L, Chen N, Luo Y, Wang X, Liu K, et al. Construction of brain atlases based on a multi-center MRI dataset of 2020 Chinese adults. *Sci Rep.* (2015) 5:18216. doi: 10.1038/srep18216
72. Kukreja SL, Löfberg J, Brenner MJ. A least absolute shrinkage and selection operator (LASSO) for nonlinear system identification. *IFAC Proc.* (2006) 39:814–9. doi: 10.3182/20060329-3-AU-2901.00128
73. Ayodele TO. Machine learning overview. *N Adv Machine Learn.* (2010) 2:9–18.
74. Kramer O, Kramer O. K-nearest neighbors. *Dimensional Reduct Unsupervised Near Neighbors.* (2013) 2:13–23. doi: 10.1007/978-3-642-38652-7_2
75. Schapire RE. The boosting approach to machine learning: an overview. *Nonlin Estim Classif.* (2003) 9:149–71. doi: 10.1007/978-0-387-21579-2_9
76. Marcus DS, Fotenos AF, Csernansky JG, Morris JC, Buckner RL. Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. *J Cogn Neurosci.* (2010) 22:2677–84. doi: 10.1162/jocn.2009.21407
77. IDA. *Alzheimer's Disease Neuroimaging Initiative* (2022) Available online at: <https://ida.loni.usc.edu> (accessed November 01, 2022).
78. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* (2002) 16:321–57. doi: 10.1613/jair.953
79. Zhou J, Gandomi AH, Chen F, Holzinger A. Evaluating the quality of machine learning explanations: a survey on methods and metrics. *Electronics.* (2021) 10:593. doi: 10.3390/electronics10050593
80. Fushiki T. Estimation of prediction error by using K-fold cross-validation. *Stat Comput.* (2011) 21:137–46. doi: 10.1007/s11222-009-9153-8
81. Dalianis H, Dalianis H. Evaluation metrics and evaluation. *Clin Text Min.* (2018) 6:45–53. doi: 10.1007/978-3-319-78503-5_6



OPEN ACCESS

EDITED BY

Hikmat A. M. Abdeljaber,
Applied Science Private University, Jordan

REVIEWED BY

Zia Ush Shamszaman,
Teesside University, United Kingdom
Mazen Al-Asali,
Cairo University, Egypt

*CORRESPONDENCE

Abdu H. Gumaei

✉ a.gumaei@psau.edu.sa

Mogeeb A. A. Mosleh

✉ mogeebmoleh@taiz.edu.ye

RECEIVED 24 March 2024

ACCEPTED 24 June 2024

PUBLISHED 24 July 2024

CITATION

Khurshid D, Wahid F, Ali S,
Gumaei AH, Alzanin SM and
Mosleh MAA (2024) A deep neural
network-based approach for seizure activity
recognition of epilepsy sufferers.
Front. Med. 11:1405848.
doi: 10.3389/fmed.2024.1405848

COPYRIGHT

© 2024 Khurshid, Wahid, Ali, Gumaei, Alzanin
and Mosleh. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

A deep neural network-based approach for seizure activity recognition of epilepsy sufferers

Danial Khurshid¹, Fazli Wahid^{2,3}, Sikandar Ali¹,
Abdu H. Gumaei^{4*}, Samah M. Alzanin⁴ and
Mogeeb A. A. Mosleh^{5,6*}

¹Department of Information Technology, The University of Haripur, Haripur, Khyber Pakhtunkhwa, Pakistan, ²College of Science and Engineering, School of Computing, University of Derby, Derby, United Kingdom, ³School of Computing Sciences, University of East Anglia, Norwich, United Kingdom, ⁴Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia, ⁵Faculty of Engineering and Information Technology, Taiz University, Taiz, Yemen, ⁶International University of Technology Twintech, Sana'a, Yemen

Epilepsy is one of the most frequent neurological illnesses caused by epileptic seizures and the second most prevalent neurological ailment after stroke, affecting millions of people worldwide. People with epileptic disease are considered a category of people with disabilities. It significantly impairs a person's capacity to perform daily tasks, especially those requiring focusing or remembering. Electroencephalogram (EEG) signals are commonly used to diagnose people with epilepsy. However, it is tedious, time-consuming, and subjected to human errors. Several machine learning techniques have been applied to recognize epilepsy previously, but they have some limitations. This study proposes a deep neural network (DNN) machine learning model to determine the existing limitations of previous studies by improving the recognition efficiency of epileptic disease. A public dataset is used in this study and classified into training and testing sets. Experiments were performed to evaluate the DNN model with different dataset classification ratios (80:20), (70:30), (60:40), and (50:50) for training and testing, respectively. Results were evaluated by using different performance metrics including validations, and comparison processes that allow the assessment of the model's effectiveness. The experimental results showed that the overall efficiency of the proposed model is the highest compared with previous works, with an accuracy rate of 97%. Thus, this study is more accurate and efficient than the existing seizure detection approaches. DNN model has great potential for recognizing epileptic patient activity using a numerical EEG dataset offering a data-driven approach to improve the accuracy and reliability of seizure detection systems for the betterment of patient care and management of epilepsy.

KEYWORDS

deep learning, deep neural network, electroencephalogram, epilepsy disability, epilepsy detection, seizure activity recognition

1 Introduction

Epilepsy is a prevalent neurological condition that affects millions of people worldwide. It is considered a kind of disability, where epileptic patients are considered a category of people with disabilities. Different techniques are used to detect c activities and their shortcomings.

EEG is the manual way of diagnosing seizures by pinning many electrodes everywhere on the head, making it difficult to pinpoint where the electrical activity in the brain originates. Additionally, medical professionals' reading of EEG signals is slow, time-consuming, and subject to human mistakes during the diagnosis process. Machine learning techniques are also used to identify epileptic seizures. Different methods have been adapted for epilepsy detection, such as CNN, K-NN, Naïve Bayes, and DWT, and briefly discussed in the literature of the study. However, most existing state-of-the-art methods are considered complex, time-consuming, and suffer from some limitations in terms of accuracy performance.

Early epilepsy detection can help society, health sectors, and medical specialists. Human activity recognition (HAR) is the automatic detection of numerous physical actions people perform daily. It is used to identify the actions that are carried out by a person, given a set of observations of themselves and the nearby environment. Activity recognition can be attained by exploiting the information retrieved from various sources, such as environmental (1) or body-worn sensors (2). Multiple approaches have adapted dedicated motion sensors in different body parts such as the wrist, waist, and chest. These sensors are primarily uncomfortable for users and do not provide long-term results for activity monitoring, e.g., sensor repositioning after dressing (3). A HAR system aids in the recognition of a person's activities and the provision of intervention responses. Most activities that keep track of everyday fitness exercises, such as walking, jogging, walking upstairs, and walking downstairs, are done daily. Taking phone calls, sweeping, making food, combing hair, washing hands, brushing teeth, wearing coats and shoes, and writing and reading are all tasks that everyone does daily. Also growing demand for wearable devices with sensing abilities (smart watches, intelligent bands) used to take out important information (4). Figure 1 shows some of the daily activities of human life.

Through wearable devices, human activity recognition (HAR) is currently considered an essential tool for health care in the future. Tracking patient activities not only helps medical professionals to

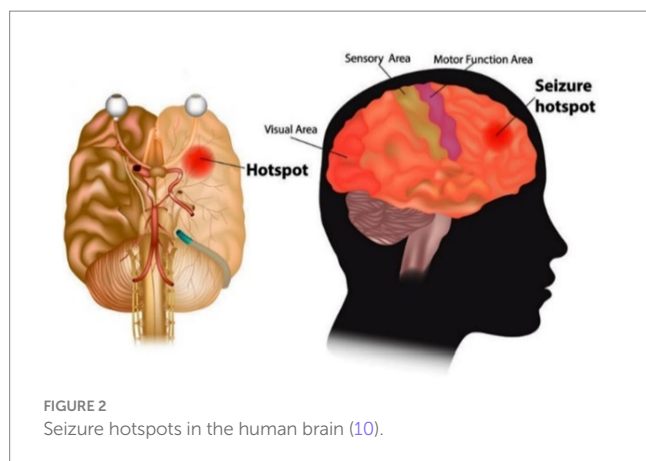
provide hospital care services to patients across any distance with the latest technology of communication and information but also provides facilities for patients to be monitored online (6). The advantages are the prevention of hospitalization, the cost, and improving human health. Patient activity recognition PAR includes monitoring Vital Data (VD) such as blood pressure, pulse, and blood glucose (7).

Different sensors are used to monitor various activities to improve patients' health. The developments in wearable and cell phone devices have made it possible to gather information from built-in smartphones and health trackers, including microphones, magnetometers, gyroscopes, GPS, and accelerometers. An epileptic seizure is a usual neurological disorder that happens because of unexpected discharge of neurons of the brain and stress influence. It is a condition distinguished by repeated (two or more) epileptic seizures. A single event is considered as numerous seizures occurring within a 24-h time or an episode of status epilepticus (SE). It is one of the world's oldest conditions of humankind, and still, it is the most typical neurological condition that affects people of all ages. About 50 million people worldwide have a diagnosis of epilepsy (8). A clinical device, an electroencephalogram (EEG) signal, plays a vital role in diagnosing epilepsy. It gives a photograph of the human brain while doing a cognitive task or even resting. The EEG is gathered by putting electrodes on the patient's scalp. Then, electro-activity is recorded, produced by the brain, and can identify epilepsy, but this method for examining an EEG signal for epileptic seizure recognition is time-consuming (9). Figure 2 visualizes the hotspot of seizure in the human brain.

Machine learning techniques have been proposed to switch this typical method. There are two fundamental stages of extraction and classification of data involved in machine learning. The traditional system of consulting doctors is time-consuming and more costly, also leading to fatigue-based diagnostic mistakes and subject to the absence of diagnostic facilities in regions of the world where physicians are not available. Recently, machine learning methods have been capable of attaining skilled-level performance in health care and the medical field (11). Different deep learning approaches are used to



FIGURE 1
Human activities of daily life (5).



detect seizures, like support vector machines, convolutional neural networks, and deep convolutional neural networks. Still, these techniques use complex algorithms and image data extracted from EEG.

There are several reasons behind the development of a Deep Neural Network-based method for identifying seizure activity in epilepsy patients. First and foremost, it tackles the pressing issue of prompt and accurate seizure detection, which is necessary for both patient care and efficient treatment. Furthermore, improvements in machine learning—especially in deep learning—present the possibility of very precise pattern identification in EEG data, which might improve detection rates. This strategy also seeks to enhance the quality of life for individuals with epilepsy by facilitating more targeted intervention techniques, which in turn lowers the frequency and intensity of seizures.

This study used a deep neural network (DNN) based model to recognize seizures. Since patterns of EEG seizures differ significantly between patients, it is challenging to recognize seizures. Thus, most of the automated methods that will be discussed in the literature review use complex algorithms and substantial image data sets, which is time-consuming and inefficient. The focus is on creating a model to swiftly and accurately detect epilepsy. Our main aim is to develop a fast and precise system. Through thorough testing and training, we aim to achieve high accuracy while also considering speed. Ultimately, our goal is to improve epilepsy diagnosis, potentially benefiting patients with better and faster care.

2 Literature review

Many deep learning and machine learning methods and algorithms are used for the detection of human activities, patient activities, and epileptic patient activities. In this section, some previous work that has been done recently will be discussed. Hassan et al. (12) proposed research on a smartphone inertial sensors-based approach for HAR. Effectual attributes are first taken out via raw data. The attributes contain median, mean, autoregressive coefficients, etc. The attributes are processed through a linear discriminant examination and kernel principal component analysis (KPCA) and (LDA) to make them extra robust. Lastly, the attributes are trained by a Deep Belief Network (DBN) for effective activity detection. The system comprises

three central portions: sensing, attribute recognition, and extraction. The sensing part collects the sensor's information as input to the HAR system. Attribute extraction removes noise to isolate signals. Finally, where DBN is used, a key aspect is modeling actions from attributes via deep learning with an overall accuracy of 95.85%.

Gul et al. (13) researched abnormal human activity recognition as a Tool for Patient Monitoring. The You Look Only Once (YOLO) network, which is based on CNN architecture, is used as a backbone CNN model. To train the CNN model, a large dataset of patient films is constructed by labeling each frame with the positions and behaviors of the patient. For 32 epochs, a CNN model with 23,040 tagged photos of the patient's actions was used. The model assigned a unique action label and a confidence score for video orders by identifying the recurring action label in each frame. The study found that aberrant action recognition is 96.8% accurate. For patient nursing, the proposed framework can benefit hospitals and elder care homes. Murad et al. (14) performed a study on deep recurrent neural networks (DRNN) and built a model that can capture distant dependencies in variable-length input arrangements. The model has bidirectional, unidirectional, and cascaded structural design, which is built on long short-term memory (LSTM). The approach exceeds other modern methods because it is capable of taking out more particular attributes via deep layers in end-to-end and task-dependent fashion and has an overall accuracy of 96.7%. Uddin et al. (15) performed research on Activity Recognition for Cognitive Assistance Using Body based sensor data and Deep Convolutional Neural Networks in which signals are examined from body wearable sensors for Medicare like gyroscope, ECG, accelerometer, and magnetometer sensors. The deep CNN is trained once attributes are extracted from sensor data using Gaussian kernel-based PCA and Z-score normalization. Lastly, trained deep CNN is utilized to detect activities in examining data. The method provides cognitive aid in wearable sensor-based intelligent medical care systems. The proposed method has an average accuracy of 93.90%.

Ouichka et al. (16) conducted research on prediction of seizures using DNN methods. In which five models (1-CNN, 2-CNN, 3-CNN, 4-CNN, and Transfer learning with ResNet50) for the prediction of epileptic seizures were proposed. The findings show that both methods, one using a fusion of three CNNs (3-CNN) and the other using four CNNs (4-CNN), achieve an accuracy of 95%. Specifically, the 3-CNN method yields an accuracy of 95.0%, a recall of 94.5%, and an F1-score of 95.0%. The 4-CNN method provides an accuracy of 95.5%, a recall of 95.5%, and an F1-score of 95.0%. Ibrahim et al. (17) presented two patient-specific CNN models for prediction and detection of seizure in which spectrogram images of EEG signal segments was used. The third CNN model is designed for patient non-specific scenarios and can classify two and three EEG signal states. It operates effectively on both spectrogram and PSR images of EEG segments. Experiments showed the highest classification performance when using PSR images, due to their superior representation of EEG signals. In contrast, the first two models are suitable for patient-specific uses, but their reliance on spectrogram images somewhat restricts their performance.

Poorani et al. (18) performed a research on a one-dimensional, patient-specific scheme for detecting epilepsy seizures addresses binary classification (seizure vs. non-seizure). The 1D-CNN and CNN-LSTM models offer a computationally efficient approach by

processing EEG data through pooling and dense layers. Abderrahim et al. (19) conducted an experiment in which they introduces four models: S-CNN, Modif-CNN, CNN-SVM, and Comb-2CNN, each demonstrating high accuracy in predicting epileptic seizures. The Modif-CNN model stands out with an impressive accuracy rate of 97.96%, making the results from all models both promising and interesting.

The presented study also addresses the challenges identified and some limitations of recent studies and machine learning techniques such as many models struggle to handle EEG data in real time and need large amounts of computing power. Additional problems include handling undesired data in the EEG, individual variations in seizure patterns, and an imbalance in data classes. Specifically for other deep learning models Long-term dependency maintenance is a hurdle for RNNs, non-image dataset adaptation may be a barrier for CNNs, training and parameter optimization are issues for RL so the current model that is using to identify Epileptic activities by using multiple hidden layers that allows to learn complex patterns and data representation the depth of these layers allows to capture the complicated features resulting in enhanced performance. DNN algorithm is more efficient because of its computational complexity, deep architecture and its ability to learn complicated patterns from the data as Compare to other deep learning models.

3 Problem identification and solution

There are various methods used to detect epileptic seizures; one of the most common and manual ways is EEG, which is a very time-consuming process. Computer-aided diagnosis methods, automatic detection, deep learning, and machine learning methods exist. The conventional technique of identifying different brain disorders has been inspected manually for centuries. Still, those manual methods have some limitations, such as inaccuracy, slow diagnosing process, and various outcomes of the same inputs. Manual identification needs more resources and time. So, to achieve high accuracy and fast diagnosis, computer-aided disease detection methods have been used for the last few decades. This method will assist medical professionals in the clarification of medical imaging. Medical computer-aided diagnosis methods are limited by noise, fuzziness, and uncertainty in medical images, so such limitations may affect decisions of disease diagnosis while determining the disease type. The main idea of this research is to detect epileptic seizures using a Deep Neural Network (DNN), which is more powerful and optimistic. A simple numerical model that is built on deep learning has applications in the fields of bioinformatics, healthcare, and computer science. The personal monitoring system for the detection of epilepsy with high accuracy is becoming popular for the improvement of human life. Researchers can achieve their targeted objectives and improve their expertise through this research. In the current study, the DNN model contains several layers of neurons that build up an output layer.

4 Proposed methodology

The proposed method consists of four main stages, illustrated in Figure 3. Initially, data acquisition involves collecting the necessary

data. This is followed by the data cleaning stage, where irrelevant or redundant features are eliminated to ensure the dataset is optimized for further analysis. Once cleaned, the dataset is divided into two subsets: one for training the model and the other for testing its performance.

In the activity recognition phase, a deep neural network is employed to identify brain activities related to seizures. This involves the model learning patterns and distinguishing between different types of brain activity. Finally, in the performance evaluation phase, the model's effectiveness is assessed using various metrics. These metrics include the F1-score, which balances precision and recall, precision itself, the confusion matrix that shows the performance of the classification, accuracy indicating the proportion of correctly classified instances, and the Receiver Operating Characteristic (ROC) curve, which illustrates the true positive rate against the false positive rate across different threshold values. This structured approach ensures that each phase contributes to building a robust and reliable model for recognizing epileptic seizures, with thorough evaluation to validate its performance.

4.1 Data loading

The data of this study is publicly available and uploaded to the model for cleaning, splitting and classification. After uploading the historical data, the valuable data will be extracted, and then irregular, null, garbage, and inconsistent values will be eliminated, which may lead to many difficulties. Data cleaning removes unwanted features that do not belong to the proposed study. In the next stage, data transformation is done, in which the raw data is turned into a format or structure that is more suited for the model or algorithm.

4.2 Data splitting

The data is split into two parts: the first part of the dataset is used for training, and the remaining part will be used for data testing. The proposed model will split DNN's dataset into different training and testing ratios to achieve high accuracy.

4.2.1 Training set

The data samples are used to fit the model, and a subset of the dataset is used to train the model (in the context of neural networks, calculating weights and biases). The model sees and learns this data, allowing it to improve its parameters.

4.2.2 Test set

The data set objectively evaluates a final model's fit to the training data. It is used once the model has been adequately trained with training and validation.

4.3 Model architecture

ANN's model architecture includes the creation of layers, which are input layers, dense layers, and output layers. Each neuron

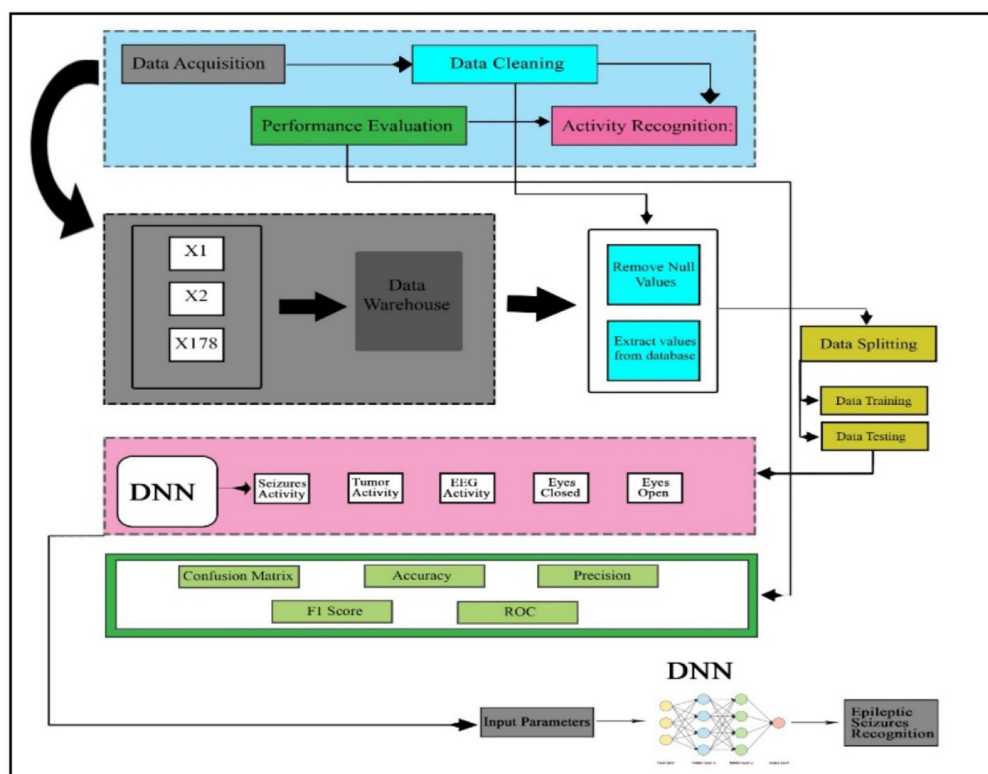


FIGURE 3
Adopted methodology.

in the dense layer receives input from all neurons in the previous layer, making it a deep-connected neural network layer. The thick layer is revealed to be the layer that is most usually utilized. The size of the input layers and output layers are also defined in this section.

4.4 Model compilation

Compilation is the last stage in the model creation process. The model will be ready to move to the training and testing phase at this stage. The model compilation uses some functions, such as the loss function, to find errors or deviations in the learning procedure. Moreover, the optimizer is used to optimize the weights of the inputs by comparing the loss function and prediction. The evaluation metrics are applied to evaluate the model's performance.

4.5 Model training

The training set consisted of sample output data and the input data sets that affect the outcome. The training model is utilized to process the input data using the algorithm to match the processed result to the sample output. NumPy arrays using the fit function are used to train models. The main aim of the fit function is to evaluate the model during the training stage (20).

4.6 Model testing

After the training model moves toward the testing phase, testing of the model is the process of analyzing a fully trained model's performance on a testing set. The testing set is a collection of samples separated from the training and validation sets, but it has the same probability distribution as the training set (21).

4.7 Model evaluation

In this stage, performance evaluation will be done to improve the system. Confusion matrix, F1-score, Precision, recall, and accuracy in a rigorously statistical manner are the parameters utilized for performance evaluation.

4.7.1 Confusion matrix

A Confusion Matrix is an $n \times n$ matrix used to assess the model's classification performance, where N represents the number of target classes. The matrix differentiates the actual values from the machine learning model's predictions. This gives us a clear picture of how efficiently our classification method works and the types of errors it generates (22).

4.7.2 Accuracy

Model accuracy is a metric for determining which model is the most effective in detecting patterns and correlations among

variables in a dataset using training or input data. The greater a model's generalization to 'unseen' data is, the more accurate insights and predictions it can deliver, and hence the additional commercial value it can provide. The accuracy of classification models is one of the factors to consider while evaluating the (23). Accuracy represents the percentage of correct predictions made by our model. Equation (1) below is the formal definition of:

$$\text{Accuracy} = \frac{\text{Number of right predictions}}{\text{Total number of predictions}} \quad (1)$$

Equation (2) below determined binary classification accuracy regarding negatives and positives.

$$\text{Accuracy} = \text{TP} + \text{TN} / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (2)$$

TP stands for True Positives, TN stands for True Negatives, FP stands for False Positives, and FN stands for False Negatives.

4.7.3 Precision

Precision is a statistic that measures the accuracy of a machine learning model's positive prediction. Precision (i.e., the total number of true positives plus the number of false positives) is the ratio of true positives to total positive predictions as shown in Equation (3) below (24).

$$\text{Precision} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Positive (FP)}} \quad (3)$$

4.7.4 Recall

The model's recall indicates how successfully it finds True Positives. As an outcome, recall tells us how many patients we correctly identified as having illness out of the total number of patients with disease (25). Mathematically shown in Equation (4) below.

$$\text{Recall} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}} \quad (4)$$

4.7.5 F1-score

The F1 score represents a balance of precision and recall. The harmonic mean of accuracy and recall is used to compute the F1 score. The harmonic mean is a measure that can be used instead of the arithmetic mean. Calculating an average rate is especially beneficial (26). The average accuracy and recall are computed using the F1-score. Because they are both rates, the harmonic mean makes sense. It is calculated using the Equation (5) below:

$$\text{F1-score} = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (5)$$

5 Experimental setup

The experiments that are done are related to epilepsy detection using deep neural networks and will be deeply discussed in this section.

5.1 System specification

The system that is used in this research is an HP Intel core i5-fourth generation Desktop with 8 GB RAM, 1.90GHz processor, and 500 GB hard drive—Windows 10 64-bit operating system. In the proposed research, Python language is used to simulate Epileptic patient activity recognition. Google COLAB is used to execute the Python code.

5.2 Dataset description

The dataset used in this study is publicly available on the KAGGLE platform at the following link: <https://www.kaggle.com/datasets/harunshimanto/epileptic-seizure-recognition>. The reference's original dataset is separated into five categories, each containing 100 files, each representing a particular subject/person. For 23.6 s, each file records brain activity. A 4097 of data points are taken from the linked time series. Each data point represents the value of the EEG recording at a certain instant in time. So it has an overall of 500 people, each with 4,097 data points collected over 23.5 s.

All 4,097 data points are split and scrambled into 23 portions, each holding 178 data points for 1 s, with each data point reflecting the amplitude of the EEG recording at a certain point. So, it has $23 \times 500 = 11,500$ pieces of data (row), each data point containing 178 data points for 1 s (column), and the last column represents the labels y , which are 1, 2, 3, 4, and 5. In column 179, the response variable is y , and the explanatory variables are $X_1, X_2 \dots X_{178}$. The 178-dimensional input vector's category is stored in y . In particular, 1, 2, 3, 4, and 5. Seizure activity is recorded. They took an EEG recording from the tumor's location. They located the tumor in the brain and captured EEG activity in a normal brain region.

Eyes closed, which suggests the patient's eyes were closed while the EEG signal was being recorded. Also, eyes open refers to the patient's eyes being open while the EEG signal of the brain is being recorded.

There are 178 EEG characteristics and five potential classes, as mentioned before. The dataset's purpose is to detect epilepsy from EEG data correctly. There are five classes in the dataset. The class label 1 is for patients who have an epileptic seizure (seizure activity). The other classes, 2, 3, 4, and 5, are for the patients who did not have epileptic seizures (non-seizure activity). In this study, we classify the patients with seizure activity from those with non-seizure activity. Hence, a binary classification task is conducted among class label 1, encoded as class label 1 for patients with seizure activity, and the other classes 2, 3, 4, and 5, encoded as class label 0 for patients with non-seizure activity. Let us specify the dependent variable (Y) and independent variables (X) to train the model (f).

5.3 DNN structure

DNNs are capable of identifying complex patterns within data due to their deep architecture, which includes multiple layers of

neurons. Proposed model is highly adaptable and can be applied to various tasks, including natural language processing and numerical data processing. This versatility makes them a strong candidate for diverse research applications. When trained with large datasets, DNNs often achieve higher accuracy compared to other models. Their ability to model complex functions and relationships within data is advantageous for tasks requiring precise and detailed analysis. The model consists of three dense layers in which each input layer to each output layer is fully connected. The activation function, rectified linear unit (ReLU), is used in dense layers for the output layer activation function. Sigmoid is used because the model works on binary classification. The dropout with each dense layer temporarily ignores/deactivates the network's neurons.

5.4 Results with 80% training and 20% test sets

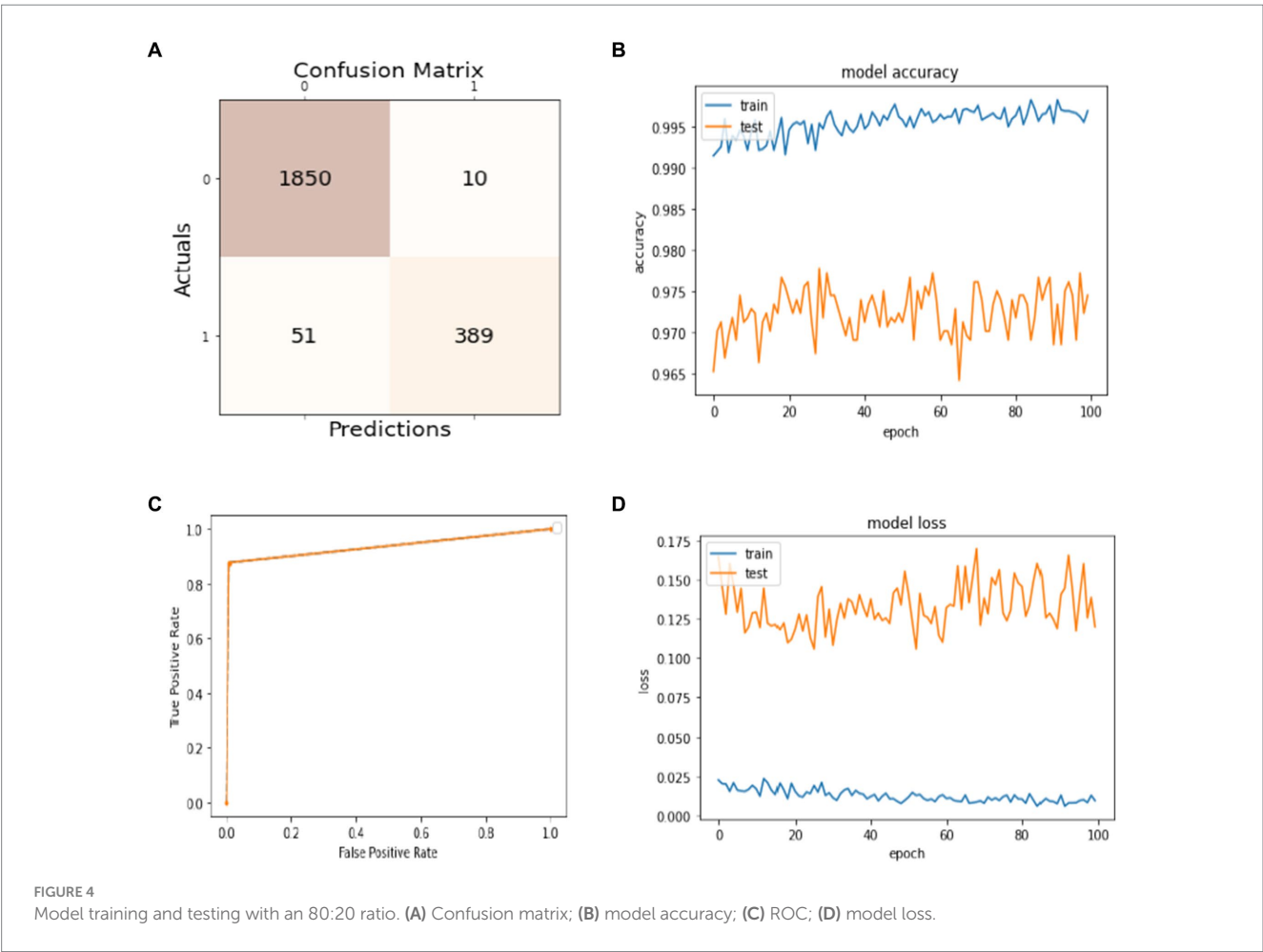
Data splitting is performed with a ratio of 80% for training and 20% for model testing. The results of the experiments are evaluated regarding the true positive examples in the confusion matrix, which are 1850, false positive examples, which are 10. False negative examples, which are 51, and true negative examples, which are 389, as shown in Figure 4A. In Figure 4B, we can see that the accuracy

curve of the model for training differs from 99% and above, and the accuracy of testing varies from 96 to 97%. In Figure 4C, the ROC moves from 0.9 of true positive toward 1.0 of false positive rate. In the model loss graph, as shown in Figure 4D, the loss of testing is about 10–17%, and the loss of training is approximately in the range of 0–2%.

The excellence of the developed model can be obtained by the values of precision, recall, F1-score, and accuracy shown in Table 1. The precision of the model is 97% for non-seizure activity, whereas the precision of seizure activity is 98%. The recall results are 99% for non-seizure activity and 88% for seizure activity. Also, we can see that the F1-score for non-seizure activity is 98%, and for seizure activity is 93%, regarding 1860 instances of non-seizure activity and 440 instances of seizure activity.

TABLE 1 Performance evaluation of non-seizure and seizure activity with an 80:20 ratio.

Class label	F1-score (%)	Precision (%)	Recall (%)	Accuracy (%)	Support
0	98	97	99	97	1860
1	93	98	88		440



5.5 Results with 70% training and 30% test sets

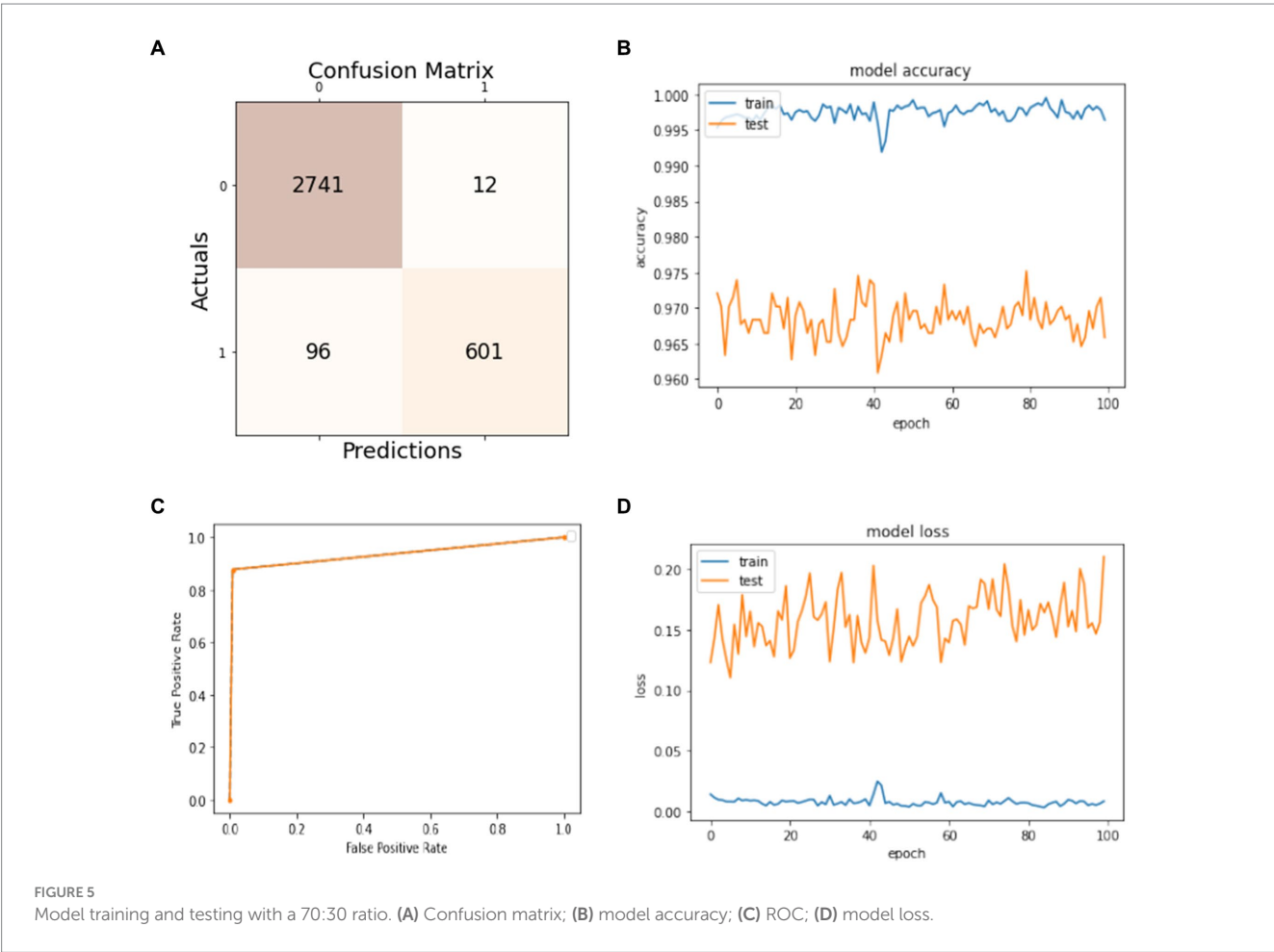
The results of splitting data into 70% for training and 30% for testing the model are discussed in this subsection. As shown in Figure 5A, the confusion matrix displays the first row-wise value to represent the true positive instances, which are patients who do not have epileptic seizures, and the model classifies them correctly as true positive instances. The second value of the confusion matrix is for the false positive instances, which are the model classified incorrectly as patients not having epileptic seizures, but in actuality, they have. The third value of the confusion matrix is several false negative instances, which the model classified as patients having epileptic seizures, but in actuality, they do not have the disease. The last value of the confusion matrix is for the true negative instances that the model classified as patients who have epileptic seizures and have epileptic seizures. In Figure 5B, the blue curve represents the training accuracy of the model, and the orange curve indicates the testing accuracy. It shows that the maximum accuracy of training reaches 99%, and the testing accuracy reaches 97.5% during the different number of epochs. Figure 5C shows the ROC curve that represents the trade-off between specificity (1 – FPR) and sensitivity (or TPR) (27). Basically, it is the relation between the true positive rate and the

false positive rate. It shows that when the true positive rate is 0.8, the false positive is 0.0, and when the true positive is 1.0, the true positive is 0.93. Figure 5D visualizes the training and testing model loss, showing how much data is lost at different epochs. The model has 97% overall accuracy, as seen in Figure 5B.

The precision, recall, F1-score, and accuracy values shown in Table 2 show the study's proficiency. The accuracy of the model is 97% for overall activity recognition. The precision of non-seizure activity is 97 and 98% for seizure activity, whereas the recall for seizure activity is 99% and for non-seizure activity is 86%. The F1-score for non-seizure activity is 98 and 92% for seizure activity. This experiment's test instances (support) are 2,753 for non-seizure activity and 697 for seizure activity.

TABLE 2 Performance evaluation of non-seizure and seizure activity with a 70–30 ratio.

Class label	F1-score (%)	Precision (%)	Recall (%)	Accuracy (%)	Support
0	98	97	99	97	2753
1	92	98	86		697



5.6 Results with 60% training and 40% test sets

In this subsection, the experiment uses 60% of the dataset for training the model and 40% for the models' test. The obtained results are presented in Figure 6. The confusion matrix is given in Figure 6A. It shows that 3,670 instances are classified as true positives, 21 instances are classified as false positives, 131 instances are classified as false negatives, and 778 instances are classified as true negatives. Figure 6B visualizes the model accuracy rates during the training process, which are above 99% for training accuracy and between 96 and 97% for testing accuracy. Figure 6C shows the ROC of the model at different numbers of true and false positive rates for the splitting data with a 60:40 ratio. For model loss, Figure 6D shows that the testing loss varies from 15 to 25% and from 0 to 5% for the training loss.

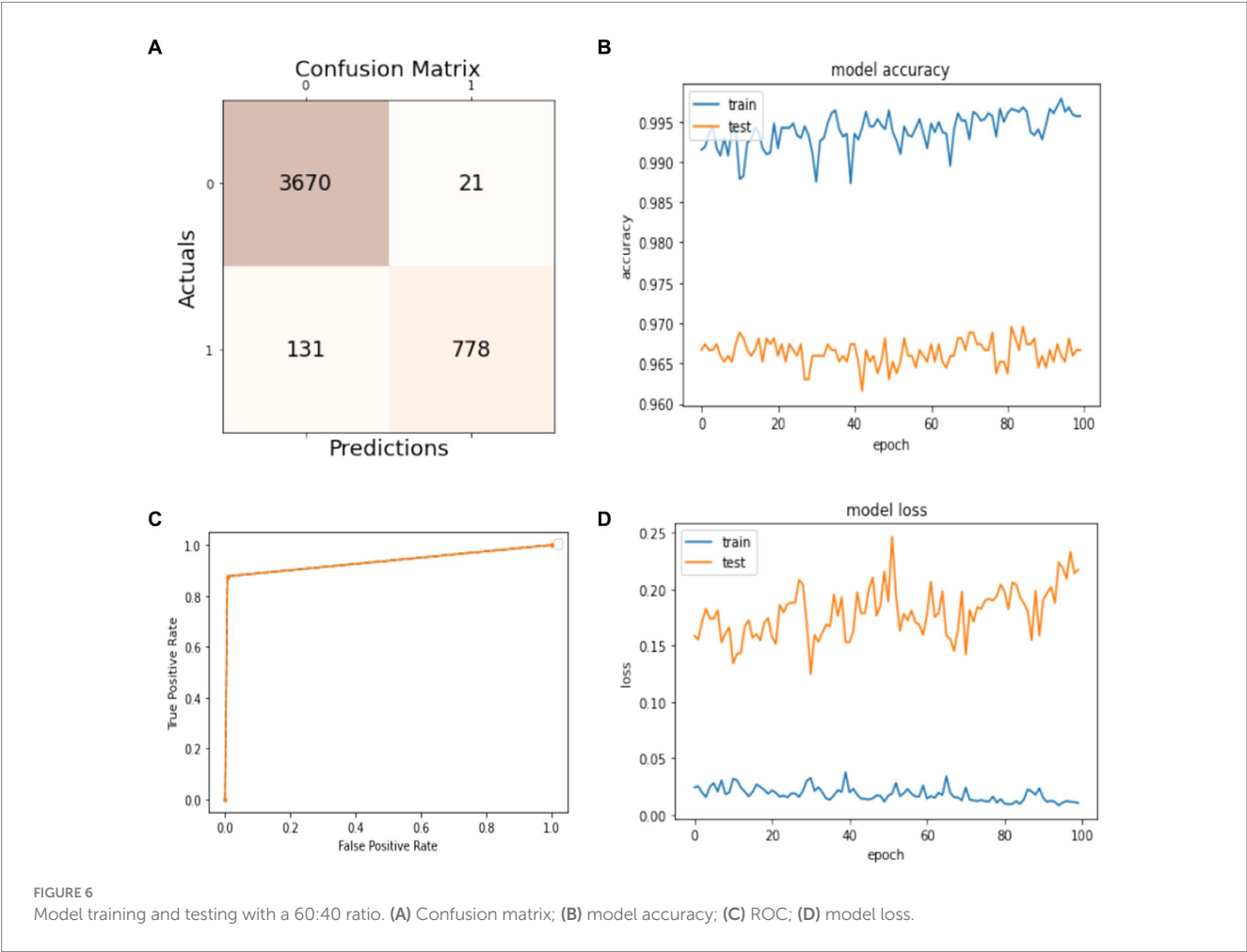
The efficiency of the model can be assessed by the values of precision, recall, F1-score, and accuracy shown in Table 3. The accuracy of the developed model is 96% for classifying both activities, while the precision for non-seizure activity is 97 and 96% for seizure. The recall for non-seizure is 99 and 88% for seizure activity. The F1-score for non-seizure activity is 96 and 91% for seizure activity. The number of instances is 3,663 for non-seizure activity and 937 for seizure activity.

5.7 Results with 50% training and 50% testing sets

Figure 7 presents the model's results trained on 50% of the dataset and tested on the remaining 50%. In Figure 7A, the confusion matrix shows that the number of true positives is 4,561 and the number of false positives is 37, measuring the model's ability to predict the non-seizure activity truly. The false negative and true negative instances in the confusion matrix, which are 152 and 1,000, mean that the model can predict 152 cases from 1,152 as they have non-seizure activity, but actually, they have seizure activity. Similarly, the model can predict 1,000 instances as they have had seizure activity since 1,152, and they have had seizure activity. The accuracy of training and testing during the training phase are given in Figure 7B. It shows the model's accuracy fluctuation from 0 to 100 epochs. The same is true

TABLE 3 Performance evaluation of non-seizure and seizure activity with a 60:40 ratio.

Class label	F1-score (%)	Precision (%)	Recall (%)	Accuracy (%)	Support
0	96	97	99	96	3,663
1	91	96	88		937



for the model's loss, which is given in Figure 7D. Figure 7C shows the ROC of the model at different numbers of true and false positive rates for the splitting data with a 50:50 ratio.

The results of precision, recall, F1-score, and accuracy are listed in Table 4. It illustrates the effectiveness of the model. We can see that the precision for non-seizure activity is 98% and for seizure is 94%, the recall for non-seizure is 98%, and for seizure is 87%, and the F1-score for non-seizure is 94 and 94% for seizure. The number of test instances (support) is 4,598 for non-seizures and 1,152 for seizure activity. The results of evaluation metrics for the model in overall splitting ratios are presented in Table 5.

5.8 10-fold cross validation

A 10-fold cross-validation technique is applied to the whole dataset to evaluate the model's performance further, as shown in Figure 8. The total number of instances in the dataset is 11,500. It is divided into 10 equal parts for the 10-fold cross-validation. In each part, 1150 instances are used to test the model. The obtained results are introduced in this subsection. Figure 8 illustrates the strategy of a 10-fold cross-validation technique for splitting the data for training and validation sets.

In Table 6, we present a comparison of different models' accuracy results using the holdout and 10-fold cross-validation techniques. As

we can see, the lowest accuracies are for the logistic regression model, which is 82.5% using a holdout technique, and 80.1% using a 10-fold cross-validation technique, while the highest accuracies are for the proposed model, which is 97% using a holdout technique and 95.5% using a 10-fold cross-validation technique. Also, we can notice that the accuracy of different models using a holdout technique is slightly higher compared to a 10-fold validation technique.

Table 7 compares different models' F1-score results using the holdout and 10-fold cross-validation techniques. As we can see, the lowest F1-scores are for the logistic regression model, with 81.5% using a holdout technique and 80.1% using a 10-fold cross-validation technique, while the highest F1-scores are for the proposed model, which is 93% using a holdout technique and 90.5% using a 10-fold cross-validation technique. Also, we can notice that the F1-score of different models using a holdout technique is a little bit higher when compared with a 10-fold validation technique.

TABLE 4 Performance evaluation of non-seizure and seizure activity with a 50–50 ratio.

Class label	F1-score (%)	Precision (%)	Recall (%)	Accuracy (%)	Support
0	94	95	98	95	4,598
1	90	94	87		1,152

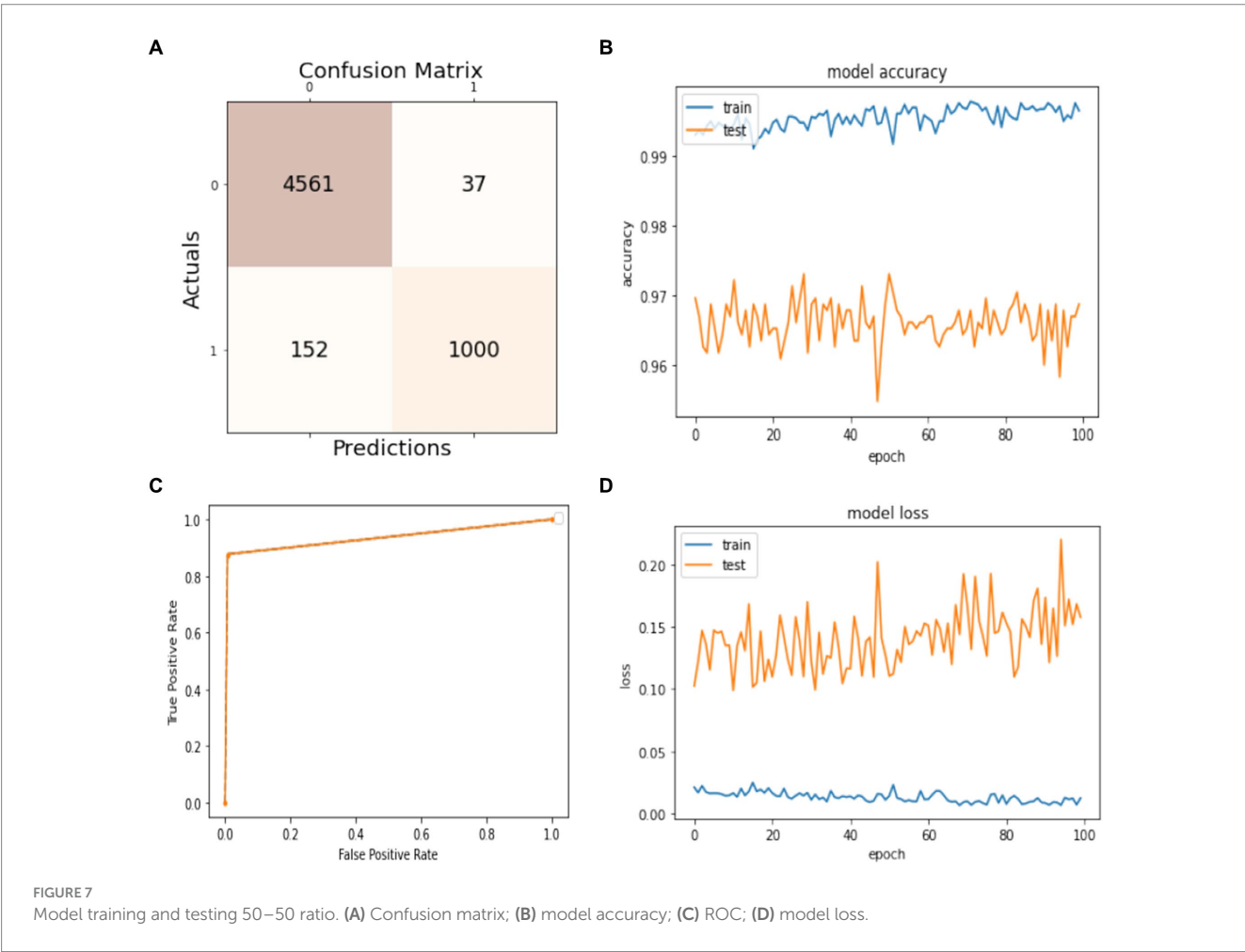


TABLE 5 Overall performance of the model with different ratios.

Training	Testing	F1-score (%)	Precision (%)	Recall (%)	Accuracy (%)
80%	20%	93	98	88	97
70%	30%	92	98	86	97
60%	40%	91	96	88	96
50%	50%	90	94	87	95

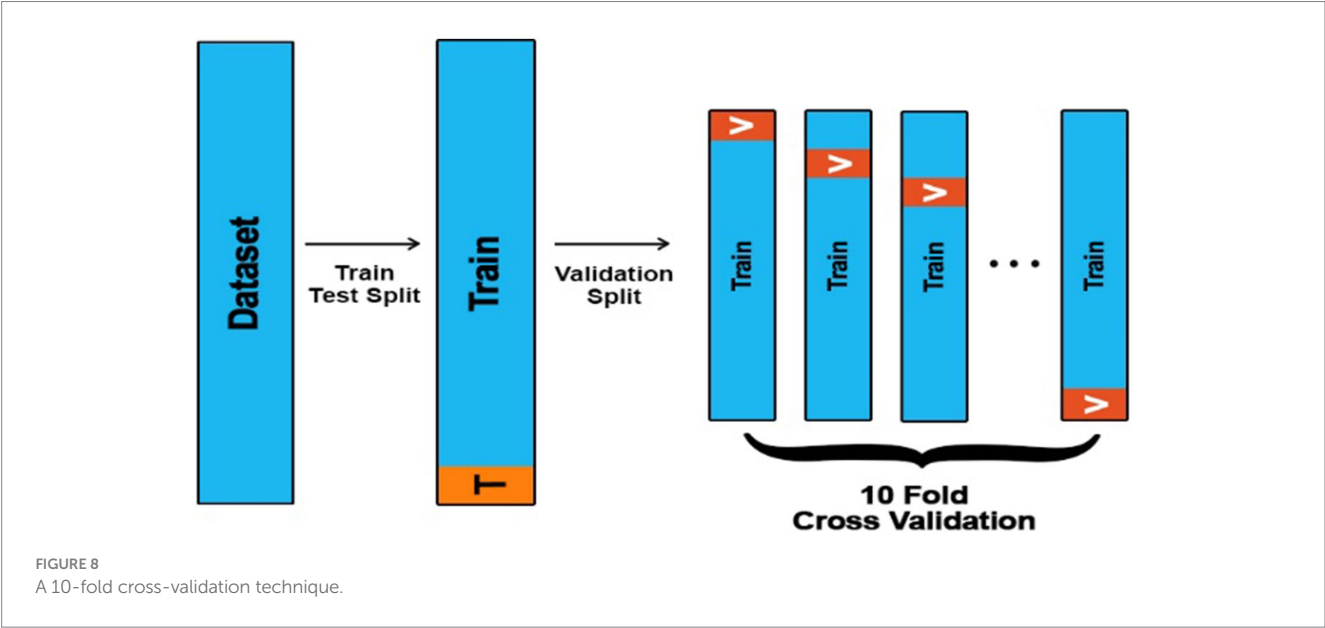


TABLE 6 Accuracy of different models using the holdout and 10-fold cross-validation techniques.

Model	Accuracy of holdout technique (%)	Accuracy of 10-fold validation technique (%)
ANN	95.7	93.4
Naive Bayes	95	94.3
KNN	93.1	91.6
Logistic regression	82.5	80.1
DNN	97	95.5

TABLE 7 F1-score of different models using the holdout and 10-fold cross-validation techniques.

Model	F1-score of holdout technique (%)	F1-score of 10-fold validation technique (%)
ANN	92.3	90.4
Naive Bayes	89.2	87.3
KNN	90	91.6
Logistic regression	81.5	80.1
DNN	93	90.5

Similarly, in Table 8, we compare the precision results of different models using the holdout and 10-fold cross-validation techniques. As we can see, the lowest precisions are for the logistic regression model, with 81.5% using a holdout technique, and 80.1% using a 10-fold cross-validation technique, while the highest precisions are for the proposed model, with 93% using a holdout technique and 90.5% using a 10-fold cross-validation technique. Also, we can notice that the precision of different models using a holdout technique is slightly higher compared to a 10-fold validation technique.

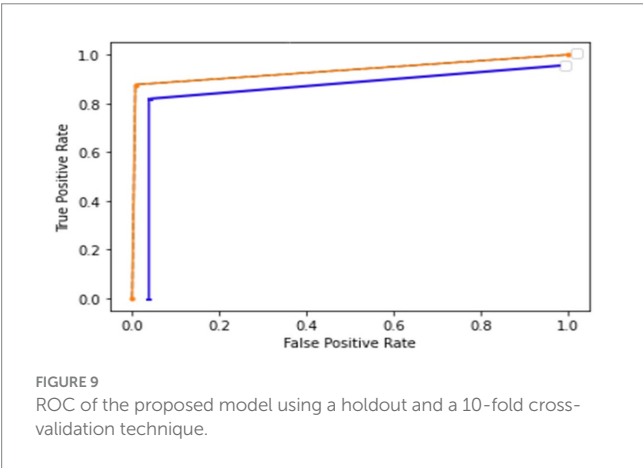
Figure 9 shows the receiver operator characteristic curve (ROC). The orange curve indicates the ROC of the proposed model using a holdout technique. It is shown that when the true positive rate is 0.9, the false positive is 0.0, and when the true positive is 0.93, the false positive is 1.0. The blue curve represents the ROC of the proposed

model using a 10-fold cross-validation technique. It is 0.0 when it starts, but when the curve reaches 0.8, the graph achieves a rate of 0.98. The ROC curve presents how well the model can differentiate among positive and negative classes by plotting the true positive rate against the false positive rate at several thresholds. The performance of the model is summarized by a single value by the area under the ROC curve (AUC). When the cost of false positives and false negatives fluctuates, the ROC curve provides a balanced assessment of the model's performance by taking into account both true positive and false positive rates.

Figures 10, 11 show the proposed model's loss and accuracy using a 10-fold cross-validation technique. The error or model loss graph indicates the overall loss of 10-fold cross-validation during testing and training. In the case of testing, the loss is 0.16% at the first epoch, and

TABLE 8 The precision of different models using the holdout and 10-fold cross-validation techniques.

Model	Precision of Holdout Technique (%)	Precision of 10-fold Validation Technique (%)
ANN	95	93.4
Naive Bayes	96.4	94.3
KNN	92	91.6
Logistic regression	85.1	83.5
DNN	98	95.5



it goes higher at 50 and 100 epochs. The loss is 0.23%. For training, the loss is 0.03% on the first epoch and goes higher on the epoch number 40; when it reaches the epoch number 100, the loss is 0.01%.

The accuracy graph for the training of 10-fold cross-validation is shown above as it can be seen that the graph started from 90% accuracy on 0 epoch and remained almost the same at 90 epoch, but an abrupt increase in accuracy can be seen after 90 epoch and achieve 93% accuracy. The accuracy graph for testing of 10-fold cross-validation in which the graph fluctuates between 65 and 100%.

6 Discussion

The present study aims to achieve high accuracy by using a numerical data set for our model. The model is trained and tested using different dataset ratios for the best results. Before this study, most of the previous methods used image data sets to execute their research work, but in this study, historical numerical data was employed, which is not complex compared to other methods. Furthermore, a binary classifier (non-seizure or seizure) is used, which does not predefine more specific seizure categories to provide a more generalizable classifier. The DNN algorithm has more than one hidden layer between the input and output layers; the data will be passed through these hidden layers' functions, in which the function applies weights to the inputs and sends them as the output using an activation function. The activation function used in this study is Sigmoid.

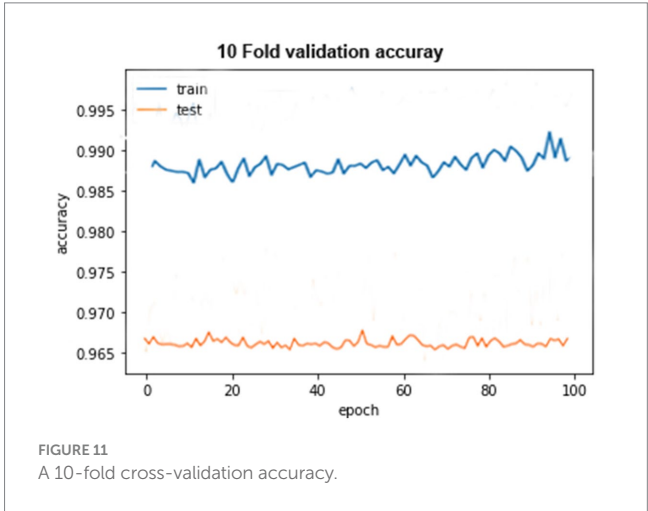
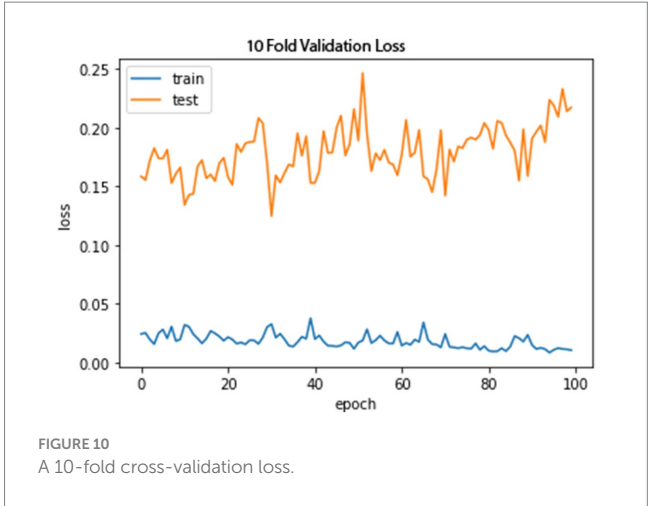


TABLE 9 Accuracy (%) comparison results in the percentage of the proposed DNN model with the other models at different splitting ratios.

Model	80–20%	70–30%	60–40%	50–50%
ANN	95.7	95	94	92
Naive Bayes	95	94	93	91.5
KNN	93	92.5	91.5	91
Logistic regression	82.5	82	81	80
DNN	97	97	96.5	95

7 Comparative analysis

This section will compare the proposed model with the other machine learning approaches concerning accuracy, precision, F1-score, and recall. The comparison of machine learning models with different training and testing ratios, i.e., 80–20%, 70–30%, 60–40%, and 50–50%, will be done through graphs and tables.

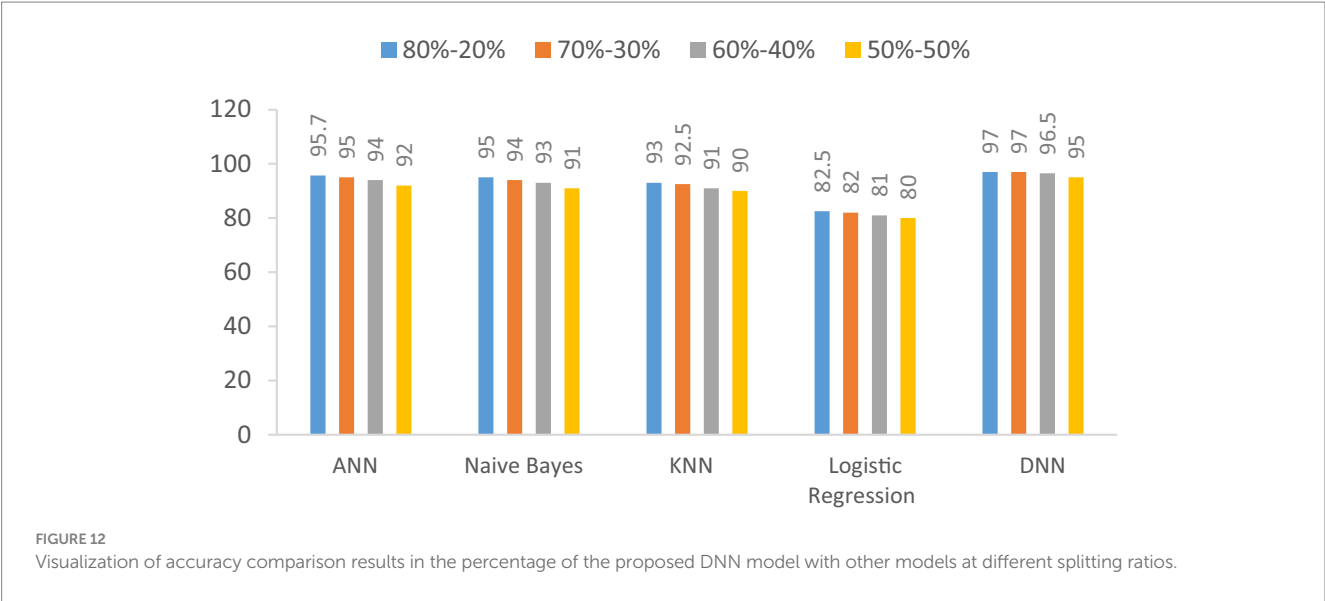
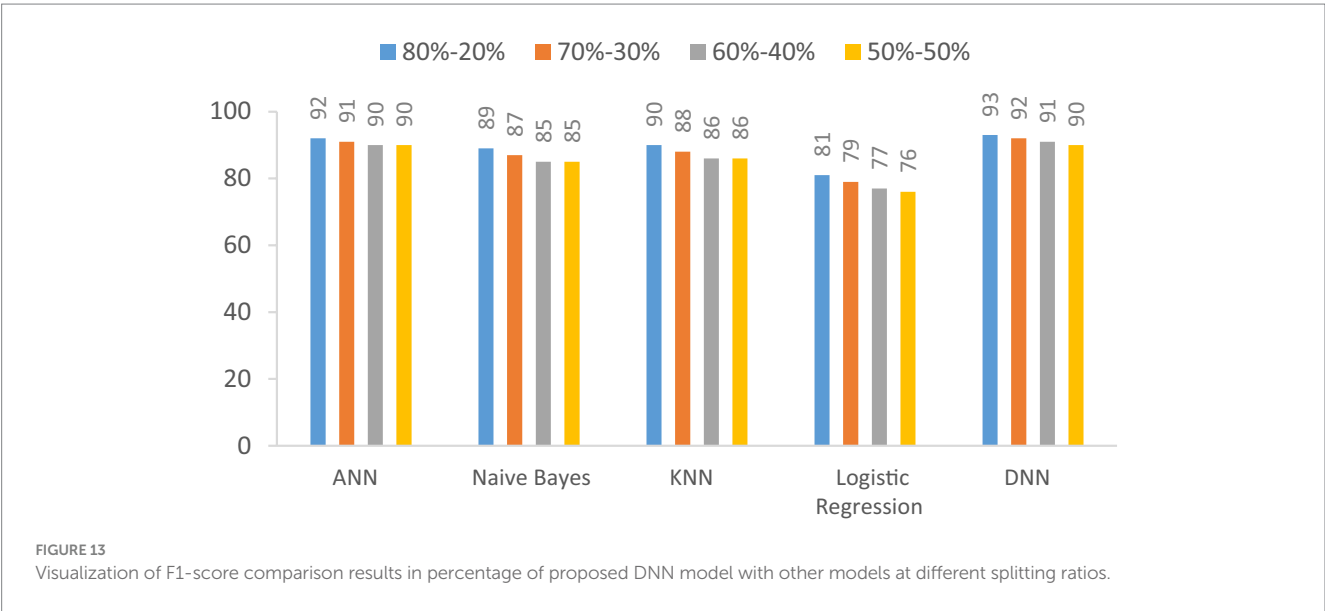


TABLE 10 F1-score (%) comparison results in a percentage of the proposed DNN model with other models at different splitting ratios.

Model	80–20%	70–30%	60–40%	50–50%
ANN	92	91	90	90
Naive Bayes	89	87	85	85
KNN	90	88	86	86
Logistic regression	81	79	77	76
DNN	93	92	91	90



7.1 Accuracy

The accuracy of the proposed DNN model is compared with the other models, such as Logistic regression, KNN, ANN, and Naïve

Bayes, using different splitting ratios as given in Table 9 and visualized in Figure 12. We can see that the proposed DNN model achieves the highest accuracy result compared to other models. Despite a general decline in accuracy across all models when the training data is

reduced, the DNN model exhibits notable resilience, maintaining comparatively high accuracy even with a balanced 50–50 data split. This suggests that the DNN model is capable of delivering strong performance even with a smaller amount of training data.

7.2 F1-score

The F1-score of the proposed DNN model is compared with the other models, such as Logistic regression, KNN, ANN, and Naïve Bayes, using different splitting ratios as given in Table 10 and visualized in Figure 13. Logistic Regression consistently shows the lowest F1-scores for all data splits, indicating its limited effectiveness for this task. On the other hand, while ANN, KNN, and Naïve Bayes deliver decent results, they still fall short compared to the performance achieved by the DNN model.

7.3 Precision

We examine the precision of the proposed Deep Neural Network (DNN) model in contrast to several established models: Artificial Neural Network (ANN), Naïve Bayes, K-Nearest Neighbors (KNN), and Logistic Regression. This evaluation encompasses various data splitting ratios, including 80–20%, 70–30%, 60–40%, and 50–50%.as shown in Table 11 and Figure 14. The values will be changed when the training and testing data ratios are changed.

7.4 Recall

In this section, we investigate the recall capabilities of the proposed Deep Neural Network (DNN) model when compared to alternative models across diverse data splitting ratios. The outcomes are illustrated in Table 12 and Figure 15. Recall assesses a model's proficiency in correctly recognizing all pertinent instances among the total relevant instances. As we manipulate the proportions between training and testing datasets, the figures in the table will adapt accordingly.

8 Conclusion

The primary objective of this study is to optimize the accuracy and performance of our research outcomes. To accomplish this, we have employed a sophisticated deep neural network (DNN) algorithm while systematically manipulating the ratios of training and testing datasets to discern optimal results. The results showed substantial advancements over previous research endeavors, boasting a remarkable 97% accuracy rate, a precision rate of 98%, an F1-score of 92%, and a recall rate of 80%. Furthermore, our commitment to robust validation methodologies is evident in applying a rigorous 10-fold cross-validation technique designed to further enhance the model's performance and bolster its reliability across the dataset. Integrating EEG data with other physiological measurements, such as heart rate and movement data, may enhance the accuracy of seizure detection. Future research could investigate methods for combining these diverse

TABLE 11 Precision (%) comparison results in a percentage of the proposed DNN model with other models at different splitting ratios.

Model	80–20%	70–30%	60–40%	50–50%
ANN	95	92	91	91
Naive Bayes	96	95	85	85
KNN	92	93	90	89
Logistic regression	85	83	83	80
DNN	98	97	95	90

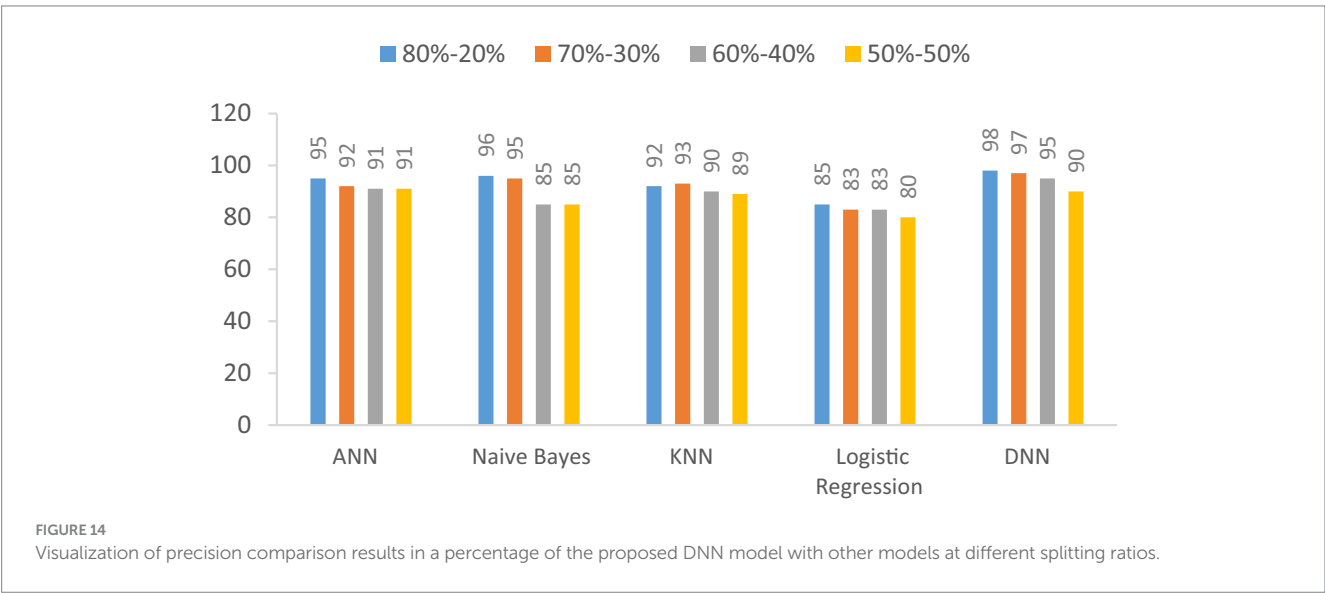
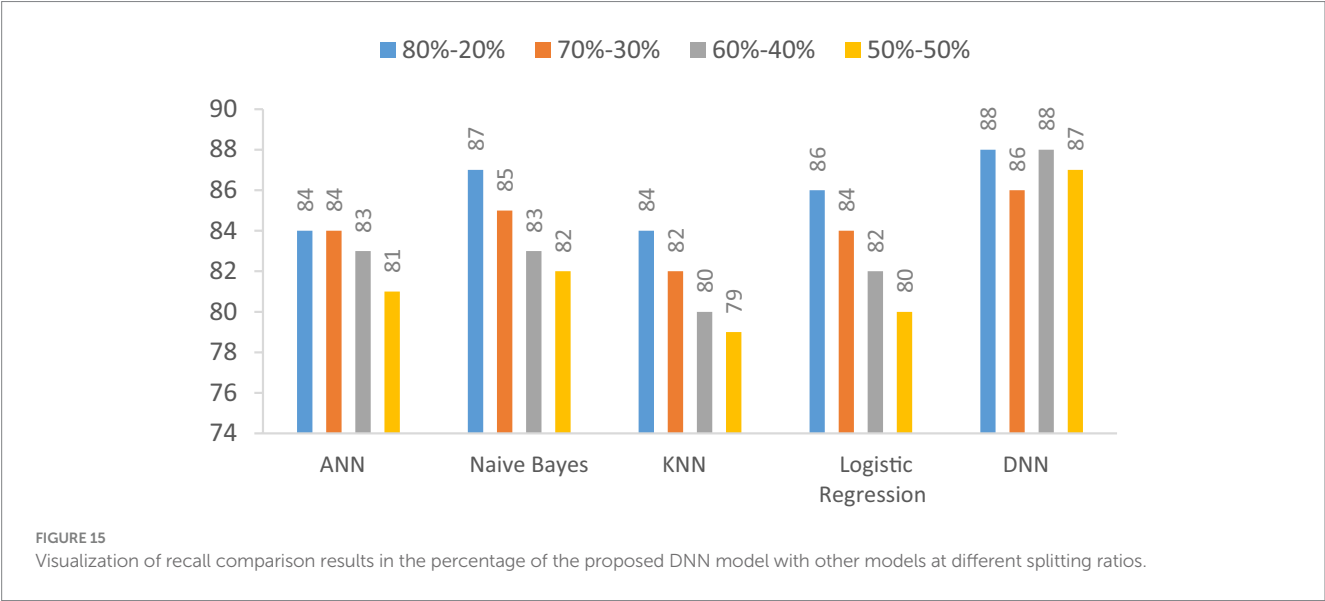


TABLE 12 Recall (%) comparison results in the percentage of the proposed DNN model with other models at different splitting ratios.

Model	80–20%	70–30%	60–40%	50–50%
ANN	84	84	83	81
Naive Bayes	87	85	83	82
KNN	84	82	80	79
Logistic regression	86	84	82	80
DNN	88	86	88	87



data types to utilize the unique benefits of each. Additionally combining the seizure detection system with electronic health records to enhance patient history tracking and care management could also be Upcoming research.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: <https://www.kaggle.com/datasets/harunshimanto/epileptic-seizure-recognition>.

Author contributions

DK: Conceptualization, Data curation, Formal analysis, Investigation, Validation, Visualization, Writing – original draft, Writing – review & editing. FW: Conceptualization, Methodology, Project administration, Resources, Software, Supervision, Writing – original draft, Writing – review & editing. SiA: Conceptualization, Data curation, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. AG: Conceptualization, Data curation, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. SaA: Formal analysis, Investigation, Methodology, Project administration, Resources, Software,

Supervision, Visualization, Writing – original draft, Writing – review & editing. MM: Conceptualization, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by King Salman Center for Disability Research through Research Group (No. KSRG-2023-544).

Acknowledgments

The authors extend their appreciation to the King Salman Center for Disability Research for funding this work through Research Group No. KSRG-2023-544.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Zhang S, Wei Z, Nie J, Huang L, Wang S, Li Z. A review on human activity recognition using vision-based method. *J Healthcare Eng.* (2017) 2017:1–31. doi: 10.1155/2017/3090343
- Moya Rueda F, Grzeszick R, Fink GA, Feldhorst S, Ten Hoppel M. Convolutional neural networks for human activity recognition using body-worn sensors. *Informatics.* (2018) 5:26. doi: 10.3390/informatics5020026
- Iqbal A, Ullah F, Anwar H, Rehman AU, Shah K, Baig A, et al. Wearable internet-of-things platform for human activity recognition and health care. *Int J Distrib Sensor Net.* (2020) 16:1550147720911561. doi: 10.1177/1550147720911561
- Nweke HF, Teh YW, al-garadi MA, Alo UR. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges. *Expert Syst Appl.* (2018) 105:233–61. doi: 10.1016/j.eswa.2018.03.056
- istockphoto. (2020,). Crowd of people in outdoor activities isolated on white stock illustration. Available at: <https://www.istockphoto.com/vector/crowd-of-people-in-outdoor-activities-isolated-on-white-gm1286951995-383302651>
- Kim S.-I., Jo Eunhyung, Ryu Myeonghan, Cha Inha, Kim Young Ho, Yoo Heejung, et al., "Toward becoming a better self: understanding self-tracking experiences of adolescents with autism spectrum disorder using custom trackers," in Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare, (2019), pp. 169–178.
- Malasinghe LP, Ramzan N, Dahal K. Remote patient monitoring: a comprehensive study. *J Ambient Intell Humaniz Comput.* (2019) 10:57–76. doi: 10.1007/s12652-017-0598-x
- Liang S, Fan X, Zhao M, Shan X, Li W, Ding P, et al. Clinical practice guidelines for the diagnosis and treatment of adult diffuse glioma-related epilepsy. *Cancer Med.* (2019) 8:4527–35. doi: 10.1002/cam4.2362
- Liu C-L, Xiao B, Hsiao W-H, Tseng VS. Epileptic seizure prediction with multi-view convolutional neural networks. *IEEE access.* (2019) 7:170352–61. doi: 10.1109/ACCESS.2019.2955285
- shutterstock. Seizures epilepsy - seizure hotspot. Medical illustration of a brain with epilepsy. Available at: <https://www.shutterstock.com/image-vector/seizures-epilepsy-seizure-hotspot-medical-illustration-1354943030>
- Richens JG, Lee CM, Johri S. Improving the accuracy of medical diagnosis with causal machine learning. *Nat Commun.* (2020) 11:3923. doi: 10.1038/s41467-020-17419-7
- Hassan MM, Uddin MZ, Mohamed A, Almogren AJFGCS. A robust human activity recognition system using smartphone sensors and deep learning. *Futur Gener Comput Syst.* (2018) 81:307–13. doi: 10.1016/j.future.2017.11.029
- Gul MA, Yousaf MH, Nawaz S, Ur Rehman Z, Kim HJE. Patient monitoring by abnormal human activity recognition based on CNN architecture. *Electronics.* (2020) 9:1993. doi: 10.3390/electronics9121993
- Schmitt M, Schuller B. Deep recurrent neural networks for emotion recognition in speech. *Proc DAGA.* (2018) 44:1537–40.
- Uddin MZ, Hassan MMJISJ. Activity recognition for cognitive assistance using body sensors data and deep convolutional neural network. *IEEE Sensors J.* (2018) 19:8413–9. doi: 10.1109/JSEN.2018.2871203
- Quichka O, Echtioui A, Hamam HJE. Deep learning models for predicting epileptic seizures using iEEG signals. *Electronics.* (2022) 11:605. doi: 10.3390/electronics11040605
- Ibrahim FE, Emara HM, el-Shafai W, Elwekeil M, Rihan M, Eldokany IM, et al. Deep-learning-based seizure detection and prediction from electroencephalography signals. *Commun Numer Methods Eng.* (2022) 38:e3573. doi: 10.1002/cnm.3573
- Poorani S, Balasubramanie PIJIOSAE. Deep learning based epileptic seizure detection with EEG data. *Comput Intell Neurosci.* (2023) 2023:1–10. doi: 10.1155/2023/9814248
- Abderrahim N, Echtioui A, Khemakhem R, Zouch W, Ghorbel M, Hamida AB. Epileptic seizures detection using iEEG signals and deep learning models. *CSSP.* (2024) 43:1597–626. doi: 10.1007/s00034-023-02527-8
- Leenings R, Winter NR, Plagwitz L, Holstein V, Ernsting J, Sarink K, et al. PHOTONAI—A Python API for rapid machine learning model development. *PLoS I.* (2021) 16:e0254062. doi: 10.1371/journal.pone.0254062
- Braiek HB, Khomf F. On testing machine learning programs. *J Syst Softw.* (2020) 164:110542. doi: 10.1016/j.jss.2020.110542
- Luque A, Carrasco A, Martín A, de Las Heras AJPR. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recogn.* (2019) 91:216–31. doi: 10.1016/j.patcog.2019.02.023
- Yin M., Wortman Vaughan J., Wallach H., "Understanding the effect of accuracy on trust in machine learning models," in Proceedings of the 2019 chi conference on human factors in computing systems, (2019), pp. 1–12.
- Luan H, Tsai C-CJET. A review of using machine learning approaches for precision education. *Educ Technol Soc.* (2021) 24:250–66.
- Wu M, Yang Y, Wang H, Xu Y. A deep learning method to more accurately recall known lysine acetylation sites. *Bioinformatics.* (2019) 20:49–11. doi: 10.1186/s12859-019-2632-9
- Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *Genomics.* (2020) 21:6–13. doi: 10.1186/s12864-019-6413-7
- Kannan R, Vasanthi V. Machine learning algorithms with ROC curve for predicting and diagnosing the heart disease In: NB Muppalaneni, M Ma and S Gurumoorthy, editors. Soft computing and medical bioinformatics. Singapore: Springer (2019). 63–72.



OPEN ACCESS

EDITED BY

Hikmat A. M. Abdeljaber,
Applied Science Private University, Jordan

REVIEWED BY

Inam Ullah,
Shandong Jianzhu University, China
Razaullah Khan,
University of Engineering and Technology,
Mardan, Pakistan

*CORRESPONDENCE

Alia Tabassum
✉ alia.tabassumcs@gmail.com

RECEIVED 24 April 2024

ACCEPTED 03 July 2024

PUBLISHED 30 July 2024

CITATION

Jian W, Tabassum A and Li JP (2024)
Systematic survey on data security in wireless
body area networks in IoT healthcare system.
Front. Med. 11:1422911.
doi: 10.3389/fmed.2024.1422911

COPYRIGHT

© 2024 Jian, Tabassum and Li. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Systematic survey on data security in wireless body area networks in IoT healthcare system

Wang Jian¹, Alia Tabassum^{2*} and Jian Ping Li³

¹School of Artificial Intelligence, Neijiang Normal University, Neijiang, Sichuan, China, ²Department of Computer Science and Software Engineering International Islamic University Islamabad, Islamabad, Pakistan, ³School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China

In the Internet of Things (IoT) healthcare sector, the wireless body area network (WBAN) is being used to optimize medical results by tracking and treating patients as they go about their daily lives. Health insurance has also been one of the cybercriminal's main goals. The Systematic Review of IoT Healthcare systems particularly wireless body area networks is significant, to reach the benefits and challenges faced by existing methods in the domain. This study provides a systematic survey of WBAN data protection. Various types of devices are used in medical science to detect and diagnose diseases. The network is an integral part of medical science in today's era. In medical sciences, sensors take data from a problematic place like cancerous cells. This research discussed a lot of techniques in the literature review. Most of them are not able to fulfill the requirements. If an unauthorized person reaches the data that can be a severe issue, like the diagnosed disease was blood cancer, and after unauthorized access manipulation can change even the diagnosed issue in the database. A doctor can prescribe the medication based on provided data that has been manipulated by unauthorized persons. Several existing schemes are explored in the literature to determine how the protection of sharing patients' healthcare data can be improved. The systematic literature review (SLR) of multiple security schemes for WBAN is presented in this survey paper.

KEYWORDS

WBAN, data security, encryption decryption, SLR, healthcare

1 Introduction

The security of WBAN (1) is essential and should not be forgotten. It is confidential, sensitive medical information and must be shielded from unauthorized persons who can use details that may be harmful to the person (2). By using WBAN with the use of various varieties of sensors to track the patients to detect any life-threatening diseases healthcare applications are enhanced. This technology aims to increase the quality of medical services delivered and reduce certain related costs. BAN has a broad spectrum of applications, like tracking the medical conditions of patients and optimizing their response to clinical guidelines, but protection and privacy are some of the main concerns in BAN-based healthcare systems at the same time medical data must be kept protected from risk factors and hackers during storage and transmission (3, 4).

The existing literature discusses data privacy and protection (5, 6), but it doesn't go into depth about the SLR and the requirements for collecting data. There is a lot of literature on security strategies, but it isn't focused on security research. This study used three different Databases including IEEE, ACM, and Springers. The existing literature discusses data privacy and protection, but it doesn't go into depth about the SLR and the requirements for collecting data. There is a lot of literature on security strategies, but it isn't focused on security research. This study used three different Databases including IEEE, ACM, and Springers. Develop a string by using the objective of all papers and then used 3 synonyms of each keyword in the string. This research discussed inclusion criteria in which we have two parts one is included and another is excluded (not included). The thesis, newspaper, books and are not included in the inclusion criteria and title-based, abstract-based, and objective-based research papers are included. After that design, this study performs objective-based filtering and abstract-based filtering. Aim objectives and methodologies of each paper are discussed below. And also provide critical analysis. The conclusion of this research is to provide an efficient way for data security in WBAN. Privacy in WBAN is important and should not be forgotten. Medical data is important and must be shielded from unauthorized access. The motivation is to compile and research papers that deal with security issues in depth. In this research papers were identified after an extensive search using strings in different databases. The papers were then screened using title and abstract-based evaluations to determine if the study was appropriate or not. We present the comparative analysis of data in tabular form in this section. The study concluded that WBAN is a more effective approach to exchanging data between doctors and patients by doing this survey. In Table 1, existing surveys on data security in WBAN are discussed. This shows the strength of this survey paper with already existing survey papers using the comparison method. Compare all survey papers in terms of communication cost, energy consumption, storage, etc.

1.1 Motivation of the study

The existing literature discusses data privacy and protection, but it doesn't go into depth about the SLR and the requirements for collecting data. Many methods for improving technical efficiency have already been established in this area, but current work required more accuracy. Another relevant and high-quality SLR survey has been rationalized, but it used a limited amount of established literature, which could impact methodology comparisons. Additionally, a systematic analysis is based on comparing and highlighting study gaps; however, this survey does not include many details regarding current WBAN literature schemes. The research conducts a systematic literature review, which is used to support the proposed SLR in the survey. The major contributions of this study are as follows:

1. To create a taxonomy that covers the security encryption techniques that are required in the WBAN setting. Existing work has been addressed in depth in each section of the taxonomy to address a variety of issues, including time, cost, and predicting network attacks.
2. From 2017 to 2024, we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow chart to search the literature, delete duplicate information, screen, exclude, and include articles.
3. SLR can be used for very relevant schemes that concentrate on protecting healthcare data by preventing security threats while using less memory.
4. To encourage researchers to provide effective solutions to problems, a security review with criticism is performed.

The rest of the paper is arranged as follows. Section 1 presents the introduction to WBAN. Section 2 shows the Systematic literature review (SLR). Section 3 shows the Detailed Literature Review and Section 4 concludes this work.

2 Systematic literature review

This study chose a year range(2017–2024), selected three synonyms for each string keyword, searched three databases (ACM, Springer, IEEE), and then conducted random searches against strings. This research created a string containing all of the papers' objectives and then utilized three synonyms for each keyword in the string. Then this research talked about inclusion criteria, where one component is included and the other is excluded (not included). The thesis, newspaper, books and are not included in the inclusion criteria, and title-based, abstract-based, and objective-based research papers are included.

2.1 Research selection procedure

The PRISMA (23) flow chart in Figure 1 demonstrates our survey's systematic review procedure. In the selection process, research papers from the years 2017 to 2024 are included. Currently, 130 papers are being considered. publications that fulfill the study criteria are selected after searching for similar publications in various databases. During an initial review, 75 papers were shortlisted, and 30 relevant articles that met the requirements were included in the survey.

2.1.1 Detailed literature review

WBAN is a multifaceted network that includes a variety of sensor hubs that track and relay data in real-time in a variety of situations. Sensor nodes collect vital information and send it to a medical server for further analysis. Since data includes highly confidential and important patient information, data security and safety is a critical challenge. WBAN information security is being investigated over a long period, from 2017 to 2024. This research literature focuses on various schemes such as SHA (Secure Hashing Algorithm), AES(Advanced Encryption Standards), and many others.

2.1.2 AES based schemes

This research paper's (11, 24) goal is to build and apply a safe end-to-end PMS by focusing on the secure wireless

TABLE 1 Overview of existing research.

References	Year	Objective	Techniques	Criticism
Rehman et al. (7)	2020	The specification for storage, the cost of computing and communication, time, and cost computational.	Internet protection protocol and automatic validation (AVISPA) platform.	The lack of usability features, and different routing threats, like a privileged insider, user, and server impersonation, do not provide an effective password change point.
Jabeen et al. (8)	2020	Complexity of time, key generation development, computational time algorithm for encryption.	Novel data protection genetic-based encryption scheme, AES.	AES also has disadvantages, like high cost of calculation for hardware, large in size, use of CPUs.
Hasan et al. (9)	2020	Reliability, Trust, provide less cost and time.	-	Execute more slowly, The more complicated mathematical model.
Jabeen et al. (8)	2020	The study examines a variety of current strategies to determine how patient health data protection can be improved.	-	
Parvez et al. (10)	2019	Electronic health records (EHR) systems handle the internet's most private information, improve security.	Electronic Healthcare Repository (EHR).	Changes in workflow temporary loss of productivity.
Chowdhury et al. (11)	2018	Portable support in the form of consuming minimum energy consumption creates a larger framework for a PMS end-to-end safe communication way for different medications.	AES, MQTT	AES also has disadvantages, like high costs of computation, and high specification for hardware. More difficult, wide footprint, use of CPUs, high storage and power.
Shanthapriya and Vaithianathan (12)	2018	Integrity, reliability, medical privacy, and security.	The polynomial curve and the Steganography technique are generated.	The key concept behind these strategies is that without being found, edges will bear more variation than smooth areas.
Braham et al. (4)	2018	The energy-efficient system, energy-efficient protection protocol, reducing the cost of healthcare.	-	No formal syntax or semantic logic is restricted to authentication protocol analysis. Does not have proper encryption accounts.
Malik et al. (13)	2018	Resolving variously defined gaps in storage requirements, enhancing privacy, improving security.	Internet protection protocol and application automatic validation (AVISPA) platform.	Lack of functionality features various known attacks, such as privileged insider, user, and server impersonation, do not include.
Anwar et al. (14)	2018	-	AES and MQTT	Theoretically, file-type attacks are successful DOS, IoT attacks are not suitable for high-complexity sensor networks.
Roy et al. (15)	2017	-	AES, ECC	The implementation of AES and ECC is challenging. Complex, time-consuming, and hard to execute.
Hasan et al. (9)	2020	Achieve high throughputs and low latency for emergency traffic.	SDN (Software-defined Networking)	Vulnerabilities in security, inconsistency SDES's flow demands rise as the network becomes more complex.
Ren et al. (16)	2019	only the designated person has access to the user's data.	DVSSA	Complexity and Time Consumption.
Zhang and Ma (17)	2018	Improve user privacy at a low cost.	A security mechanism that is aware of the channel	It takes a long time to authenticate each node.
Shanmugavadivel et al. (18)	2021	In a cloud environment improve data security using the AES.	AES, Genetic Algorithm for Task Flow Scheduling.	High complexity.
Singh and Prasad (19)	2021	Provide detailed study of different systems and protocols for dealing with energy efficiency and security.	-	
Soni and Singh (20)	2021	Lower execution cost, calculation time, and power consumption when compared to other protocols.	LAKA Lightweight Authentication and Key Agreement Protocol.	Higher power consumption.
Sandhu and Malik (21)	2020	The goal of this article is to efficiently transmit data based on the priority of the data.	PAP(Priority aware protocol)	As the priority level rises, so does the amount of energy expended.
Li et al. (22)	2017	The goal of this article is to use physical channel information to eliminate the need for additional hardware requirements.	RSSI(Received Signal Strength Indicator).	An algorithm is difficult to implement because it is complex.

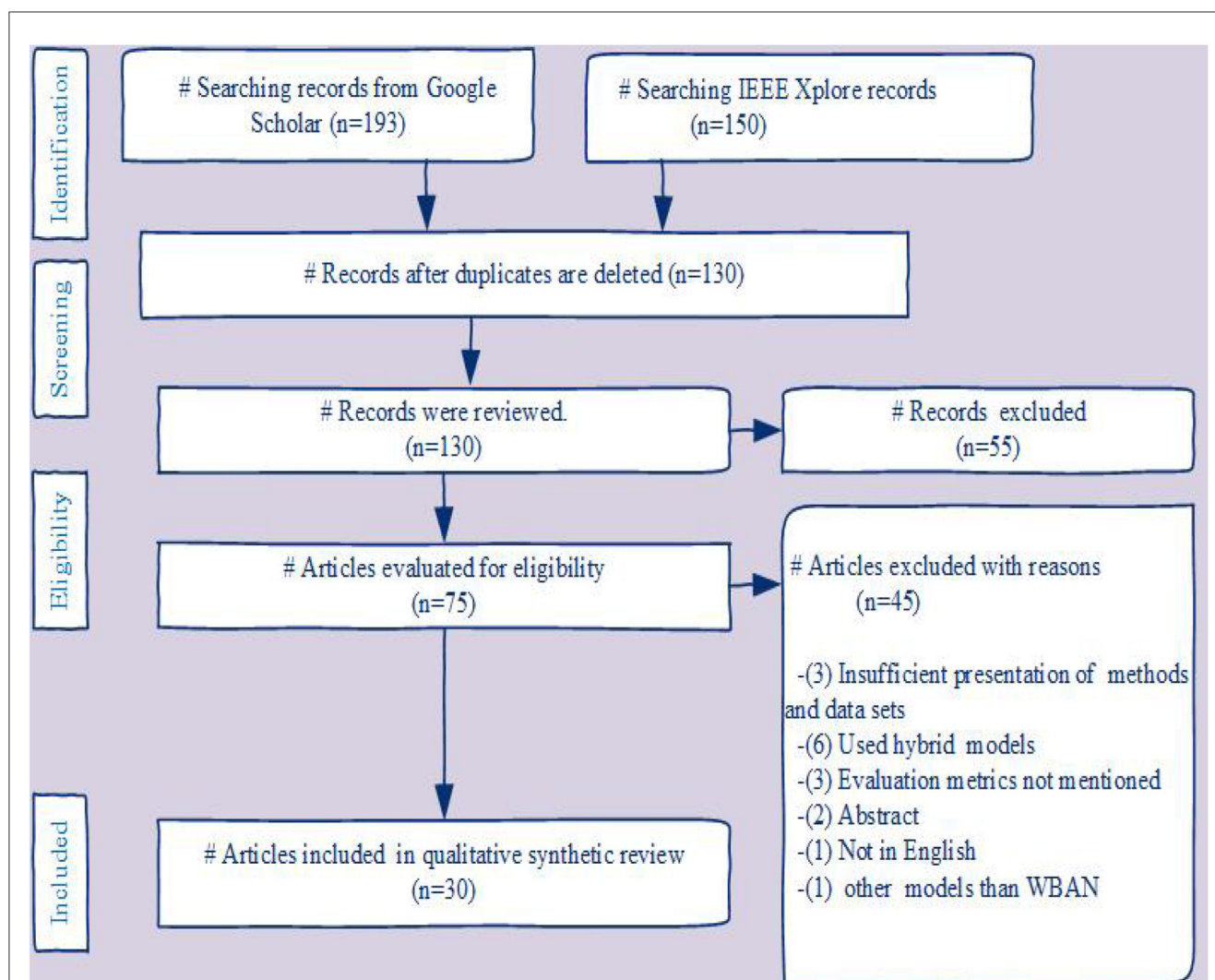


FIGURE 1

The PRISMA flow diagram depicts the procedure by which we searched the literature, removed duplicate records, screened, excluded, and included articles in our systematic review.

connection gateway sensors with a lightweight encryption protocol that consumes minimal power. The research goal should be to provide protection and authorization processes to ensure that during the entire communication route, the data is not disclosed to an external observer nor damaged by a malicious sensor inside or in the vicinity of the WBAN. Lightweight encryption protocol, low energy consumption, a wider system for different medications, and an end-to-end safe communication network for a PMS are some of the key objectives of this study. The wireless body area network (WBAN) which is also used to capture the sensitive medical information of the patients is the access network of the users through a server in which the data of the patients is processed. In comparison to the literature approaches, the research work (8) aims to have less computational time complexity and a cost-effective genetic-based algorithm. This method also introduces a new algorithm for a key generation that has fewer steps and fewer computational methods. After generating the patient's data, the genetic-based lightweight encryption algorithm was applied over the nanosensors

units. Genetic-based light encryption algorithm applied after producing the data of patients and over nanosensors devices. The encrypted information is then transmitted to the server, which further transforms it through a wireless network. Patients can also be tested with remote medical nanosensors nowadays, also for the collection of ongoing patients records, WBAN includes connected small sensors that are distributed via the networks for further processing. Cloud-based WBAN has recently gotten a lot of interest, but the cloud has many disadvantages in terms of data management and security. Consider these issues using the Advanced Encryption Standard (AES) and the Genetic Algorithm, this (Shanmugavade) research provides improved data security and efficient task flow scheduling (GA). WBAN should address two critical criteria to deliver reliable services data security and privacy. Fake data and information in medical records can lead to major problems. If a person alters the values of gathered information and the physician prescribes medication based on the changed information, significant health problems and even death might occur (18).

2.1.3 Data authentication

An essential component of Wireless Body Area Networks (WBAN) security and privacy protocols is the authentication of sensitive health-related data transmitted through the network. It is essential that both WBAN nodes have data authentication, and the coordinator must be able to confirm that the data is being sent from a reliable source and not a fraudster. Symmetric approaches, which generate the MAC (Message Authentication Code) of the whole set of data using a shared key, are used to verify the source of the data (25).

2.1.4 Data authorization

A user's identity, role, or permissions determine which data or resources they can access or cannot access. This process is known as data authorization. It involves verifying users' identities and figuring out if they are authorized to view, edit, or remove data (25).

2.1.5 Block cipher based schemes

This research work (7) aims to defend from various known cyberattacks, in particular, the vulnerability attack also on the base station and the dos attacks on the sensor node. These research findings and safety review show that in terms of storage needs, computing, and communication costs, the suggested improved system has overcome various established gaps. The goal of this paper is to establish a framework for safeguarding patients' health data from all safety difficulties. Requirements for storage, cost of computing and connectivity, time, and cost of computing. The suggested security system demonstrates its effectiveness in protecting against various known cyber-attacks, especially the compromise attack on the base station and the doc attacks on the sensor node. This paper's (15) goal is to propose a cost-effective framework that prevents unauthorized attackers from removing data packets or forwarding false data. This paper's goal is to present novel data protection mechanisms for WBAN that are capable of detecting getting into trouble relay nodes or links. The process refers to the routing algorithm for AOVD. The non-homogenous pattern of Poisson is used here to describe the possibility of malicious actions. The protection does not add any new packets of controls. To access performance, SLR on AODV is simulated and the results are compared with AODV. At a low cost, it is used to detect harmful intruders. The wireless body area networks (WBAN) are common options for a wide variety of health, sports activities, and recovery current study applications. In providing secure identification using an encryption mechanism, some existing WBAN routing protocols can be found, but they do not provide a lightweight communication solution. An energy-efficient framework is proposed in this paper that stops unauthorized intruders by dropping data packets or forwarding fake data. While it can communicate with any other reactive WBAN routing algorithm, the algorithm can be applied on the Adhoc On-demand distance-vector machine (AODV) protocol. In detecting malicious nodes with minimal latency, the protocol is simulated and results show its effectiveness.

2.1.6 ZigBee

This research paper (13) first gives an overview of WBAN, how it was used for medical surveillance, then highlights its design, significant security, and privacy specifications, and attacks on specific network layers in a WBAN, and finally talks about different encryption protocols and laws to provide WBAN data protection solutions. Provides WBAN protections sensors are used to capture a patient's confidential and valuable medical data, are they may even be used in sports. WBANs connect with the device and other applications such as ZigBee, WI-FI, cellular networks, and applications for the wireless personal area network (WPAN). The wireless body area network is a series of wireless sensors that can be mounted in or out of the body of the human or living person, thus detecting or tracking the body's functions and adjacent circumstances.

2.1.7 BAN detection

This paper (4) aims to review BAN communication standards, security risks, and BAN-based applications weaknesses, as well as current privacy and security processes. Privacy and security problems and the internet technology used in a BAN are outlined in the report. This technology aims to increase the quality of medical services rendered and reduce certain related costs. BAN has a wide variety of uses, such as tracking the health conditions of patients and optimizing the response to treatment plans, but protection and safety are among the main concerns in BAN-based healthcare systems at the same time, as medical data must be kept protected from adverse reactions and threats during stroke and transmission. Reducing healthcare cost, and energy-efficient climate, protocols for energy-efficient protection. Many studies have shown that if diseases are identified in their initial phases, there is a way to detect them.

2.1.8 Hashing algorithm

This paper (14) aims to design Safe hashing algorithms (SHA) and encryption techniques used in research reviews to make data transfer more secure and efficient (14). It creates digital signatures using a hash method to move patient data more stably and authentically. This proposed algorithm makes use of an asymmetric key generation technique, which uses a pair of public and private keys, making the algorithms slow and more complex. Protecting Data Communication in WBAN through Digital Signatures, the proposed technique is based on a combination of different methods for securing data in WBAN by using protected keys and digital signatures. BNC digitally signs each data packet to SK and sends it to all sensor nodes in the network. WBAN (Wireless Body Area Network) is a special form of sensor network that connects patients with medical service providers via the Internet to exchange crucial health data. WBAN offers several advantages, including location-independent monitoring, no influence on patients' movement, early illness diagnosis and prevention, remote patient support, and so on. To ensure security, researchers have proposed several health data transmission techniques. The author (Soni) proposes a low-cost health authentication and key agreement technique that is both secure and lightweight. The suggested protocol uses a one-way hashing algorithm (SHA-256), and the National Institute of

Standards and Technology (NIST) has determined that it is safe against the polynomial-time method (20).

2.1.9 Multiple scheme

Mehmood et al. (26) aims to design a framework for the portable authentication process and session key arrangement between sensor nodes and health professionals that discuss both patterns of communication. The safety review shows that required security features are maintained. The purpose of this paper is to implement lightweight user security mechanisms that facilitate internal and external information exchange to build a safe session key between a health professional and a particular sensor node linked to the body of the patient. In the future, the scheme will be applied in an actual system in which the sensor nodes mostly on the patient's body communicate with mobile devices, cloud services, authentic gateway, and health professionals. The wireless body area network (WBAN) is also an IOT-based health service that greatly improves health treatment by allowing patients' health conditions to be tracked remotely. This paper (12) aim is an attempt to examine that IOT based WBAN security infrastructure on a base of the main security agreement scheme. Key encryption techniques are extremely inefficient in terms of computing, processing, and energy usage. In tier 1 of WBAN, this paper mainly focuses on various primary agreement frameworks. Four different groups separate the private key agreement schemes, conventional key framework, physiological key strategy, hybrid scheme, and private key agreement strategy. The Internet of things (IoT) (27) is one of the newest technologies these days that has consumed a lot of possibilities. Wireless body area network (WBAN) also is such emerging field that provides a remote ability to prevent and collect patients' health data using IoT based wearable biosensors. In IoT devices that are extremely resources constraints, their architectures are discovered to be ineffective. This study is an attempt to examine the IoT-based. The goal of this paper (12) is to create a polynomial-based curve for a safe system that helps the patient with dignity, authenticity, confidentiality, and privacy. An attacker can access the medical data of the patients that are stored in the controller or hack the data while communicating through wireless communication, without any of the patient's permission. An attacker can alter the message produced within the BAN before they are transmitted to a receiver (such as location, layout, quality, query, etc.) or change the communication content being transferred from the BAN to an external entity(e.g doctor). Farooq et al. (28) proposed a method to secure physical layer (PHY) transmission. This approach encrypts data without requiring the keys. Physical Layer Security The sensor nodes in multi-hop WBAN use the MTFG (Multi-Hop Topology Formation Game) algorithm to create a spanning tree for multi-hop communication in the uplink of the WBAN. This algorithm can be implemented in a distributed manner, among each sensor being aware of the presence of its neighbors to choose the best direction. The system's performance is evaluated in a variety of situations, and the results show that the suggested scheme has the best performance, which can be tailored to meet the competing needs of protection and latency for different applications. This article offers software-defined networking (SDN)-based WBAN (SDWBAN) architecture

for application-specific traffic control to address these challenges. The suggested system achieves high throughput and low latency for emergency traffic in SDWBANs, according to the results of the paper's experiments. The objective of this paper is a scalable and adaptable SDWBAN framework that allows for dynamic network control as the number of apps on the network grows (traffic management) (9). WBAN is a sensor network with nodes that may be attached outside or within the body. Priority aware protocol (PAP) was proposed in this (Sandhu) paper to deal with smart healthcare systems. PAP is made up of three primary components: sensor, controller, and medical server. The sensor module detects the data, assigns a dynamic priority to the data packet based on the estimated values, and then delivers it to the controller unit according to the data packet's determined priority. The major goal of this article is to send data from a node to a coordinator node and then to a patient database in a timely and reliable manner (21). Radio waves on the receiving end are used to calculate the (RSSI). RSSI used 128 bits of size for data. Implementation of RSSI is complex and requires high memory. The restricted data density of RSSI-based key generation and agreement is a major problem. Unlike them, the research presents a physical layer-based security strategy in this work that uses physical channel information and eliminates the need for additional hardware (22).

2.1.10 Blockchain

WBAN provides a quick approach to gathering patient data, but they also introduces severe issues, the most important of which is the secure storage of the data obtained. WBAN devices' data storage and data security do not fulfill the demands of WBAN customers. As a result, the (Ren) paper uses a blockchain database to collect data, which increases the data's security. In addition, the research paper solution proposes a blockchain-based storage architecture for WBAN. The blockchain's storage space is limited, and the data it stores is exposed to unwanted access. To address these issues, the article presented a sequential aggregate signature method with a specified verifier (DVSSA), which ensures that a user's data may only be read by the authorized person and protects WBAN users' privacy (16).

2.1.11 Characteristic of the channel

The essential component of many telemedicine applications, such as customized medicine and home-based smartphone apps is a wireless body area network (WBAN) that uses wireless media to offer data transfer services. WBAN is an important field that is used to transmit patients related important information. Because of the wide accessibility of media in WBAN, malicious tapping or tampering attacks can readily occur, stealing personal information or introducing incorrect data. To avoid this type of attack (Zhang) proposed a mechanism that is used to channel characteristic aware privacy protection method for WBAN is suggested to improve user privacy at a relatively low cost and with great flexibility. Tempering attacks, malicious node attacks, and inserting fake data attacks may all be possible as a result of the great accessibility of resources.

2.1.12 Survey scheme

The systematic literature review (SLR) of multiple protection schemes for WBAN is presented in this survey paper. The study came up with a research question to look at the possibilities of multiple attacks while keeping memory constraints in mind. The study used quality valuation to ensure that the schemes were relevant to the research question. Furthermore, the schemes are examined from 2016 to 2020 to concentrate on recent work. Several current systems are investigated in the literature to determine how the protection of sharing patients' healthcare data can be improved. The study degree of confidence and satisfaction required by patients (29). Also examines the protection of various attack scenarios. The efficient transmission of data over a wireless channel may be disrupted by a variety of attacks (29). Existing studies include an overview of data protection in the medical environment, but the research concentrated on data security schemes in WBAN that reduce various attacks to provide the degree of confidence and satisfaction required by patients (29). WBAN is highly beneficial in today's environment, but it faces a variety of issues that must be overcome before it can be used. This (Singh) research considers different systems and protocols for dealing with energy efficiency, security, and privacy in depth. WBAN is a type of Wireless Sensor Network that comprises tiny bio-medical types of equipment known as nodes that are dedicated to guaranteeing continuously patient monitoring based on certain essential criteria. Because of its benefits, including portability, flexibility, and simplicity of patient monitoring, smart healthcare has gotten a lot of attention. WBAN is made up of a variety of heterogeneous devices, thus the amount of data and bandwidth required varies depending on their characteristics (19). First, the research discussed the article title, year, and references in this research Table 1. Then discuss the objective of each paper and also the technique or methodology. Finally, critical analyses were discussed for each of them. Gathered a variety of literary techniques to give us the ability to come up with new ways to defend against attacks that are vulnerable to the schemes. Because of their complex algorithms, the majority of research methods are time and cost-intensive. AES is difficult to implement on software in a way that is both fast. Table 2 compares the security of various schemes in the literature based on eavesdropping (30), denial of service (DoS) (31), malicious nodes (32), and execution time and cost. Several schemes have been proposed in the literature to examine the strengths of these security mechanisms to reduce attacks in the WBAN scenario.

3 Research gap

The research gaps according to the literature review are reported here. In the research Papers (7, 13, 29) uses the AES algorithm and that is a very complex algorithm. AES is a complex and costly algorithm and not suitable for sensor networks. This algorithm is complicated to implement. Encryption is difficult with large key sizes. Furthermore, decrypting data with this algorithm takes a longer time. And also these schemes are affected by DOS and IoT-based attacks. In this research paper (22) RSSI scheme is presented and this algorithm is difficult to implement because it is complex. In research paper (21) PAP(Priority Aware Protocol) as the priority level rises, so does the amount of energy

TABLE 2 Comparative analysis of techniques.

Scheme	Time	Eavesdropping	DOS	MN
Chowdhury et al. (11)	+	+	-	-
Rehman et al. (7)	+	-	+	-
Shanthapriya and Vaithianathan (12)	-	+	-	+
Braham et al. (4)	-	-	-	-
Jabeen et al. (8)	+	-	-	-
Jabeen et al. (8)	-	-	-	+
Mehmood et al. (26)	+	+	-	-
Parvez et al. (10)	+	-	-	-
Roy et al. (15)	+	-	+	-
Anwar et al. (14)	+	+	-	-
Farooq et al. (28)	-	-	+	+
Jabeen et al. (29)	+	+	-	-
Hasan et al. (9)	+	-	+	+
Ren et al. (16)	-	-	-	+
Zhang and Ma (17)	-	+	-	+
Shanmugavadivel et al. (18)	+	-	-	-
Singh and Prasad (19)	+	-	-	-
Soni and Singh (20)	+	-	-	+
Sandhu and Malik (21)	+	+	-	-
Li et al. (22)	-	-	-	-

expended. In the paper, Ren et al. (16) DVSSA has proposed it is a time-consuming and complex technique also data tampering attack is possible on it. In paper, Roy et al. (15) SDN(Software Defined Networking) is proposed. This scheme is affected by two attacks and that is DOS, MINA. Vulnerabilities in security, and inconsistency SDESW's flow demands rise as the network becomes more complex. In this research paper (20) LAKA is presented and that is increased energy use.

4 Conclusions

The protection of data in WBAN is important and should not be neglected. WBAN is used for gathering the medical conditions of patients and is sent to any portable device that is linked to databases that can store patient details. Because of the critical importance of the health issue, it must be kept hidden from unauthorized persons. In addition to highlighting security and privacy problems, a number of approaches for a WBAN utilizing IoT systems are thoroughly evaluated. Only a few research methodologies are considered viable due to the multifaceted nature of WBAN, and there are some extremely challenging and difficult research methodologies. This literature focuses on various approaches to

information security however, only a few are considered to be superior to others in terms of information security. Various current strategies are observed in the literature to understand how the security of patient's health data is upgraded.

Author contributions

WJ: Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Writing – original draft. AT: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft. JL: Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science

Foundation of China (Grant No. 61370073), the National High Technology Research and Development Program of China, and the project of the Science and Technology Department of Sichuan Province (Grant No. 2021YFG0322).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Priyadarshi R. Energy-efficient routing in wireless sensor networks: a meta-heuristic and artificial intelligence-based approach: a comprehensive review. In: *Archives of Computational Methods in Engineering*. (2024). p. 1–29. doi: 10.1007/s11831-023-10039-6
- Kumar R, Mukesh R. State of the art: Security in wireless body area networks. *Int J Comput Sci Eng Technol*. (2013) 4:622–30.
- Pattanayak A, Dhal S, Addya SK. Automatic privacy-preserving contact tracing of novel coronavirus infection by cloud-enabled wban using blockchain. *Cryptology ePrint Archive*. (2020).
- Braham TG, Butakov S, Ruhl R. Reference security architecture for body area networks in healthcare applications. In: *2018 International Conference on Platform Technology and Service (PlatCon)*. IEEE (2018). p. 1–6. doi: 10.1109/PlatCon.2018.8472765
- Oyewole AT, Oguejiofor BB, Eneh NE, Akpuokwe CU, Bakare SS. Data privacy laws and their impact on financial technology companies: a review. *Comput Sci IT Res J*. (2024) 5:628–50. doi: 10.51594/csitj.v5i3.911
- El Mestari SZ, Lenzini G, Demirci H. Preserving data privacy in machine learning systems. *Comput Secur*. (2024) 137:103605. doi: 10.1016/j.cose.2023.103605
- Rehman ZU, Altaf S, Iqbal S. An efficient lightweight key agreement and authentication scheme for WBAN. *IEEE Access*. (2020) 8:175385–97. doi: 10.1109/ACCESS.2020.3026630
- Jabeen T, Ashraf H, Khatoon A, Band SS, Mosavi A. A lightweight genetic based algorithm for data security in wireless body area networks. *IEEE Access*. (2020) 8:183460–9. doi: 10.1109/ACCESS.2020.3028686
- Hasan K, Ahmed K, Biswas K, Islam MS, Sianaki OA. Software-defined application-specific traffic management for wireless body area networks. *Fut Gener Comput Syst*. (2020) 107:274–85. doi: 10.1016/j.future.2020.01.052
- Parvez K, Zohra FT, Jahan M. A secure and lightweight user authentication mechanism for wireless body area network. In: *Proceedings of the 6th International Conference on Networking, Systems and Security*. (2019). p. 139–143. doi: 10.1145/3362966.3362981
- Chowdhury FS, Istiaque A, Mahmud A, Miskat M. An implementation of a lightweight end-to-end secured communication system for patient monitoring system. In: *2018 Emerging Trends in Electronic Devices and Computational Techniques (EDCT)*. IEEE (2018). p. 1–5. doi: 10.1109/EDCT.2018.8405076
- Shanthapriya R, Vaithianathan V. ECG-based secure healthcare monitoring system in body area networks. In: *2018 Fourth International Conference on Biosignals, Images and Instrumentation (ICBSII)*. IEEE (2018). p. 206–212. doi: 10.1109/ICBSII.2018.8524714
- Malik MSA, Ahmed M, Abdullah T, Kousar N, Shumaila MN, Awais M. Wireless body area network security and privacy issue in e-healthcare. *Int J Adv Comput Sci Applic*. (2018) 9:433. doi: 10.14569/IJACSA.2018.090433
- Anwar M, Abdullah AH, Butt RA, Ashraf MW, Qureshi KN, Ullah F. Securing data communication in wireless body area networks using digital signatures. *Tech J*. (2018) 23:50–5.
- Roy M, Chowdhury C, Kundu A, Aslam N. Secure lightweight routing (SLR) strategy for wireless body area networks. In: *2017 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*. IEEE (2017). p. 1–4. doi: 10.1109/ANTS.2017.8384119
- Ren Y, Leng Y, Zhu F, Wang J, Kim HJ. Data storage mechanism based on blockchain with privacy protection in wireless body area network. *Sensors*. (2019) 19:2395. doi: 10.3390/s19102395
- Zhang P, Ma J. Channel characteristic aware privacy protection mechanism in WBAN. *Sensors*. (2018) 18:2403. doi: 10.3390/s18082403
- Shanmugavadivel G, Gomathy B, Ramesh S. An enhanced data security and task flow scheduling in cloud-enabled wireless body area network. *Wirel Pers Commun*. (2021) 120:849–67. doi: 10.1007/s11277-021-08493-1
- Singh S, Prasad D. Wireless body area network (WBAN): a review of schemes and protocols. *Mater Today*. (2022) 49:3488–96. doi: 10.1016/j.matpr.2021.05.564
- Soni M, Singh DK. LAKA lightweight authentication and key agreement protocol for internet of things based wireless body area network. *Wirel Pers Commun*. (2022) 127:1067–84. doi: 10.1007/s11277-021-08565-2
- Sandhu A, Malik A. PAP priority aware protocol for healthcare applications in wireless body area network. *Parameters*. (2020) 2:3.
- Li Z, Wang H, Fang H. Group-based cooperation on symmetric key generation for wireless body area networks. *IEEE Internet Things J*. (2017) 4:1955–63. doi: 10.1109/JIOT.2017.2761700
- Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group* t. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Internal Med*. (2009) 151:264–9. doi: 10.7326/0003-4819-151-4-200908180-00135
- Derbez P, Fouque PA, Isobe T, Rahman M, Schrottenloher A. Key committing attacks against AES-based AEAD schemes. *IACR Trans Sym Cryptol*. (2024) 2024:135–57. doi: 10.46586/tosc.v2024.i1.135-157
- Narwal B, Mohapatra AK. A survey on security and authentication in wireless body area networks. *J Syst Arch*. (2021) 113:101883. doi: 10.1016/j.sysarc.2020.101883
- Mehmood G, Khan MZ, Waheed A, Zareei M, Mohamed EM. A trust-based energy-efficient and reliable communication scheme (trust-based ERCS) for remote

patient monitoring in wireless body area networks. *IEEE Access*. (2020) 8:131397–413. doi: 10.1109/ACCESS.2020.3007405

27. Hasan MK, Weichen Z, Safie N, Ahmed FRA, Ghazal TM. A survey on key agreement and authentication protocol for internet of things application. *IEEE Access*. (2024). doi: 10.1109/ACCESS.2024.3393567

28. Farooq S, Prashar D, Jyoti K. Hybrid encryption algorithm in wireless body area networks (WBAN). In: *Intelligent Communication, Control and Devices: Proceedings of ICICCD 2017*. Springer (2018). p. 401–410. doi: 10.1007/978-981-10-5903-2_41

29. Jabeen T, Ashraf H, Ullah A. A survey on healthcare data security in wireless body area networks. *J Ambient Intell Humaniz Comput*. (2021) 12:9841–54. doi: 10.1007/s12652-020-02728-y

30. Dai HN, Wang Q, Li D, Wong RCW. On eavesdropping attacks in wireless sensor networks with directional antennas. *Int J Distr Sensor Netw*. (2013) 9:760834. doi: 10.1155/2013/760834

31. Welteji BT, Tiwari B, Kebede SD, Gupta S, Tiwari V. DDoS attack detection using predictive machine learning (ML) algorithms in wireless body area network environments. In: *IoT in Healthcare Systems*. CRC Press (2023). p. 191–216. doi: 10.1201/9781003145035-11

32. Kumar M. *Secured Key Agreement Schemes in Wireless Body Area Network-A Review*. Available at SSRN 4213580. (2022). doi: 10.2139/ssrn.4213580



OPEN ACCESS

EDITED BY

Amin Ul Haq,
University of Electronic Science and
Technology of China, China

REVIEWED BY

Tassadaq Hussain,
King's College London, United Kingdom
Jan Kubicek,
VSB-Technical University of Ostrava, Czechia
Fatma Elsarkawy,
Suez Canal University, Egypt

*CORRESPONDENCE

Imad Ali
✉ imad.ali@uswat.edu.pk

RECEIVED 03 June 2024

ACCEPTED 06 August 2024

PUBLISHED 29 August 2024

CITATION

Ali I, Saleem N, Alhussein M, Zohra B,
Aurangzeb K and Haq QMu (2024)
DeepCGAN: early Alzheimer's detection with
deep convolutional generative adversarial
networks. *Front. Med.* 11:1443151.
doi: 10.3389/fmed.2024.1443151

COPYRIGHT

© 2024 Ali, Saleem, Alhussein, Zohra,
Aurangzeb and Haq. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC
BY\)](#). The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

DeepCGAN: early Alzheimer's detection with deep convolutional generative adversarial networks

Imad Ali^{1*}, Nasir Saleem², Musaed Alhussein³, Benazeer Zohra^{4,5},
Khursheed Aurangzeb³ and Qazi Mazhar ul Haq⁶

¹Department of Computer Science, University of Swat, Swat, KP, Pakistan, ²Department of Electrical Engineering, Faculty of Engineering & Technology (FET), Gomal University, Dera Ismail Khan, Pakistan, ³Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia, ⁴Department of Anatomy, School of Medical Sciences and Research, Sharda University, Greater Noida, UP, India, ⁵Department of Anatomy, Noida International Institute of Medical Sciences, Noida International University, Greater Noida, UP, India, ⁶Department of International Bachelor Program in Informatics and Computer Science and Engineering, Yuan Ze University, Taoyuan City, Taiwan

Introduction: Alzheimer's disease (AD) is a neurodegenerative disorder and the most prevailing cause of dementia. AD critically disturbs the daily routine, which usually needs to be detected at its early stage. Unfortunately, AD detection using magnetic resonance imaging is challenging because of the subtle physiological variations between normal and AD patients visible on magnetic resonance imaging.

Methods: To cope with this challenge, we propose a deep convolutional generative adversarial network (DeepCGAN) for detecting early-stage AD in this article. The DeepCGAN is an unsupervised generative model that expands the dataset size in addition to its diversity by utilizing the generative adversarial network (GAN). The Generator of GAN follows the encoder-decoder framework and takes cognitive data as inputs, whereas the Discriminator follows a structure similar to the Generator's encoder. The last dense layer uses a softmax classifier to detect the labels indicating the AD.

Results: The proposed model attains an accuracy rate of 97.32%, significantly surpassing recent state-of-the-art models' performance, including Adaptive Voting, ResNet, AlexNet, GoogleNet, Deep Neural Networks, and Support Vector Machines.

Discussion: The DeepCGAN significantly improves early AD detection accuracy and robustness by enhancing the dataset diversity and leveraging advanced GAN techniques, leading to better generalization and higher performance in comparison to traditional and contemporary methods. These results demonstrate the efficacy of DeepCGAN in enhancing early AD detection, thereby potentially improving patient outcomes through timely intervention.

KEYWORDS

GAN, CNN, Alzheimer's disease, deep learning, cognitive features

1 Introduction

Alzheimer's disease (AD) is a neurodegenerative condition primarily affecting the elderly, characterized by memory, behavioral, and cognitive impairments that disrupt daily life (1). This devastating disease is projected to have a staggering impact on global health in the coming decades. Epidemiological studies indicate a disturbing trend, with expectations of a fourfold surge in the worldwide prevalence of AD by 2050,

potentially exceeding 100 million cases (2). The impending prevalence of AD raises critical concerns for individuals, families, and healthcare systems worldwide. The burden of AD extends beyond the individual, affecting the very fabric of society. Some studies employ mathematical modeling to forecast the trends and growth of AD, considering factors such as increasing life expectancy, shifting mortality patterns, and the prevalence of cardiovascular diseases. Unfortunately, these projections collectively suggest a growing proportion of the population will be impacted by AD in the future (3).

Detecting AD in its early stages is of paramount importance for effective intervention and treatment. AD diagnosis is a complex endeavor, demanding the accurate identification of different dementia subtypes (4). While the challenges are substantial, recent research highlights the central role of AD in dementia cases, constituting approximately two-thirds of all diagnoses (5). One of the pressing issues in AD management is the lack of effective pharmacological treatments in clinical practice. This shortfall has prompted a paradigm shift in therapeutic strategies, emphasizing the early-stage detection of AD as a promising avenue for intervention (6, 7). Identifying individuals in the early stages of cognitive decline or Mild Cognitive Impairment, whether stable or progressive, is pivotal for understanding high-risk populations and potentially delaying AD progression. The combination of the increasing prevalence of AD and the complexity of its diagnosis underscores the urgent need for advancements in early detection methods and comprehensive care strategies to address the growing global challenge of AD.

The AD research landscape has shifted significantly due to deep learning (DL) models, including stacked auto-encoders, recurrent neural networks, support vector machines, and convolutional neural networks (CNN). The bi-directional gated recurrent units (BiGRUs) layers consist of 2,048 units, with 1,024 units in each direction. The BiGRUs capture long-term temporal cues from the cognitive data, which is crucial for identifying patterns and changes indicative of early AD which have emerged as potent tools in this endeavor (8). However, limitations exist in feature quality, especially from image processing (9), requiring DL models adaptable to diverse data types. Simultaneously, generative adversarial networks (GANs), (which is a class of machine learning frameworks where two neural networks, a generator, and a discriminator, compete against each other to produce more accurate results) originally designed for images, have found their place in AD classification (9). DL models with GANs are proficient in classifying AD states and enhancing image-based AD tasks, like denoising images and precise brain segmentation (10, 11). These advances drive understanding, detection, and treatment of AD, a pressing neurodegenerative disease. Although, these architectures have made sufficient advancement in AD detection; however, these existing AD detection models have primarily focused on neuroimaging data, resulting in the underutilization of critical cognitive features. Moreover, temporal information, which is highly relevant for understanding AD progression, has been largely neglected in the literature. Additionally, the well-known challenge of training instability in these models remains a significant concern.

Existing models often struggle with limited dataset sizes and lack diversity, leading to overfitting and poor generalization. Traditional GAN-based methods, primarily designed for image data, fail to leverage cognitive data crucial for early AD detection. This article introduces a groundbreaking method for the early detection of AD—the deep convolutional generative adversarial network (DeepCGAN), which is an unsupervised generative model designed to leverage cognitive (clinical) data for AD detection. DeepCGAN addresses these issues by using a deep convolutional GAN framework to expand and diversify the dataset, generating high-quality synthetic data that improves detection accuracy and robustness. DeepCGANs generate high-quality synthetic medical images, crucial for augmenting limited datasets like MRI and PET scans and enhancing model generalization. They create diverse synthetic samples, augmenting training data in medical imaging where labeled data is scarce, improving model performance. DeepCGANs' convolutional layers learn complex features for accurate early Alzheimer's detection, and their flexibility across imaging modalities makes them versatile beyond disease detection, which makes DeepCGANs a powerful and effective choice for early AD detection.

To address the aforementioned gaps, the proposed model effectively incorporates and analyzes cognitive data, offering a more comprehensive understanding of AD. Also, the proposed model integrates temporal information using BiGRUs to capture long-term patterns and introduces mechanisms like gradient penalty and relativistic average loss to stabilize training, thereby enhancing the stability and reliability of AD detection with GANs. Operating through a dual structure, the Generator follows an encoder-decoder framework that takes cognitive data as input, while the Discriminator mirrors the architecture of the Generator's encoder. Moreover, the proposed model employs two distinct loss functions, Wasserstein and Relativistic loss, ensuring stable training and improved performance. The pivotal component of the model is the last dense layer, employing a softmax classifier to detect AD labels. The proposed DeepCGAN undergoes comprehensive training using cognitive data, demonstrating promising results in the early prediction of AD, achieving a remarkable 97.32% accuracy on cognitively labeled data from the ADNI dataset, surpassing recent state-of-the-art models. The contributions of this article include:

1. For detecting early-stage AD, this article proposes a DeepCGAN, an unsupervised generative model that extends the cognitive features of the data and its diversity by utilizing the GAN framework.
2. To optimize the detection performance of DeepCGAN, a novel convolutional encoder-decoder-based GAN is proposed and trained on the cognitive features.
3. Our comprehensive experiments on the ADNI dataset show that the proposed DeepCGAN performs better in detecting early-stage AD compared to start-of-the-art models.

The remainder of this article is structured as follows: Section 2 reviews related work. Section 3 presents the proposed DeepCGAN. Section 4 describes the experimental setup, and Section 5 discusses the evaluation of the proposed model. Finally, Section 6 concludes the article.

2 Related work

The current gold standard for detecting and prognosing neurodegenerative AD relies on clinical assessments of symptoms and their severity. However, early disease detection before clinical symptoms manifest is critical for effective disease management and timely therapeutic intervention. Research indicates that analyzing structural and functional changes in patients during the early stages of AD can provide valuable insights (12). Machine learning approaches offer a rapid and robust means to interpret medical examinations, aiding in the early detection of AD. Early detection is paramount, allowing for proactive intervention and potentially improving patient outcomes. Machine learning enhances the diagnostic process by uncovering subtle patterns and anomalies that may precede clinical symptoms. It transcends the limitations of conventional clinical assessments, which often rely on symptomatic markers that become evident at later disease stages. Integrating machine learning into AD detection represents a paradigm shift, emphasizing the significance of early and accurate diagnosis in transforming AD research and treatment strategies.

CNNs are deep learning models (13) known for their ability to extract complex patterns (14–16). They excel in body part segmentation, surpassing traditional methods like logistic regression and support vector machines (17). CNN-based computer-aided diagnosis (CAD) systems are effective in neurodegenerative disease detection (18). In AD detection, methods combining the dual-tree complex wavelet transform with neural networks show promise (19). Architectures like GoogleNet and ResNet deliver strong results in distinguishing healthy subjects from those with AD and mild cognitive impairment (20). LeNet-5 has been effectively employed for AD vs. normal control (NC) brain classification (21). Hosseini et al. extended previous work by proposing a Deeply Supervised Adaptive 3D-CNN (DSA-3D-CNN) for AD prediction (22). They trained this model on the CAD-Dementia dataset without skull stripping preprocessing and rigorously evaluated its performance through 10-fold cross-validation. In addition to CNNs, ensemble learning (EL) has proven invaluable in the detection and prognosis of neurodegenerative diseases. Given the often limited availability and the inherent 3D nature of medical imaging data, training classifiers can be a challenge (23). EL offers a promising solution by combining the strengths of multiple trained models, making it particularly useful for classification tasks involving heterogeneous datasets. To harness the power of ensemble learning, individual classifiers are trained on various subsets of the data and subsequently combined. EL with bootstrapping techniques becomes especially beneficial when relevant data is scarce, such as cognitive features. Additionally, when dealing with limited data, common practices include data augmentation to enhance the performance of ensemble models. This combined approach of CNNs and ensemble learning offers a robust and adaptable framework for tackling the complexities of neurodegenerative disease detection and prognosis.

GANs are a prominent method for enhancing imaging data by creating synthetic data that competes with a discriminator aiming to distinguish real from synthetic data (24). When generative networks excel, they can replicate data based on the inherent

structure of real data. In the field of medical imaging, GANs have found success in tasks like MRI and CT reconstruction and unconditional image synthesis (25, 26). Furthermore, GANs exhibit a wide array of applications in AD-related image processing. They are proficient in denoising low-dose positron emission tomography (PET) scans to yield high-quality images (10, 11, 27). Accurate brain image segmentation, facilitated by GANs, aids in locating features critical for AD diagnosis and research across various image modalities (28–30). Despite the promise of GANs in AD image processing, the existing models for detecting AD have predominantly centered around neuroimaging data, leading to the insufficient utilization of vital cognitive features. Furthermore, the valuable temporal dimension, crucial for comprehending the progression of AD, has been notably overlooked in the existing literature. Additionally, the persisting issue of training instability in these models continues to pose a noteworthy challenge.

3 Materials and methods

The DeepCGAN model, proposed in this study, is designed for AD detection. It leverages a Generative Adversarial Network architecture, specifically tailored to the analysis of cognitive features and temporal information, which are often overlooked in existing AD detection models.

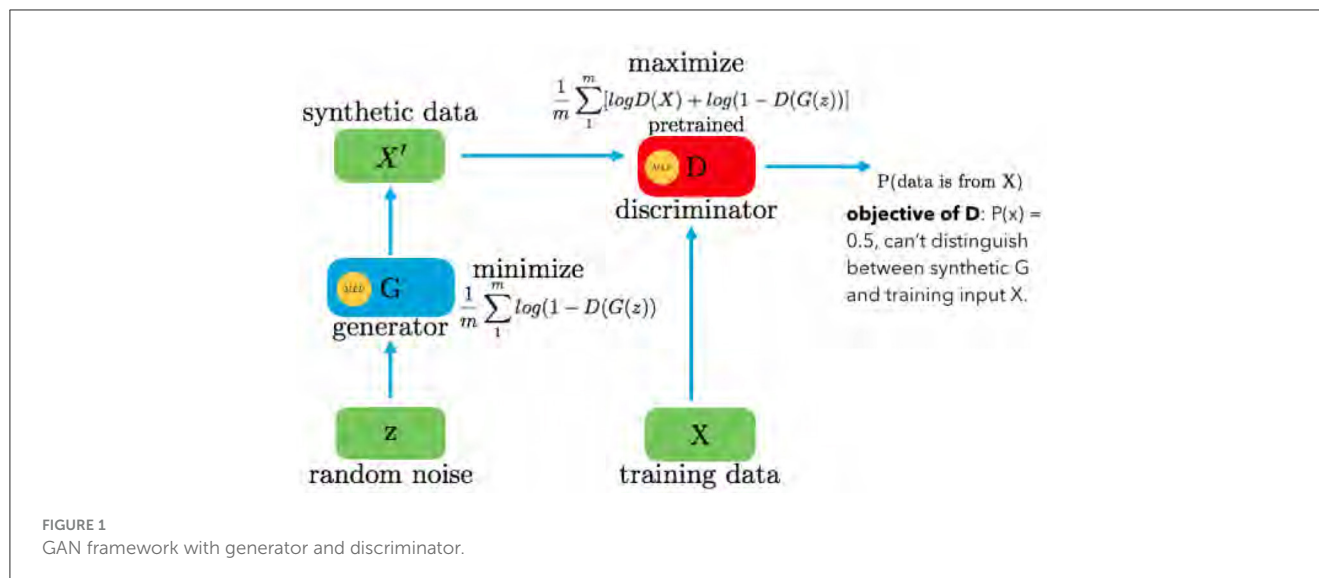
3.1 Generative adversarial networks

GAN is a fundamental architecture in machine learning, composed of two primary components: the Generator $G(z)$ and the Discriminator $D(x)$, as shown in Figure 1. The GAN framework is designed for generative tasks, aiming to produce synthetic data that closely resembles real data distributions. The Generator $G(z)$ is responsible for creating new data samples. It takes random noise $N(z)$ as input, typically drawn from a uniform or normal distribution. Through a learned transformation process, the Generator generates data that mimics real training data. This process relies on adjusting internal parameters to produce data samples that are increasingly realistic.

The Discriminator $D(x)$ acts as an adversary to the Generator. Its primary role is to differentiate between genuine data from the training set and data generated by the Generator. The Discriminator evaluates each input and assigns a probability score, indicating the likelihood of the input being real. If an input is genuine, $D(x)$ approaches 1, whereas if it is generated, $D(x)$ tends toward 0. The GAN operates as a two-player minimax game, optimizing the value function $V(G, D)$. The objective function is given in Equation 1:

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim P_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

Here, $D(\cdot)$ provides the probability that a given sample belongs to the training data X . The Generator aims to minimize $\log(1 - D(G(z)))$, making $D(G(z))$ as high as possible, essentially fooling



the Discriminator into considering $G(z)$ as real data. Conversely, the Discriminator seeks to maximize $D(X)$ and $1 - D(G(z))$, driving its optimal state toward $P(x) = 0.5$. In practice, GANs continually refine the Generator to produce data that is indistinguishable from real data, representing a powerful framework for generating synthetic data in various domains.

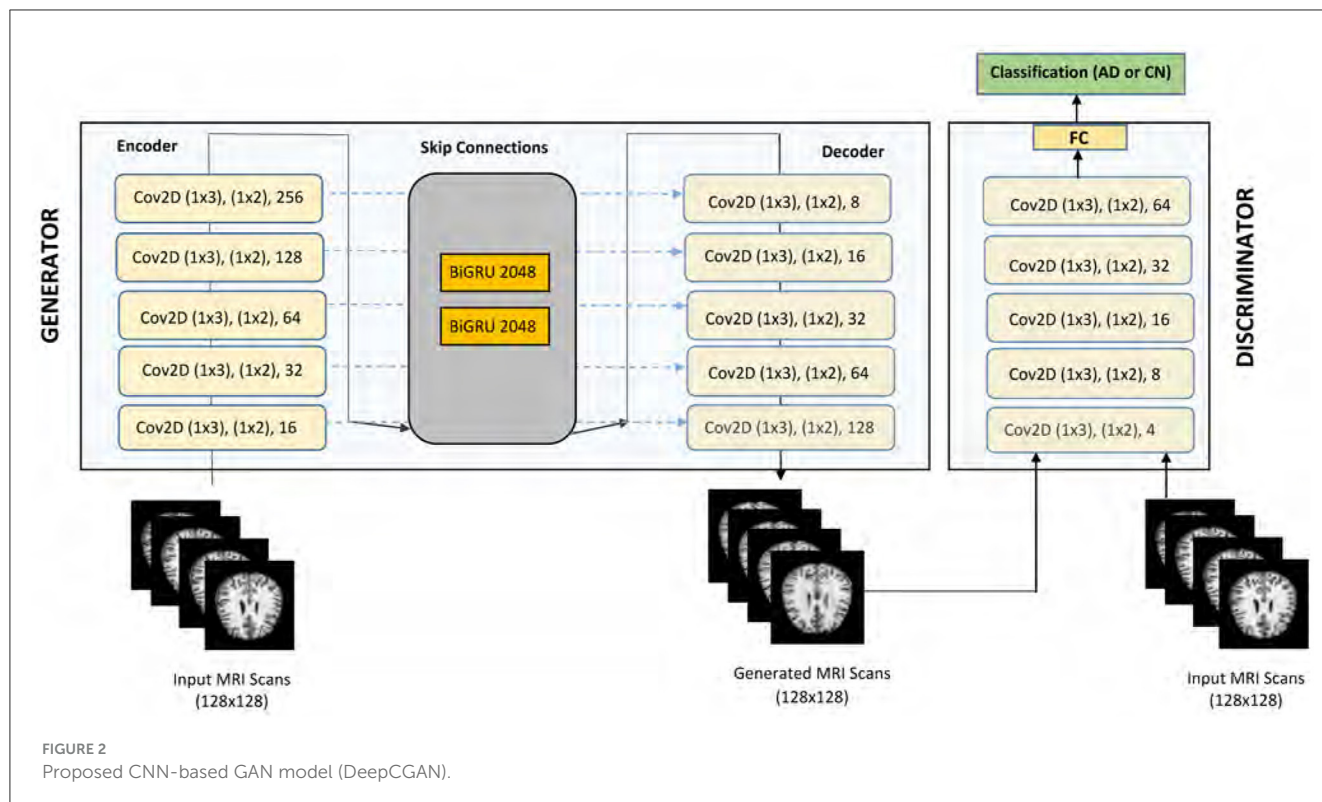
3.2 DeepCGAN for AD detection

The architecture of our proposed GAN model for AD detection is illustrated in Figure 2. This model is carefully designed to effectively utilize cognitive features in the detection process. The Generator component of our model is based on an encoder-decoder framework, optimized for processing cognitive features as inputs. The encoder in our model is designed to extract meaningful features from the input data through a series of convolutional layers. The encoder comprises five 2-D convolutional layers, strategically placed to extract local correlations within the input features. A reshape layer is employed to appropriately format the encoded features. These layers progressively downsample the input, capturing local correlations and essential patterns. Each convolutional layer is followed by batch normalization and Leaky Rectified Linear Unit (ReLU) activation functions to stabilize training and introduce non-linearity. Positioned in the middle of the Generator architecture, the BiGRU layers are crucial for capturing long-term dependencies and temporal dynamics in the cognitive features. Each BiGRU layer consists of 2,048 units (1,024 in each direction), enabling the model to learn bidirectional temporal patterns that are significant for early Alzheimer's detection.

The decoder mirrors the encoder's structure but performs the inverse operation. It utilizes deconvolutional (transposed convolution) layers to reconstruct the input data from the encoded features. The skip connections between corresponding layers of the encoder and decoder facilitate fine-grained feature integration,

enhancing the model's ability to preserve important information during reconstruction. The Generator's primary function is to produce synthetic data that closely resembles the real cognitive feature data. By transforming random noise inputs through the encoder-BiGRU-decoder pipeline, the Generator learns to create realistic data samples that help augment the training set and improve the robustness of the Discriminator. The input to our DeepCGAN model consists of cognitive features derived from the ADNI dataset. The input to the model is a three-dimensional tensor with a batch size of 32, 50-time steps, and 128 features. Thus, the input shape is [32, 50, 128], specifically tailored to capture the temporal and cognitive aspects critical for Alzheimer's detection.

The data preprocessing steps include normalization and sequence padding to ensure uniform input dimensions. The preprocessing steps include: Normalization: The cognitive features are normalized to ensure consistent scales and improve model training stability. Padding: Sequences are padded to a fixed length (e.g., 50 time steps) to ensure uniform input dimensions across different samples. Handling Missing Values: Features with more than 40% missing values are removed. For the remaining features, missing values are imputed using appropriate statistical methods (e.g., mean imputation). In addition, two BiGRUs layers are thoughtfully inserted in the middle of the Generator architecture, which enhances the model's ability to capture long-term temporal cues from the cognitive data. This integration addresses a critical gap in existing models that primarily focus on neuroimaging data, thereby improving the detection of early AD. The decoder of our model mirrors the encoder's structure and consists of five 2-D deconvolutional layers, also known as transposed convolution layers. Batch normalization is consistently applied following each convolutional and deconvolutional operation. ReLU functions are used as activation functions within the hidden layers, while a sigmoid activation function is applied to the output layer. To facilitate fine-grained feature integration, we have incorporated skip connections within the Generator. These skip connections concatenate the outputs of each convolutional layer in the encoder with the corresponding inputs of the deconvolutional layers in



the decoder. This design element enhances the model's ability to capture intricate feature cues effectively.

The Discriminator component, denoted as D , shares a similar structure with the encoder of the Generator. However, a flattened layer is introduced after the fifth convolutional layer to streamline feature processing. Finally, a fully connected dense layer with softmax activation is integrated into the Discriminator to enable classification tasks. Notably, the Discriminator provides two types of outputs, $D(y)$ and $D_k(y)$, with $D(y)$ representing sigmoidal output and $D_k(y)$ signifying linear output, linked by the sigmoid non-linearity function $\lambda(D_k(y)) = D(y)$. The proposed GAN model for AD detection leverages cognitive features and exhibits a sophisticated architecture, comprising convolutional, deconvolutional, and recurrent layers, skip connections, and a dual-output Discriminator. These design innovations collectively contribute to the model's efficacy in AD detection. Most AD detection models predominantly focus on neuroimaging data, neglecting cognitive features. Our model efficiently incorporates and exploits these underutilized data sources. By including BiGRUs, our model accounts for long-term temporal cues, a crucial aspect often overlooked in AD progression analysis. The Discriminator's architecture, featuring dual output types and skip connections, introduces novel enhancements to improve the model's performance in distinguishing real and synthetic data.

The Discriminator is designed to differentiate between real and synthetic data. It shares a similar structure with the encoder and includes an additional fully connected layer with softmax activation for classification. The dual outputs of the Discriminator, $D(y)$ and $D_k(y)$, provide sigmoidal and linear outputs, respectively, enhancing the model's ability to distinguish between genuine

and generated data. The DeeCGAN model is specifically tailored for Alzheimer's detection by focusing on cognitive features and temporal information, which are often underutilized in traditional models. By leveraging the strengths of DeeCGANs in generating realistic synthetic data and incorporating bidirectional GRUs for temporal analysis, our model is able to achieve high accuracy in early Alzheimer's detection.

3.3 Loss function

In the realm of GANs, choosing appropriate loss functions plays a pivotal role in achieving stable training and optimal performance. Our proposed GAN model incorporates and thoroughly investigates two distinct loss functions to determine the one that yields superior results.

3.3.1 Wasserstein loss

The Wasserstein loss, denoted as L_D for the Discriminator and L_G for the Generator, offers significant advantages in stabilizing and enhancing the robustness of GAN models (18). The Wasserstein loss function is used to train the DeepCGAN model due to its ability to provide a smoother gradient, leading to more stable training. This stability is crucial for AD detection, as it ensures that the model effectively learns from the subtle and complex patterns in the cognitive data. These loss functions are defined in Equations 2–4:

$$L_D = -\mathbb{E}_{y \sim P(y)}[D_k(y)] + \mathbb{E}_{x \sim P(x)}[D_k(G(x))], \quad (2)$$

$$L_G = -\mathbb{E}_{x \sim P(x)}[D_k(G(x))], \quad (3)$$

$$L_{GP} = \mathbb{E}_{\tilde{y} \sim \tilde{y}} \left[\left(\|\nabla_{\tilde{y}} D_k(\tilde{y})\|_2 - 1 \right)^2 \right], \quad (4)$$

where $\nabla_{\tilde{y}} D_k(\tilde{y})$ represents the gradient of the Discriminator output with respect to $\tilde{y}$.

3.3.2 Relativistic loss

The second loss function incorporated into our GAN model is the Relativistic loss. It computes the probability of real data features being classified as real and the probability of synthetic data features being classified as real. This is achieved by considering the difference between the Discriminator's outputs for real and synthetic input features. The loss functions for the Discriminator and Generator are given by Equations 5, and 6, respectively.

$$L_D = -\mathbb{E}_{(x,y) \sim P(x,y)} \left[(x, y) [\log(v D_k(y) - D_k(G(x)))], \quad (5)$$

$$L_G = -\mathbb{E}_{(x,y) \sim P_{data}(x,y)} \left[\log(v D_k^k(G(x)) - D_k(y)) \right] \quad (6)$$

However, the the Relativistic loss in Equations 5 and 6 exhibits high variance, primarily when the Generator significantly influences the Discriminator. To address this, we consider the average loss functions for the Discriminator and Generator are given by Equations 7, and 8, respectively.

$$L_D = -\mathbb{E}_{y \sim P(y)} [\log(D_{\tilde{y}}(y))] - \mathbb{E}_{x \sim P(x)} [\log(1 - D_{G(x)}(x))], \quad (7)$$

$$L_G = -\mathbb{E}_{x \sim P(x)} [\log(D_{G(x)}(x))] - \mathbb{E}_{x \sim P(x)} [\log(1 - D_{\tilde{y}}(y))], \quad (8)$$

where $D_{\tilde{y}}(y)$ and $D_{G(x)}(x)$ represent the relativistic Discriminator outputs for real and synthetic data, respectively. Thus, our GAN model incorporates both Wasserstein and Relativistic loss functions, each with its distinct advantages. These loss functions are carefully chosen and utilized to optimize the model's training stability and performance in AD detection.

We selected the Wasserstein loss and Relativistic loss due to their efficacy in stabilizing GAN training and enhancing the quality of generated data. The Wasserstein loss addresses mode collapse and provides meaningful gradients for GAN convergence. The Relativistic loss improves the model's discriminative power by comparing real and generated data in a relativistic manner, aligning with the goal of distinguishing subtle differences in medical data. These loss functions balance sensitivity and specificity in Alzheimer's detection, with the Wasserstein loss ensuring high-quality synthetic data and the Relativistic loss enhancing the discriminator's accuracy and robustness.

4 Experiments

This section provides insights into the dataset, experimental settings, and an evaluation of the proposed model.

4.1 Dataset

We utilized the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (20), consisting of three distinct stages. The ADNI dataset encompasses cognitive test scores and records of 5,013 instances, corresponding to 819 different AD patients. The cognitive features selected for this study include memory recall tests, attention assessments, and executive function evaluations, which are clinically relevant as they have been shown to be significant indicators of early cognitive decline associated with AD. Patients visited the clinic multiple times during clinical trials, resulting in new cognitive test scores generated and stored as additional records in the dataset for each visit. Among these records, there are 1,643 belonging to cognitively normal individuals and 3,370 related to AD patients. However, the dataset exhibited missing values and underwent initialization through an Iterative Imputer technique to impute the missing values using a round-robin method. This ensures that the most clinically significant features are retained and accurately represented in the dataset. The irrelevant features were removed during the data cleaning and preprocessing.

In the ADNI1 dataset, each record comprises 113 features. The data includes various cognitive assessments (e.g., MMSE scores, ADAS-Cog scores), biomarkers (e.g., cerebrospinal fluid biomarkers, amyloid-beta levels), and potentially neuroimaging features (e.g., MRI and PET scan data). These features are chosen for their relevance in assessing cognitive decline and AD progression. The input data is organized as a temporal sequence, capturing changes in cognitive features over time. This is crucial for modeling the progression of AD, which involves gradual cognitive decline. The dataset was divided into 80% for training and 20% for testing, resulting in 5,000 samples for training and 1,250 samples for testing. Some of these features had excessive missing values, prompting the removal of those with more than 40% missing values. The remaining features underwent initialization through an Iterative Imputer technique to impute the missing values using a round-robin method. After preprocessing, the final dataset comprised 4,500 samples. Additionally, the dataset contained features with varying value ranges, which were normalized to a range of 0–1 using the min-max scaling method. Primary filtering of cognitive features was performed using Pearson's correlation coefficient to identify those most correlated with AD diagnosis. Features with a correlation coefficient above a predefined threshold were selected for further analysis. Performance evaluation utilized metrics including Accuracy, Sensitivity, and F1-Score, which are computed by Equations 9, 10, and 11, respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (10)$$

$$F1 - Score = \frac{2TP}{2(TP + FP + FN)} \quad (11)$$

Here, TP represents True Positives, TN stands for True Negatives, FP is False Positives, and FN represents False Negatives.

4.2 Network settings

The DeepCGAN model architecture for AD detection incorporates carefully chosen parameters to optimize performance across multiple metrics. The feature maps in the Generator’s encoder are structured with fixed sizes of 16, 32, 64, 128, and 256 in successive convolutional layers, with specific kernel sizes and strides tailored to enhance feature extraction efficiency. The kernel size is set to (1,3) for the first 2D-Conv layer and (2,3) for subsequent layers, all with a stride of (1,2). Convolutional layers are utilized for their strength in extracting local patterns and hierarchical features from the cognitive data, which are essential for distinguishing subtle differences between normal and AD-affected individuals. Similarly, the BiGRU layers are configured with 2,048 units, with 1,024 units in each direction, split into forward and backward directions, operating over a fixed time step of 50. The BiGRUs were selected for their ability to capture long-term dependencies and temporal patterns in cognitive data, which are crucial for accurately modeling the progression of AD over time. For the Generator’s decoder, these parameters are inversely set to reconstruct the input features faithfully. Moreover, the Discriminator (*D*) employs deconvolutional layers with gradually increasing feature maps from 4 to 64, designed to discriminate between real and generated samples effectively. The proposed AD detection models, incorporating these two distinct losses, undergo training and optimization using the Adam optimizer for 1,000 epochs, with a learning rate of 0.005 and a batch size of 32 samples. The combination of convolutional layers and BiGRUs in the DeepCGAN architecture leverages both spatial and temporal features, providing a robust framework for early AD detection by capturing complex patterns in cognitive data that simpler architectures might miss. This setup ensures robust optimization and convergence of the DeepCGAN model. To assess the performance of our proposed model, we conducted a comprehensive comparison with several other models, including DeciTree, RanForest, KNN, Linear Regression (LR), SVM, DNN, AdaBoost, and Adaptive Voting, utilizing various metrics such as Accuracy, Precision, Recall, and F1-Score.

5 Results and analysis

In this section, we present the results of our experiments and provide a comprehensive analysis of the findings.

5.1 Model performance comparison

Table 1 displays the results obtained from our proposed DeepCGAN model along with other DL models trained on similar cognitive features for detecting cognitive normal and AD. We measure the model’s performance using Accuracy, Precision, Recall, and F1-Score as evaluation metrics. Notably, the results demonstrate that our proposed DeepCGAN outperforms all other competing models in terms of these metrics. The DeepCGAN achieved an Accuracy of 97.32%, Precision of 95.31%, Recall of

TABLE 1 Performance analysis using various measures for ADNI cognitive features dataset.

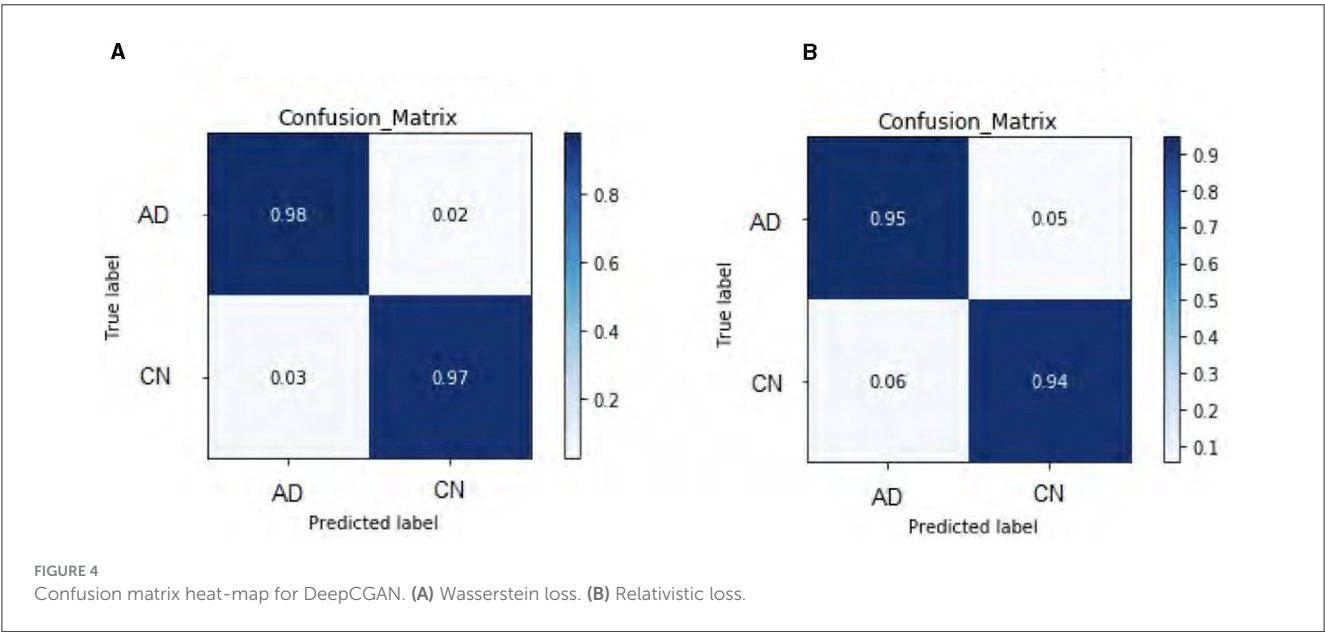
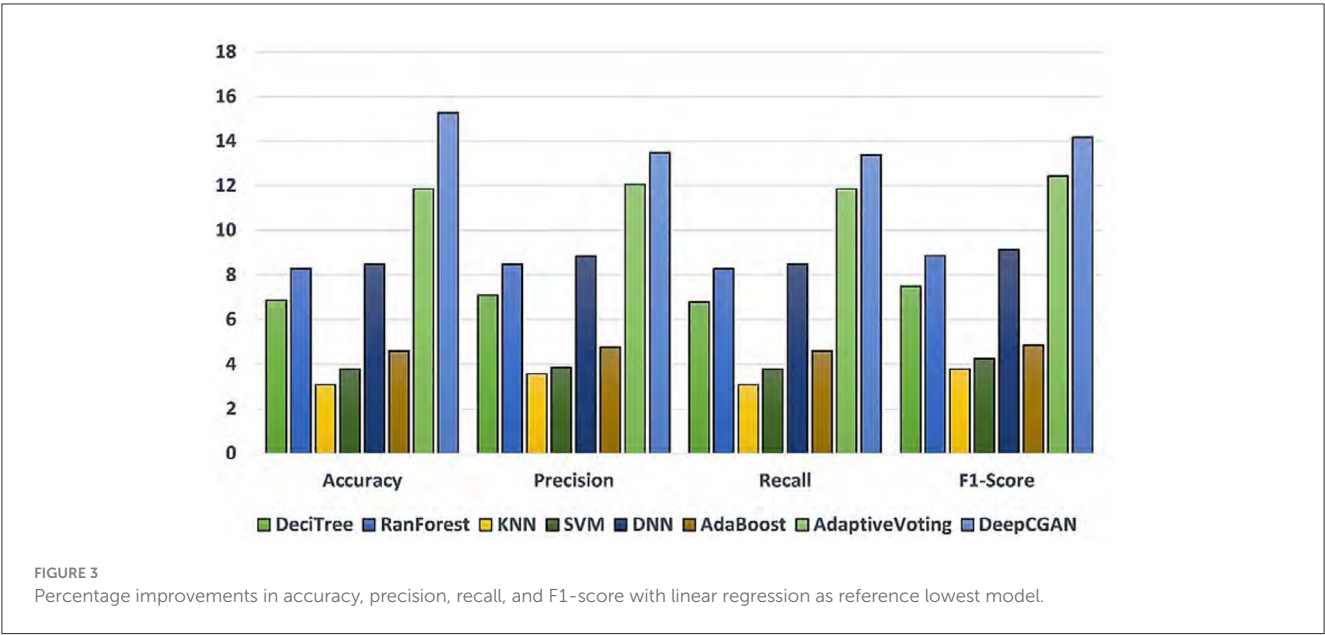
DL model	Acc. (%)	Pre. (%)	Rec. (%)	F1-S. (%)
DeciTree	88.93	88.93	88.93	88.93
RanForest	90.33	90.30	90.33	90.31
KNN	85.14	85.40	85.14	85.24
LR	82.05	81.83	82.05	81.45
SVM	85.84	85.68	85.84	85.71
DNN	90.53	90.67	90.53	90.59
AdaBoost	86.64	86.61	86.64	86.31
AdapVoting	93.92	93.89	93.92	93.89
DeepCGAN	97.32	95.31	95.43	95.61

95.43%, and F1-Score of 95.61%, respectively. In contrast, the lowest-performing model, linear regression, achieved only 82.05% Accuracy, 81.83% Precision, and 81.45% F1-Score.

To highlight the improvements made by our proposed model, we chose linear regression as a reference model. DeepCGAN substantially improved Accuracy by 15.27%, Precision by 13.48%, and F1-Score by 14.16% compared to linear regression. Moreover, when compared to the second-best model, Adaptive Voting, DeepCGAN showed a 3.40% improvement in Accuracy. It also outperformed DNN and Random Forest by 6.79 and 6.99% in Accuracy, respectively, which is a significant performance gain. Furthermore, our DeepCGAN model demonstrated substantial improvements in Recall and F1-Score compared to competing models. The Recall increased from 88.93% (DeciTree) to 95.43% with DeepCGAN, and the F1-Score increased from 86.31% (AdaBoost) to 95.61%. These results signify the superior ability of DeepCGAN to correctly identify AD cases while minimizing false negatives. The errors are vastly improved over other models. To highlight the effectiveness of the proposed model, we present the overall improvements depicted in Figure 3. The linear regression is the reference model that has achieved the lowest performance among DL models. Figure 3 indicates the best performance of the proposed DeepCGAN.

5.2 Loss function analysis

Our DeepCGAN model was trained and optimized using two different loss functions: Wasserstein and Relativistic loss. The Wasserstein loss function was chosen for its ability to provide a smoother gradient, thereby stabilizing the training process. This stability is particularly important for AD detection, where the model must accurately learn from subtle variations in cognitive data. Figure 4 presents the confusion matrices for both losses, revealing that the Wasserstein loss function results in better performance. Figure 4A illustrates the predicted labels when trained with Wasserstein loss, while Figure 4B shows the outcomes with Relativistic loss. It is evident



that the model trained with Wasserstein loss provides more accurate predictions.

To further assess the performance, we compared the model's Accuracy on training and validation data. The Wasserstein loss outperformed the relativistic loss by a significant margin, indicating faster convergence and better Accuracy. Figure 5 displays the loss curves over 1,000 epochs, illustrating the superior performance of the DeepCGAN model in achieving its detection task. We also evaluated the Area Under the Curve (AUC), which measures the model's ability to differentiate between labels. DeepCGAN exhibited higher AUC values compared to models trained with relativistic loss, further confirming its superior discriminatory capability. Figure 6 illustrates the AUC curves for both loss functions.

5.3 Comparison against state-of-the-art models

In this subsection, we compare our proposed DeepCGAN model with state-of-the-art (SOTA) models in the literature, including AlexNet (31), VGG-16 (32), GoogleNet (33), and ResNet (34), using cognitive features from the ADNI dataset. This comparison aims to benchmark the performance of DeepCGAN under similar experimental settings and datasets. Table 2 presents the results in terms of Accuracy, Precision, Recall, F1-scores, and AUC.

DeepCGAN surpassed all SOTA models in terms of Accuracy, achieving an Accuracy of 97.32%, which is a 5.59% improvement over GoogleNet. Similarly, Precision improved from 90.20% (GoogleNet) to 95.31%, reflecting a 5.11% boost in Precision.

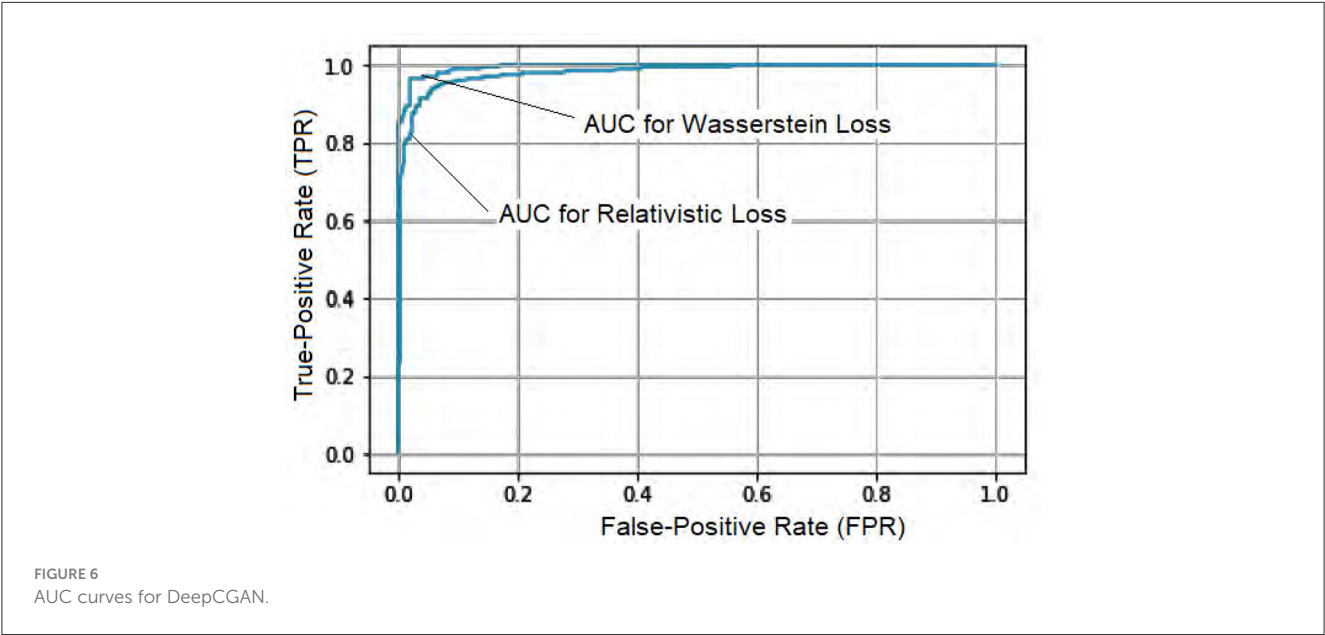
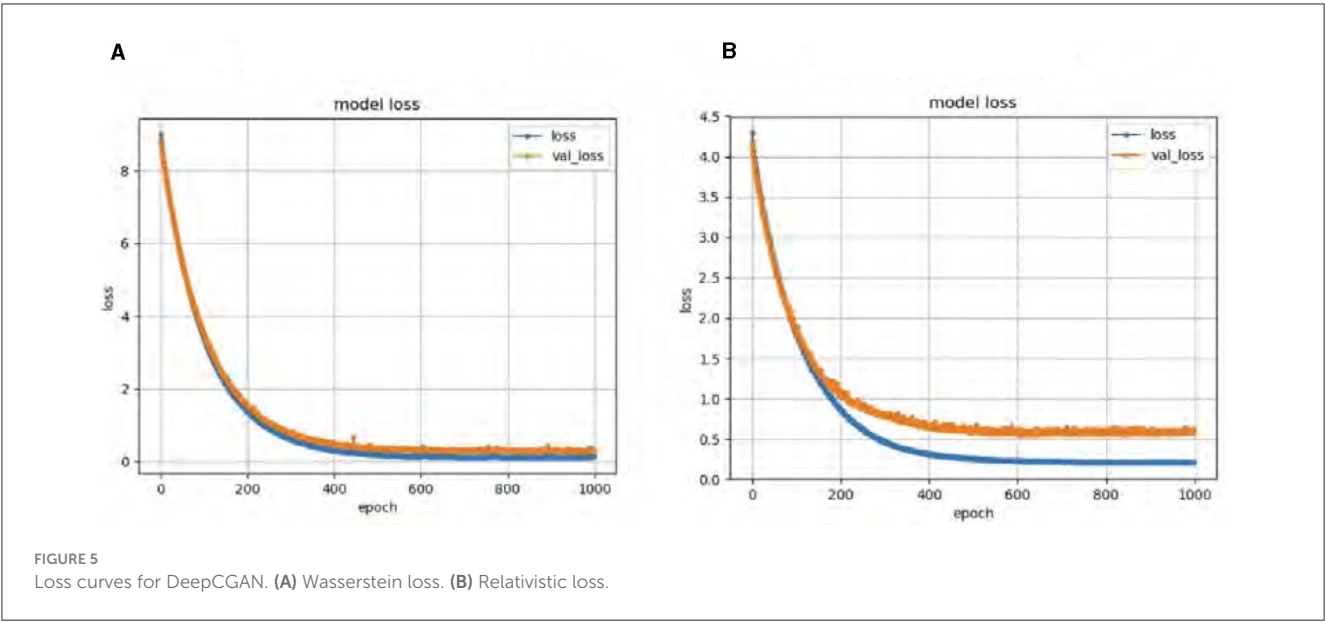


TABLE 2 Performance analysis (in %) for SOTA using ADNI Cognitive features dataset.

Model	Accuracy	Precision	Recall	F1-score	AUC
AlexNet	93.75	94.98	92.28	93.61	93.68
VGG-16	94.96	94.02	95.43	94.97	94.96
GoogleNet	91.73	90.20	93.50	91.82	91.79
ResNet	94.96	93.00	97.15	95.03	94.98
DeepCGAN	97.32	95.31	95.43	95.61	99.51

When compared to AlexNet, DeepCGAN demonstrated a 3.57% increase in Accuracy. Furthermore, DeepCGAN achieved the highest AUC among all models, with a 99.51% AUC, outperforming ResNet by 4.53% and VGG-16 by 4.55%, highlighting its superior discriminatory power. Regarding Recall, DeepCGAN

exhibited substantial improvements over SOTA models except for ResNet, where the results were marginally lower. Specifically, the Recall increased from 92.28% (AlexNet) to 95.43% with DeepCGAN. The F1-Score achieved with GoogleNet was 91.82%. DeepCGAN’s superior performance can be attributed to its

TABLE 3 Performance comparison of the proposed model on ADNI dataset cognitive features.

Model	Records	Accuracy	AUC
Gill et al. (35)	600	81.80%	79.0%
AdaptiveVoting	5013	93.92%	99.3%
DeepCGAN	5013	97.32%	99.5%

novel architecture, which combines convolutional layers for effective feature extraction and BiGRUs for capturing temporal dependencies. This dual approach enables the model to detect subtle changes and patterns in cognitive data more accurately than models that rely solely on neuroimaging data or simpler architectures. Additionally, the use of GANs for data augmentation increases the dataset's size and diversity, enhancing the model's generalizability and robustness. The core innovation lies in expanding the cognitive features dataset and enhancing its diversity through GANs.

DeepCGANs significantly enhance Alzheimer's detection due to their ability to generate realistic synthetic images, crucial for augmenting limited MRI and PET scan datasets. Their deep convolutional layers extract complex features, improving diagnostic accuracy by capturing subtle disease indicators. Adversarial training refines synthetic images iteratively, ensuring they closely resemble real patient data. DCGANs' adaptability across imaging modalities and superior performance in comparative evaluations underline their transformative role in improving diagnostic accuracy and clinical outcomes for AD.

5.4 Comparison with existing techniques

In this section, we compare our proposed DeepCGAN model with a recently reported technique by Gill et al. (35) that used cognitive features for AD detection. Both studies utilized the same ADNI dataset, and the results are presented in Table 3. Our proposed DeepCGAN model outperformed the model proposed by Gill et al. (35) and Adaptive Voting using cognitive features from the ADNI dataset. DeepCGAN achieved the highest Accuracy of 97.32%, representing a 15.52% improvement over Adaptive Voting and a 3.4% improvement over Gill et al.'s technique for early AD detection. This improvement is due to its ability to generate synthetic data that closely resembles the actual cognitive features, thus reducing overfitting and improving the model's ability to generalize to new, unseen data.

6 Conclusion

In this study, we propose a novel convolutional encoder-decoder-based GAN for early AD detection using cognitive features. This model leverages a Generator module with Conv2D and Deconv2D layers in an encoder-decoder architecture to optimize Accuracy, Precision, Recall, F1-Score, and AUC metrics. Our experimental results demonstrate the superior performance of

DeepCGAN, which significantly advances early AD detection, and outperforms several state-of-the-art models and benchmarks across various measures, achieving an outstanding 97.32% Accuracy compared to most other DL models in this study's SOTA comparison. Moreover, We find that using the Wasserstein loss is superior for training the proposed GAN. While our GAN excels, it is important to acknowledge the potential of SOTA DL models for early AD detection, which offer advantages over non-DL techniques like Gill's study. These DL models can expedite diagnosis, making them valuable tools in the detection of neurodegenerative diseases like Alzheimer's. The unique contribution of DeepCGAN lies in its novel use of GANs to enhance the dataset's size and diversity, coupled with a sophisticated architecture that integrates convolutional layers and BiGRUs. This approach significantly improves accuracy, precision, and overall performance metrics in detecting AD at early stages, demonstrating the model's superior capability in distinguishing subtle cognitive changes indicative of early AD. In the future, we aim to develop even more robust and streamlined DL models for detecting early and various stages of AD. Our proposed DeepCGAN model significantly advances early AD detection by leveraging a convolutional encoder-decoder-based GAN with Wasserstein loss, achieving superior performance metrics compared to SOTA models such as AlexNet, VGG-16, GoogleNet, and ResNet. This novel approach enhances the diversity and richness of cognitive features, resulting in a remarkable improvement in accuracy, precision, and discriminatory power, as demonstrated through comprehensive comparisons with existing techniques and models.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://adni.loni.usc.edu/data-samples/adni-data/>.

Author contributions

IA: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Conceptualization. NS: Writing – review & editing, Resources, Project administration, Methodology, Investigation. MA: Writing – review & editing, Validation, Supervision, Software, Funding acquisition. BZ: Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation. KA: Writing – review & editing, Supervision, Project administration, Funding acquisition, Formal analysis, Data curation, Conceptualization. QH: Writing – review & editing, Visualization, Validation, Software, Methodology, Formal analysis.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This Research was funded by Researchers Supporting Project

Number (RSPD2024R553), King Saud University, Riyadh, Saudi Arabia.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Pan Y, Liu M, Lian C, Zhou T, Xia Y, Shen D. Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis. In: *Medical Image Computing and Computer Assisted Intervention-MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part III*. Cham: Springer (2018). p. 455–63.
- Yan Y, Lee H, Somer E, Grau V. Generation of amyloid PET images via conditional adversarial training for predicting progression to Alzheimer's disease. In: *Predictive Intelligence in Medicine: First International Workshop, PRIME 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 1*. Cham: Springer (2018). p. 26–33.
- Wang Y, Zhou L, Yu B, Wang L, Zu C, Lalush DS, et al. 3D auto-context-based locality adaptive multi-modality GANs for PET synthesis. *IEEE Trans Med Imaging*. (2019) 38:1328–39. doi: 10.1109/TMI.2018.2884053
- Wang Y, Yu B, Wang L, Zu C, Lalush DS, Lin W, et al. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. *Neuroimage*. (2018) 174:550–62. doi: 10.1016/j.neuroimage.2018.03.045
- Ouyang J, Chen KT, Gong E, Pauly J, Zaharchuk G. Ultra-low-dose PET reconstruction using a generative adversarial network with feature matching and task-specific perceptual loss. *Med Phys*. (2019) 46:3555–64. doi: 10.1117/12.2512946
- Islam J, Zhang Y. GAN-based synthetic brain PET image generation. *Brain Inf*. (2020) 7:1–12. doi: 10.1186/s40708-020-00104-2
- Kim HW, Lee HE, Lee S, Oh KT, Yun M, Yoo SK. Slice-selective learning for Alzheimer's disease classification using a generative adversarial network: a feasibility study of external validation. *Eur J Nucl Med Mol Imaging*. (2020) 47:2197–206. doi: 10.1007/s00259-019-04676-y
- Kang H, Park JS, Cho K, Kang D-Y. Visual and quantitative evaluation of amyloid brain PET image synthesis with the generative adversarial network. *Appl Sci*. (2020) 10:2628. doi: 10.3390/app10072628
- Kang SK, Seo S, Shin SA, Byun MS, Lee DY, Kim YK, et al. Adaptive template generation for amyloid PET using a deep learning approach. *Hum Brain Mapp*. (2018) 39:3769–78. doi: 10.1002/hbm.24210
- Jha D, Kim JI, Kwon GR. Diagnosis of Alzheimer's disease using dual-tree complex wavelet transform, PCA, and feedforward neural network. *J Healthc Eng*. (2017) 2017:9060124. doi: 10.1155/2017/9060124
- Sarraf S, Tofighi G. Classification of Alzheimer's disease structural MRI data by deep learning convolutional neural networks. *arXiv [preprint]* (2016). doi: 10.48550/arXiv.1607.06583
- Jaiswal S, Gupta P, Prasad LN, Kulkarni R. An empirical model for the classification of diabetes and diabetes_types using ensemble approaches. *J Artif Intell Technol*. (2023) 3:181–6. doi: 10.37965/jait.2023.0220
- Ganaie MA, Hu M, Tanveer M, Suganthan PN. Ensemble deep learning: a review. *arXiv [preprint]* (2021). doi: 10.1016/j.engappai.2022.105151
- Wu X, Xu K, Hall P. A survey of image synthesis and editing with generative adversarial networks. *Tsinghua Sci Technol*. (2017) 22:660–74. doi: 10.23919/TST.2017.8195348
- Wolterink JM, Dinkla AM, Savenije MHF, Seevinck PR, van den Berg CAT, Išgum I. Deep MR to CT synthesis using unpaired data. In: Gooya A, Frangi AF, Tsafaris SA, Prince JL, editors. *Simulation and Synthesis in Medical Imaging*. Vol. 10557. Cham: Springer (2017). p. 14–23.
- Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: a review. *Med Image Anal*. (2019) 58:101552. doi: 10.1016/j.media.2019.101552
- Bjorck N, Gomes CP, Selman B, Weinberger KQ. Understanding batch normalization. In: *Advances in Neural Information Processing Systems*. Vol. 31. Red Hook, NY: Curran Associates, Inc (2018).
- Dukler Y, Li W, Lin A, Montúfar G. Wasserstein of Wasserstein loss for learning generative models. In: *International Conference on Machine Learning*. Cambridge, MA: PMLR (2019). p. 1716–25.
- Jolicoeur-Martineau A. The relativistic discriminator: a key element missing from standard GAN. *arXiv [preprint]* (2018). doi: 10.48550/arXiv.1807.00734
- de Vent NR, Agelink van Rentergem JA, Schmand BA, Murre JM, Huizenga HM, Consortium A. Advanced Neuropsychological Diagnostics Infrastructure (ANDI): a normative database created from control datasets. *Front Psychol*. (2016) 7:1601. doi: 10.3389/fpsyg.2016.01601
- Weiner MW, Veitch DP, Miller MJ, Aisen PS, Alcala B, Beckett LA, et al. Increasing participant diversity in AD research: plans for digital screening, blood testing, and a community-engaged approach in the Alzheimer's Disease Neuroimaging Initiative 4. *Alzheimers Dement*. (2023) 19:307–17. doi: 10.1002/alz.12797
- Hosseini-Asl E, Ghazal M, Mahmoud A, Aslantas A, Shalaby AM, Casanova ME, et al. Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. *Front Biosci*. (2018) 23:584–96. doi: 10.2741/4606
- Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv [preprint]* (2014). doi: 10.48550/arXiv.1412.6980
- Hosseini-asl E, Keynton R, El-baz A. Alzheimer's disease diagnostics by adaptation of 3D convolutional network. In: *Proc. - Int Conf Image Process ICIP*. Piscataway, NJ (2016). 502 p.
- Zhan B, Li D, Wu X, Zhou J, Wang Y. Multi-modal MRI image synthesis via GAN with multi-scale gate merge. *IEEE J Biomed Health Inf*. (2021) 26:17–26. doi: 10.1109/JBHI.2021.3088866
- Suganthi K. Review of medical image synthesis using GAN techniques. In: *ITM Web of Conferences*. Vol. 37. Les Ulis: EDP Sciences (2021). p. 01005.
- Wang J, Shi M, Zhang X, Li Y, Yuan Y, Yang C, et al. Incremental data stream classification with adaptive multi-task multi-view learning. *Big Data Mining Anal*. (2023) 7:87–106. doi: 10.26599/BDMA.2023.9020006
- KV S, Ravi V, Almeshari M, Alzamil Y, DS SK, et al. A new deep learning model based on neuroimaging for predicting Alzheimer's disease. *Open Bioinf J*. (2023) 16:1–9. doi: 10.2174/0118750362260635230922051326
- Dong X, Yu Z, Cao W, Shi Y, Qianli MA. A survey on ensemble learning. *Front Comput Sci*. (2020) 14:241–58. doi: 10.1007/s11704-019-8208-z
- Prakash D, Madusanka N, Bhattacharjee S, Park HG, Kim CH, Choi HK, et al. comparative study of Alzheimer's disease classification using multiple transfer learning models. *J Multimed Inf Syst*. (2019) 6:209–16. doi: 10.33851/JMIS.2019.6.4.209
- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. (2012) 25:1–9. doi: 10.1145/3065386
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv [preprint]* (2014). doi: 10.48550/arXiv.1409.1556
- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ (2015). p. 1–9.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ (2016). p. 770–8.
- Gill S, Mouches P, Hu S, Rajashekar D, MacMaster FP, Smith EE, et al. Using machine learning to predict dementia from neuropsychiatric symptom and neuroimaging data. *J Alzheimers Dis*. (2020) 75:277–88. doi: 10.3233/JAD-191169

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



OPEN ACCESS

EDITED BY

Surbhi Bhatia Khan,
University of Salford, United Kingdom

REVIEWED BY

Krishna Kumar Mohbey,
Central University of Rajasthan, India
Chandrasekar Ravi,
National Institute of Technology Puducherry,
India
Sohail Mohammed,
Toronto Metropolitan University, Canada
Nur Al Hasan Haldar,
Curtin University, Australia

*CORRESPONDENCE

Mustafa Haider Abidi
✉ mabidi@ksu.edu.sa

RECEIVED 03 May 2024

ACCEPTED 12 August 2024

PUBLISHED 03 September 2024

CITATION

Alkhalefah H, Preethi D, Khare N,
Abidi MH and Umer U (2024) Deep learning
infused SIRVD model for COVID-19
prediction: XGBoost-SIRVD-LSTM approach.
Front. Med. 11:1427239.
doi: 10.3389/fmed.2024.1427239

COPYRIGHT

© 2024 Alkhalefah, Preethi, Khare, Abidi and
Umer. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Deep learning infused SIRVD model for COVID-19 prediction: XGBoost-SIRVD-LSTM approach

Hisham Alkhalefah¹, D. Preethi², Neelu Khare³,
Mustafa Haider Abidi^{1*} and Usama Umer¹

¹Advanced Manufacturing Institute, King Saud University, Riyadh, Saudi Arabia, ²Department of
Computer Science and Engineering, Faculty of Engineering and Technology, SRM Institute of Science
and Technology, Chennai, Tamil Nadu, India, ³School of Computer Science Engineering and
Information Systems (SCORE), Vellore Institute of Technology, Vellore, Tamil Nadu, India

The global impact of the ongoing COVID-19 pandemic, while somewhat contained, remains a critical challenge that has tested the resilience of humanity. Accurate and timely prediction of COVID-19 transmission dynamics and future trends is essential for informed decision-making in public health. Deep learning and mathematical models have emerged as promising tools, yet concerns regarding accuracy persist. This research suggests a novel model for forecasting the COVID-19's future trajectory. The model combines the benefits of machine learning models and mathematical models. The SIRVD model, a mathematical based model that depicts the reach of the infection via population, serves as basis for the proposed model. A deep prediction model for COVID-19 using XGBoost-SIRVD-LSTM is presented. The suggested approach combines Susceptible-Infected-Recovered-Vaccinated-Deceased (SIRVD), and a deep learning model, which includes Long Short-Term Memory (LSTM) and other prediction models, including feature selection using XGBoost method. The model keeps track of changes in each group's membership over time. To increase the SIRVD model's accuracy, machine learning is applied. The key properties for forecasting the spread of the infection are found using a method called feature selection. Then, in order to learn from these features and create predictions, a model involving deep learning is applied. The performance of the model proposed was assessed with prediction metrics such as R^2 , root mean square error (RMSE), mean absolute percentage error (MAPE), and normalized root mean square error (NRMSE). The results are also validated to those of other prediction models. The empirical results show that the suggested model outperforms similar models. Findings suggest its potential as a valuable tool for pandemic management and public health decision-making.

KEYWORDS

deep learning, extreme gradient boosting (XGBoost), susceptible-infected-recovered-vaccination-deceased (SIRVD), long short-term memory (LSTM), feature selection, COVID-19, prediction

1 Introduction

The COVID-19 epidemic has presented a serious threat to civilization worldwide. The virus has killed millions of people and spread quickly. World Health Organization (WHO) at the end of 2019 announced COVID-19 as global epidemic disease, since its outbreak worldwide. As of November 6, 2023, reported by WHO, there are 775,335,916 confirmed

cases and 7,045,569 deaths worldwide (1). Based on WHO data, 13.59bn vaccine doses have been given as of May 2, 2024. COVID-19 immensely affected daily life, health, and the economy at the global level. Governments and public health experts have put in place a number of strategies to prevent the epidemic, including social isolation, mask use, and vaccine. However, the transmission of the virus has not totally been halted by these precautions. Predicting how the pandemic will develop in the future is one of the difficulties in combatting COVID-19. This is significant for various reasons. First, it can assist governments and public health experts in making choices regarding the distribution of resources and pandemic response. Second, it can assist organizations and people in making decisions regarding how to run and safeguard themselves. The upcoming course of COVID-19 transmission is forecasted using various techniques. Making use of mathematical models is one strategy. The transmission of the virus and its effects on various populations can be predicted using mathematical models. Mathematical modeling is a crucial device for analyzing epidemic infectious diseases, presented in 1927 by Kermack (2). Since the outbreak of the pandemic, various mathematical models have been employed in predicting the diseases, which are epidemic. The widely used mathematical models include SIR (3), which assesses susceptible, infected, and recovered rates (4), and SEIR (5), which evaluates based on susceptible, exposed, infected, and recovered rates. Furthermore, most of the research studies are the enhanced models derived from these two models. However, using mathematical models can be challenging and complex.

Machine learning is a different strategy for forecasting COVID-19's future trajectory. Machine learning, a form of artificial intelligence, possesses the ability to gain information from data and produce predictions. The efficacy of models involving machine learning in predicting transmission of various illnesses, including influenza, has been established through empirical evidence. Many studies are available on predicting and transmitting the virus's spread (6).

This paper introduces a novel deep learning model named Extreme Gradient Boosting-Susceptible-Infected-Recovered-Vaccinated-Deceased-Long Short-Term Memory (XGBoost-SIRVD-LSTM), which is designed to forecast the quantity of COVID-19 cases. The suggested XGBoost-SIRVD-LSTM model operates in four stages: (1) Data pre-processing, (2) XGBoost feature importance score feature selection, (3) SIRVD epidemic model design, and (4) LSTM prediction. The suggested model is tested using datasets from John Hopkins University's CSSE (7) and Our World in Data (8). The dataset is first pre-processed using the min-max normalization technique. Second, the XGBoost is used for feature selection, which is done using the feature importance score. Finally, the optimal features are supplied into the SIRVD model to estimate the COVID-19 transmission with respect to time. Finally, the LSTM model is applied to the dataset for disease prediction. The empirical results suggest that the suggested model exhibits superior performance in relation to accuracy for predicting outcomes compared to alternative deep learning models.

The following are the research study's contributions:

- In this study, we introduce a deep learning model that utilizes XGBoost-SIRVD-LSTM model to predict COVID-19 infection cases.
- The outcomes of the suggested model assessed in comparison with existing deep learning models and utilizing performance measures for prediction.

The remaining sections of the paper are structured as follows: Section 2 presents a summary of the current body of literature. Section 3 delves into background information of the techniques employed in the proposed model. Section 4 outlines the methodology proposed in detail. Section 5 explores the dataset, presents the experimental results, and includes a comparative analysis with other models.

2 Literature review

This section elaborates on numerous models for COVID-19 prediction found in the literature. A standard SIR model for predicting COVID-19 pandemic progression was proposed in Kartono et al. (9). The model was tested using the most recent confirmed cases from the WHO dashboard. The authors used this approach to forecast instances in Singapore, Saudi Arabia, Indonesia, and the Philippines. In their study, Kumar et al. (10) employed recurrent neural network (RNN) models, including gated recurrent unit (GRU) and LSTM cells, to predict the future patterns of COVID-19 cases. The researchers utilized the publicly accessible COVID-19 dataset from Johns Hopkins University and emphasized the importance of factors such as age, population density, healthcare infrastructure, and disease-prevention efforts in the rapid progression of the COVID-19 outbreak. To analyze the COVID-19 pandemic, the study conducted exploratory data analysis using machine-learning techniques, followed by the implementation of the SIR model (11). The most popular John Hopkins dataset for COVID-19 was used for experiments, with just data from the Kingdom of Saudi Arabia used to forecast instances. The researchers analyzed three possibilities for anticipating the progression of the outbreak and its possible resolution, namely new medicine, lockdowns, and no actions. The simulation results demonstrate that interventions such as new drugs and lockdowns outperform no-action scenarios. To forecast the COVID-19 instances, the MLP with feature selection (MLPFS) classification model was presented (12). This study was based on the characteristics and symptoms of Electronic Medical Records (EMR) patients. Three separate datasets and eight alternative models were utilized to evaluate the provided model. According to the experimental findings, the suggested MLPFS outperformed the other seven models chosen for comparison in terms of accuracy indicators, extracted number of features, and time required to implement the model. The SIRVD model, an extension of classic epidemiological models, incorporates vaccination and time-dependent fatality rates (13). Analyzing exact solutions and approximations, it reveals crucial insights into epidemic dynamics, offering benchmarks for numerical simulations. By applying analytical approximations, particularly effective for low cumulative infection rates, it elucidates the impact of vaccination and time-varying fatality rates, enabling precise parameter extraction from COVID-19 data, essential for pandemic management. Babaei et al. (14) explores integrability conditions and exact analytical solutions for the SIRV model, crucial for understanding COVID-19 dynamics, using a partial Hamiltonian approach. Analyzing two cases based on model parameters and considering different phase spaces, it provides insights into the dynamics of susceptible, infected, recovered, and vaccinated populations over time through graphical representations. Federico (15) addresses an optimal vaccination strategy within an SIRS compartmental model, aiming to minimize social and economic costs while reducing susceptibility. Theoretical contributions include a non-smooth verification theorem and

conditions for well-posed closed-loop equations, while numerical implementations highlight the effectiveness of vaccination policies in long-term infection control, particularly with low reproduction and reinfection rates.

In another study, researchers suggested a three-stage COVID-19 prediction, namely pre-processing, feature selection, and classification (16). Wrapper-based feature selection using Recursive Feature Extraction and embedded-based feature selection using Extra Tree Classifier were the two methods used. The naive bayes and restricted Boltzmann Machine models were employed for classification. The proposed approach was implemented using WHO data. According to the authors, the model worked well and produced better prediction results with feature selection than models without feature selection. In their previous work, the researchers put forth COVID-19 prediction models utilizing Susceptible_Infected_Recovered (SIR) and Susceptible_Exposed_Infected_Quarantined_Recovered (SEIQR) epidemic models for several countries, including Australia, United Kingdom, and Italy (3). To enhance parameters in these epidemic models (L-BFGS-B), they employed optimization algorithms such as Conjugate Gradient (CG), Nelder–Mead, restricted memory bound constrained, and the Broyden–Fletcher–Goldfarb–Shanno (BFGS). The performance of these two models was compared to the performance of two machine learning methods, prophet and logistic function. The authors discovered that the prophet model outperformed the logistic function and provided a superior prediction model for Italy and the United Kingdom than for Australia. The prediction accuracy was significantly increased once the models such as SIR and SEIQR were optimized. In their findings, the authors observed that the prophet model demonstrated superior performance compared to the logistic function, particularly in predicting the COVID-19 trends for United Kingdom and Italy, while its performance in the case of Australia was relatively less favorable. The accuracy of predictions was notably improved by optimizing the SIR and SEIQR models. In a separate study conducted by the authors of Chandra et al. (17), deep learning-based LSTM models were explored for predicting the future trajectory of COVID-19 in specific Indian states that experienced a high incidence of the disease. Various LSTM models, including LSTM, bidirectional, and encoder-decoder models, were developed for disease spread prediction. The authors highlighted that the encoder-decoder LSTM model exhibited superior prediction accuracy compared to other models. In Alassafi et al. (18), a comparison study was undertaken to assess the efficacy of RNN and LSTM models in predicting the spread of the coronavirus. The dataset utilized for this analysis consisted of data from Malaysia, Morocco, and Saudi Arabia, sourced from the European Center for Disease Prevention and Control. The authors examined the models' effectiveness in predicting positive cases, recoveries, and COVID-19-related mortality rates. Also, estimating the potential quantity of cases over the next 7 days. Another research study (19) proposed an XGBoost-DNN classifier model for detecting network intrusions. The model employed XGBoost feature importance scores to select relevant features and utilized DNN for classification of network intrusions.

In a separate study, researchers introduced a feature selection based on ensemble approach with LSTM for network intrusion classification (20). Their method aimed to improve the accuracy of network invasion detection by utilizing LSTM along with ensemble-based feature selection. Youssef et al. (4) employed the SEIQR model and utilized real data of Saudi Arabia for predicting

the transmission of COVID-19 cases. The results demonstrated the efficiency of the model suggested in analyzing epidemic spread, thus providing a basis for framing effective government policies.

The COVID-19 pandemic has significantly accelerated research on the development of predictive models for the pandemic's future trajectory. Numerous models, including mathematical, machine learning, and hybrid models, have been put forth. The propagation of the virus can be simulated and the effects on various populations can be predicted using mathematical models that are based on epidemiological principles. However, using mathematical models can be challenging and complex. An artificial intelligence that can learn from data and predict the future is known as a machine-learning model. It has been demonstrated that machine-learning models are useful for forecasting the spread of other diseases, such as influenza. Nevertheless, machine-learning models can often rely heavily on the specific data they were trained on, resulting in potential challenges when attempting to generalize to new data. To overcome these limitations, hybrid models (21–23) merge the advantages of both mathematical models and machine learning approaches. By combining these two techniques, hybrid models have the potential to offer greater precision and accuracy compared to using either method in isolation. However, the development of hybrid models can be intricate and pose significant challenges. Despite the extensive research conducted thus far, there remains a need for more precise and reliable models to effectively forecast the future trajectory of COVID-19. Thus, this study endeavors to fill this research void by proposing a novel model that leverages the strengths of both mathematical and machine learning methods.

3 Methodology

3.1 Xgboost feature selection

Extreme Gradient Boosting (XGBoost) is a scalable machine learning technique used for tree boosting, which falls within the class of scalable machine learning approaches (24). This method, known as a distributed optimized library for gradient boosting, is capable of analyzing the relevance of each feature in the dataset. It has been demonstrated as a reliable and practical approach in machine learning research (19, 25). In comparison to earlier boosting methods, XGBoost excels at selecting a robust classifier from a set of weaker classifiers. It offers advantages such as effective handling of missing values, avoidance of overfitting, and faster computation times for parallel and distributed models. The primary objective of XGBoost utilizes an optimized gradient descent approach with versatile differentiable loss functions is to employ an optimized gradient descent method with arbitrary differentiable loss functions. This is achieved by incorporating weak learners to minimize the loss function, thus defining and optimizing the overall objective function.

Extreme gradient boosting strives to reduce the objective function in the following manner (as shown in Equation 1).

$$obj(\theta) = \sum_i L\left(\hat{y}_i, y_i\right) + \sum_k \Omega(f_k), f_k \in F \quad (1)$$

The training loss function, denoted as L , quantifies the disparity between the predicted value y_i by the model proposed and actual value of y_i . Overfitting is prevented thanks to the regularization function Ω , which estimates the model's complexity. The set of all possible regression trees is represented by the function f in the functional space F . By using parameters and a greedy search method, XGBoost determines the optimal tree structure to minimize the objective function.

3.2 SIRVD epidemic model

The SIRVD is derived from the SIR epidemic model (26). This model encompasses dynamics of the virus's interaction during transmission with the host and classifies individuals into five distinct groups: susceptible, infected, recovered, vaccinated, and deceased (27). The SIRVD expands upon the existing SIR framework by including the new states of vaccinated and deceased. Vaccinated persons are those who have been inoculated against the disease, while deceased individuals are those who have died after becoming sick in the community (28). The ordinary differential equations below represent the mathematical formulation of the SIRVD model (Equations 2–7).

$$\frac{dS_r(t_i)}{d(t_i)} = -\frac{\beta I_r(t_i) S(t_i)}{N} + \sigma R_r(t_i) - \alpha S_r(t_i) \quad (2)$$

$$\frac{dI_r(t_i)}{d(t_i)} = \frac{\beta I_r(t_i) S_r(t_i)}{N} - \gamma I_r(t_i) - \delta I_r(t_i) \quad (3)$$

$$\frac{dR_r(t_i)}{d(t_i)} = \gamma I_r(t_i) - \sigma R_r(t_i) \quad (4)$$

$$\frac{dV_r(t_i)}{d(t_i)} = \alpha S_r(t_i) \quad (5)$$

$$\frac{dD_r(t_i)}{d(t_i)} = \delta I_r(t_i) \quad (6)$$

$$N = S_r(t_i) + I_r(t_i) + R_r(t_i) + V_r(t_i) + D_r(t_i) \quad (7)$$

where,

β —Infection rate, encompasses the spread of the infection in a susceptible state.

γ —Recovery rate consists of the transferal from the infected to the recovered state.

δ —Rate of death, represents the transferal from the infected to the deceased state.

α —Rate of vaccination consists of the transferal from susceptible to the vaccinated condition.

σ —rate of susceptibility depicts the transferal from recovered to a susceptible state.

It is stated that the transference cycle of the virus is characterized by $\frac{\beta I_r(t_i) S_r(t_i)}{N}$ depicts the number of individuals per unit of time

who transmitted from the susceptible individuals (S_r) to the infected individuals (I_r). The five parameters of the SIRVD epidemic model such as $\beta, \gamma, \delta, \alpha$, and σ are considered to be constant, as these are dynamic and thereby, this model neglects their time-dependent characteristics. To predict the growth of the disease trend efficiently and effectively, a time-dependent SIRVD model was proposed, which includes these factors of the SIRVD epidemic model with respect to time t_i . The proposed SIRVD epidemic model can reasonably trace the COVID-19 disease transmission and also predicts the future spread of the disease.

3.3 SIRVD epidemic time-dependent COVID-19 model

The SIRVD model, which is dependent on time, incorporates five parameters that change over time: the infection rate β , the recovery rate γ , the death rate δ , the vaccination rate α , and the susceptibility rate σ as in Liao et al. (27). These parameters are represented as functions of time, denoted as $\beta(t_i), \gamma(t_i), \delta(t_i), \alpha(t_i)$, and $\sigma(t_i)$ (27). The differential equations have been adjusted as follows (Equations 8–12):

$$\frac{dS_r(t_i)}{d(t_i)} = -\frac{\beta(t_i) I_r(t_i) S_r(t_i)}{N} + \sigma(t_i) R_r(t_i) - \alpha(t_i) S_r(t_i) \quad (8)$$

$$\frac{dI_r(t_i)}{d(t_i)} = \frac{\beta(t_i) I_r(t_i) S_r(t_i)}{N} - \gamma(t_i) I_r(t_i) - \delta(t_i) I_r(t_i) \quad (9)$$

$$\frac{dR_r(t_i)}{d(t_i)} = \gamma(t_i) I_r(t_i) - \sigma(t_i) R_r(t_i) \quad (10)$$

$$\frac{dV_r(t_i)}{d(t_i)} = \alpha(t_i) S_r(t_i) \quad (11)$$

$$\frac{dD_r(t_i)}{d(t_i)} = \delta(t_i) I_r(t_i) \quad (12)$$

N is a constant across the population, then the sum of each population's gain or decrease in the state equals to zero (as shown in Equation 13).

$$\frac{dS_r(t_i)}{d(t_i)} + \frac{dI_r(t_i)}{d(t_i)} + \frac{dR_r(t_i)}{d(t_i)} + \frac{dV_r(t_i)}{d(t_i)} + \frac{dD_r(t_i)}{d(t_i)} = 0 \quad (13)$$

Since the COVID-19 data are updated regularly on daily basis, the Equations 8–12 can be changed to differential Equations 14–18.

$$S_r(t_{i+1}) - S_r(t_i) = -\frac{\beta(t_i)I_r(t_i)S_r(t_i)}{N} + \sigma(t_i)R_r(t_i) - \alpha(t_i)S_r(t_i) \quad (14)$$

$$I_r(t_{i+1}) - I_r(t_i) = \frac{\beta(t_i)I_r(t_i)S_r(t_i)}{N} - \gamma(t_i)I_r(t_i) - \delta(t_i)\delta I_r(t_i) \quad (15)$$

$$R_r(t_{i+1}) - R_r(t_i) = \gamma(t_i)I_r(t_i) - \sigma(t_i)R_r(t_i) \quad (16)$$

$$V_r(t_{i+1}) - V_r(t_i) = \alpha(t_i)S_r(t_i) \quad (17)$$

$$D_r(t_{i+1}) - D_r(t_i) = \delta(t_i)\delta I_r(t_i) \quad (18)$$

Since the human body would create antibodies to the virus, it is believed that the COVID-19 reinfection rate during transmission was approximately equal to zero (29).

Subsequently, the formula of $\gamma(t)$ can be expressed as (Equations 19–21):

$$\gamma(t_i) = \frac{R_r(t_{i+1}) - R_r(t_i)}{I_r(t_i)} \quad (19)$$

Similarly,

$$\alpha(t_i) = \frac{S_r(t_{i+1}) - S_r(t_i)}{V_r(t_i)} \quad (20)$$

$$\delta(t_i) = \frac{D_r(t_{i+1}) - D_r(t_i)}{I_r(t_i)} \quad (21)$$

Once the rate of death and recovery is computed, add up with Equation 13. Thus, $\beta(t_i)$, the time dependent parameter can be obtained using Equation 22.

$$\beta(ii) = \frac{(I_r(t_{i+1}) - I_r(t_i) + R_r(t_{i+1}) - R_r(t_i) + D_r(t_{i+1}) - D_r(t_i)) \times N}{I_r(ii) \times S_r(ii)} \quad (22)$$

3.4 Long short-term memory

Long Short Term Memory (LSTM) is a specialized deep learning-based RNN architecture that finds extensive use in practical applications of time series models (30). As a subclass of artificial neural networks, RNNs display dynamic behavior over time due to their interconnected nodes forming a directed graph along a temporal sequence. RNNs can process input sequences of varying lengths by leveraging their internal state or memory. An RNN can be precisely defined as a collection of analogous networks, each transmitting information to a different recipient, enabling them to connect prior knowledge with the current context. However, as this gap widens, RNNs may struggle to learn to establish meaningful relationships in the data, particularly focusing on short-term memory over long-term memory's influence.

To address the challenges of long-term dependencies, LSTM networks were introduced by Hochreiter and Schmidhuber (30). LSTMs have demonstrated exceptional proficiency in classifying and predicting from time series data. These networks are constructed as chains of replicated modules, each equipped with a unique structure. A typical LSTM unit comprises of memory cell, and three gates say, forget, input, and output. The memory cell possesses the ability of retaining information across extended time intervals, while the three gates discussed earlier controls the information flow in the cell. The output gate determines which value should be stored as the expected output, the input gate decides which additional information to record, and the forget gate selectively discards certain information from the cell state. Figure 1 illustrates the LSTM's structure, where lines connect entire vectors from one node's output to another node's input. The circles represent pointwise operations, while the yellow boxes denote the layers of the previously trained neural network.

The output of LSTM gates, which use sigmoid activation functions to process information, is either 0 or 1. "0" indicates that the gates are blocking everything, and "1" indicates that everything is able to pass past the gates. In the LSTM, the equations of gates are:

$$f_t = \sigma(w_f \cdot [a_{t-1}, z_t] + b_f) \quad (23)$$

$$i_t = \sigma(w_i \cdot [a_{t-1}, z_t] + b_i) \quad (24)$$

$$o_t = \sigma(w_o \cdot [a_{t-1}, z_t] + b_o) \quad (25)$$

From Equations 23–25, i_t , o_t and (30) represents three gates say, forget, input and output. The sigmoid function is denoted by the symbol σ , and x , represents the relevant weight for each LSTM block. a_{t-1} represents the preceding output at $t-1$, timestamp, while z_t denotes the current input vector at timestamp, t and b_x represents bias neurons for gate z . The formulas for the final output, candidate cell state, and cell state are given as follows:

$$\tilde{c}_t = \tanh(w_c \cdot [a_{t+1}, z_t] + b_c) \quad (26)$$

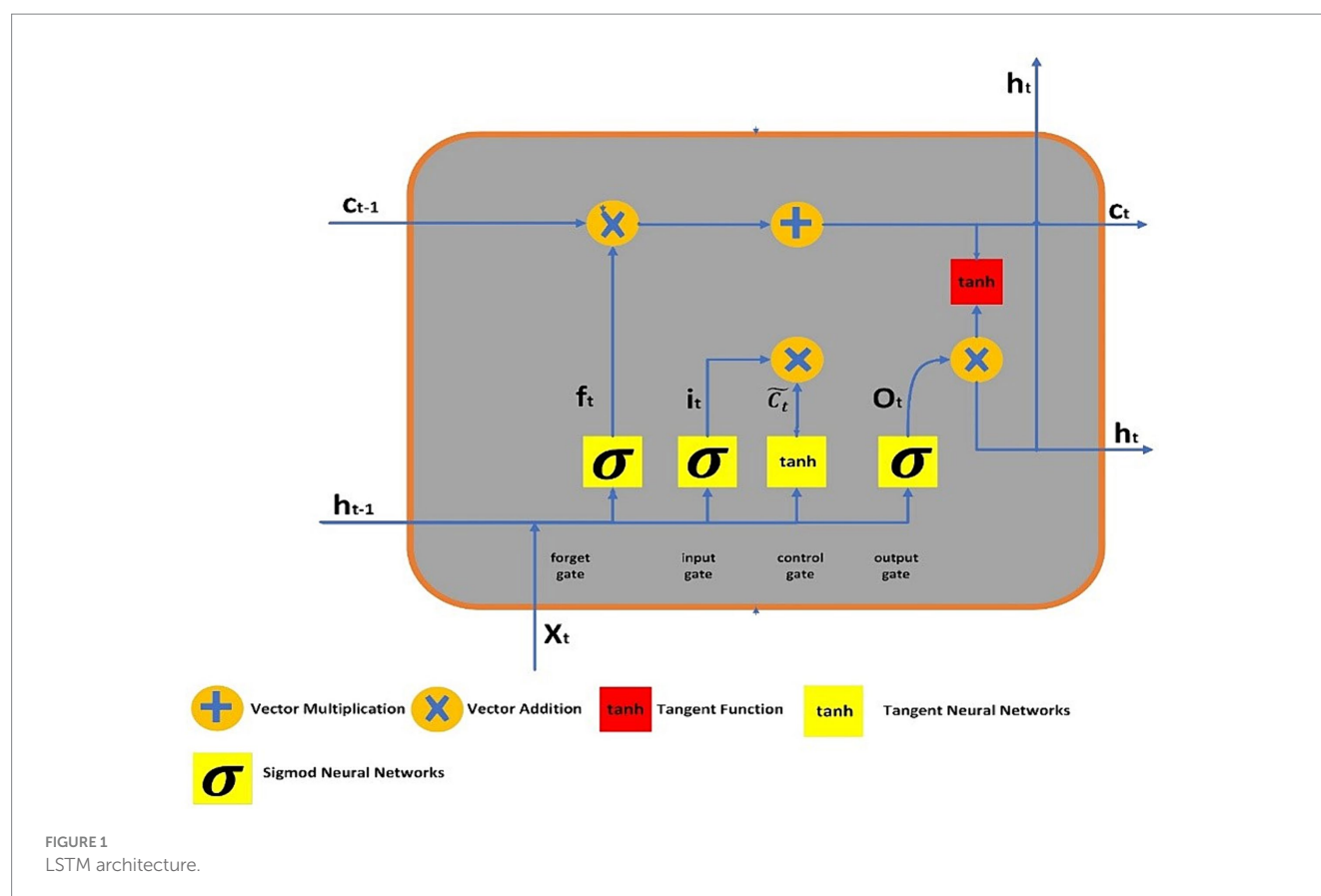
$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (27)$$

$$a_t = o_t * \tanh(c_t) \quad (28)$$

From the Equations 26–28, c_t and c_{t-1} depicts the current and preceding cell states or memory at t and $t-1$ timestamps, respectively. The term $\tilde{c}_t$ expresses to the output of the $\tanh$ function, which represents the potential cell state at timestamp t . The symbol $*$ denotes element-wise multiplication between vectors.

3.5 Proposed XGBoost-SIRVD-LSTM model

Figure 2 shows the suggested model's workflow details. The proposed XGBoost-SIRVD-LSTM model works in four phases: (1)



Data preprocessing, (2) Feature selection using XGBoost feature importance score, (3) SIRVD epidemic model construction, and (4) Prediction using LSTM. This model focuses mainly on the prediction of the recent trends of the epidemic based on the evaluation of the parameter changes in the epidemic. The remainder of this section explains the various stages of the suggested prediction model.

The steps for the proposed XGBoost-SIRVD-LSTM model are as follows:

Input: COVID-19 dataset containing confirmed cases, susceptible cases, recovered cases, deceased cases, and vaccination.

Output: estimating/predicting the COVID-19 infection rate.

Algorithm steps:

- 1 Implement data pre-processing techniques on the COVID-19 dataset.
- 2 Utilize the Min-Max approach to normalize the dataset.
- 3 For feature selection, use XGBoost feature importance score.
- 4 Develop the SIRVD epidemic model with the selected features from step 3.
- 5 Using step 4, the quantity of COVID-19 infection cases using LSTM are predicted.
- 6 Evaluate the proposed model using predictive performance metrics.

This section discusses the detailed steps involved in the proposed XGBoost-SIRVD-LSTM model for prediction.

3.5.1 Data pre-processing

The min-max normalization suggested in this paper to pre-process the COVID-19 data. Using below Equation 29, the feature values are normalized between [0, 1].

$$\text{min-max normalization, } y_i = \frac{y_i - \min}{\max - \min} \quad (29)$$

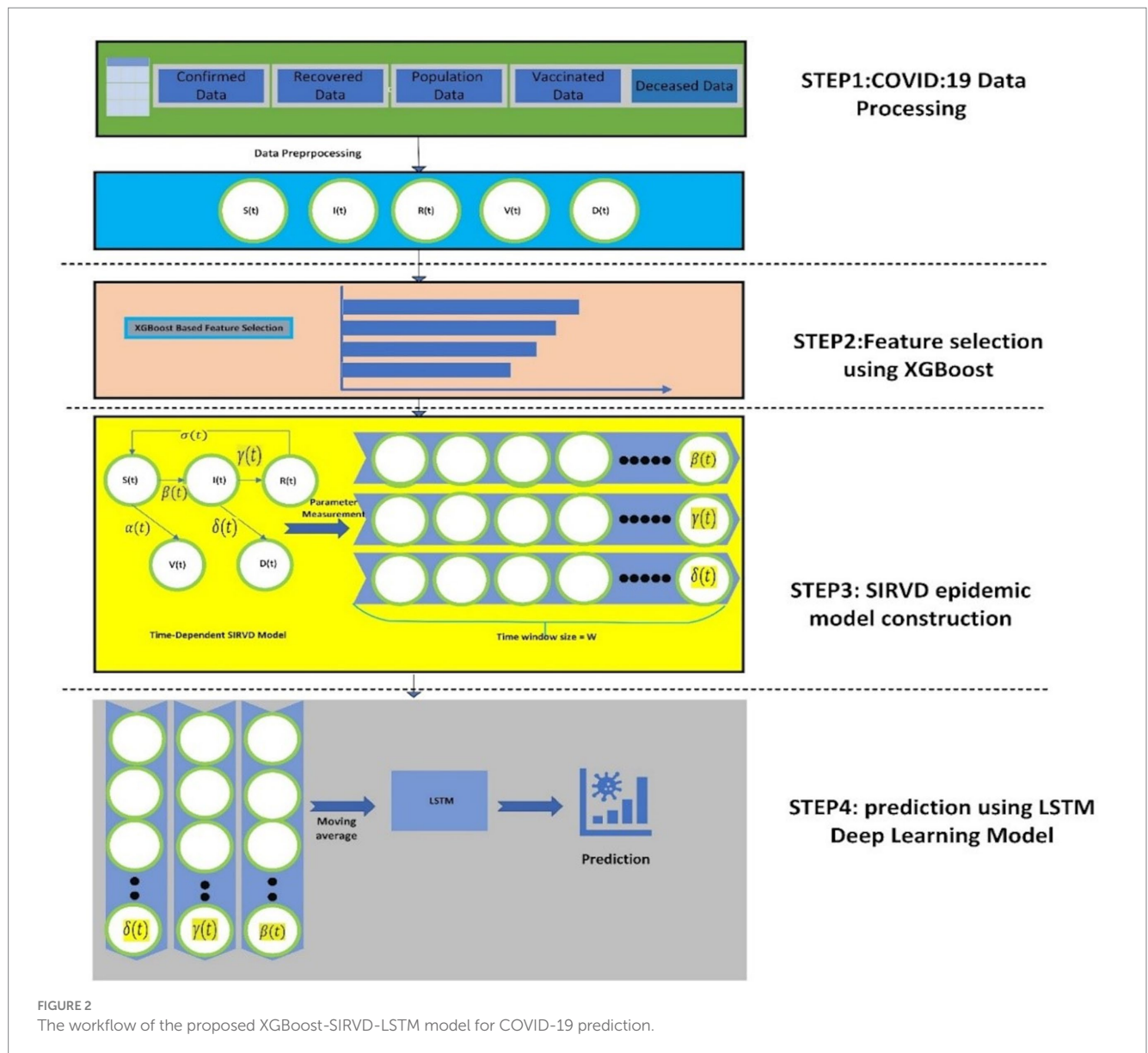
Where max denotes the highest value and min denotes the least value.

3.5.2 Feature selection using XGBoost feature importance score

The dataset pre-processed after step 1 used for feature optimization in this step. XGBoost feature importance score computed for the optimal selection of features from the COVID-19 dataset (19). Feature importance scores are normalized so that they sum up to 1 across all features. Higher scores indicate more important features relative to others in the dataset. Feature importance scores are useful for feature selection and understanding which features contribute most to the predictions made by the model.

3.5.3 SIRVD epidemic model construction

In this stage, the reduced-feature dataset obtained from step 2 is employed to construct the SIRVD epidemic model. The model incorporates five parameters: (infection) β , (recovery) γ , (death) δ , (vaccination) α , and (susceptibility) σ , which varies



over time represented by t (27). The dataset is prepared and formatted according to the specifications of the SIRVD model. The construction of the suggested SIRVD occurs once; dataset has been processed and transformed into the desired format.

3.5.4 Prediction using LSTM

The SIRVD model from step 3 is used for prediction using LSTM in this stage. In this study, single day prediction is computed for predicting the COVID-19 infection, and the model is tested with third, seventh, fourteenth, twenty-first- and twenty-eighth-days' prediction to evaluate the developed model's efficacy.

4 Results

This section describes the dataset in depth, including the evaluation metrics and efficacy evaluation of the suggested model.

4.1 Dataset

Extreme Due to the outbreak of COVID-19, multiple governments worldwide have made public their actions or measures and undertaken real-time data analysis to determine the disease's up-to-date trends. In this research study, two research data, which are publicly available are collected for experimentation of the proposed model, namely CSSE from Johns Hopkins University (7) and Our World in Data (8). The John Hopkins dataset comprises cumulative cases, including confirmed, recovered, and deceased at a global level. This dataset includes country, province, longitude, latitude, and total affected patients on a specified date as its features.

The data source from Our World in Data includes potential features of interest, namely confirmed and deceased cases, hospitalizations, vaccinations, and testing. The vaccination data obtained from this data source includes various information such as the country name (location), country code (iso_code), date of observation (date), total number of administered doses (total vaccinations), and the count of

vaccinated individuals (people_vaccinated). These data, in combination with the data from John Hopkins University, are utilized to implement and assess the proposed model.

4.2 Evaluation metrics

The performance of the XGBoost-SIRVD-LSTM model's performance involves comparing the observed and forecasted values. The evaluation metrics employed in this study include R^2 (determination coefficient) (Equation 32), normalized root mean square error (NRMSE) (Equation 31), root mean square error (RMSE) (Equation 30), and mean absolute percentage error (MAPE) (Equation 33) (31). The validation of the suggested model computed with the following formulas for calculating these metrics.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (30)$$

$$NRMSE = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}}{\bar{y}} \quad (31)$$

$$R^2 = \frac{\left(\sum_{i=1}^N ((x_i - \bar{x})(y_i - \bar{y})) \right)^2}{\sum_{i=1}^N (x_i - \bar{x})^2 \times \sum_{i=1}^N (y_i - \bar{y})^2} \quad (32)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right| * 100\% \quad (33)$$

4.3 Performance evaluation

The evaluation of the proposed model involves the utilization of datasets mentioned above. The experiments are conducted using Python, with deep learning libraries: numpy, pandas, keras, and tensorflow. The experimentation is performed on hardware with the following specifications: Intel (R) Core i7-8750H CPU @ 2.20 GHz, 64-bit operating system, RAM of 8.00 GB, and with GPU.

The architecture of deep learning models is determined by their hyper-parameters, which play a crucial role in achieving high-quality models. In this study, the optimal hyper-parameters are determined using a grid search approach. Table 1 presents the hyper-parameters utilized in the developed model. The dataset is split as training and testing sets in the ratio of 70:30 and implemented in training and testing the proposed COVID-19 infection case prediction model. The evaluation metrics described in the equations above are used in this study, and Table 2 compares the single-day prediction results of the developed model with existing models in literature.

The effectiveness of the proposed model is assessed by comparing its outcomes with those of existing literature on recurrent deep learning models, including bidirectional LSTM, GRU, Stacked LSTM, Vanilla

TABLE 1 Hyper-parameters for the proposed model.

Hyper-parameter	Test values
Optimizer	{SGD, ADAGRAD, Adam}
Learning rate	{0.01, 0.1, 0.5}
Batch size	{64, 128, 256}
Epochs	{1,000, 2,000, 3,000}

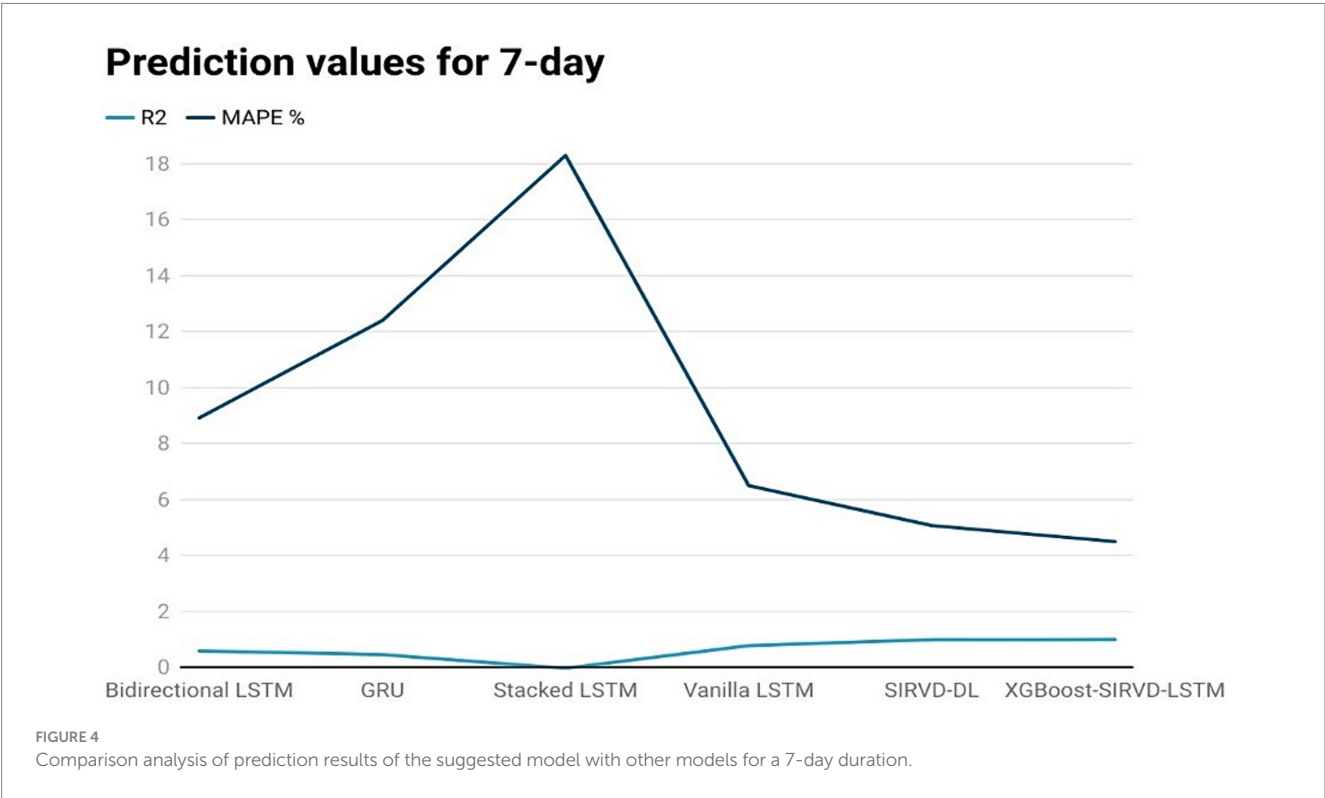
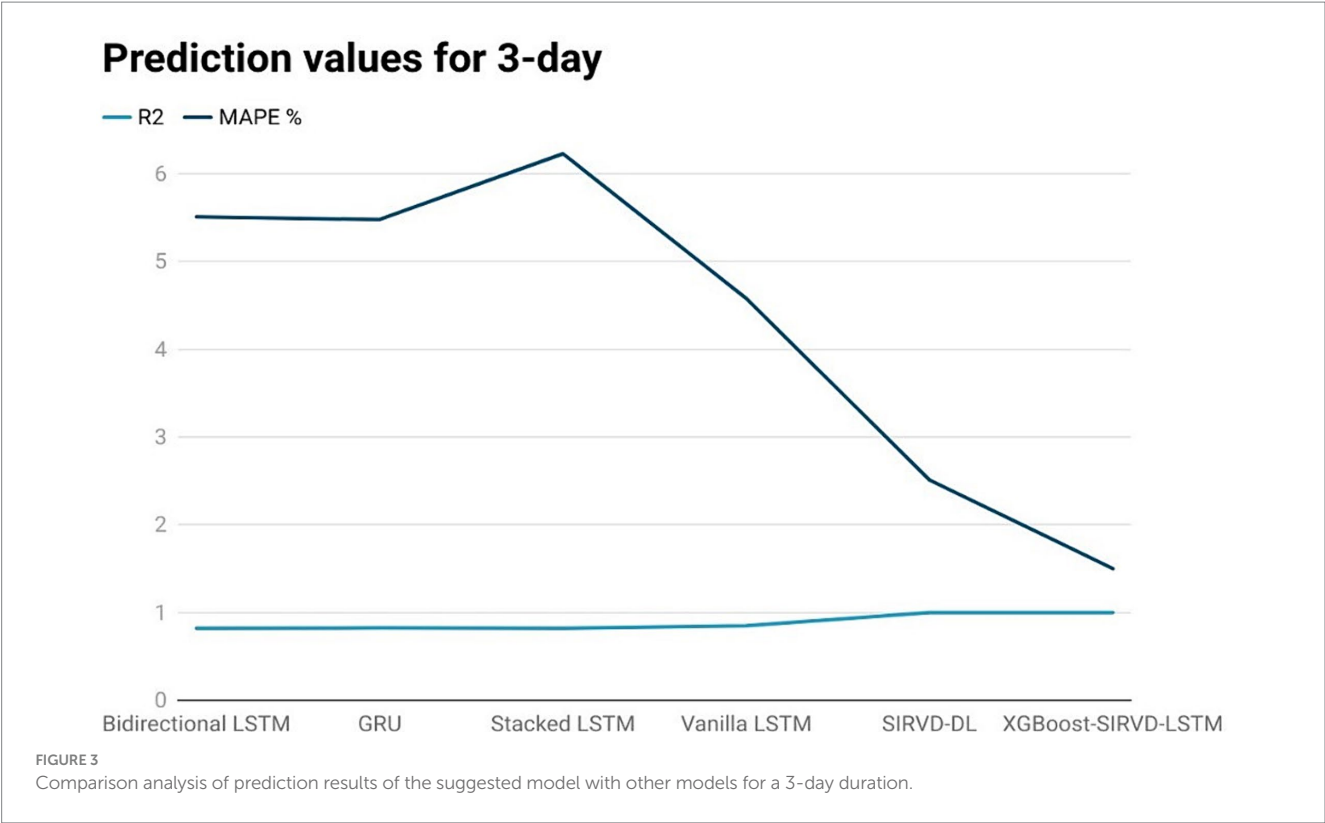
TABLE 2 Results depicting prediction for a single day with the proposed model as well as other models.

Model	R^2	MAPE	RMSE	NRMSE
Bidirectional LSTM	0.92	3.66	145,200	0.05
GRU	0.96	1.89	89,782	0.03
Stacked LSTM	0.96	2.15	92,065	0.03
Vanilla LSTM	0.92	3.29	151,580	0.05
SIRVD-DL	0.99	0.92	38,519	0.01
XGBoost-SIRVD-LSTM	0.99	0.90	35,025	0.01

LSTM, and SIRVD-DL (27). The unique combination of machine learning and mathematical modeling makes the XGBoost-SIRVD-LSTM model better than others. First, using XGBoost for feature selection helps the model find and prioritize key variables, enhancing prediction accuracy. Second, adding the SIRVD model captures COVID-19 transmission dynamics between susceptible, infected, recovered, vaccinated, and deceased populations. Thirdly, LSTM's sequential data learning allows it to capture COVID-19 temporal patterns and trends. Our comprehensive strategy combines the benefits of each component, resulting in improved prediction accuracy in empirical data. This integrative approach yields more accurate estimates than machine learning or epidemiological models. The experiments were specifically conducted to predict outcomes for the third, seventh, fourteenth, twenty-first, and twenty-eighth days. The experimental results are presented in Figures 3–7. To evaluate the performance of the proposed model, the obtained results are compared to those of other recurrent deep learning models, such as bidirectional LSTM, GRU, stacked LSTM, vanilla LSTM, and SIRVD-DL (27). The experiments were accurately performed to predict outcomes for the third day, seventh day, fourteenth day, twenty-first day, and twenty-eighth day. The experimental findings are displayed in Figures 4–7. Similarly, the proposed model resulted with the R^2 score of 0.999 on the 3-day, 0.997 on the 7-day, 0.956 on the 14-day, 0.64 on the 21-day, and 0.19 on the 28-day. When compared to other models that were taken into consideration for evaluation, the R^2 score grows comparatively as the number of predicting days' rises, demonstrating the effectiveness of the suggested model. The other models consequently displayed negative values as the number of days increased, indicating that the fitting function's prediction error was higher than the mean function. As a result, the prediction models' performance when combined with other models is ineffective.

From the preceding discussion, the contributions of the proposed model can be summarized as follows:

- 1 A new XGBoost-SIRVD-LSTM model is introduced for predicting COVID-19 infection cases. This model combines XGBoost for feature selection and integrates the SIRVD epidemic model with LSTM for disease prediction.

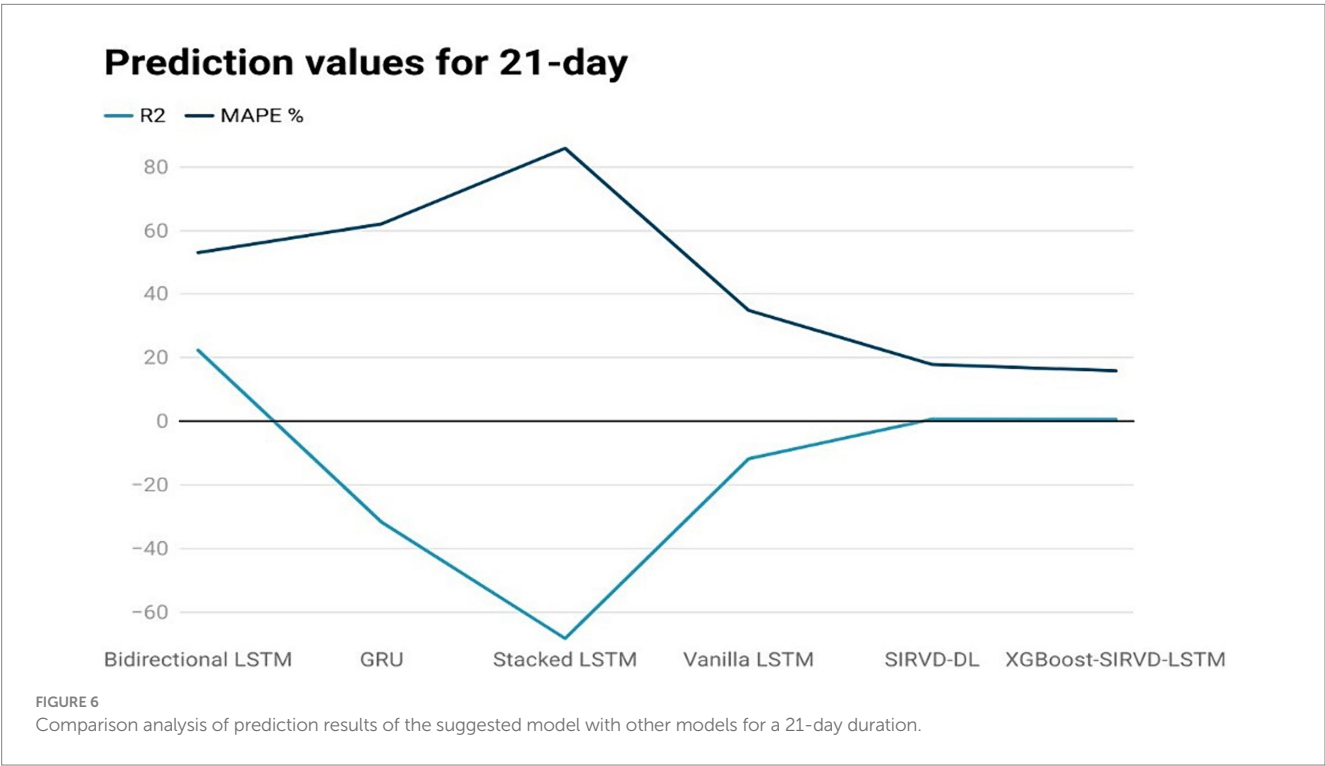
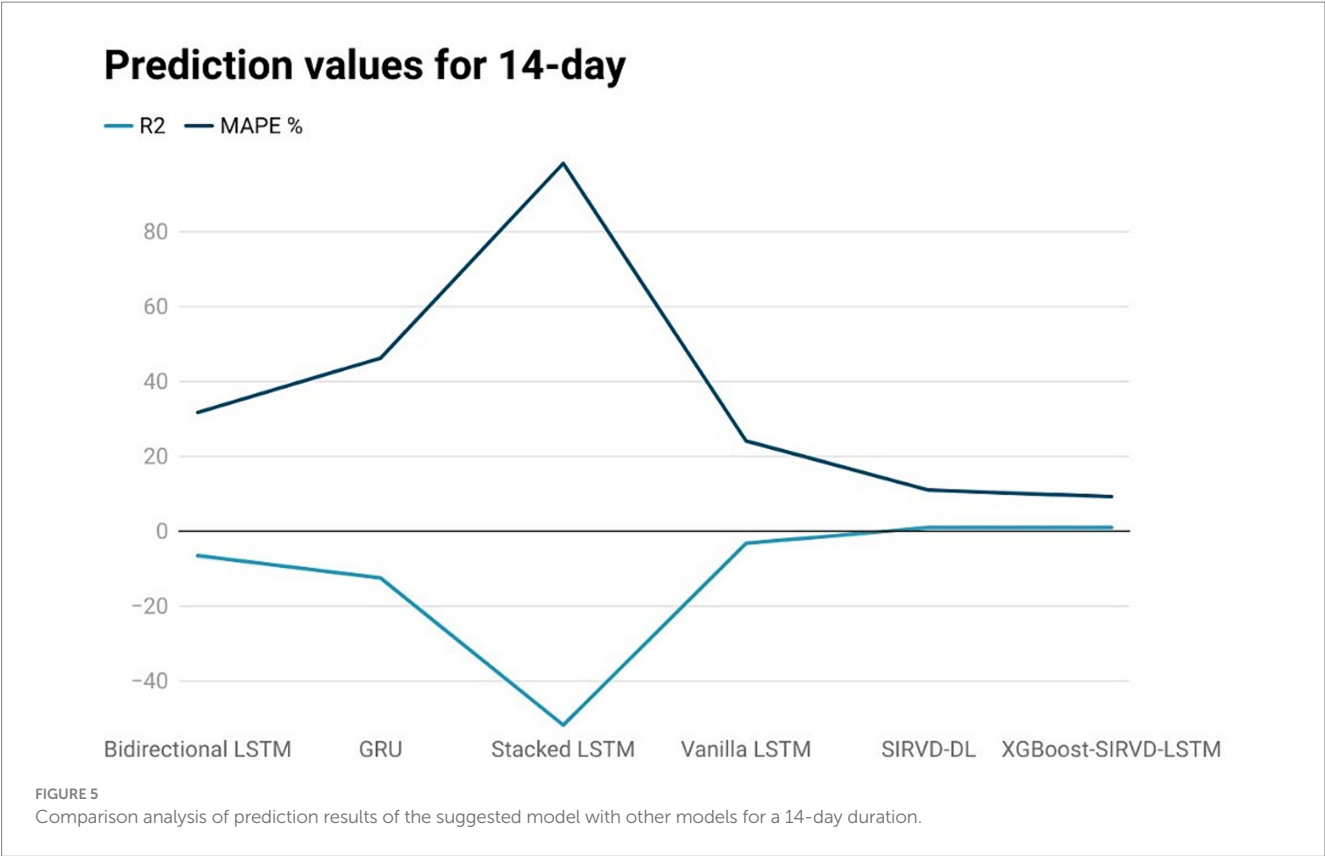


2 SIRVD-DL and other recurrent deep learning models were used to compare the efficacy of the suggested model.

When compared to previous models, the performance of the proposed XGBoost-SIRVD-LSTM produced improved predictions.

5 Conclusion

This research work introduces an innovative model that merges mathematical and machine learning methodologies to forecast the future trajectory of COVID-19. The XGBoost-SIRVD-LSTM model



represents a significant advancement in forecasting the course of COVID-19, offering a solution to the critical challenge of precise prediction in the face of a dynamically evolving pandemic. By harmonizing the strengths of XGBoost for feature selection with the SIRVD model's capacity to track COVID-19 transmission over time, this research provides a comprehensive approach for pandemic forecasting. The dataset is processed using LSTM to provide disease predictions. The model is evaluated using the Our World in Data and

Prediction values for 28-day



FIGURE 7

Comparison analysis of prediction results of the suggested model with other models for a 28-day duration.

CSSE datasets from John Hopkins University. The experimental findings illustrate that the suggested model surpasses alternative deep learning models in terms of performance, exhibiting superior prediction accuracy and precision. These findings suggest that the model proposed will be one of a valuable resource for forecasting the future course of COVID-19. It has the potential to assist governments and public health experts in making informed decisions and formulating effective strategies to combat the pandemic.

Here are some specific potential future research trajectories:

- 1 Increase the model's precision and accuracy. More data, more advanced machine learning algorithms, or a mix of the two may be used to achieve this.
- 2 Improve the model's usability. This could be achieved by creating a user interface that makes it simple for users to enter data and generate predictions.
- 3 Predict the efficacy of various therapies using the model. Governments and public health professionals may utilize this information to assist in choosing which actions to prioritize.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material; further inquiries can be directed to the corresponding author.

Author contributions

HA: Conceptualization, Funding acquisition, Project administration, Resources, Writing – original draft, Writing

– review & editing. DP: Conceptualization, Data curation, Formal analysis, Software, Validation, Writing – original draft. NK: Conceptualization, Methodology, Supervision, Validation, Writing – review & editing. MA: Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. UU: Conceptualization, Methodology, Resources, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was funded by King Abdulaziz City for Science and Technology (KACST) via the Fast Track Funding Path for COVID-19 Research Projects, grant number 5-21-01-001-0036.

Acknowledgments

The authors extend their appreciation to King Abdulaziz City for Science and Technology (KACST) to fund this work via the Fast Track Funding Path for COVID-19 Research Projects, grant number 5-21-01-001-0036.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- WHO (2023). WHO Coronavirus (COVID-19) Dashboard 2023. Available online at: <https://covid19.who.int/> (Accessed on May 2, 2024).
- Kermack WO, McKendrick AG, Walker GT. A contribution to the mathematical theory of epidemics. *Proc Roy Soc London Ser A Contain Papers Math Phys Charact.* (1997) 115:700–21.
- Rahimi I, Gandomi AH, Asteris PG, Chen F. Analysis and prediction of COVID-19 using SIR, SEIQR, and machine learning models: Australia, Italy, and UK cases. *Inform.* (2021) 12. doi: 10.3390/info12030109
- Youssef H, Alghamdi N, Ezzat MA, El-Bary AA, Shawky AM. Study on the SEIQR model and applying the epidemiological rates of COVID-19 epidemic spread in Saudi Arabia. *Infect Dis Model.* (2021) 6:678–92. doi: 10.1016/j.idm.2021.04.005
- He S, Peng Y, Sun K. SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dynam.* (2020) 101:1667–80. doi: 10.1007/s11071-020-05743-y
- Fatimah B, Aggarwal P, Singh P, Gupta A. A comparative study for predictive monitoring of COVID-19 pandemic. *Appl Soft Comput.* (2022) 122:108806. doi: 10.1016/j.asoc.2022.108806
- Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis.* (2020) 20:533–4. doi: 10.1016/S1473-3099(20)30120-1
- Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C, et al. A global database of COVID-19 vaccinations. *Nat Hum Behav.* (2021) 5:947–53. doi: 10.1038/s41562-021-01122-8
- Kartono A, Karimah SV, Wahyudi ST, Setiawan AA, Sofian I. Forecasting the long-term trends of coronavirus disease 2019 (COVID-19) epidemic using the susceptible-infectious-recovered (SIR) model. *Infect Dis Rep.* (2021) 13:668–84. doi: 10.3390/idr13030063
- ArunKumar KE, Kalaga DV, Kumar CMS, Kawaji M, Brenza TM. Forecasting of COVID-19 using deep layer recurrent neural networks (RNNs) with gated recurrent units (GRUs) and long short-term memory (LSTM) cells. *Chaos Solitons Fract.* (2021) 146:110861. doi: 10.1016/j.chaos.2021.110861
- Alanazi SA, Kamruzzaman MM, Alruwaili M, Alshammari N, Alqahtani SA, Karime A. Measuring and preventing COVID-19 using the SIR model and machine learning in smart health care. *J Healthcare Eng.* (2020) 2020:1–12. doi: 10.1155/2020/8857346
- El-Attar N-E, Sabbah S-F, Fasihuddin H, Awad W-A. An improved DeepNN with feature ranking for Covid-19 detection. *Comput Mater Contin.* (2022) 71:2249–69. doi: 10.32604/cmc.2022.022673
- Schlickeiser R, Kröger M. Mathematics of epidemics: on the general solution of SIRVD, SIRV, SIRD, and SIR compartment models. *Mathematics.* (2024) 12:941. doi: 10.3390/math12070941
- Amiri Babaei N, Özer T. On exact integrability of a Covid-19 model: SIRV. *Math Methods Appl Sci.* (2024) 47:3529–46. doi: 10.1002/mma.8874
- Federico S, Ferrari G, Torrente M-L. Optimal vaccination in a SIRS epidemic model. *Economic Theory.* (2024) 77:49–74. doi: 10.1007/s00199-022-01475-9
- Ali RH, Abdulsalam WH. The prediction of COVID 19 disease using feature selection techniques. *J Phys Conf Ser.* (2021) 1879:022083. doi: 10.1088/1742-6596/1879/2/022083
- Chandra R, Jain A, Singh CD. Deep learning via LSTM models for COVID-19 infection forecasting in India. *PLoS One.* (2022) 17:e0262708. doi: 10.1371/journal.pone.0262708
- Alassafi MO, Jarrah M, Alotaibi R. Time series predicting of COVID-19 based on deep learning. *Neurocomputing.* (2022) 468:335–44. doi: 10.1016/j.neucom.2021.10.035
- Devan P, Khare N. An efficient XGBoost–DNN-based classification model for network intrusion detection system. *Neural Comput & Applic.* (2020) 32:12499–514. doi: 10.1007/s00521-020-04708-x
- Devan P, Khare N. EFS-LSTM (ensemble-based feature selection with LSTM) classifier for intrusion detection system. *Int J e-Collab.* (2020) 16:72–86. doi: 10.4018/IJec.2020100106
- Singh P, Gupta A. Generalized SIR (GSIR) epidemic model: an improved framework for the predictive monitoring of COVID-19 pandemic. *ISA Trans.* (2022) 124:31–40. doi: 10.1016/j.isatra.2021.02.016
- Cooper I, Mondal A, Antonopoulos CG. A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons Fractals.* (2020) 139:110057. doi: 10.1016/j.chaos.2020.110057
- Devaraj J, Elavarasan RM, Pugazhendhi R, Shafiullah GM, Ganesan S, Jeysree AK, et al. Forecasting of COVID-19 cases using deep learning models: is it reliable and practically significant? *Results Phys.* (2021) 21:103817. doi: 10.1016/j.rinp.2021.103817
- Liew XY, Hameed N, Clos J. An investigation of XGBoost-based algorithm for breast cancer classification. *Mach Learn Appl.* (2021) 6:100154. doi: 10.1016/j.mlwa.2021.100154
- Zheng Y, Zhu Y, Ji M, Wang R, Liu X, Zhang M, et al. A learning-based model to evaluate hospitalization priority in COVID-19 pandemics. *Patterns.* (2020) 1:100092. doi: 10.1016/j.patter.2020.100092
- Farooq J, Bazaz MA. A novel adaptive deep learning model of Covid-19 with focus on mortality reduction strategies. *Chaos, Solitons Fractals.* (2020) 138:110148. doi: 10.1016/j.chaos.2020.110148
- Liao Z, Lan P, Fan X, Kelly B, Innes A, Liao Z. SIRVD-DL: a COVID-19 deep learning prediction model based on time-dependent SIRVD. *Comput Biol Med.* (2021) 138:104868. doi: 10.1016/j.combiomed.2021.104868
- Usherwood T, LaJoie Z, Srivastava V. A model and predictions for COVID-19 considering population behavior and vaccination. *Sci Rep.* (2021) 11:12051. doi: 10.1038/s41598-021-91514-7
- Phelan AL. COVID-19 immunity passports and vaccination certificates: scientific, equitable, and legal challenges. *Lancet.* (2020) 395:1595–8. doi: 10.1016/S0140-6736(20)31034-5
- Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* (1997) 9:1735–80. doi: 10.1162/neco.1997.9.8.1735
- Kafieh R, Arian R, Saeedizadeh N, Amini Z, Serej ND, Minaee S, et al. COVID-19 in Iran: forecasting pandemic using deep learning. *Comput Math Methods Med.* (2021) 2021:1–16. doi: 10.1155/2021/6927985



OPEN ACCESS

EDITED BY

Sultan Ahmad,
Prince Sattam Bin Abdulaziz University,
Saudi Arabia

REVIEWED BY

Neeraj Kumar Pandey,
Graphic Era University, India
Jackson Burton,
Biogen Idec, United States
Irfan Uddin,
Kohat University of Science and Technology,
Pakistan

*CORRESPONDENCE

Nesren Farhah
✉ n.farhah@seu.edu.sa

RECEIVED 23 June 2024

ACCEPTED 23 August 2024

PUBLISHED 04 September 2024

CITATION

Farhah N (2024) Utilizing deep learning
models in an intelligent spiral drawing
classification system for Parkinson's disease
classification.
Front. Med. 11:1453743.
doi: 10.3389/fmed.2024.1453743

COPYRIGHT

© 2024 Farhah. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Utilizing deep learning models in an intelligent spiral drawing classification system for Parkinson's disease classification

Nesren Farhah*

Department of Health Informatics, College of Health Sciences, Saudi Electronic University, Riyadh, Saudi Arabia

Introduction: Parkinson's disease (PD) is a neurodegenerative illness that impairs normal human movement. The primary cause of PD is the deficiency of dopamine in the human brain. PD also leads to several other challenges, including insomnia, eating disturbances, excessive sleepiness, fluctuations in blood pressure, sexual dysfunction, and other issues.

Methods: The suggested system is an extremely promising technological strategy that may help medical professionals provide accurate and unbiased disease diagnoses. This is accomplished by utilizing significant and unique traits taken from spiral drawings connected to Parkinson's disease. While PD cannot be cured, early administration of drugs may significantly improve the condition of a patient with PD. An expeditious and accurate clinical classification of PD ensures that efficacious therapeutic interventions can commence promptly, potentially impeding the advancement of the disease and enhancing the quality of life for both patients and their caregivers. Transfer learning models have been applied to diagnose PD by analyzing important and distinctive characteristics extracted from hand-drawn spirals. The studies were carried out in conjunction with a comparison analysis employing 102 spiral drawings. This work enhances current research by analyzing the effectiveness of transfer learning models, including VGG19, InceptionV3, ResNet50v2, and DenseNet169, for identifying PD using hand-drawn spirals.

Results: Transfer machine learning models demonstrate highly encouraging outcomes in providing a precise and reliable classification of PD. Actual results demonstrate that the InceptionV3 model achieved a high accuracy of 89% when learning from spiral drawing images and had a superior receiver operating characteristic (ROC) curve value of 95%.

Discussion: The comparison results suggest that PD identification using these models is currently at the forefront of PD research. The dataset will be enlarged, transfer learning strategies will be investigated, and the system's integration into a comprehensive Parkinson's monitoring and evaluation platform will be looked into as future research areas. The results of this study could lead to a better quality of life for Parkinson's sufferers, individualized treatment, and an early classification.

KEYWORDS

Parkinson's disease, transfer learning models, deep learning, hand drawing, E-health

1 Introduction

Parkinson's disease (PD) is a chronic deteriorating illness that primarily affects the motor system of the central nervous system. Its indications often manifest gradually, and as the disease progresses, non-motor indications become more prevalent. The primary indications are tremors, stiffness, bradykinesia, and gait disturbances. PD may also result in dysphoria, apprehension, sleep disturbances, sensory impairments, and alterations in behavior. Environmental factors and genetic inheritance are significant contributors to the development of PD (1, 2).

In 2019, a World Health Organization research reported that approximately 8.5 million individuals are diagnosed with PD (3). The prevalence of this condition increases with age, with only 4% of afflicted persons younger than 50 years old. PD is a highly prevalent neurological disorder worldwide, ranking as the second most common condition after Alzheimer's disease. It affects a significant number of people, as evidenced by the data from sources (4, 5). Currently, therapists have limitations in effectively treating the symptoms of this condition as interventions are still in their early stages (6). The main tool used to determine a PD classification (PDD) is the patient's medicinal past; however, such classification remains uncertain (3). Thus, it is critical to offer a simple and reliable method for detecting this disease in order to save time and money on invasive classification and treatment (7, 8).

Patients with PD may exhibit a broad variety of non-motor symptoms, including mood disorders and depression, among others. These symptoms, including language and other relevant aspects, may manifest in the patient's facial expressions (9). The present study aims to analyze the effect of PD on both motor and non-motor abilities by applying handwriting modeling methodologies, with a special focus on spirals. This study seeks to fill a current knowledge gap by exploring the potential of spiral drawing as a tool for PD assessment.

Spiral drawing is a sophisticated and intricate motor skill that requires coordination. Consequently, it is regarded as an accurate evaluation of motor function. The Motion Rating Scale and its subcategory, The Unified PD Rating Scale (UPDRS-III), are the predominant and universally acknowledged rating scales for assessing PD. PD impacts a range of bodily processes, including speaking, handwriting, walking, and coordination, all of which are classified as motor functions. Various methods for measuring motor decline and non-motor biomarkers have been proposed to assess the severity of PD, which is considered a motor condition resulting from neurodegeneration. Both the classification and intensive care of PD are expensive and challenging because of two primary factors: (1) the inconvenience faced by caregivers in transporting the patient to the clinic and (2) the need for skilled medical professionals to conduct physical examinations and make diagnoses based on their observations. Clinical invasive techniques are only accessible at the early stage of the disease, and they carry risks and require considerable resources, especially in underdeveloped regions of the world. These techniques are only beneficial if early classification is achieved (10, 11).

At present, there is no accurate standard for making an objective finding of PD. When a non-specialist makes the classification, the likelihood of a mistake increases dramatically. There is a 20% chance of making a wrong classification in such instances (12). The accuracy of the classification is improved by carefully analyzing the main indications, which include tremors, bradykinesia, and stiffness. Having said that,

physician bias may creep into clinical assessments. Medical choice support systems are attracting interest for their capability to enhance objectivity and facilitate early classification. An early identification of PD will enable the development of tailored interventions for people with PD (13, 14). A crucial objective in the study of neurodegenerative illnesses is to discover precise biomarkers (15). Within the literature, several research have been conducted to diagnose PD by analyzing speech. These studies (16–18) mostly use sustained vowels and natural speech for diagnostic purposes. Motor symptoms may also be identified and monitored by analyzing patients' motions and gait (19, 20).

Several techniques have been created to examine the handwriting of patients with PD (21). Both static and dynamic characteristics are intriguing, including factors such as speed and the lowering of pen pressure throughout the handwriting (22). Numerous recent review studies have been published (23, 24). The legibility of an individual's handwriting is influenced by their visual acuity, writing technique, and linguistic proficiency, resulting in significant differences across individuals. A viable substitute for handwriting is the use of illustrations. Deep learning (DL) models have greatly revolutionized biomedical and medical image analysis (25). DL approaches have been applied in different domains, including segmentation, detection, classification, and classification (11), owing to their exceptional capability to extract sophisticated features, leading to enhanced accuracy in illness categorization. This may mostly be ascribed to their remarkable ability to generalize. Convolutional neural networks (CNNs) have been crucial in promoting the progress of the medical imaging field, achieving notable success in several medical image classification tasks (19, 20).

1.1 Main contribution

Spiral drawing is a sophisticated and intricate motor skill that requires coordination. Accordingly, it is regarded as an accurate evaluation of motor function and an initial examination for early indications of PD. This article proposes a method for PDD by analyzing spiral drawings and employing transfer learning models. The method categorizes an individual as either healthy or diagnoses them with PD based on their spiral drawing. A spiral drawing produced by a healthy individual will closely resemble a typical spiral form. By contrast, a spiral created by an individual with PD will exhibit significant deviation from a flawless spiral form and appear twisted because of the individual's sluggish motor movements and diminished synchronization between the hand and the brain.

2 Related works

Drotar and colleagues planned the utilization of a feature selection algorithm and support vector machine (SVM) approach to analyze the handwriting of patients with PD (26, 27). Their study is one of the first efforts to analyze the results of hand motions in the air or on a surface for diagnosing motor disorders associated with neurodegenerative illnesses. The findings revealed that these motions have a significant influence on the evaluation of handwriting and achieve a prediction accuracy of 85.61% (26). The work featured the PaHaW handwriting database, which was created by having individuals with PD complete eight distinct handwriting challenges, one being the Archimedean spiral. Basnin et al. (27) demonstrated their approach by using deep

transfer learning, achieving a testing accuracy of 91.36%. The research only used a dataset consisting of 800 hand-drawn spiral pictures. Das et al. (28) investigated a sophisticated technique for identifying PD using pictures that were hand-drawn by the patients. The authors combined discrete wavelet transform coefficients with histograms of oriented gradient data to enhance the accuracy of detection rate. They revealed the effectiveness of integrating these methods to extract pertinent information and identify vital coefficients, resulting in improved accuracy in disease detection using machine learning techniques. They specifically highlighted the efficacy of random forest (RF) and SVM approaches when applied to spiral pattern features of images.

Researchers have discovered that studying handwriting or hand drawings is a more efficient method for identifying PD (29). Shaban (30) advocated for the use of a meticulously adjusted VGG19 model that applies spiral and wave handwriting patterns to diagnose conditions. The dataset used was of limited size and comprised 102 wave photos and 102 spiral images. Data augmentation, such as applying picture rotation, was used to alleviate the problem of model overfitting. After implementing 10-fold cross-validation, the CNN model demonstrated impressive accuracies of 88 and 89% for the wave and spiral pictures, respectively. Megha Kamble et al. (31) proposed a comprehensive examination of the static and dynamic spirals created by people with Parkinson's disease. To do this, we extracted kinematic characteristics related to movement in the air and on the surface from data files created for 25 patients and 15 healthy controls. We utilized mathematical models for this purpose. Gil-Martín (32) this study contributes to the ongoing endeavor by examining a convolutional neural network (CNN) for the purpose of detecting PD based on drawing gestures. The analysis was conducted with a publicly available dataset: Digitized graphics are utilized to create spiral drawings for Parkinson's disease. Donalto Impedovo et al. (33) have proposed handwriting as a robust indicator for the development of a diagnostic tool for Parkinson's disease. The authors have applied a machine learning classification framework to the PaHaW dataset and achieved high specificity performance scores. Marta San Lucianol et al. (34) proposed the utilization of spiral drawing for computerized analysis of PD, as digitized spirals demonstrate a correlation with motor scores. The indices that are generated or calculated that have a correlation with the overall execution of a spiral include severity, shape, and kinematic irregularity. Kinematic irregularity includes second order smoothness and first order zero crossing. Other indices include tightness, mean speed, and variability of spiral width. Theyazn H. H. Aldhyani et al. (10) study makes a contribution by utilizing deep learning models to diagnose PD using photos of spiral and wave drawings. Manju Singh et al. (35) aims to provide a method for detecting PD utilizing spiral sketching and convolutional neural networks (CNN). The core concept is to examine an individual's spiral drawings and categorize them as either indicative of good health or indicative of Parkinson's disease. The spiral doodles produced by individuals in good health bear a striking resemblance to conventional helical forms. Table 1 presents a concise summary of the key attributes of prior studies on PD identification using drawings and other datasets.

3 Materials and methods

This section details the planned methodology applied to develop a PDD system based on DL techniques, specifically designed to detect PD from features extracted from spiral drawing images. This methodology includes dataset collection, data preprocessing, DL

classification models, evaluation metrics, and results analysis. The framework of this methodology is shown in Figure 1.

3.1 Dataset collection

For our experimental study, we employed a dataset of spiral drawing images obtained from the Kaggle platform. This dataset, which was created by Adriano et al. (36) based on the NIATS of the Federal University, includes digital records of 102 spiral image samples, with 51 from Parkinson's disease patients (PDP) and 51 from healthy persons. The images have been pre-split into a training set and a testing set (Figure 2).

3.2 Data preprocessing

For our experimental work on PDD using drawn spiral images, we utilized a comprehensive dataset from the Kaggle platform. This dataset includes digital drawings from 51 PDPs and 51 healthy individuals. The processing steps are presented in Figure 3.

3.2.1 Data loading and preparation

The dataset was divided into two classes: "healthy" and "parkinson." Each image was resized to 100×100 pixels and converted to array format for consistency. Labels were encoded into binary format, where "healthy" was labeled as 0 and "parkinson" as 1. This preparation step ensured uniform input data for the model.

3.2.2 Data augmentation

To increase the diversity and robustness of the training dataset, we applied data augmentation techniques using the Image Data Generator module, including rotation, shifting, and flipping of images. We likewise introduced variations that prevent overfitting and enhance the model's capability to generalize to novel, unnoticed image data.

3.2.3 Data splitting

In this step, we split the dataset into a training set and a testing set using an 80-20 split ratio. This stratification ensures a balanced representation of both classes in the training and testing phases.

3.2.4 Normalization and label encoding

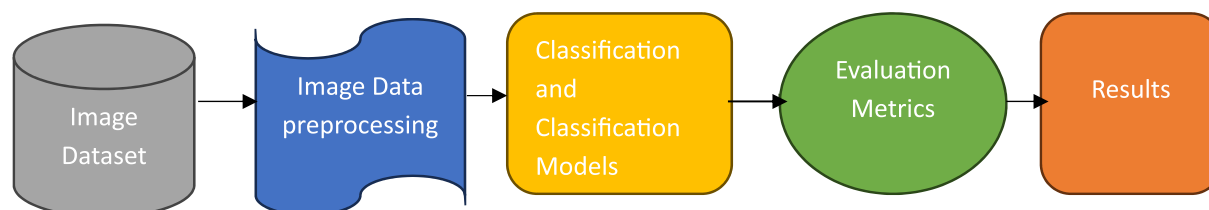
The pixel values of the images were standardized to the range [0, 1] to expedite the training process and improve model performance. Additionally, the labels were one-hot encoded to facilitate categorical classification.

3.3 Diagnoses and classification models

For the classification and classification of drawn spiral images into the "Parkinson" and "healthy" classes, we applied several advanced CNN architectures, including VGG19, InceptionV3, ResNet50v2, and DenseNet169. These models were pre-trained on the ImageNet dataset, which comprises over 14 million images across 1,000 categories. ImageNet provides a robust foundation for transfer learning due to its diverse range of visual concepts, although it does not inherently include clinical images.

TABLE 1 Overview of the current state of the art in employing various types of publicly available datasets based on artificial intelligence techniques.

Authors	Datasets	Approaches	Object of study
Vanegas et al. (42)	Parkinson's dataset in EGG	Decision tree approaches	For this study, authors employed machine learning techniques to create a model that can accurately detect the most significant indicators from the EEG spectra during visual stimulation. The purpose of this model is to aid in the classification of PD.
Oh et al. (43)	Parkinson's dataset in EGG	CNN model	This study utilized the electroencephalogram (EEG) data of twenty individuals with PD and twenty individuals without PD. An established CNN architecture consisting of thirteen layers effectively eliminates the requirement for traditional feature representation stages.
Prasuhn et al. (44)	Parkinson's dataset in using MRI images	SVM approach	The proposed work suggests utilizing computer-aided methods and a highly reproducible method, as opposed to manually segmenting Substantia nigra (SN) to enhance the dependability and precision of Diffusion Tensor Imaging (DTI) of the measurements employed for categorisation.
Rasheed et al. (45)	Parkinson's dataset in using voice	BPVAM	This study presents two classification algorithms aimed at enhancing the accuracy of identifying PD cases based on voice measures. Initially, implemented the BPVAM algorithm, which is a variable adaptive moment-based backpropagation algorithm of artificial neural networks (ANN).
Gunduz et al. (46)	Parkinson's dataset in using voice	GB model	This study presents two frameworks utilizing CNNs method to accurately classify PD by analyzing sets of vocal (voice) data. Both frameworks are used to combine different feature sets, but they differ in how they combine these sets.
Pdisher et al. (47)	Parkinson's dataset collected using sensor device	CNN model	Employed DL techniques to categorize motion data obtained from a solitary IMU sensor worn on the wrist, which was recorded in unstructured settings. In order to validate the results, patients were followed by a specialist in movement disorders, and their motor condition was assessed regularly and without active participation every minute.
Taliki et al. (48)	Parkinson's dataset in using sensory	Random forest	This article explores instances of misclassification and presents a proposed system for obtaining a second opinion. The system relies on wearable sensors and artificial intelligence. To address this issue, authors developed several standardized tasks and collected movement data using wearable sensors worn by persons diagnosed with PD other extrapyramidal illnesses.
Shaban et al. (30)	Parkinson's dataset using hand drawing	DL-based VGG16	This work explores the application of a fine-tuned VGG-19 model to screen for PD using a Kaggle handwriting dataset. The study involves conducting experiments to test the effectiveness of this approach. The dataset consisted of 102 wave and 102 spiral handwriting patterns.
Robin (49)	Parkinson's dataset using hand drawing	RestNet50	Developing RestNet50 to detect PD using of 102 of 102 wave and 102 spiral
Stpete_ishii (50)	Parkinson's dataset using hand drawing	CNN model	Developing online for classification of PD by using spiral images
Shaban et al. (30)	Handwriting dataset (same dataset)	CNN model	Developing online for classification of PD by using spiral images
Adrian (36)	Parkinson's dataset using hand drawing	CNN model	Developing online for classification of PD by using spiral images

FIGURE 1
Framework of the proposed methodology.

3.3.1 VGG19 model

We employed a CNN using the pre-trained VGG19 model to identify PD (37). The input layer accepts images resized to 100×100 pixels with three color channels. The model, pre-trained on the ImageNet dataset and excluding its top categorization layer, assists as

a feature mining with average pooling. This is followed by a custom dense layer with 64 units and ReLU activation to introduce nonlinearity. The final layer is a dense output layer with 2 units and softmax activation, designed for binary classification between healthy individuals and PDPs. The model is compiled with the Adam

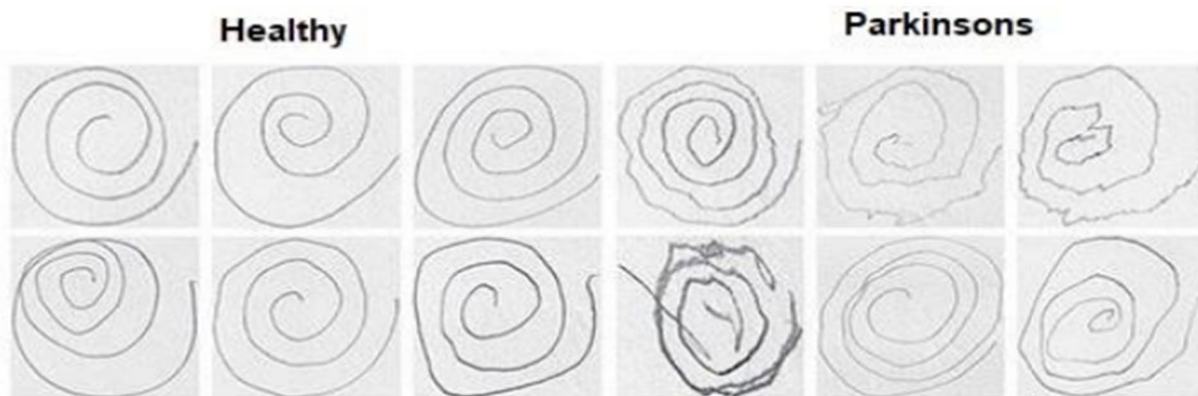


FIGURE 2
Samples of spiral drawing images dataset.

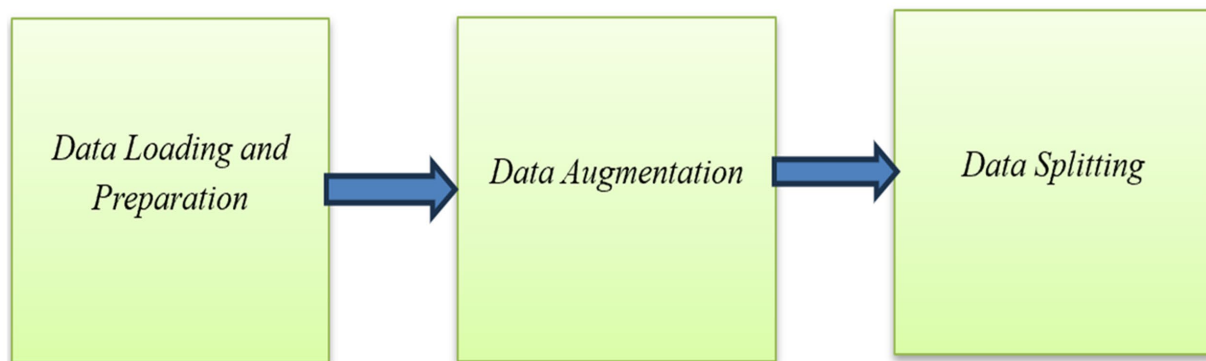


FIGURE 3
Preprocessing steps.

optimizer, where categorical cross-entropy is the loss function and accuracy is the assessment metric. The data training process was conducted over 50 epochs with a batch size of 16 samples in each iteration, utilizing augmented training data. Figure 4 shows the VGG19 model architecture. The parameters of the VGG19 model are presented in Table 2.

3.3.2 InceptionV3 model

We also employed the pre-trained InceptionV3 model (38), whose inception modules are well known for their effective multi-scale feature extraction capabilities for PD detection by analyzing spiral drawing image features. Images with three color channels and a resizing of 100×100 pixels are accepted by the input layer. With average pooling, the InceptionV3 model functions as the feature extractor, omitting its top classification layer. To add nonlinearity, a bespoke dense layer with 128 units and a ReLU activation function is applied. The last layer is a dense output layer for binary classification among individuals without PD and those with the condition. Figure 5 depicts the Inception model structure.

InceptionV3 has two units in the output layer to represent the dataset classes, namely, Parkinson and Healthy, as well as softmax activation applied for the classification task. The Adam optimizer is

used to create the model. Model training is carried out using a batch size of 32 utilizing augmented training data across 50 epochs. Table 3 summarizes the inception model parameters and their values used to develop and implement the model.

3.3.3 DenseNet169 model

We also applied the pre-trained DenseNet169 (39, 40) model for PD detection and classification based on spiral drawing image features. This model is known for having a dense pattern of connectivity that promotes improved feature reuse and maximum information flow across layers. Images with three color channels and a resizing of 224×224 pixels are accepted by the input layer. With average pooling, the pre-trained DenseNet169 model functions as the feature extractor, omitting its top classification layer. A bespoke dense layer with 128 units and a ReLU activation function is applied to add nonlinearity. Figure 6 illustrates the DenseNet169 model structure.

The final layer is a dense output layer used for binary classification between individuals without PD and those with the condition. Also known as the output or last layer, this layer has two units to represent the dataset classes and uses a Softmax activation function to calculate the probability of each sample being either PPD or Healthy. The model

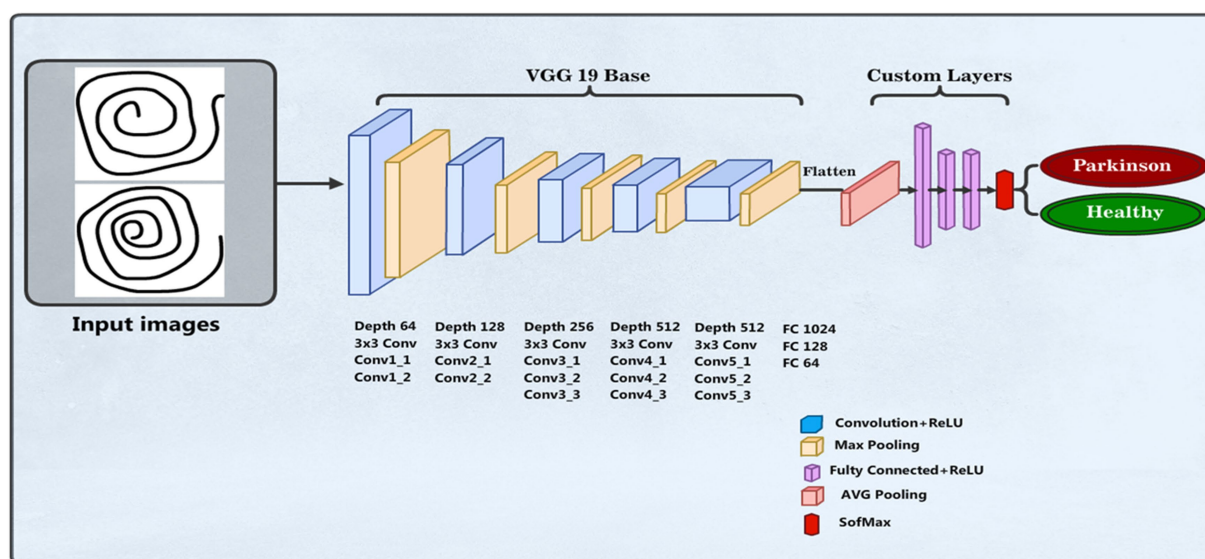


FIGURE 4
Structure of the VGG19 model.

TABLE 2 Summary of the VGG19 model parameters.

Layer	Parameters
Input Layer	(100, 100, 3)
VGG19 Base Model	Pre-trained on ImageNet, include_top=False, average pooling
Dense Layer	64 units, ReLU activation
Output Layer	2 units, Softmax activation
Optimizer	Adam
Loss Function	Categorical Cross-entropy
Metrics	Accuracy
No. of Epochs	50
Batch Size Used	16

utilizes accuracy as the evaluation measure, categorical cross-entropy as the loss function, and the Adam optimizer for training. Table 4 presents the summary of the model parameters used.

3.3.4 ResNet50v2 model

A DL framework called residual network (ResNet) was presented by Kaiming He et al. (41). The capability of this architecture to effectively train deep neural networks has attracted huge interest. The main breakthrough in ResNet is the use of residual connections, or skip connections, which improve gradient flow and lessen the problem of vanishing gradients. The residual blocks make up the bulk of the ResNet architecture. These blocks are made up of multiple convolutional layers, an activation function (usually ReLU), and batch normalization. The skip link, which enables the direct addition of the block's input to its output, is what distinguishes a residual block. This method enhances gradient flow during backpropagation and helps the network learn residual functions. We applied the ResNet50v2 model structure in our experimental work for PD detection and classification based the features of spiral drawing images. The images were scaled to

224×224 pixels with three color channels and can be loaded into the input layer. The feature extractor with average pooling is the pre-trained ResNet50v2 model without its top classification layer. Nonlinearity is added by adding a customized dense layer with 128 neurons and a ReLU activation function. Figure 7 depicts the model architecture.

The final layer is an output layer with two neurons and softmax activation function for binary classification of patients with PD and healthy people. Categorical cross-entropy is used as the loss function, accuracy is the assessment measure, and the model is assembled based on the Adam optimizer. Using supplemented training data, the training process was run across 50 epochs with a batch size of 32. Table 5 outlines the model parameters used.

To evaluate the models' performance on our current dataset, we first trained these pre-trained models on the spiral image dataset before testing them. We recorded performance metrics such as accuracy, precision, recall, and F1-score.

3.4 Evaluation metrics

Assessing the performance and testing results obtained by the proposed DL models, namely, VGG19, DenseNet169, Inception, and ResNet50v2, are crucial for gauging the effectiveness of the models. Several metrics are used to quantify performance, including precision, recall, accuracy, F1-score, and ROC curve, which are calculated from the confusion matrix. The evaluation measures provide an alternative perspective on the advantages and disadvantages of the model.

$$Accuracy = \frac{TP + TN}{FP + FN + TP + TN} \times 100 \quad (1)$$

$$F1 - score = 2 * \frac{precision \times Recall}{precision + Recall} \times 100\% \quad (2)$$

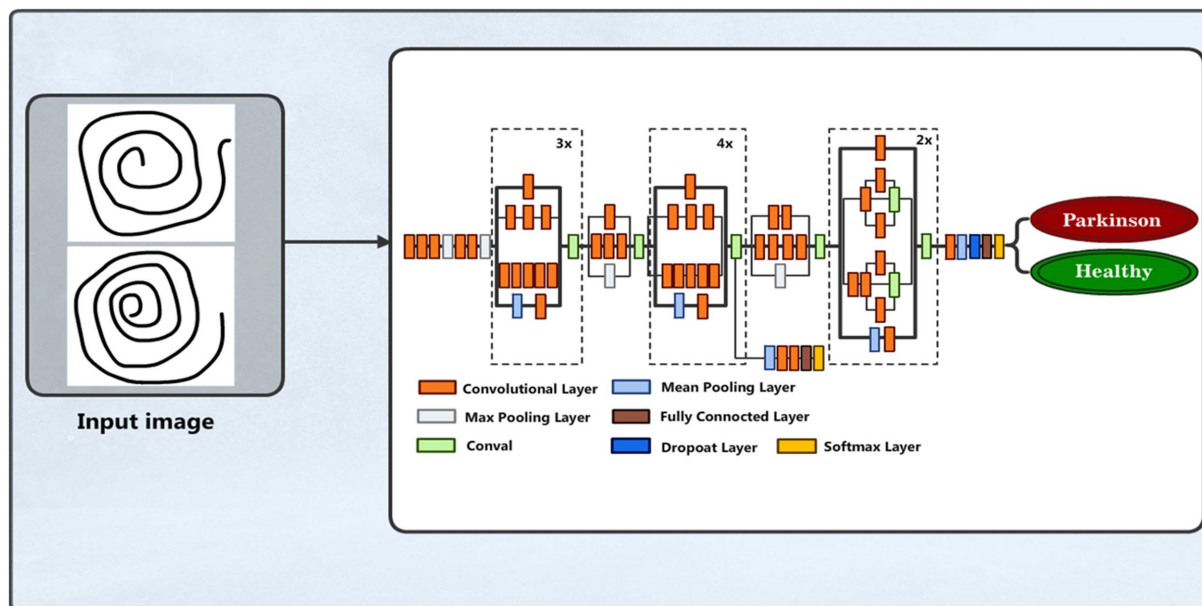


FIGURE 5
Inception model structure.

TABLE 3 Summary of the Inception model parameters.

Layer	Parameters
Input layer	(100, 100, 3)
Dense layer	128 units, ReLU activation
Output layer	2 units, Softmax activation
Optimizer	Adam
Loss function	Categorical Cross-entropy
Metrics	Accuracy
No. of epochs	50
Batch size used	32

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \times 100\% \quad (3)$$

$$Recall = Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negatives} \times 100\% \quad (4)$$

4 Experimental results

This section reports the findings obtained from various experiments carried out for PD recognition and classification using various DL models, namely, VGG19, ResNet50, InceptionV3, and DenseNet169. Each model was assessed based on its ability to accurately categorize spiral drawn images from patients with PD and healthy individuals.

4.1 Testing results of the VGG19 model

As revealed in Table 6 below, an overall accuracy of 72% is shown in the testing classification results for PD recognition utilizing the

VGG19 model. With a recall of 86% and precision of 60% for Parkinson's cases, the model successfully recognized the majority of Parkinson's cases with a small number of false positives.

Recall was 64% and precision was 88% for healthy persons, indicating a higher classification accuracy for healthy cases but with some false negatives. For Parkinson's patients, the F1-score was 71, while for healthy cases it was 74. The macro averages for precision, recall, and F1-score were 74, 75, and 72%, respectively. These findings point to areas where the model might be improved to lower classification mistakes while also demonstrating how well it detects PD. Figure 8 shows a graphical representation of the performance results for the VGG19 model.

Figure 8A illustrates the validation and training accuracies of the model over 50 epochs, presenting how well it learned to distinguish between Parkinson's and healthy cases. Figure 8B presents the model's loss over the training period, indicating the reduction in prediction error as training progressed. Figure 8C depicts the area under the curve (AUC) of the VGG19 model, providing a quantity of the model's capacity to distinguish between the two classes with an AUC value of 81%. The AUC is a valuable metric for evaluating the overall results of the classification model.

4.2 Testing results of the inception v3 model

The testing classification findings utilizing the InceptionV3 model for PD identification are given in Table 7. The InceptionV3 model attained an overall accuracy of 89%. For Parkinson's cases, the model achieved a precision of 78% and a recall of 100%, indicating it accurately recognized all true Parkinson's occurrences but included some false positives. For healthy individuals, the precision was 100% and the recall was 82%, showing exceptional precision but missing some real healthy examples. The F1-score for PD was 88%, and for healthy persons, it was 90%.

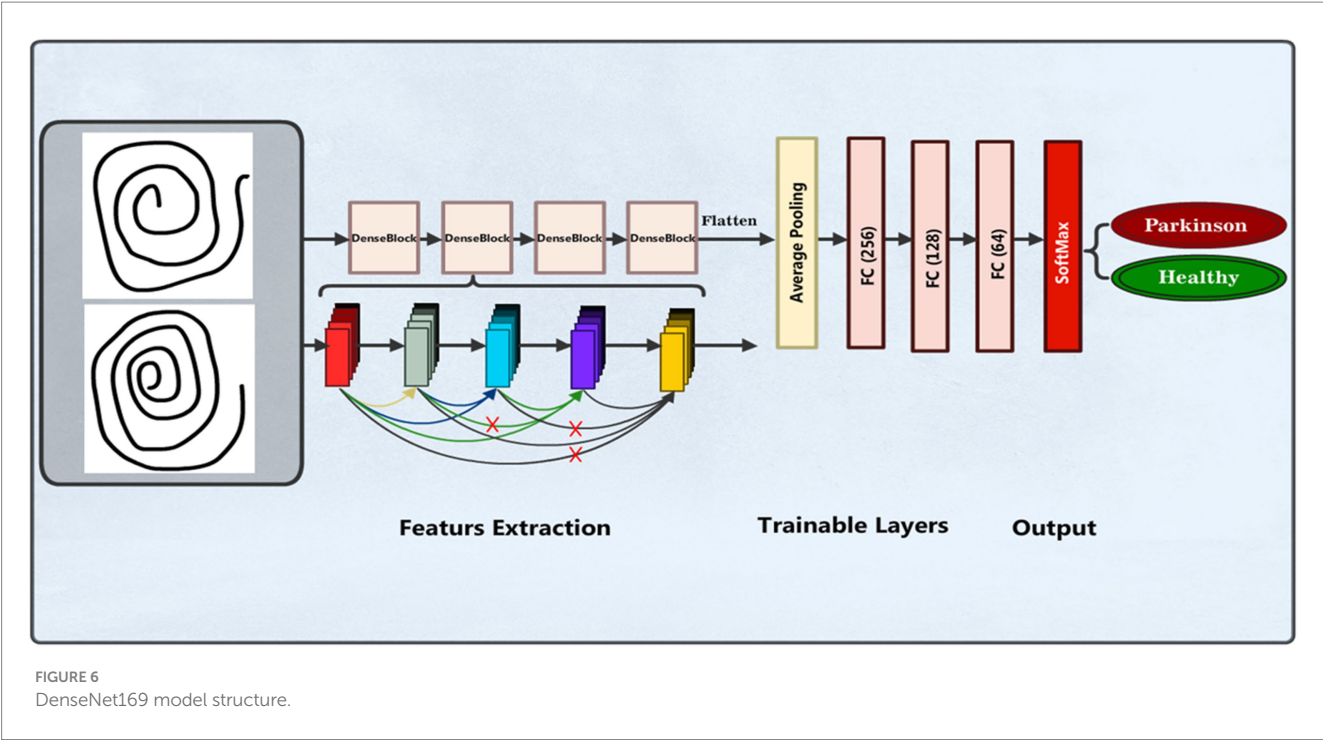


TABLE 4 Summary of the DenseNet169 model parameters.

Layer	Parameters
Input layer	(224, 224, 3)
DenseNet169 base model	Pre-trained on ImageNet, include_top = False, average pooling
Dense layer	128 units, ReLU activation
Output layer	2 units, Softmax activation
Optimizer	Adam
Loss function	Categorical cross-entropy
Metrics	Accuracy
No. of epochs	50
Batch size used	16

The overall averages of the metrics are 91% for precision, 89% for recall, and 89% for F1-score, demonstrating the balanced performance of the model across both classes. These results suggest that InceptionV3 is highly effective for PD detection, particularly excelling in correctly identifying true cases of the disease. Figure 9 shows a graphical representation of the performance results for the Inceptionv3 model.

Figure 9A shows the validation and training accuracies, which started at 55% and ended at 79% for training and the validation started at 45% and ended at 89%. The significant improvement from the initial to the final epoch indicates effective learning. Figure 9B illustrates the model's loss over the training period, with a notable reduction from an initial loss of 1.2840 to a final loss of 0.4486 for training and 0.3879 for validation, indicating the increased ability of the model to make accurate predictions. Figure 9C depicts the AUC of the InceptionV3 model, which reached an impressive value of 95,

demonstrating the robust discriminative ability of the model between Parkinson's and healthy cases.

4.3 Testing results of the ResNet50v2 model

This subsection presents the outcomes of our experiments utilizing the ResNet50v2 model for the detection and classification of Parkinson's Disease (PD). The model achieved an overall accuracy of 80%. For instances of Parkinson's, the ResNet50v2 model exhibited a precision of 79% and a recall of 92%. This indicates that the model correctly identified 92% of Parkinson's cases within the testing set, though it produced some false positives. In contrast, for healthy individuals, the model attained a precision of 83% and a recall of 62%, signifying a reasonable accuracy in classifying healthy cases but missing some true healthy instances. The F1-scores were 85% for Parkinson's cases and 71% for healthy cases. The testing classification performance of the ResNet50V2 model is summarized in Table 8.

The macro average precision, recall, and F1-score were 81, 77, and 78%, respectively. These metrics underscore the model's efficacy in distinguishing between PD and healthy individuals, although there remains room for improvement, particularly in increasing the recall for healthy cases. Figure 10 graphically represents the performance of the ResNet50V2 model over 50 epochs.

Figure 10A shows the validation and training accuracies, which improved significantly from 40% initially to 90% for training and 85% for validation by the final epoch, indicating effective learning. Figure 10B illustrates the model's loss over the training period, with a reduction from a preliminary loss of 1.20 to an ending loss of 0.20 for training and 0.40 for validation, reflecting good enhanced prediction accuracy of the model.

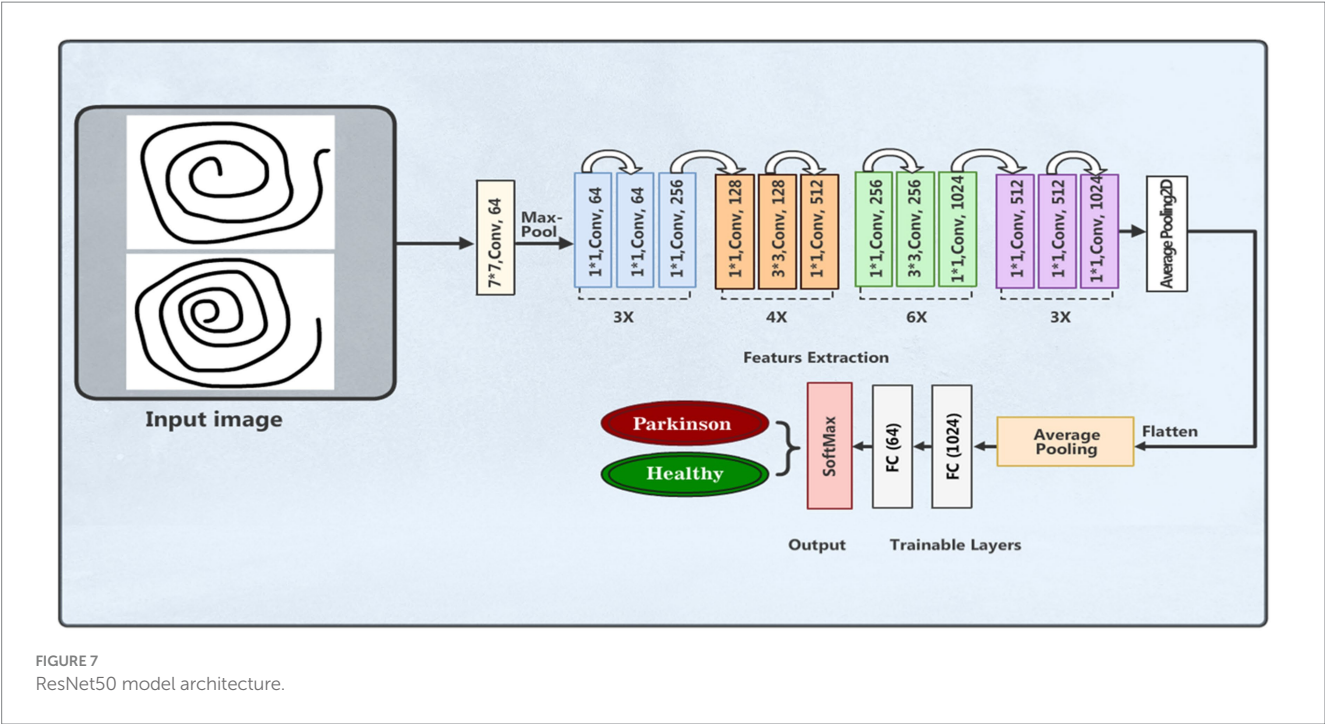


TABLE 5 Summary of the ResNet50 model parameters.

Layer	Parameters
Input layer	(224, 224, 3)
ResNet50v2 base model	Pre-trained on ImageNet, include_top = False, average pooling
Dense layer	128 units, ReLU activation
Output layer	2 units, Softmax activation
Optimizer	Adam
Loss function	Categorical Cross-entropy
Metrics	Accuracy
No. of epochs	50
Batch size used	32

4.4 Testing results of the DenseNet169 model

The testing classification results for the DenseNet169 model in detecting PD using spiral drawing images are summarized in Table 9. The DenseNet169 model achieved an overall accuracy of 85%, indicating a high level of performance in distinguishing between PD patients and healthy individuals based on their spiral drawing patterns. The model showed 80% precision and 100% recall for Parkinson's cases. This implies that there were no false negatives in the model's identification of all actual cases of PD. However, as the precision score shows, the model did generate some erroneous positives. For Parkinson's cases, the F1-score was 89%, indicating a fair trade-off between recall and precision for this class. The model's precision for healthy individuals was 100%, meaning that it was always accurate when it projected a case to be healthy. The

recall rate for healthy patients was 62%, indicating that some genuine healthy instances were overlooked by the algorithm, leading to misleading negative results. Compared to the Parkinson's class, the F1-score for healthy persons was 77%, indicating a reduced but still acceptable balance between precision and recall. The macro averages of 81% for recall, 83% for F1-score, and 90% for accuracy show how well the model performed generally in both classes. The recall macro average shows that there is still need for growth in accurately recognizing every instance across both classes, but the high precision macro average shows how well the model can make positive predictions. Figure 10 shows a graphical representation of the performance plots of the DensNet169 model. As seen in Figure 10, the training accuracy of the model started at 50% and steadily increased to 89% by the last epoch. Simultaneously, there was an upward trend in the validation accuracy, starting at 60% and reaching 83%. The training loss was reduced significantly from 90 to 20% in terms of model loss. In a similar vein, the validation loss significantly decreased, going from 100 to 55%. Collectively, these indicators show how the model's performance and capacity for generalization have increased during the training phase.

5 Discussion of the results

PD is a neurodegenerative condition that progresses over time and is characterized by both motor and non-motor symptoms. Accurate identification of PD is essential for timely intervention. Conventional diagnostic methods often rely on subjective neurological exams and clinical evaluations, leading to potential inaccuracies. Therefore, there is growing interest in leveraging advanced computational and machine learning methods to enhance diagnostic precision. Figure 11 shows performance of DenseNet169.

TABLE 6 Testing classification results of the VGG19 model.

	Precision %	Recall %	F1-score%	Support	Accuracy%
Parkinson	60	86	71	7	72
Healthy	88	64	74	11	
Macro average	74	75	72	18	

TABLE 7 Testing classification results of the InceptionV3 model.

	Precision %	Recall %	F1-score%	Support	Accuracy%
Parkinson	78	100	88	7	89
Healthy	100	82	90	11	
Macro average	91	89	89	18	

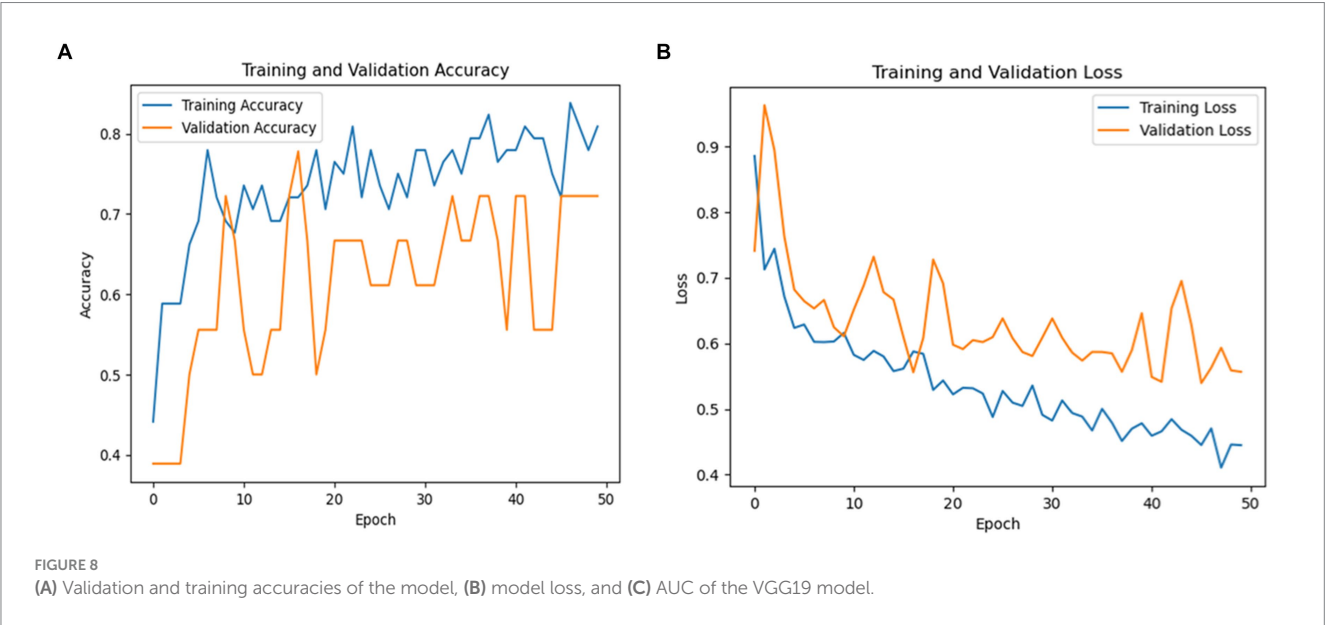


TABLE 8 Testing classification results of the ResNet50v2 model.

	Precision %	Recall %	F1-score %	Support	Accuracy %
Parkinson	79	92	85	12	80
Healthy	83	62	71	8	
Macro average	81	77	78	20	

In this study, we assessed the performance of several deep learning models VGG19, InceptionV3, ResNet50V2, and DenseNet169 in identifying PD from spiral drawing tests. The results highlight the strengths and limitations of each model. The VGG19 model achieved a total accuracy of 72%, demonstrating the lowest performance in detecting PD cases and a higher rate of false positives and false negatives compared to the other models.

The DenseNet169 model demonstrated an accuracy rate of 85%, whereas the InceptionV3 model achieved a higher accuracy of 89%, both surpassing the performance of the ResNet50V2 model. The InceptionV3 model, in particular, exhibited excellent

sensitivity and minimal false positives, making it highly effective in identifying both Parkinson's disease (PD) and healthy cases. In contrast, ResNet50V2 achieved an accuracy of 80%, with notable precision in identifying PD cases but less efficacy in classifying healthy individuals. Collectively, these findings indicate that transfer learning models based CNN architectures have capability to classify Parkinson's disease status using intelligent spiral drawings features, especially InceptionV3 and DenseNet169, that showed substantial potential for enhancing PD classification. Future research should focus on optimizing these models further, exploring additional data sources, and validating these findings in real-world clinical settings. Figure 12 displays the ROC of the

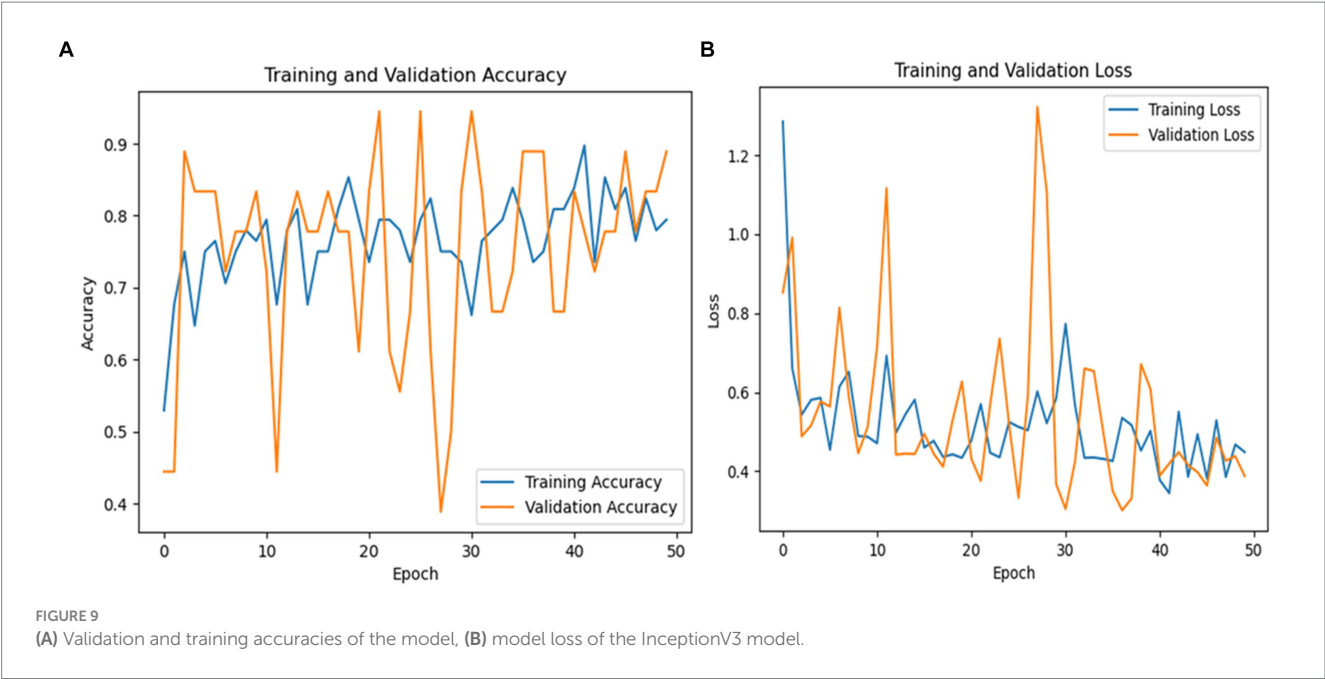
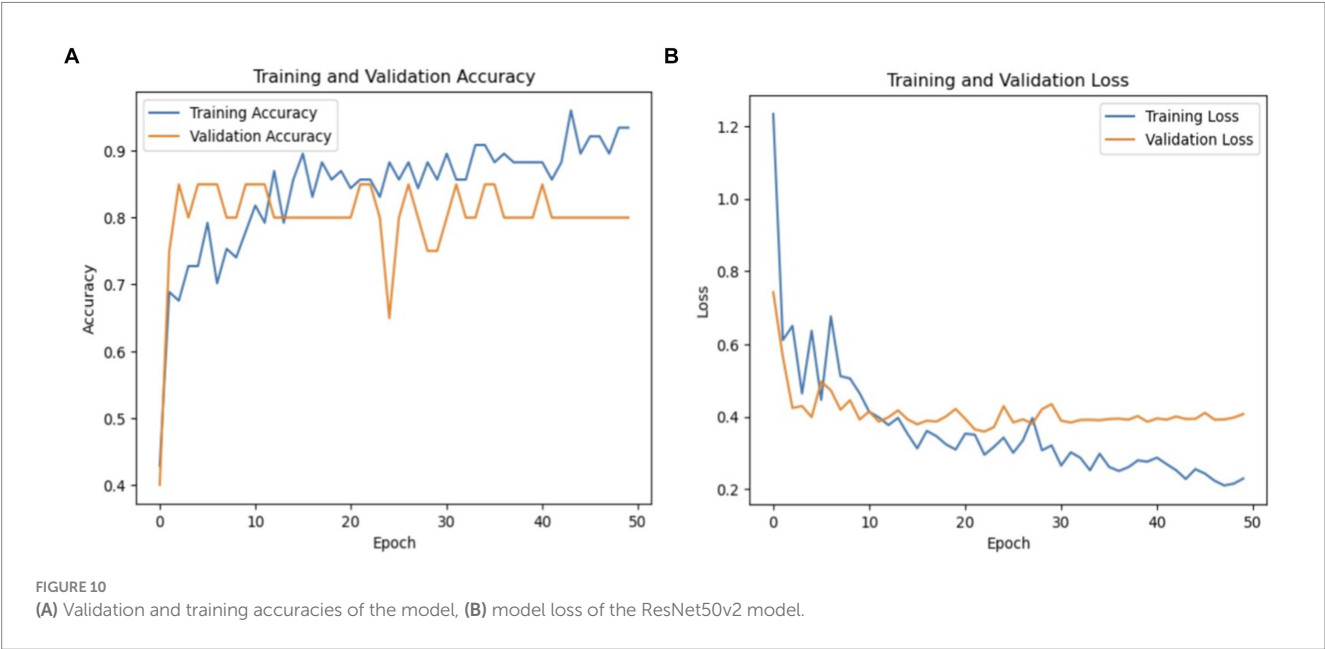


TABLE 9 Testing classification results of the DensNet169 model.

	Precision %	Recall %	F1-score %	Support	Accuracy %
Parkinson	80	100	89	12	85
Healthy	100	62	77	8	
Macro average	90	81	83	20	



proposed models, where the InceptionV3 model is found to achieve a high percentage of 91%.

This subsection highlights the variations in accuracy outcomes by providing an analysis of several techniques used on the same dataset of 102 spiral images. The authors reported a 67% accuracy rate using the RF technique in (38). According to Haq et al. (39), lightning CNNs achieved an accuracy of 63.33%, while in Huang et al. (41), a standard CNN approach demonstrated a significant increase with an accuracy of 83%. By comparison, the InceptionV3 model we used in our investigation produced the best accuracy of 89%. This better

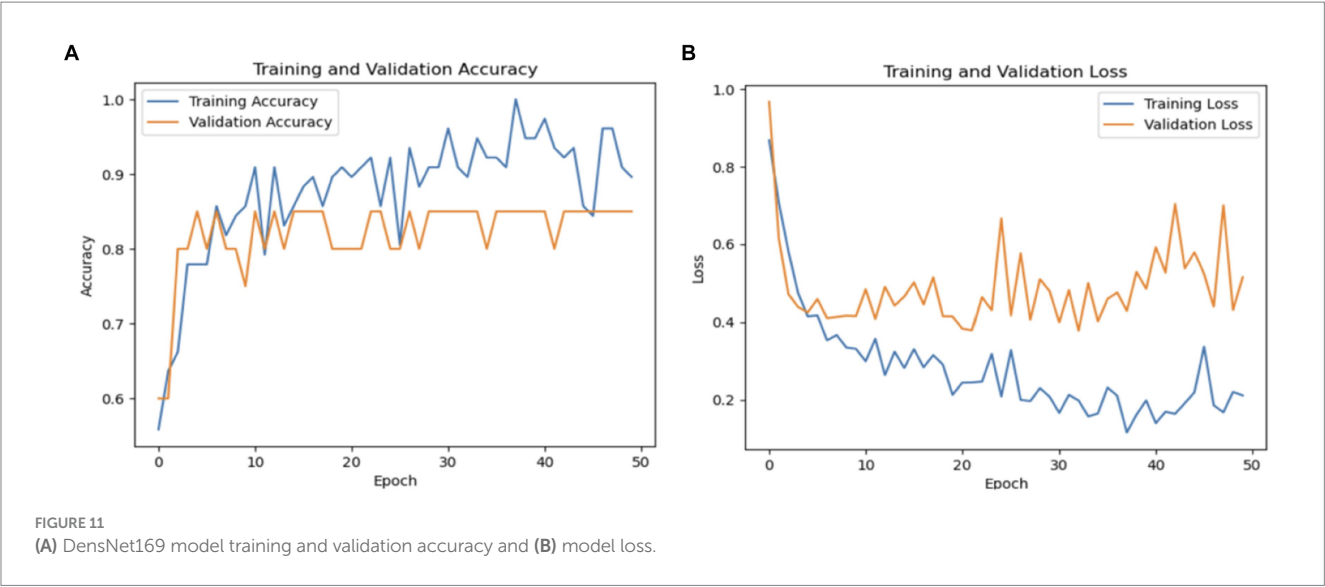


TABLE 10 Comparison of the contribution of the present study with existing research.

Reference ID	Approach	Dataset	Accuracy
Prasuhn et al. (44)	RF	Same dataset 102 spiral images	67%
Rasheed et al. (45)	Lightning CNNs	Same dataset 102 spiral images	63.33%
Pfister et al. (47)	CNNs	Same dataset 102 spiral images	83%
Our study	InceptionV3	Same dataset 102 spiral images	89%

performance highlights the potential of sophisticated DL architectures above more conventional machine learning and simpler neural network approaches, proving their effectiveness in correctly detecting PD using spiral drawing images. Table 10 displays the comparative analysis between our study results and existing ones based on the same dataset and accuracy metric.

6 Conclusion

The timely detection of PD is of utmost significance. The complexity of identifying PD necessitates the development of effective diagnostic instruments. In this work, PDD was determined by examining the Parkinson's spiral test. In contrast to other investigations in the literature, this study regarded the Parkinson's spiral test as an issue of recognition. Furthermore, pattern recognition approaches can yield favorable outcomes when used in the analysis of spiral images in PD. This strategy can enhance the effectiveness of diagnosing PD, a condition that is challenging to detect in its early stages. The proposed approach utilized a standardized dataset of 102 spiral samples obtained from individuals diagnosed with PD. The implementation involved the use of VGG19, InceptionV3, ResNet50v2, and DenseNet169 models for the detection of PD utilizing spiral drawings. The aim of this

work was to improve the diagnostic process of PD by utilizing transfer learning models. The approach shows promising results in diagnosing PD by analyzing the movement patterns of patients with PD. The classifier, trained on photos of the spiral drawing challenge, achieved an accuracy of 89% and an ROC score of 91% using the InceptionV3 and ResNet50v21 models. The use of DL-based analysis can enhance the efficiency and accessibility of spiral drawing assessment in clinical and research contexts due to its automated and scalable nature. Creating a deep learning system that utilizes spiral drawing images to detect PD can be a valuable method for aiding clinical decision making and advancing drug research. It can improve the diagnostic process, assist in selecting and monitoring patients in clinical trials, and offer objective measures of outcomes, ultimately leading to better patient care and the progress of PD research. The limitation of this research is that it did not investigate the possibility of use spiral drawings to identify other associated movement disorders; instead, it concentrated on utilizing them to create a system for diagnosing PD. The study showed that spiral image analysis is a useful tool for diagnosing PD, but it did not look into whether the technique can distinguish PD from other disorders that can similarly impair motor function, such as essential tremor. Another key limitation is that the data utilized was based on previously diagnosed PD participants, thereby making it more challenging to apply this AI approach as PD diagnostic criteria, given that the classification is already known. However, this research demonstrates that more sophisticated transfer learning architectures can improve on previous deep learning approaches for PD classification. As additional study data becomes available, especially spiral drawing data that can be collected in a general population of prodromal PD or those displaying motor symptoms, such architectures can be readily adapted.

Overall, although spiral image analysis for PD classification shows promise in the current research, more investigation is required to examine the approach's more extensive prospective applications and prove its efficacy for a larger range of movement disorders and patient demographics. Future research addressing these limitations may result in an even more potent and therapeutically valuable tool to aid in the differential classification and early detection of PD and associated

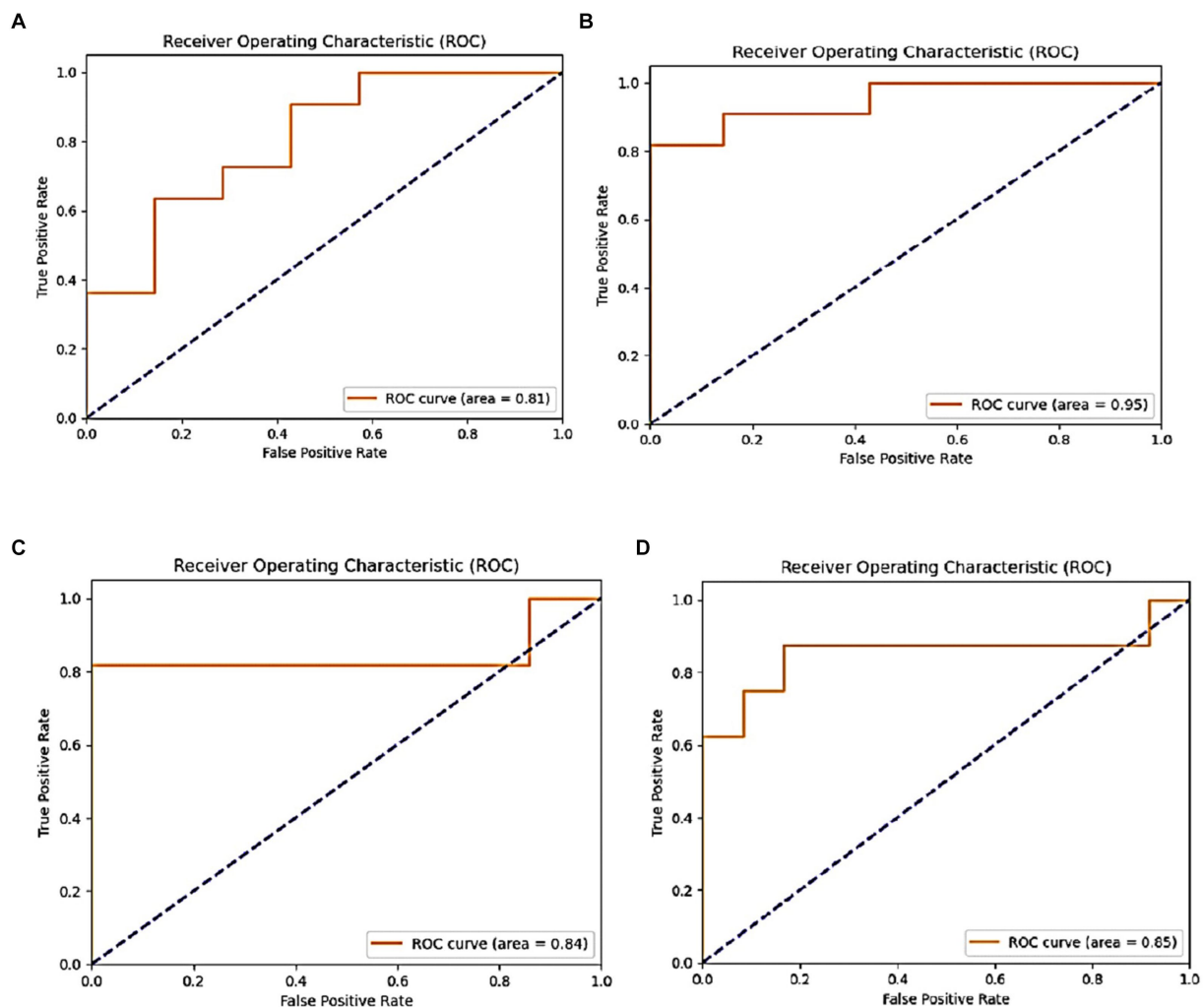


FIGURE 12
ROC metric of the proposed models: (A) VGG19, (B) Inception, (C) ResNet50v2, and (D) DensNet169.

disorders. In Future studies will try to solve this issue for improving the system.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: the dataset is available in this link, <https://www.kaggle.com/datasets/kmader/parkinsons-drawings>.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the patients/participants or patients/participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

NF: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author declares that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Goetz CG. The history of Parkinson's disease: early clinical descriptions and neurological therapies. *Pers Med.* (2011) 1:a008862. doi: 10.1101/cshperspect.a008862
- Vassar SD, Bordelon YM, Hays RD, Diaz N, Rausch R, Mao C, et al. Confirmatory factor analysis of the motor unified Parkinson's disease rating scale. *Parkinsons Dis.* (2012) 2012:719167. doi: 10.1155/2012/719167
- World Health Organization. (2023). Parkinson disease. Available online at: <https://www.who.int/news-room/fact-sheets/detail/parkinson-disease>.
- de Rijk MC, Launer LJ, Berger K, Breteler MM, Dartigues JF, Baldereschi M, et al. Prevalence of Parkinson's disease in Europe: a collaborative study of population-based cohorts: neurologic diseases in the elderly research group. *Neurology.* (2000) 54:S21–3. doi: 10.1212/WNL.54.11.21A
- Cantürk İ, Karabiber F. A machine learning system for the diagnosis of Parkinson's disease from speech signals and its application to multiple speech signal types. *Arab J Sci Eng.* (2016) 41:5049–59. doi: 10.1007/s13369-016-2206-3
- Singh N, Pillay V, Choonara YE. Advances in the treatment of Parkinson's disease. *Prog Neurobiol.* (2007) 81:29–44. doi: 10.1016/j.pneurobio.2006.11.009
- Rana A. S., Rawat A., Bijalwan H. Bahuguna. (2018). Application of multi-layer (perceptron) artificial neural network in the diagnosis system: a systematic review. 2018 international conference on research in intelligent and computing in engineering (RICE). Piscataway, NJ, 1–6.
- Gomez-Gomez LF, Morales A, Fierrez J, Orozco-Arroyave JR. Exploring facial expressions and affective domains for Parkinson detection. *arXiv.* (2020). doi: 10.48550/arXiv.2012.06563
- García AM, Arias-Vergara T, Vazquez-Correa JC, Nöth E, Schuster M, Welch AE, et al. Cognitive determinants of dysarthria in Parkinson's disease: an automated machine learning approach. *Mov Disord.* (2021) 36:2862–73. doi: 10.1002/mds.28751
- Aldhyani THH, Al-Nefae AH, Koundal D. Modeling and diagnosis Parkinson disease by using hand drawing: deep learning model. *AIMS Mathematics.* (2024) 9:6850–77. doi: 10.3934/math.2024334shu
- Al-Nefae AH, Aldhyani THH, Koundal D. Developing system-based voice features for detecting Parkinson's disease using machine learning algorithms. *JDR.* (2024) 3:1. doi: 10.57197/JDR-2024-0001
- Rizzo G, Copetti M, Arcuti S, Martino D, Fontana A, Logroscino G. Accuracy of clinical diagnosis of Parkinson disease: a systematic review and meta-analysis. *Neurology.* (2016) 86:566–76. doi: 10.1212/WNL.0000000000002350
- Ammenwerth E, Nykanen P, Rigby M, de Keizer N. Clinical decision support systems: need for evidence, need for evaluation. *Artif Intell Med.* (2013) 59:1–3. doi: 10.1016/j.artmed.2013.05.001
- Dreiseitl S, Binder M. Do physicians value decision support? A look at the effect of decision support systems on physician opinion. *Artif Intell Med.* (2005) 33:25–30. doi: 10.1016/j.artmed.2004.07.007
- Mattison HA, Stewart T, Zhang J. Applying bioinformatics to proteomics: is machine learning the answer to biomarker discovery for PD and MSA? *Mov Disord.* (2012) 27:1595–7. doi: 10.1002/mds.25189
- Lahmiri S, Shmuel A. Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine. *Biomed. Signal Process. Control.* (2018) 49:427–33. doi: 10.1016/j.bspc.2018.08.029
- Gómez-García JA, Moro-Velázquez L, Godino-Llorente JJ. On the design of automatic voice condition analysis systems. Part I: review of concepts and an insight to the state of the art. *Biomed. Signal Process. Control.* (2019) 51:181–99. doi: 10.1016/j.bspc.2018.12.024
- Gómez-García JA, Moro-Velázquez L, Godino-Llorente JJ. On the design of automatic voice condition analysis systems. Part II: review of speaker recognition techniques and study on the effects of different variability factors. *Biomed. Signal Process. Control.* (2019) 48:128–43. doi: 10.1016/j.bspc.2018.09.003
- Viteckova S, Kutilek P, Svoboda Z, Krupicka R, Kauler J, Szabo Z. Gait symmetry measures: a review of current and prospective methods. *Biomed Signal Process Control.* (2018) 42:89–100. doi: 10.1016/j.bspc.2018.01.013
- San-Segundo R, Navarro-Hellín H, Torres-Sánchez R, Hodgins J, De la Torre F. Increasing robustness in the detection of freezing of gait in Parkinson's disease. *Electronics.* (2019) 8:119. doi: 10.3390/electronics8020119
- Thomas M, Lenka A, Kumar Pal P. Handwriting analysis in Parkinson's disease: current status and future directions. *Mov Disord Clin Pract.* (2017) 4:806–18. doi: 10.1002/mdc3.12552
- Rosenblum S, Samuel M, Zlotnik S, Erikh I, Schlesinger I. Handwriting as an objective tool for Parkinson's disease diagnosis. *J Neurol.* (2013) 260:2357–61. doi: 10.1007/s00415-013-6996-x
- Impedovo D, Pirlo G. Dynamic handwriting analysis for the assessment of neurodegenerative diseases: a pattern recognition perspective. *IEEE Rev Biomed.* (2019) 12:209–20. doi: 10.1109/RBME.2018.2840679
- Impedovo D, Pirlo G. Chapter 7: online handwriting analysis for the assessment of Alzheimer's disease and Parkinson's disease: overview and experimental investigation In: Series on language processing, pattern recognition, and intelligent systems. Pattern recognition and artificial intelligence. Singapore: World Scientific Publishing (2019). 113–28.
- Dolz J, Desrosiers C, Ayed IB. 3D fully convolutional networks for subcortical segmentation in MRI: a large-scale study. *neuro. Image.* (2018) 170:456–70. doi: 10.1016/j.neuroimage.2017.04.039
- Drotár P, Mekyska J, Rektorová I, Masarová L, Smékal Z, Faundez-Zanuy M, et al. Analysis of in-air movement in handwriting: a novel marker for Parkinson's disease. *Comput Methods Prog Biomed.* (2014) 117:405–11. doi: 10.1016/j.cmpb.2014.08.007
- Basnin N., Sumi A.T., Hossain S.M., Andersson K. (2021). Early detection of Parkinson's disease from micrographic static hand drawings. In Proceedings of the 2021 international conference on brain informatics, virtual; 12960, pp. 433–447.
- Das HS, Das A, Neog A, Mallik S, Bora K, Zhao Z. Early detection of Parkinson's disease using fusion of discrete wavelet transformation and histograms of oriented gradients. *Mathematics.* (2022) 10:4218. doi: 10.3390/math10224218
- Kamran I, Naz S, Razzak I, Imran M. Handwriting dynamics assessment using deep neural networks for early identification of Parkinson's disease. *Futur Gener Comput Syst.* (2020) 117:234–44. doi: 10.1016/j.future.2018.01.013
- Shaban M. (2020). Deep convolutional neural network for Parkinson's disease based handwriting screening. Proceedings of the IEEE 17th international symposium on biomedical imaging workshops (ISBI workshops), Iowa City, IA. doi: 10.1109/ISBIWorkshops50223.2020.915340
- Kamble M, Shrivastava P, Jain M. Digitized spiral drawing classification for Parkinson's disease diagnosis. *Measurement: Sensors.* (2021) 16:100047. doi: 10.1016/j.measen.2021.100047
- Gil-Martín M, Montero JM, San-Segundo R. Parkinson's disease detection from drawing movements using convolutional neural networks. *Electronics.* (2019) 8:907. doi: 10.3390/electronics8080907
- Impedovo D, Pirlo G, Vessio G. Dynamic handwriting analysis for supporting earlier Parkinson's disease. *Information.* (2018) 9:247. doi: 10.3390/info910024710
- San Luciano M, Wang C, Ortega RA, Yu Q, Boschung S, Soto-Valencia J, et al. Digitized spiral drawing: a possible biomarker for early Parkinson's disease. *PLoS One.* (2016) 11:e0162799. doi: 10.1371/journal.pone.0162799
- Singh M, Khare V. Detection of Parkinson's disease using the spiral diagram and convolutional neural network. *Ingénierie des Systèmes d'Informat.* (2022) 27:991–7. doi: 10.18280/isi.270616
- Rosebrock Adrian. Detecting Parkinson's Disease with OpenCV, Computer Vision, and the Spiral/Wave Test. (2019). Available online at: <https://pyimagesearch.com/2019/04/29/detecting-parkinsons-disease-with-opencv-computer-vision-and-the-spiral-wave-test/>
- Szegedy C., Liu W., Jia Y., Sermanet P, Reed S., Anguelov D., et al. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA. 1–9.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv.* (2014) 1409
- Haq AU, Li JP, Ahmad S, Khan S, Alshara MA, Alotaibi RM. Diagnostic approach for accurate diagnosis of COVID-19 employing deep learning and transfer learning techniques through chest X-ray images clinical data in E-healthcare. *Sensors.* (2021). doi: 10.3390/s21248219
- Ruchi S, Singh D, Singla J, Rahmani MKI, Ahmad S, Rehman MU, et al. Lumbar spine disease detection: enhanced CNN model with improved classification accuracy. *IEEE Access.* (2023) 11:141889–901. doi: 10.1109/ACCESS.2023.3342064
- Huang G., Liu Z., Van Der Maaten L., Weinberger K.Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI; pp. 4700–4708.

42. Vanegas M.I., Ghilardi M.F., Kelly S.P., Blangero A. (2018). Machine learning for EEG-based biomarkers in Parkinson's disease. Proceedings of the 2018 IEEE international conference on bioinformatics and biomedicine (BIBM), Madrid, Spain; pp. 2661–2665.
43. Oh S.L., Hagiwara Y., Raghavendra U., Yuvaraj R., Arunkumar N., Murugappan M., et al. A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Comput Applic.* (2018) 32:10927–33. doi: 10.1007/s00521-018-3689-5
44. Prasuhn J., Heldmann M., Münte TF, Brüggemann N. A machine learning-based classification approach on Parkinson's disease diffusion tensor imaging datasets. *Neurol Res Pract.* (2020) 2:46. doi: 10.1186/s42466-020-00092-y
45. Rasheed J., Hameed A.A., Ajlouni N., Jamil A., Ozyavas A., Orman Z. (2020). Application of adaptive Back-propagation neural networks for Parkinson's disease prediction. In Proceedings of the 2020 international conference on data analytics for business and industry: Way towards a sustainable economy (ICDABI), Sakhr, Bahrain
46. Gunduz H. Deep learning-based Parkinson's disease classification using vocal feature sets. *IEEE Access.* (2019) 7:115540–51. doi: 10.1109/ACCESS.2019.2936564
47. Pfister FMJ, Um T.T, Pichler DC, Goschenhofer J, Abedinpour K, Lang M, et al. High-resolution motor state detection in Parkinson's disease using convolutional neural networks. *Sci Rep.* (2020) 10:5860. doi: 10.1038/s41598-020-61789-3
48. Talitckii A., Kovalenko E., Anikina A., Zimniakova O., Semenov M., Bril E., et al. Avoiding misdiagnosis of Parkinson's disease with the use of wearable sensors and artificial intelligence. *IEEE Sens.* (2021).
49. White Robin. (2020). Classifying Parkinson's disease through image analysis: Part 2. Available online at: <https://towardsdatascience.com/classifying-parkinsons-disease-through-image-analysis-part-2-ddbbf05aac21>
50. STPETE, ISHII. Parkinsons Drawing Pytorch Lightning CNN. (2023). Available online at: <https://www.kaggle.com/code/stpeteishii/parkinsons-drawing-pytorch-lightning-cnn>



OPEN ACCESS

EDITED BY

Hikmat A. M. Abdeljaber,
Applied Science Private University, Jordan

REVIEWED BY

Gaurav Gupta,
Shoolini University, India
Arwa Mashat,
King Abdulaziz University, Saudi Arabia
Adarsh Kumar,
University of Petroleum and Energy Studies,
India
Dayananda Pruthviraja,
Manipal Institute of Technology Bengaluru,
India

*CORRESPONDENCE

Purushottam Kumar Pandey
✉ purushottam9999@manavrachna.net
Saeed Alshahrani
✉ salhariri@ksu.edu.sa

RECEIVED 07 June 2024

ACCEPTED 20 August 2024

PUBLISHED 20 September 2024

CITATION

Pandey PK, Pruthi J, Alzahrani S, Verma A and
Zohra B (2024) Enhancing healthcare
recommendation: transfer learning in deep
convolutional neural networks for Alzheimer
disease detection.
Front. Med. 11:1445325.
doi: 10.3389/fmed.2024.1445325

COPYRIGHT

© 2024 Pandey, Pruthi, Alzahrani, Verma and
Zohra. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Enhancing healthcare recommendation: transfer learning in deep convolutional neural networks for Alzheimer disease detection

Purushottam Kumar Pandey^{1*}, Jyoti Pruthi¹, Saeed Alzahrani^{2*},
Anshul Verma³ and Benazeer Zohra^{4,5}

¹Department of Computer Science, Manav Rachna University, Faridabad, Haryana, India,

²Management Information System Department, College of Business Administration, King Saud University, Riyadh, Saudi Arabia, ³Department of Computer Science, Banaras Hindu University, Varanasi, India, ⁴Department of Anatomy, School of Medical Sciences & Research, Sharda University, Greater Noida, Uttar Pradesh, India, ⁵Department of Anatomy, Noida International Institute of Medical Sciences, Noida International University, Greater Noida, Uttar Pradesh, India

Neurodegenerative disorders such as Alzheimer's Disease (AD) and Mild Cognitive Impairment (MCI) significantly impact brain function and cognition. Advanced neuroimaging techniques, particularly Magnetic Resonance Imaging (MRI), play a crucial role in diagnosing these conditions by detecting structural abnormalities. This study leverages the ADNI and OASIS datasets, renowned for their extensive MRI data, to develop effective models for detecting AD and MCI. The research conducted three sets of tests, comparing multiple groups: multi-class classification (AD vs. Cognitively Normal (CN) vs. MCI), binary classification (AD vs. CN, and MCI vs. CN), to evaluate the performance of models trained on ADNI and OASIS datasets. Key preprocessing techniques such as Gaussian filtering, contrast enhancement, and resizing were applied to both datasets. Additionally, skull stripping using U-Net was utilized to extract features by removing the skull. Several prominent deep learning architectures including DenseNet-201, EfficientNet-B0, ResNet-50, ResNet-101, and ResNet-152 were investigated to identify subtle patterns associated with AD and MCI. Transfer learning techniques were employed to enhance model performance, leveraging pre-trained datasets for improved Alzheimer's MCI detection. ResNet-101 exhibited superior performance compared to other models, achieving 98.21% accuracy on the ADNI dataset and 97.45% accuracy on the OASIS dataset in multi-class classification tasks encompassing AD, CN, and MCI. It also performed well in binary classification tasks distinguishing AD from CN. ResNet-152 excelled particularly in binary classification between MCI and CN on the OASIS dataset. These findings underscore the utility of deep learning models in accurately identifying and distinguishing neurodegenerative diseases, showcasing their potential for enhancing clinical diagnosis and treatment monitoring.

KEYWORDS

deep learning, Densenet, EfficientNet-B0, Resnet, skull stripping, healthcare, clustering, decision making

1 Introduction

Neurodegenerative illnesses like AD affect brain cognitive function. It is one of the most common cause of Dementia. The exact cause of disease is still not fully discovered and so the cure. It is believed that it happens due to a combination of genetic, environmental, and lifestyle factors. The protein accumulation in the brain which is Amyloid Plaques is the main cause. The plaques accumulate between the neurons because of which the death of neuron starts. Inflammation in the brain and oxidative damage to neurons are also believed to play roles in the development and progression of Alzheimer's disease. These processes can further contribute to neuronal dysfunction and death. These disorders cause problems with brain function and impair cognition (1). Progressive decline in cognitive function, including memory loss and diminished cognitive ability, characterizes AD, the most prevalent form of dementia. Conversely, MCI is a transitional stage between typical cognitive aging and AD, distinguished by observable deterioration in cognitive functions that do not significantly impede routine tasks (2). These conditions impose a burden on healthcare organizations as well as society at large, in addition to endangering the health and safety of those affected.

The efficient detection of AD and MCI has become a crucial area of interest in medical research. The progress in neuroimaging methods, including magnetic resonance imaging (MRI), has improved the ability to diagnose these conditions. MRI scans are used to diagnose Alzheimer's and MCI by examining structural abnormalities, which often require advanced image processing to increase clarity and extract relevant features (3).

The ADNI (4) and OASIS (5) datasets are renowned for their efficacy in diagnosing Alzheimer's and MCI, both used in this analysis and recognized for their vast human macroscopic MRI data. These datasets cover healthy and AD/MCI patients. MRI images from both datasets are used to identify anatomical changes connected to neurodegenerative illnesses, such as brain volume and cortical thickness (4, 5).

Multiple methods are utilized to preprocess MRI data to increase AD and MCI diagnosis accuracy and comprehension. A Gaussian filter reduces noise and decreases artifacts and electrical noise to improve visual clarity (6), contrast-limited responsive Histogram Equalization (CLAHE) enhances contrast, the image is resized to 224×224 pixels for consistency (7), and CNN model compatibility and intensity levels are normalized across scans (8). Skull stripping eliminates non-brain tissues to focus further investigations on the importance of brain regions, and then Tissue segmentation segments the brain into gray matter, white matter, and cerebrospinal fluid, providing more precise data for study (9). In ADNI additional preprocessing, we performed skull stripping using U-Net (10) to remove the cranium. The brain is cut cross-sectionally along three axes—axial, coronal, and sagittal. The slices are evaluated for quality, and three are selected to show the most essential MRI imaging areas while reducing noise (9, 11, 12).

Many researches have been working in the area however the work done so far has limitation that this paper is trying to address. The work done by researcher's is focused on one dataset, where as we have used the multiclass dataset for the research. The improvised preprocessing model that can work on MRI from different datasets and the prediction model provides the consistency accuracy while predicting.

This investigation offers several substantial improvements to the existing Research on the identification of Alzheimer's and MCI:

- To extract and preprocess the renowned datasets, ADNI and OASIS, from the neuroimaging discipline for the investigation.
- To propose a framework model for early detection of AD using different deep learning techniques such as DenseNet-201, EfficientNet-B0, ResNet-50, ResNet-101, and ResNet-152 for the classification of MCI detection
- To evaluate and analyze the performance of prominent deep learning using performance metrics for making recommendations in healthcare organizations.

The complexity of MRI images for AD and MCI identification highlights cutting-edge deep learning processes. This work contributes to neuroimaging studies and AD/MCI diagnosis as the discipline progresses.

The paper not only addresses binary classification but also emphasizes multiclass classification. The predictive model extends beyond determining whether a subject has AD or not; it also predicts the stage of the disease, such as AD, CN, or MCI.

In binary classification, the model's output provides a straightforward yes or no answer regarding the presence of AD or another condition. However, in multiclass classification, the model distinguishes between different stages of the disease, offering a more nuanced understanding of the individual's cognitive health status. This approach is crucial for clinical applications as it allows healthcare providers to not only diagnose the presence of AD but also to categorize the severity or progression of the disease. Such detailed predictions can significantly aid in early intervention, personalized treatment planning, and monitoring of disease progression over time.

The paper is structured around the materials and methods outlined in Section 3, encompassing preprocessing techniques, transfer learning, and notable CNN architectures. Section 4 presents the dataset details and outcomes of the proposed approach. Lastly, Section 5 encapsulates the conclusion and outlines future avenues for the model's development.

2 Literature review

The analysis of the research done so far is represented in this section.

Modern deep-learning architectures are used to identify subtle patterns from the datasets to create powerful AD and MCI detection applications/models. These architectures ensure and advance neurodegenerative condition research. Most prominent advanced deep-learning architectures such as DenseNet-201 (13), EfficientNet-B0 (14), ResNet-50, ResNet-101, and ResNet-152 (15) have been investigated to develop efficient models for detecting Alzheimer's and MCI. The architectures often extract detailed patterns from complicated datasets and are used with transfer learning.

The DenseNet-201 design operates by establishing dense connections between each layer and all subsequent layers in a feed-forward manner to recycle features efficiently. The connection mentioned above improves the transmission of features and promotes the reuse of features, resulting in more effective use of parameters (13). EfficientNet-B0 prioritizes enhancing model efficiency by scaling the

network in many dimensions (depth, breadth, and resolution) to achieve an optimal trade-off (14). ResNet-50, ResNet-101, and ResNet-152 belong to the ResNet (Residual Network) family. This network family includes skip connections, also known as shortcuts, which enable bypassing one or more layers. This technique helps to address the vanishing gradient issue and facilitates the training of intense networks. These skip connections further enhance the propagation of gradients during backpropagation, enabling the model to learn more efficiently (15). Each of these designs offers distinct methods for extracting features and optimizing parameters, making them suitable for various elements of Alzheimer's MCI detection in MRI datasets.

A 3D-CNN model was trained using ADNI MRI data to distinguish AD from CN. An AD brain mask was found using a genetic algorithm-based Occlusion Map technique, and Backpropagation-based explainability methods. The recommended model had 87% accuracy in 5-fold cross-validation, mirroring prior findings, whereas an updated 3D-CNN model with 29 brain regions achieved a high validation accuracy using the `lrp_z_plus_fast` explainability technique (16). The assessment process exploits shallow CNN architecture on 2D T1-weighted MR brain images. This pipeline rapidly and accurately identifies normal, MCI, and AD. The technique is labeled MCI prodromal AD. They tested it against DenseNet121, ResNet50, and EfficientNetB7 (17). A unique ensemble deep-learning AD classification technique was developed. Soft-NMS consolidates candidate data and improves detection in the Faster R-CNN architecture. Enhanced ResNet50 extracts more complicated visual data. For sequence data processing, the feature extraction network employed Bi-GRU. Improved Faster R-CNN did the classification well (18). Researchers created EfficientNetB2 for AD, MCI, and NC. Front-end Global Attention Mechanism (GAM) in EfficientNetB2 took crucial features. Coordination Attention helped get channel and location data from two-dimensional slice data for appropriate diagnosis. Micro-designing using the ConvNeXt network reduced model complexity and improved categorization. The recommended method outperformed CNNs on AD/NC, AD/MCI, and MCI/NC dichotomous data (19). Investigators created an integrated automated method for guided machine learning-driven selection using K-Means++. A sophisticated deep learning framework using EfficientNetV2S transfer learning and learned features. Trials utilized ADNI and OASIS benchmark datasets. In research and validation, the integrated design outperformed all other models. Model validation was 20-fold. On the ADNI dataset, CN showed 83.64% accuracy against AD, 82.69% against MCI, 71.40% against MCI, and 91.54% on the OASIS dataset (20).

3 Materials and methods

The research approach used in this study centers on utilizing the ADNI and OASIS datasets, which are well-known for their extensive human macroscopic MRI data. These datasets include people who are in good health as well as those who have been diagnosed with AD and MCI (4, 5).

3.1 Methodology

MRI data is preprocessed using Gaussian filters (6), CLAHE for contrast enhancement, standardized image dimensions, and

normalizing intensity levels (7, 8). U-Net removes the cranium for the ADNI dataset, and the brain is sliced along three axes for cross-sectional slices. These slices undergo a quality evaluation to provide the best depiction while minimizing noise and highlighting significant regions of MRI imaging (9–13). State-of-the-art deep-learning architectures like DenseNet-201, EfficientNet-B0, ResNet-50, ResNet-101, and ResNet-152 extract intricate patterns from the datasets, hence aiding in creating effective models for AD and MCI identification. This mechanism guarantees the strength and dependability of the analysis performed on the ADNI and OASIS datasets, enabling progress in comprehending and identifying neurodegenerative disorders (13–15). Figure 1 presents the methodology for detecting AD and MCI using ADNI and OASIS datasets.

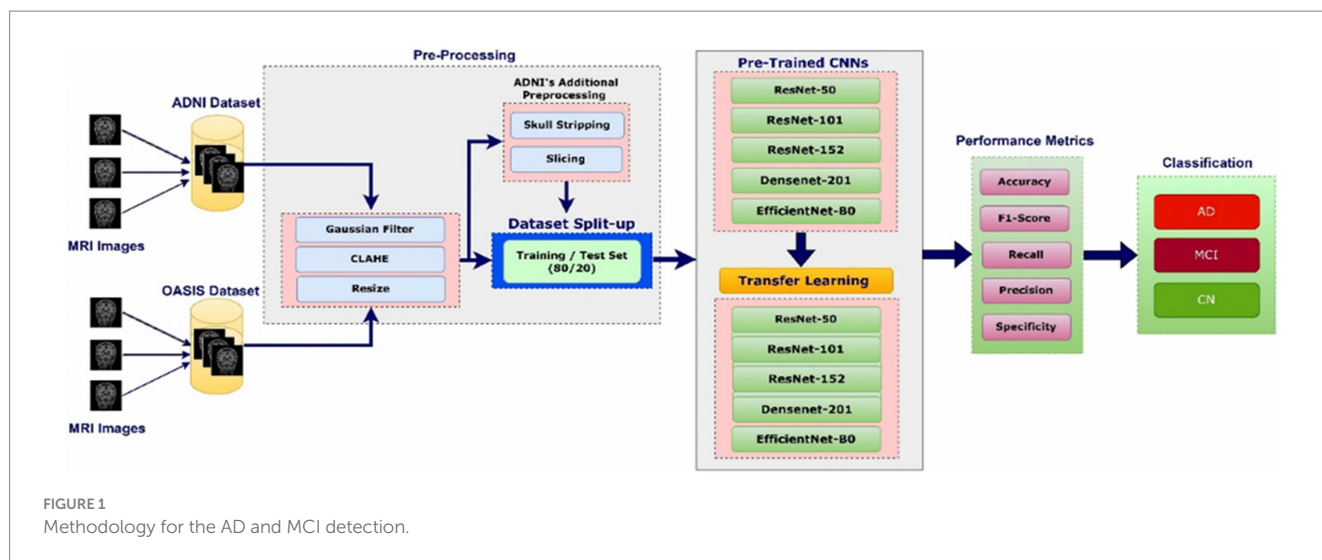
During the implementation the steps followed will be explained in the paragraph. The dataset will be provided to preprocessing model. The preprocessing model will make sure each image goes through Gaussian filter, Clahe and resizing. The ADNI images will go through additional two steps which are skull stripping and slicing. Once the data is preprocessed the images will be split in training and testing data in 80:20 ratio. For the model training the training dataset will be provide to the model. In the model image features will be extracted through different models and post that it will go through the transfer learning models. Once the model is trained the images from the test dataset will be provided and the prediction will be done by model. The efficacy of the model will be judged on F1 score, accuracy, recall value and precision. The model will categorize the images in the three buckets as CN, AD & MCI.

3.2 Data preprocessing

The preprocessing approaches explored for identifying AD and MCI include noise reduction, CLAHE, Image resizing, and normalization. Noise reduction in MRI scans is achieved by using a Gaussian filter. This filter effectively reduces noise caused by different sources, better depicting the images for analysis. Gaussian filtering reduces intensity fluctuations and maintains structural information, enhancing MRI data quality (6). CLAHE improves the contrast of specific areas by adjusting the intensity levels according to local histograms. This leads to a more detailed representation of the essential structural features of AD research (7). Resizing an image to a defined dimension, such as 224×224 pixels, guarantees consistency and compatibility with CNN models. This process maintains the structural data of the image for analytical purposes. Normalization is a process that makes intensity levels similar across MRI scans. This helps in accurate and comparative analysis by guaranteeing that intensity distributions are the same (8).

The designated preprocessing techniques enhance the accuracy and comprehensibility of MRI data in identifying AD and MCI in both the ADNI and OASIS datasets. In addition, some prominent steps of preprocessing involved for the ADNI dataset:

Skull Stripping: The U-Net architecture has an encoder-decoder structure incorporating skip links, similar to ResNets (10). Regarding skull stripping, the network takes a 3D MRI image as input and produces a binary mask that identifies the brain area. During learning, the network can divide the brain into distinct segments by predicting which pixels are part of the brain and which are not. The encoder



component of the U-Net collects features from the input image at various scales, while the decoder component increases the resolution of these features to produce a segmentation mask that matches the resolution of the input image. Spatial information is using skip links that facilitate conserved and accurate localization. The neural network is taught using a dataset consisting of MRI images and their matching manually generated skull-stripped masks (11). After training, the U-Net may automatically perform skull stripping on newly acquired MRI images, making it a significant asset in neuroimaging research and clinical practice (9, 12).

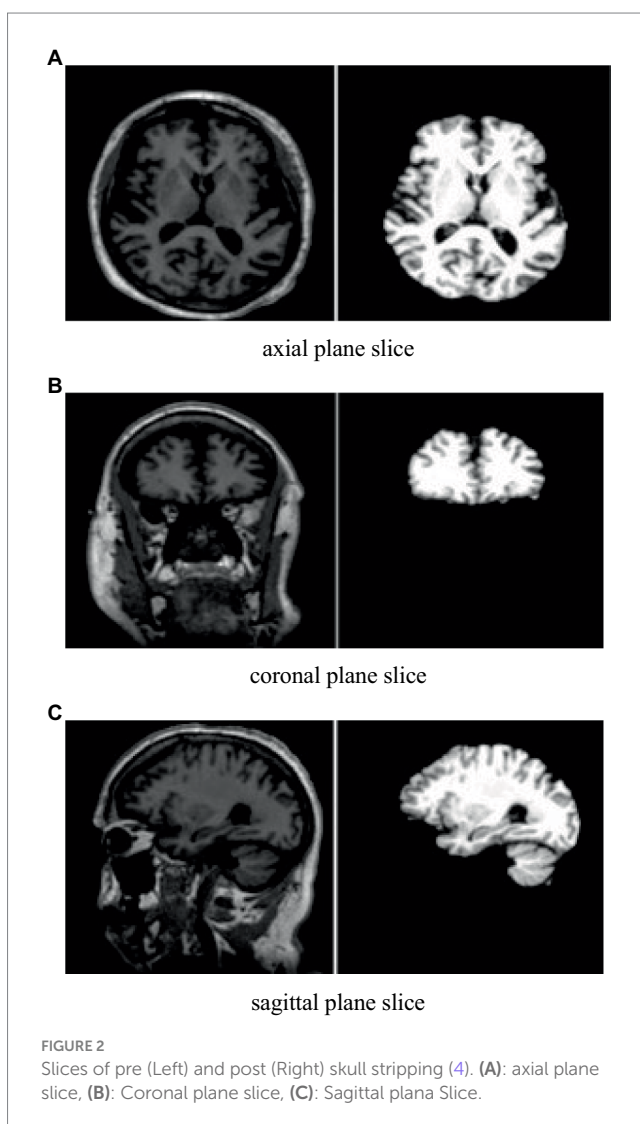
Slicing: After removing the skull, slice the brain along the three axes (axial, coronal, and sagittal) to get cross-sectional slices. This procedure entails segmenting the three-dimensional (3D) pictures into two-dimensional (2D) slices, which record distinct brain structure viewpoints. After getting the slices, visually assess their quality. Select three slices for a better qualitative representation than the rejected ones. This reduces noise and highlights the most essential areas of MRI imaging (11).

The images in Figure 2 show cross-sectional views before and after skull stripping, demonstrating the effects of the preprocessing method.

3.3 Prominent CNNs

ResNets: Residual Networks (ResNets) utilize shortcut connections between layers to facilitate residual learning. The residual learning approach entails acquiring knowledge about the residual mapping rather than the direct mapping of the input data, thereby enabling the efficient training of intense networks. Shortcut connections facilitate the propagation of gradients across layers and effectively address the disappearing gradients often seen in deep neural networks. ResNet comprises numerous residual blocks, including several convolutional layers and shortcut connections. This design enables the network to capture intricate input data aspects effectively (15).

ResNet can employ shortcut connections, bypassing one or more layers. The shortcut connections merely execute identity mapping; the results of these connections are aggregated with those of the layered layers. When many layers are appended, vanishing gradient issues frequently arise, preventing backpropagation from updating the



weights of the initial layers. The problem might be remedied through the incorporation of an identity link. The ResNet architecture facilitates the direct propagation of gradients in the opposite direction,

allowing them to traverse from the later layers to the initial filters via an identity link. By incorporating residual learning, the method improves the CNN architecture and renders it more applicable to the training of deep networks. A plain and simple network with a more significant number of layers tends to have more errors, but ResNet, which has specific layer configurations such as 50 and 101, has a superior capacity to handle deeper networks (15, 21, 22). Figure 3 presents the concept of shortcut connections.

ResNet familiarizes the conception of residual learning, where the layer transforms the input into a layer, and a shortcut connection bypasses one or more layers. Equation 1 presents the residual learning through Shortcut connection, understanding the basic building blocks of ResNets and how they are combined to form the architecture. This is expressed mathematically as:

$$\text{Output}(Y) = F(\text{Input}(X)) + \text{Input}(X) \quad (1)$$

Here, $F(\text{input}(X))$ represents the transformation performed by the layer.

A ResNet block typically consists of two convolutional layers followed by a shortcut connection. Let us denote input to the block as X , output as Y , and the residual function as $F(X)$ (15). In Equation 2 the output y is computed as:

$$Y = F(X, \{W_i\}) + X \quad (2)$$

where W_i are the weights of the convolutional layers.

ResNet has several layers, and these basic blocks are stacked together. The architecture consists of convolutional layers, batch normalization, ReLU activations, and residual blocks.

Let us consider a single convolutional layer within the residual function to simplify and derive this equation. Equation 3 calculates the output Y_1 of the convolutional layer is given by:

$$Y_1 = \sigma(W_1 * X + b_1) \quad (3)$$

Here, W_1 is the weights, b_1 is the bias, σ is the activation function (commonly ReLU), and $*$ denotes convolution.

Now, let us consider another convolutional layer with output Y_2 which can be calculated as Equation 4:

$$Y_2 = \sigma(W_2 * Y_1 + b_2) \quad (4)$$

The residual function $F(X)$ can be represented as the composition of these two layers:

$$F(X) = \sigma(W_2 * (\sigma(W_1 * X + b_1)) + b_2) \quad (5)$$

Substituting the expression for $F(X)$ of Equation 5 into the Equation 1, we get:

$$Y = \sigma(W_2 * (\sigma(W_1 * X + b_1)) + b_2) + X \quad (6)$$

Equation 6 represents the forward pass through a single residual block.

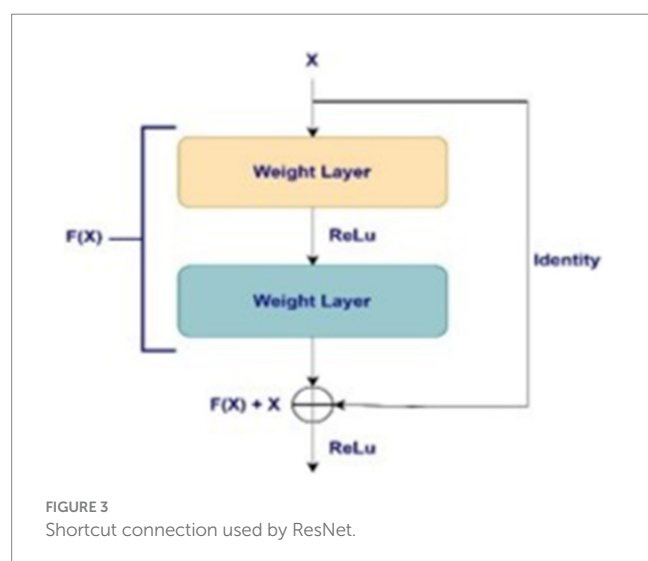
The beauty of ResNet architecture lies in the ability to learn the identity mapping (i.e., $Y = X$) if needed. If the optimal transformation for a block is close to the identity mapping, the weights of the convolutional layers can be adjusted to approach the identity function, allowing for easier optimization during training (15, 23).

Each ResNet network consists of numerous convolutional layers, pooling layers, and fully connected layers with varying output sizes and numbers of filters. The advantages include improved accuracy with increased depth and overcoming the degradation problem observed in shallower networks. The disadvantages may include higher computational complexity, as indicated by the increase in floating-point operations (FLOPs), which measures the number of floating-point operations a neural network performs during inference or training with deeper networks (15).

ResNet-50: ResNet-50 uses residual learning to solve the degradation issue of deeper neural networks by creating skip connections or shortcuts that enable information to move directly across layers. The model consists of 50 layers, which include convolutional, pooling, and fully linked layers, using residual blocks as the fundamental components. Each residual block has many convolutional layers and a shortcut link to help the network learn abstract features (15).

ResNet-50 can train deeper networks without the vanishing gradient issue, improving performance on complex datasets. ResNet-50's skip connections simplify training optimization, speeding convergence and improving generalization. ResNet-50 can be helpful for image classification and feature extraction because of its novel design and efficient training processes (18, 21).

ResNet-101: ResNet-101 is a CNN composed of precisely 101 layers. The construction of this architecture utilizes bottleneck blocks, which consist of three layers. It entails laying out various convolutional blocks with unique weights and additional elements, such as batch normalization and ReLU activations. The method employs residual learning to tackle degradation issues and enhance accuracy by leveraging higher depth. The network incorporates shortcut connections to facilitate residual learning, offering the choice between identity mapping or projection shortcuts. The model is trained using batch normalization,



stochastic gradient descent (SGD), weight decay, and dropout. This approach achieves high accuracy and successfully addresses optimization challenges encountered in regular networks. ResNet-101 consists of recurring blocks with different filter quantities and other properties. The design guarantees that the number of parameters, depth, breadth, and computing cost remain identical to those of plain networks (15, 24).

ResNet-152: Its 152 layers make ResNet-152 one of the deepest convolutional neural networks. Multiple convolutional layers and identity mappings in residual blocks enable feature extraction at various abstraction levels. Skips in ResNet-152 let information flow directly from previous layers to subsequent ones, maintaining gradient flow and simplifying training optimization. Deep and skip connections improve this architecture's image recognition performance, including accuracy, convergence during training, and the ability to handle vanishing gradient issues in deep neural networks (15, 24).

DenseNet-201: The Dense Convolutional Network (DenseNet) is characterized by a dense connection structure, which enables effective feature reuse and rapid model generation. The DenseNet-201 connects layers feed forwardly by utilizing feature maps from previous levels as inputs and producing feature maps for subsequent layers. The network has a total of $a(a+1)/2$ direct connections for nodes, i.e., a , which successfully alleviates the vanishing-gradient problem, improves feature propagation, encourages feature reuse, and decreases parameter count. The architecture comprises many compact blocks, including convolutional layers alternated with transition layers, which reduce dimensionality and regulate the complexity of the model. This architectural design facilitates extracting features and propagating gradients, effectively tackling the issue of disappearing gradients in deep neural networks (13).

The main benefits of this approach are eliminating unnecessary features, less computational burden, and increased understanding of the model's behavior due to dense connections. These advantages result in enhanced accuracy and efficiency while performing deep-learning tasks (11).

EfficientNet-B0: EfficientNet-B0 uses compound scaling to adjust the network's depth, breadth, and resolution equally. This leads to the creation of smaller and more precise models. The fundamental idea is to attain the best possible performance within computing limitations by carefully managing the model's depth, breadth, and resolution. The design incorporates inverted bottleneck blocks, squeeze-and-excitation blocks, and movable inverted bottleneck blocks, which optimize the use of parameters and processing resources. EfficientNet-B0 demonstrates exceptional performance, increased precision, reduced processing requirements, and adaptability, making it suitable for resource-limited settings such as mobile devices and edge computing (14).

3.4 Transfer learning

Transfer learning is a potent approach in ML that entails adjusting a pre-trained model from one task to another associated task, thereby capitalizing on the information acquired during the initial training. This strategy dramatically enhances the model's performance while decreasing the need for extensive datasets in the target domain (25). Transfer learning allows researchers to optimize specific tasks by refining pre-existing CNN models such as ResNet-50, ResNet-101,

ResNet-152, DenseNet-201, and EfficientNet-B0. These models have been extensively trained on datasets like ImageNet (11).

Fine-tuning, an essential component of transfer learning, is modifying the model's characteristics to match the patterns and correlations of the target problem. For example, ResNet and EfficientNet topologies sometimes improve by including extra dense layers, usually 256 and 128 units. This method allows models to specialize in activities beyond their initial training goals. Transfer learning allows models to use the information from pre-training on extensive datasets, enabling quicker convergence and enhanced generalization when fine-tuned on particular datasets. This strategy simplifies the process of developing models and improves performance in different applications (25, 26). The demonstration of the employment of transfer learning on ResNet-101 is presented in Figure 4, which removes the top layer and adds a new layer.

4 Results and discussion

The experiment was conducted utilizing a system including quadruple NVIDIA RTX A6000 GPUs, each equipped with 32 GB of memory, resulting in a combined processing capability of 194.8 TFLOPS. The system also included 64 GB of RAM and an AMD EPYC 7232P Octa Core CPU.

4.1 Dataset

The ADNI and OASIS datasets comprise human macroscopic MRI data, encompassing both individuals in good health and those who have received a diagnosis of Alzheimer's and MCI disease. The ADNI and OASIS datasets employed in the investigation, renowned for their unrestricted access, provide researchers with invaluable resources for examining the human brain's structural characteristics via MRI imaging. These datasets enable inquiries into both typical brain anatomy and pathological alterations that are linked to Alzheimer's and MCI disease. The robust prediction model can be integrated with the MRI system so that it act as a helpful resource to the doctors.

ADNI: The ADNI dataset, a vast resource for AD progression research, uses MRI images to reveal deep brain anatomy. The ADNI longitudinal study uses MRI, PET, and other biological markers to identify biomarkers for early detection and tracking of AD. It allows in-depth analysis of brain area using bottom-to-top brain scanning axial visuals, with T1-weighted images improving anatomical structure analysis and problem detection. The ADNI incorporates several methods for participant and phantom scans. Participants undergo scanning utilizing a variety of sequences, including axial T2 STAR, axial 3D PASL, accelerated sagittal MPRAGE, sagittal 3D FLAIR, axial DTI, field mapping, axial rsfMRI with eyes open, and HighResHippocampus. The specific sequences may differ depending on the scanner's manufacturer, for example, GE Systems for axial DTI scans and Philips Systems for resting-state fMRI and axial T2-FLAIR scans. This overview offers essential information on the imaging procedures and sequences used in the ADNI dataset. The dataset includes MRI scans from over 1,200 participants, each having multiple scans over time (4).

The age cohort-specific Alzheimer's disease progression analysis is possible from 20 to 90. Further processing involved extracting 2D slices from the original T1-weighted MRI scans and a processed

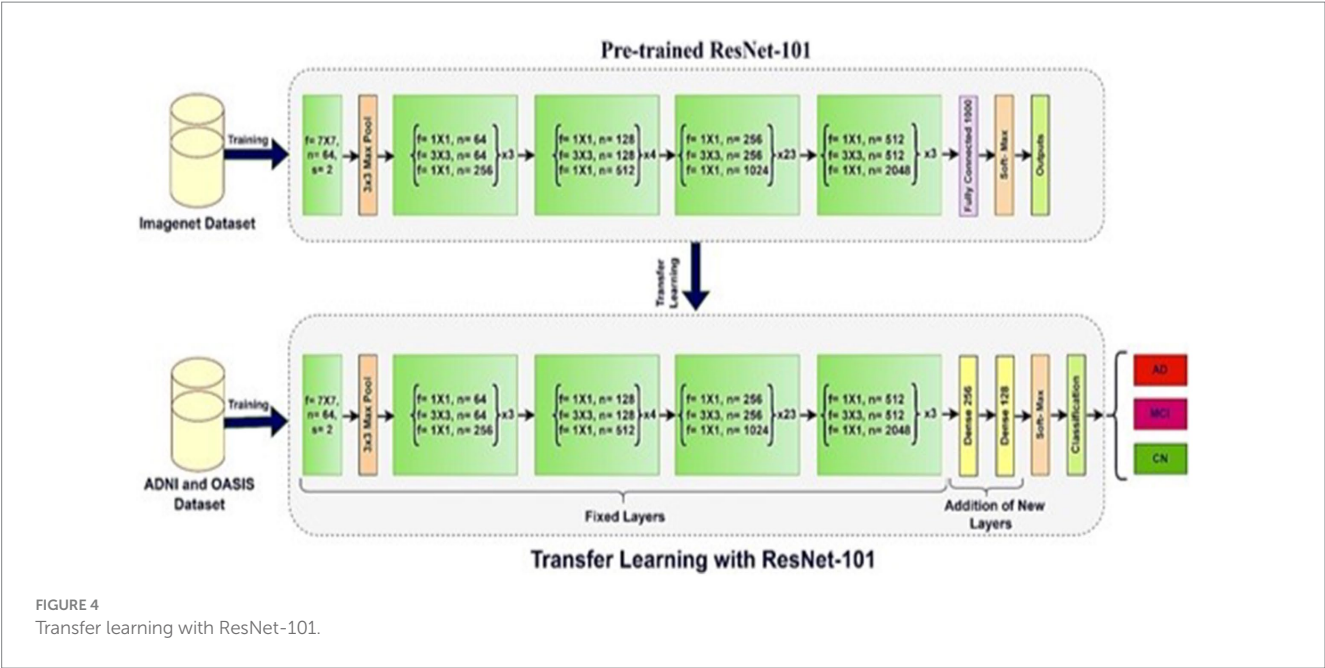


FIGURE 4
Transfer learning with ResNet-101.

Class	Train	Test	Total
AD	4,980	1,244	6,224
MCI	4,162	1,040	5,202
CN	6,605	1,651	8,256

collection after skull stripping. The distribution of these slices across different anatomical planes was recorded as follows and given in Tables 1, 2.

This breakdown provides detailed insight into the composition of the dataset, which is crucial for understanding the distribution of data for training and testing purposes across different classes and anatomical planes (4). When working with MRI images many computational complexity's need to be considered like image size and resolutions. Data augmentation need to be applied so that model has good number of images for the training. The preprocessing and feature extraction should be robust so that noise can be handled. Optimizing these factors is crucial for achieving efficient and effective analysis of MRI data in medical applications.

This investigation focused on three classes: CN, AD, and MCI, which had corresponding MRI scan counts of 159, 123, and 100, respectively. The dataset was partitioned into training and testing sets to facilitate deep learning tasks, wherein training comprised 80% of the data and testing included 20%. The CN group allocated 127 scans for training and 32 scans for testing, whereas the AD group utilized 99 scans for training and 24 scans for testing. For assessment purposes, there are 20 scans for testing of the MCI and 80 scans for training.

OASIS: The OASIS dataset is accessible to the public for investigation. It comprises cross-sectional MRI data from 416 people aged 18 to 96years. Among these individuals, 100 have been diagnosed with AD at a very low to moderate stage. The dataset comprises T1-weighted MRI images for each participant, enabling a

broad spectrum of analytical methodologies. The dataset has undergone de-identification, meticulous quality screening, and post-processing to provide standardized anatomical measurements. The inclusion of measures such as estimated total intracranial volume (eTIV) and normalized whole-brain volume (nWBV) offers valuable insights into the structural changes in the brain associated with aging and AD (5, 27). Table 3 shows the train-test (80–20%) split and quantity of MRI images for AD, MCI, and CN.

4.2 Performance metrics

Performance metrics quantify deep learning model performance. Many performance indicators include accuracy, precision, recall, and F1 score. Accuracy is the ratio of real positives and negatives to data points. Predicting the majority class may give the model high accuracy with imbalanced datasets, which may be misleading. The F1 score is a metric that combines recall and precision (26). On these metrics, precision, and recall calculations are predicated. Recall is the percentage of positive instances from the overall count of positive cases. At the same time, precision denotes the ratio of accurate optimistic predictions to the overall count of positive predictions. Incorporating false positives and false negatives, the F1 score is an exceptionally effective metric for assessing the performance of datasets containing unbalanced classes (28).

The specified CNN model employs the following Hyperparameters: Although RMSprop is renowned for its capability to modify learning rates and manage sparse gradients, a learning rate of 0.02 may be excessively high and could be improved. The detection of AD and MCI are examples of multi-class classification tasks amenable to categorical cross-entropy. The batch size 64 frequently balances model stability and computational efficiency. Although 50 training epochs are a reasonable starting point, the validation loss must be closely monitored to prevent overfitting, and early halting should be considered.

4.3 Analysis

This investigation included three types of classification: first, multi-class classification in the AD vs. CN vs. MCI classes. The second is the Binary classification of AD and CN, and the third is MCI and CN. Comparing the predicted and observed labels yielded the accuracy of classification.

4.3.1 Multi-class classification (AD vs. CN vs. MCI)

Table 4 compares employed CNNs onto the specified two datasets in the investigation of multi-class classification, i.e., AD, MCI, and CN. The outcomes presented in Table 4 demonstrate the model’s remarkable capacity to differentiate between cases of AD and MCI. The experimental results for the three-class classification experiment revealed that models become increasingly proficient in handling multi-class problems, as evidenced by their superior performance.

Quantification was performed on three cerebral components—white matter, gray matter, and cerebrospinal fluid—as part of the assessment of malady severity. The findings presented in Table 5 illustrate that the group comparing AD to CN to MCI attained exceptional levels of ResNet-101 Accuracy, Precision, Recall, and F1 Score. The findings show significant variations in performance across the CNN models in the multi-class classification test for AD, CN, and MCI. ResNet-101 scored the maximum accuracy and F1 score across both datasets, 98.21 and 94.78% for ADNI and 97.45 and 93.45% for OASIS, respectively, proving its ability to discriminate between the classes. ResNet-152 followed closely, achieving 97.89% accuracy for ADNI and 96.91% for OASIS. EfficientNet-B0, despite performing satisfactorily and scoring 89.67% for ADNI and 88.56% for OASIS. The F1 scores, which consider both accuracy and recall, reflected the patterns found in the individual measures, with ResNet-101 getting

TABLE 2 ADNI sagittal planes slices (4).

Class	Train	Test	Total
AD	4,644	1,160	5,804
MCI	3,522	1,200	4,402
CN	3,912	978	4,890

TABLE 3 OASIS class wise instances (27).

Class	Train	Test	Total
AD	390	98	488
MCI	4,800	1,200	6,000
CN	4,800	1,200	6,000

TABLE 4 AD vs. CN vs. MCI (multi-class classification).

CNN	Accuracy		Precision		Recall		F1 Score	
	A	O	A	O	A	O	A	O
ResNet-50	83.45	80.12	82.56	79.34	82.34	80.45	82.65	80.23
ResNet-101	98.21	97.45	94.67	93.12	94.89	93.67	94.78	93.45
ResNet-152	97.89	96.91	92.01	90.89	91.78	91.45	91.89	91.23
DenseNet-201	78.23	76.56	78.45	76.34	77.89	76.78	78.01	76.45
EfficientNet-B0	89.67	88.56	89.23	88.45	89.45	88.67	89.34	88.56

the most significant F1 scores for both datasets, followed by ResNet-152 and EfficientNet-B0. These findings indicate that ResNet-101 is the best model for this multi-class classification job, followed by ResNet-152, with EfficientNet-B0 trailing behind in performance. ResNet-101 has the highest accuracy level, meaning it can correctly put cases into each class. ResNet-152 and EfficientNet-B0 have lower accuracy measurements and fewer correct results. ResNet-101 did better than the others in memory to catch more true positives. ResNet-152 and EfficientNet-B0 had lower scores, which means they missed more false positives. Figure 5 illustrates the confusion matrix of the ResNet-101.

4.3.2 Binary classification (AD vs. CN)

The outcomes of the binary classification test, distinguishing between AD and CN individuals, demonstrate diverse degrees of performance and are presented in Table 6. The AD versus CN group exhibited the subsequent categorization for the assessment primarily because of notable disparities in brain tissue region. The AD versus MCI group indices showed a reasonably high value but somewhat lower than the AD versus CN group. This observation aligns with predictions since MCI is pathologically more similar to AD than CN. Consequently, distinguishing between MCI and AD may be slightly more challenging. The findings indicated that the ResNet-101 model attained a notable level of accuracy in accurately categorizing the AD. The approach had a 92.34% accuracy in differentiating normal controls from AD patients. Again, ResNet-152 closely follows as the second-highest achiever, exhibiting robust performance across all criteria. DenseNet-201 has commendable performance but could be a lot better than ResNet-152. Conversely, EfficientNet-B0 has the least favorable performance compared to the other models. Figure 6 illustrates the confusion matrix of the ResNet-101 with the ADNI dataset.

4.3.3 Binary classification (MCI vs. CN)

The outcomes of the binary classification of MCI and CN individuals demonstrate diverse degrees of performance and are presented in Table 7. In the endeavor of classifying MCI from CN, the efficacy of CNN models varied across metrics and datasets. ResNet-152 demonstrated the most exceptional overall performance among the assessed models, attaining an accuracy of 90.11% on the OASIS dataset and 89.56% on the ADNI dataset. Furthermore, the model exhibited high precision, recall, and F1 Score values across both datasets, signifying its resilient capability to differentiate between cases of MCI and CN. ResNet-101 demonstrated commendable performance. According to these findings, deeper CNN architectures, namely ResNet-152 and ResNet-101, exhibit notable efficacy in

distinguishing between MCI and CN. Figure 7 illustrates the confusion matrix of the ResNet-152 with the OASIS dataset.

The investigation found that ResNet-101 is best performed in the multi-classification and binary classification for the ADNI dataset; it is also well achieved with the OASIS dataset. The ResNet-101 model's accuracy and loss were used to track and assess the training and validation process and presented through Figure 8. ResNet-101's multi-class classification performance improves with time, as shown in Figures 8A,B, by showing training and validation accuracy and loss throughout epochs.

The training and validation accuracy consistently rises, while the training and validation loss consistently decreases, indicating that the model successfully integrates information from the training data.

4.4 Ablation study

The present investigation leverages the ADNI and OASIS datasets, which include comprehensive human macroscopic MRI data on healthy people and Alzheimer's MCI patients. The ADNI dataset uses U-Net to remove the skull and brain, slicing along three axes for cross-sectional slices. These slices are quality-checked to minimize noise and highlight important MRI imaging areas. Prominent deep learning architectures like DenseNet-201, EfficientNet-B0, ResNet-50, ResNet-101, and ResNet-152 extract complex patterns to identify AD and MCI. The approach allows reliable ADNI and OASIS dataset processing, improving neurodegenerative condition comprehension and detection. The research examined how preprocessing techniques, deep learning architectures, and transfer learning methodologies affect the performance of models and compared their effectiveness. While Gaussian filters are frequently employed to reduce image noise, they might not be the most optimal approach to accentuate critical features in MRI data, especially when identifying Alzheimer's and MCI. However, transfer learning is a highly effective method in deep learning, the results obtained from fine-tuning with specified

TABLE 5 Hyperparameters.

Sr No	Parameter	Value
1	Optimizer	RMSprop
2	Learning Rate	0.02
3	Loss Function	Categorical Cross-Entropy
4	Batch Size	64
5	Number of Epoch	20

TABLE 6 AD vs. CN (binary classification).

CNN	Accuracy		Precision		Recall		F1 Score	
	A	O	A	O	A	O	A	O
ResNet-50	87.45	85.32	86.78	88.23	86.23	89.45	86.78	87.89
ResNet-101	92.34	90.12	90.02	92.89	90.01	91.34	90.17	91.89
ResNet-152	89.67	88.23	88.78	90.45	88.34	91.23	88.78	89.67
DenseNet-201	85.67	84.12	84.78	86.23	84.45	87.12	84.78	85.67
EfficientNet-B0	78.23	77.45	77.78	79.56	77.89	80.12	77.78	78.23

pre-trained models and their effectiveness. While these models have been extensively trained on datasets like ImageNet, the performance of the transfer learning approach may have needed to be improved. The research examined the effects of incorporating dense layers of 256 and 128 units into each specified deep learning architecture after transfer learning from ImageNet-trained models to the ADNI and OASIS datasets. By comparing the efficacy of each architecture with and without additional layers, the research seeks to identify the

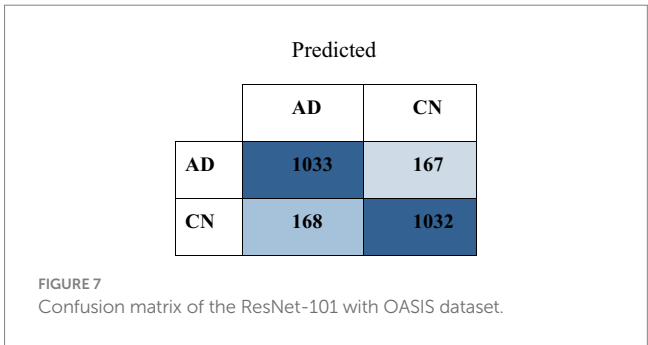
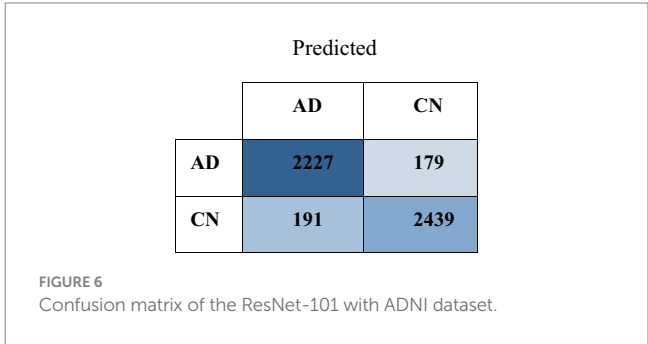
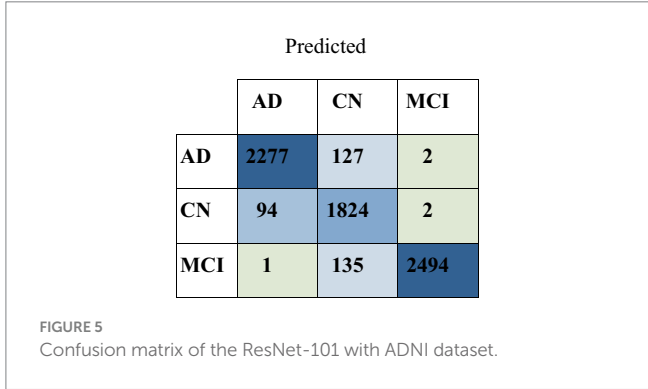
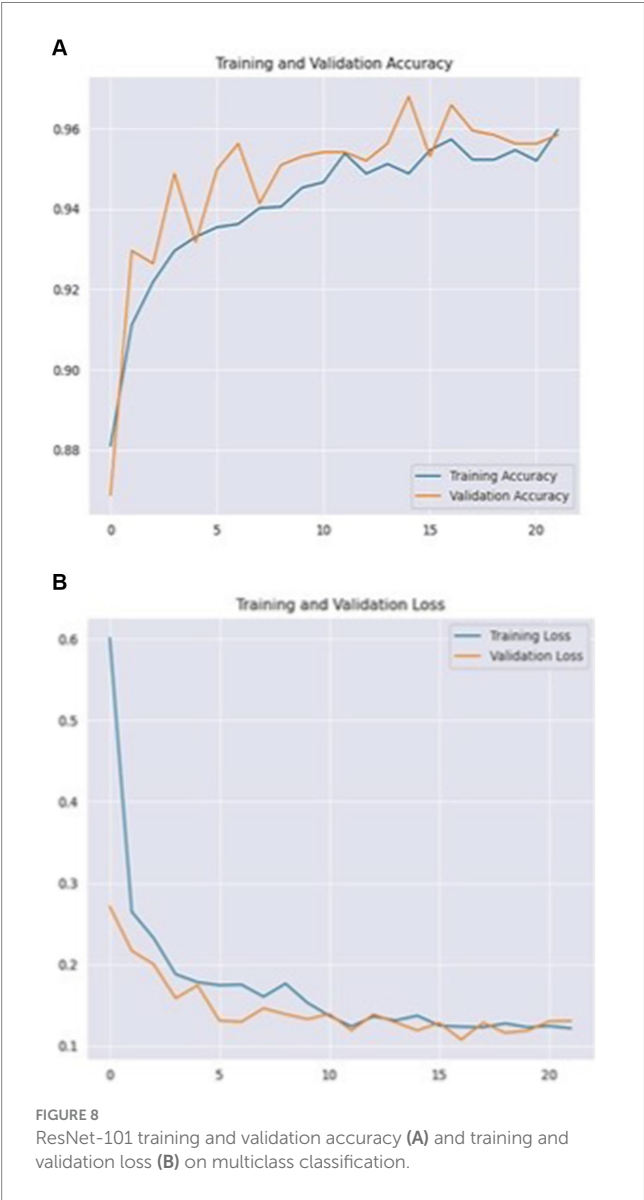


TABLE 7 MCI vs. CN (binary classification).

CNN	Accuracy		Precision		Recall		F1 Score	
	A	O	A	O	A	O	A	O
ResNet-50	75.23	69.45	80.34	74.56	75.34	68.23	78.45	71.23
ResNet-101	86.57	79.45	92.34	85.23	86.87	78.99	88.76	81.23
ResNet-152	89.56	90.11	93.12	86.12	87.34	79.45	89.23	82.34
DenseNet-201	84.32	77.89	90.45	82.67	84.45	76.78	87.34	79.56
EfficientNet-B0	82.45	76.23	89.12	81.34	82.56	74.93	85.12	78.32

A, ADNI; O, OASIS.



structure that positively influences the distinction between CN, AD, and MCI groups the most. It can be deduced from the analysis that ResNet-101 exhibited the highest performance among the CNNs, with ResNet-152 following suit, whereas EfficientNet-B0 demonstrated the lowest performance. Across both datasets, ResNet-101 consistently attained the highest accuracy and F1 score, showcasing its efficacy in identifying AD and MCI. In the same way that ResNet-101 outperformed ResNet-152, albeit marginally,

TABLE 8 Comparative analysis.

# Classes	Deep learning architecture	Acc. (%)	Ref.
2	3D-CNN	93.00	(16)
2	EfficientNet-B2	93.30	(19)
3	ResNet18, AlexNet, SqueezeNet, VGG16, InceptionV3 & DenseNet	82.53	(21)
2	DenseNet	96.51	(30)
3	ResNet-18	69.10	(31)
3	VGG-16 & 19	95.35	(32)
3	DenseNet-201, EfficientNet-B0, ResNet-50, ResNet-101, and ResNet-152	98.21	Our

EfficientNet-B0 demonstrated subpar performance, suggesting limitations in its ability to classify data, particularly when coupled with transfer learning and preprocessing utilizing Gaussian filters.

Table 8 depicts an empirical comparison of AD and MCI identification using prominent deep-learning architectures, showing that our investigation achieved the maximum efficacy on both datasets.

The analysis of deep learning architectures shows differing degrees of performance across various deep learning models. Table 8 presents the number of classes, the deep learning architecture used, and the attained accuracy. EfficientNet-B2 and 3D-CNN both demonstrated excellent accuracy in binary classification tests, suggesting their usefulness in the task. DenseNet demonstrated superior performance in a multi-class classification job with three classes, highlighting its resilience in addressing intricate classification challenges compared to other models. ResNet-18 obtained lower accuracy in a different multi-class classification scenario, indicating its shortcomings in hard classification tasks compared to other models. While in our investigation, the employed ResNet-101 obtained the highest accuracy in the multi-class classification challenge, showcasing enhanced performance. The findings emphasize the significance of choosing a suitable deep learning architecture according to the particular classification problem and the intricacy of the dataset. The research highlights the subtle variations in performance across different deep learning architectures, stressing the need to make well-informed choices to enhance model performance for specific tasks.

5 Conclusion

The investigation highlights the crucial use of modern neuroimaging and deep learning approaches in diagnosing and comprehending neurodegenerative disorders like Alzheimer's and

MCI. The extraction of valuable insights from complicated brain imaging using employed datasets ADNI and OASIS, which provide comprehensive MRI data, and implementation of advanced preprocessing techniques like skull stripping and segmentation on ADNI.

The U-Net architecture performed skull stripping on MRI images, successfully eliminating non-brain tissues. Specific deep learning models, such as DenseNet-201, EfficientNet-B0, ResNet-50, ResNet-101, and ResNet-152, were assessed for their ability to detect AD and MCI. Transfer learning is a powerful method for improving models, especially in situations with little datasets. Performance research shows that ResNet-101 regularly outperforms other models, followed closely by ResNet-152 with the datasets. ResNet-101 stands out as the best performer, attaining the most significant Accuracy levels and F1 Score on both datasets. This demonstrates its ability to effectively differentiate between people with AD, MCI, and CN instances, highlighting its resilience. ResNet-152 performed most in distinguishing between MCI and CN instances in a binary classification exercise with the OASIS dataset. The findings indicated that the CNN models performed well in this multi-class (29) and binary classification.

The study yielded promising results, yet several constraints and areas for future research remain to be addressed. Variations in model performance could stem from dataset characteristics and preprocessing methods. Further exploration of diverse preprocessing techniques and datasets is crucial to achieving a more comprehensive evaluation of model efficacy. While the research focused on a limited range of deep learning architectures, investigating additional structures and ensemble techniques may further enhance performance. Moreover, delving into model architecture choices and identifying biomarkers specific to Alzheimer's and MCI could deepen our understanding of the underlying mechanisms of these disorders. Integrating clinical data with neuroimaging holds potential to improve diagnostic accuracy and prognostic predictions for Alzheimer's and MCI. Future studies could benefit from combining these complementary sources of information to develop more robust and reliable predictive models. In the future research the research should focus on incorporating multimodality images like PET Scan with MRI for the more precise prediction and should try to use generative AI models on generating future brain images and use if for prediction.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding author.

References

1. What is mild cognitive impairment? Available at: <https://www.nia.nih.gov/health/memory-loss-and-forgetfulness/what-mild-cognitive-impairment>. November 22, 2023.
2. Angelucci F, Spalletta G, Iulio F, Ciaramella A, Salani F, Varsi A, et al. Alzheimers disease (AD) and mild cognitive impairment (MCI) patients are characterized by increased BDNF serum levels. *Curr Alzheimer Res.* (2010) 7:15–20. doi: 10.2174/156720510790274473
3. Kulasiri D, Aberathne I, Samarasinghe S. Detection of Alzheimer's disease onset using MRI and PET neuroimaging: longitudinal data analysis and machine learning. *Neural Regen Res.* (2023) 18:2134–40. doi: 10.4103/1673-5374.367840
4. ADNI. Alzheimer's Disease Neuroimaging Initiative. Available at: <https://adni.loni.usc.edu/>. February 20, 2023.
5. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J Cogn Neurosci.* (2007) 19:1498–507. doi: 10.1162/jocn.2007.19.9.1498
6. Cadena L, Zotin A, Cadena F, Espinosa N. Noise removal of the x-ray medical image using fast spatial filters and GPU. *Applications of Digital Image Processing XLI.* (2018). doi: 10.1117/12.2319327

Author contributions

PP: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. JP: Project administration, Supervision, Writing – review & editing. SA: Methodology, Validation, Writing – review & editing. AV: Formal analysis, Visualization, Writing – review & editing. BZ: Formal analysis, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship and/or publication of this article. This support was received by the Researchers Supporting Program at King Saud University. Researchers Supporting Project number (RSPD2024R867), King Saud University, Riyadh, Saudi Arabia.

Acknowledgments

The authors extend their appreciation to the Researchers Supporting Program at King Saud University. Researchers Supporting Project Number (RSPD2024R867), King Saud University, Riyadh, Saudi Arabia.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2024.1445325/full#supplementary-material>

7. Kumar S., Kumar H. PneuML: a novel sequential convolutional neural network-based x-ray diagnostic system for pneumonia in contrast to machine learning and pre-trained networks. Available at: https://www.scientificbulletin.upb.ro/rev_docs_arhiva/rez1f3_911998.pdf, (accessed on 2024/01/11).
8. Pei X, Zhao YH, Chen L, Guo Q, Duan Z, Pan Y, et al. Robustness of machine learning to color, size change, normalization, and image enhancement on micrograph datasets with large sample differences. *Mater Des.* (2023) 232:112086. doi: 10.1016/j.matdes.2023.112086
9. Rempe M, Mentzel F, Pomykala KL, Haubold J, Nensa F, Kroeninger K, et al. K-strip: a novel segmentation algorithm in k-space for the application of skull stripping. *Comput Methods Prog Biomed.* (2024) 243:107912. doi: 10.1016/j.cmpb.2023.107912
10. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation In: N Navab, J Hornegger, W Wells and A Frangi, editors. Medical image computing and computer-assisted intervention – MICCAI 2015. MICCAI 2015. Lecture notes in computer science, vol. 9351. Cham: Springer (2015)
11. Stoleru GI, Iftene A. Transfer learning for Alzheimer's disease diagnosis from MRI slices: a comparative study of deep learning models. *Procedia Comp Sci.* (2023) 225:2614–23. doi: 10.1016/j.procs.2023.10.253
12. Hoopes A, Mora JS, Dalca AV, Fischl B, Hoffmann M. SynthStrip: skull-stripping for any brain image. *NeuroImage.* (2022) 260:119474. doi: 10.1016/j.neuroimage.2022.119474
13. Huang G., Liu Z., Weinberger K.Q. Densely connected convolutional networks. Available at: <https://arxiv.org/abs/1608.06993v5>. January 28 2018.
14. Tan M. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. doi: 10.48550/arXiv.1905.11946
15. He K., Zhang X., Ren S., Sun J.. Deep residual learning for image recognition. Available at: <https://arxiv.org/abs/1512.03385v1>.
16. Shojaei S, Saniee Abadeh M, Momeni Z. An evolutionary explainable deep learning approach for Alzheimer's MRI classification. *Expert Syst Appl.* (2023) 220:119709. doi: 10.1016/j.eswa.2023.119709
17. EL-Geneedy M, Moustafa HE-D, Khalifa F, Khater H, Abdelhalim E. An MRI-based deep learning approach for accurate detection of Alzheimer's disease. *Alex Eng J.* (2023) 63:211–21. doi: 10.1016/j.aej.2022.07.062
18. Chen Y, Wang L, Ding B, Shi J, Wen T, Huang J, et al. Automated Alzheimer's disease classification using deep learning models with soft-NMS and improved ResNet50 integration. *J Radiat Res Appl Sci.* (2024) 17:100782. doi: 10.1016/j.jrras.2023.100782
19. Li H, Tan Y, Miao J, Liang P, Gong J, He H, et al. Attention-based and micro designed EfficientNetB2 for diagnosis of Alzheimer's disease. *Biomed Signal Process Control.* (2023) 82:104571. doi: 10.1016/j.bspc.2023.104571
20. Inan MSK, Sworna NS, Islam AKMM, Islam S, Alom Z, Azim MA, et al. A slice selection guided deep integrated pipeline for Alzheimer's prediction from structural brain MRI. *Biomed Signal Process Control.* (2024) 89:105773. doi: 10.1016/j.bspc.2023.105773
21. Odusami M, Maskeliūnas R, Damaševičius R, Krilavičius T. Analysis of features of Alzheimer's disease: detection of early stage from functional brain changes in magnetic resonance images using a Finetuned ResNet18 network. *Diagnostics.* (2021) 11:1071. doi: 10.3390/diagnostics11061071
22. Nguyen D, Nguyen H, Ong H, Le H, Ha H, Duc NT, et al. Ensemble learning using traditional machine learning and deep neural network for diagnosis of Alzheimer's disease. *IBRO Neurosci Reports.* (2022) 13:255–63. doi: 10.1016/j.ibneur.2022.08.010
23. Ioffe S. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. doi: 10.48550/arXiv.1502.03167
24. Yaseliani M, Ijadi Maghsoodi A, Hassannayebi E, Aickelin U. Diagnostic clinical decision support based on deep learning and knowledge-based systems for psoriasis: from diagnosis to treatment options. *Comput Ind Eng.* (2024) 187:109754. doi: 10.1016/j.cie.2023.109754
25. Iman M, Arabnia HR, Rasheed K. A review of deep transfer learning and recent advancements. *Secur Technol Des.* (2023) 11:40. doi: 10.3390/technologies11020040
26. Kumar S, Kumar H. Classification of COVID-19 X-ray images using transfer learning with visual geometrical groups and novel sequential convolutional neural networks. *MethodsX.* (2023) 11:102295. doi: 10.1016/j.mex.2023.102295
27. OASIS. Alzheimer's detection. Available at: <https://www.kaggle.com/datasets/ninadaithal/imagesoasis>.
28. Kumar S, Kumar H, Kumar G, Singh SP, Bijalwan A, Diwakar M. A methodical exploration of imaging modalities from dataset to detection through machine learning paradigms in prominent lung disease diagnosis: a review. *BMC Med Imaging.* (2024) 24:30. doi: 10.1186/s12880-024-01192-w
29. Pruthviraja D, Nagaraju SC, Mudligiriyappa N, Raisinghani MS, Khan SB, Alkhaldi NA, et al. Detection of Alzheimer's disease based on cloud-based deep learning paradigm. *Diagnostics.* (2023) 13:2687. doi: 10.3390/diagnostics13162687
30. Saleh AW, Gupta G, Khan SB, Alkhaldi NA, Verma A. An Alzheimer's disease classification model using transfer learning Densenet with embedded healthcare decision support system. *Decision Analytics J.* (2023) 9:100348. doi: 10.1016/j.dajour.2023.100348
31. Oktavian M.W. Classification of Alzheimer's Disease Using the Convolutional Neural Network (CNN) with Transfer Learning and Weighted Loss. doi: 10.48550/arXiv.2207.01584
32. Suk H-I, Lee S-W, Shen D. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage.* (2014) 101:569–82. doi: 10.1016/j.neuroimage.2014.06.077



OPEN ACCESS

EDITED BY

Sultan Ahmad,
Prince Sattam Bin Abdulaziz University,
Saudi Arabia

REVIEWED BY

Pankaj Dadheech,
Swami Keshvanand Institute of Technology,
India
Md Alimul Haque,
Veer Kunwar Singh University, India

*CORRESPONDENCE

Xiang Li
✉ lixiangzzchxy@163.com

RECEIVED 14 May 2024

ACCEPTED 26 December 2024

PUBLISHED 08 January 2025

CITATION

Wang L, Li X, Zhang Z, Yuan H, Lu P and
Li Y (2025) Comparing circular and
flexibly-shaped scan statistics for disease
clustering detection.
Front. Public Health 12:1432645.
doi: 10.3389/fpubh.2024.1432645

COPYRIGHT

© 2025 Wang, Li, Zhang, Yuan, Lu and Li. This
is an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Comparing circular and flexibly-shaped scan statistics for disease clustering detection

Lina Wang¹, Xiang Li^{2*}, Zhengbin Zhang³, Haoxun Yuan¹,
Pengfei Lu² and Yaru Li¹

¹School of Computer Science and Technology, Zhengzhou University of Light Industry, Zhengzhou, China, ²Institute of Surveying and Mapping, Information Engineering University, Zhengzhou, China, ³Tuberculosis Prevention and Control Office, Wuhan Institute for Tuberculosis Control, Wuhan, China

The accuracy of spatial clustering detection is crucial for public health policy development and identifying etiological clues. Circular and flexibly-shaped scan statistics are widely used for disease cluster detection, but differences in results arise mainly due to parameter sensitivity and variations in the scanning window shapes. This study aims to analyze the impact of parameter settings on the results of these methods and compare their performance in disease clustering detection. Using tuberculosis data from Wuhan, China (2015–2019), the study identified the optimal parameter settings—MSWS and K -value—for each method to ensure accurate clustering. A comprehensive comparison was made using two quantitative indicators, the LLR value and cluster size, as well as clustering visualizations. The results show that the optimal MSWS parameter for SaTScan is determined through a Gini coefficient-based stepwise-threshold-reduction approach, while a K -value of 30 is ideal for FlexScan. SaTScan tends to produce more regular clusters, while FlexScan often generates more irregular clusters. FlexScan detects fewer clusters but with higher LLR values and larger average cluster sizes, although the maximum cluster size is smaller. These findings provide valuable insights for optimizing disease clustering detection methods and enhancing public health interventions.

KEYWORDS

spatial scan statistics, disease cluster detection, SaTScan, FlexScan, Gini coefficient, log-likelihood ratio (LLR), cluster size

1 Introduction

The purpose of spatial cluster detection of diseases is to identify whether clustering disease exists and to locate the areas where these clusters occur. This information can provide clues for further etiological investigation. Spatial scan statistics have been widely used as a technique for detecting disease clusters (1–4). This method was first introduced by Kulldorff, along with the freely available SaTScan software, and has since been extended with several different statistical models. The method utilizes a likelihood ratio test statistic to evaluate a large number of different and overlapping scanning windows. The test statistic is formulated based on a probability model depending on the data type, such as the Poisson model for count data. However, this method is limited to circular scan windows for detecting compact clusters, which may struggle to accurately identify non-circular clusters. Consequently, other researchers have proposed alternative approaches that employ different scanning window selection schemes, such as elliptical (5–7) and flexibly-shaped windows (8–10).

A popular alternative for detecting clusters with arbitrary shapes is the flexibly shaped spatial scan statistic proposed by Tango and Takahashi, which is implemented in the FlexScan software. This method employs an adjacency expansion search, scanning adjacent units in the

spatial region to detect irregularly shaped clusters (8–10). However, the selection of scanning windows in FleXScan is dependent on an exhaustive search strategy, which leads to exponential runtime scaling as the K -value increases. Here, the K -value is a constant that indicates the maximum number of sub-regions allowed within a preset window, serving as a crucial parameter in the implementation of the FleXScan method. Due to computational constraints, the K -value is typically limited to 30, with a default of 15. To address these limitations, Tango and Takahashi (10) proposed a restricted version of the flexibly-shaped scan statistic that focuses exclusively on regions with elevated risk. This modified approach offers improved computational efficiency and enables the use of a K -value up to 30 (10, 11).

In addition, Speakman et al. (12) developed the GraphScan method for detecting connected clusters of arbitrary shapes in graph or network data. This method improves search efficiency by incorporating a branch-and-bound depth-first search approach, which enhances the brute force algorithm used in FleXScan. Cadena et al. (13) presented a framework for network anomaly detection based on scan statistics that outperforms existing methods in terms of performance and scalability. Meysami M et al. (6) proposed the flexible–elliptical scan method, which combines the flexible and elliptic scan methods to address their respective limitations and leverage their advantages. However, for non-statistical users such as epidemiologists and public health researchers, user-friendly software may be more practical than introducing new algorithms. Currently, the most commonly used methods are still circular and flexibly-shaped scan statistics, which can be directly implemented in SaTScan and FleXScan, respectively.

However, both methods face the challenge of parameter setting during implementation, as the cluster results are highly sensitive to these parameters. For example, the performance of circular spatial scan statistics is influenced by the selection of the maximum scanning window size (MSWS) (14–16). If the MSWS is too large, the detected clusters may be overly large and may include areas with non-elevated risk. Conversely, if it is too small, numerous small clusters may be detected. Different MSWS values yield varying cluster sizes, locations, and numbers within the same dataset. Although it is common to use 50% of the total population as the default setting for MSWS in SaTScan, this may result in an overly large cluster. Therefore, determining the optimal MSWS value is crucial for the SaTScan method. Performance indicators such as sensitivity, specificity, positive predictive value (PPV), and Youden's index (YDI) are typically used to select the optimal MSWS (16), but these metrics are often only available in simulation studies. Han et al. (17) proposed the Gini coefficient as an effective criterion for determining optimal cluster reporting sizes, which helps avoid unnecessarily large and less informative clusters. This approach has been implemented in SaTScan version 9.3 and has shown success with both simulated and real data (18, 19). Another indicator, called the maximum clustering heterogeneous set-proportion (MCHS-P), was introduced by Wang et al. (16) for selecting suitable MSWS. However, the Gini coefficient remains widely used due to its direct application through SaTScan, despite some limitations pointed out by Li et al. (15) and Wang et al. (16).

The FleXScan method suggests that setting $K = 30$ theoretically helps achieve the optimal maximum likelihood clustering (MLC). However, it is important to understand the impact of different K -values on the final clustering results. For example, if we set $K = 15$, can we still achieve a good MLC, and what are the differences between the two clustering results? Evaluating the influence of K -value requires practical analysis and comparison.

The accuracy of spatial cluster detection results is of great significance for the formulation of prevention and control policies in the region and the detection of further etiology. Spurious cluster results, however, may have unnecessary negative impacts on the socio-economic development of that region (20). Therefore, selecting the appropriate parameter settings is important for accurate cluster identification. Unfortunately, there is currently no standard reference criterion for parameter selection.

Previous studies have demonstrated that different research purposes require different parameter combinations for analysis. However, most previous studies are based on simulated data with specific assumptions, and the conclusions drawn may not be fully applicable to real data, which has certain limitations. The optimal parameter combination varies with different data, and the conclusions of simulation research are often difficult to extend to more complex and variable real-world scenarios without sufficient prior knowledge.

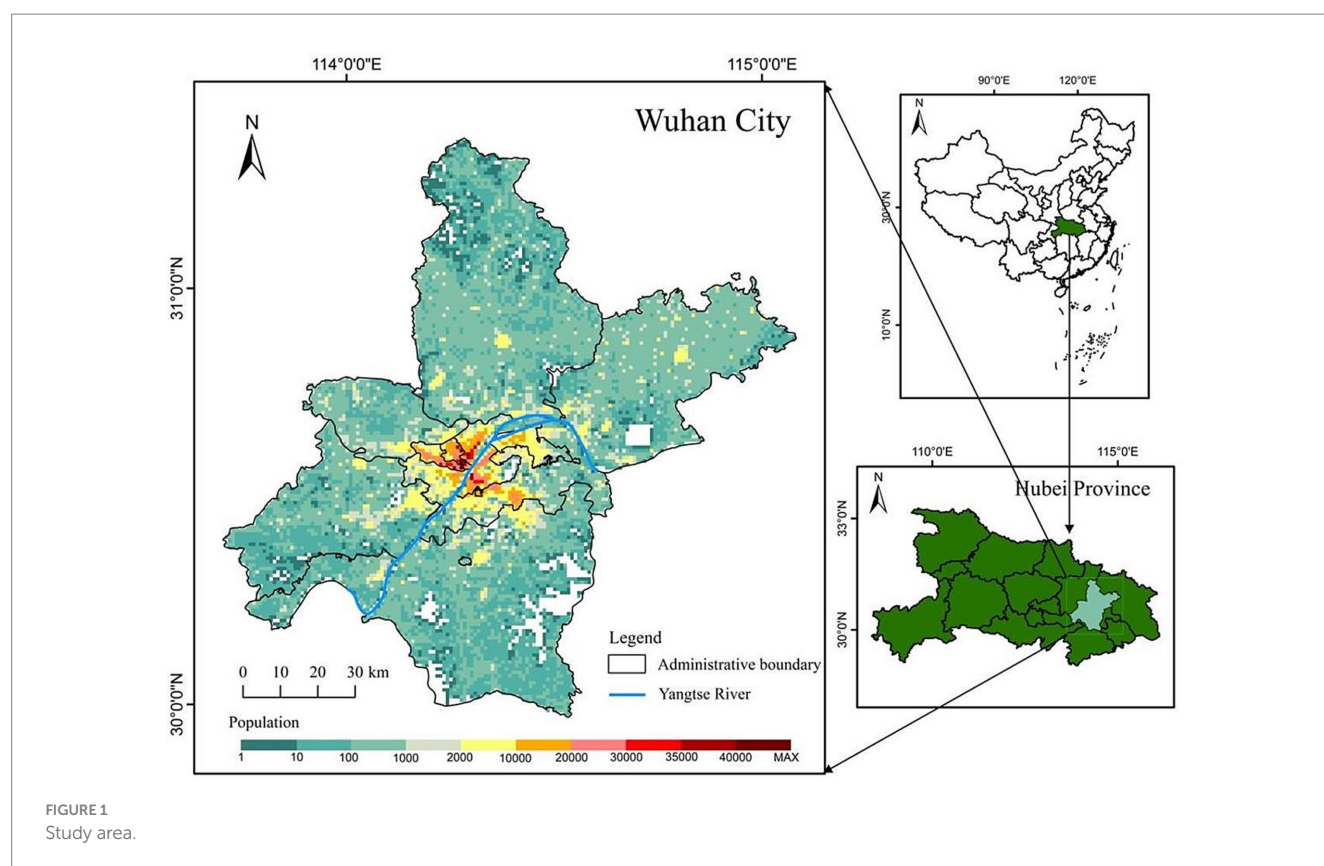
Therefore, the purpose of our study is to compare the differences between the two different scanning window methods, and to determine the optimal parameter settings for each method. This will clarify the impact of parameter settings on the results and provide a reference for other researchers. We will utilize real pulmonary tuberculosis disease data (at the township level) from Wuhan, spanning 2015 to 2019, as our research dataset. We will provide optimal parameter settings for the two scanning window types in different years and compare the spatial clustering results obtained from these methods.

2 Study area and data

Our study area is Wuhan City, the capital of Hubei Province, located in central China. Known as “the River City,” Wuhan is situated at the confluence of the Yangtze River and the Han River the largest tributary of the Yangtze. This strategic location has made Wuhan a crucial transportation hub, connecting various parts of China through its extensive network of railways, highways, and waterways. Wuhan City comprises 13 county-level units and 164 town-level units, with a total area of 8,569.15 square kilometers. As of the end of 2021, according to official information, Wuhan had a permanent population of 12.3265 million, with its population spatial distribution shown in Figure 1. It can be observed from the figure that the central urban area is densely populated, while the peripheral rural areas are sparsely populated.

The case information for pulmonary tuberculosis in this study was obtained from the National Tuberculosis Management Information System, specifically the registered and managed medical records of pulmonary tuberculosis patients based on their initial diagnosis locations from 2015 to 2019. A total of 30,486 pulmonary tuberculosis patients were included in the study. We first employed geocoding techniques to spatially encode the addresses of the cases and then combined this with population demographic data to obtain the incidence rate at the township level. Thus, the final research data used

Abbreviations: LLR, Log-Likelihood Ratio; RR, Risk Ratio; MSWS, Maximum Scanning Window Size; MLC, Maximum Likelihood Clustering; MCHS-P, Maximum Clustering Heterogeneous Set-Proportion; PPV, Positive Predictive Value; YDI, Youden's index.



in this study consisted of the pulmonary tuberculosis incidence rates in Wuhan at the township level from 2015 to 2019.

3 Methods

The study involved determining optimal parameters, visualizing incidence and disease clustering results, and conducting a comparative analysis. We assessed and compared the performance of both methods in detecting disease clusters by evaluating LLR values and cluster size. To facilitate comparability, both methods were implemented using the same statistical model, specifically the Poisson statistical model.

3.1 Evaluation metrics for comparison

3.1.1 The LLR value

The LLR value quantifies the deviation of observed data from random spatial distribution. A higher LLR value suggests a higher likelihood of non-random clustering, indicating the presence of genuine spatial clusters. Comparing LLR values allows us to assess the strength and significance of detected clusters, helping to identify meaningful and informative clusters in the analysis.

3.1.2 The cluster size

The cluster size represents the number of sub-regions contained within a cluster. Restricting the cluster size may help reduce the likelihood of misclassifying random noise as clusters. Tango et al. (8) pointed out that it is unlikely for the size of a true cluster to be larger

than 10–15 percent of the total number of regions. However, this is not a fixed rule and may vary depending on the specific research field and data characteristics.

3.2 MSWS settings for the circular scan statistic

Initially, we attempted to use the same MSWS value for different years within the same spatial region. However, this approach proved to be unreasonable, as the spatial distribution of diseases varied significantly across different years. To address this issue, we utilized Gini coefficients to assist in identifying the optimal clusters. The Gini coefficient is a statistical measure of data inequality, which helps evaluate the quality of clustering results under different MSWS values. A higher Gini coefficient indicates a more uneven spatial distribution of the clustering result, suggesting that the clustering results achieved at that particular MSWS value possess greater distinctiveness and significance in terms of differentiation.

In this study, we tested MSWS values of 5, 10, 15, 20, 25, 30, 40, and 50% for each year from 2015 to 2019. We then calculated the corresponding Gini coefficients for each MSWS value. The MSWS value associated with the highest Gini coefficient was selected as the optimal choice. Our analysis revealed that even within the same spatial region, the optimal MSWS values varied due to differences in the spatial distribution of diseases across different years.

However, we have found in practice that simply using the Gini coefficient as a criterion for determining MSWS is insufficient. According to the Gini coefficient, the optimal MSWS in 2018 should

be 25%. However, the spatial clustering result at this value includes too many sub-regions, with the most likely cluster (MLC) containing 31 sub-regions and the secondary cluster containing 39 sub-regions. Together these two clusters cover nearly 50% of the total number of sub-regions. Clearly, the clusters are too large and may contain non-clustered areas. Therefore, we predetermined the MSWS value to be 10% when the Gini coefficient was the second largest. At this MSWS value, the MLC is divided into two small clusters and non-cluster regions, resulting in a significant reduction in the number of intra-cluster sub-regions. Although the LLR value of the MLC decreased, the LLR values of other clusters increased. The results are presented in [Table 1](#) and [Figure 2](#).

Finally, considering both the Gini coefficient and the number of sub-regions included in the clusters, we determined the optimal MSWS values for this study, as shown in [Table 2](#).

3.3 K-value setting for the flexibly-shaped scan statistic

Theoretically, a larger K -value increases the number of candidate scan windows that need to be calculated, but it also enhances the likelihood of identifying clusters with higher LLR values, indicating a higher probability of detecting true clusters. From this perspective, a K -value of 30 is ideal. However, to assess the effect of the K -value on the final results, we compared the results for $K = 15$ and $K = 30$. These results are presented in [Table 3](#), and the spatial clusters are also presented on the map in [Figure 3](#).

The results indicate that the spatial distribution of clusters is roughly the same when $K = 15$ and $K = 30$, but there are differences in cluster levels, as ordered by descending LLR values. Notably, there are variations in the spatial distribution of the MLC and the number of sub-regions included. When $K = 30$, the MLC contains more sub-regions and has a higher LLR value. Due to the limitation of the K -value, when $K = 15$, clusters 3 and 5 are identified as two separate clusters, with the LLR values of 29.72 and 16.27, respectively. However, these two clusters merge into a single, larger cluster when $K = 30$, with the LLR value increasing to 50.02.

3.4 Cluster visualization

To facilitate understanding of the results, we visualize the cluster analysis results on the map, using the color brightness to indicate the

magnitude of the LLR statistical value. Darker colors correspond to higher LLR values, suggesting a greater likelihood of true clustering. Additionally, we employ different color lightness on the map to represent the incidence rates of townships and streets as a reference for the spatial clustering results.

4 Results

4.1 Comparison of SaTScan and FleXScan in evaluation metrics

[Tables 4–8](#) present detailed comparison results of the SaTScan and FleXScan methods from 2015 to 2019. These tables include information on cluster level (ordered by descending LLR values), number of sub-regions (i.e., cluster size), number of cases, expected number of cases, population, RR value, LLR value, and p value.

By comparing the relevant information of the two methods in [Tables 4–8](#), particularly the LLR values and cluster sizes, we find that FleXScan identifies fewer clusters than SaTScan, but generally with higher LLR values. This suggests that the FleXScan method applies stricter criteria for defining clusters, reducing the likelihood of falsely identifying non-cluster areas as clusters. Consequently, while FleXScan may detect fewer true clusters, it is likely more accurate in identifying statistically significant clusters. The higher LLR values associated with FleXScan indicate stronger clustering signals, reflecting a greater probability of detecting true clusters.

In addition, we measure cluster size by the number of sub-regions covered. [Tables 4–8](#) show that FleXScan identifies clusters with a larger average size but a smaller maximum size compared to the SaTScan method. This suggests that FleXScan tends to recognize larger, more consistent clusters but with a less extreme maximum size. In contrast, SaTScan produces results with greater variability in cluster sizes, indicating more dispersed and variable cluster sizes. Consequently, FleXScan demonstrates higher stability in cluster size compared to SaTScan, as it produces more consistent cluster sizes across different datasets.

Both methods generally produce MLCs of similar sizes, typically containing fewer than 16 sub-regions, representing less than 10% of the total 164 sub-regions. However, 2017 is an exception, with SaTScan's MLC size reaching 31 sub-regions, compared to 20 sub-regions for FleXScan. This discrepancy is mainly due to SaTScan's higher MSWS value of 20% in 2017, which was larger than in other years and resulted in a larger MLC size.

TABLE 1 Comparison of cluster results with different MSWS values in SaTScan.

MSWS = 25% (2018)						MSWS = 10% (2018)					
Level of clusters	Number of Sun-region	Number of cases	Population in risk	LLR	p -value	Level of clusters	Number of Sun-region	Number of cases	Population in risk	LLR	p -value
MLC	31	1,270	1,861,018	38.70	<0.001	MLC	12	614	815,615	32.70	<0.001
2	39	1,517	2,398,624	22.66	<0.001	2	4	242	279,704	23.58	<0.001
3	2	143	181,846	9.00	<0.001	3	3	111	111,182	17.33	<0.001
						4	1	69	60,229	15.43	<0.001
						5	13	586	895,986	10.76	0.002

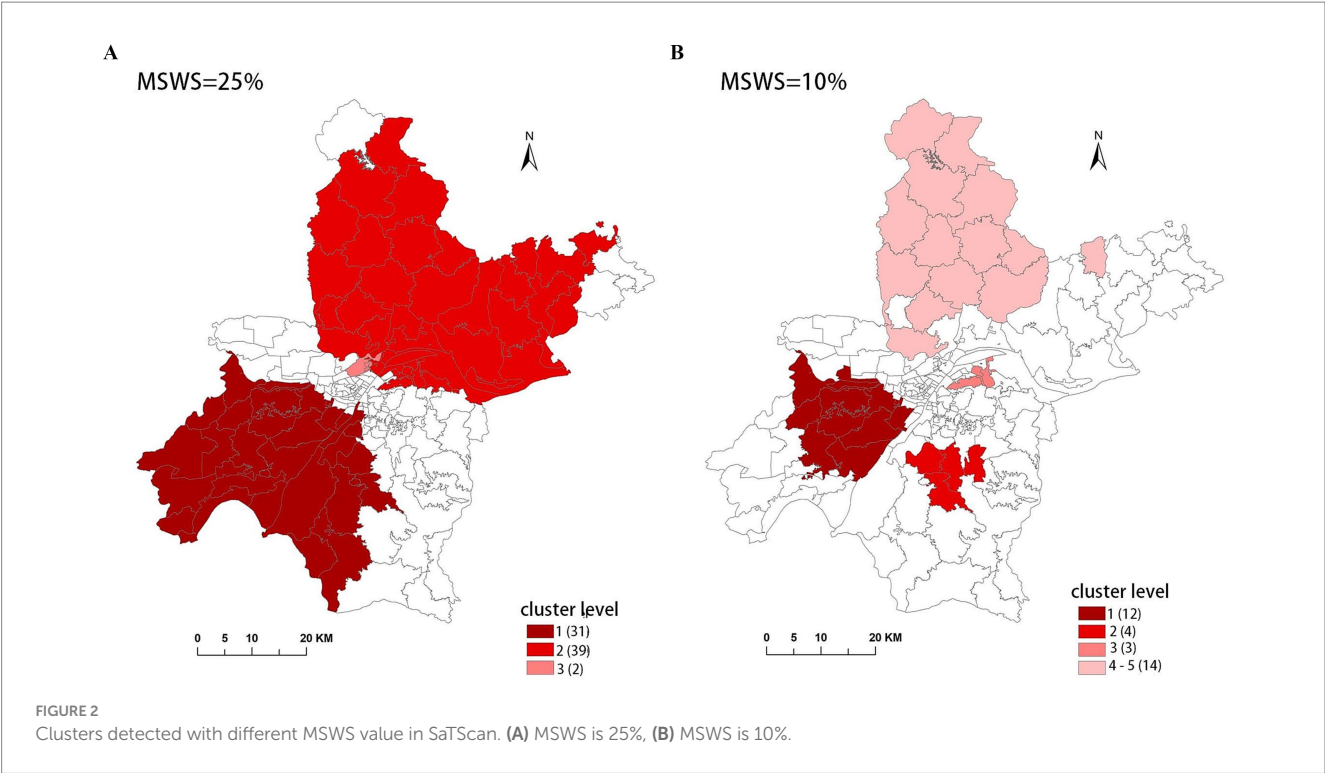


TABLE 2 The selected optimal MSWS value in different years.

Year	2015	2016	2017	2018	2019
MSWS	15%	10%	20%	10%	15%

This difference in performance arises from FlexScan’s use of an exhaustive algorithm, which evaluates all potential scan windows to pinpoint those with the highest LLR values. This approach allows FlexScan to be more precise in detecting clusters and identifying significant clustering patterns, as it thoroughly assesses a wide range of possible cluster configurations. By contrast, SaTScan utilizes a circular scanning window, which can constrain its ability to capture irregularly shaped or more complex clustering patterns. The circular window’s limitations can result in less accurate cluster detection and higher variability in the sizes of detected clusters. Furthermore, the SaTScan method, which involves scanning regions with progressively larger circles, might miss clusters that are not well-aligned with the circular shape or that have non-uniform spatial distributions. This can lead to less consistent results and a greater variability in cluster sizes, as observed in the data.

4.2 Comparison of SaTScan and FlexScan in cluster visualization

Figures 4–8 show the clustering results generated by the SaTScan and FlexScan methods in Wuhan from 2015 to 2019 on maps. Additionally, pulmonary tuberculosis incidence maps are provided for comparison and reference.

To accurately represent the cluster areas, we used the cluster regions comprising all polygons whose centroids are enclosed by the cluster circle, rather than directly using the cluster circles generated by SaTScan.

Furthermore, since certain years have numerous cluster levels with only a few regions per level (e.g., in 2016, SaTScan identified a total of 10 cluster levels, many of which included only one sub-region). To improve the legibility of the visualization, we categorized the original clustering regions into four categories based on LLR values and the number of included sub-regions. The legend specifies the number of sub-regions in each category, and the MLC was assigned a single distinct category. This classification converted the original data into ordinal data, represented by different lightness of color in Figures 4–8. The color intensity in each map corresponds to the clustering area level determined by the LLR value, with darker colors indicating higher likelihoods.

From Figures 4–8, it is evident that the clusters obtained by SaTScan are more regular in shape, whereas those identified by FlexScan exhibit excessive irregularity. Clusters with highly irregular shapes may be less meaningful, as they complicate the assessment of geographical significance for practitioners (9, 17).

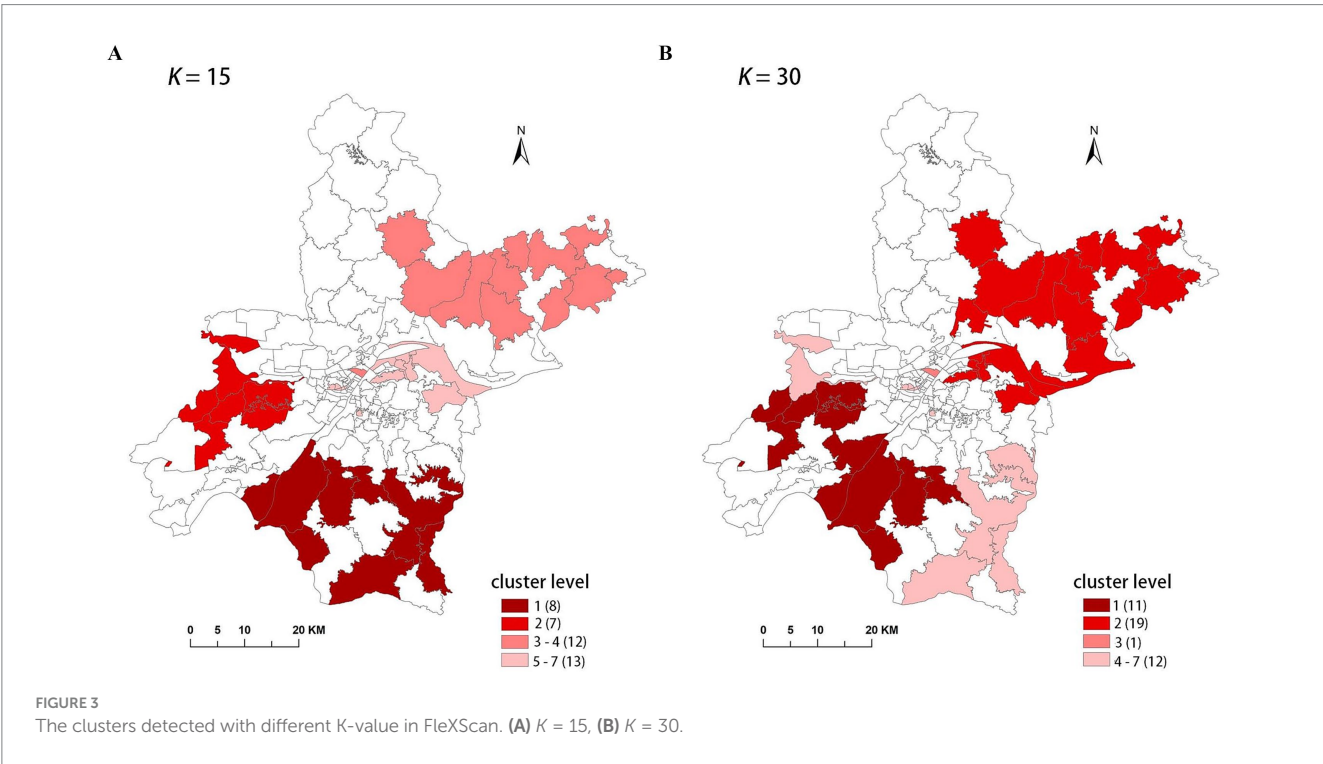
Overall, the clusters identified by SaTScan and FlexScan generally align with the spatial distribution of high-incidence areas, primarily located in the suburban districts of Wuhan, which are sparsely populated and economically underdeveloped. However, some high-incidence areas were not identified as clusters, suggesting that the elevated incidence rates in these regions may be random. Although the spatial coverage of clusters identified by both methods is largely similar, there are significant differences in the LLR values and the specific sub-regions included. Notably, the spatial distribution of MLCs identified by the two methods from 2015 to 2019 differs considerably, with the exception of 2017, where the distributions are similar.

5 Discussion

In this study, we undertook a detailed comparative analysis of the SaTScan and FlexScan methods for disease clustering using real data

TABLE 3 Comparison of cluster results with different *K*-values in FlexScan.

2015 (<i>K</i> = 15)						2015 (<i>K</i> = 30)					
Level of clusters	Number of Sun-region	Number of cases	Population in risk	LLR	<i>p</i> - value	Level of clusters	Number of Sun-region	Number of cases	Population in risk	LLR	<i>p</i> - value
MLC	8	455	477,903	46.16	0.001	MLC	11	645	688,811	62.79	0.001
2	7	314	314,546	37.30	0.001	2	19	738	856,888	50.02	0.001
3	11	490	577,467	29.72	0.001	3	1	70	46,702	22.71	0.001
4	1	70	46,702	22.71	0.001	4	2	72	57,870	15.75	0.003
5	7	215	240,358	16.27	0.001	5	1	72	58,570	15.30	0.004
6	1	72	58,570	15.30	0.001	6	4	111	105,171	15.04	0.004
7	5	223	257,809	14.40	0.001	7	5	223	257,809	14.40	0.006



from 2015 to 2019. This comparison aimed to explore the strengths and limitations of each method in accurately detecting disease clusters.

5.1 Real disease data

Unlike most previous studies that relied on simulated data, our research utilized real disease data. Although this limited our ability to use common quantitative metrics, such as sensitivity, specificity, PPV, and YDI, to determine the optimal parameter settings and compare method performance, using LLR values and cluster size as our analysis metrics is still appropriate, albeit not entirely comprehensive. Nevertheless, real disease data better reflect actual disease distribution and trends, enhancing the realism and generalizability of our results and providing more reliable support for public health decision-making.

5.2 Parameter settings

The Gini coefficient, traditionally used to determine the MSWS in SaTScan, has been validated as effective by some researchers (17, 18). However, our findings align with those of Li et al. (15), who identified limitations in this approach. Specifically, the Gini coefficient measures overall distribution uniformity across the entire region and may not capture the nuances of smaller, individual clusters when multiple clusters are present. This may lead to multiple small clusters being combined into one large cluster, resulting in distorted results, as confirmed in our study illustrated in Section 3.2. To address this issue, we recommend combining the Gini coefficient approach with a gradual reduction in the threshold to accurately identify and separate individual clusters, thereby obtaining more reliable clustering results.

FlexScan identifies clusters using LLR values. While setting *K* = 30 is theoretically optimal, our study reveals that focusing solely

TABLE 4 Comparison of SaTScan clusters and FleXScan clusters in 2015.

SaTScan-2015 (MSWS = 15%)								Restricted FleXScan-2015 (K = 30)							
Cluster level	Number of sub-regions	Number of cases	Expected cases	Population in risk	RR	LLR	p - value	Cluster level	Number of sub-regions	Number of cases	Expected cases	Population in risk	RR	LLR	p - value
MLC	12	529	351.0	591,071	1.56	41.78	<0.0001	MLC	11	645	409.0	688,811	1.64	62.79	0.001
2	7	324	214.9	361,950	1.54	24.93	<0.0001	2	19	738	508.8	856,888	1.51	50.02	0.001
3	29	1,035	841.7	1,417,641	1.28	24.225	<0.0001	3	1	70	27.7	46,702	2.54	22.71	0.001
4	1	70	27.7	46,702	2.54	22.70	<0.0001	4	2	72	34.4	57,870	2.11	15.75	0.003
5	1	41	14.8	24,884	2.79	15.68	<0.0001	5	1	72	34.8	58,570	2.08	15.30	0.004
6	1	72	34.8	58,570	2.08	15.29	<0.0001	6	4	111	62.5	105,171	1.79	15.04	0.004
7	5	224	164.7	277,338	1.37	9.89	0.006	7	5	223	153.1	257,809	1.47	14.40	0.006

TABLE 5 Comparison of SaTScan clusters and FleXScan clusters in 2016.

SaTScan-2016 (MSWS = 10%)								Restricted FleXScan-2016 (K = 30)							
Cluster level	Number of sub-regions	Number of cases	Expected cases	Population in risk	RR	LLR	p - value	Cluster level	Number of sub-regions	Number of cases	Expected cases	Population in risk	RR	LLR	p - value
MLC	8	431	264.5	432,035	1.67	46.21	<0.0001	MLC	9	566	348.3	568,916	1.69	61.09	0.0005
2	1	91	36.4	59,445	2.52	29.02	<0.0001	2	6	388	227.5	371,545	1.75	48.79	0.0005
3	17	604	459.2	749,933	1.35	22.54	<0.0001	3	7	357	245.3	400,657	1.48	23.30	0.0005
4	3	169	109.2	178,290	1.56	14.31	<0.0001	4	4	200	135.9	221,932	1.49	13.53	0.01
5	1	59	27.8	45,394	2.13	13.28	0.0001	5	1	59	27.8	45,394	2.13	13.29	0.01
6	1	14	2.7	4,407	5.20	11.76	<0.0007	6	1	14	2.7	4,407	5.20	11.77	0.03
7	4	248	186.2	304,145	1.35	9.58	0.005								
8	1	63	35.8	58,518	1.77	8.44	0.01								
9	1	34	25.6	25,420	2.19	8.16	0.02								
10	1	52	29.0	47,400	1.80	7.39	0.03								

TABLE 6 Comparison of SaTScan clusters and FleXScan clusters in 2017.

SaTScan-2017 (MSWS = 20%)								Restricted FleXScan-2017 ($K = 30$)							
Cluster level	Number of sub-regions	Number of cases	Expected cases	Population in risk	RR	LLR	p - value	Cluster level	Number of sub-regions	Number of cases	Expected cases	Population in risk	RR	LLR	p - value
MLC	31	1,374	1048.8	1,829,721	1.40	56.64	<0.0001	MLC	20	1,126	768.4	1,340,576	1.57	85.03	0.0005
2	3	111	62.7	109,295	1.79	15.34	<0.0001	2	11	511	340.5	594,032	1.55	39.53	0.0005
3	4	215	145.1	253,094	1.50	15.07	<0.0001	3	4	188	128.4	223,975	1.48	12.40	0.02
4	1	64	33.9	59,207	1.90	10.61	<0.0001								
5	5	214	163.2	284,795	1.32	7.40	<0.0001								

TABLE 7 Comparison of SaTScan clusters and FleXScan clusters in 2018.

SaTScan-2018 (MSWS = 10%)								Restricted FleXScan-2018 ($K = 30$)							
Cluster level	Number of sub-regions	Number of cases	Expected cases	Population in risk	RR	LLR	p - value	Cluster level	Number of sub-regions	Number of cases	Expected cases	Population in risk	RR	LLR	p - value
MLC	12	614	441.7	815,615	1.44	32.70	<0.0001	MLC	13	703	495.0	913,950	1.48	42.75	0.0005
2	4	242	151.5	279,704	1.62	23.58	<0.0001	2	11	504	337.7	623,635	1.54	38.05	0.0005
3	3	111	60.2	111,182	1.86	17.33	<0.0001	3	10	509	355.2	655,898	1.47	31.51	0.0005
4	1	69	32.6	60,229	2.13	15.43	<0.0001	4	7	336	228.1	421,252	1.50	23.28	0.0005
5	13	586	485.2	895,986	1.23	10.76	0.002	5	3	213	150.2	277,388	1.43	11.95	0.03

TABLE 8 Comparison of SaTScan clusters and FlexScan clusters in 2019.

SaTScan-2019 (MSWS = 15%)								Restricted FlexScan-2019 (K = 30)							
Cluster level	Number of sub-regions	Number of cases	Expected cases	Population in risk	RR	LLR	p - value	Cluster level	Number of sub-regions	Number of cases	Expected cases	Population in risk	RR	LLR	p - value
MLC	6	364	227.5	422,571	1.64	36.26	<0.0001	MLC	12	706	427.6	794,267	1.74	82.93	0.0005
2	7	248	140.5	260,952	1.80	34.46	<0.0001	2	13	851	545.8	1,013,967	1.65	81.73	0.0005
3	7	329	205.4	381,515	1.64	32.77	<0.0001	3	8	362	238.8	443,597	1.55	28.78	0.0005
4	1	81	32.8	60,941	2.49	25.22	<0.0001	4	6	197	125.4	232,961	1.59	17.84	0.003
5	2	166	99.0	183,995	1.70	19.16	<0.0001	5	5	130	81.1	150,606	1.62	12.67	0.02
6	6	192	131.3	243,977	1.48	12.57	0.00048								
7	4	212	152.3	283,010	1.41	10.71	0.0026								
8	1	77	43.9	81,566	1.76	10.25	0.0039								
9	3	172	121.4	225,519	1.43	9.55	0.0073								

on the highest LLR values can lead to clusters with highly irregular shapes (21–23). Duczmal et al. (21, 22) have noted that such irregular shapes can complicate geographic interpretation and suggest that both LLR values and cluster shapes should be considered together to achieve clusters that are both statistically significant and meaningful. Irregular cluster shapes may arise from specific geographic features, population distribution, or data noise and might not accurately reflect the actual disease distribution. Therefore, considering the regularity of cluster shapes is important to avoid misleading interpretations and ineffective public health interventions. The current version of FlexScan lacks features to control or modify cluster shapes, highlighting the need for further developments in this area.

5.3 Computational efficiency

Although both methods can be implemented through software, it is essential to discuss their computational efficiency to gain a deeper understanding of the differences in their results. The efficiency primarily depends on the number of scanning windows that need to be calculated.

In the SaTScan method, let the entire study area contain *m* sub-regions. For each region, the scanning radius varies systematically from 0 to a predefined maximum (MSWS value), centered on each region. If each region has *T* concentric circular windows, the maximum number of windows that need to be calculated is *m* × *T*.

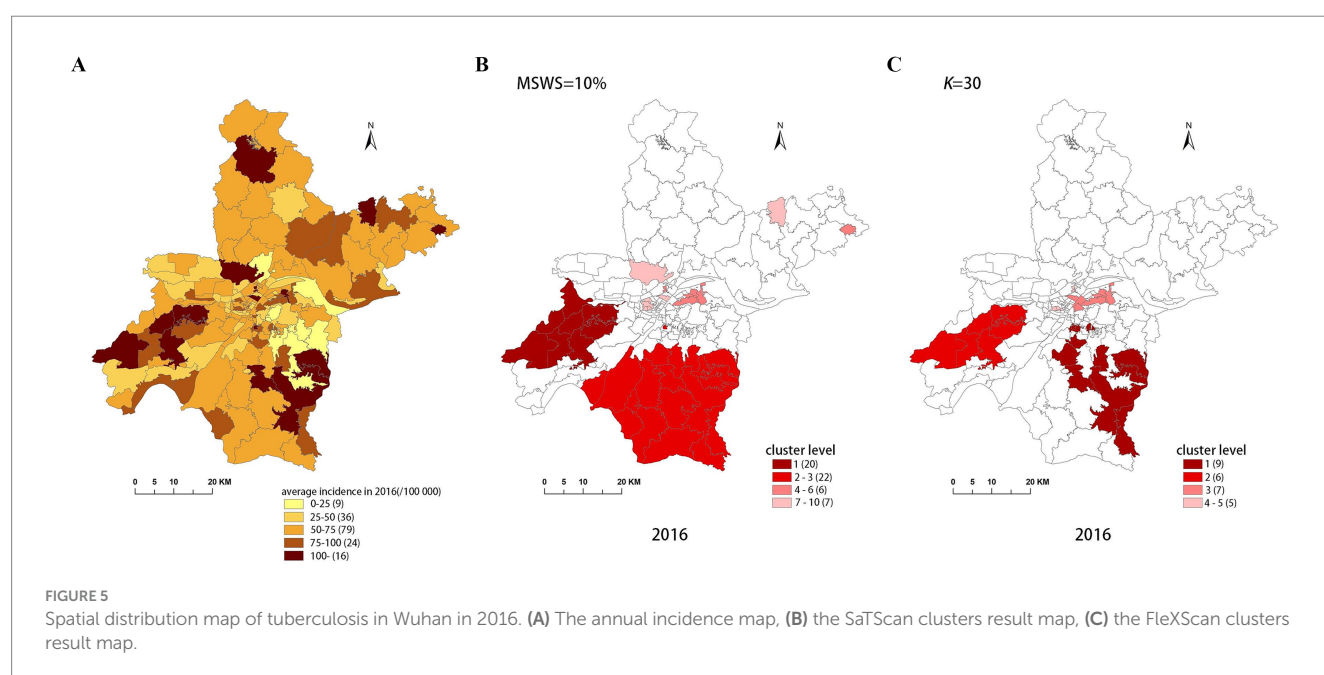
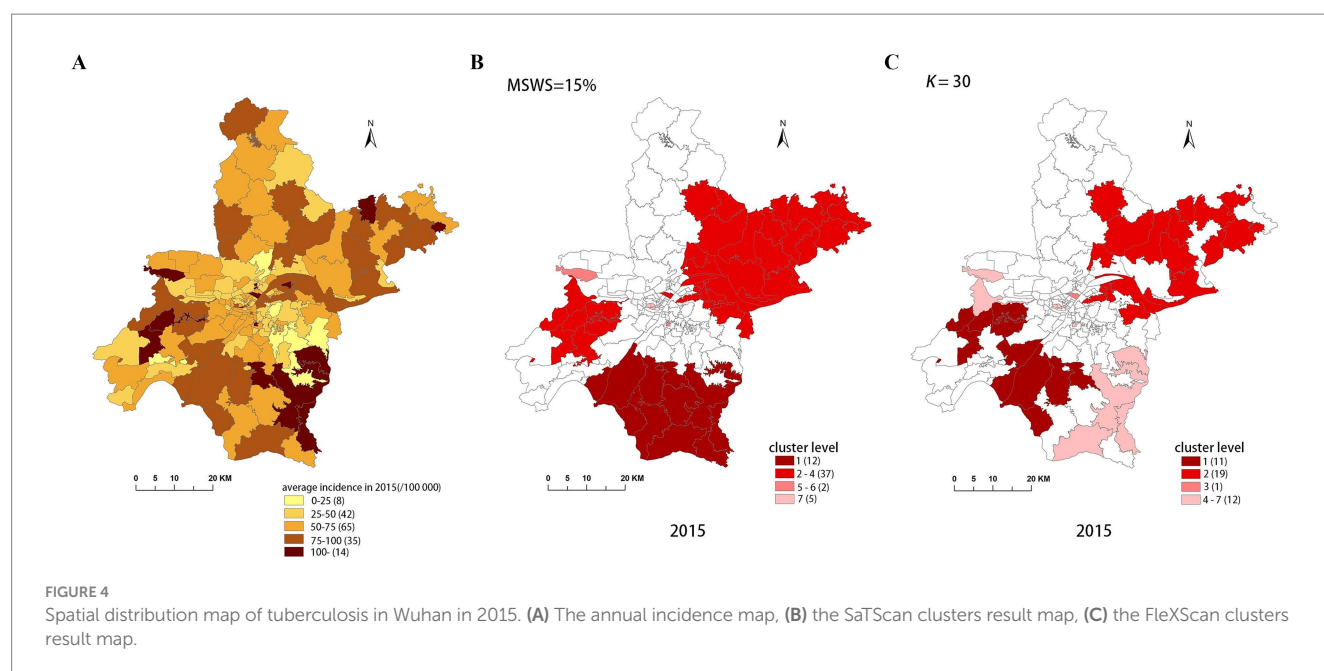
In contrast, the FlexScan method requires calculating a greater number of scanning windows. The process is as follows:

Let *Z_i* represent region *i* (1 ≤ *i* ≤ *m*), and *Z_{ik}* denote the scanning window formed by sub-region *Z_i* and its *k*-1 connected neighboring sub-regions. The basic method for determining these *k*-1 sub-regions is:

- (1) Calculate the *K*-1 nearest neighboring sub-regions of *Z_i* (which may not necessarily be adjacent to *Z_i*).
- (2) From these *K*-1 neighboring sub-regions, select *k*-1 (noting that 1 ≤ *k* ≤ *K*) while ensuring that they form a “connected” scanning window with *Z_i*.

For example, with *k* = 4, this means that the scanning window consists of *Z_i* and three neighboring sub-regions. In the worstcase, the selection of these three sub-regions can result in *C_{K-1}³* combinations. Therefore, theoretically, the FlexScan method may need to calculate *m* * 2^{*K*−1} windows in the worst case. Although the requirement for “connectivity” among sub-regions means that the actual number of scanning windows calculated will be lower, it remains substantial. This is why the FlexScan software typically recommends that the value of *K* should not exceed 30, with a default value of 15.

From a computational efficiency perspective, the SaTScan method demonstrates higher efficiency, while the FlexScan method is comparatively less efficient. Thus, enhancing the computational efficiency of the FlexScan method presents a valuable area for further research. Both classic methods can currently be implemented directly through software, allowing researchers to focus less on their computational efficiency. However, any optimizations or



improvements based on these methods must inevitably consider computational efficiency.

5.4 Result visualizations

This study employed a map visualization method to display the spatial distribution of disease clusters, using color brightness to indicate risk levels. However, differences in cluster distributions from the SaTScan and FleXScan methods are not intuitively discernible. Introducing interactive visualization tools would enhance the comparison of distribution differences among clusters with varying risk levels, improving the clarity and practicality of the analysis.

5.5 Limitations

Despite the in-depth comparison and analysis of the SaTScan and FleXScan methods, our study has several limitations:

- (1) Before applying FleXScan, obtaining a complete spatial adjacency matrix for the specific geographic area is crucial. Missing spatial adjacency relationships can bias clustering results, making preliminary topological checks essential to ensure the integrity of the adjacency matrix. In this study, we defined the spatial adjacency matrix using queen adjacency, which considers shared vertex connections. This may explain the irregular cluster shapes produced by FleXScan. Since queen adjacency only considers regions sharing a vertex as neighbors,

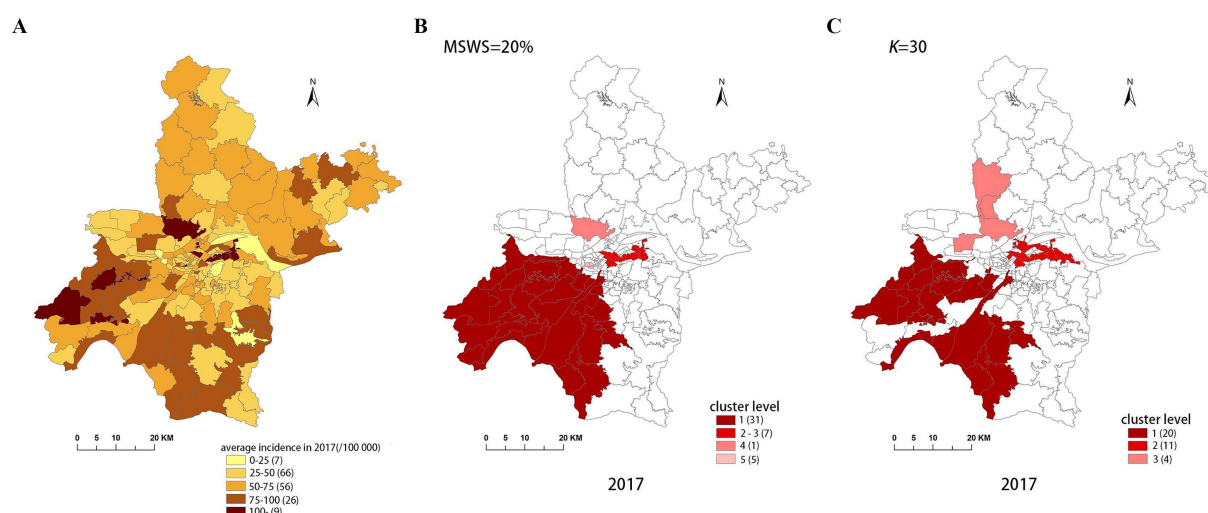


FIGURE 6

Spatial distribution map of tuberculosis in Wuhan in 2017. (A) The annual incidence map, (B) the SaTScan clusters result map, (C) the FleXScan clusters result map.

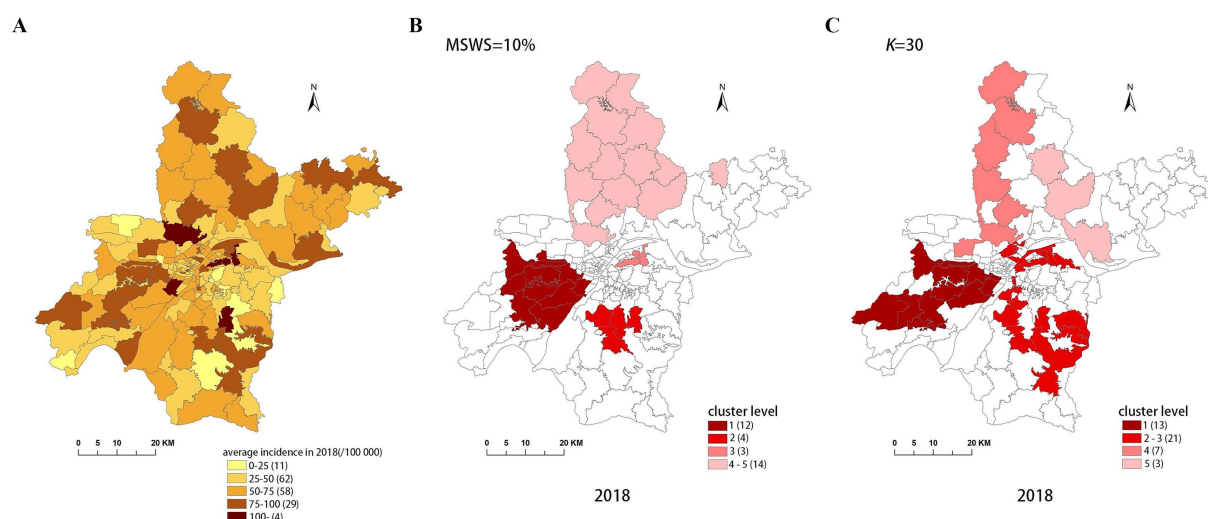


FIGURE 7

Spatial distribution map of tuberculosis in Wuhan in 2018. (A) The annual incidence map, (B) the SaTScan clusters result map, (C) the FleXScan clusters result map.

it may lack precision, especially for irregular or complex cluster shapes. To enhance the accuracy and interpretability of clustering results, future research could explore alternative adjacency definitions, such as rock adjacency (shared edge adjacency) or bishop adjacency (considering both shared vertices and edges).

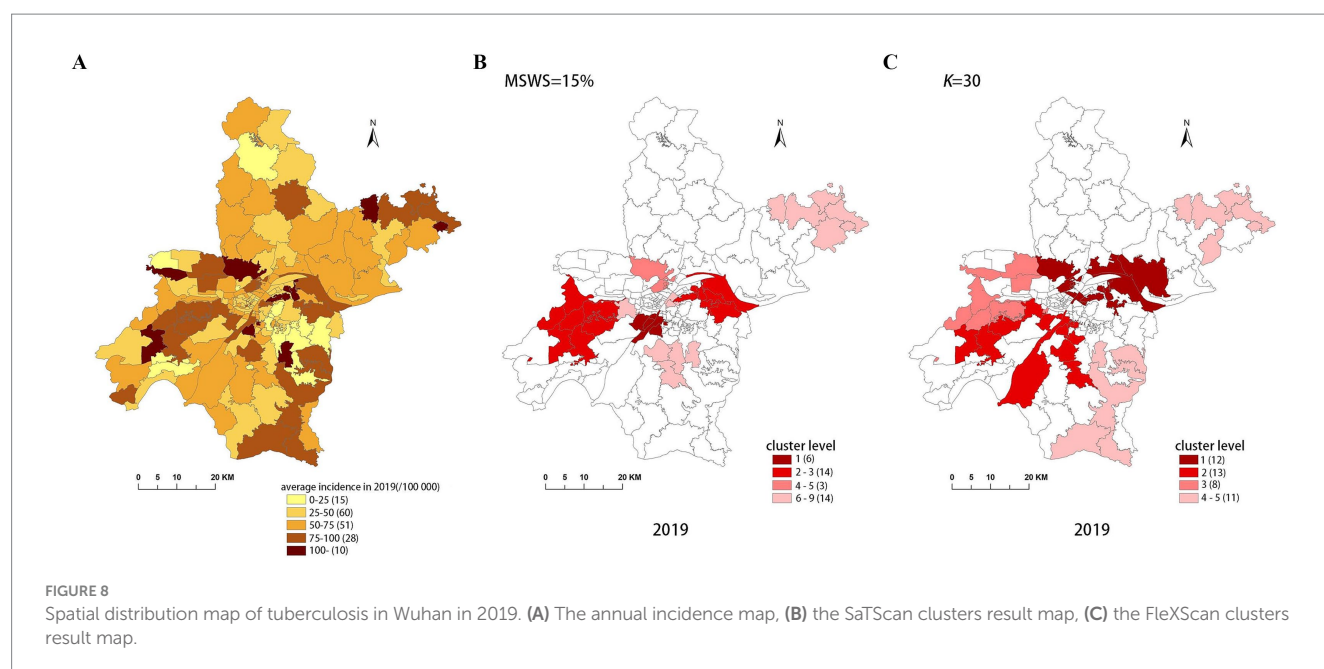
- (2) Although our study used multi-year disease data, the cross-sectional nature of the data limited the use of space-time scan statistics, restricting a full assessment of SaTScan and FleXScan's spatiotemporal sensitivity and precision. Additionally, our analysis was limited to Wuhan City and did not include data from broader scales like Hubei Province. Future research should assess these methods across different geographical scales, such as provincial or national levels, to

provide a more comprehensive evaluation and increase the generalizability of the results.

- (3) Our comparison focused on circular and flexible-shaped scan windows. However, elliptical scan windows, which can adjust their radii in two directions to better fit non-uniform spatial distributions (6), warrant further exploration and evaluation in future research.

6 Conclusion

In this study, we concentrated on determining the optimal parameter settings for circular and flexible-shaped scan statistics and their effects on clustering results. We also explored the characteristics



of these two methods and the influence of different scan window shapes on accuracy and reliability, offering valuable insights for future research.

While the FleXScan method may offer advantages in terms of result accuracy, disease spatial clustering patterns are highly complex. To mitigate the limitations of a single method, it is advisable to use a combination of methods to determine the final clustering results. Furthermore, the exploration of disease spatial clustering characteristics should be integrated with the study of influencing factors. Investigating clustering patterns not only aids in developing more effective prevention and control strategies but also reveals the factors and dynamics that influence disease occurrence and spread. By integrating these research methods, a more comprehensive understanding of disease transmission and its impact can be achieved, leading to more targeted and effective intervention measures.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

LW: Writing – original draft. XL: Writing – original draft. ZZ: Writing – review & editing. HY: Visualization, Writing – review &

editing. PL: Visualization, Writing – review & editing. YL: Visualization, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was supported by the National Natural Science Foundation of China (Grant No. 42201490), the Natural Science Foundation of Henan Province (Grant No. 212300410303), the Scientific Research Projects from Wuhan Municipal Health Commission (Grant No. WX23B39).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Patil GP, Taillie C. Geographic and network surveillance via scan statistics for critical area detection. *Stat Sci.* (2003) 18:457–65. doi: 10.1214/ss/1081443229
- Ozdenrol E, Williams BL, Kang SY, Magsumbol MS. Comparison of spatial scan statistic and spatial filtering in estimating low birth weight clusters. *Int J Health Geogr.* (2005) 4:19. doi: 10.1186/1476-072X-4-19
- Moon J, Kim M, Jung I. Optimizing the maximum reported cluster size for the multinomial-based spatial scan statistic. *Int J Health Geogr.* (2023) 22:30. doi: 10.1186/s12942-023-00353-4
- Kulldorff M. A spatial scan statistic. *Commun Stat-Theor M.* (1997) 26:1481–96. doi: 10.1080/03610929708831995

5. Kulldorff M, Huang L, Pickle L, Duczmal L. An elliptic spatial scan statistic. *Stat Med.* (2006) 25:3929–43. doi: 10.1002/sim.2490
6. Meysami M, French JP, Lipner EM. Flexible-elliptical spatial scan method. *Mathematics.* (2023) 11:3627. doi: 10.3390/math11173627
7. French JP, Meysami M, Hall LM, Weaver NE, Nguyen MC, Panter L. A comparison of spatial scan methods for cluster detection. *J Stat Comput Sim.* (2022) 92:3343–72. doi: 10.1080/00949655.2022.2065676
8. Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. *Int J Health Geogr.* (2005) 4:11. doi: 10.1186/1476-072X-4-11
9. Tango T. Spatial scan statistics can be dangerous. *Stat Methods Med Res.* (2021) 30:75–86. doi: 10.1177/0962280220930562
10. Tango T, Takahashi K. A flexible spatial scan statistic with a restricted likelihood ratio for detecting disease clusters. *Stat Med.* (2012) 31:4207–18. doi: 10.1002/sim.5478
11. Lee M, Jung I. Modified spatial scan statistics using a restricted likelihood ratio for ordinal outcome data. *Comput Stat Data An.* (2019) 133:28–39. doi: 10.1016/j.csda.2018.09.005
12. Speakman S, McFowland Iii E, Neill DB. Scalable detection of anomalous patterns with connectivity constraints. *J Comput Graph Stat.* (2015) 24:1014–33. doi: 10.1080/10618600.2014.960926
13. Cadena J, Chen F, Vullikanti A. Near-optimal and practical algorithms for graph scan statistics with connectivity constraints. *Acm T Knowl Discov D.* (2019) 13:1–33. doi: 10.1145/3309712
14. Ribeiro SHR, Costa MA. Optimal selection of the spatial scan parameters for cluster detection: a simulation study. *Spat Spatiotemporal Epidemiol.* (2012) 3:107–20. doi: 10.1016/j.sste.2012.04.004
15. Li M, Shi X, Li X, Ma W, He J, Liu T. Sensitivity of disease cluster detection to spatial scales: an analysis with the spatial scan statistic method. *Int J Geogr Inf Sci.* (2019) 33:2125–52. doi: 10.1080/13658816.2019.1616741
16. Wang W, Zhang T, Yin F, Xiao X, Chen S, Zhang X, et al. Using the maximum clustering heterogeneous set-proportion to select the maximum window size for the spatial scan statistic. *Sci Rep.* (2020) 10:4900. doi: 10.1038/s41598-020-61829-y
17. Han J, Zhu L, Kulldorff M, Hostovich S, Stinchcomb DG, Tatalovich Z, et al. Using Gini coefficient to determining optimal cluster reporting sizes for spatial scan statistics. *Int J Health Geogr.* (2016) 15:27. doi: 10.1186/s12942-016-0056-6
18. Kim J, Jung I. Evaluation of the Gini coefficient in spatial scan statistics for detecting irregularly shaped clusters. *PLoS One.* (2017) 12:e0170736. doi: 10.1371/journal.pone.0170736
19. Yoo H, Jung I. Optimizing the maximum reported cluster size for normal-based spatial scan statistics. *Commun Stat Appl Methods.* (2018) 25:373–83. doi: 10.29220/CSAM.2018.25.4.373
20. Xie Y, Shekhar S, Li Y. Statistically-robust clustering techniques for mapping spatial hotspots: a survey. *ACM Comput Surv.* (2022) 55:1–38. doi: 10.1145/3487893
21. Duczmal L, Kulldorff M, Huang L. Evaluation of spatial scan statistics for irregularly shaped clusters. *J Comput Graph Stat.* (2006) 15:428–42. doi: 10.1198/106186006X112396
22. Duczmal L, Ferreira S J, da Cunha Santos M. A topological correction for a graph based spatial scan cluster detection algorithm. (2022) Available at: https://www.researchgate.net/profile/Sabino-Ferreira/publication/229047815_A_topological_correction_for_a_graph_based_spatial_scan_cluster_detection_algorithm/links/00b495385ef3d0034c000000/A-topological-correction-for-a-graph-based-spatial-scan-cluster-detection-algorithm.pdf (Accessed April 15, 2024).
23. Cançado ALF, Duarte AR, Duczmal LH, Ferreira SJ, Fonseca CM, Gontijo ECDM. Penalized likelihood and multi-objective spatial scans for the detection and inference of irregular clusters. *Int J Health Geogr.* (2010) 9:55. doi: 10.1186/1476-072X-9-55

Frontiers in Medicine

Translating medical research and innovation into
improved patient care

A multidisciplinary journal which advances our
medical knowledge. It supports the translation
of scientific advances into new therapies and
diagnostic tools that will improve patient care.

Discover the latest Research Topics

[See more →](#)

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact



Frontiers in Medicine

