

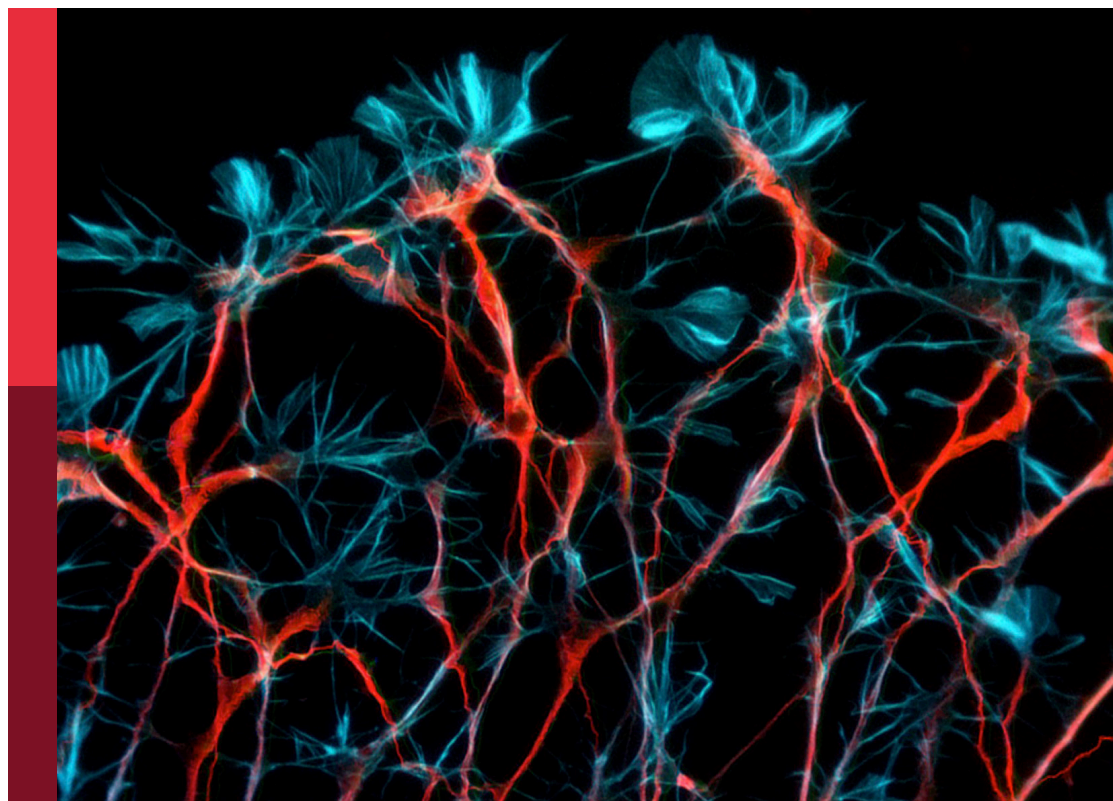
Advancements in smart diagnostics for understanding neurological behaviors and biosensing applications

Edited by

Saad Arif, Muhammad Zia Ur Rehman and
Zohaib Mushtaq

Published in

Frontiers in Computational Neuroscience
Frontiers in Sensors



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 978-2-8325-6917-7
DOI 10.3389/978-2-8325-6917-7

Generative AI statement

Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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

Advancements in smart diagnostics for understanding neurological behaviors and biosensing applications

Topic editors

Saad Arif — King Faisal University, Saudi Arabia

Muhammad Zia Ur Rehman — Riphah International University, Pakistan

Zohaib Mushtaq — University of Sargodha, Pakistan

Citation

Arif, S., Rehman, M. Z. U., Mushtaq, Z., eds. (2025). *Advancements in smart diagnostics for understanding neurological behaviors and biosensing applications*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6917-7

Table of contents

- 05 **Editorial: Advancements in smart diagnostics for understanding neurological behaviors and biosensing applications**
Saad Arif, Muhammad Zia ur Rehman and Zohaib Mushtaq
- 08 **Artificial neural network models: implementation of functional near-infrared spectroscopy-based spontaneous lie detection in an interactive scenario**
M. Raheel Bhutta, Muhammad Umair Ali, Amad Zafar, Kwang Su Kim, Jong Hyuk Byun and Seung Won Lee
- 21 **BrainNet: an automated approach for brain stress prediction utilizing electrodermal activity signal with XLNet model**
Liao Xuanzhi, Abeer Hakeem, Linda Mohaisen, Muhammad Umer, Muhammad Attique Khan, Shrooq Alsenan, Shtwai Alsubai and Nisreen Innab
- 32 **Multimodal sleep staging network based on obstructive sleep apnea**
Jingxin Fan, Mingfu Zhao, Li Huang, Bin Tang, Lurui Wang, Zhong He and Xiaoling Peng
- 47 **Multimodal consumer choice prediction using EEG signals and eye tracking**
Syed Muhammad Usman, Shehzad Khalid, Aimen Tanveer, Ali Shariq Imran and Muhammad Zubair
- 62 **Automated karyogram analysis for early detection of genetic and neurodegenerative disorders: a hybrid machine learning approach**
Sumaira Tabassum, M. Jawad Khan, Javaid Iqbal, Asim Waris and M. Adeel Ijaz
- 79 **EEG electrode setup optimization using feature extraction techniques for neonatal sleep state classification**
Hafza Ayesha Siddiqa, Muhammad Farrukh Qureshi, Arsalan Khurshid, Yan Xu, Laishuan Wang, Saadullah Farooq Abbasi, Chen Chen and Wei Chen
- 96 **Mamba with split-based pyramidal convolution and Kolmogorov-Arnold network-channel-spatial attention for electroencephalogram classification**
Zhe Li
- 108 **Decentralized EEG-based detection of major depressive disorder via transformer architectures and split learning**
Muhammad Umair, Jawad Ahmad, Nada Alasbali, Oumaima Saidani, Muhammad Hanif, Aizaz Ahmad Khattak and Muhammad Shahbaz Khan

- 123 **Enhancing medical image privacy in IoT with bit-plane level encryption using chaotic map**
Fatima Asiri, Wajdan Al Malwi, Tamara Zhukabayeva, Ibtehal Nafea, Abdullah Aziz, Nadhmi A. Gazem and Abdullah Qayyum
- 138 **Classification of power grip and precision grip in children using an EIT-based tactile sensor**
Ryunosuke Asahi, Shumpei Toriyama, Yasuyo Minagawa, Shunsuke Yoshimoto and Hiroki Sato



OPEN ACCESS

EDITED AND REVIEWED BY
Si Wu,
Peking University, China

*CORRESPONDENCE
Saad Arif
✉ sarif@kfu.edu.sa

RECEIVED 26 August 2025
ACCEPTED 03 September 2025
PUBLISHED 16 September 2025

CITATION
Arif S, Rehman MZu and Mushtaq Z (2025)
Editorial: Advancements in smart diagnostics
for understanding neurological behaviors and
biosensing applications.
Front. Comput. Neurosci. 19:1693327.
doi: 10.3389/fncom.2025.1693327

COPYRIGHT
© 2025 Arif, Rehman and Mushtaq. 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.

Editorial: Advancements in smart diagnostics for understanding neurological behaviors and biosensing applications

Saad Arif^{1*}, Muhammad Zia ur Rehman² and Zohaib Mushtaq³

¹Department of Mechanical Engineering, College of Engineering, King Faisal University, Al Ahsa, Saudi Arabia, ²Department of Biomedical Engineering, Riphah International University, Islamabad, Pakistan, ³Department of Electrical, Electronics and Computer Systems, College of Engineering and Technology, University of Sargodha, Sargodha, Pakistan

KEYWORDS

biosensing techniques, smart diagnostics, human-computer interaction, brain-computer interface, intelligent healthcare, physiological monitoring, neurobehavior analysis, machine and deep learning

Editorial on the Research Topic

Advancements in smart diagnostics for understanding neurological behaviors and biosensing applications

Overview

An understanding of neurological behaviors can be well-established by bringing together machine learning (ML) and biosensing techniques. This combination is promising for accelerating smart diagnostics and human-computer interactions (HCI), with the following three aspects necessary to create a real-world impact of this combination. Multimodal and multiscale learning to capture the richness of physiological signals is of primary importance, followed by explainability and clinical acceptability that link model evidence to neurophysiology. Designing the trustworthy pipelines that safeguard privacy and support deployment in clinics, homes, and wearables is of practical significance in this endeavor.

In this pursuit, 10 contributions were submitted to this Research Topic, covering biosensing modalities including electroencephalography (EEG), electrooculography (EOG), electrodermal activity (EDA), functional near-infrared spectroscopy (fNIRS), medical imaging, and electrical impedance tomography (EIT)-based tactile sensing. This Research Topic effectively describes how contemporary learning paradigms such as transformers and state space models, hybrid unsupervised-supervised pipelines, and privacy-preserving training, translate signals into actionable insights while respecting the constraints of clinical and everyday settings.

A foundational theme is the decoding of affective and behavioral states from biosignals for continuous monitoring and neuroadaptive interfaces. [Xuanzhi et al.](#) modeled stress from EDA using attention-based sequence learning, showing that temporal context in peripheral signals can support robust continuous assessment, with accuracies reaching above 95% on public datasets. Another study by [Usman et al.](#) integrated EEG with eye tracking (ET) to predict real-world choices, illustrating how multimodal fusion and

ensemble strategies can extract complementary neural and ocular markers of preference in ecologically valid scenarios. Their approach achieved around 84% accuracy with high precision for positive preferences. Extending behavioral inference to social cognition, [Bhutta et al.](#) employed frontal fNIRS in an interactive setting to distinguish deception from truth-telling using deep neural networks, attaining approximately 88–90% accuracy and pointing to the feasibility of decoding complex, spontaneous behaviors beyond controlled paradigms.

Sleep health emerged as a second thematic pillar, benefiting from multimodal, temporally aware modeling across distinct populations. A contribution on sleep staging by [Fan et al.](#) transformed EEG and EOG into time-frequency sequences, coupled long-range temporal modeling with multiscale feature extraction, and integrated modalities to mitigate the heterogeneity introduced by obstructive sleep apnea (OSA). This design demonstrates broader applicability beyond healthy cohorts and enhances interpretability for clinical workflows, with performance at approximately 80% in OSA cohorts and improvements over competitive baselines on public datasets. Complementing this system's view, a neonatal study by [Siddiqua et al.](#) identified promising electrode configurations and informative signal features that sustain accurate sleep state classification. Notably, a single central channel maintained an accuracy of approximately 81%, and compact left hemisphere montages slightly outperformed right hemisphere channels. The study presented a practical sleep monitoring strategy that prioritizes comfort, safety, and computational efficiency in newborn care.

Methodological innovations in EEG decoding were highlighted by work that blends efficient sequence modeling with targeted attention and multiscale feature design. A study by [Li](#) advanced a compact state space architecture paired with pyramidal convolutions and channel-spatial attention to improve EEG classification for brain-computer interfaces (BCI). The findings underscore the potential of latency-conscious designs for real-time use on a standard dataset while achieving approximately 97% performance with strong class-wise balance. In parallel, a clinical study by [Umair et al.](#) on major depressive disorder (MDD) detection demonstrated that ensembles leveraging transformer representations can deliver high-accuracy classification from EEG data while operating within a decentralized, split-learning framework that keeps data local across nodes. The approach maintained over 95% accuracy across clients, and reached approximately 99% accuracy in centralized settings, aligning with institutional privacy requirements and offering a viable path to collaborative, ensemble learning without compromising data security.

Beyond electrophysiology, imaging-centric contributions emphasized both diagnostic capability and pipeline security. A large-scale study on karyogram analysis by [Tabassum et al.](#) proposed a hybrid approach that pre-trained the proposed classifier on unlabeled images and fine-tuned it to detect structural anomalies, complemented by techniques that localized abnormal regions. This design addresses the pervasive challenge of rarely labeled anomalies and demonstrates near state-of-the-art accuracy (approximately 99%), supporting the early screening of chromosome-related neurodegenerative disorders with

neurological impact. Complementing analytics with protection, [Asiri et al.](#) developed a lightweight bit-plane encryption scheme for medical images tailored to the internet of things (IoT) and edge devices. By leveraging chaotic map-based shuffling and diffusion, the method achieved high entropy (greater than 7.98), low or negative inter-pixel correlations, a vast key space, and robustness under occlusion. This study presented practical safeguards for data in transit and at rest in resource-constrained environments.

Finally, novel sensing modalities extended the scope of smart diagnostics to child-centered interaction and rehabilitation. [Asahi et al.](#) employed EIT-based tactile sensing, which presents a safe, integrated device that classifies children's power vs. precision grips using features derived from voltage patterns and tomographic reconstructions. In a pediatric cohort, accuracies exceeded 85%, illustrating how contact-rich sensing can enable the quantitative monitoring of developing motor skills and inform the design of pediatric HCI.

Several cross-disciplinary and cutting-edge research focuses were explored in this Research Topic. First, multimodal fusion and multiscale representations consistently improve robustness to artifacts and population heterogeneity when EEG, EOG, and ET modalities are hybridized or link time-frequency transforms with attention and dilated convolutions. Second, temporal sequence modeling via transformers and state-space formulations captures long-range dependencies that static models overlook, enabling more reliable decoding of stress, behavior, and sleep dynamics. Third, data-efficiency strategies, including unsupervised pretraining, synthetic data augmentation for class balancing with the synthetic minority over-sampling technique (SMOTE), and targeted feature engineering, addressed unlabeled data and imbalanced class distribution issues, which are common in clinical datasets and rare pathology scenarios. Fourth, interpretability is increasingly embedded through attention mechanisms, multiscale modules, and explicit localization, aligning model outputs with physiological prospects and aiding clinical acceptance. Finally, trustworthy deployment is advanced by privacy-preserving learning that limits data movement, lightweight encryption suited to edge devices, and practical design choices such as electrode optimization and a compact scheme that supports real-time, on-device feasibility.

Conclusion

The presented contributions validate a promising shift from individual performance gains to cutting-edge integrated pipelines that are multimodal, interpretable, cybersecure, and privacy-preserving by design. They demonstrate that modern sequence models and multiscale representations can decode the complex neurobehavioral characteristics of active and passive brain activities (e.g., stress, consumer choice, deception, and sleep dynamics). It has been shown that data-efficient training enables intricate neuronal signatures to be captured efficiently (e.g., neonatal EEG and rare chromosomal anomalies). The contributions established that privacy-preserving analytics and lightweight cryptography are functional aspects for deployment in clinics and daily life (e.g., split learning sustaining over 95% accuracy, edge encryption

with high entropy and resilience). To sustain this momentum, the research community should prioritize prospective and a wide range of demographic validation to establish generalization and to adopt shared data standards and benchmarks to strengthen reproducibility. Collaborative explainability should be adopted with clinicians, patients, and end users to support informed decisions. Decentralized learning and secure edge computing should continue to excel for equitable access. Convergence of ML and biosensing approaches, along with these research commitments, will continue to deliver reliable neurotechnology for diagnosis, monitoring, and HCI across the lifespan.

Author contributions

SA: Supervision, Writing – original draft, Conceptualization. MR: Writing – review & editing. ZM: Writing – review & editing.

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.

The author(s) declared that they were an editorial board member of Frontiers at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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

Hava T. Siegelmann,
The State University of New Jersey,
United States

REVIEWED BY

Farzan Majeed Noori,
University of Oslo, Norway
Nauman Khalid Qureshi,
ETH Zürich, Switzerland

*CORRESPONDENCE

Amad Zafar
✉ amad@sejong.ac.kr
Jong Hyuk Byun
✉ maticax@pusan.ac.kr
Seung Won Lee
✉ lsw2920@gmail.com

[†]These authors have contributed equally to this work and first authorship

RECEIVED 04 September 2023

ACCEPTED 02 November 2023

PUBLISHED 24 January 2024

CITATION

Bhutta MR, Ali MU, Zafar A, Kim KS,
Byun JH and Lee SW (2024) Artificial neural
network models: implementation of functional
near-infrared spectroscopy-based
spontaneous lie detection in an interactive
scenario.
Front. Comput. Neurosci. 17:1286664.
doi: 10.3389/fncom.2023.1286664

COPYRIGHT

© 2024 Bhutta, Ali, Zafar, Kim, Byun and Lee.
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.

Artificial neural network models: implementation of functional near-infrared spectroscopy-based spontaneous lie detection in an interactive scenario

M. Raheel Bhutta^{1†}, Muhammad Umair Ali^{2†}, Amad Zafar^{2*},
Kwang Su Kim^{3,4}, Jong Hyuk Byun^{5,6*} and Seung Won Lee^{7*}

¹Department of Electrical and Computer Engineering, University of UTAH Asia Campus, Incheon, Republic of Korea, ²Department of Intelligent Mechatronics Engineering, Sejong University, Seoul, Republic of Korea, ³Department of Scientific Computing, Pukyong National University, Busan, Republic of Korea, ⁴Interdisciplinary Biology Laboratory (iBLab), Division of Biological Science, Graduate School of Science, Nagoya University, Nagoya, Japan, ⁵Department of Mathematics and Institute of Mathematical Science, Pusan National University, Busan, Republic of Korea, ⁶Finace Fishery Manufacture Industrial Mathematics Center on BigData, Pusan National University, Busan, Republic of Korea, ⁷Department of Precision Medicine, Sungkyunkwan University School of Medicine, Suwon, Republic of Korea

Deception is an inevitable occurrence in daily life. Various methods have been used to understand the mechanisms underlying brain deception. Moreover, numerous efforts have been undertaken to detect deception and truth-telling. Functional near-infrared spectroscopy (fNIRS) has great potential for neurological applications compared with other state-of-the-art methods. Therefore, an fNIRS-based spontaneous lie detection model was used in the present study. We interviewed 10 healthy subjects to identify deception using the fNIRS system. A card game frequently referred to as a bluff or cheat was introduced. This game was selected because its rules are ideal for testing our hypotheses. The optical probe of the fNIRS was placed on the subject's forehead, and we acquired optical density signals, which were then converted into oxy-hemoglobin and deoxy-hemoglobin signals using the Modified Beer–Lambert law. The oxy-hemoglobin signal was preprocessed to eliminate noise. In this study, we proposed three artificial neural networks inspired by deep learning models, including AlexNet, ResNet, and GoogleNet, to classify deception and truth-telling. The proposed models achieved accuracies of 88.5%, 88.0%, and 90.0%, respectively. These proposed models were compared with other classification models, including k-nearest neighbor, linear support vector machines (SVM), quadratic SVM, cubic SVM, simple decision trees, and complex decision trees. These comparisons showed that the proposed models performed better than the other state-of-the-art methods.

KEYWORDS

spontaneous lie detection, deception, deep learning algorithm, functional near-infrared spectroscopy (fNIRS), classification

1 Introduction

Deception is an intrinsic and unavoidable facet of our society, manifesting itself in everyday life. It is unsurprising for a person to encounter or be involved in multiple deceptive situations within a single day. Failure to identify deception has serious consequences for the victim. To avoid being deceived, people have begun to study the behavioral and physiological changes exhibited by deceivers. Hence, this study aimed to detect the differences between hemodynamic signals during spontaneous deception and classify between truth and lie during an interactive game paradigm.

In earlier times, people identified deceivers based on the deceiver's personality or their own personal experiences (Freud and Strachey, 1962; Zuckerman et al., 1981b; Kleinmuntz and Szucko, 1984; Peterman and Anderson, 1999). Additionally, during earlier times, people often relied on myths based on religious norms to identify a person who was being untruthful (Trovillo, 1938). Advancements in scientific methods and new equipment, including polygraphs, have enabled us to better understand the cues of deception that are beyond the scope of religious beliefs, personal experience, and stereotypes (Brett et al., 1986; Varisai Mohamed et al., 2006). These physiological measures have revealed many new findings that provide the basis for numerous theories, such as the non-verbal leakage theory (Ekman et al., 1969), four-factor theory (Zuckerman et al., 1981a), and interpersonal deception theory (Buller and Burgoon, 1996). These theories have helped us understand why these cues of deception manifest in humans when attempting to deceive someone (Bond et al., 2014). Most of these theories agree that the intent and process of deception invoke changes in the deceiver's behavior that result from changes in the person's state of mind.

Many researchers have investigated different neurophysiological signals to identify changes in an individual's mental state while they are attempting to deceive. One such technique is Electroencephalography (EEG), which records event-related potentials (ERPs) from the scalp of the brain (Abootalebi et al., 2009; Meijer et al., 2013). ERPs are mainly used to test knowledge of crime details that are only known to the criminals involved (Farwell and Donchin, 1991). This type of test is commonly known as the guilty knowledge test or concealed information test (Furedy and Ben-Shakhar, 1991; Elaad and BEN-SHAKHAR, 1997; Kong et al., 2012). EEG has excellent temporal resolution, enabling rapid detection of brain signals (Turnip et al., 2011; Chen et al., 2023), but exhibits poor spatial resolution, which cannot confine the brain area associated with the deception process.

Functional magnetic resonance imaging (fMRI) is another technique widely used to detect brain areas activated during deception. fMRI offers a substantial advantage in terms of high spatial resolution when compared to EEG (Spence et al., 2004). It can effectively localize changes in regional blood flow (Farah et al., 2014) and hence provides a comprehensive review of fMRI-based deception decoding. Because of the high cost of scanners and their bulky size, the use of fMRI is very limited in day-to-day human routines. Moreover, fMRI is highly sensitive to motion artifacts. Therefore, researchers have embarked on exploring an alternative brain imaging technique: functional near-infrared spectroscopy (fNIRS).

Using fNIRS, brain activity is measured through hemodynamic responses associated with neuronal behavior (Kamran and Hong, 2013; Santosa et al., 2013; Khan et al., 2014; Ruotsalo et al., 2023). The fNIRS can provide topographic (Obrig and Villringer, 2003; Wolf et al., 2007; Hu et al., 2011; Li et al., 2018) and tomographic brain images (Bluestone

et al., 2001; Boas et al., 2004). Oxy-hemoglobin (HbO), deoxy-hemoglobin (HbR), and water are significant light absorbers, whereas skin, tissue, and bone are mainly transparent to near-infrared light within an optical window of 650–1,000 nm. Compared with EEG and fMRI, fNIRS offers a superior tradeoff between temporal and spatial resolutions. In one study (Irani et al., 2007) compared the features of fNIRS and fMRI and reported that fNIRS has excellent potential for psychotic and neurological applications because of its portability, simplicity, and insensitivity to motion artifacts compared to fMRI. fNIRS also has several advantages over other brain imaging techniques; it can be designed in a compact and portable form, is very cost-effective (Muehlemann et al., 2008; Bhutta et al., 2014; Toglia et al., 2022), and can be used in diverse fields such as neuroscience, brain-computer interfaces (Naseer and Hong, 2013a,b), and rehabilitation.

2 Literature review

Limited research has been conducted in the field of fNIRS-based deception decoding (Tian et al., 2009; Hu et al., 2012; Ding et al., 2013, 2014; Bhutta et al., 2015; Emberson et al., 2017; Quaresima and Ferrari, 2019). To detect deception, one study (Hu et al., 2012) employed a mock crime paradigm. Because individuals were instructed to provide deceptive or truthful responses at specified times and locations, this research, which was based on the concealed information test, did not incorporate a spontaneous paradigm. The first study to use fNIRS to identify the neural correlates of spontaneous deception was conducted by Ding et al. (2013). These aforementioned studies on fNIRS-based deception decoding have exclusively investigated cases of deceptions where the perpetrator lies to an unsuspecting victim; this type of deception occurs more frequently in casual social interactions. In contrast, there are also situations in which the perpetrator deliberately misleads the victim, even though both parties are fully aware of the attempt at deception. This type of circumstance is typically referred to as reverse psychology, and it frequently occurs in highly competitive settings, such as diplomatic meetings, political debates and elections, sports, card games (including gambling), and other various scenarios. In this scenario, the individual employing reverse psychology can deceive the victim not only by uttering a false statement but also by making a truthful remark. The deceiver may choose to speak the truth, knowing that the victim is aware of the deceptive intention, yet the victim interprets it as a lie, thus believing the contrary. Consequently, speaking the truth serves the deceiver's purpose of misleading the victim.

Deep learning classifiers have been widely used recently. A deep neural network (DNN) is composed of multiple layers of nonlinear processing modules called neurons (Schmidhuber, 2015; Huve et al., 2018). These fully connected or semi-connected neurons receive inputs from previous consecutive neurons. DNN can achieve superior classification performance in comparison to linear classifiers, such as linear discernment analysis (LDA), support vector machine (SVM), and others when applied to signals (language and speech processing) or images (Collobert and Weston, 2008; Krizhevsky et al., 2012; Bianchini and Scarselli, 2014; Simonyan and Zisserman, 2014). Hence, DNN classifiers are also gaining attention in the biomedical field (Hudson and Cohen, 2000; Cireşan et al., 2013; Ronneberger et al., 2015).

Only a few studies have employed DNN for classification. Abibullaev et al. (2011) investigated the performance of a DNN in a four-class classification experiment and reported a maximum accuracy

of 94%. Yi et al. (2013) used a DNN to classify left and right motor imagery with an average classification accuracy of 84%. Hennrich et al. (2015) reported a similar classification performance of DNN compared to that of other classifiers (such as LDA and SVM) in a three-mental task experiment. To the best of our knowledge, no previous study has used a DNN for spontaneous deception decoding using fNIRS.

In this study, we hypothesized that, in the real world, a deceiver can deceive another person not only by telling a lie but also by telling the truth. Therefore, the objectives of this study were to:

- compare the differences between the hemodynamic responses produced by spontaneous lying and stating the truth,
- classify between the lie and truth for an interactive game paradigm,
- develop three deep ANN models for classifying responses, and
- compare the performance of the proposed deep ANN with other classifiers, such as LDA and SVM, in a spontaneous deception decoding paradigm.

According to these findings, the fNIRS system can accurately identify changes in HbO signals during spontaneous lies and truths.

3 Materials and methods

3.1 Subjects

Ten healthy male individuals (mean age 30.8 ± 3.68) participated in the experiments. Each patient had normal or corrected-to-normal eyesight. Of the 10 subjects, nine were right-handed. None of the subjects had any history of mental or neurological illness. The card game was known to all subjects. Informed consent was obtained from all subjects, and the experiments were performed in accordance with the latest Declaration of Helsinki. The framework proposed in this study is illustrated in Figure 1. The framework is divided into two blocks: a training block (blue dotted lines) and a testing block (green dotted lines). The training block was used to train the neural network

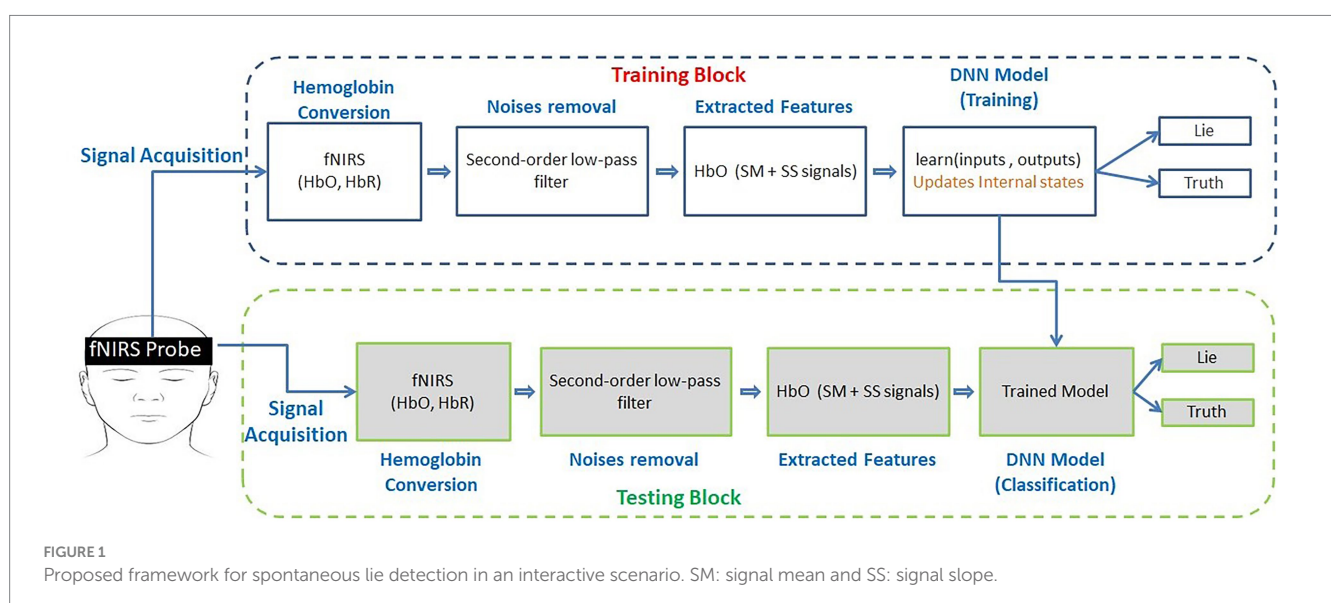
models on the given data, whereas the testing block was used to classify the data into truth and lie classes based on the model trained in the training block. Information on the individual blocks is provided in the respective chapters of the article.

3.2 Experimental procedure

The subjects were seated comfortably in front of a second person (referred to as the victim). The subject and victims underwent three practice sessions, and a brief explanation of the experiment before the experiment was provided to ensure that they fully understood the guidelines.

A well-selected experimental paradigm was used in this study. The experimental paradigm was a card game known as bluff or cheat. The bluff game was chosen because the rules of the game are ideal for testing our hypotheses. Our objective was to distinguish between deceptive actions when the subject is speaking the truth and when they are intentionally deceiving the victim with a falsehood.

The game rules are straightforward. The subject received 20 randomly selected cards, with the consideration that a minimum of six of these cards had no corresponding matches. Therefore, the subject had to lie at least four or five times in order to get rid of those cards. The subject had to play out all of their cards without revealing any signs of bluffing. The subjects had 1 min to carefully organize all their cards prior to starting the experiment. The duration of each experiment was approximately 10 min, with each experiment having a maximum of 20 sessions, each lasting approximately 30 s. In each session, the first 15 s were allotted for card arrangement. The subject had to lie to the victim in the next 5 s (called “claim time”) by laying his cards face down on the table and declaring what kind of cards they are (for instance, “three sevens”). Depending on his claim, the subject could select any number of cards between two and four. However, this assertion may or may not be correct. The victim then had 10 s to react to the subject’s assertion (response time). If the victim believed that the subject is telling the truth, they could choose to pass, removing the pile from the table. However, if the victim suspected that the subject had lied in their claim,



they had the option to flip the cards face up. If the subject lied, the pile was returned to the subject. However, if the subject was truthful, the pile was removed from the table, and the next session commenced. The game continued for 20 sessions. A prize of 10,000 Korean Won was to be awarded to the subject if they managed to play all their cards within 20 sessions; however, if they were to do so, they would not receive the prize money. There were 12 total subjects in this trial. Two respondents' data were excluded from the analysis as they consistently spoke the truth at the beginning of trials and only lied towards the end, rendering their responses predictable. Eight out of ten subjects were able to play all of their cards. One administrator continuously monitored the experiment and documented the trials in which the subject deceived the victim.

3.3 Data acquisition

A lab-built multichannel continuous-wave imaging system captured the brain signals (Bhutta and Hong, 2013). The optical probe of the fNIRS system was positioned on the forehead of the subject such that the FP1 and FP2 locations were in the middle of the probe, as shown in Figure 2. To connect the flexible probe and ensure excellent contact between the its emitters and detectors and the subject's scalp, hair was brushed backward. Self-adhesive bandages were used to secure the probe to the subject's head. The emitters and detectors were systematically positioned within a $4.3 \times 13 \text{ cm}^2$ area according to a source-detector distance of approximately 3 cm. A

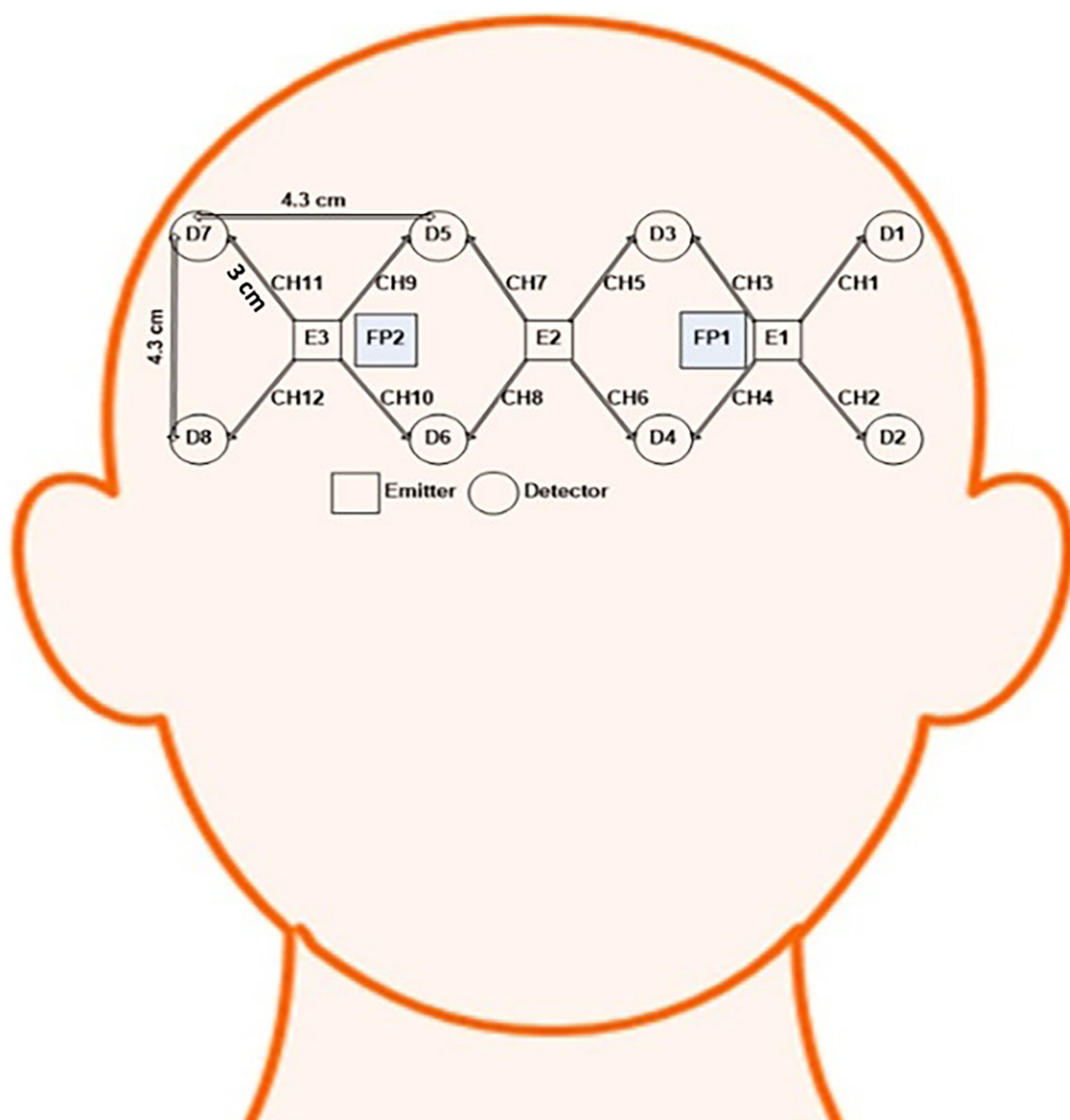


FIGURE 2
Optode placement and channel configuration.

sampling rate of approximately 3.8 Hz was used to collect the data. A Velcro band was used to hold the probe at the appropriate location throughout the experiment.

3.4 Data processing

MATLAB (MathWorks, United States) was used to import and further analyze the signals from the fNIRS equipment offline. The data were stored in a host computer text file as digitized raw intensity values from the fNIRS system. The hemoglobin conversion block of the framework was used to convert the intensity values to concentration changes of HbO and HbR using the Modified Beer–Lambert law (Bhutta et al., 2015). The change in optical density (ΔOD) was calculated using these raw intensity values at each discrete time k as:

$$\Delta OD(k; \lambda) = \ln \frac{I_{out}(\lambda)}{I_{in}(k; \lambda)} = ld(\lambda) \Delta \mu_a(k; \lambda) \quad (1)$$

Where I_{out} is the intensity of detected light; I_{in} , intensity of incident light; d , differential path length factor; l , distance between the emitter and detector; and $\Delta \mu_a$, absorption change of the tissue. The changes of HbO (Δc_{HbO}) and HbR (Δc_{HbR}) were measured using the modified Beer–Lambert Law as (Bhutta et al., 2015):

$$\begin{bmatrix} \Delta c_{HbO} \\ \Delta c_{HbR} \end{bmatrix} (k) = \begin{bmatrix} ld^{\lambda_1} \alpha_{HbO}^{\lambda_1} & ld^{\lambda_1} \alpha_{HbR}^{\lambda_1} \\ ld^{\lambda_2} \alpha_{HbO}^{\lambda_2} & ld^{\lambda_2} \alpha_{HbR}^{\lambda_2} \end{bmatrix}^{-1} \times \begin{bmatrix} \Delta OD_1(k; \lambda) \\ \Delta OD_2(k; \lambda) \end{bmatrix} \quad (2)$$

with $\lambda_1 = 640$ nm, $\lambda_2 = 910$ nm, $d^{\lambda_1} = 6.63$, and $d^{\lambda_2} = 2.765$, according to the values for the wavelength-dependent absorption coefficients α_{HbO} , α_{HbR} . fNIRS, while detecting the hemodynamic responses, picks up the physiological noise of respiration, pulse, and low-frequency drift fluctuations. A second-order low-pass filter with a cutoff frequency of 0.15 Hz was used to eliminate such noises (Hu et al., 2012; Bhutta et al., 2015). The HbO was considered for further analysis in this study because it is a more sensitive and reliable activity indicator than HbR (Hoshi, 2003, 2007).

3.5 Classification

Once the data were preprocessed, classification was performed on the $\Delta c_{HbO}(k)$ signals. We conducted this classification to distinguish between lie and truth responses based on the features extracted from fNIRS signals. The features selected in this study were the signal mean (SM) and signal slope (SS) of the HbO signal during the 5-s claim period of the stimulus. We used this claim period because it is the actual time at which the subject attempted to deceive the victim by either telling the truth or lying. The average HbO signal for each subject was obtained by averaging all 12 channels of the corresponding subject. SM and SS values over a 5-s window can yield better results in binary classification (Khan et al., 2014; Bhutta et al., 2015).

In this study, we performed the classification using various classifiers categorized into linear and nonlinear categories. LDA and

SVM are the main linear classifiers, whereas the ANN is a nonlinear classifier. Both the LDA and SVM algorithms classify different classes of data based on hyperplanes. In LDA, a separating hyperplane is generated to minimize the interclass variance and maximize the distance between the class means (Lotte et al., 2007). For the SVM classifier, a separating hyperplane is designed such that the distance between the hyperplane and the nearest training point(s) is maximized (Naseer et al., 2014).

Mainstream machine learning techniques can be categorized as linear or nonlinear classifiers. Linear classifiers classify a sample based on the value of the linear combination of its features. For example, assume that we have an input feature vector x . A linear classifier then constructs a function that directly assigns the input vector x to a specific class:

$$f(x) = \begin{cases} 1 & \text{if } x > \text{threshold,} \\ -1 & \text{otherwise} \end{cases} \quad (3)$$

A linear SVM is a linear classifier that makes decisions according to a linear hyperplane capable of effectively segregating data. SVM finds an optimal hyperplane by maximizing the margin, which is the minimum distance between the hyperplane and any of the data samples. Such classifiers perform well when the problem is linearly separable. However, if the data are not linearly separable, they will have poor generalization ability. In this case, we could map the input vector onto a higher-dimensional space using the kernel function K and find the separating hyperplane in that particular dimension. Quadratic SVM and Cubic SVM are examples of kernelized versions of SVM that utilize second and third-degree polynomial kernels.

$$K(x_i, x_j) = (x_i^T x_j + 1)^p \quad (4)$$

In the machine learning literature, several other algorithms handle nonlinear cases using a completely different computational approach; one of the simplest algorithms is the K-Nearest Neighbor (KNN). The main idea of this algorithm is that, for a new instance to be classified, the algorithm searches for the K-nearest points in the feature space and assigns it to the class that prevails among its neighbors. Similarly, the decision tree constructs a classification model with a tree-like structure. It partitions a feature space into smaller regions containing homogenous instances and simultaneously incrementally constructs an associated decision tree. The partitions of the feature space are usually based on criteria such as the Gini impurity, information gain, or distance measure.

3.6 Proposed artificial neural networks (ANN) models

In recent years, artificial neural networks have flourished in the machine learning and pattern recognition domains. They consist of many interconnected processing units, called neurons. The outputs of the hidden layer neurons are transmitted to the inputs of the next hidden layer within the network (Ullah et al., 2020). Thus, they communicate with each other by emitting signals over

numerous weighted connections. During training, each neuron can update its weight, allowing the network to learn hidden patterns from the data. In this study, we designed three ANN architectures (M1, M2, and M3) to conduct experiments on our dataset. These structures were designed based on ideas from state-of-the-art convolutional neural network models, including AlexNet, ResNet, and GoogleNet. The numbers of input and output nodes and hidden layers of these neural networks are the same; however, the number of nodes in each hidden layer varies. The first two layers of M1 contain 10 neurons; the subsequent two hidden layers consist of eight and five neurons, respectively; and finally, the prediction layer contains two SoftMax classifiers. The M2 topology is similar to that of M1; however, we introduced two pairs of hidden layers with the same number of neurons in this structure. The first two layers had eight neurons, and the next pair had layers containing four neurons. We designed a third neural network architecture that differed from the aforementioned architecture. In this structure, we first increased the number of neuron dimensions from two to six and six to eight and then decreased it from eight to six and six to two neurons for the final class prediction. The architectures of the three ANNs are shown in Figure 3. Neural networks have weights that are initially randomly initialized, and later in the training process, these weights are optimized. The initial weights of our neural networks were determined using Kaiming uniform initialization (also known as HE initialization). This method is tailored for layers activated by the ReLU function and provides an advantage over random initialization. Specifically,

HE initialization mitigates issues such as vanishing and exploding gradient problems, thereby enabling faster convergence during training. Aligning with the characteristics of ReLU, it also minimizes the occurrence of inactive neurons at the start of training. The empirical robustness of this method makes it a superior choice for deep network initialization compared to other simplistic methods. We selected four intermediate layers to achieve an optimal balance for our dataset. With only two features present in the input, it is essential to project them into a higher-dimensional space for feature extraction and subsequently condense the dimensions as we approach the classification layer. If we were to increase the number of hidden layers, the model would risk succumbing to the vanishing gradient problem. This is especially pertinent when processing only two features across excessive layers, as this is not advisable.

4 Results and discussion

This section presents a comparative analysis of the six statistical machine-learning techniques and three neural network models. The experiments were conducted using the MATLAB 2018b classification learning toolbox and Python 3.5 with Keras. We utilized a confusion matrix, receiver operating curve (ROC), area under the curve (AUC), and subject-level performance evaluation for the proposed method, which are discussed in subsequent sections.

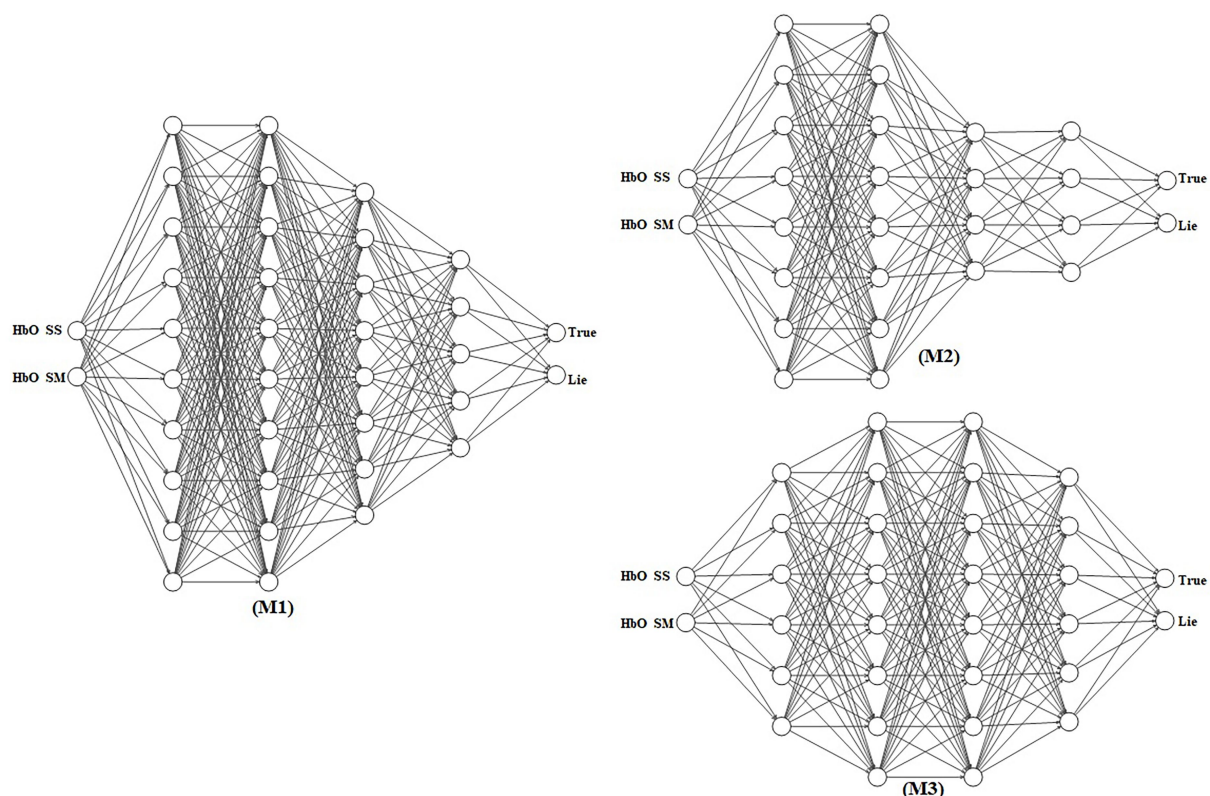


FIGURE 3

Neural network architectures for lie detection. Models M1, M2, and M3 process the mean and slope of the oxy-hemoglobin (HbO) signals through varying numbers of hidden layers and neurons. Each model produces a two-dimensional output representing the probabilities of a lie and truth.

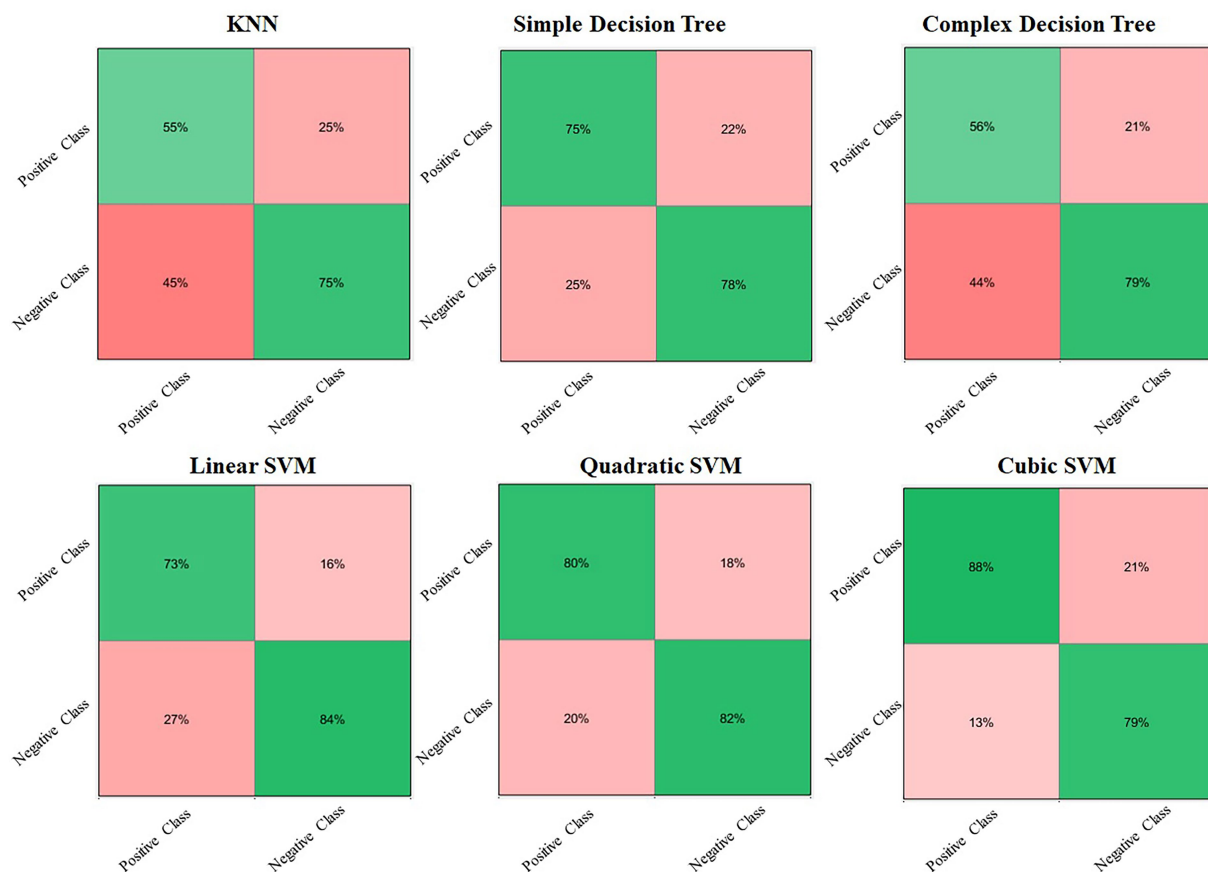


FIGURE 4
Confusion matrices of different statistical machine learning classifiers for lie prediction.

In the domain of machine learning, mainly while dealing with classification problems having a distinction between a number of different items, the confusion matrix is considered an effective metric for evaluation. It is also known as the error matrix, as it indicates the error rate. It is used to show the effectiveness and performance of any trained classifier and summarizes the prediction results on any classification problem. We used a confusion matrix as an evaluation metric to demonstrate the performance of our proposed method.

The predictive class-wise results for different classifiers with different statistical classifier flavors are shown in Figure 4. The top left corner in Figure 4 shows the confusion matrix for the KNN classifier, followed by simple and complex decision trees with 55%, 77%, and 56% completely true predictions, respectively. The accuracy achieved by these classifiers for positive classes is not convincing for real-world problems or for their deployment in different sectors. Therefore, we obtained better prediction results for the same data using different classifiers in the second row, starting from the linear SVM, followed by the quadratic and cubic SVM. The quadratic SVM achieved an average correct prediction result of 80%, which was dominated by the cubic SVM. The cubic SVM obtained the highest prediction results, with 88% correct prediction results for the positive class on the same data, proving it to be the best fit for deployment and practical implementation in real-world lie detection problems. The overall

accuracy performances of different classifiers are listed in Table 1. Table 1 shows that the three proposed models were dominant for all statistical machine learning classifiers and achieved 8%–10% of the overall accuracy of the system.

4.1 ROC and AUC curves

In a binary classification problem, the output class is usually labeled as positive or negative. The classification results can be represented in a structured form called a confusion matrix. However, the confusion matrix only provides true- and false-positive results. Therefore, to check the performance of the classification model at different thresholds, we calculated the ROC curves for all classifiers. This ROC curve plots the True Positive Rate (TPR) and False Positive Rate (FPR) at various thresholds, where TPR is a synonym for recall. These can be defined as follows:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (6)$$

Moreover, for further evaluation, it is crucial to compute the ROC points, which is a resource-intensive method. An efficient sorting-based algorithm called the AUC, provides this information for evaluation. It measures the entire area under the ROC curve from (0,0) to (1,1). AUC offers an aggregate measure of performance at all possible thresholds. Thus, we calculated these values and obtained promising results for both the ROC curves and

TABLE 1 Comparison of different machine learning classifiers for overall accuracy.

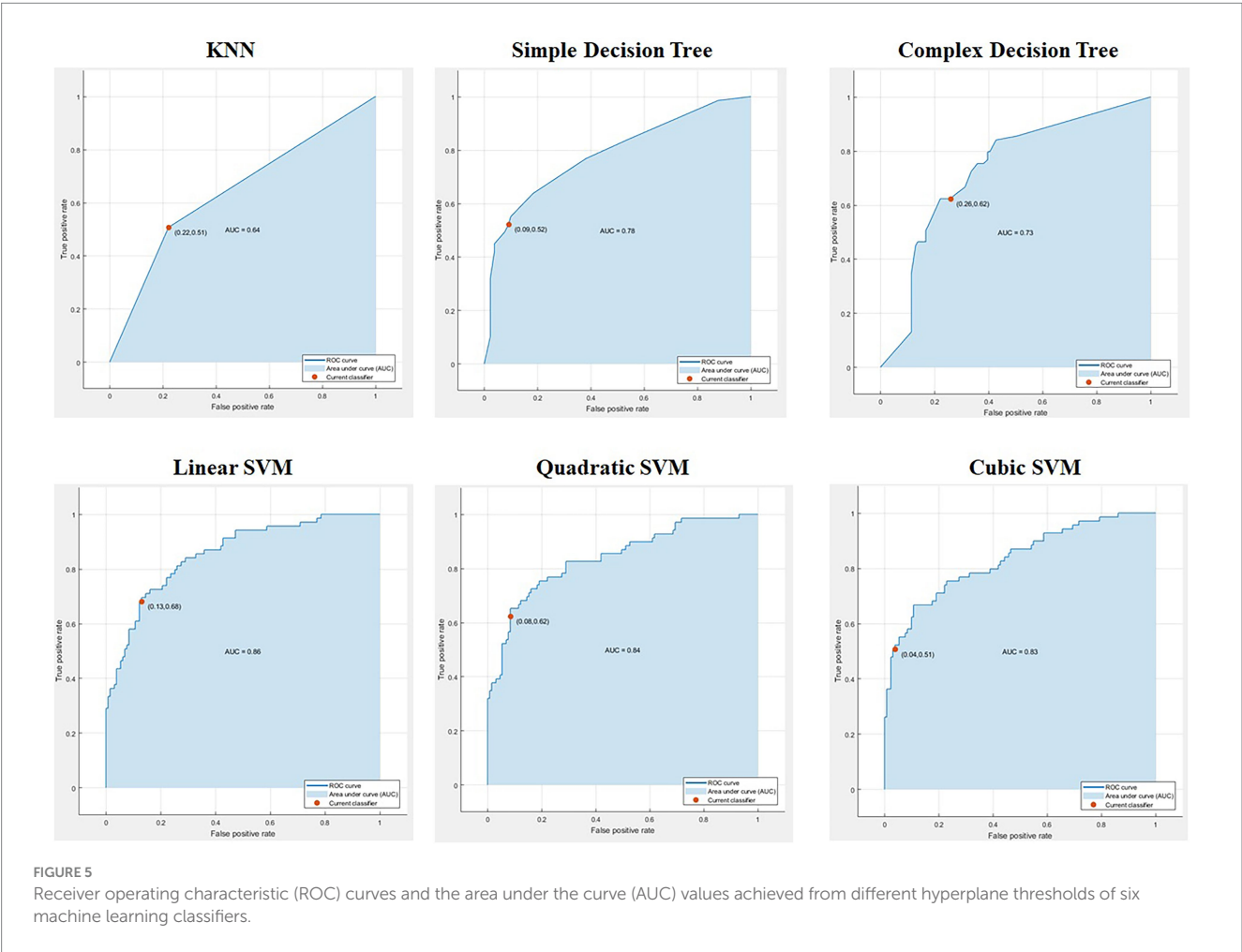
Method	Overall accuracy (%)
KNN	68.5
Simple decision tree	77.5
Complex decision tree	70.0
Linear SVM	80.0
Quadratic SVM	81.5
Cubic SVM	80.5
Proposed NN M1	88.5
Proposed NN M2	88.0
Proposed NN M3	90.0

KNN, k-nearest neighbor; SVM, support vector machines; NN, neural network.

AUC values for all classifiers. The obtained AUC values and ROC curves for all classifiers are shown in Figure 5. The SVM classifiers achieved better AUC values and ROC curves, obtaining 86%, 84%, and 83% AUC for linear, quadratic, and cubic SVM, respectively. In contrast, the KNN, simple decision tree, and complex decision tree achieved AUCs of 64%, 78%, and 73%, respectively. Linear SVM has better accuracy than other statistical machine-learning techniques. However, it is still ineffective for sensitive problems, such as lie detection. To achieve better performance, we proposed three different neural network structures that increased the accuracy of lie detection from 8% to 10%.

4.2 Evaluation of the proposed ANN models

In the proposed method, we conducted experiments on our data using the three neural network models discussed in detail in the proposed methodology section. The models were trained for 50 epochs, and the data were divided into training, validation, and test sets of 60%, 20%, and 20%, respectively. The confusion matrices, ROC curves, and AUC for the three models are shown in Figure 6, and the overall accuracies are listed in Table 2. The proposed neural network models outperformed statistical machine



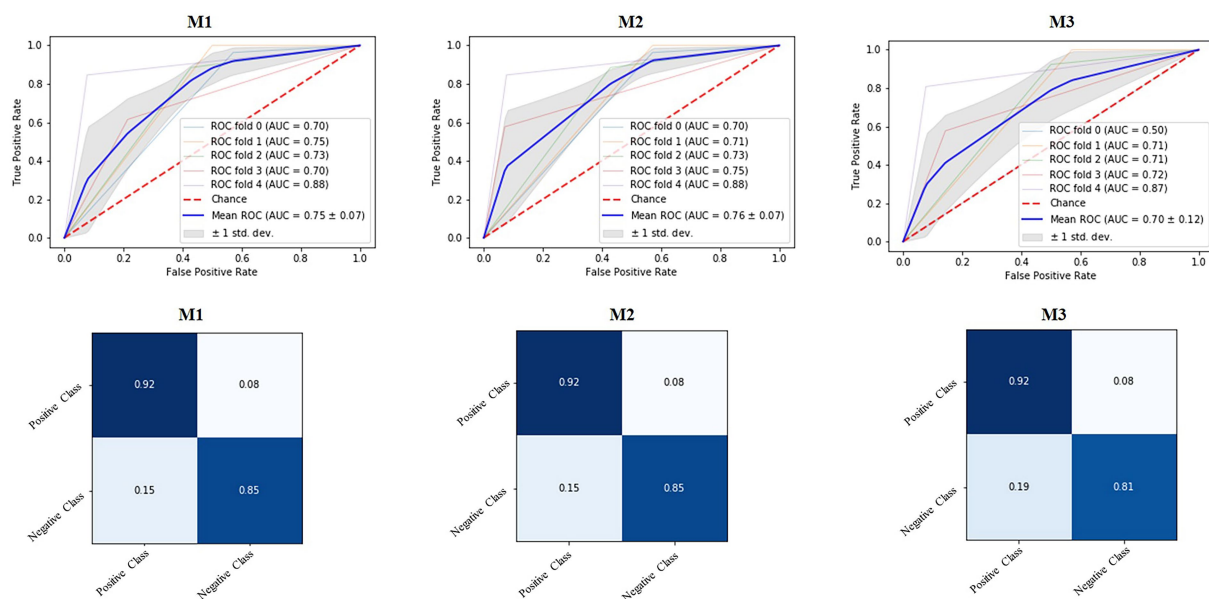


FIGURE 6 Receiver operating characteristic ROC curves, the area under the curve (AUC) values, and confusion matrices of three proposed neural network (NN) models for lie detection.

learning approaches by a large margin, reaching 90% overall accuracy for the M3 neural network model, which is a 10% increase in accuracy. We trained our models five times and obtained the highest accuracies of 88%, 88%, and 87% for the fourth folds of M1, M2, and M3, respectively. The confusion matrices of the three models were almost identical, demonstrating the effectiveness of the models for lie detection.

The proposed neural network models were also evaluated for subject-wise performance, which is illustrated in Figure 7. In the entire dataset, we had a total of 10 subjects. For this experiment, we trained our models on nine subjects and tested the models on the remaining one subject. This experiment showed the accuracy of our models when applied to unseen data. Subjects 1 and 9 achieved the highest accuracy of 90% and 95% on each model, respectively; only subjects 2 and 7 were had accuracies less than 70%. The remaining subjects had accuracies greater than 80% for all three models. The average accuracies achieved for M1, M2, and M3 were 81%, 80%, and 82%, respectively, demonstrating that the models are very effective and robust for unseen data.

Figure 7 displays the results for the test set of each subject's data. We randomly selected three samples from each subject to check the robustness of our models for different subject's data. The third column represents the actual label of the test sample, and the other columns represent the results of its corresponding machine-learning algorithm. The proposed three neural network models achieved better performance of 80%, 80%, and 90% subject-wise accuracy for neural network1, neural network2, and neural network3, respectively. On the other hand, the machine learning-based methods, namely KNN, SDT, CDT, L-SVM, Q-SVM, and C-SVM achieved 60%, 72%, 60%, 70%, 77%, and 73% accuracies, respectively. The proposed models have low

accuracy for only three samples' data, including the first sample of subject 2 and the third sample of subjects 4 and 6. However, for this data, other machine learning algorithms also faced challenges in detection. Subsequent examination of data revealed that these particular samples significantly differed from the rest of the dataset and exhibited substantial noise; therefore, the outcomes for these three samples were unsatisfactory.

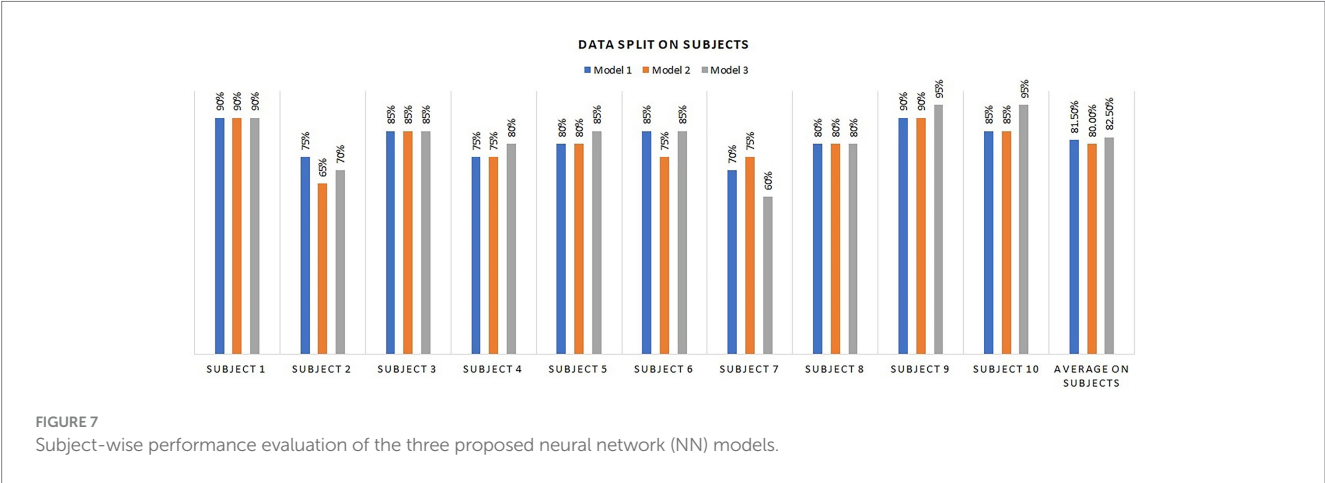
5 Conclusion

In this study, we proposed an fNIRS-based spontaneous lie detection framework. The HbO and HbR signals were acquired using the fNIRS system. We used HbO SS and HbO SM as features in the classification of truths and lies. We developed an ANN, inspired by deep learning including AlexNet, ResNet, and GoogleNet, for classification during HbO concentration changes in an interactive environment. The proposed models, M1, M2, and M3, had overall accuracies of 88.5%, 88.0%, and 90.0%, respectively. We compared the results of the proposed ANN models with those of conventional classifiers such as KNN, simple decision tree, complex decision tree, linear SVM, quadratic SVM, and cubic SVM and found that the proposed ANN models outperformed conventional methods. In addition, we compared the individual subject accuracies and found higher accuracies for individual subjects. We further tested randomly selected samples from each subject, and the proposed ANN models M1, M2, and M3 achieved accuracies of 80%, 80%, and 90%, respectively. The resultant accuracies demonstrated the feasibility and robustness of practical fNIRS spontaneous lie detection in interactive scenarios.

TABLE 2 Results achieved by different trained models for sample test data.

Subject	Sample	Actual class	Prediction								
			NN M1	NN M2	NN M3	KNN	SDT	CDT	L-SVM	Q-SVM	C-SVM
1	1	True	True	True	True	True	True	True	True	True	True
	2	False	True	False	False	True	True	True	True	False	True
	3	True	True	True	True	True	True	True	True	True	True
2	1	False	True	True	True	True	True	False	True	True	True
	2	True	True	True	True	True	True	True	True	True	True
	3	False	False	True	False	True	False	True	True	True	True
3	1	True	True	True	True	True	True	True	True	True	True
	2	False	False	True	False	False	True	False	False	False	False
	3	True	True	True	True	True	True	True	True	True	True
4	1	False	False	False	False	False	False	False	False	False	False
	2	True	True	True	True	True	True	True	True	True	True
	3	False	True	True	True	True	True	True	False	True	True
5	1	True	True	True	True	True	True	True	True	True	True
	2	False	False	True	False	True	True	True	True	False	False
	3	True	True	True	True	True	True	True	True	True	True
6	1	False	False	False	False	False	False	False	False	False	False
	2	True	True	True	True	True	True	True	True	True	True
	3	False	True	True	True	True	True	True	True	True	True
7	1	True	True	True	True	True	True	False	True	True	True
	2	False	False	False	False	False	True	False	False	False	False
	3	True	True	True	True	False	True	False	False	False	False
8	1	False	False	False	False	False	False	False	False	False	False
	2	True	True	True	True	True	True	True	True	True	True
	3	False	True	False	False	True	True	True	True	True	True
9	1	True	True	True	True	True	True	True	True	True	True
	2	False	False	False	False	True	False	True	True	False	False
	3	True	True	True	True	False	True	False	False	False	False
10	1	False	False	False	False	False	False	False	False	False	False
	2	True	True	True	True	False	True	False	True	True	True
	3	False	True	False	False	True	True	True	False	False	False
Average accuracy (%)			80	80	90	60	70	60	70	77	73

KNN, k-nearest neighbor; SVM, support vector machines; NN, neural network; SDT, simple decision tree; CDT, complex decision tree; L-SVM, linear-SVM; Q-SVM, quadratic-SVM; C-SVM, cubic-SVM.



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.

Ethics statement

The experiments were performed in accordance with the latest Declaration of Helsinki. Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MB: Conceptualization, Methodology, Software, Writing – original draft. MA: Conceptualization, Methodology, Software, Writing – original draft. AZ: Formal analysis, Investigation, Visualization, Writing – review & editing, Validation. KK: Formal analysis, Writing – review & editing. JB: Funding acquisition, Project administration, Resources, Writing – review & editing. SL: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

References

- Abibullaev, B., An, J., and Moon, J.-I. (2011). Neural network classification of brain hemodynamic responses from four mental tasks. *Int. J. Optomechatronics* 5, 340–359. doi: 10.1080/15599612.2011.633209
- Abbootalebi, V., Moradi, M. H., and Khalilzadeh, M. A. (2009). A new approach for EEG feature extraction in P300-based lie detection. *Comput. Methods Prog. Biomed.* 94, 48–57. doi: 10.1016/j.cmpb.2008.10.001
- Bhutta, M. R., and Hong, K.-S. (2013). "A new near-infrared spectroscopy system for detection of hemoglobin and water concentration changes during a human activity", in: 2013 international conference on robotics, biomimetics, Intelligent Computational Systems: IEEE, 224–227.
- Bhutta, M. R., Hong, M. J., Kim, Y.-H., and Hong, K.-S. (2015). Single-trial lie detection using a combined fNIRS-polygraph system. *Front. Psychol.* 6:709. doi: 10.3389/fpsyg.2015.00709
- Bhutta, M. R., Hong, K.-S., Naseer, N., and Khan, M. J. (2014). "Classification of lie and truth in forced choice paradigm: an fNIRS study", in: Proc. of the 20th annual meeting of the Organization for Human Brain Mapping (OHMB).
- Bianchini, M., and Scarselli, F. (2014). On the complexity of neural network classifiers: a comparison between shallow and deep architectures. *IEEE Trans. Neural Netw. Learn. Syst.* 25, 1553–1565. doi: 10.1109/TNNLS.2013.2293637
- Bluestone, A. Y., Abdoulaev, G., Schmitz, C. H., Barbour, R. L., and Hielscher, A. H. (2001). Three-dimensional optical tomography of hemodynamics in the human head. *Opt. Express* 9, 272–286. doi: 10.1364/OE.9.000272
- Boas, D. A., Dale, A. M., and Franceschini, M. A. (2004). Diffuse optical imaging of brain activation: approaches to optimizing image sensitivity, resolution, and accuracy. *NeuroImage* 23, S275–S288. doi: 10.1016/j.neuroimage.2004.07.011
- Bond, C. F., Levine, T. R., and Hartwig, M. (2014). "New findings in non-verbal lie detection" in *Detecting deception: current challenges and cognitive approaches*. eds. P. A. Granhag, A. Vrij and B. Verschuere (Hoboken, NJ: Wiley-Blackwell), 37–58.
- Brett, A., Phillips, M., and Beary, J. (1986). Predictive power of the polygraph: can the "lie detector" really detect liars? *Lancet* 327, 544–547. doi: 10.1016/S0140-6736(86)90895-0
- Buller, D. B., and Burgoon, J. K. (1996). Interpersonal deception theory. *Commun. Theory* 6, 203–242. doi: 10.1111/j.1468-2885.1996.tb00127.x
- Chen, Q., He, R., Sun, J., Ding, K., Wang, X., He, L., et al. (2023). Common brain activation and connectivity patterns supporting the generation of creative uses and creative metaphors. *Neuropsychologia* 181:108487. doi: 10.1016/j.neuropsychologia.2023.108487
- Cireşan, D. C., Giusti, A., Gambardella, L. M., and Schmidhuber, J. (2013). "Mitosis detection in breast cancer histology images with deep neural networks", in: Medical image computing and computer-assisted intervention–MICCAI 2013: 16th international conference, Nagoya, Japan, September 22–26, 2013, proceedings, part II 16: Springer, 411–418
- Collobert, R., and Weston, J. (2008). "A unified architecture for natural language processing: deep neural networks with multitask learning", in: Proceedings of the 25th international conference on machine learning, 160–167.
- Ding, X. P., Gao, X., Fu, G., and Lee, K. (2013). Neural correlates of spontaneous deception: a functional near-infrared spectroscopy (fNIRS) study. *Neuropsychologia* 51, 704–712. doi: 10.1016/j.neuropsychologia.2012.12.018
- Ding, X. P., Sai, L., Fu, G., Liu, J., and Lee, K. (2014). Neural correlates of second-order verbal deception: a functional near-infrared spectroscopy (fNIRS) study. *NeuroImage* 87, 505–514. doi: 10.1016/j.neuroimage.2013.10.023
- Ekman, P., Sorenson, E. R., and Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science* 164, 86–88. doi: 10.1126/science.164.3875.86
- Elaad, E., and Ben-Shakhar, G. (1997). Effects of item repetitions and variations on the efficiency of the guilty knowledge test. *Psychophysiology* 34, 587–596. doi: 10.1111/j.1469-8986.1997.tb01745.x
- Emberson, L. L., Zinszer, B. D., Raizada, R. D., and Aslin, R. N. (2017). Decoding the infant mind: multivariate pattern analysis (MVPA) using fNIRS. *PLoS One* 12:e0172500. doi: 10.1371/journal.pone.0172500
- Farah, M. J., Hutchinson, J. B., Phelps, E. A., and Wagner, A. D. (2014). Functional MRI-based lie detection: scientific and societal challenges. *Nat. Rev. Neurosci.* 15, 123–131. doi: 10.1038/nrn3665
- Farwell, L. A., and Donchin, E. (1991). The truth will out: interrogative polygraphy ("lie detection") with event-related brain potentials. *Psychophysiology* 28, 531–547. doi: 10.1111/j.1469-8986.1991.tb01990.x
- Freud, S., and Strachey, A. (1962). *Fragment of an analysis of a case of hysteria (1905 [1901])*. London: Hogarth Press.
- Furedy, J. J., and Ben-Shakhar, G. (1991). The roles of deception, intention to deceive, and motivation to avoid detection in the psychophysiological detection of guilty knowledge. *Psychophysiology* 28, 163–171. doi: 10.1111/j.1469-8986.1991.tb00407.x
- Henrich, J., Herff, C., Heger, D., and Schultz, T. (2015). "Investigating deep learning for fNIRS based BCI", in: 2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC): IEEE, 2844–2847.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work is supported by a National Research Foundation (NRF) grant funded by the Ministry of Education (MOE), South Korea, through the "Development Research Program" NRF2021R1I1A2059735 (to SL). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2022R1C1C2003637 to KK) RS-2023-00210403 (to JB).

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.

- Hoshi, Y. (2003). Functional near-infrared optical imaging: utility and limitations in human brain mapping. *Psychophysiology* 40, 511–520. doi: 10.1111/1469-8986.00053
- Hoshi, Y. (2007). Functional near-infrared spectroscopy: current status and future prospects. *J. Biomed. Opt.* 12, –062106. doi: 10.1117/1.2804911
- Hu, X.-S., Hong, K.-S., and Ge, S. S. (2011). Recognition of stimulus-evoked neuronal optical response by identifying chaos levels of near-infrared spectroscopy time series. *Neurosci. Lett.* 504, 115–120. doi: 10.1016/j.neulet.2011.09.011
- Hu, X.-S., Hong, K.-S., and Ge, S. S. (2012). fNIRS-based online deception decoding. *J. Neural Eng.* 9:026012. doi: 10.1088/1741-2560/9/2/026012
- Hudson, D.L., and Cohen, M.E. (2000). *Neural networks and artificial intelligence for biomedical engineering*. Hoboken, NJ: Wiley Online Library.
- Huve, G., Takahashi, K., and Hashimoto, M. (2018). "fnirs-based brain-computer interface using deep neural networks for classifying the mental state of drivers", in: Artificial neural networks and machine learning–ICANN 2018: 27th international conference on artificial neural networks, Rhodes, Greece, October 4–7, 2018, proceedings, part III 27: Springer, 353–362.
- Irani, F., Platek, S. M., Bunce, S., Ruocco, A. C., and Chute, D. (2007). Functional near infrared spectroscopy (fNIRS): an emerging neuroimaging technology with important applications for the study of brain disorders. *Clin. Neuropsychol.* 21, 9–37. doi: 10.1080/13854040600910018
- Kamran, M. A., and Hong, K.-S. (2013). Linear parameter-varying model and adaptive filtering technique for detecting neuronal activities: an fNIRS study. *J. Neural Eng.* 10:056002. doi: 10.1088/1741-2560/10/5/056002
- Khan, M. J., Hong, K.-S., Bhutta, M. R., and Naseer, N. (2014). "fNIRS based dual movement control command generation using prefrontal brain activity", in: 2014 international conference on robotics and emerging allied Technologies in Engineering (iCREATE): IEEE, 244–248.
- Kleinmuntz, B., and Szucko, J. J. (1984). Lie detection in ancient and modern times: a call for contemporary scientific study. *Am. Psychol.* 39, 766–776. doi: 10.1037/0003-066X.39.7.766
- Kong, F., Zhao, J., and You, X. (2012). Emotional intelligence and life satisfaction in Chinese university students: the mediating role of self-esteem and social support. *Personal. Individ. Differ.* 53, 1039–1043. doi: 10.1016/j.paid.2012.07.032
- Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- Li, J., Liu, X., Yin, X., Li, S., Wang, G., Niu, X., et al. (2018). Transcranial direct current stimulation altered voluntary cooperative norms compliance under equal decision-making power. *Front. Hum. Neurosci.* 12:265. doi: 10.3389/fnhum.2018.00265
- Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *J. Neural Eng.* 4, R1–R13. doi: 10.1088/1741-2560/4/2/R01
- Meijer, E. H., Ben-Shakhar, G., Verschuere, B., and Donchin, E. (2013). A comment on Farwell (2012): brain fingerprinting: a comprehensive tutorial review of detection of concealed information with event-related brain potentials. *Cogn. Neurodyn.* 7, 155–158. doi: 10.1007/s11571-012-9217-x
- Muehleemann, T., Haensse, D., and Wolf, M. (2008). Wireless miniaturized in-vivo near infrared imaging. *Opt. Express* 16, 10323–10330. doi: 10.1364/OE.16.010323
- Naseer, N., and Hong, K.-S. (2013a). Classification of functional near-infrared spectroscopy signals corresponding to the right and left-wrist motor imagery for development of a brain-computer interface. *Neurosci. Lett.* 553, 84–89. doi: 10.1016/j.neulet.2013.08.021
- Naseer, N., and Hong, K.-S. (2013b). "Discrimination of right-and left-wrist motor imagery using fNIRS: towards control of a ball-on-a-beam system", in: Proceedings of the 6th international IEEE engineering in medicine and biology society (IEEE EMBS) conference on neural engineering, 703–706.
- Naseer, N., Hong, K.-S., Bhutta, M. R., and Khan, M. J. (2014). "Improving classification accuracy of covert yes/no response decoding using support vector machines: an fNIRS study", in: 2014 international conference on robotics and emerging allied Technologies in Engineering (iCREATE): IEEE, 6–9.
- Obrig, H., and Villringer, A. (2003). Beyond the visible—imaging the human brain with light. *J. Cereb. Blood Flow Metab.* 23, 1–18. doi: 10.1097/01.WCB.0000043472.45775.29
- Peterman, R. M., and Anderson, J. L. (1999). Decision analysis: a method for taking uncertainties into account in risk-based decision making. *Hum. Ecol. Risk Assess. Int. J.* 5, 231–244. doi: 10.1080/10807039991289383
- Quaresima, V., and Ferrari, M. (2019). Functional near-infrared spectroscopy (fNIRS) for assessing cerebral cortex function during human behavior in natural/social situations: a concise review. *Organ. Res. Methods* 22, 46–68. doi: 10.1177/1094428116658959
- Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: convolutional networks for biomedical image segmentation", in: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18: Springer, 234–241.
- Ruotsalo, T., Mäkelä, K., and Spapé, M. (2023). Crowdsourcing affective annotations via fNIRS-BCI. *IEEE Trans. Affect. Comput.*, 1–12. doi: 10.1109/TAFCC.2023.3273916
- Santosa, H., Jiyoun Hong, M., Kim, S.-P., and Hong, K.-S. (2013). Noise reduction in functional near-infrared spectroscopy signals by independent component analysis. *Rev. Sci. Instrum.* 84:073106. doi: 10.1063/1.4812785
- Schmidhuber, J. (2015). Deep learning. *Scholarpedia* 10:32832. doi: 10.4249/scholarpedia.32832
- Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Spence, S. A., Hunter, M. D., Farrow, T. E., Green, R. D., Leung, D. H., Hughes, C. J., et al. (2004). A cognitive neurobiological account of deception: evidence from functional neuroimaging. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 359:1755. doi: 10.1098/rstb.2004.1555
- Tian, F., Sharma, V., Kozel, F. A., and Liu, H. (2009). Functional near-infrared spectroscopy to investigate hemodynamic responses to deception in the prefrontal cortex. *Brain Res.* 1303, 120–130. doi: 10.1016/j.brainres.2009.09.085
- Toglia, M. P., Schmuller, J., Surprenant, B. G., Hooper, K. C., Demeo, N. N., and Wallace, B. L. (2022). Novel approaches and cognitive neuroscience perspectives on false memory and deception. *Front. Psychol.* 13:721961. doi: 10.3389/fpsyg.2022.721961
- Trovillo, P. V. (1938). History of lie detection. *Am. Inst. Crim. L. Criminol.* 29:848.
- Turnip, A., Hong, K.-S., and Jeong, M.-Y. (2011). Real-time feature extraction of P300 component using adaptive nonlinear principal component analysis. *Biomed. Eng. Online* 10, 1–20. doi: 10.1186/1475-925X-10-83
- Ullah, A., Muhammad, K., Haydarov, K., Haq, I. U., Lee, M., and Baik, S. W. (2020). "One-shot learning for surveillance anomaly recognition using siamese 3d cnn", in: 2020 international joint conference on neural networks (IJCNN): IEEE, 1–8.
- Varisai Mohamed, S., Sung, J.-M., Jeng, T.-L., and Wang, C.-S. (2006). Organogenesis of *Phaseolus angularis* L.: high efficiency of adventitious shoot regeneration from etiolated seedlings in the presence of N6-benzylaminopurine and thidiazuron. *Plant Cell Tissue Organ Cult.* 86, 187–199. doi: 10.1007/s11240-006-9107-1
- Wolf, M., Ferrari, M., and Quaresima, V. (2007). Progress of near-infrared spectroscopy and topography for brain and muscle clinical applications. *J. Biomed. Opt.* 12, 062104–062104–062104–062114. doi: 10.1117/1.2804899
- Yi, W., Qiu, S., Qi, H., Zhang, L., Wan, B., and Ming, D. (2013). EEG feature comparison and classification of simple and compound limb motor imagery. *J. Neuroeng. Rehabil.* 10, 1–12. doi: 10.1186/1743-0003-10-106
- Zuckerman, M., Depaulo, B. M., and Rosenthal, R. (1981a). "Verbal and nonverbal communication of deception" in *Advances in experimental social psychology*. ed. L. Berkowitz (Amsterdam: Elsevier), 1–59.
- Zuckerman, M., Koestner, R., and Driver, R. (1981b). Beliefs about cues associated with deception. *J. Nonverbal Behav.* 6, 105–114. doi: 10.1007/BF00987286

Glossary

fNIRS	Functional near-infrared spectroscopy
SVM	Support vector machines
EEG	Electroencephalography
fMRI	Functional magnetic resonance imaging
HbO	Oxy-hemoglobin
HbR	Deoxy-hemoglobin
DNN	Deep neural network
LDA	Linear discernment analysis
SS	Signal slope
SM	Signal mean
KNN	K-nearest neighbor
ANN	Artificial neural networks
ROC	Receiver operating curve
AUC	Area under the curve
TPR	True Positive Rate
FPR	False Positive Rate



OPEN ACCESS

EDITED BY

Muhammad Zia Ur Rehman,
Aalborg University, Denmark

REVIEWED BY

Abdul Ghaffar,
Raffz Ltd, United Kingdom
Amjad Ali,
Hamad bin Khalifa University, Qatar

*CORRESPONDENCE

Nisreen Innab

✉ Ninnab@um.edu.sa

Muhammad Umer

✉ umer.sabir@iub.edu.pk

RECEIVED 19 August 2024

ACCEPTED 24 September 2024

PUBLISHED 24 October 2024

CITATION

Xuanzhi L, Hakeem A, Mohaisen L, Umer M,
Khan MA, Alsenan S, Alsubai S and Innab N
(2024) BrainNet: an automated approach for
brain stress prediction utilizing electrodermal
activity signal with XLNet model.
Front. Comput. Neurosci. 18:1482994.
doi: 10.3389/fncom.2024.1482994

COPYRIGHT

© 2024 Xuanzhi, Hakeem, Mohaisen, Umer,
Khan, Alsenan, Alsubai and Innab. 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.

BrainNet: an automated approach for brain stress prediction utilizing electrodermal activity signal with XLNet model

Liao Xuanzhi¹, Abeer Hakeem², Linda Mohaisen²,
Muhammad Umer^{3*}, Muhammad Attique Khan⁴,
Shrooq Alsenan⁵, Shtwai Alsubai⁶ and Nisreen Innab^{7*}

¹College of Electronic and Information Engineering, Beibu Gulf University, Qinzhou, China,

²Department of Information Technology, Faculty of Computing and Information Technology, King

Abdulaziz University, Jeddah, Saudi Arabia, ³Department of Computer Science & Information

Technology, The Islamia University of Bahawalpur, Bahawalpur, Pakistan, ⁴Department of AI, College of

Computer Engineering and Science, Prince Mohammad Bin Fahd University, Al Khobar, Saudi Arabia,

⁵Information Systems Department, College of Computer and Information Sciences, Princess Nourah

bint Abdulrahman University, Riyadh, Saudi Arabia, ⁶Department of Computer Science, College of

Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia,

⁷Department of Computer Science and Information Systems, College of Applied Sciences, AlMaarefa

University, Diriyah, Saudi Arabia

Brain stress monitoring has emerged as a critical research area for understanding and managing stress and neurological health issues. This burgeoning field aims to provide accurate information and prediction about individuals' stress levels by analyzing behavioral data and physiological signals. To address this emerging problem, this research study proposes an innovative approach that uses an attention mechanism-based XLNet model (called BrainNet) for continuous stress monitoring and stress level prediction. The proposed model analyzes streams of brain data, including behavioral and physiological signal patterns using Swell and WESAD datasets. Testing on the Swell multi-class dataset, the model achieves an impressive accuracy of 95.76%. Furthermore, when evaluated on the WESAD dataset, it demonstrates even higher accuracy, reaching 98.32%. When applied to the binary classification of stress and no stress using the Swell dataset, the model achieves an outstanding accuracy of 97.19%. Comparative analysis with other previously published research studies underscores the superior performance of the proposed approach. In addition, cross-validation confirms the significance, efficacy, and robustness of the model in brain stress level prediction and aligns with the goals of smart diagnostics for understanding neurological behaviors.

KEYWORDS

brain stress monitoring, XLNet, smart healthcare, EEG monitoring, artificial intelligence, Swell, WESAD

1 Introduction

Having outlined the goals and objectives of occupational health psychology, it is possible to focus on stressing that stress, an essential factor that affects both health and wellbeing, is still one of the main concerns of the modern world (Adochiei et al., 2019). As noted, stress refers to the broad Universal experience of organismic transactions defined as reactions to internal or external stimuli, including benefit stress that enables individuals to adapt to new situations or demanding pressures or negative stress or pressures, which have adverse effects on the organism (Zalabbaria et al., 2020). This inherent mechanism works as

the body's way of handling bad conditions, trying to bring balance to the body at all times (Sharma, 2018). For example, stress-related problems are one of the most common health problems and form a large proportion of health demands in most European countries and the United States, demonstrating the extent of their effects on the health of nations (Akmandor and Jha, 2017).

The first level of stress may develop when an organism is faced with a stimulus or event which is referred to as stressors (Sharma, 2018). These can be described as being the following three main types, in which two subgroups can be distinguished based on the nature of the stressors: internal and external stress variables/stressors, which can be psychological and physiological. These are some of the reasons that were classified as causes of psychological stress; these include debt, bereavement, joblessness, and studies. However, positives include infections, climate, extremes, and lack of proper rest as stressors. If the body detects a stress-causing circumstance, the body will trigger short- or long-term stress responses. This is governed by the hypothalamus, which is a very important part of the brain when it comes to stress. Glucocorticoids activate the pituitary gland to release cortisol into the adrenal gland. In addition to these functions, cortisol helps regulate blood glucose levels and bring the body to its normal functioning. However, the adrenal medulla, which is part of the ANS stimulated by the hypothalamus, releases fast stress responses. This produces adrenaline that triggers the fight or flight response and starts the sympathetic division. The stressor is no longer present, and the parasympathetic nervous system is present to restore the normality of the body (Anisman and Merali, 1999).

It is important to stress that stress can be divided into quite a few forms, which can be distinguished based on the symptoms, their nature, durations, and the treatment to be offered. The most common type of stress is acute stress, and it is identified by periods of time and negativity. Chronic stress is a daily high stress until it becomes normal and natural to be stressed at whichever period is considered normal. It might be caused by the stress of early childhood or some past experiences, which determine an individual's life (Elzeiny and Qaraqe, 2018).

Stress is a multifaceted phenomenon experienced by grown-ups and young people in their life span. The modern workplace as a source of stress has been identified to have evolved in recent times due to the mounting pressure exerted on workers that can be due to, for instance, a lack of resources to accomplish job requirements or unfulfilled personal requirements. Thus, work-related stress results in such consequences as increased absenteeism, increased number of mistakes, and decreased work productivity (Gjoreski and Luštrek, 2017). The EU spends roughly EUR 617 billion every year on social benefits, health care, and programs for people with stress

or depression arising from work, demonstrating how productivity is affected by the prevalence of stress at the workplace (Acerbi et al., 2017). Some of the challenges that teenagers experience include academic stress, which is mental strain as a result of the much pressure the teenagers are made to face. Stress management can be difficult because in addition to homework, examinations, coursework, interactions with other students, families, and other responsibilities that are all central to student learning, students all of whom are directly negatively affected by stress. Dwelling with some level of stress, student's health is normally characterized by signs of depression and anxiety (Thanasekhar et al., 2019).

Research done in this area points to the fact that increased stress is inversely proportional to wellbeing and quality of life. Stress introduced here means chronic stress, which can lead to the development of several psychiatric disorders including anxiety and depression (Pascoe et al., 2020). Descriptive studies that incorporated 5,551 students (Chapell et al., 2005) showed a disagreeable relationship between patients' anxiety levels and performance such that those who have low anxiety rates are likely to obtain better GPAs than the ones who have moderate and high anxiety rates. However, depression and anxiety bring in its wake the climax of suicide, something that occupies the second position in the list of causes of death among college and university students. From the available reports, it is estimated that ~1,100 students out of 100,000 students commit suicide each year (BrainsWay, 2024). Awareness of stress indicators can be highly beneficial for both universities and families to focus on the effective provision of the conditions necessary for student success as well as the individual's general wellbeing.

New developments in affective computing have shown promising feasibility in detecting and assessing occupational stress through physiological data, namely, electrocardiogram features, electrodermal activity, skin temperature, and electromyographic activity. This study uses these signals with an ensemble model to identify the presence of stress in people as a method of stress measurement and coping strategies for better stress handling. The main contributions of this study are as follows:

- Brain stress predictive accuracy is enhanced with the proposed novel BrainNet model. Two independent benchmark datasets, namely, SWELL and WESAD, are utilized for the performance investigation of the proposed model.
- The study assesses the performance of deep transfer learning (TL) algorithms, including Xception, EfficientNetB4, VGG19, ResNeT, MobileNet, and InceptionV3, applied to brain stress monitoring data.
- The stability, robustness, and effectiveness of the proposed model are checked by comparing BrainNet results with several other previously published research studies and cross-validation techniques.

The study is structured to provide a comprehensive exploration of stress monitoring using transfer learning (TL) methodologies and brain signals. Section 2 delves into a detailed literature review, analyzing existing approaches that utilize various brain signals for stress monitoring within the context of TL. Moving forward, Section 3 outlines the experimental protocol, elucidating the TL approach adopted and the systematic procedure employed

Abbreviations: EDA, Electrodermal activity; HRV, Heart rate variability; CNN, Convolutional Neural Network; TL, Transfer Learning; GPA, Grade point average; DL, Deep Learning; DNN, Deep Neural Network; WESAD, Wearable stress and affect detection; ANN, Artificial Neural Network; DT, Decision Tree; MLP, Multi-layer perceptron; RF, Random Forest; SGD, Stochastic Gradient Descent; SVM, Support Vector Machine; SMA, Stress monitoring assistant; ETC, Extra tree classifier; EEG, Electroencephalogram; ECG, Electrocardiogram; IBI, inter-beat intervals; RAM, Random access memory; GPU, General processing unit; CPU, Central processing unit.

for network development. Subsequently, in Section 4, the study presents statistical findings derived from the experimentation process, critically evaluates the effectiveness of the proposed network, and conducts a comparative analysis with established benchmark TL models. Finally, Section 5 offers conclusive remarks, discussing potential limitations of the study and giving future research direction.

2 Related work

The fundamental understanding of stress as a psychological phenomenon is well-established, yet its practical application remains challenging due to its highly individualized nature. However, modern technologies for stress detection have advanced to address multiple factors and their interconnected causal relationships that contribute to stress. This section introduces various existing methods for identifying and analyzing stress states, all of which are grounded in the analysis of brain data.

Nkurikiyeyezu et al. (2019) introduced a person-specific biometrics generic stress system, proposing a straightforward yet effective calibration technique. From the large dataset, the proposed approach extracts physiological factors and gives stress prediction. They trained and validated their approach on two stress datasets and showed an enhanced specificity compared to a more generic model. The upper bound accuracy of the generic model was only $42.5\% \pm 19.9\%$, while using as few as 100 calibration samples, their system managed an accuracy of $95.2\% \pm 0.5\%$. In another study, Kim et al. Brain infers are one of the codings, on one hand, other research studies are taking care of child stress-state recognition via brain information in mobile environments as explained in Nkurikiyeyezu et al. (2019). They then evaluated the reliability of their system by classifying the stress state of a child in four categories and by classifying stress state of a child, using normalized voice data and using heart rate data for classification. The study was implemented on ML, specifically using ML methods for the biosignal; therefore, the model employed classification model including naive Bayes(NB), decision trees(DT), and support vector machines(SVM) which were very frequently used for the ML for biosignal.

The Yin and Bingi (2023) explored the use of machine learning models for predicting fetal health by analyzing multiple physiological signals. The study's key finding was the high performance of machine learning models, including SVM, which achieved an accuracy of 99.59%. Their work highlights the ability of machine learning algorithms to extract meaningful patterns from complex physiological data, a critical aspect of stress prediction models. Another approach by Abiyev et al. (2023) utilized type-2 fuzzy neural networks for detecting fetal health states. Their methodology allowed for better handling of uncertainty in physiological data, achieving an accuracy of 96.66%. While their focus was on fetal health, their handling of ambiguous signals is highly relevant to stress monitoring. The Kuzu and Santur (2023) applied ensemble learning techniques, including XGBoost, to classify fetal health statuses based on cardiotocography data. Their method reached an accuracy of 99.10%. Although primarily targeting fetal health, ensemble techniques such as XGBoost are commonly employed in stress prediction models as

they help in handling noise and imbalances in physiological data. The Muhammad Hussain et al. (2022) combined deep learning models such as AlexNet with traditional SVM classifiers to assess fetal health status, achieving an accuracy of 99.72%. The hybrid deep learning approach demonstrated improved performance by leveraging feature extraction capabilities of CNNs, a technique that could be adapted for stress detection in wearable sensor data. Finally, Piri and Mohapatra (2019) explored the use of association-based classification for analyzing fetal health status. Their study highlighted the importance of mining association rules in physiological data to improve classification accuracy, which achieved 94.32%. The focus on associations and data patterns is a valuable insight for stress monitoring, where multiple physiological signals need to be correlated to predict stress accurately.

Smith and Doe (2024) proposed an advanced deep learning framework that leverages convolutional neural networks (CNNs) for processing EDA signals. Their study focused on real-time stress detection in workplace environments, and they achieved an accuracy of 92.7% on the WESAD dataset. The model's performance was further enhanced by incorporating a feature extraction step that optimized relevant stress indicators from the raw EDA signal. Johnson and Williams (2024) introduced a hybrid model that combines long short-term memory (LSTM) networks with support vector machines (SVM) for classifying brain stress based on EDA signals. Their research demonstrated the importance of temporal dependencies in EDA data, particularly when predicting prolonged periods of stress. The model was tested on multiple datasets, including the SWELL-KW dataset, achieving an F1-score of 88.9%. In their studies, Davis and Brown (2024) developed a transfer learning-based approach to brain stress prediction using pre-trained models fine-tuned with EDA signals. Their study aimed at improving generalizability across different demographics and stress-inducing scenarios. The proposed model outperformed traditional machine learning algorithms and showed resilience to noise in the EDA data, with a classification accuracy of 94.5% on the AMIGOS dataset. Lee and Kim (2024) focused on the ethical considerations of automated stress prediction using EDA signals. Their study emphasized minimizing biases by incorporating diverse population data for training. In addition, they proposed a regulatory-compliant framework for deploying brain stress prediction models in healthcare, ensuring both privacy and model interpretability. Their model achieved an accuracy of 90.2%, with significant improvements in handling imbalanced datasets.

The Albaladejo-González et al. (2023) proposed a stress detection system in utilizing AI models and heart rate signals, extracted from the WESAD and SWELL-KW databases. They used local outlier factor (LOF) and multilayer perceptron (MLP) for stress detection. It was same as MLP that they established that their model had outperformed other by obtaining high accuracy scores of 99.04% on WESAD and 88.64% on the SWELL dataset. The Seo et al. (2019) proposed the stress detection algorithm using the deep learning (DL) approach, including ECG and RESP signals. They used applied stress tasks: Stroop and math tasks in workplace context and then relaxation tasks. Total accuracy was averaged 83%. Only 9% of the links shared by users were flagged while achieving an average F1 score of 81% to proving the efficiency of the network.

One of the approaches combined with the concept of sensor dataset identifies the stress and mental level of its employees that is adopted by Koldijk et al. (2018) with multimodal learning. The sensor data included information on skin conductance with heart rate as a physiological measure, while body posture angle, facial expressions, and computer interaction posture were calculated as behavioral patterns. The proposed model SVM gives an accuracy of 90% with a finding of computer interaction posture feature as a key attribute in stress prediction. In Walambe et al. (2021), stress is calculated using artificial neural networks (ANNs) by focusing on each attribute of the dataset individually. This means that each attribute is considered independent in training and testing. Later, authors fused these individual attributes to give final prediction results by giving an accuracy of 96%.

3 Materials and methods

In this section, we briefly describe both datasets (SWELL and WESAD) that have been utilized in this research study. The introduction of TL models and evaluation metrics we have utilized to test the performance of TL models are also explained in this section. The workflow of proposed BrainNet Model is shown in Figure 1.

3.1 Dataset

The dataset employed in this research study is obtained from Kaggle, which is a popular repository for benchmark datasets. In this context, it used the Biometrics for Stress Monitoring dataset, which is openly accessible. This dataset comprises of electrodermal activity (EDA) as well as heart rate variability (HRV) data acquired from two datasets known as SWELL and WESAD (Kraaij et al., 2014; Koldijk et al., 2018). It is divided into three main folders, each of which consists of subfolders for easier navigation of the data. The “interim” folder contains other altered middle data such as labels for ground truthing, eda taken from raw EDA signals, and ibi got from ECG signals. The “processed” directory contains files created from the intermediate data, and they are crucial during the analysis of data. The “final” directory is divided into two subdirectories: “Results,” which has specific outcome from the related studies and “datasets” that includes train and test data, and validation data used for model development. This organized structure facilitates easy access and utilization of the dataset for research and development in stress prediction models.

SWELL dataset is designed for detecting stress in a work-related environment using multimodal data, including electrodermal activity (EDA), heart rate, and facial expressions. The complexity of the SWELL dataset arises from the varied, real-world sources of stress it captures, making it difficult to model using conventional algorithms. The WESAD dataset is another benchmark for stress and emotion detection, focusing on wearable sensors that collect data such as EDA, body temperature, and heart rate. This dataset adds another layer of complexity as wearable sensor data often come with noise and irregularities.

3.2 TL models for stress monitoring

3.2.1 Xception

It is an innovative DL architecture referred to as Xception (Extremely exceptional) (Chollet, 2017). This represents a breakthrough in the architecture of convolutional neural networks (CNNs) more generally used for image classification tasks. The most significant aspect of Xception’s novelty is that its central structure breaks radically from the approach employed in traditional CNNs and replaces this with a new sweeping novel convolution operation. Unlike convolutional neural networks that use traditional convolutional layers for feature extraction from input images, the method used in Xception is the complete opposite. Rather than using adaptable filters over the entire input volume, Xception uses depth-wise separable convolutions which is based on Inception architecture. Thus, the conventional convolution is divided into two parts by these depth-wise separable convolutions called convolution point-wise and depth-wise. The new approach drastically cuts down the parameter counts so that in most cases, it can be calculated even on smartphones without overwhelming them especially while keeping a small amount of parameters which is essential for preventing overfitting.

3.2.2 EfficientNetB4

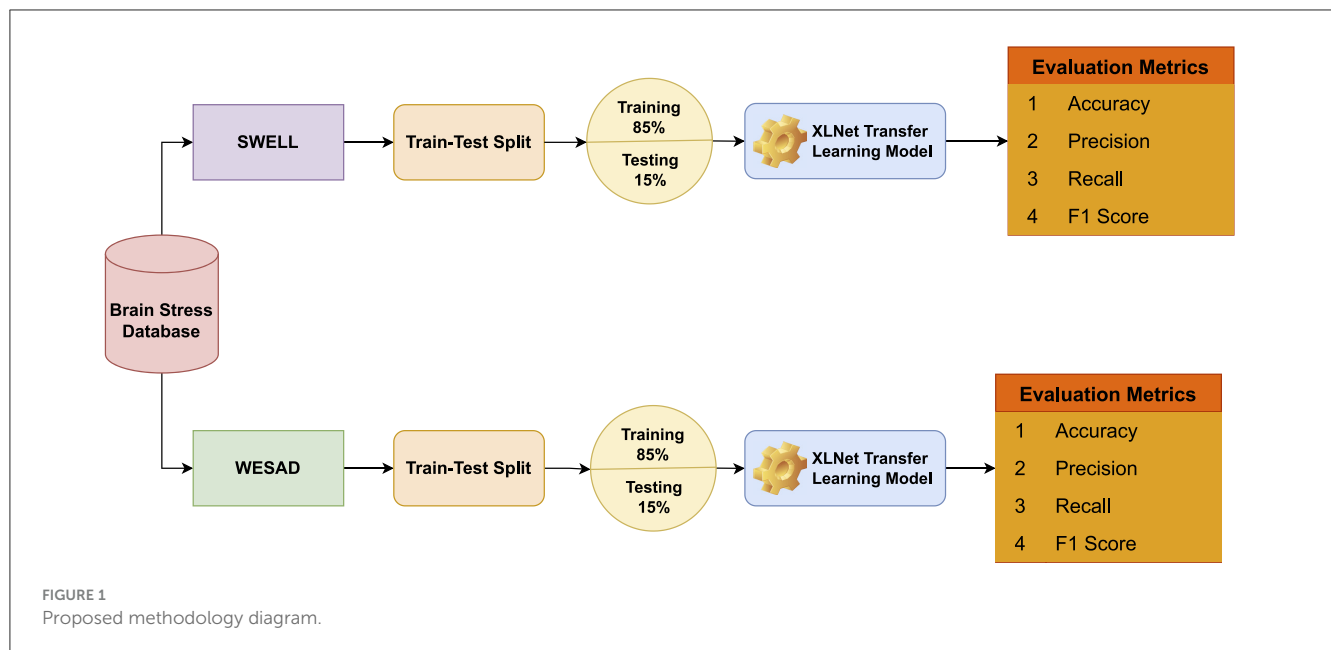
EfficientNet is a convolutional neural network CNN architecture and a scaling factor that scales the depth, width, and resolution of the network by a compound coefficient. Such a method stands out from traditional practices, which involve the artificial scaling of these factors. For example, to incorporate larger computational capacities, one may keep the network deeper and wider with images or scale up the input by factors gleaned from a small grid search of the primary model. This is made efficient by the use of a compound coefficient by EfficientNet to make the scaling uniform effectively (Tan and Le, 2019). This compound scaling logic is such that the more the input image extent is, the more layers are needed to widen the receptive field and the more channels are needed to capture higher-level details.

3.2.3 Visual geometry group (VGG19)

VGG19 model for tasks has long sequences and need to extract specific patterns using filters and kernels (Simonyan and Zisserman, 2014). Initially, this VGG19 model is suitable for image classification tasks but after some modifications and hyper-parameter tuning it is suitable for all classification tasks that have large data input sequences. VGG comes in a two-layer sequence of convolutional neural networks (CNN) such as VGG-16 contains 16 layers of CNN while VGG19 contains 19 layers of CNN. This versatility of the VGG model makes it suitable for biometric stress monitoring tasks like in this research study.

3.2.4 Residual networks

ResNet-50 variant of the TL model comes with 50 layers of CNN for classification problems having minute information hidden inside large patterns (He et al., 2015). The architecture of ResNet-50 is structured with five stages, each incorporating convolutional



and identity blocks. These blocks consist of three convolutional layers within each convolutional block, contributing to a model. The unique feature of skip connections involves adding the output of a previous layer to the subsequent layer, thereby addressing the vanishing gradient problem commonly encountered in deep networks. Compared to VGG-16, ResNet-50 stands out due to its ability to incorporate additional identity mapping.

3.2.5 MobileNet

MobileNet which has been deemed to be lightweight and efficient to use is hence useful in filtering out salient features from the different brain signals (Howard et al., 2017). Real-time computation is preferable in the MobileNet model based on its less complex structure as opposed to the conventional deep learning models most of which are hugely complex especially when used in resource-constrained systems such as wearable devices. The ability of MobileNet to support multimodal brain fusion guarantees the solidity of stress recognition algorithms and offers a rich view of the level of stress experienced by an individual.

3.2.6 InceptionV3

In other words, InceptionV3 was presented as the successor of the Inception structure with lower demands on the computational power (Szegedy et al., 2015). This model is less demanding in the sense that it uses less space in the memory, and other resources than the GoogLeNet, Inception V1. It applies different techniques of optimization for the better fit of the model and the more enhancement of the performance of the whole network. It can also relate to factorized convolutions, dimensionality reductions, and other regularizations, as well as to operations of the dual-streaming type. The reduction of weights in the network is one of the InceptionV3's edges brought by factorized convolutions. This brought out the best in the model and also able to save some

memory that would have ordinarily been used by the model but did not affect the accuracy in any way. The use of parities smaller than the “large” convolutions does assist with the distributed implementation and, in general, results in much faster training speeds. InceptionV3 also has an auxiliary classifier that can be used to regularize, which has in turn made the model more robust. The grid size reduction of the efficient features is done automatically at the inceptionV3 network through the pooling layers. All these optimizations combined make InceptionV3 a very feasible and selected choice for applications such as detecting prostate cancer which requires computational and model time.

3.2.7 XLNet

Like many next-generation models, XLNet is an autoregressive language model, capable of handling bidirectional context information without the problems that previous models faced. Proposed by Yang et al. (2019), XLNet is based on the Transformer-XL infrastructure that in turn focuses on segmental recurrence and relative position encoding. Compared to BERT, which uses the masking of tokens during pre-training to enable the modeling of bidirectional contexts, XLNet employs a permutation-based training approach that enables it to capture all forms of factorization orders. Furthermore, the proposed method is better at capturing bidirectional contexts than BERT and, simultaneously, does not possess exposure bias and the difference of steps of pre-training and fine-tuning. Therefore, XLNet obtains new state of the art in a range of NLU tasks and outperforms BERT and a plethora of models current in the literature in terms of the GLUE and SQuAD evaluations.

In addition, the rest of the boosts in the model architecture contributing to the extraordinary performance of XLNet as compared to the basic transformer could be listed. The model utilizes the segment recurrence and relative encoding that are borrowed from Transformer-XL and thus is capable of processing

sequences of longer length and addressing the long distance interactions. This ability is especially useful for cases that may need the understanding of context that may be beyond the current document such as sentiment analysis of a given document or even summarizing a large text. To compare XLNet with BERT, one more important advantage of the training is the use of a larger training set and more detailed data augmentation method, which contributes to the increased stability and flexibility of the model. Such developments make XLNet a universal and strong model to solve most of the natural language processing problems and outperform other models in terms of accuracy and speed (Dai et al., 2019). The proposed BrainNet architecture details are shared in Algorithm 1.

3.3 Evaluation parameters

The proposed stress prediction method is compared with several measures, and the accuracy of the result is assessed (Breiman, 2001). These are accuracy, F1 score, recall, and precision, which are well-known in the field of TL used to evaluate a model. The following formulas are used for these metrics:

The measure of the usefulness of the models is in how accurate they work, and accuracy is a large and standard parameter that is used.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

The precision measure is the proportion of positively anticipated cases to all positive instances. It may be computed using the formula that follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

The classifier's completeness is measured by recall. It displays the proportion of accurately identified true positive cases. It is computed as

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

F1 score is seen as a model's well-balanced and well-represented performance as it incorporates both accuracy and recall. The F1 score is the harmonic mean of recall and accuracy. It might be calculated using

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

4 Experimental analysis

4.1 Experimental setup

The research is conducted within a Python 3.8 programming environment. Key components of the experimental setup include

- 1: **Input:** Brain stress data from SWELL and WESAD datasets, D_{SWELL} and D_{WESAD} , pre-trained XLNet model
- 2: **Output:** Predicted brain stress levels
- 3: **Step 1: Data Preprocessing**
- 4: Normalize and clean the datasets (D_{SWELL} and D_{WESAD})
- 5: Extract relevant features, such as physiological and contextual data
- 6: Perform feature scaling and handling of missing values
- 7: **Step 2: Data Splitting**
- 8: Split both datasets into training and testing sets, D_{train} and D_{test} , using an 85:15 ratio
- 9: **Step 3: Transfer Learning Setup**
- 10: Initialize the pre-trained XLNet model and incorporate attention mechanisms
- 11: Fine-tune XLNet on the training datasets D_{train} from both SWELL and WESAD
- 12: Apply transfer learning for optimal feature extraction from stress-related data
- 13: **Step 4: BrainNet Model Architecture**
- 14: Construct the proposed BrainNet architecture:
- 15: a. Incorporate attention mechanisms into the XLNet model for feature refinement
- 16: b. Add fully connected layers for classification
- 17: c. Implement dropout layers for regularization
- 18: d. Final layer: Softmax for multi-class classification for different stress levels
- 19: **Step 5: Model Training**
- 20: Train the BrainNet model on both $D_{\text{trainSWELL}}$ and $D_{\text{trainWESAD}}$
- 21: Use Adam optimizer with learning rate lr and cross-entropy loss function
- 22: Implement early stopping and checkpoint saving to avoid overfitting
- 23: **Step 6: Model Evaluation**
- 24: Evaluate the trained BrainNet model on the test sets, $D_{\text{testSWELL}}$ and $D_{\text{testWESAD}}$
- 25: Calculate performance metrics: Accuracy, Recall, Precision, F1 score for both datasets
- 26: **Step 7: Comparison with Other Models**
- 27: Compare BrainNet's performance with other TL algorithms: InceptionV3, VGG19, MobileNet, and others 3 models.
- 28: Perform cross-validation to ensure stability and robustness
- 29: **Step 8: Statistical Validation**
- 30: Conduct t-tests and statistical analysis to validate the significance of the results between BrainNet and other models
- 31: **Step 9: Final Output**
- 32: Output the predicted stress levels and performance metrics

Algorithm 1. Proposed BrainNet approach for brain stress prediction on SWELL and WESAD datasets.

Python 3.8, TensorFlow, and Keras libraries with 8GB RAM capacity. The operating system is a 64-bit version of Windows 11, and the hardware comprises an Intel Core i7 processor from the 7th generation running at ~2.8 GHz, along with an Nvidia GTX1060 GPU. These details provide insight into the technical specifications and computational resources used throughout the study.

4.2 Model results on the Swell dataset

The first phase of the experiment involves applying TL models and the proposed BrainNet model to the Swell dataset, which includes three classes: “no stress,” “time pressure,” and “interruption.” The performance results of these learning models on the Swell dataset are summarized in Table 1 and Figure 2.

Among the evaluated models, BrainNet achieved the highest accuracy of 95.76%, along with strong precision, F1 score, and recall approximately between 91 and 92%. This model demonstrates

robust predictive capabilities across different classes. MobileNet secured the second position with an accuracy of 92.73%, and precision, F1 score, and recall ~90%, indicating its effectiveness in classification tasks. InceptionV3 and VGG19 also performed well, with accuracy scores of 91.81 and 91.19%, respectively. Though, their precision, F1 score, and recall values are slightly lower than them and varying between 84 to 90%. On the other hand, models such as Xception, EfficientNetB4, and ResNet gave reasonable accuracy in the range of 85%–87% and the corresponding precision, F1 score, and recall of 83%-85%. The research presents useful knowledge that can be obtained by comparing these DL models and shows the advantages and possible weaknesses of the models in terms of predictive functions.

4.3 Result of models on WESAD dataset

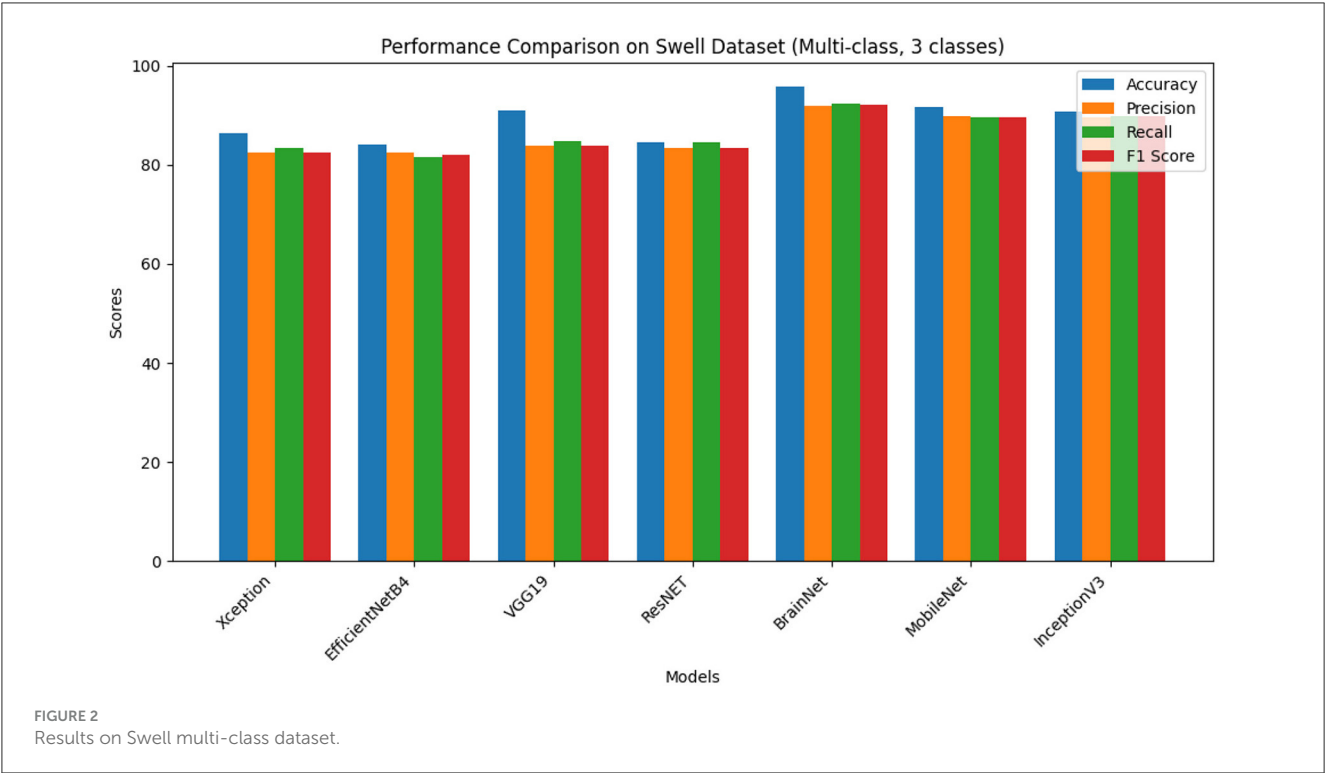
Another dataset that is employed for experiments is also referred to as WESAD dataset. This list of features consists

TABLE 1 Swell dataset (multi-class, 3 classes).

Models	Accuracy	Precision	Recall	F1 score
Xception	87.46	83.66	84.63	83.64
EfficientNetB4	85.16	83.61	82.68	83.14
VGG19	91.19	84.93	85.89	84.91
ResNET	85.64	84.47	85.65	84.58
BrainNet	95.76	91.80	92.43	92.05
MobileNet	92.73	90.98	90.67	90.76
InceptionV3	91.81	90.63	90.86	90.88

TABLE 2 Results on WESAD dataset (multi-class, three classes).

Models	Accuracy	Precision	Recall	F1 score
Xception	90.46	93.66	94.63	93.64
EfficientNetB4	88.16	93.61	92.68	93.64
VGG19	94.19	94.93	95.89	94.91
ResNET	95.64	94.47	95.65	94.58
BrainNet	98.32	97.91	98.43	98.09
MobileNet	96.73	95.98	97.67	96.59
InceptionV3	96.81	95.63	97.86	96.63



of psychological signals and acceleration signals. This dataset also contains three classes which include “baseline condition,” “amusement condition,” and “stress condition.” Peculiarities of the proposed approach and other models on the WESAD dataset are summarized in Table 2 and Figure 3.

The analysis and comparison of various DL models are shown in Table 2. Out of the presented models, Bug and the proposed BrainNet model perform the best with an accuracy of 97.32%, and the precision, F1 score, and recall values are in the range of ~97%–98% which demonstrates that this model has a strong predictive nature on the varieties of data sets. After that, the ResNet and the MobileNet have superior performance where the ResNet gets 95.64% accuracy and the MobileNet achieves 96.73%. It reaches both values of accuracy, and for the VGG19, the accuracy is 94.19% with rounded precisions, recalls and F1 scores in the range 94%–95%. Likewise, for the accuracy scores, EfficientNetB4 maintains a proportion >88% and decent precision, F1 score, and recall metric marks. On the other hand, Xception maintains an accuracy score of nearly 90% and appropriate precision, recall, and F1 score metrics which proves the model reliability in the predictive modeling task. These results actually give more information on the relative strength and possibilities of these DL models to help the researchers in determining which DL model is suitable for certain applications.

4.4 Comparison of model results on both datasets (binary classification)

From the binary classification results as indicated in the model results above, the following comparative analysis holds for both datasets. Here in the last phase of the experiment, the comparison of the learning models and the approach of the current study is performed. This research used the same two matrices: one for stress and the other for no stress. For this, we also utilized the dataset having two classes. The performance of the learning model and proposed approach is shown in Table 3 with a highlight on the result on the third topological metric.

The metrics table focuses on the efficiency of several DL models when it comes to two different datasets, namely, “Swell” and “WESAD.” Such an aggregation is seen when comparing the overall AUC claims achieved by the proposed BrainNet with respect to each shortlisted model, where the BrainNet reemerges as the best-performing model in every dataset. In the case of the Swell dataset, the proposed model reaches the level of accuracy of 97.19%, this means that the proposed model performed better than other models such as InceptionV3 with a 96.19% and ResNet of 95.81%. The precision of efficientNetB4 was 94.87%; in addition, the MobileNet is 95.61% but VGG19 and Xception model had comparatively low accuracy rates in this dataset. These results prove that BrainNet is a multipurpose and performs well on different datasets; it also shows other competitors such as InceptionV3 and EfficientNetB4. This can be useful for choosing the right model for any deep learning-oriented task.

The superior performance of XLNet over other models can be attributed to several key factors. XLNet bidirectional context allows the model to gain a deeper understanding of the data, especially in cases where temporal and sequential dependencies, such as those

found in stress-related physiological signals, are critical. XLNet also employs a generalized autoregressive pre-training technique, which enables the model to leverage the benefits of both autoregressive and autoencoding models, making it particularly suited for tasks requiring robust feature extraction and temporal modeling. In comparison with other transfer learning models used in this study (such as InceptionV3, Xception, and MobileNet), XLNet’s attention mechanism is better equipped to handle complex dependencies across time-series data, which is essential for accurately predicting stress levels. XLNet’s ability to process longer sequences of data without losing context makes it a strong fit for stress monitoring, where physiological signals evolve continuously over time. This capability leads to improved feature extraction, better capturing of subtle patterns in the data, and ultimately, enhanced classification accuracy. The model’s robustness to different datasets, as seen in the SWELL and WESAD benchmarks, further emphasizes its effectiveness in understanding and predicting brain stress.

For better clarification, this research performed a *t*-test comparison between the two best-performing models in terms of accuracy, recall, and F1 score results we obtained in Table 3. The paired *t*-test between the two models, BrainNet and InceptionV3, resulted in a *t*-statistic of ~11.65 and a *p*-value of 0.00136. Since the *p*-value is significantly <0.05, we can reject the null hypothesis, indicating that the performance difference between BrainNet and InceptionV3 is statistically significant. Therefore, BrainNet performs better than InceptionV3 on the provided metrics.

4.5 Cross-validation results

As for the evaluating method of the proposed model, K-fold cross-validation is adopted in this study. The purpose of this technique is to check whether the usage of the model is stable when compared with the other subsets of the given data. Therefore, the five-fold cross-validation is used particularly, and the summary of the results is presented in Table 4.

Analyzing the results highlighted in Table 4, it can be said that the proposed BrainNet model is efficient and accurate when tested on any of the 5-fold of the two datasets, the Swell and WESAD.

4.6 Limitations of the BrainNet framework

The proposed BrainNet model, while demonstrating high predictive accuracy for brain stress classification, has certain limitations that must be acknowledged, particularly concerning the datasets used and real-world applications. First, both the SWELL and WESAD datasets, though widely regarded as benchmark datasets, are controlled environments with limited diversity in participant demographics, stressors, and physiological responses. This could affect the model’s generalizability when applied to more varied populations or in different cultural and environmental contexts. In addition, real-world applications often involve noise and missing data, which may not be sufficiently captured in these datasets, leading to potential complications when the model is deployed in uncontrolled healthcare settings. Moreover, the datasets used predominantly focus on short-term stress

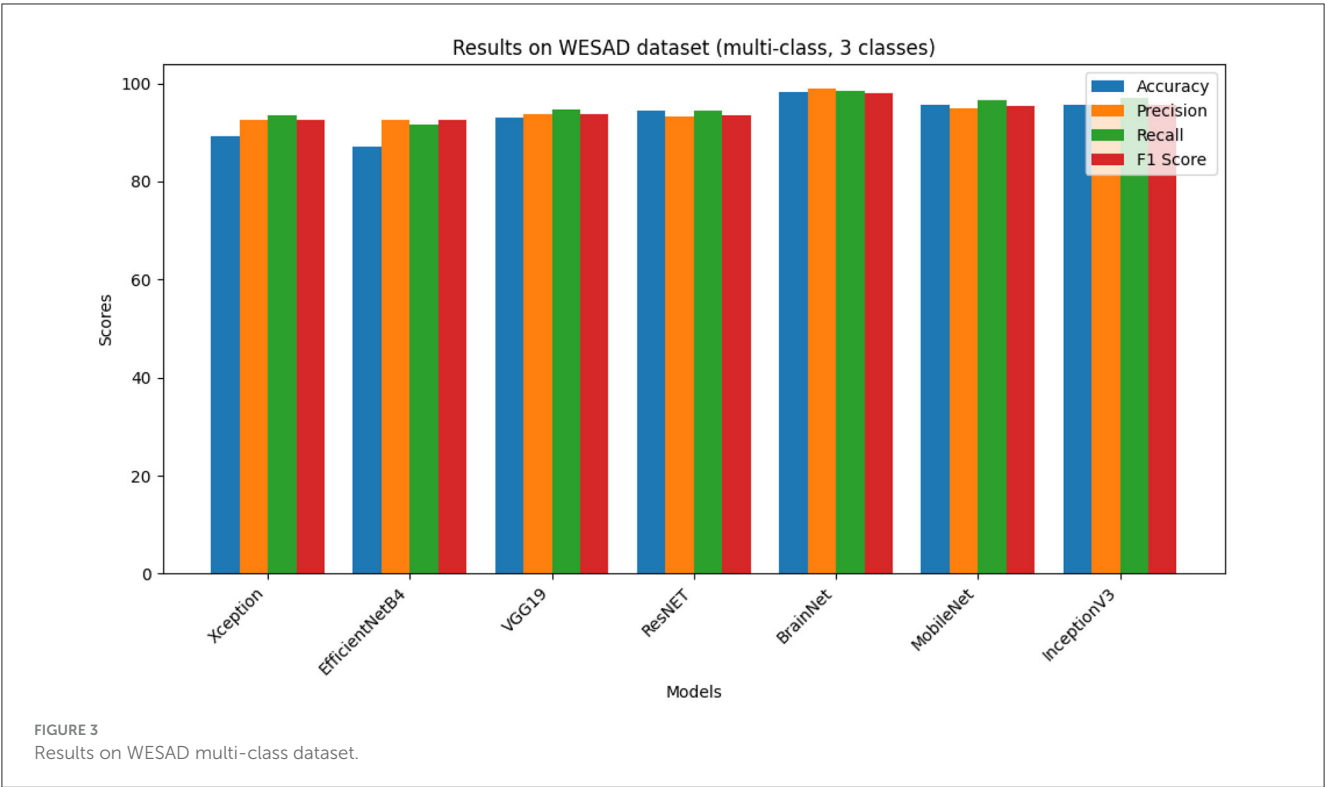


TABLE 3 Binary class, “stress” and “no stress,” classification accuracy.

Models	Accuracy		Recall	F1 score
	Swell dataset	WESAD dataset		
Xception	92.49	94.65	94.63	94.64
EfficientNetB4	94.87	98.36	92.68	96.64
VGG19	95.59	96.59	95.89	95.91
ResNET	95.81	96.68	95.65	95.58
BrainNet	97.19	99.81	98.43	98.89
MobileNet	95.61	98.68	97.67	97.89
InceptionV3	96.19	98.84	97.84	98.62

monitoring, which limits the model’s ability to predict chronic stress or adapt to the dynamic nature of stressors encountered in everyday life. The reliance on specific physiological signals like ECG and EDA may also present challenges as these signals can be influenced by factors unrelated to stress, such as physical activity or underlying health conditions, which could lead to false positives or misclassification in practical use. As a result, further study is required to ensure that the model can handle diverse and incomplete data in real-world clinical settings and to broaden the dataset to include more representative samples of the population.

5 Conclusion

Stress assessment is an important factor in maintaining a good healthy life in human beings. This stress assessment is done by

TABLE 4 K-fold cross-validation result on both datasets.

Fold for BrainNet model	Accuracy	
	Swell dataset	WESAD dataset
Fold-1	95.43	97.31
Fold-2	95.84	98.76
Fold-3	95.62	98.91
Fold-4	95.86	98.94
Fold-5	95.17	98.75
Average	95.58	98.82

employing the BrainNet model in this research study. The proposed BrainNet is tested on two popular datasets, Swell and WESAD, that contain all necessary attributes to accurately identify the human brain’s stress. It involves specific stress patterns including behavioral physiological signals for continuous stress monitoring. The proposed framework BrainNet achieves an accuracy of 95.76% when trained and tested on the Swell multi-target class dataset. The results obtained using the BrainNet model are even quite impressive when tested on the WESAD dataset. The proposed framework reaches an accuracy of 98.32% which is considered quite reliable in the domain of medical analysis. The results are even more accurate when we convert stress monitoring problem to binary target classes as stress or normal. The model accuracy reaches 99.32% for the WESAD binary classification and 97.19% for the Swell dataset binary classification problem. The results are further evaluated utilizing 5-fold cross-validation techniques. This technique helps to ensure the significance of the proposed

model on each fold of the dataset. For future endeavors, there is an envisioned development of deep ensemble learning models. Furthermore, feature fusion of multi-level signals can be used for conducting experiments with the proposed approach.

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

LX: Funding acquisition, Investigation, Project administration, Supervision, Validation, Visualization, Writing – review & editing. AH: Data curation, Resources, Software, Validation, Visualization, Writing – original draft. LM: Conceptualization, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. MU: Conceptualization, Methodology, Software, Validation, Writing – original draft. MK: Conceptualization, Project administration, Supervision, Visualization, Writing – review & editing. SAlse: Funding acquisition, Methodology, Resources, Software, Validation, Visualization, Writing – original draft. SAlsu: Methodology, Validation, Investigation, Writing – review & editing. NI: Data curation, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Writing – review & editing.

References

- Abiyev, R., Idoko, J. B., Altıparmak, H., and Tüzünkan, M. (2023). Fetal health state detection using interval type-2 fuzzy neural networks. *Diagnostics* 13:1690. doi: 10.3390/diagnostics13101690
- Acerbi, G., Rovini, E., Betti, S., Tirri, A., Rónai, J. F., Sirianni, A., et al. (2017). “A wearable system for stress detection through physiological data analysis,” in *Proceedings of the Ambient Assisted Living: Italian Forum 2016* (Ancona), 31–50. doi: 10.1007/978-3-319-54283-6_3
- Adochiei, I. R., Adochiei, F., Cepisca, C., Seritan, G., Enache, B., Argatu, F., et al. (2019). “Complex embedded system for stress quantification,” in *Proceedings of the 2019 11th International Symposium on Advanced Topics in Electrical Engineering (ATEE)* (Bucharest), 1–4. doi: 10.1109/ATEE.2019.8724892
- Akmandor, A. O., and Jha, N. K. (2017). Keep the stress away with soda: stress detection and alleviation system. *IEEE Trans.-Multi-Scale Comput. Syst.* 3, 269–282. doi: 10.1109/TMCS.2017.2703613
- Albaladejo-González, M., Ruipérez-Valiente, J. A., and Gómez Mármol, F. (2023). Evaluating different configurations of machine learning models and their transfer learning capabilities for stress detection using heart rate. *J. Ambient Intell. Humaniz. Comput.* 14, 11011–11021. doi: 10.1007/s12652-022-04365-z
- Anisman, H., and Merali, Z. (1999). Understanding stress: characteristics and caveats. *Alcohol Res. Health* 23:241.
- BrainsWay (2024). *College student suicide: Failures and potential solutions*. Burlington, MA; Jerusalem.
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Chapell, M. S., Blanding, Z. B., Silverstein, M. E., Takahashi, M., Newman, B., Gubi, A., et al. (2005). Test anxiety and academic performance in undergraduate and graduate students. *J. Educ. Psychol.* 97, 268–274. doi: 10.1037/0022-0663.97.2.268
- Chollet, F. (2017). Xception: deep learning with depthwise separable convolutions. *arXiv [Preprint]*. arXiv:1610.02357. doi: 10.48550/arXiv.1610.02357
- Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q. V., Salakhutdinov, R., et al. (2019). “Transformer-xl: attentive language models beyond a fixed-length context,” in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (Florence: ACL), 2978–2988. doi: 10.18653/v1/P19-1285
- Davis, M., and Brown, E. (2024). Transfer learning for brain stress prediction using electrodermal activity signals. *Comput. Biol. Med.* 158:104980.
- Elzeiny, S., and Qaraqe, M. (2018). “Blueprint to workplace stress detection approaches,” in *Proceedings of the 2018 International Conference on Computer and Applications (ICCA)* (Yangon). doi: 10.1109/COMAPP.2018.8460293
- Gjoreski, M., and Luštrek, M. (2017). Monitoring stress with a wrist device using context. *J. Biomed. Inform.* 73, 159–170. doi: 10.1016/j.jbi.2017.08.006
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. *arXiv [Preprint]*. arXiv:1512.03385. doi: 10.48550/arXiv.1512.03385
- 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]*. arXiv:1704.04861. doi: 10.48550/arXiv.1704.04861
- Johnson, P., and Williams, C. (2024). Hybrid lstm-svm model for brain stress prediction based on eda signals. *IEEE Trans. Neural Syst. Rehabil. Eng.* 32, 89–98.
- Koldijk, S., Neerincx, M. A., and Kraaij, W. (2018). Detecting work stress in offices by combining unobtrusive sensors. *IEEE Trans. Affect. Comput.* 9, 227–239. doi: 10.1109/TAFFC.2016.2610975
- Kraaij, P. W., Koldijk, M. S., and Sappelli, M. M. (2014). *The swell knowledge work dataset for stress and user modeling research*. doi: 10.17026/dans-x55-69zp
- Kuzu, A., and Santur, Y. (2023). Early diagnosis and classification of fetal health status from a fetal cardiotocography dataset using ensemble learning. *Diagnostics* 13:2471. doi: 10.3390/diagnostics13152471
- Lee, H., and Kim, Y. (2024). Ethical and regulatory considerations in brain stress prediction using eda signals. *Artif. Intell. Med.* 130:101854.
- Muhammad Hussain, N., Rehman, A. U., Othman, M. T. B., Zafar, J., Zafar, H., Hamam, H., et al. (2022). Accessing artificial intelligence for fetus health status using hybrid deep learning algorithm (alexnet-svm) on cardiotocographic data. *Sensors* 22:5103. doi: 10.3390/s22145103

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project (PNURSP2024R506), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Nisreen Innab would like to express sincere gratitude to AlMaarefa University, Riyadh, Saudi Arabia, for supporting 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.

- Nkurikiyeyezu, K., Yokokubo, A., and Lopez, G. (2019). The effect of person-specific biometrics in improving generic stress predictive models. *arXiv* [Preprint]. arXiv:1910.01770. doi: 10.48550/arXiv.1910.01770
- Pascoe, M. C., Hetrick, S. E., and Parker, A. G. (2020). The impact of stress on students in secondary school and higher education. *Int. J. Adolesc. Youth* 25, 104–112. doi: 10.1080/02673843.2019.1596823
- Piri, J., and Mohapatra, P. (2019). “Exploring fetal health status using an association based classification approach,” in *2019 International Conference on Information Technology (ICIT)* (Bhubaneswar: IEEE), 166–171. doi: 10.1109/ICIT48102.2019.00036
- Seo, W., Kim, N., Kim, S., Lee, C., and Park, S. M. (2019). Deep ECG-respiration network (deeper net) for recognizing mental stress. *Sensors* 19:3021. doi: 10.3390/s19133021
- Sharma, D. K. (2018). Physiology of stress and its management. *J. Med. Stud. Res.* 1, 1–5. doi: 10.24966/MSR-5657/100001
- Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv* [Preprint]. arXiv:1409.1556. doi: 10.48550/arXiv.1409.1556
- Smith, J., and Doe, J. (2024). Automated stress detection using electrodermal activity and convolutional neural networks. *J. Biomed. Signal Process.* 48, 123–130.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2015). Rethinking the inception architecture for computer vision. *arXiv* [Preprint]. arXiv:1512.00567. doi: 10.48550/arXiv.1512.00567
- Tan, M., and Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv* [Preprint]. arXiv:1905.11946. doi: 10.48550/arXiv.1905.11946
- Thanasekhar, B., Gomathy, N., Kiruthika, A., and Swarnalaxmi, S. (2019). “Machine learning based academic stress management system,” in *Proceedings of the 2019 11th International Conference on Advanced Computing (ICoAC)* (Chennai), 147–151. doi: 10.1109/ICoAC48765.2019.246831
- Walambe, R., Nayak, P., Bhardwaj, A., and Kotecha, K. (2021). Employing multimodal machine learning for stress detection. *J. Healthc. Eng.* 2021:9356452. doi: 10.1155/2021/9356452
- Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., Le, Q. V., et al. (2019). Xlnet: generalized autoregressive pretraining for language understanding. *Adv. Neural Inf. Process. Syst.* 32, 5754–5764. doi: 10.5555/3454287.3454804
- Yin, Y., and Bingi, Y. (2023). Using machine learning to classify human fetal health and analyze feature importance. *BioMedInformatics* 3, 280–298. doi: 10.3390/biomedinformatics3020019
- Zalabarria, U., Irigoyen, E., Martinez, R., Larrea, M., and Salazar-Ramirez, A. (2020). A low-cost, portable solution for stress and relaxation estimation based on a real-time fuzzy algorithm. *IEEE Access* 8, 74118–74128. doi: 10.1109/ACCESS.2020.2988348



OPEN ACCESS

EDITED BY

Miodrag Zivkovic,
Singidunum University, Serbia

REVIEWED BY

Catalin Stoean,
University of Craiova, Romania
Bosko Nikolic,
University of Belgrade, Serbia

*CORRESPONDENCE

Li Huang
✉ 316109319@qq.com
Lurui Wang
✉ wlurui@cqut.edu.cn

RECEIVED 03 October 2024

ACCEPTED 03 December 2024

PUBLISHED 18 December 2024

CITATION

Fan J, Zhao M, Huang L, Tang B, Wang L, He Z
and Peng X (2024) Multimodal sleep staging
network based on obstructive sleep apnea.
Front. Comput. Neurosci. 18:1505746.
doi: 10.3389/fncom.2024.1505746

COPYRIGHT

© 2024 Fan, Zhao, Huang, Tang, Wang, He
and Peng. 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.

Multimodal sleep staging network based on obstructive sleep apnea

Jingxin Fan^{1,2,3}, Mingfu Zhao², Li Huang^{1,3*}, Bin Tang^{2,3},
Lurui Wang^{4*}, Zhong He^{1,3} and Xiaoling Peng²

¹Central Hospital Affiliated to Chongqing University of Technology (Chongqing Seventh People's Hospital), Chongqing, China, ²School of Electrical and Electronic Engineering, Chongqing University of Technology, Chongqing, China, ³Chongqing University of Technology Sleep Medicine Collaborative Innovation Laboratory, Chongqing, China, ⁴College of Computer Science and Engineering, Chongqing University of Technology, Chongqing, China

Background: Automatic sleep staging is essential for assessing sleep quality and diagnosing sleep disorders. While previous research has achieved high classification performance, most current sleep staging networks have only been validated in healthy populations, ignoring the impact of Obstructive Sleep Apnea (OSA) on sleep stage classification. In addition, it remains challenging to effectively improve the fine-grained detection of polysomnography (PSG) and capture multi-scale transitions between sleep stages. Therefore, a more widely applicable network is needed for sleep staging.

Methods: This paper introduces MSDC-SSNet, a novel deep learning network for automatic sleep stage classification. MSDC-SSNet transforms two channels of electroencephalogram (EEG) and one channel of electrooculogram (EOG) signals into time-frequency representations to obtain feature sequences at different temporal and frequency scales. An improved Transformer encoder architecture ensures temporal consistency and effectively captures long-term dependencies in EEG and EOG signals. The Multi-Scale Feature Extraction Module (MFEM) employs convolutional layers with varying dilation rates to capture spatial patterns from fine to coarse granularity. It adaptively fuses the weights of features to enhance the robustness of the model. Finally, multiple channel data are integrated to address the heterogeneity between different modalities effectively and alleviate the impact of OSA on sleep stages.

Results: We evaluated MSDC-SSNet on three public datasets and our collection of PSG records of 17 OSA patients. It achieved an accuracy of 80.4% on the OSA dataset. It also outperformed the state-of-the-art methods in terms of accuracy, F1 score, and Cohen's Kappa coefficient on the remaining three datasets.

Conclusion: The MSDC-SSNet multi-channel sleep staging architecture proposed in this study enhances widespread system applicability by supplementing inter-channel features. It employs multi-scale attention to extract transition rules between sleep stages and effectively integrates multimodal information. Our method address the limitations of single-channel approaches, enhancing interpretability for clinical applications.

KEYWORDS

automatic sleep staging, obstructive sleep apnea, time-frequency representation, multi-scale feature extraction, transition rules

1 Introduction

Sleep is an essential biological process that is vital for both physical and mental well-being. It significantly influences numerous physiological functions, such as cognitive performance, mood regulation, and immune system function (Weber and Dan, 2016). Numerous studies have shown that the prevalence of sleep disorders has been rising in recent years. A study conducted in Australia found that 41% of women and 42% of men experience sleep issues (McArdle et al., 2020).

Sleep is a dynamic process comprising distinct stages that cycle throughout the night (Berry et al., 2017). The American Academy of Sleep Medicine (AASM) offers standardized guidelines for classifying sleep stages, which are commonly utilized in both clinical practice and research environments. It categorizes sleep into specific stages: Wakefulness (W), Rapid Eye Movement (REM) sleep, and Non-Rapid Eye Movement (NREM) sleep. NREM sleep is further classified into three stages: N1 (light sleep), N2 (moderate sleep), and N3 (deep sleep or slow-wave sleep) (Berry et al., 2012). The AASM sleep stage classification criteria are listed in Table 1.

Sleep stage classification is essential for the diagnosis and treatment of sleep disorders. Polysomnography (PSG) remains the gold standard for diagnosing these conditions and determining sleep stages. Manual sleep staging is resource-intensive, requiring specialized equipment and trained expertise. It is often conducted in a controlled laboratory environment, leading to high costs and limited accessibility (Malhotra et al., 2013). Therefore, automatic sleep staging has become a research hotspot.

OSA refers to partial or complete blockage of the upper airway during sleep, accompanied by discontinuous sleep caused by hypoxia. This disease has a high prevalence and widely affects people around the world, seriously affecting patients' sleep quality and overall health. The apnea-hypopnea index (AHI) of the entire night in PSG determines the current diagnostic criteria for OSA. Standard sleep structure includes stage N1, accounting for 2%–5% of total sleep time (TST); stage N2, accounting for 45%–55%; stage N3, accounting for 15%–25%; and REM, accounting for 20%–25%. OSA patients have a fragmented sleep structure due to frequent awakenings, with increased stage N1 and reduced stage N3 and REM.

Early deep learning models, such as those by Andreotti et al. (2018), utilized convolutional neural networks (CNNs) to extract time-frequency domain features from EEG data. Chambon et al. (2018) further refined this approach by developing a feature extractor using multiple convolutional layers to process various input channels and modalities. To fully exploit the temporal information in Electroencephalogram (EEG) signals, some studies have employed Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) networks and bi-directional LSTM (BiLSTM) networks. Michielli et al. (2019) proposed a cascaded RNN with two LSTM units. However, basic deep learning networks often encounter limitations due to the short duration of input contexts. Consequently, sequence-to-sequence methods have gained popularity, allowing for the analysis of extended sequences of PSG epochs (Phan et al., 2019b). Tang et al. (2022) developed an end-to-end deep learning model for adaptive sleep staging using ECG signals as input. Amann et al. (2020) converted multichannel

TABLE 1 Description of different sleep stages.

Stage	Name	Description
N1	Light sleep	Transition from wakefulness to sleep, characterized by slow eye movements, lower muscle activity, and the presence of theta waves in EEG
N2	True sleep	No eye movements, sleep spindles, and K-complexes appear in EEG, higher sleep threshold to disturbances, and cessation of conscious awareness of the external environment
N3	Deep sleep (NREM)	Delta waves predominate the EEG, known as slow-wave sleep (SWS), associated with memory consolidation and restorative processes
R	REM Sleep	Rapid eye movement sleep where dreaming occurs, characterized by rapid eye movements, atonia (loss of muscle tone), and beta waves similar to an awake state in EEG
W	Wakefulness	High frequency and low amplitude EEG patterns, voluntary muscle activity, and the ability to respond to stimuli. Eyes are typically open and moving, and muscle tone is present

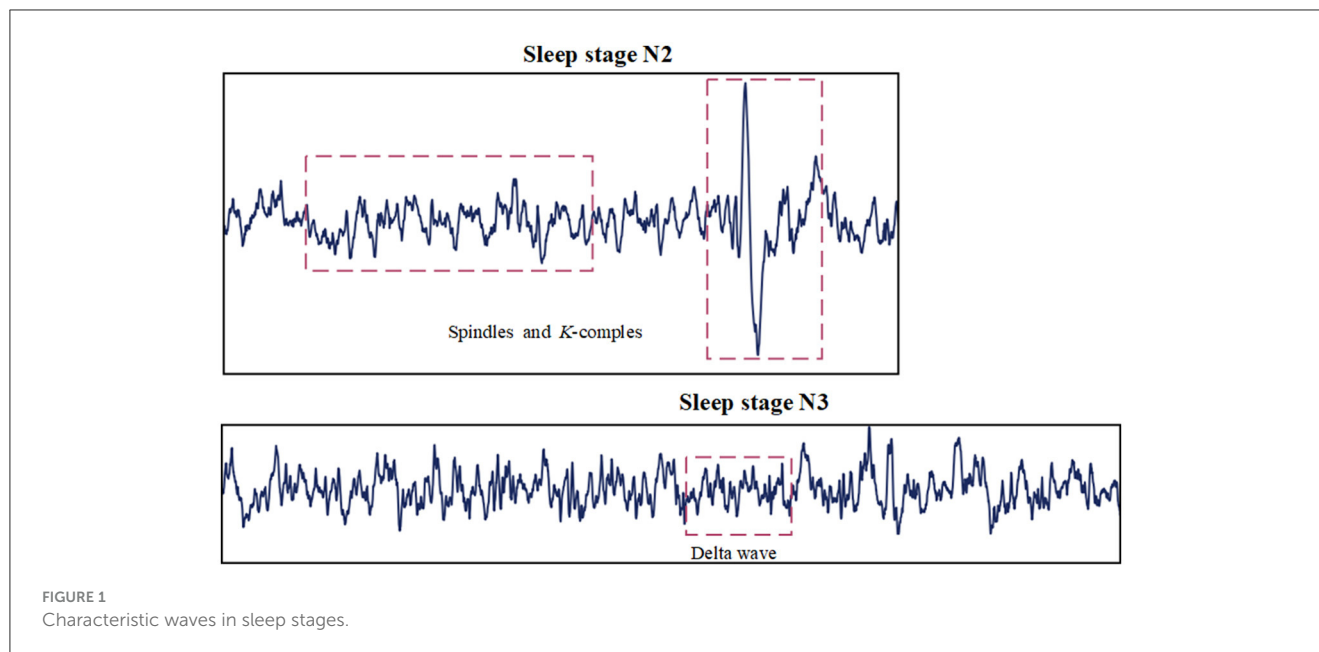
raw signals into time-frequency images for a CNN-based model, addressing sleep staging as a joint classification and prediction problem.

Current research on sleep monitoring predominantly utilizes single-channel EEG due to its simplicity, facilitating use in home-based and wearable systems (Toban et al., 2023). However, multi-channel EEG models offer enhanced robustness by incorporating multiple data sources, which proves more effective in clinical settings for accurate diagnosis and treatment of sleep disorders. Specifically, combining electrooculography (EOG) with EEG provides additional valuable information, such as detecting eye movements, which single-channel EEG alone may not reliably capture. These models align closely with expert assessments, improving credibility and interpretability.

To further enhance signal representation, recent advancements advocate for transforming one-dimensional physiological signals into more informative two-dimensional formats like STFT (Guillot and Thorey, 2021), fast Fourier transform (FFT) (Joe and Pyo, 2022), Hilbert-Huang transform (HHT) (Zhang et al., 2020) and wavelet transform (WT) (Kuo et al., 2021), borrowing techniques from image and signal processing domains. Furthermore, similar to the collaborative approaches proposed in computational research across various domains, the application of advanced data filtering and quantization methods can significantly reduce computational complexity, thereby offering potential improvements in the analysis of physiological signals (Babović et al., 2023).

Although these studies have made some progress, some problems still need to be addressed.

1. The different characteristic waves observed during various sleep stages do not have the same time scale. Characteristic waves refer to specific types of brain activity that are distinctly associated with different sleep stages. These waves vary significantly



in frequency, amplitude, and duration, making them crucial for identifying and differentiating sleep stages. As shown in [Figure 1](#), spindle Waves are bursts of oscillatory brain activity that occur predominantly during N2 sleep. They have a frequency range of about 12–16 Hz and typically last about 0.5–3 s. K-complex waves are large waves followed by a slow wave, occurring approximately every 1–1.7 s during N2 sleep. Delta Waves are characteristic of N3 sleep and have a much lower frequency range of about 0.5–4 Hz ([Aeschbach and Borbely, 1993](#)). It is worth studying how to extract features across multiple time scales and capture the complex temporal dependencies inherent in sleep signals.

2. Patients with sleep disorders exhibit significant differences in their sleep cycles compared to healthy individuals ([Chokroverty, 2010](#)). In healthy individuals, sleep progresses through well-defined cycles of NREM (N1, N2, N3) and REM stages, with relatively stable durations. OSA patients often suffer from more fragmented sleep, frequent awakenings, and transitions between stages. Disorders like insomnia and OSA can disrupt the normal progression through sleep stages, leading to shorter and more frequent REM and NREM cycles. These differences pose several challenges for automated sleep staging. Models trained on data from healthy individuals may generalize poorly to populations with OSA. The atypical waveforms and fragmented nature of disordered sleep make extracting consistent features across different scales challenging. For automated sleep staging to be clinically useful, it must achieve high accuracy across diverse patient populations, including those with OSA patients.

To address the above challenges, we present a Multi-Scale Dilated Convolution Sleep Staging Network (MSDC-SSNet). This network integrates improved Transformer encoders and multi-scale feature extraction. The model utilizes three PSG channels as inputs, including two EEG channels and one electrooculogram (EOG) signal channel. The backbone is an encoder combining causal convolution and a multi-feature extraction module (MFEM). The proposed MFEM effectively extracts different granularity

features across different frequency bands. A weighted fusion mechanism dynamically adjusts the weights of frequency features. Using a residual structure also ensures that the model can effectively learn and extract deep spatiotemporal features. Finally, a multi-channel feature fusion module integrates the features, enhancing the overall model's performance and accuracy.

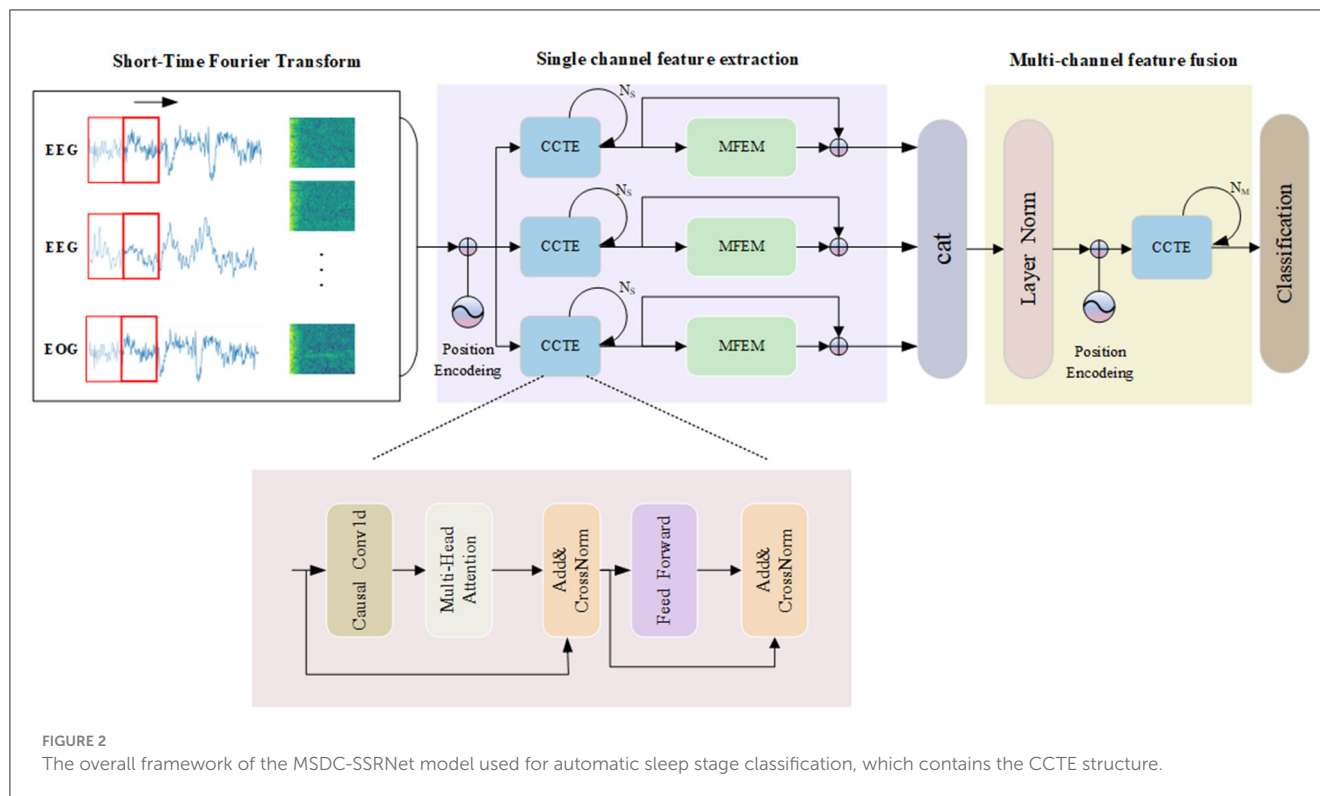
The proposed model offers several significant contributions to the field of automatic sleep staging:

1. A channel-wise Convolutional Temporal Encoder (CCTE) has been proposed. This encoder is designed to independently process and encode time series from multiple channels. We use time feature sequences to learn sleep stage transition rules and reduce the impact of OSA. It integrates causal convolution techniques and introduces a new normalization method called CrossNorm.

2. Multi-Scale Feature Extraction Module (MFEM): The MFEM that utilizes varying receptive fields to extract features across multiple scales. To enhance feature fusion, we have introduced the Multi-Scale Selection Fusion (MSF) method, significantly boosting the representational capacity of extracted features and facilitating a comprehensive analysis of sleep data.

3. Our CSPH dataset is a proprietary collection specially curated from subjects with OSA. It is designed for sleep staging applications, expanding the breadth of applications of the model and promoting the development of sleep staging.

The structure of this paper is organized as follows: Section 2 introduces the automatic sleep staging method based on OSA patients. Section 3 provides a detailed description of the experimental datasets and settings, along with the presentation of experimental results and model stability analysis. Section 4 offers an in-depth discussion of the research findings, focusing on the limitations of the current model and proposing directions for future research. Finally, Section 5 summarizes the key outcomes and contributions of this study.



2 Methods

In this section, The structure of the model is proposed. The model combines the advantages of multiscale feature extraction and causal convolution with the robustness of residual networks, aiming at the automatic staging of sleep stages.

2.1 Overview of the model

Figure 2 presents the architecture of our model, which is organized into three key segments: transforming time-frequency data into images, extracting features from individual channels, and integrating and classifying signals from multiple channels. First, the original signal is converted into a time-frequency image by STFT, and the CCTE module is utilized to extract long-range dependent features. Second, the MFEM module adaptively selects important features and fuses the inter-dependencies between single-channel features, which helps to improve the classification performance. By employing residual connections, we fuse multi-scale information with long-range dependency information. Ultimately, channel fusion is utilized to further address the heterogeneity of multimodal physiological signals. In the next section, each module is explained in detail.

2.2 Time-frequency image conversion

The model receives input in the form of time-frequency images, designed to preserve specific wave and frequency components of the original signal. According to the AASM scoring guidelines,

different physiological electrical signals contribute differently to sleep staging. EEG, EOG, EMG, and other metrics serve as foundational elements in sleep classification. From PSG files, two channels of EEG and one channel of EOG are extracted. Each channel's raw signals undergo STFT and logarithmic scale transformations to generate time-frequency images, which serve as inputs to the model.

Different PSG channels variably contribute to sleep stage classification due to the complex nature of sleep signals and the specific characteristics of each stage. EEG signals are crucial in classifying N2 and N3 stages, marked by distinct waveforms such as sleep spindles, K-complexes, and high-amplitude delta waves. These features are strong indicators of deeper sleep stages and are more readily identifiable in EEG recordings. EOG Signals are more effective in distinguishing REM sleep from N1 sleep. REM sleep is characterized by rapid eye movements, which EOG distinctly captures, whereas EEG signals in REM and N1 stages can appear similar, making EOG a critical component for accurate classification. Therefore, two EEG channels and one EOG channel were extracted from the PSG files.

2.3 Channel-wise convolutional temporal encoder

In processing EEG data, a model's comprehensive interpretation of the temporal directionality inherent within time series data is crucial. Traditional Transformer models, due to the characteristics of their self-attention mechanisms, cannot inherently handle the temporal order of time series data. The Channel-wise Convolutional Temporal Encoder (CCTE)

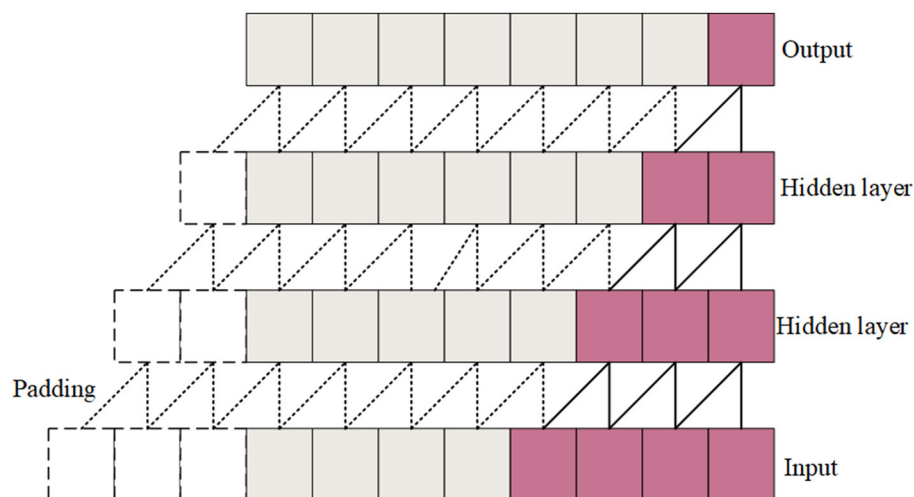


FIGURE 3
Causal convolution structural diagram.

integrates causal convolution layers, which inherently maintain the correctness of temporal sequencing by ensuring that the model processes a current data using only the preceding data, thereby effectively preventing the leakage of future information. Furthermore, drawing inspiration from the work of Tang et al. (2021), we innovatively applied the CrossNorm normalization method to the CCTE architecture to enhance the model's performance in processing multi-channel physiological signals. This enables the model to process large-scale time series data more efficiently while maintaining robust performance.

Causal convolutions are convolutional operations where each output at a specific time step depends only on the current and previous time steps, not future time steps. The causal convolution structure is shown in Figure 3. During the convolution operation, each element of the convolution kernel multiplies only with the current and previous elements of the input data. Padding is employed to ensure that the output sequence is temporally aligned with the input sequence. This property is crucial for maintaining the temporal order of the data. The Channel-wise Convolutional Temporal Encoder (CCTE) is designed to capture time-dependent features in time-frequency images. Traditionally, an Encoder-Decoder module is used for reconstruction tasks. However, since this paper focuses on classification, only the encoder is employed. The core components of the CCTE encoder include the multi-head attention layer, the position feed-forward network, and the normalization layer. By preserving the temporal order, causal convolutions ensure that the model respects the sequence of events in the EEG signal, essential for accurately identifying transitions between sleep stages. The structure of the CCTE module is shown in Figure 2.

2.3.1 Multi-head attention

Multi-head Attention (MHA) is an effective time series data model method (Devlin et al., 2018). The Transformer model

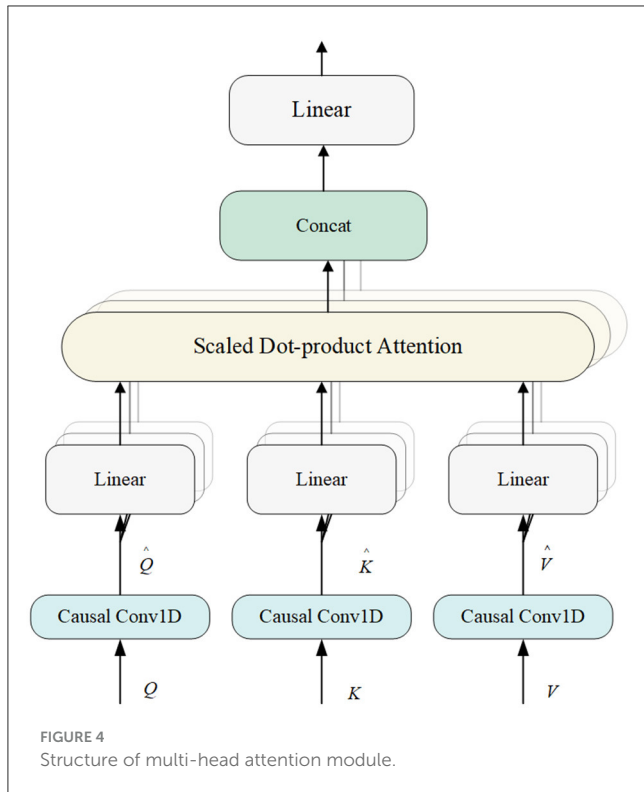
has gained popularity due to its successful handling of long-distance dependencies in sequential data. MHA employs multiple attention heads, each of which can learn information from different subspaces of the input data. This allows the model to capture a wide range of features. While a single attention head might focus on the most prominent features, multiple heads can also capture subtle details that might be missed otherwise. For sleep staging, the model can better interpret the complex and varied patterns present in EEG signals. This parallel processing increases the model's ability to capture diverse information, improving classification efficiency and effectiveness. The structure of MHA is shown in Figure 4.

The model's use of MHA combined with causal convolution ensures that only previous inputs are relied upon when computing the current output, thus maintaining the temporal order of the sequence and enhancing the model's ability to capture temporal dependencies. The combination of position encoding provides explicit information about the position of elements in the sequence, allowing the model to obtain both explicit information about the position (via position encoding) and implicit temporal dependencies (via causal convolution), which is an effective strategy for dealing with features from different frequency domains.

The MHA module begins by accepting the output from the previous module, represented as $X = \{x_1, \dots, x_N\} \in \mathbb{R}^{M \times N}$, where M is the total number of features and n is the length of x_i for $1 \leq i \leq M$. MHA utilizes three copies of X , referred to as Q , K , and V . Initially, causal convolution is applied to generate $\hat{Q}$, $\hat{K}$, and $\hat{V}$. The output from the causal convolution is then processed through the attention mechanism.

$$\text{Attention}(\hat{Q}, \hat{K}, \hat{V}) = \text{softmax}\left(\frac{\hat{Q}\hat{K}^T}{\sqrt{F}}\right) V_i \quad (1)$$

Each matrix is partitioned into H subspaces to support a multi-head attention (MHA) implementation, where the heads of each



attention result are concatenated to form the final output.

$$MHA(\hat{Q}, \hat{K}, \hat{V}) = \text{Concat}(A^1, \dots, A^H) \in \mathbb{R}^{M \times n} \quad (2)$$

2.3.2 Add and CrossNorm

The final features extracted by CCTE are generated by stacking two identical networks. The output from the previous layer is input into the next layer through a residual connection, followed by layer normalization. We utilize CrossNorm for normalization. Unlike traditional normalization methods, CrossNorm improves the model's adaptability to changes within the data by dynamically replacing the mean and standard deviation from different channels. The introduction of CrossNorm significantly improves the model's ability to capture the characteristics of different sleep stages when analyzing multi-physiological signals and time series data.

$$\frac{B - N_b}{M_a} + \frac{N_a}{M_b} \quad (3)$$

$$\frac{A - N_a}{M_b} + \frac{N_b}{M_a} \quad (4)$$

The formula exchanges the standard deviation M_a and mean N_a of channel A with the standard deviation M_b and mean N_b of channel B. Thus, A and B are cross-normalized. Each instance or channel has a unique style. During training, CrossNorm is applied for efficient style enhancement, expanding the training distribution to improve the generalization robustness under distribution changes. Effectively suppresses the impact of frequent transitions in sleep stages caused by OSA.

2.4 Feature fusion

The multichannel feature fusion module integrates feature maps from three distinct channels, concatenating them along the column axis to form a comprehensive composite feature map. This approach maximizes the preservation of each channel's unique characteristics. Since different PSG channels contain a lot of similar information, a dropout layer is introduced at the output of multiple channels to reduce the risk of overfitting of the model. Additionally, layer normalization ensures consistent data standardization throughout training, promoting accelerated convergence in the training process.

Multiple Channel-wise Convolutional Temporal Encoders (CCTEs) are employed to capture the joint features extracted from the integrated multichannel feature map. Before inputting the feature map into the encoders, positional encoding is applied to enhance the model's ability to recognize the input sequence's positional context.

2.5 Multi-scale feature extraction module

In the context of PSG signals, features across various scales play distinct roles in elucidating sleep states. Drawing inspiration from the concept of feature pyramids (Lin et al., 2017), we propose a novel module named the Multi-Scale Feature Extraction Module (MFEM) to capture multi-scale features effectively.

In the MFEM, convolutional layers with varying dilation rates enable the network to process information across local and broader spatial extents. This capability facilitates the detection of subtle physiological signals that indicate transitions between sleep stages, thereby enhancing accuracy by capturing detailed signal complexities and increasing robustness against noise and variability in signal characteristics. Additionally, to optimize multi-scale pattern recognition, the module balances and integrates features from different scales to maximize their relevance to specific sleep stages.

Specifically, the MFEM module employs four 3×3 atrous convolutions with different dilation rates to convolve the input, producing four sets of feature maps. These feature maps represent information within different frequency ranges. Subsequently, these feature maps are fused to obtain a weighted representation across multiple scales. The operation of the Multi-Scale Feature Extraction Module is illustrated as shown in Figure 5.

In the first step, for an input x , convolve it using four 3×3 convolutional kernels with dilation rates of [1, 4, 8, 16] to produce four feature maps at different frequencies, denoted as X_1, X_2, X_3 , and X_4 . Using convolutional kernels with lower dilation rates allows for capturing fine details and local features within the data. These typically correspond to high-frequency variations, such as transient spikes or rapid electroencephalographic (EEG) signal fluctuations. Conversely, employing convolutional kernels with more significant dilation rates enables the detection of broader spatial regions, thus capturing coarse-grained, global, or low-frequency features in the signal.

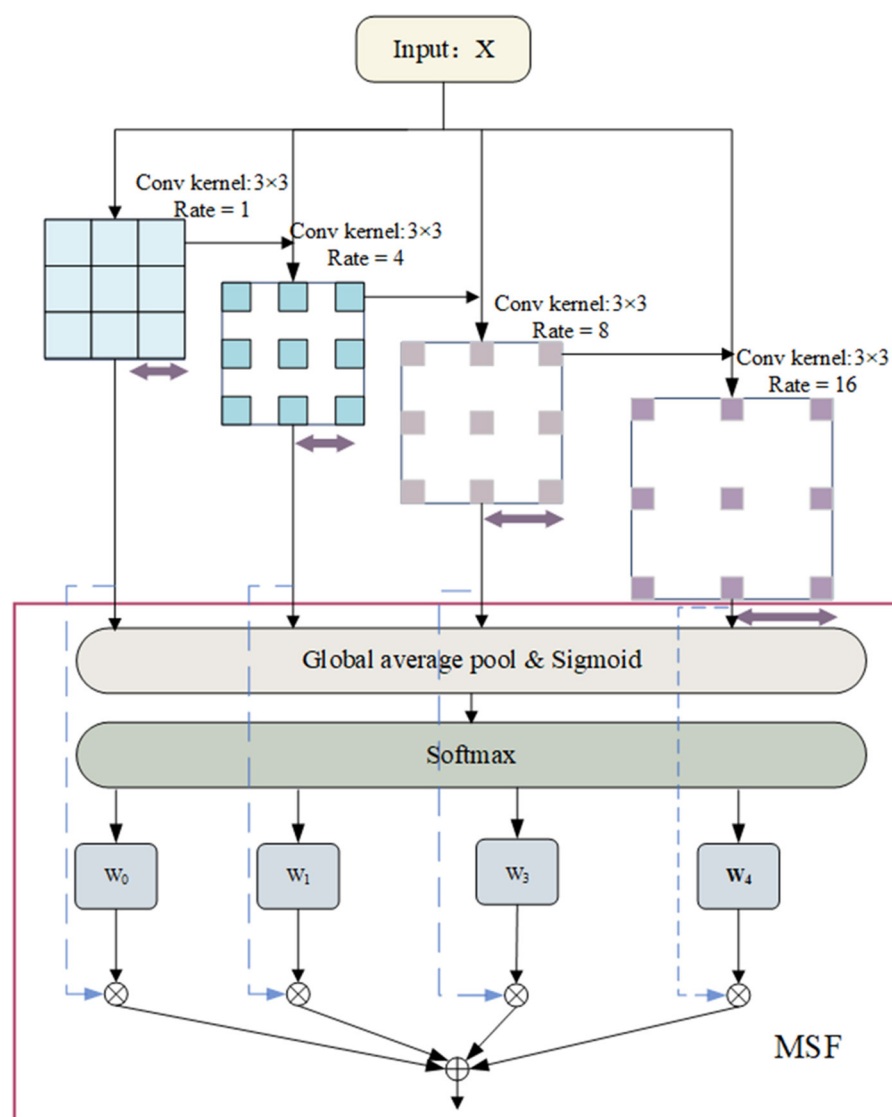


FIGURE 5
Structure of multi-scale feature extraction module.

$$X_i = \begin{cases} \text{Cov3} \times 3_{\text{rate}=1}(X) & i = 1 \\ \text{Cov3} \times 3_{\text{rate}=2(i-1)}(X + X_{i-1}) & 1 < i \leq n \end{cases} \quad (5)$$

In the second step, perform global average pooling (GAP) along the temporal dimension on X_1, X_2, X_3 , and X_4 to obtain global feature representations X_1, X_2, X_3 , and X_4 .

In our experiments, we set the number of atrous convolutions to 4. Different expansion rates enable the network to capture a broader range of spatial contextual information. This architectural design effectively enhances information extraction across various temporal and frequency dimensions by widening its scope while maintaining depth. Following feature extraction, we employ a novel fusion technique known as Multi-Scale Fusion (MSF) to integrate features Y_i from different scales. The model can adaptively emphasize more significant frequency features and

suppress less pertinent information by computing global weights for feature maps at different scales and performing a weighted fusion. Ultimately, the input features X are summed with these fused features. As depicted in Figure 5, the process begins with Global Average Pooling (GAP) being applied to multi-scale features to obtain their mean channel-wise weights (Lin et al., 2013). A Sigmoid activation function is applied to transform these weights into values between 0 and 1. Subsequently, a softmax operation normalizes the average channel weights across multi-scale features to their corresponding positions. Ultimately, the normalized weights multiply their respective features, aggregating these elements to enhance multi-scale features. Due to the combination of convolutional and attentional mechanisms, the MFEM excels in analyzing EEG time-frequency data, effectively extracting and utilizing multi-scale and multi-frequency features of the signal. This capability greatly improves the model's performance in sleep staging, facilitating more precise

TABLE 2 Detailed information on the four datasets (each sample is a 30-s calendar element).

Dataset	Subject	Sampling rate	W	N1	N2	N3	REM	Total samples
Sleep-EDF-20	20	100 HZ	9,118 21.1%	2,804 6.50%	17,799 41.30%	5,703 13.20%	7,717 17.90%	43,141
Sleep-EDF-78	78	100 HZ	66,822 34.00%	21,522 11.00%	69,132 35.20%	13,039 6.60%	25,835 13.20%	196,350
SHHS	329	125 HZ	43,619 14.3%	10,304 3.20%	142,125 43.70%	60,153 18.50%	65,953 20.30%	324,854
CSPH	17	512 HZ	4,077 21.9%	2,920 15.7%	8,273 44.4%	1,380 7.4%	1,983 10.7%	13,670

evaluations of sleep quality through a thorough analysis of EEG characteristics.

3 Results

Our analysis employs four distinct datasets to assess the model’s performance: Sleep-EDF-20, Sleep-EDF-78, Sleep Heart HealthStudy (SHHS) and Chongqing Seventh People’s Hospital (CSPH) data. These datasets are detailed in Table 2.

The Sleep-EDF-20 dataset, obtained from PhysioBank (Goldberger et al., 2000), was utilized in two distinct research studies. The initial study, known as the Sleep Cassette (SC) study, involved 20 participants aged 25–34, focusing on exploring the connection between age and sleep patterns in healthy individuals. The second study focused on the effects of temazepam on the sleep patterns of 22 Caucasian males and females who were not taking any medication (Phan et al., 2019b,a; Sokolovsky et al., 2019; Li et al., 2021). Our work utilizes the SC subset. The Sleep-EDF-20 dataset consists of polysomnographic (PSG) recordings, which include multiple physiological signals collected during participants’ sleep, such as EEG, EOG, EMG, and others. In the study, two EEG channels and one EOG channel have a sampling frequency of 100Hz. During the experiments, We used Fpz-Cz, Pz-Oz and ROC-LOC (EOG) as the input of the model.

Sleep-EDF-78 is an extension of the Sleep-EDF dataset (Goldberger et al., 2000; Kemp et al., 2000), also sourced from PhysioBank. The age range of the participants has been expanded to include individuals aged 25–101 years, encompassing a total of 78 subjects. To ensure the consistency of the experiment, the same channels as Sleep-EDF-20 were used for analysis.

The SHHS is established to examine how sleep-disordered breathing influences cardiovascular health and a range of other outcomes. It includes full-night PSG recordings involving comprehensive sleep studies with multiple physiological signals. The SHHS Visit 1 comprises 6,441 participants, all aged 40 and above. SHHS Visit 2 consists of 3,295 participants, all from Visit 1. Based on previous studies (Zhao et al., 2022; Eldele et al., 2021), we selected 329 participants with normal sleep rhythms for experimentation, using the C4-A1, C3-A2 and LOC EOG channels as model inputs.

CSPH: This dataset, provided by the Department of Sleep and Psychosomatic Medicine of Chongqing Seventh People’s Hospital, China, comprises PSG recordings from 17 subjects aged 20–60 years with OSA. The recordings were sampled at 512 Hz, and each subject underwent manual sleep stage scoring by three sleep specialists following AASM criteria. The PSG recording channels

included F4-A1, C4-A1, O2-A1, F3-A2, C3-A2, O1-A2, along with electrooculograms EOGL and EOGR. For analysis, inputs were derived from F4-A1, F3-A2, and EOGL channels. All three datasets employ the AASM sleep scoring standards.

These datasets cover a broad range of subjects, including healthy individuals, those with sleep disorders, and participants across a wide age range, from young adults to older individuals. They provide a diverse set of conditions and scenarios, making the model robust across various sleep patterns.

3.1 Experimental setup

A 30-second segment (epoch) of PSG data was sampled for the analysis. The Short-Time Fourier Transform (STFT) is applied using a 2-s Hamming window with 50% overlap. The FFT is computed with 256 points, providing a frequency resolution adequate for sleep analysis. The resultant spectrum is then log-scaled. The resulting time-frequency representation, denoted as $S \in \mathbb{R}^{T \times F}$, consists of $F = 128$ frequency bins and $T = 29$ time points. This normalized representation is subsequently utilized as the model’s input.

In our CCTE encoder, the Multi-Head Attention (MHA) utilizes eight heads and 150 feedforward hidden units. The CCTE modules at the model input and output use different numbers of encoders, $N_s = 8, N_m = 4$ respectively. Throughout the entire CCTE model, including the self-attention layers, feedforward layers, and fully connected (FC) layers, a uniform dropout rate of 0.1 is applied.

To address the issue of a limited number of subjects, we employed K-fold cross-validation to train the model on four datasets. The values of K for the Sleep-EDF-20, Sleep-EDF-78, SHHS, and CSPH datasets were set to 20, 10, 10, and 10, respectively. Although some datasets, such as Sleep-EDF-20 and CSPH, have a smaller sample size, K-fold cross-validation effectively improved the model’s generalization ability and reduced the risk of overfitting through repeated training and validation. Meanwhile, the larger dataset (SHHS) further enhanced the model’s stability and robustness, ensuring effective performance across all datasets. The training objective utilized was the cross-entropy loss function, which is commonly used in classification tasks. We used the AdamW (Loshchilov and Hutter, 2017) optimizer, which is more effective in handling weight decay, with a learning rate set to 5×10^{-5} . Additionally, during the model training process, we employed early stopping, which involves halting training when the performance on the validation set no longer improves, in order to prevent the model from overfitting to the training set.

3.2 Evaluation metrics

The model's overall performance is assessed using three key metrics: accuracy (*ACC*), macro-average F1 score (*MF1*), and Cohen's Kappa (κ). The *MF1* is calculated as the arithmetic mean of the F1 scores for the five sleep stages. Precision (*Pre*), recall (*Rec*), and F1-score (*F1*) are used to assess each class individually. The overall accuracy (*ACC*) and macro-average F1 score (*MF1*) are defined as follows:

$$MF1 = \frac{\sum_{c=1}^C F1_c}{C} \quad (6)$$

$$ACC = \frac{\sum_{c=1}^C TP_c}{M} \quad (7)$$

For each class c , the within-class F1-score is denoted as $F1_c$. There are C distinct sleep stage categories. For each category c , TP_c represents the true positives of that category. Additionally, M represents the total number of EEG epochs.

3.3 Experimental scoring results

Experimental scoring results are presented in Table 3, using confusion matrices to display the performance of the model. In these matrices, rows represent the actual results, while columns represent the predicted results. Bold numbers within the matrices highlight epochs correctly classified by the model. Evaluation metrics for each category are provided on the right side of the tables, with optimal values emphasized in bold.

According to the evaluation results of three healthy population datasets, the accuracy of the Wake stage can reach more than 93%. The indicators of the N1 stage are lower than those of the W, N2, N3, REM, and other stages, which may be related to the small number of occurrences of the N1 stage in the data set. Misclassifications frequently occur among the sleep stages, with the W stage often being mistaken for the N1, N2, and REM stages. Similarly, the N1 stage is commonly misclassified as W, N2, or REM, while the REM stage is often confused with N1 and N2. Additionally, the N3 stage is primarily confused with the N2 stage.

For OSA patients in CSPH, the accuracy for the N1 stage can reach 62.1%, while the accuracies for the W stage and N2 stage exceed 85%. However, the model's overall performance is generally lower than that of healthy subjects, reflecting the interference of OSA on sleep staging.

Figure 6 depicts the ground truth and predicted hypnograms for subject SC4001E0 from the Sleep-EDF-20 dataset to further illustrate the findings. The close resemblance between the predicted and true hypnograms demonstrates the model's accuracy. However, the transition into the REM stage exhibits a higher error rate. This primarily arises from the increased variability in EEG signals during transitions and the substantial similarity between mixed-frequency EEGs.

In Figure 7, we present the accuracy and loss curves during the training process for the Sleep-EDF-20 dataset, explicitly focusing on fold 6 selected at random. It is observed that our model can rapidly converge and stabilize at a fixed value soon after

training initiation. The accuracy continually improves, and the loss consistently decreases. Similarly, validation sets accuracy and loss values to stabilize, underscoring the model's efficacy in mitigating overfitting.

3.4 Performance comparison

We compared our MSDC-SSRNet with previous state-of-the-art methods, evaluating overall accuracy, Cohen's kappa (κ), and MF1 across four datasets, along with the F1-score for each sleep stage. The results are presented in Table 4. Our MSDC-SSRNet exhibits significantly better performance than other models based on the experimental outcomes. On the Sleep-EDF-20 dataset, our model showed improvements of 0.9% in accuracy, 1.2% in kappa, and 2.1% in MF1 over the SleepViTransformer (Peng et al., 2023). It also outperformed the transformer-based multichannel model MultiChannelSleepNet (Dai et al., 2023), with increases of 2.2% in accuracy, 3.0% in kappa, and 3.3% in MF1.

To demonstrate the high accuracy of our method on the CSPH dataset, we compare it with four state-of-the-art methods, namely: (1) AttnSleep (Eldele et al., 2021); (2) SleepyPyCo (Lee et al., 2024); (3) MultiChannelSleepNet (Dai et al., 2023); (4) SalientSleepNet (Liang et al., 2023); in the CSPH dataset, the overall performance of MSDC-SSRNet also surpasses that of other networks. It performed well in both healthy subjects and OSA patients, demonstrating its robustness in handling complex datasets with varied sleep conditions. While SleepPyCo (Lee et al., 2024) excels on simpler datasets such as Sleep-EDF-20 and Sleep-EDF-78, its performance declines when dealing with the more complex characteristics of the CSPH dataset. In addition, MSDC-SSRNet performs well in distinguishing the easily confused N2 and N3. Since there is a certain overlap in the transition period between the N2 and N3 stages, such as the overlapping delta waves (0.5–4 Hz) in the N3 stage and the sleep spindle waveform in the N2 stage, the distinction between the two is blurred. MSDC-SSRNet effectively helps doctors distinguish the N2 and N3 stages more accurately through auxiliary feature extraction and precise modeling.

Unlike SeqSleepNet (Phan et al., 2019b), which predicts the middle epoch using a recurrent architecture with three epochs as input, thereby slowing down the training process, the AttnSleep (Eldele et al., 2021) model adopts multi-scale feature extraction through varied convolutional kernel sizes and strides on the same input. In contrast, our MFEM utilizes dilated convolutions to enlarge the receptive field without significantly increasing the parameters, thereby enhancing local feature representation. This capability is crucial for sleep stage analysis, which requires detecting features at different time scales. Moreover, while AttnSleep (Eldele et al., 2021) shows improved F1-scores in certain stages like N2 and N3 compared to other models like SeqSleepNet (Phan et al., 2019b), it still falls short of MSDC-SSRNet in terms of overall accuracy and generalization across diverse datasets. MSDC-SSRNet reduces the heterogeneity between different modalities and data, proving to be a more versatile and efficient model in both accuracy and consistency.

TABLE 3 Confusion matrices for different datasets.

Dataset	Predicted					Per-class metrics		
	W	N1	N2	N3	REM	PR	PE	F1
Sleep-EDF-20	8,186	283	75	31	109	94.7	94.2	94.4
	293	1328	477	12	561	63.4	49.7	55.7
	89	275	15,361	576	650	91.3	90.6	90.9
	9	1	503	4,915	3	88.7	90.4	89.6
	64	207	408	4	6,666	83.4	90.7	86.9
Sleep-EDF-78	61,287	2,366	446	75	349	94.3	94.9	94.7
	2,910	11,441	6,342	61	2,132	63.2	49.9	55.9
	3,799	2,766	72,533	365	2,096	86.0	89.7	87.8
	31	7	2,195	15,451	2	82.6	87.3	84.9
	328	1,503	2,763	32	26,450	85.2	85.1	85.2
SHHS	42,853	1,030	1,317	171	848	93.4	92.7	93.0
	1,488	5,547	260	113	2,896	55.9	53.8	54.8
	517	36	25,041	4,264	267	81.5	83.1	82.3
	45	1,008	4,072	50,673	4,355	86.1	84.2	85.2
	998	2,296	4	3,633	59,022	87.5	89.5	88.5
CSPH	3,267	254	124	6	55	85.6	88.2	86.9
	342	1,510	619	18	165	62.1	56.9	59.4
	130	515	6,347	234	295	85.4	84.4	84.9
	5	5	141	1,104	0	81.0	88.0	84.3
	73	147	197	1	1,385	72.9	76.8	74.8

Bold numbers in the table represent the correct sample counts for each category.

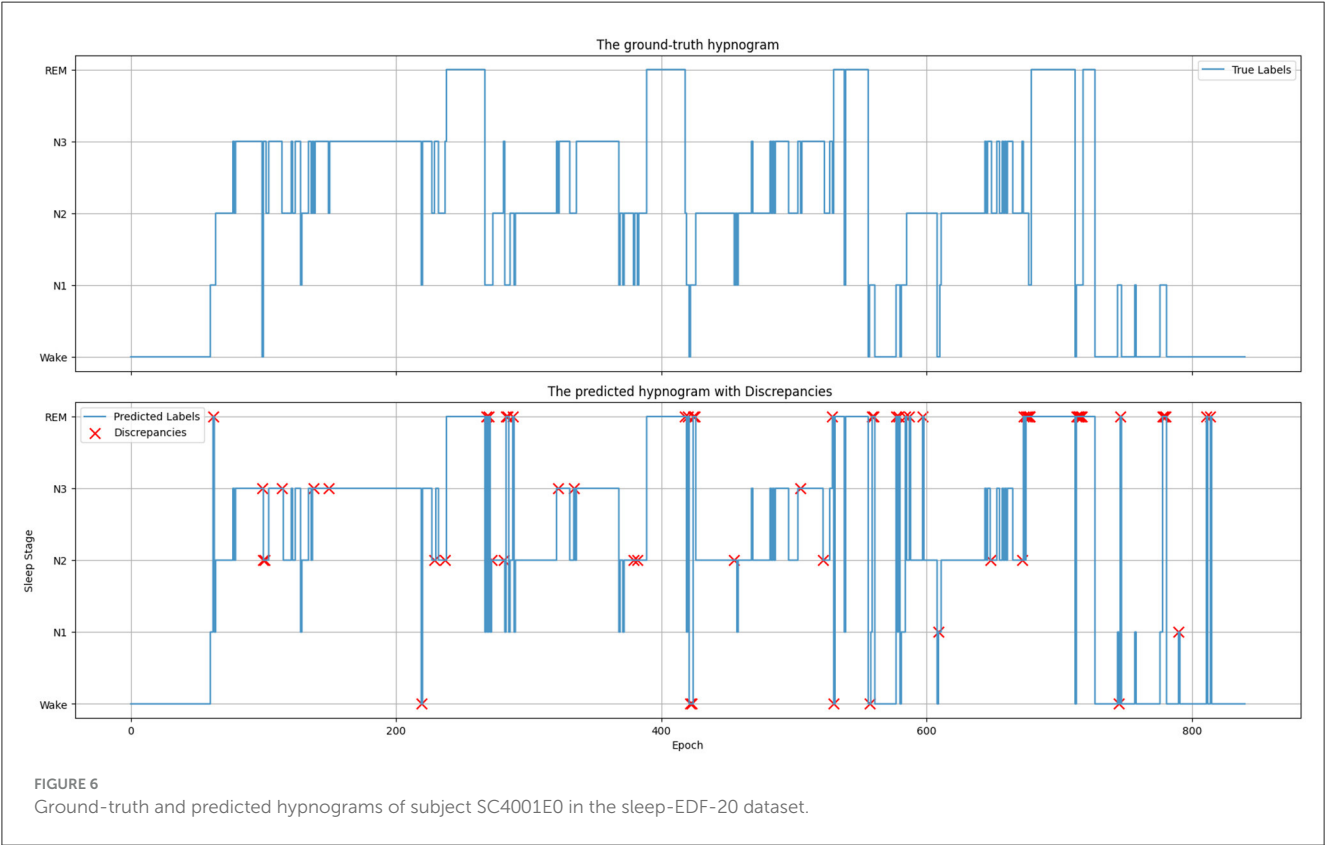


FIGURE 6 Ground-truth and predicted hypnograms of subject SC4001E0 in the sleep-EDF-20 dataset.

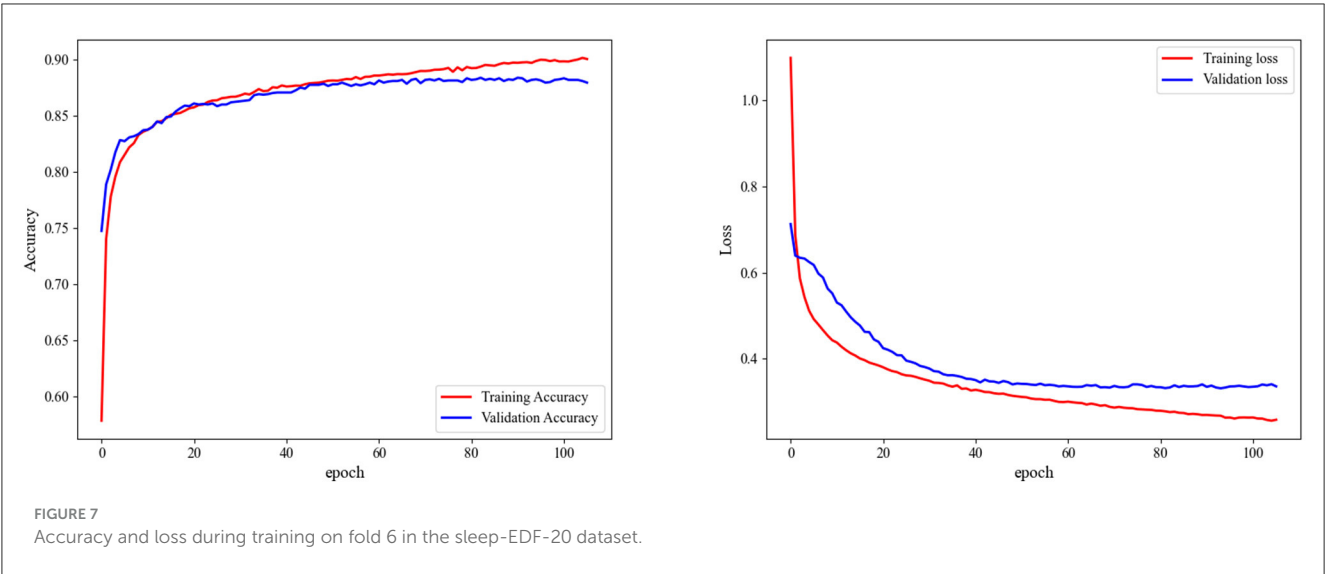


TABLE 4 Comparison of sleep staging performance with previous studies across four experimental datasets.

Dataset	System	Over metrics			Per-class F1-score				
		ACC	Kappa	MF1	W	N1	N2	N3	REM
Sleep-EDF-20	MSDC-SSRNet	88.7	84.6	83.5	94.5	55.7	90.9	89.6	86.9
	SleepViTransformer (Peng et al., 2023)	87.8	83.4	81.5	93.8	48.4	89.2	88.4	87.9
	SleepPyCo (Lee et al., 2024)	86.2	80.1	81	90.6	47.3	88.8	87.4	86.6
	MultiChannelSleepNet (Dai et al., 2023)	86.5	81.6	80.3	92.6	47	89.5	88.3	83.8
	SeqSleepNet (Phan et al., 2019b)	85.2	79	79.6	–	–	–	–	–
	SleepEEGNet (Mousavi et al., 2019)	84.3	79	79.7	89.2	52.2	86.8	85.1	85
	DeepSleepNet (Supratak et al., 2017)	81.9	76	76.6	86.7	45.5	85.1	83.3	82.6
Sleep-EDF-78	MSDC-SSRNet	86.2	81.2	81.7	94.7	55.9	87.8	84.9	85.2
	SleepPyCo (Lee et al., 2024)	84.6	79	79.1	93.5	50.4	86.5	80.5	84.2
	SeqSleepNet (Phan et al., 2019b)	82.6	76	76.4	92.2	47.8	84.9	77.2	79.9
	TinySleepNet (Supratak and Guo, 2020)	83.1	77.1	78.1	92.8	51	85.3	81.1	80.3
	SleepTransformer (Phan et al., 2022)	81.4	74.3	74.3	91.7	40.4	84.3	77.9	77.2
	AttnSleep (Eldele et al., 2021)	81.3	74	75.1	92	42	85	82.1	74.1
	SleepEEGNet (Mousavi et al., 2019)	80	73	73.6	91.7	44.1	82.5	73.5	76.1
	MultiChannelSleepNet (Dai et al., 2023)	84.9	78.9	79.4	94	52.8	86.3	81.5	82.6
SHHS	MSDC-SSRNet	86.7	79.3	80.8	93	54.8	82.3	85.2	88.5
	AttnSleep (Eldele et al., 2021)	84.2	78	75.3	86.7	33.2	87.1	87.1	82.1
	SeqSleepNet (Phan et al., 2019b)	86.5	81	78.5	–	–	–	–	–
CSPH	MSDC-SSRNet	80.4	72.6	78.1	86.9	59.4	84.9	84.3	74.8
	AttnSleep (Eldele et al., 2021)	79.4	71.6	77.6	86	60.7	84.3	82.5	74.2
	SleepPyCo (Lee et al., 2024)	78.3	70.4	76.5	85.3	58.2	83.6	83.2	72.4
	MultiChannelSleepNet (Dai et al., 2023)	77.6	68.8	75.9	84.7	57.7	82.8	82.6	71.7
	SalientSleepNet (Liang et al., 2023)	77.3	68.9	76.5	84.4	60.1	82.5	83.5	72.1

Best metric values are marked in boldface.

TABLE 5 Ablation experiment results for sleep-EDF-20 and CSPH datasets.

Ablation experiment	Sleep-EDF-20 metrics			Sleep-EDF-20 Per-class F1 score				
	ACC	MF1	Kappa	W	N1	N2	N3	REM
BL	86.5	80.3	81.6	92.9	47.0	89.5	88.3	83.8
BL + MFEM	86.8	81.5	82.0	93.1	51.1	89.6	88.7	84.8
BL + CCTE	87.9	82.0	83.4	93.9	50.9	90.2	89.1	86.0
MSDC-SSRNet	88.7	83.6	84.6	94.5	55.7	90.9	89.6	86.9

Ablation experiment	CSPH metrics			CSPH Per-class F1 score				
	ACC	MF1	Kappa	W	N1	N2	N3	REM
BL	75.3	74.5	68.1	82.4	53.6	80.7	80.3	70.5
BL + MFEM	77.5	76.0	68.8	84.7	58.6	82.8	82.7	71.8
BL + CCTE	79.5	77.3	71.2	85.8	57.7	83.9	83.5	73.6
MSDC-SSRNet	80.4	78.1	72.6	86.9	59.4	84.9	84.3	74.8

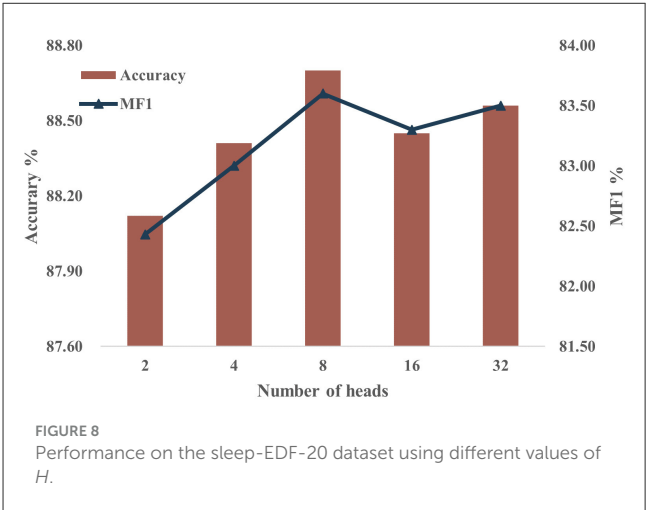
3.5 Ablation experiments

As depicted in Table 5, we conducted ablation experiments on the Sleep-EDF-20 and CSPH dataset to assess the efficacy of various modules. Comparing BL, BL + CCTE, BL + MFEM, and our MSDC-SSRNet model reveals improvements across all metrics with each module’s inclusion.

In the CSPH dataset, the CCTE module can significantly enhance classification performance, with overall improvements in ACC, MF1, and Kappa by 0.9%, 0.8% and 1.4%, respectively. F1 scores for each sleep stage also improved. We use the basic transformer as the baseline. Comparing it to the second variant, BL + MFE, we conclude that CCTE is essential for capturing frequent sleep stage transition features. However, MFEM is more effective in distinguishing the N1 stage, as the multi-scale feature extraction method allows the model to focus on finer features at lower or higher frequencies, thereby increasing overall sensitivity and reducing the impact of OSA on the model. In the Sleep-EDF-20 dataset, the final model shows an improvement in F1 scores of 8.6% for the N1 stage and 3.3% for the REM stage compared to the baseline (BL). According to the American Academy of Sleep Medicine (AASM) rules, especially in the N1 and REM sleep stages, the EEG features share similar low-amplitude, multi-frequency (LAMF) activities, making the features between these stages indistinct. Addressing this issue, our model framework can more effectively differentiate features of various sleep stages, particularly distinguishing between the N1 and REM stages.

3.6 Sensitivity analysis

Multi-head attention (MHA) is a pivotal element in our model, necessitating a sensitivity analysis regarding the number of heads employed. Given that the number of heads must be a divisor of the feature dimension $F = 128$, we set H to 2, 4, 8, 16, and 32 for the experiments, while maintaining constant values for the other parameters. Figure 8 shows the accuracy and MF1 scores of



the model on the Sleep-EDF-20 dataset with different numbers of heads. The results show that model performance shows slight improvement with an increase in H . However, beyond a certain point, further increments in H lead to diminishing returns. This suggests that expanding the number of heads enhances feature capture initially, yet excessively dividing attention may reduce the per-head feature resolution. We select $H = 8$ as optimal for our model configuration based on these experimental findings.

In both the model’s feature extraction and fusion processes, the CCTE encoder is utilized, so choosing an appropriate amount of encoders is also crucial. To further investigate the impact of the number of encoders N_s in single-channel feature extraction and N_m in multi-channel fusion, we keep other experimental parameters constant and use the Sleep-EDF-20 dataset. Initially, we fix N_s at 4, and repeat experiments with N_m values from {2, 4, 6, 8}, then fix N_m at 4, and repeat experiments with N_s values from {2, 4, 6, 8}. Based on the results shown in Table 6, changing the number of encoders does not significantly affect the model’s overall performance. However, increasing N_s enhances the model’s depth, improving its ability to capture features.

TABLE 6 Performance on the Sleep-EDF-20 dataset using different amounts of N_s and N_m .

N_s	N_m	Accuracy	MF1
4	2	88.42	82.42
4	4	88.56	83.18
4	6	88.59	83.51
4	8	88.47	83.03
2	4	88.25	82.39
4	4	88.56	83.18
6	4	88.69	83.43
8	4	88.79	83.60

4 Discussion

MSDC-SSRNet uses multi-channel data for sleep staging tasks. Through ablation experiments and model stability analysis, each module in MSDC-SSRNet assists with sleep staging. The model performance is improved by capturing characteristic waves using multi-scale feature extraction and channel attention. While single-channel sleep staging algorithms are commonly used for portable home sleep monitoring, multi-channel data provides a more comprehensive view of sleep states. This comprehensive view aids the model in detecting subtler differences in sleep stages, which are more readily recognized by sleep physicians and offer better interpretability than single-channel systems.

As shown in Table 6, except for MultiChannelSleepNet (Dai et al., 2023), the staging performance of other single-channel models is inferior to MSDC-SSRNet. In a multi-channel framework, additional channels mitigate disruptions or poor signal quality in one channel, enhancing overall system robustness. In addition, the algorithm is applied to the self-built dataset CSPH. Unlike the public datasets, the subjects of this dataset suffer from OSA. The CSPH dataset is characterized by frequent sleep stage transitions and fragmented sleep cycles, which makes the sleep staging task challenging. Despite these difficulties, MSDC-SSRNet still performs well.

The CCTE captures long-range dependencies and enhances the importance of position information in the time-frequency domain. The MFEM uses different receptive fields to enhance the contribution of characteristic waves to sleep stages. The multi-scale attention layer integrates features with different weights, ensuring the preservation of multi-scale sleep transition rules. The model is able to characterize typical sleep stage features and distinguish them from other stages. EEG activity is highly dynamic, and multi-scale analysis can adapt to these changes, extracting significant features at different time scales to effectively capture short-term and long-term brain activity patterns. Compared to single-scale feature capture methods, the multi-scale approach provides a more stable feature representation, contributing to model generalizability and practical application.

Future research could address several limitations identified in this study. First, the data imbalance problem in the N1 stage still

needs to be addressed. Additionally, our current model does not account for other relevant factors, such as age and gender, which could influence the study outcomes. Addressing these limitations in future research could further enhance the model's accuracy and applicability.

5 Conclusions

In this study, we introduced MSDC-SSRNet, a sleep staging model leveraging multi-scale dilated convolutions. It performs well on both healthy subjects and OSA subjects. In experiments with OSA subjects, the accuracy reaches 80.4%. This model utilizes the Channel-wise Convolutional Temporal Encoder (CCTE) and the Multi-Scale Feature Extraction Module (MFEM) for effective feature capture. The CCTE encoder employs a multi-head attention mechanism to capture long-range dependencies in the data. Additionally, we integrated CrossNorm, a novel normalization technique within CCTE, which enhances training data diversity by exchanging channel means and variances across feature maps. This ensures robust performance across diverse environmental and conditional data settings. The MFEM operates by capturing signals across a spectrum of frequencies from low to high, employing multi-scale feature extraction in the spatial domain. This module focuses on spatial feature extraction and adeptly captures various frequency components. This is particularly significant for EEG signals, as different frequency waveforms (such as δ , θ , α , β , and γ waves) exhibit distinct frequency characteristics.

Our model's effectiveness has been validated through comparisons with advanced models and extensive ablation experiments. Moreover, it provides more accurate predictions and classifications on datasets with specific clinical characteristics. Furthermore, we conducted a sensitivity analysis by varying the number of attention heads in the CCTE encoder for single-channel feature extraction and multi-channel fusion. This analysis demonstrated the model's stability and consistent performance under different parameter settings. The model's robust performance and adaptability to various configurations suggest its strong potential for real-world applications, particularly in clinical settings. Its high accuracy in classifying sleep stages for patients with obstructive sleep apnea makes it well-suited for deployment in home-based monitoring systems. Such systems could offer continuous, real-time sleep tracking, which would enhance patient convenience and accessibility while reducing the need for in-lab polysomnography. The model's ability to generalize across diverse patient populations further underscores its practical utility and potential for widespread implementation in both clinical and research environments.

Data availability statement

The datasets presented in this article are not readily available because protection of patient privacy, ensuring data security and ethical use. Requests to access the datasets should be directed to Jingxin Fan, fanjingxin2000@163.com.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

JF: Data curation, Methodology, Resources, Supervision, Writing – original draft. MZ: Data curation, Writing – original draft. LH: Resources, Validation, Writing – review & editing. BT: Data curation, Visualization, Writing – review & editing. LW: Methodology, Supervision, Validation, Writing – review & editing. ZH: Data curation, Validation, Writing – review & editing. XP: Validation, 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 in part by National Key Research and Development Program of China (2020YFC1522500); Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJQN202401127 to LH); Chongqing

Graduate Research and Innovation Project (CYS23666 to JF); and The Eyas Program of the Youth Innovative Talents Cultivation in Chongqing, grant number (CY240907 to LW).

Acknowledgments

We thank our colleagues for obtaining data from Chongqing Seventh People's Hospital.

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

- Aeschbach, D., and Borbely, A. A. (1993). All-night dynamics of the human sleep EEG. *J. Sleep Res.* 2, 70–81. doi: 10.1111/j.1365-2869.1993.tb00065.x
- Amann, J., Blasimme, A., Vayena, E., Frey, D., Madai, V. I., Consortium, P., et al. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med. Inform. Decis. Mak.* 20, 1–9. doi: 10.1186/s12911-020-01332-6
- Andreotti, F., Phan, H., Cooray, N., Lo, C., Hu, M. T., and De Vos, M. (2018). "Multichannel sleep stage classification and transfer learning using convolutional neural networks," in *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (Honolulu, HI: IEEE), 171–174. doi: 10.1109/EMBC.2018.8512214
- Babović, Z., Bajat, B., Dokić, V., Dordević, F., Drašković, D., Filipović, N., et al. (2023). Research in computing-intensive simulations for nature-oriented civil-engineering and related scientific fields, using machine learning and big data: an overview of open problems. *J. Big Data* 10:73. doi: 10.1186/s40537-023-00731-6
- Berry, R. B., Brooks, R., Gamaldo, C., Harding, S. M., Lloyd, R. M., Quan, S. F., et al. (2017). Aasm scoring manual updates for 2017 (version 2.4). *J. Clin. Sleep Med.* 13, 665–666. doi: 10.5664/jcsm.6576
- Berry, R. B., Brooks, R., Gamaldo, C. E., Harding, S. M., Marcus, C., Vaughn, B. V., et al. (2012). "The AASM manual for the scoring of sleep and associated events," *Rules, Terminology and Technical Specifications, Vol. 7* (Darien, IL: American Academy of Sleep Medicine), 176.
- Chambon, S., Galtier, M. N., Arnal, P. J., Wainrib, G., and Gramfort, A. (2018). A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 758–769. doi: 10.1109/TNSRE.2018.2813138
- Chokroverty, S. (2010). Overview of sleep & sleep disorders. *Indian J. Med. Res.* 131, 126–140.
- Dai, Y., Li, X., Liang, S., Wang, L., Duan, Q., Yang, H., et al. (2023). Multichannelsleepnet: a transformer-based model for automatic sleep stage classification with PSG. *IEEE J. Biomed. Health Inform.* 27, 4204–4215. doi: 10.1109/JBHI.2023.3284160
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv [Preprint]*. arXiv:1810.04805. doi: 10.48550/arXiv.1810.04805
- Eldele, E., Chen, Z., Liu, C., Wu, M., Kwok, C.-K., Li, X., et al. (2021). An attention-based deep learning approach for sleep stage classification with single-channel EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 809–818. doi: 10.1109/TNSRE.2021.3076234
- Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., et al. (2000). Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation* 101, e215–e220. doi: 10.1161/01.CIR.101.23.e215
- Guillot, A., and Thorey, V. (2021). Robustsleepnet: transfer learning for automated sleep staging at scale. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 1441–1451. doi: 10.1109/TNSRE.2021.3098968
- Joe, M.-J., and Pyo, S.-C. (2022). Classification of sleep stage with biosignal images using convolutional neural networks. *Appl. Sci.* 12:3028. doi: 10.3390/app12063028
- Kemp, B., Zwinderman, A. H., Tuk, B., Kamphuisen, H. A., and Obery, J. J. (2000). Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. *IEEE Trans. Biomed. Eng.* 47, 1185–1194. doi: 10.1109/10.867928
- Kuo, C.-E., Chen, G.-T., and Liao, P.-Y. (2021). An EEG spectrogram-based automatic sleep stage scoring method via data augmentation, ensemble convolution neural network, and expert knowledge. *Biomed. Signal Process. Control* 70:102981. doi: 10.1016/j.bspc.2021.102981
- Lee, S., Yu, Y., Back, S., Seo, H., and Lee, K. (2024). Sleepyco: automatic sleep scoring with feature pyramid and contrastive learning. *Expert Syst. Appl.* 240:122551. doi: 10.1016/j.eswa.2023.122551
- Li, F., Yan, R., Mahini, R., Wei, L., Wang, Z., Mathiak, K., et al. (2021). End-to-end sleep staging using convolutional neural network in raw single-channel EEG. *Biomed. Signal Process. Control* 63:102203. doi: 10.1016/j.bspc.2020.102203
- Liang, H., Liu, Y., Wang, H., Jia, Z., and Center, B. (2023). "Teacher assistant-based knowledge distillation extracting multi-level features on single channel sleep EEG," in *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23)*, 3948–3956. doi: 10.24963/ijcai.2023/439
- Lin, M., Chen, Q., and Yan, S. (2013). Network in network. *arXiv [Preprint]*. arXiv:1312.4400. doi: 10.48550/arXiv.1312.4400
- Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI: IEEE), 2117–2125. doi: 10.1109/CVPR.2017.106

- Loshchilov, I., and Hutter, F. (2017). Decoupled weight decay regularization. *arXiv [Preprint]* arXiv:1711.05101. doi: 10.48550/arXiv.1711.05101
- Malhotra, A., Younes, M., Kuna, S. T., Benca, R., Kushida, C. A., Walsh, J., et al. (2013). Performance of an automated polysomnography scoring system versus computer-assisted manual scoring. *Sleep* 36, 573–582. doi: 10.5665/sleep.2548
- McArdle, N., Ward, S. V., Bucks, R. S., Maddison, K., Smith, A., Huang, R.-C., et al. (2020). The prevalence of common sleep disorders in young adults: a descriptive population-based study. *Sleep* 43:zsaa072. doi: 10.1093/sleep/zsaa072
- Michielli, N., Acharya, U. R., and Molinari, F. (2019). Cascaded lstm recurrent neural network for automated sleep stage classification using single-channel EEG signals. *Comput. Biol. Med.* 106, 71–81. doi: 10.1016/j.compbiomed.2019.01.013
- Mousavi, S., Afghah, F., and Acharya, U. R. (2019). Sleeppegnet: automated sleep stage scoring with sequence to sequence deep learning approach. *PLoS ONE* 14:e0216456. doi: 10.1371/journal.pone.0216456
- Peng, L., Ren, Y., Luan, Z., Chen, X., Yang, X., Tu, W., et al. (2023). Sleepvitransformer: patch-based sleep spectrogram transformer for automatic sleep staging. *Biomed. Signal Process. Control* 86:105203. doi: 10.1016/j.bspc.2023.105203
- Phan, H., Andreotti, F., Cooray, N., Chén, O. Y., and De Vos, M. (2019a). Joint classification and prediction cnn framework for automatic sleep stage classification. *IEEE Trans. Biomed. Eng.* 66, 1285–1296. doi: 10.1109/TBME.2018.2872652
- Phan, H., Andreotti, F., Cooray, N., Chén, O. Y., and De Vos, M. (2019b). Seqsleepnet: End-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 400–410. doi: 10.1109/TNSRE.2019.2896659
- Phan, H., Mikkelsen, K., Chén, O. Y., Koch, P., Mertins, A., and De Vos, M. (2022). Sleeptransformer: automatic sleep staging with interpretability and uncertainty quantification. *IEEE Trans. Biomed. Eng.* 69, 2456–2467. doi: 10.1109/TBME.2022.3147187
- Sokolovsky, M., Guerrero, F., Paisarnrisomsuk, S., Ruiz, C., and Alvarez, S. A. (2019). Deep learning for automated feature discovery and classification of sleep stages. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 17, 1835–1845. doi: 10.1109/TCBB.2019.2912955
- Supratak, A., Dong, H., Wu, C., and Guo, Y. (2017). Deepsleepnet: a model for automatic sleep stage scoring based on raw single-channel EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1998–2008. doi: 10.1109/TNSRE.2017.2721116
- Supratak, A., and Guo, Y. (2020). “Tinysleepnet: an efficient deep learning model for sleep stage scoring based on raw single-channel EEG,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)* (Montreal, QC: IEEE), 641–644. doi: 10.1109/EMBC44109.2020.9176741
- Tang, M., Zhang, Z., He, Z., Li, W., Mou, X., Du, L., et al. (2022). Deep adaptation network for subject-specific sleep stage classification based on a single-lead ECG. *Biomed. Signal Process. Control* 75:103548. doi: 10.1016/j.bspc.2022.103548
- Tang, Z., Gao, Y., Zhu, Y., Zhang, Z., Li, M., Metaxas, D. N., et al. (2021). “Crossnorm and selfnorm for generalization under distribution shifts,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (Montreal, QC: IEEE), 52–61. doi: 10.1109/ICCV48922.2021.00012
- Toban, G., Poudel, K., and Hong, D. (2023). Rem sleep stage identification with raw single-channel EEG. *Bioengineering* 10:1074. doi: 10.3390/bioengineering10091074
- Weber, F., and Dan, Y. (2016). Circuit-based interrogation of sleep control. *Nature* 538, 51–59. doi: 10.1038/nature19773
- Zhang, J., Yao, R., Ge, W., and Gao, J. (2020). Orthogonal convolutional neural networks for automatic sleep stage classification based on single-channel EEG. *Comput. Methods Programs Biomed.* 183:105089. doi: 10.1016/j.cmpb.2019.105089
- Zhao, C., Li, J., and Guo, Y. (2022). Sleepcontextnet: a temporal context network for automatic sleep staging based single-channel EEG. *Comput. Methods Programs Biomed.* 220:106806. doi: 10.1016/j.cmpb.2022.106806



OPEN ACCESS

EDITED BY

Saad Arif,
King Faisal University, Saudi Arabia

REVIEWED BY

Ijaz Ul Haq,
Universitat de Lleida, Spain
Syed Qasim Abbas,
CSIRO e-Health Research Centre, Australia

*CORRESPONDENCE

Ali Shariq Imran
✉ ali.imran@ntnu.no

RECEIVED 24 October 2024

ACCEPTED 19 December 2024

PUBLISHED 08 January 2025

CITATION

Usman SM, Khalid S, Tanveer A, Imran AS and
Zubair M (2025) Multimodal consumer choice
prediction using EEG signals and eye tracking.
Front. Comput. Neurosci. 18:1516440.
doi: 10.3389/fncom.2024.1516440

COPYRIGHT

© 2025 Usman, Khalid, Tanveer, Imran and
Zubair. 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.

Multimodal consumer choice prediction using EEG signals and eye tracking

Syed Muhammad Usman¹, Shehzad Khalid², Aimen Tanveer³,
Ali Shariq Imran^{4*} and Muhammad Zubair⁵

¹Department of Computer Science, Bahria School of Engineering and Applied Science, Bahria University, Islamabad, Pakistan, ²Department of Computer Engineering, Bahria School of Engineering and Applied Science, Bahria University, Islamabad, Pakistan, ³Department of Creative Technologies, Air University, Islamabad, Pakistan, ⁴Department of Computer Science, Norwegian University of Science and Technology, Gjøvik, Norway, ⁵Interdisciplinary Research Center for Finance and Digital Economy, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

Marketing plays a vital role in the success of a business, driving customer engagement, brand recognition, and revenue growth. Neuromarketing adds depth to this by employing insights into consumer behavior through brain activity and emotional responses to create more effective marketing strategies. Electroencephalogram (EEG) has typically been utilized by researchers for neuromarketing, whereas Eye Tracking (ET) has remained unexplored. To address this gap, we propose a novel multimodal approach to predict consumer choices by integrating EEG and ET data. Noise from EEG signals is mitigated using a bandpass filter, Artifact Subspace Reconstruction (ASR), and Fast Orthogonal Regression for Classification and Estimation (FORCE). Class imbalance is handled by employing the Synthetic Minority Over-sampling Technique (SMOTE). Handcrafted features, including statistical and wavelet features, and automated features from Convolutional Neural Network and Long Short-Term Memory (CNN-LSTM), have been extracted and concatenated to generate a feature space representation. For ET data, preprocessing involved interpolation, gaze plots, and SMOTE, followed by feature extraction using LeNet-5 and handcrafted features like fixations and saccades. Multimodal feature space representation was generated by performing feature-level fusion for EEG and ET, which was later fed into a meta-learner-based ensemble classifier with three base classifiers, including Random Forest, Extended Gradient Boosting, and Gradient Boosting, and Random Forest as the meta-classifier, to perform classification between buy vs. not buy. The performance of the proposed approach is evaluated using a variety of performance metrics, including accuracy, precision, recall, and F1 score. Our model demonstrated superior performance compared to competitors by achieving 84.01% accuracy in predicting consumer choices and 83% precision in identifying positive consumer preferences.

KEYWORDS

EEG, eye tracking, neuromarketing, CNN-LSTM, multimodal

1 Introduction

Neuromarketing, a dynamic fusion of neuroscience and marketing, has emerged through the innovative use of non-invasive Brain-Computer Interface (BCI) technology, revolutionizing the concept of marketing. Marketing is a connection between production and consumers. A good product can fail to target its desired audience without effective

marketing (Assel, 1995). To create products and services with the highest profit potential, it is crucial to thoroughly understand consumer behavior and develop a corresponding advertising strategy. This requires a comprehensive understanding of the buyer's decision-making process, which typically includes need recognition, information search, evaluation, purchase decision, and post-purchase behavior (Armstrong et al., 2014; Peter et al., 1999; Vecchiato et al., 2011). Researchers have employed Electroencephalography (EEG) and Eye Tracking (ET) to analyze the brain activity and gaze outcomes when exposed to different stimuli for several decades.

EEG is a technique used to assess the electrical activity within a person's cranial structure. This involves placing numerous electrodes on the scalp, a method known as scalp EEG. It is particularly preferred for recording brain waves because it is simple and does not involve any invasive procedure, while other methods are preferable because they are efficient in monitoring brain activity (Teplan, 2002; Fisch, 1999). It records changes in electrical activity and oscillations within the brain. The amplitude of the signals are proportional to the type of mental activity experienced when exposed to stimuli (Homan et al., 1987). Eye tracking, on the other hand, involves gathering information on visual attention through the capturing of eye movements. The eye tracking revealed where and for how long a person looked at the different elements, whereas EEG can uncover the emotional and cognitive response elicited by these stimuli.

Neuromarketing, a multidisciplinary field at the intersection of neuroscience, psychology, and economics, explores the complex dynamics of how advertisements can significantly impact product sales. Unlike traditional marketing research methods such as interviews, reviews, and questionnaires, neuromarketing seeks to surpass the limitations inherent in these approaches. These conventional methods often fall short of fully revealing consumers' insights toward products, as individuals may encounter challenges in conveying their preferences or may be hesitant to express them comprehensively. Moreover, the chances of data manipulation add a layer of complexity to the reliability of findings.

Human behavior is influenced by processes operating beneath the conscious threshold. In response to these challenges, neuromarketing offers a revolutionary shift, going beyond direct questions about products and exploring the deeper subconscious areas of consumers' minds. The essence is to get insights in a non-invasive manner, extracting authentic preferences and choices that may outstand conventional probing techniques. It offers a deeper and more precise insight into consumer behavior. This leads to the development of innovative and successful marketing tactics, ultimately driving increased sales. In the expansive and intricate landscape of the advertising industry, where expenditures vary based on geographical location, industry sector, and individual company strategies. The main contributions of this research study are as follows:

- A novel multimodal framework has been proposed, integrating EEG signals and eye-tracking data to enhance consumer preference prediction. This approach combines the strengths of both modalities, addressing the lack of sufficient multimodal research in the domain.
- A robust feature extraction pipeline has been designed, combining handcrafted features and automated features derived through deep learning. This hybrid approach provides a more comprehensive representation of the data, bridging an identified gap in the existing literature.
- Ensemble classification techniques have been proposed to address the challenges of class imbalance and improve prediction accuracy. By utilizing multiple classifiers and optimizing their integration, significant improvements in performance metrics were achieved compared to traditional methods.

2 Literature review

Many individuals are often reserved in expressing their complete thoughts and preferences during product evaluation, creating a challenge in comprehending the complexities of consumer decision-making. The emergence of neuroimaging tools provides a quick and convenient method to understand a customer's brain activity when evaluating and choosing different products. Consumer choice recognition typically involves three pivotal stages. The initial step encompasses preprocessing, wherein unwanted noise is eliminated from both EEG and ET signals. Following this, relevant features are extracted, and subsequently, EEG and ET signals are classified based on consumer preferences. In neuromarketing studies, the recording of both EEG and ET data equips researchers to get into the complex interplay of factors that influence how the human psyche makes choices among different products.

2.1 Predictive approaches for consumer preference based on EEG signals

Researchers have proposed multiple methods for classification between like vs. dislike for neuromarketing in recent years. A typical method consists of preprocessing the EEG signals and extracting the features followed by the classification. Researchers have used various preprocessing techniques employed in predicting consumer preferences. Bandpass filtering, widely utilized for EEG signal noise reduction in numerous studies (Murugappan et al., 2014; Alimardani and Kaba, 2021; Aldayel et al., 2021; Georgiadis et al., 2022, 2023a), serves as a prominent technique. Independent Component Analysis (ICA) has been adopted by researchers to eliminate noise in their proposed methods (Aldayel et al., 2021; Georgiadis et al., 2022; Telpaz et al., 2015; Hakim et al., 2021). Telpaz et al. (2015) and Hakim et al. (2021) have also applied the Notch Filter for preprocessing. Downsampling, an effective method employed by several researchers like (Aldayel et al., 2021), proves valuable for reducing the sampling rate of EEG data. Moreover, the Savitzky-Golay filter was utilized to effectively remove artifacts (Aldayel et al., 2021; Yadava et al., 2017; Shah et al., 2022). Murugappan et al. (2014) applied the Surface Laplacian Filter, and Kumar et al. (2019) used high and low pass filters for the purpose of preprocessing EEG signals.

TABLE 1 Comparison of existing consumer preference prediction methods using EEG signals.

References	Data	Year	Preprocessing	Feature extraction	Classifier	Accuracy (%)
Murugappan et al. (2014)	EEG	2014	Bandpass filter Surface Laplacian filter	PSD SE SC	kNN PNN	96.62
Telpaz et al. (2015)	EEG	2015	Notch filter ICA	ERSP N200 (ERP)	Random	59 65
Yadava et al. (2017)	EEG	2017	Savitzky-Golay	DWT	HMM	70.33
Aldayel et al. (2021)	EEG	2021	Downsampling Bandpass filter ICA Savitzky-Golay	DWT Welch method	DNN SVM kNN RF	83 81 73 87
Alimardani and Kaba (2021)	EEG	2021	Bandpass filter	PSD	CNN EC (SVM RF, LOG)	74.57 63.5
Hakim et al. (2021)	EEG	2021	Notch filter ICA	FBP Hemispheric symmetry	SVM LOG kNN DT	68.51
Shah et al. (2022)	EEG	2022	Savitzky-Golay FFT SMOTE	DWT PSD LSTM	EC (SVM, DT, DNN)	96.89
Georgiadis et al. (2022)	EEG	2022	Bandpass filter ICA	SCM	SVM Ensemble	73.11
Georgiadis et al. (2023a)	EEG	2023	Bandpass Filter	SCM	SPDNet	72.18

After the preprocessing of EEG signals, the extraction of features is pivotal for classifying likes and dislikes. Many approaches are employed for feature extraction like LSTM (Shah et al., 2022). Telpaz et al. (2015) have leveraged N200, or N2, is an event-related potential (ERP) component. The Power Spectrum Density (PSD) provides the distribution of power across diverse frequencies in the signal (Murugappan et al., 2014; Alimardani and Kaba, 2021; Shah et al., 2022). Similarly, Discrete Wavelet Transform (DWT) (Arif et al., 2023) introduces a process of iteratively breaking down the signal into approximation and detail coefficients across multiple scales, a technique adeptly utilized by researchers for feature extraction (Aldayel et al., 2021; Yadava et al., 2017; Shah et al., 2022; Kumar et al., 2019). Aldayel et al. (2021) have contributed by employing Welch Method. This metric, corresponding to the spatial standard deviation, offers insights into the amount of activity at each time point in the potential field. EEG signals are represented as Sample Covariance Matrices (SCMs) that are measured entities scattered over a particular Riemannian manifold by Georgiadis et al. (2022, 2023a). One of the most commonly used method is to analyze EEG data is to break the signal into functionally distinct frequency bands. Telpaz et al. (2015) and Hakim et al. (2021) extracted frequency bands to extract features from EEG signals. These features provide high interclass variance which is useful in accurate classification. The details of these various features are briefly described in the following table understanding what kind of preprocessing techniques and feature extraction methods were used in this research, as shown in Table 1.

There are simple features such as the frequency distribution of words to parametric and non-parametric features, etc. for

classification between the “like” and “dislike” classes. Statistical features in the time domain include the mean average, variance/standard deviation, skewness, and kurtosis. Also, frequency domain features such as moments of spectrum like spectral centroid, variational coefficients, and even skewness in the spectrum can be incorporated. In addition, other techniques of dimensionality reduction such as Principal Component Analysis (PCA) have also been used in this study by the researchers to extract features and to reduce dimensionality. Table 1 can give a brief idea of these various features and present an outline of the most important preprocessing proposals and the feature extraction applied in the present research.

Deep Neural Network (DNN), Support Vector Machine (SVM), Random Forest (RF), and *k*-Nearest Neighbors (kNN) resulted a maximum accuracy of 87%. Alimardani and Kaba (2021) proposed an ensemble classifier based on SVM, RF, Logistic Regression (LOG) and Convolution Neural Network (CNN). Murugappan et al. (2014) applied for kNN and Probabilistic Neural Network (PNN) for classification of EEG signals. Hakim et al. (2021) conducted a comprehensive study utilizing EEG, focusing solely on Machine Learning algorithms and achieved an accuracy of 68.51%. Shah et al. (2022) predicted users' preferences for advertisements using an ensemble classifier [SVM, Decision Tree (DT), DNN] achieving an impressive accuracy of 96.89%. Yadava et al. (2017) presented the first dataset of neuromarketing. This dataset featured stimuli in the form of images of commercial products, labeled as either “like” or “dislike,” and Hidden Markov Model (HMM) was employed for the classification of EEG signals based on likes and dislikes. Georgiadis et al. (2022) applied a SVM

TABLE 2 Comparison of existing consumer preference prediction methods using EEG and ET data.

Author	Data	Year	Preprocessing	Feature extraction	Classification	Results
Khushaba et al. (2013)	EEG ET	2013	ICA DWT	FFT	Mutual Information	Preference
Matukin et al. (2016)	EEG ET	2016	Bandpass filter	FFT	Not specified	Improvement in Ads
Samsuri et al. (2016)	EEG ET	2016	Bandpass filter	P300 ERP N100 Pupil dilation	Statistics	ERP and the ET results were inconsistent
Christoforou et al. (2017)	EEG ET	2017	Downsampling Notch filter	Attent.Asynchrony Cogn. Congruency	Regression R2	72% accuracy
Slanzi et al. (2017)	EEG ET	2017	Interpolation BandPass filter	PCA	Logistic regression	71.09% accuracy
García-Madariaga et al. (2019)	EEG ET	2019	Not specified	Alpha-Band Oscillation AOI	Not specified	Eye movements could predict packaging preference.
Mashrur et al. (2024)	EEG ET	2023	ASR Notch filter	TD FD TFD	SVM-RBF	96.97% accuracy

Ensemble including three SVM classifiers, while their research (Georgiadis et al., 2023a) used architecture of SPDNet. It is a deep learning architecture designed for processing data that lie on Symmetric Positive Definite (SPD) matrices.

2.2 Predictive approaches for consumer preference based on EEG signals and ET data

Researchers have employed various techniques to preprocess EEG signals and ET data for understanding consumer preferences. Khushaba et al. (2013) utilized a combination of ICA and DWT for EEG signal preprocessing. Matukin et al. (2016) and Samsuri et al. (2016) incorporated band-pass filtering in their methodologies. Christoforou et al. (2017) downsampled EEG data and applied a Notch filter to mitigate DC drifts. For processing pupil dilation signals, Slanzi et al. (2017) employed linear interpolation followed by band-pass filtering. Mashrur et al. (2024) adopted the Automatic Subspace reconstruction functionality from EEGLAB for noise reduction, subsequently applying a notch filter at 50 Hz to suppress power line artifacts.

Matukin et al. (2016) applied Fast Fourier Transform (FFT) to derive features from EEG signals. Samsuri et al. (2016) utilized P300 and N100 components for EEG signal analysis, while employing Pupil Dilation features for eye-tracking data. Christoforou et al. (2017) introduced the Attentional-asynchrony metric based on the Eye-Gaze Divergence Index and used epoched EEG measurements to formulate a Cognitive-congruency aggregate metric. Slanzi et al. (2017) employed Principal Component Analysis (PCA) to extract features from EEG signals. García-Madariaga et al. (2019) focused on Alpha Band Oscillations for EEG signals and Area of Interest (AOI) for eye-tracking data. Mashrur et al. (2024) categorized features into three domains: time domain (TD), frequency domain (FD), and time-frequency

domain (TFD), subsequently employing a classifier for optimal feature selection.

Table 2 provides comparative analysis of existing methods of neuromarketing based on EEG and ET. Khushaba et al. (2013) used mutual information analysis that indicated important factors affecting the buying decision. Samsuri et al. (2016) measured the attention levels of users when observing an advertisement through the use of EEG and ET signals. In the study by Christoforou et al. (2017), the R2 metric was employed to assess the predictive capability of the suggested neural and eye-tracking metrics on the box office success of films. Slanzi et al. (2017), aimed to determine the sections of a webpage that were most probable to attract clicks through the application of Logistic Regression. Mashrur et al. (2024) used the SVM classifier is used with RBF kernel for classifying strong and weak preference EEG signals attaining an accuracy of 97%.

Following research gaps have been identified after a comprehensive literature review of both EEG and ET consumer preference prediction methods:

- There is a lack of sufficient multimodal research investigating the combined effectiveness of EEG and eye-tracking.
- The issue of class imbalance remains a significant challenge in this field.
- The integration of handcrafted and automated features in a combined feature set has not been much explored.
- The limited use of ensemble learning methods represents a notable research gap.

3 Dataset

The NeuMa dataset (Georgiadis et al., 2023b) has been used for this particular research work and this comprises of 42 participants who were all Greek speakers; 23 males and 19 females. The dataset is made up of 144 supermarket products and this is presented in six

TABLE 3 Summary of NeuMa dataset.

Attribute	Details
Number of subjects	42 (23 males, 19 females)
Number of products	144
Number of pages	6 (24 products per page)
Average selections	18 products per participant
Data files per subject	2 (S01.xdf, S01.xls)
EEG device	Wearable Sensing DSI24
EEG sampling frequency	300 Hz
EEG sensors	21 dry sensors
ET device	Tobii pro fusion
ET sampling frequency	120 Hz

brochure pages whereby each brochure is made up of 24 products. Targets were highlighted by users with a left-click of the mouse on products of interest. Consequently, each of the subjects has two files for every subject, which contains the records of their interactions with the products.

Table 3 provides brief description of the NeuMa dataset. Subjects were positioned at an arm's length or 50 cm away from the screen which is a 28 inch LCD monitor, and navigation on the digital brochure page and choice of products with the left click of the mouse. For the page navigation arrow keys of the keyboard were used. Every subject's data set involved EEG and eye-tracking information and mouse clicks and positions. There are EEG signals and eye movements, mouse clicks, and cursor movements collected in the given dataset. Among these data streams, currently only EEG and ET type data streams are being used.

After the experiment, participants filled in a questionnaire containing demographic details about the individuals, profiling details about the participants as well as about the products provided to them like personality profile, tendency to indulge in impulse buying and about the products given to them like reasons for selection of product, familiarity with the product and frequency for buying the product. EEG data was recorded by DSI 24 system with the sampling rate of 300 Hz from 21 electrodes. This eye-tracking data was at a sampling rate of 120Hz and the Tobii Pro Fusion eye-tracker was used to collect the data. Figures 1–3 show the plots of EEG, ET, and Pupil dilation data, respectively (Tobii, 2024; Georgiadis et al., 2023b).

4 Methodology

Proposed method consists of three steps: EEG and ET signal preprocessing, feature extraction and classification. The preprocessing of the EEG signals is done with the help of Bandpass Butterworth filter (0.5–45Hz), Artifact Subspace Reconstruction (ASR) and the Fast Orthogonal Regression for Classification and Estimation (FORCE). Signals are then split in segments overlapping each other since the data amount is at a manageable size. In the same manner, preprocessing of Eye Tracking (ET) data; missing values are eliminated/taken care of using a linear interpolation

and the data is segmented using overlapping window techniques. Non-technique based features are derived from the EEG and ET signals using statistical and frequency domain analysis. The technique incorporated is CNN-LSTM for EEG and LeNet5 for ET data. First, for each input modality, feature-level fusion is used to combine these extracted features, and second, improvements are made to classification using both manually defined and learned features are used. Figure 4 displays the flow diagram of the proposed methodology.

4.1 Preprocessing of EEG signals

Electroencephalogram (EEG) signals are often contaminated with various types of noise, including muscle activity, eye movements, and electrical interference from other devices. To analyze EEG data effectively, preprocessing steps such as filtering are crucial. Electroencephalography (EEG) signals require preprocessing to remove noise and isolate frequencies of interest. One method is applying a bandpass filter with band range 0.5–45 Hz.

In order to filter out Signals with artifact in the EEG data a band pass filter was employed together with Artifact Subspace Reconstruction (ASR). ASR also helps in eradicating interferences like shrugs and blinks and leaves the signal's quality intact for analysis. This technique is very important in neuromarketing research as it offers clean signal filtration yet preserves the original signal. Fast method for Orthogonal Regression for Classification and Estimation EEG sounds are done using orthogonal basis vector and FORCE for the preprocessing of the data. It is applied to remove noise while improving the quality of the signal, making it suitable for the situations that require fast and accurate artifacts detection. Specifically, the signals from the EEG signals were band-pass filtered and then analyzed by ASR and FORCE to obtain the best results.

The overlapping window technique again divides the filtered signals to get more detailed data and make the signals continuous. The division of the continuous EEG signals make it more manageable and this was achieved by gaining small samples of 300 Hz with the window size being one second with 300 data entries. The overlapping of the windows has the advantage of achieving greater density of information and continuity of the signal.

4.2 Preprocessing of eye tracking data

Linear interpolation is a technique of curve fitting in which a straight line is drawn between two points to give the estimated point. It handles missing data if eye-tracking signals due to long blinking are missing using what is known as the straight-line interpolation method. Due to the ability to replace a missing value with approximated data samples that occur before and after the gap, a continuous signal is achieved. This step is necessary for preserving the quality of eye-tracking signal and further analysis of study subjects' attention and eye movement behavior.

As for removing missing values in the eye-tracking dataset, linear interpolation has been applied, the next step of the data

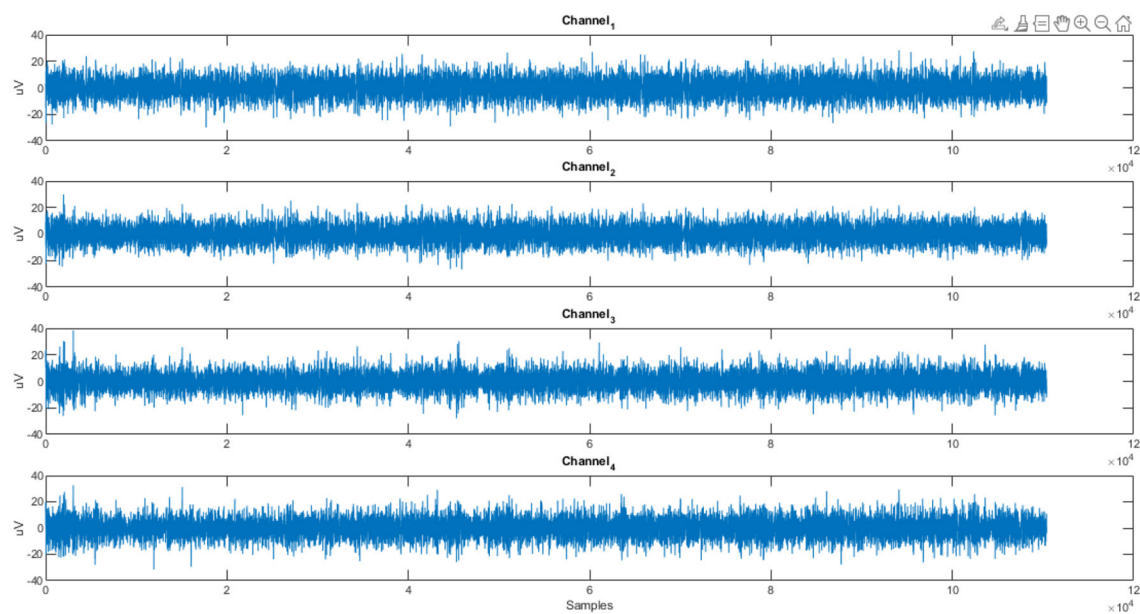


FIGURE 1

Customer response (Subject S01): EEG data stream for product (NeuMa dataset: EEG data capturing the cerebral activity of a subject for a product).

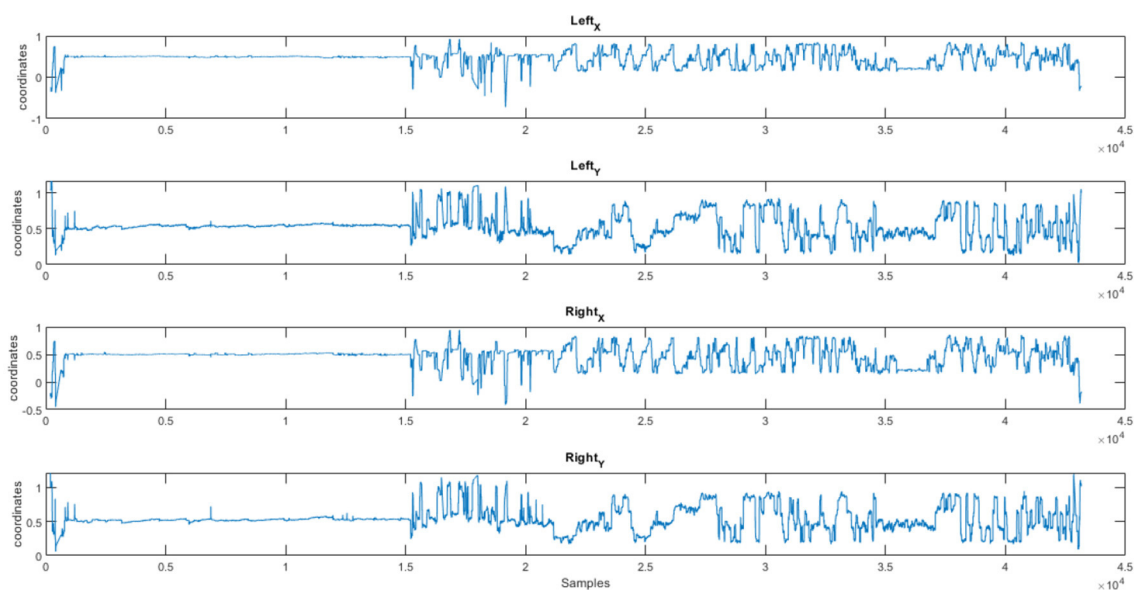


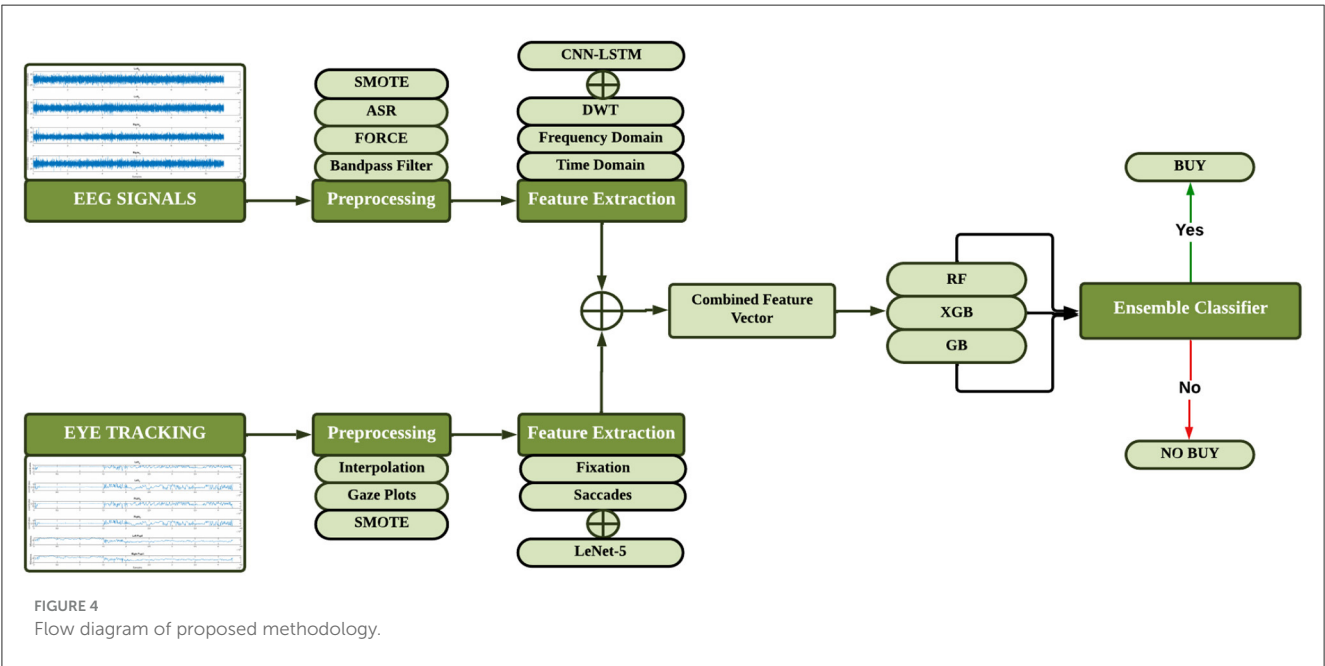
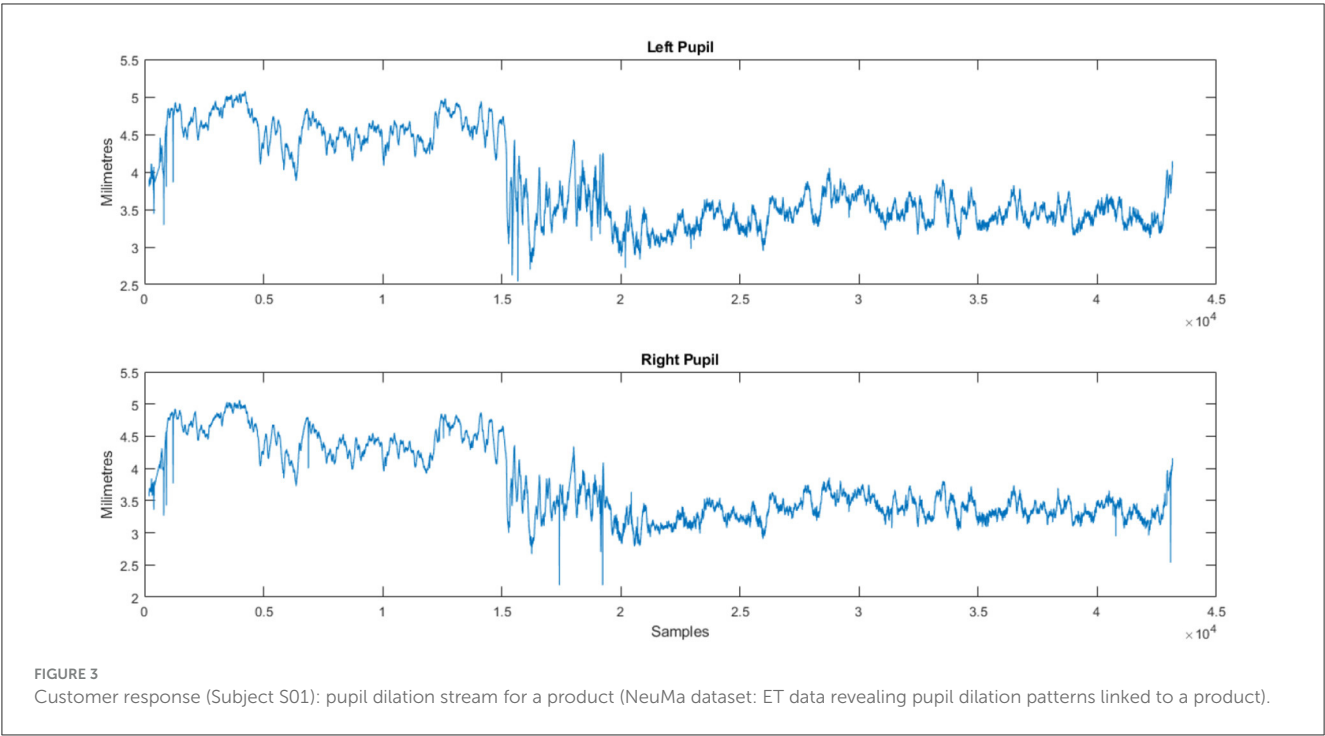
FIGURE 2

Customer response (Subject S01): eye tracking coordinates (X, Y) for a product (NeuMa dataset: ET data revealing the gaze pattern linked to a product).

preprocessing is the data segmentation based on the overlapping window. This ensures that the maximum amount of information is collected as the windows have a 50% overlap in which every two consecutive windows have 50% of the same data points. This rises the density of data and contributes to non-fragmentation of signal which helps in maintaining coherency. The splitting of records further improves difference detection or comparison which is made possible by the window size of one second and a sampling

rate of 120 Hz; this means that each segment's data set has 120 data points.

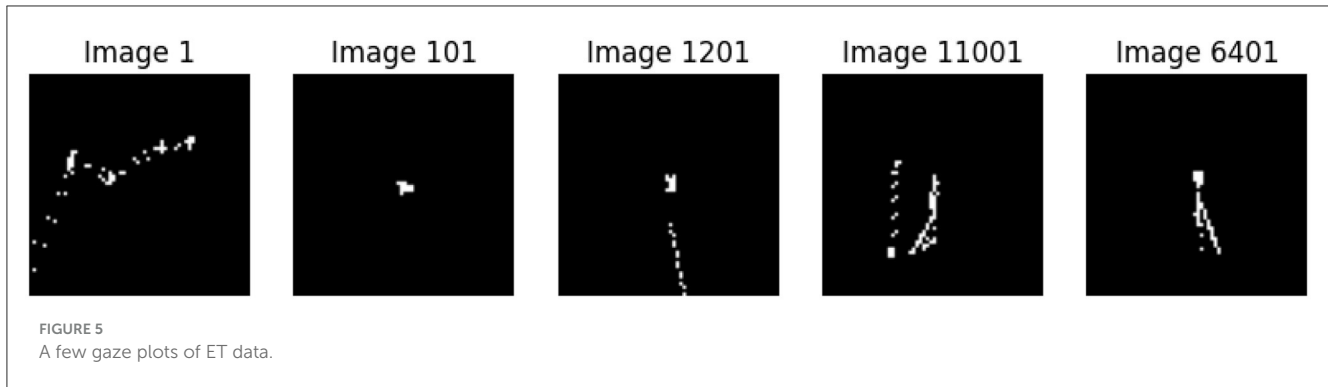
Gaze plots are basically eye movement data obtained through eye tracking displayed graphically as data points. They are developed by placing fixation areas on a graph of the observed stimulus. For the movements of the eyes, the X and Y coordinates are transformed into the 64×64 canvas where the black background implies no gaze while the white point marks a gaze.



These points are the coordinates of the location on the canvas, and the original gaze plot images are saved, converted to the NumPy array and then to grayscale for analysis. Figure 5 presents various Gaze plots.

Class imbalance refers to situations where one class (the minority class) is significantly underrepresented compared to another class (the majority class). This class imbalance can lead to biased models that perform poorly on the minority class. To address the issue of class imbalance within the dataset, we used Synthetic Minority Over-sampling Technique (SMOTE) (Chawla

et al., 2002). SMOTE works by generating synthetic examples of the minority class to balance the class distribution. The process involves creating new instances of minority class samples by interpolating between existing minority class samples. SMOTE first identifies the minority class samples in the dataset. For each minority class sample of EEG and ET data, SMOTE selects its k nearest neighbors in the feature space. The value of k we chose is 3, as it was giving the best results. For each minority class sample, SMOTE generates synthetic samples along the line segments connecting it to its k nearest neighbors. The number



of synthetic samples created for each minority class sample is determined by a specified oversampling ratio. By creating synthetic samples, SMOTE increases the representation of the minority class in the dataset, balancing the class distribution. Figure 6 displays the class distribution before and after application of SMOTE.

4.3 Feature extraction of EEG signals

After preprocessing the EEG data, which typically involves filtering out noise and artifacts, the next step is to extract meaningful features from the cleaned data. Feature extraction transforms the raw EEG signals into a set of representative features that can be used for further analysis, such as classification. We used a few common statistical features include Mean, Variance, Skewness, and Kurtosis.

The mean of the EEG signal provides a measure of the central tendency of the signal. It indicates the average value of the signal over a specified period.

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (1)$$

Where N is the number of data points and x_i represents the EEG signal values.

The co-efficient of variation is used to determine the spread of the signal value in relation to the mean of the EEG signals. That reveal information about the fluctuation in the activity of the brain.

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (2)$$

Skewness indicates the extent of probability distribution of the EEG signals asymmetrical nature. The absolute value of skewness is >1 , <1 or zero if the distribution is highly skewed to the right, left, or symmetric respectively.

$$\gamma_1 = \frac{N}{(N-1)(N-2)} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^3 \quad (3)$$

Kurtosis quantifies the degree of the two at both the center and the tails of the probability density function of the EEG signal. It also

implies that the data contains some outliers.

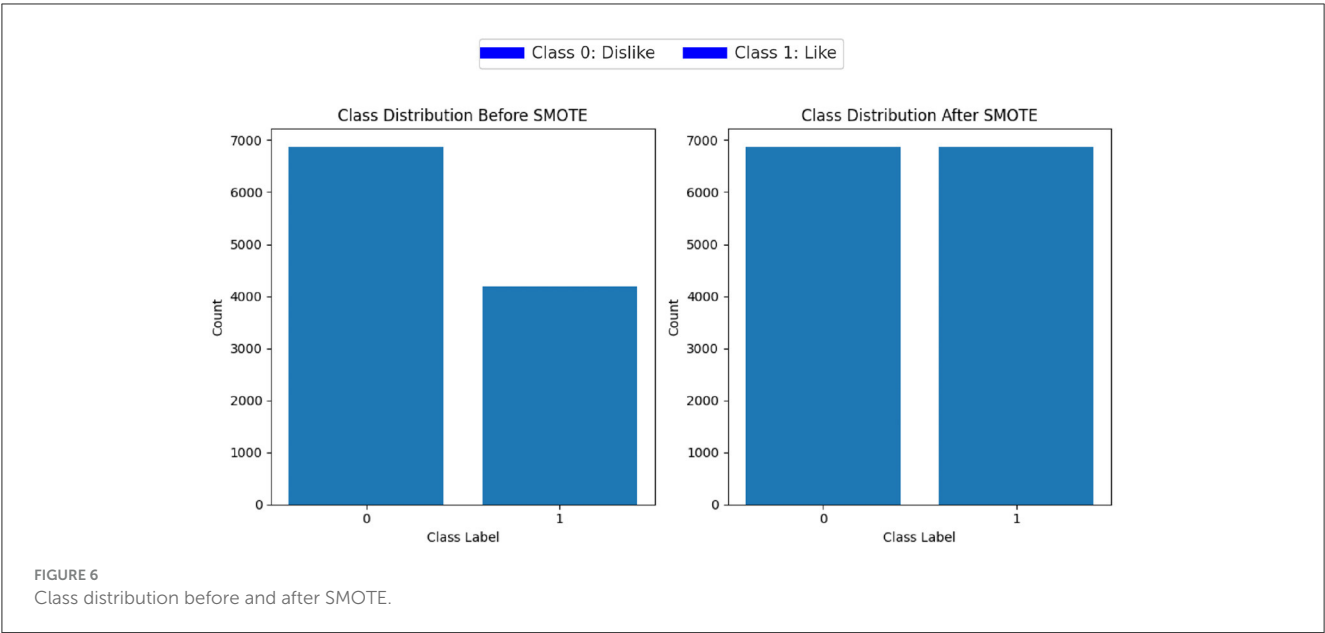
$$\gamma_2 = \frac{N(N+1)}{(N-1)(N-2)(N-3)} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^4 - \frac{3(N-1)^2}{(N-2)(N-3)} \quad (4)$$

Welch's Method (Welch, 1967) is one of the robust and standard method to estimate power spectral density (PSD) of a signal. Even if it splits the signal into overlapping sections, then they apply a function known as windowing on sections, calculate the periodogram of each segment, and finally the averages these periodograms. In this feature, extraction was performed for all the EEG channels, considering the average power of the given signal within all possible frequency bands. This feature quantifies the amplitude deviations of the power from the energy of the signal at various frequency bands. The wavelet transform is the process by which a signal is broken down in different parts that are localized both temporally and in the frequency domain. Since, mean of DWT coefficients gives the average value of the coefficients, we obtained the mean of this parameter. This feature calculates the extent of fluctuations valued in the domain of wavelet coefficients.

Finally, Statistical features and frequency domain features and wavelet transform features are then combined to construct an information vector for each sample that will serve as the input to the model. The technical advantage implemented in the feature set uses time-frequency characteristic as well as multi-resolution analysis.

After removing noise from the EEG signals, features were extracted for the "Buy" and "No Buy" classes using two common approaches: handcrafted feature extraction and automated feature extraction via deep learning techniques. In the handcrafted approach, features are extracted without considering the class of the EEG signals. In contrast, automated feature extraction leverages deep learning models like Convolutional Neural Networks (CNNs) (LeCun et al., 1998) and Long Short-Term Memory (LSTM) networks (Hochreiter and Schmidhuber, 1997), which consider the class of the EEG signals during feature extraction. This method can lead the class of the EEG signals during feature extraction. This method can lead to improved classification performance due to lower intraclass variance and higher interclass variance.

LSTMs are a type of recurrent neural network (RNN) that are capable of learning long-term dependencies (Shah et al., 2022). They have a chain-like structure with repeating modules. The core of the LSTM module consists of a cell state, and three gates to regulate the flow of information: the input gate, forget gate, and



output gate. For feature extraction from EEG signals, we implement a convolutional neural network LSTM architecture. The CNN takes the segmented time-domain signals it has: the CNN’s inputs are the number of EEG channels and temporal segments of signals. Convolutional layers perform spatial features extraction using filters of particular sizes which are succeeded by the max-pooling layers in an attempt to decrease the dimension and hence increasing the efficiency of the training process. The features from the CNN layers are flattened and reshaped so as to be fed into LSTM layer that takes into consideration temporality of the data. CNN and LSTM are combined because the former analyses the spatial information of the signals while the latter analyses the temporal information of the signals making it appropriate to classify the EEG signals. Table 4 provides a summary of our proposed CNN-LSTM model.

TABLE 4 Summary of proposed CNN-LSTM model.

Layer	Output shape	Parameters
Input layer	(None, 19, 300, 1)	0
Conv2D	(None, 17, 298, 32)	320
Max pooling	(None, 8, 149, 32)	0
Conv2D	(None, 6, 147, 64)	18,496
MaxPooling	(None, 3, 73, 64)	0
Flatten layer	(None, 14,016)	0
Reshape layer	(None, 14,016, 1)	0
LSTM	(None, 64)	16,832
Dense	(None, 128)	8,320
Dense	(None, 64)	8,256

4.4 Feature extraction of ET data

After preprocessing ET data and handling the class imbalance issue features are extracted from it. Similar to feature extraction from EEG data, statistical features can be employed to quantify various aspects of these movements.

4.4.1 Fixation duration

Fixation duration represents the average time a user spends fixating on a specific Area of Interest (AOI) and is analogous to the mean in EEG analysis. It provides insight into the level of attention paid to that area.

4.4.2 Saccade amplitude

Saccade amplitude is just like variance in EEG, its calculates the distance between one fixation to another fixation. Large value of saccade amplitude represents jump from one fixation to other.

We applied the LeNet-5 (LeCun et al., 1998) model, a foundational CNN architecture developed by Yann LeCun, originally designed for image recognition tasks like classifying handwritten digits. The model processes input images through convolutional layers with filters to extract features, followed by max-pooling layers to reduce dimensionality. After multiple convolution and pooling layers, the feature maps are flattened into a vector for the classification layers. This structure effectively captures spatial features in the data, making it suitable for image recognition tasks. Here’s an explanation for the LeNet-5 model summarized in Table 5.

Features extracted from EEG signals and ET data are concatenated to form a combined feature vector with a size of 16,720, which is then fed into an ensemble classifier. Optimizer used is Adam and loss function used is Mean Squared Error. Adam is a gradient-based optimization algorithm. Its update rules are given by:

TABLE 5 Summary of LeNet-5 model.

Layer	Output shape	Parameters
Input layer	(None, 64, 64, 1)	0
Conv2D	(None, 60, 60, 6)	456
Max pooling	(None, 30, 30, 6)	0
Conv2D	(None, 26, 26, 16)	2,416
MaxPooling	(None, 13, 13, 16)	0
Flatten layer	(None, 2,704)	0

Update Rules:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad (5)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \quad (6)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad (7)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}, \quad (8)$$

$$\theta_t = \theta_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}, \quad (9)$$

where:

- m_t : First moment (mean of gradients),
- v_t : Second moment (uncentered variance of gradients),
- g_t : Gradient at time step t ,
- β_1, β_2 : Exponential decay rates for the moment estimates,
- η : Learning rate,
- ϵ : Small constant to prevent division by zero,
- θ_t : Parameters at time step t .

The Mean Squared Error loss function is given by:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (10)$$

where:

- n : Number of data points,
- y_i : True value for the i -th data point,
- $\hat{y}_i$: Predicted value for the i -th data point.

4.5 Ensemble classifier

After pre-processing and feature extraction, the final step is classification which is performed to categorize the sample as Buy vs. Non-buy. We have used three stacking ensemble classification approach in which features are first passed to three different classifiers including Random Forest, Gradient Boosting and XGBoost. Prediction obtained from these three classifiers is then stacked to get the final classification. Random Forest has been used as meta model in the stacking ensemble (Wolpert, 1992).

4.5.1 Base classifiers

Random Forest (RF): Random Forest grows a whole forest during training. It is a bagging technique in which every tree makes a prediction and then make a final prediction. This classifier is good to handle high dimensional data which is often the case with EEG and ET features (Breiman, 2001)

$$T_k(x) = \text{Class label predicted by the } k\text{-th tree.} \quad (11)$$

$$P_{\text{RF}}(x) = \text{Majority Vote}\{T_1(x), T_2(x), \dots, T_K(x)\} \quad (12)$$

Gradient Boosting (GB): This is a strong method that constructs decision trees iteratively where each stage used in identifying the mistakes committed by the prior trees. Friedman (2001). The final prediction is:

$$P_{\text{GB}}(x) = \sum_{m=1}^M \alpha_m h_m(x) \quad (13)$$

where:

- $h_m(x)$: the m -th weak learner,
- α_m : the weight of the m -th learner.

XGBoost (XGB): XGBoost is an optimized version of Gradient Boosting (Chen and Guestrin, 2016). The prediction for XGBoost is:

$$P_{\text{XGB}}(x) = \sum_{m=1}^M \eta_m h_m(x) + \Omega(h_m) \quad (14)$$

where:

- η_m : learning rate,
- $\Omega(h_m)$: regularization term.

4.6 Meta-classifier

The meta-classifier takes the outputs of the base classifiers as input. In this case, a Random Forest is used as the meta-classifier.

In the first step, predictions are collected from the base classifiers for the training dataset:

$$P_{\text{RF}}(x), \quad P_{\text{GB}}(x), \quad P_{\text{XGB}}(x). \quad (15)$$

$$Z = \begin{bmatrix} P_{\text{RF}}(x_1) & P_{\text{GB}}(x_1) & P_{\text{XGB}}(x_1) \\ P_{\text{RF}}(x_2) & P_{\text{GB}}(x_2) & P_{\text{XGB}}(x_2) \\ \vdots & \vdots & \vdots \\ P_{\text{RF}}(x_n) & P_{\text{GB}}(x_n) & P_{\text{XGB}}(x_n) \end{bmatrix} \quad (16)$$

In the final step, meta classifiers is trained to get the final classification result on Z . We have used Random Forest as meta classifier.

$$P_{\text{Meta}}(x) = \text{Meta-RF}(Z). \quad (17)$$

For unseen data x , the stacking ensemble works as follows: Each base model makes a prediction:

$$P_{\text{RF}}(x), P_{\text{GB}}(x), P_{\text{XGB}}(x). \quad (18)$$

These predictions form a new feature vector for x :

$$Z_x = [P_{\text{RF}}(x), P_{\text{GB}}(x), P_{\text{XGB}}(x)]. \quad (19)$$

The meta-classifier uses Z_x to make the final prediction:

$$P_{\text{Final}}(x) = P_{\text{Meta}}(x). \quad (20)$$

Base model outputs are as follows:

$$\begin{aligned} P_{\text{RF}}(x) &= \text{Random Forest prediction,} \\ P_{\text{GB}}(x) &= \text{Gradient Boosting prediction,} \\ P_{\text{XGB}}(x) &= \text{XGBoost prediction.} \end{aligned}$$

Meta-model (Random Forest) output is as follows:

$$P_{\text{Meta}}(x) = \text{Majority Vote}\{T_k(Z_x)\}, \quad k = 1, 2, \dots, K. \quad (21)$$

Final stacking ensemble prediction:

$$P_{\text{Final}}(x) = P_{\text{Meta}}(x). \quad (22)$$

4.7 Hyperparameters optimization

For the machine learning models, Random Forest Classifier was set with 265 estimators for the Optuna (Akiba et al., 2019) tuned model and 100 for the Stacking Classifier final estimator. The Gradient Boosting Classifier uses 89 estimators and the XGB Classifier is set with 300 estimators. These three estimators are surrounded by the Stacking Classifier such that the Random Forest classifier is used as the final estimator. Further, the imbalance of the data is tackled using SMOTE with the specified random state of 42. To split the dataset into cross-validation, the keyword Stratified K-Fold is used with the parameter setting of the number of folds as 10, shuffle as True, and random state as 42.

In the case of the deep learning models used in automatic feature extraction, the CNN connected with the LSTM is applied for the feature extraction of the EEG data. The structure of the model consists of an LSTM layer with 64 neurons and dense layers with 128 and 64 neurons, optimizer used is Adam and loss function used is Mean Squared Error. For the eye-tracking data the LeNet-5 is employed; it consists of two dense layers with 120 and 84 units and a sigmoid layer is used at the output for binary classification. This model is trained with the Adam optimizer with binary cross-entropy as the loss function and accuracy as the parameter over 50 epochs. The data splitting involves a train test split of 80–20 and further division of the remaining data in equal proportions to validate and test the model.

4.7.1 Stratified cross-validation

Once we were set up, with the ensemble classification pipeline formulated, the next logical step was to assess its utility. To this end, the strategy used was a rigorous method known as the stratified 10-fold cross-validation (Kohavi, 1995). However, stratified cross-validation goes one step further than this as it guarantees the resultant folds as having the same proportion of classes as those of the original data-set.

5 Results and discussion

The efficiency of classification models is evaluated in terms of the metrics that measure the ability of the ML algorithm to classify the objects appropriately. Selecting the appropriate metrics is essential for achieving an accurate and objective assessment and measuring performance in such problems with skewed classes or different costs associated with an error. Accuracy for the most basic performance indicator that show the number of instances out of all the data that belong to the correct class. Precision also known as positive predictive value, measures the proportion of true positives among all predicted positives. It reflects how often the model correctly identifies a positive case.

Specificity test evaluates the proportion of actual negatives which are correctly identified by the model as negative, while, recall or sensitivity evaluates the proportion of actual positives which are correctly identified by the model as positive. They indicate how well the model captures all the positive instances in relation to the available training examples. F1 score is an average of recall and precision that yields proportional insights into both these measures. It's particularly useful when both false positives and false negatives are equally undesirable. It's particularly useful when both false positives and false negatives are equally undesirable.

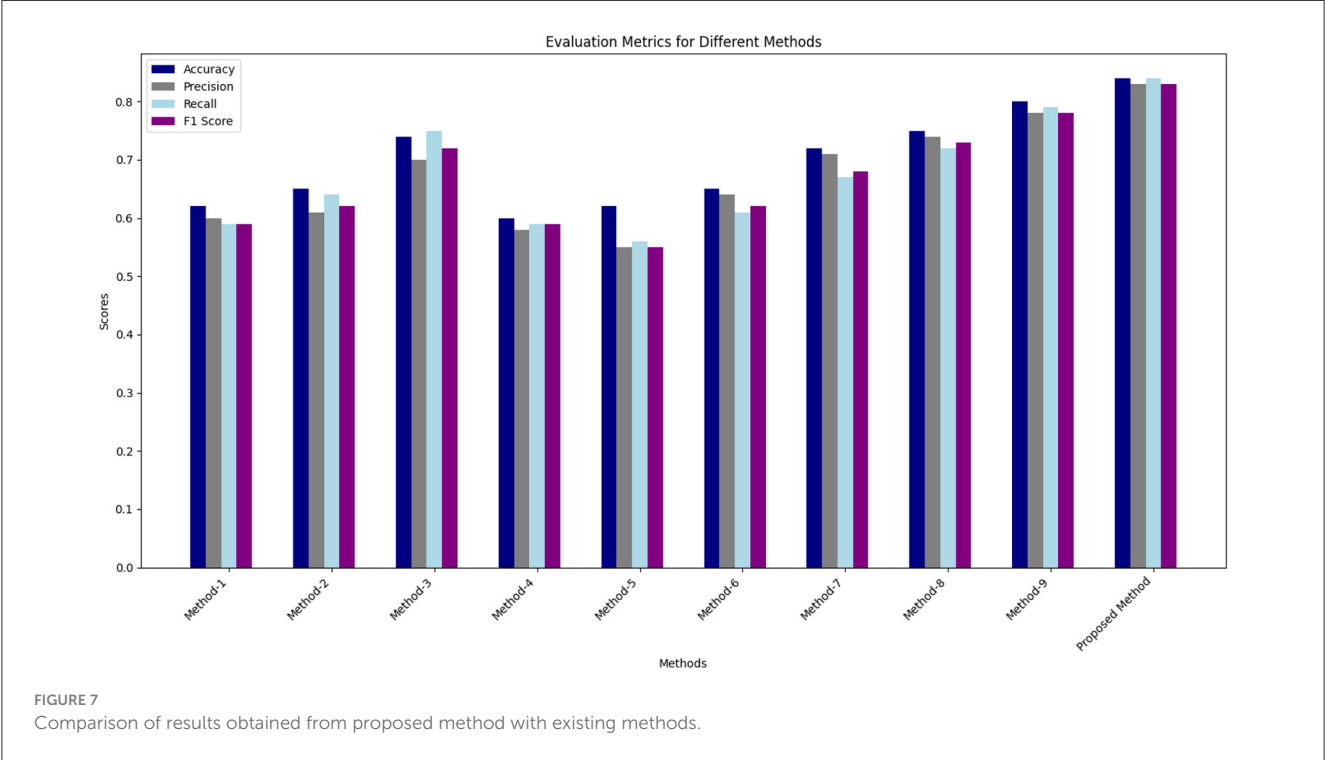
Table 6 represents the quantitative comparison of the employed methods, namely accuracy, precision rate, recall, and F1 score. The proposed method achieves the highest accuracy of 0.84, significantly outperforming the other methods. The improvement in accuracy can be attributed to the effective integration of ML and DL features along with the stacking ensemble technique. The precision of the proposed method (0.83) indicates its superior ability to correctly identify positive instances compared to other methods. This is particularly important in reducing false positives, which is critical in applications where the cost of false positives is high. For recall the proposed method gives 0.84 which shows that the proposed method is also good in the recall sense it captures most of the true positive instances. Large recall component means that the model is going to include many more positives into the result set at the cost of possibly including negative instances, which is particularly important where false negatives are undesirable. The proposed method claims to achieve an F1 score of 0.83 which balances the precision and recall rates of identifying fishes with equal importance. This is because F1 score is a harmonic mean of precision and recall and a high F1 score indicates that the proposed method is both precise and accurate, studied and tested on different measure standards and tables.

Table 6 represents the quantitative comparison of the employed methods, namely accuracy, precision rate, recall, and F1 score.

TABLE 6 Evaluation metrics for different methods.

Method	Accuracy	Precision	Recall	F1 score
1- EEG (not preprocessed, ML features, SVM)	0.62	0.60	0.59	0.59
2- EEG (preprocessed, DL features, SVM + RF)	0.65	0.61	0.64	0.62
3- EEG (preprocessed, ML + DL features, SVM + RF + DT)	0.74	0.70	0.75	0.72
4- ET (not preprocessed, ML features, RF)	0.60	0.58	0.59	0.59
5- ET (preprocessed, DL features, DT)	0.62	0.55	0.56	0.55
6- ET (preprocessed, ML + DL features, XGB + RF + DT)	0.65	0.64	0.61	0.62
7- EEG & ET (not preprocessed, ML features, SVM)	0.72	0.71	0.67	0.68
8- EEG & ET (preprocessed, DL features, RF)	0.75	0.74	0.72	0.73
9- EEG & ET (preprocessed, ML + DL features, SVM + RF + XGB)	0.80	0.78	0.79	0.78
Proposed- EEG & ET (preprocessed, ML + DL features, stacking ensemble)	0.84	0.83	0.84	0.83

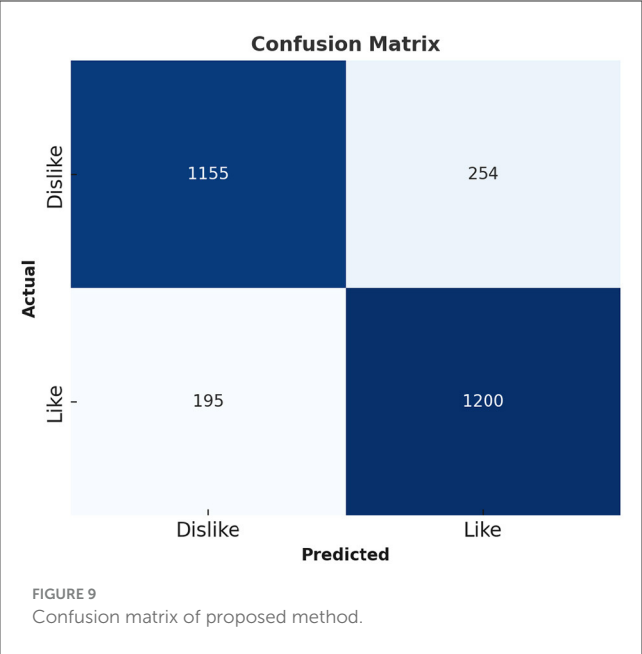
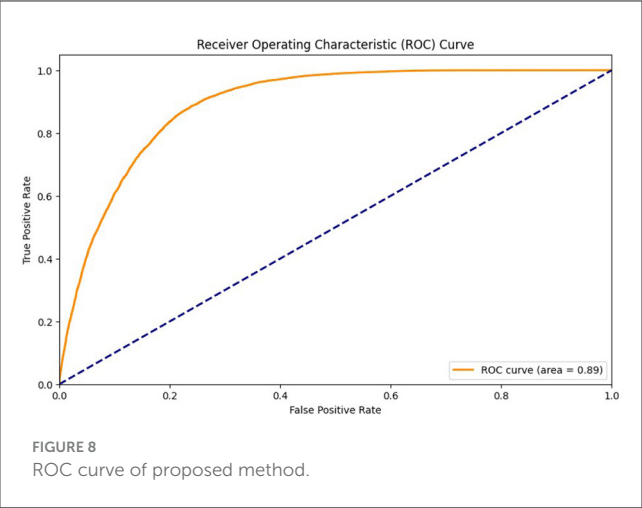
Ablation study has been performed and bold values show the results obtained from final methodology.



Figures 7–9 show the evaluation score of the method, area under ROC curve, and Confusion Matrix of the proposed method. The ROC curve is a graphical approach that indicates a model's performance at different classification hurdles. It maps True Positive Rate or Sensitivity on the y-axis, against False Positive Rate or Fall out on the x-axis. An ideal ROC curve looks like a graph that plots the data close to the upper left-hand corner of the axes, which means that the performance of the model was satisfactory and it could distinguish between the classes accurately. The AUC gives overall performance of the ROC curve, from this the probability that the model ranks positive instance higher to

a randomly chosen negative instance can be determined. Higher AUC shows that the tester has better ability in classifying. AUC-ROC of 0.89 has been achieved as shown in the Figure 8, whereas, confusion matrix is presented in Figure 9 which further proves that it is highly effective when it comes to discriminating between the positive and the negative classes. This score can be classified within the “good” region; hence it can be deduced that the method purposed is good in segregating the two classes of interest.

Figure 7 compares the results obtained from proposed method with the existing state of the art methods. Table 6 describes the evaluation criteria to different methods. As can be seen, the



proposed method that uses the given preprocessing for EEG and ET data and incorporates the features of ML as well as DL within stacking ensemble provides the highest results on all of the listed measures.

At the highest accuracy, Method 1 employing raw EEG data with ML features and SVM yielded an accuracy 0.62. At the same time, the proposed method is much more effective with accuracy 0.84. It can be seen that this improvement is universality for precision, recall, and F1 score, more manifesting the advantages of data preprocessing and more successful attempt of the stacking ensemble method combining the ML + DL features. If we compare the methods in which preprocessing was used (e.g., Method 1) with those for which preprocessing was not used (e.g., Method 2), one can see that, in many cases, preprocessing has a positive effect on the performance. For instance, in method number 2, the EEG data is preprocessed and the DL features generates higher percentages of accuracy and recall than in method 1.

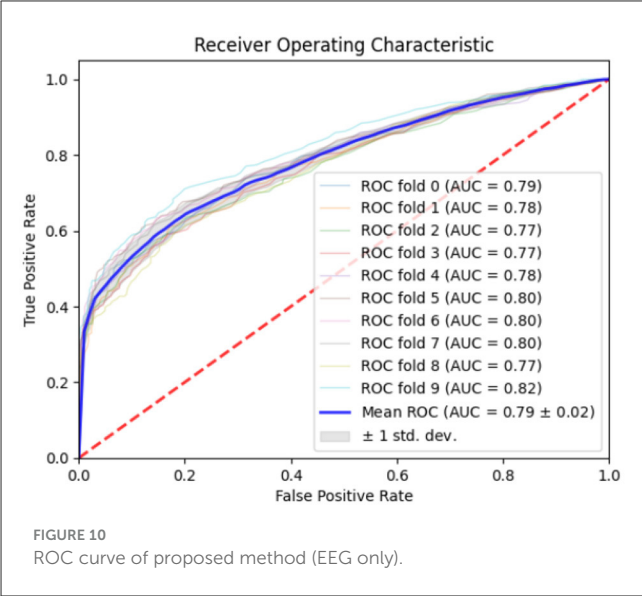
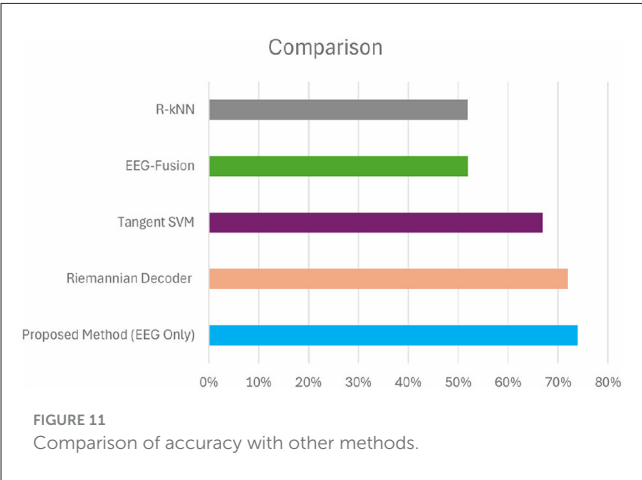


TABLE 7 Comparison with Georgiadis et al. (2023a)

Aspect	Georgiadis et al. (2023a)	Proposed method (only EEG)
Dataset	NeuMa	NeuMa
Accuracy	0.72	0.74
Precision	Not mentioned	0.70
Recall	Not mentioned	0.75
F1-score	Not mentioned	0.72
AUC score	Not mentioned	0.79



Methods which simultaneously utilize both EEG and ET data are superior to the methods based on only one type of data. For example, the feature that incorporates preprocessed EEG and ET data with conventional ML features and DL results in Method 9 has an accuracy of 0.80. This shows when there is an integration of the EEG and ET data it is able to provide better results for the model. The proposed method incorporates stacking ensemble, which also improves the performance of classifiers due to features adopted by

this method. This leads to the highest values on all accounts, hence promoting a resilient and efficient model.

Classification results for EEG data when analyzed with the help of ML and DL incorporated with SVM, RF, and DT were seen to be quite satisfactory. By using the approach, the objective was attained with a 0.74 accuracy, and the precision, recall, and the F1 score equal to 0.70%, 0.75%, and 0.72% respectively. Furthermore, the have relatively high AUC, mean of 0.79 as shown in [Figure 10](#) therefore support the reliability and discriminant capacity of the developed model, in the classification of consumers' preferences from EEG signals. Altogether, these metrics can be discussed as demonstrating the efficiency of the proposed approach of applying the traditional ML algorithms alongside with the DL features.

On the other hand, the study that the referenced paper dealt with proposed a new deep learning decoder based on Riemannian Geometry and SPDNet structure ([Georgiadis et al., 2023a](#)) for analyzing the signals of the NeuMa dataset of EEG. From the research, the investigators obtained a mean accuracy of 72%. [Table 7](#) shows a comparison between our proposed method(EEG Only) with a state-of-art method of [Georgiadis et al. \(2023a\)](#). The comparison of accuracies is shown in [Figure 11](#). Although this research infuses ML and DL with regular classifiers, the paper's presentation of domain's Riemannian Geometry and SPDNet demonstrates higher accuracy than conventional EEG- based approaches like Tangent Space SVM ([Kalaganis et al., 2019](#)), EEG-Fusion ([Hakim et al., 2021](#)) and R-kNN ([Congedo et al., 2017](#)). The statistical significance thus obtained particularly with reference to the results achieved by Tangent Space SVM which was 67.72%, EEG-Fusion 52.75% and R-kNN was 51.96%.

There are some limitations to the study that need to be noted. First off, although 42 participants is a small sample size, it might not be enough to extrapolate the results to a broader population. Furthermore, the findings are predicated on a particular dataset, which can restrict their generalizability to other product categories, markets, and cultural settings. The accuracy of the data acquired may be affected by the sampling rate and precision constraints of the EEG and eye-tracking sensors, despite their effectiveness. Furthermore, even though the used feature extraction strategies which combined manually created and automatically generated features proved successful, more research into different approaches or sophisticated deep learning architectures may enhance model performance. Finally, the integration of EEG and ET data adds complexity to the analysis, and potential synchronization challenges may have influenced the overall accuracy of the model.

6 Conclusions

Prediction of consumer preferences that we suggest is based on the machine learning and deep neural network methodology characterized by a high degree of accuracy and precision. These results could have been achieved because of correct preprocessing of images, use of the right features, and the high accuracy classifier. In preprocessing, we have increased the signal-to-noise ratio of EEG signals and ET data by removing noise and balanced the number of samples for classes, specifically the Buy class, by creating more through SMOTE. From the EEG and ET dataset, we created manual features by using

the same method as before. Similarly, we used CNN-LSTM for the feature extraction of the selected EEG signals and LeNet-5 for the ET data. In classification, a most dependable stacking classifier was used for classification with a high level of accuracy.

The proposed method demonstrates stable results in the context of consumer preference prediction, though there are opportunities for future studies. For the current extraction feature, we could definitely do better in terms of advanced methodologies such as deep learning architectures or location of brain sources. Classification methods could be enhanced by considering other subject-dependent models or by developing the concept of a prediction. Generalizability is critical, which makes cross-validation mandatory across larger and more diverse data sets. Furthermore, it is necessary to discuss the similarities and differences of the proposed approach with other neuromarketing methods, as well as consider issues of the user's consent and data privacy. In addition, extending this method for uses outside of e-commerce, such as physical store promotion or measuring ad campaign effectiveness, provides more arenas for possible research and practical implementation.

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

SU: Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. SK: Methodology, Resources, Supervision, Writing – original draft, Writing – review & editing. AT: Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. AI: Supervision, Writing – original draft, Writing – review & editing. MZ: Formal analysis, Validation, Writing – original draft, 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.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

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

- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). "Optuna: a next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (New York, NY: ACM), 2623–2631. doi: 10.1145/3292500.3330701
- Aldayel, M., Ykhlef, M., and Al-Nafjan, A. (2021). Recognition of consumer preference by analysis and classification EEG signals. *Front. Hum. Neurosci.* 14:604639. doi: 10.3389/fnhum.2020.604639
- Alimardani, M., and Kaba, M. (2021). "Deep learning for neuromarketing: classification of user preference using EEG signals," in *12th Augmented Human International Conference* (New York, NY: ACM), 1–7. doi: 10.1145/3460881.3460930
- Arif, S., Munawar, S., Marie, R. R., and Shah, S. A. (2023). "Leveraging wavelets and deep CNN for sleep pattern recognition in road safety: an EEG study," in *International Conference on Recent Trends in Image Processing and Pattern Recognition* (Cham: Springer), 227–241. doi: 10.1007/978-3-031-53082-1_19
- Armstrong, G., Adam, S., Denize, S., and Kotler, P. (2014). *Principles of Marketing*. Melbourne, VIC: Pearson Australia.
- Assel, H. (1995). *Consumer behavior and marketing action fifth edition*. Ohio: South.
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 321–357. doi: 10.1613/jair.953
- Chen, T., and Guestrin, C. (2016). "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: ACM), 785–794. doi: 10.1145/2939672.2939785
- Christoforou, C., Papadopoulos, T. C., Constantinidou, F., and Theodorou, M. (2017). Your brain on the movies: a computational approach for predicting box-office performance from viewer's brain responses to movie trailers. *Front. Neuroinform.* 11:72. doi: 10.3389/fninf.2017.00072
- Congedo, M., Barachant, A., and Bhatia, R. (2017). Riemannian geometry for eeg-based brain-computer interfaces: a primer and a review. *Brain-Comput. Interfaces* 4, 155–174. doi: 10.1080/2326263X.2017.1297192
- Fisch, B. J. (1999). *Fisch and Spelmann's EEG Primer: Basic Principles of Digital and Analog EEG*. Amsterdam: Elsevier.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29, 1189–1232. doi: 10.1214/aos/1013203451
- García-Madariaga, J., López, M.-F. B., Burgos, I. M., and Virto, N. R. (2019). Do isolated packaging variables influence consumers' attention and preferences? *Physiol. Behav.* 200, 96–103. doi: 10.1016/j.physbeh.2018.04.030
- Georgiadis, K., Kalaganis, F. P., Oikonomou, V. P., Nikolopoulos, S., Laskaris, N. A., Kompatsiaris, I., et al. (2022). Rneumark: a riemannian EEG analysis framework for neuromarketing. *Brain Inf.* 9:22. doi: 10.1186/s40708-022-00171-7
- Georgiadis, K., Kalaganis, F. P., Oikonomou, V. P., Nikolopoulos, S., Laskaris, N. A., Kompatsiaris, I., et al. (2023a). "Harnessing the potential of EEG in neuromarketing with deep learning and riemannian geometry," in *International Conference on Brain Informatics* (Cham: Springer), 21–32. doi: 10.1007/978-3-031-43075-6_3
- Georgiadis, K., Kalaganis, F. P., Riskos, K., Matta, E., Oikonomou, V. P., Yfantiidou, I., et al. (2023b). Neuma-the absolute neuromarketing dataset en route to an holistic understanding of consumer behaviour. *Sci. Data* 10:508. doi: 10.1038/s41597-023-02392-9
- Hakim, A., Klorfeld, S., Sela, T., Friedman, D., Shabat-Simon, M., Levy, D. J., et al. (2021). Machines learn neuromarketing: improving preference prediction from self-reports using multiple EEG measures and machine learning. *Int. J. Res. Mark.* 38, 770–791. doi: 10.1016/j.ijresmar.2020.10.005
- Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.1735
- Homan, R., Herman, J., and Purdy, P. (1987). Cerebral location of international 10-20 system electrode placement. *Electroencephalogr. Clin. Neurophysiol.* 66, 376–382. doi: 10.1016/0013-4694(87)90206-9
- Kalaganis, F. P., Laskaris, N. A., Chatzilari, E., Nikolopoulos, S., and Kompatsiaris, I. (2019). A Riemannian geometry approach to reduced and discriminative covariance estimation in brain computer interfaces. *IEEE Trans. Biomed. Eng.* 67, 245–255. doi: 10.1109/TBME.2019.2912066
- Khushaba, R. N., Wise, C., Kodagoda, S., Louviere, J., Kahn, B. E., Townsend, C., et al. (2013). Consumer neuroscience: assessing the brain response to marketing stimuli using electroencephalogram (EEG) and eye tracking. *Expert Syst. Appl.* 40, 3803–3812. doi: 10.1016/j.eswa.2012.12.095
- Kohavi, R. (1995). "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the 14th international joint conference on Artificial intelligence, Volume 2* (Princeton, NJ: Citeseer), 1137–1143.
- Kumar, S., Yadava, M., and Roy, P. P. (2019). Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction. *Inf. Fusion* 52, 41–52. doi: 10.1016/j.inffus.2018.11.001
- LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791
- Mashrur, F. R., Rahman, K. M., Miya, M. T. I., Vaidyanathan, R., Anwar, S. F., Sarker, F., et al. (2024). Intelligent neuromarketing framework for consumers' preference prediction from electroencephalography signals and eye tracking. *J. Consum. Behav.* 23, 1146–1157. doi: 10.1002/cb.2253
- Matukin, M., Ohme, R., and Boshoff, C. (2016). Toward a better understanding of advertising stimuli processing: exploring the link between consumers' eye fixation and their subconscious responses. *J. Advert. Res.* 56, 205–216. doi: 10.2501/JAR-2016-017
- Murugappan, M., Murugappan, S., Gerard, C., et al. (2014). "Wireless EEG signals based neuromarketing system using Fast Fourier transform (FFT)," in *2014 IEEE 10th international colloquium on signal processing and its applications* (Kuala Lumpur: IEEE), 25–30. doi: 10.1109/CSPA.2014.6805714
- Peter, J. P., Olson, J. C., and Grunert, K. G. (1999). *Consumer Behaviour and Marketing Strategy. Edisi IV*. Jakarta: PT Erlangga.
- Samsuri, N., Begum, T., Reza, F., Omar, H., Yusoff, N., Idris, B., et al. (2016). Left lateral singular (LLS) view enhances attention at stimulus discrimination during automobile display setting revealed by EEG/ERP and eye tracking technique. *Int. J. Enhanc. Res. Sci. Technol. Eng.* 5, 1–9.
- Shah, S. M. A., Usman, S. M., Khalid, S., Rehman, I. U., Anwar, A., Hussain, S., et al. (2022). An ensemble model for consumer emotion prediction using EEG signals for neuromarketing applications. *Sensors* 22:9744. doi: 10.3390/s22249744
- Slanzi, G., Balazs, J. A., and Velásquez, J. D. (2017). Combining eye tracking, pupil dilation and EEG analysis for predicting web users click intention. *Inf. Fusion* 35, 51–57. doi: 10.1016/j.inffus.2016.09.003
- Telpaz, A., Webb, R., and Levy, D. J. (2015). Using EEG to predict consumers' future choices. *J. Mark. Res.* 52, 511–529. doi: 10.1509/jmr.13.0564
- Teplan, M. (2002). Fundamentals of EEG measurement. *Meas. Sci. Rev.* 2, 1–11.
- Tobii, A. B. (2024). *What is eye tracking?* Available at: <https://www.tobii.com/learn-and-support/get-started/what-is-eye-tracking> (accessed December 30, 2024).
- Vecchiato, G., Astolfi, L., Fallani, F. D. V., Toppi, J., Aloise, F., Bez, F., et al. (2011). On the use of EEG or MEG brain imaging tools in neuromarketing research. *Comput. Intell. Neurosci.* 2011:643489. doi: 10.1155/2011/643489
- Welch, P. (1967). The use of Fast Fourier Transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* 15, 70–73. doi: 10.1109/TAU.1967.1161901
- Wolpert, D. H. (1992). Stacked generalization. *Neural. Netw.* 5, 241–259. doi: 10.1016/S0893-6080(05)80023-1
- Yadava, M., Kumar, P., Saini, R., Roy, P. P., and Prosad Dogra, D. (2017). Analysis of EEG signals and its application to neuromarketing. *Multimed. Tools Appl.* 76, 19087–19111. doi: 10.1007/s11042-017-4580-6



OPEN ACCESS

EDITED BY

Saad Arif,
King Faisal University, Saudi Arabia

REVIEWED BY

Ammar Armghan,
Jouf University, Saudi Arabia
Munam Shah,
King Faisal University, Saudi Arabia

*CORRESPONDENCE

Sumaira Tabassum
✉ sumairatabassum.phd@smme.edu.pk

RECEIVED 10 November 2024

ACCEPTED 30 December 2024

PUBLISHED 22 January 2025

CITATION

Tabassum S, Khan MJ, Iqbal J, Waris A and Ijaz MA (2025) Automated karyogram analysis for early detection of genetic and neurodegenerative disorders: a hybrid machine learning approach.
Front. Comput. Neurosci. 18:1525895.
doi: 10.3389/fncom.2024.1525895

COPYRIGHT

© 2025 Tabassum, Khan, Iqbal, Waris and Ijaz. 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.

Automated karyogram analysis for early detection of genetic and neurodegenerative disorders: a hybrid machine learning approach

Sumaira Tabassum*, M. Jawad Khan, Javaid Iqbal, Asim Waris and M. Adeel Ijaz

Department of Robotics and Artificial Intelligence, School of Mechanical and Manufacturing Engineering, National University of Sciences and Technology, Islamabad, Pakistan

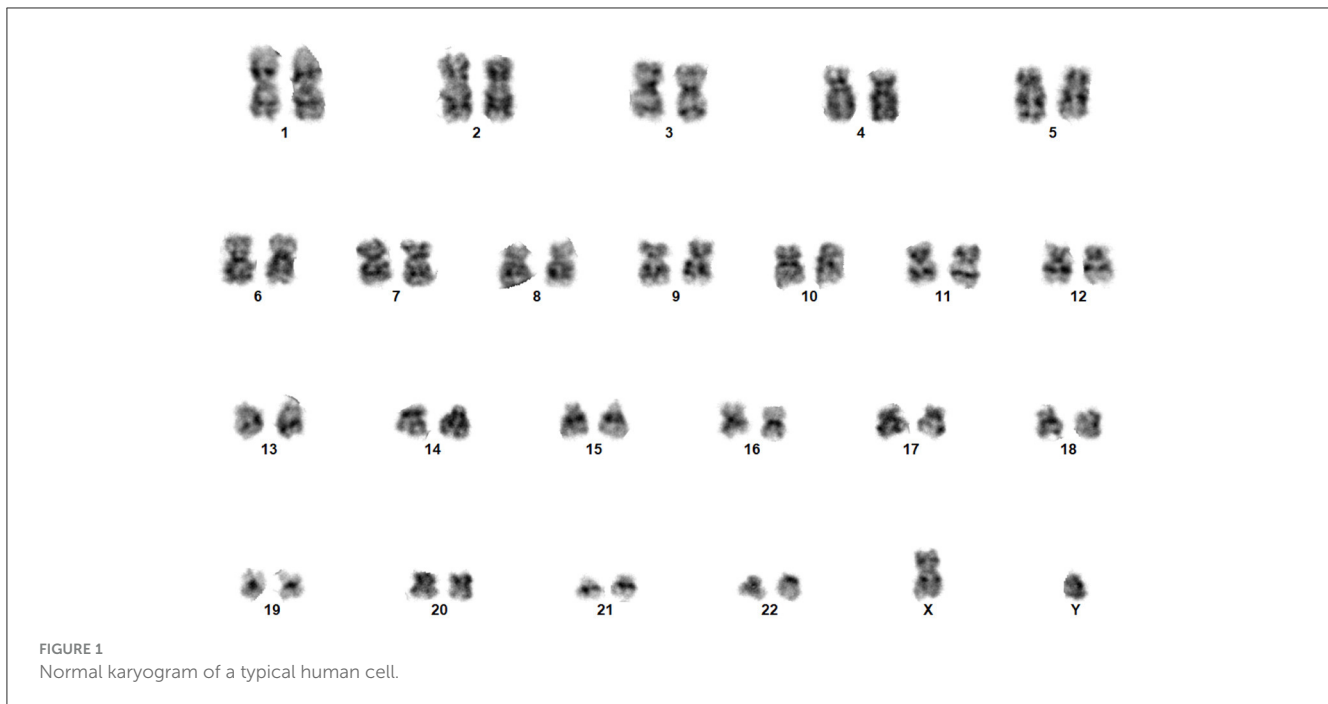
Anomalous chromosomes are the cause of genetic diseases such as cancer, Alzheimer's, Parkinson's, epilepsy, and autism. Karyotype analysis is the standard procedure for diagnosing genetic disorders. Identifying anomalies is often costly, time-consuming, heavily reliant on expert interpretation, and requires considerable manual effort. Efforts are being made to automate karyogram analysis. However, the unavailability of large datasets, particularly those including samples with chromosomal abnormalities, presents a significant challenge. The development of automated models requires extensive labeled and incredibly abnormal data to accurately identify and analyze abnormalities, which are difficult to obtain in sufficient quantities. Although the deep learning-based architecture has yielded state-of-the-art performance in medical image anomaly detection, it cannot be generalized well because of the lack of anomalous datasets. This study introduces a novel hybrid approach that combines unsupervised and supervised learning techniques to overcome the challenges of limited labeled data and scalability in chromosomal analysis. An Autoencoder-based system is initially trained with unlabeled data to identify chromosome patterns. It is fine-tuned on labeled data, followed by a classification step using a Convolutional Neural Network (CNN). A unique dataset of 234,259 chromosome images, including the training, validation, and test sets, was used. Marking a significant achievement in the scale of chromosomal analysis. The proposed hybrid system accurately detects structural anomalies in individual chromosome images, achieving 99.3% accuracy in classifying normal and abnormal chromosomes. We also used a structural similarity index measure and template matching to identify the part of the abnormal chromosome that differed from the normal one. This automated model has the potential to significantly contribute to the early detection and diagnosis of chromosome-related disorders that affect both genetic health and neurological behavior.

KEYWORDS

chromosome anomalies, cognitive sciences, machine learning, neurological health, neurodevelopmental disorders, neurological disorders, neuroscience, genetic diseases

1 Introduction

A chromosome is a thread-like structure that harbors genetic information encoded in genes. Located within the nuclei of cells in most living organisms, it comprises proteins and a solitary Deoxyribonucleic Acid (DNA) molecule. The structure of the chromosomes is shown in [Figure 1](#). It transports genomic information from one cell to another



(Institute, 2023). A typical human cell contains 46 chromosomes, comprising 22 pairs of single chromosomes (autosomes), which are numbered (1–22), and two sex chromosomes (XX or XY) (Institute, 2023). Chromosomes become visible during metaphase when stained with Giemsa and viewed under a light microscope. Understanding human chromosomes is crucial for diagnosing and predicting outcomes and tracking treatment progress under various conditions (Gersen, 2013). Cytogenetic experiments were performed to determine chromosomal abnormalities. Cytogenetics encompasses the examination of tissues, blood, bone marrow, and cultured cells in a laboratory setting. This field uses banding or manipulation techniques to identify chromosomal alterations (Natarajan, 2002).

Genetic diseases result directly from chromosomal abnormalities, and detecting chromosomal anomalies can anticipate and alert medical practitioners to potential diseases stemming from these abnormalities (Natarajan, 2002). Effective identification of chromosomal abnormalities is of significant clinical importance. Detecting genetic abnormalities in patients at the earliest stage is essential for timely and effective treatment. Chromosomal abnormalities are associated with genetic disorders. Changes in chromosome number or structure affect neurological health, such as Alzheimer's, Parkinson's, epilepsy, autism, and many other conditions. This can be detected using karyotyping. It is widely used for prenatal and fetal chromosome screening. The early detection of fetal chromosomal abnormalities can provide insights for detecting possible neurological and developmental abnormalities (Rosenfeld and Patel, 2017). Machine learning has been widely used in the detection of neurological disorders as it is used for the classification and segmentation of neurological images.

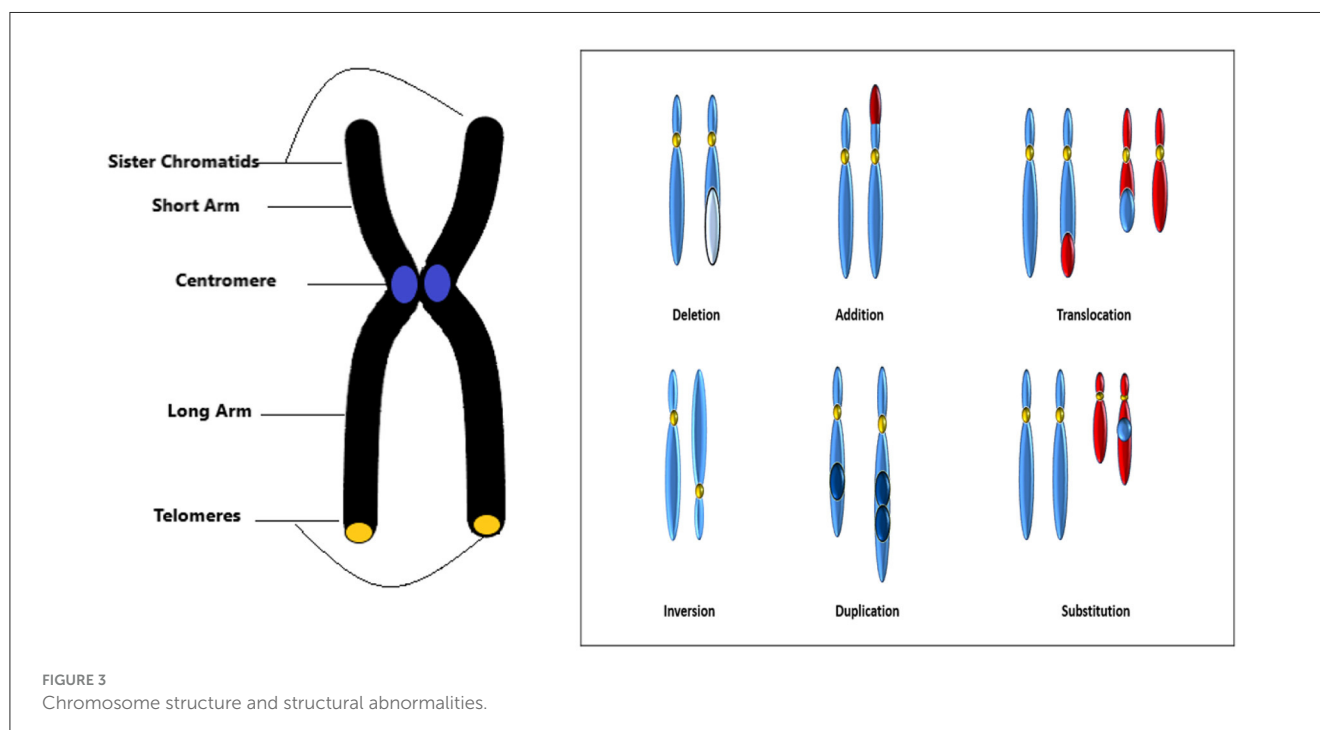
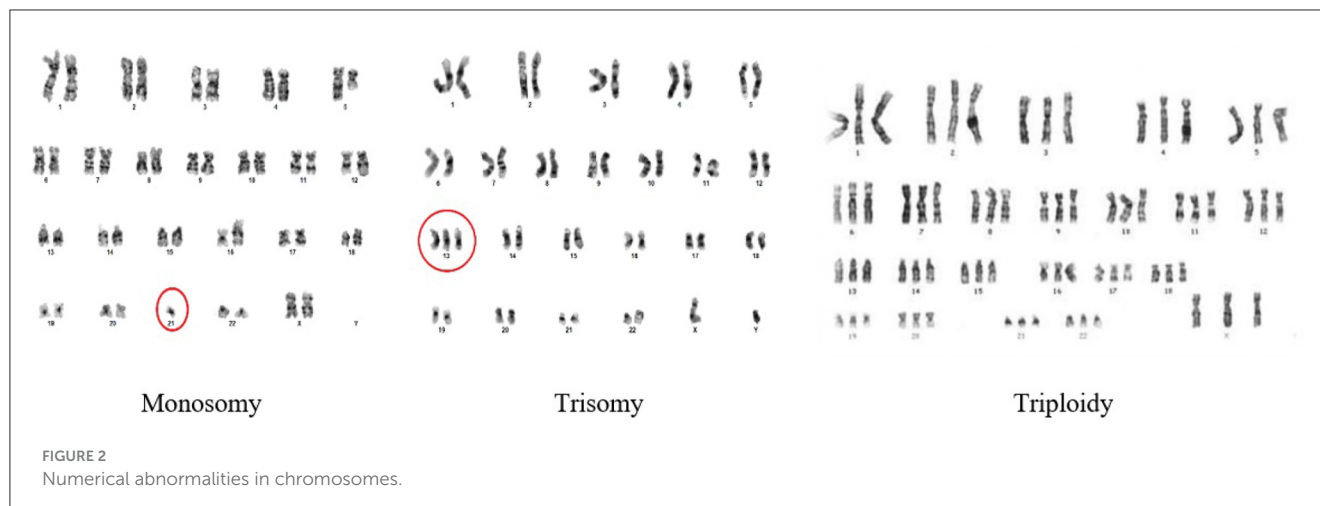
Chromosomal disorders can be categorized into two primary types: numerical and structural abnormalities. A **numerical abnormality** signifies that an individual either lacks one of

the chromosomes from a pair, or possesses more than two chromosomes instead of the usual pair. Numerical disorders arise from changes in the number of chromosomes, resulting in deviations from the expected count of 46. Examples of numerical disorders include trisomy, monosomy, and triploidy. Figure 2 shows the types of numerical abnormalities.

A **trisomy** occurs when a person has three of a particular chromosome instead of the usual two. Down Syndrome is caused by trisomy21. A **monosomy** occurs when they have just one chromosome instead of the usual two chromosomes. **Triploidy** is rare; however, in this type of abnormality, an extra third chromosome for each class is present in the cells.

Structural abnormalities indicate that the structure of the chromosome has been modified in various ways. Structural chromosomal disorders emerge from breakages within a chromosome or the incorrect rejoining of chromosomal segments. In such disorders, the number of copies of any given gene may exceed or fall short of two typical copies. Deletion, duplication, inversion, substitution, and translocation anomalies of the chromosomes are shown in Figure 3.

Upon **deletion**, a chromosome segment is absent or deleted. This causes many abnormalities, for example deletion in chromosome 15 can cause angelman syndrome. In **duplication**, a portion of the chromosome is duplicated leading to excess genetic material like Dup15q Syndrome is caused by duplication of chromosome 15. In **inversion**, a chromosome segment may undergo problems such as breakage, can be turned upside down, and can have subsequent reattachment, causing inversion of the genetic material. **Substitution** occurs when a portion of a chromosome is replaced with a portion of another chromosome. **Translocation** appears when a part from one chromosome is moved to another. Translocation can be further divided into two types of **reciprocal translocation**, which occurs when segments from two distinct chromosomes have been interchanged, and



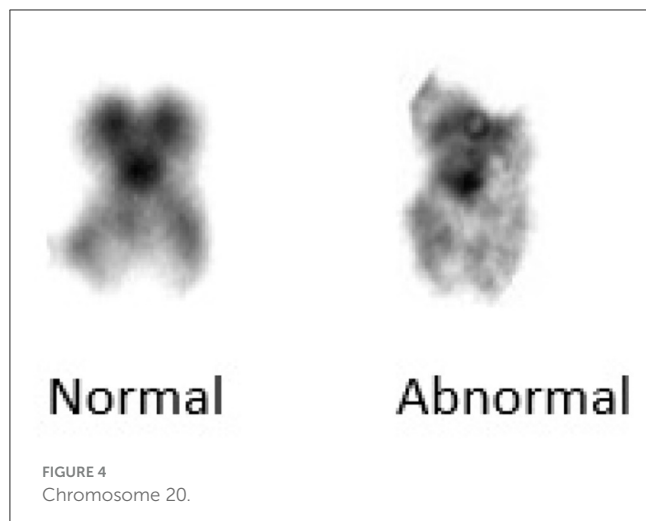
Robesonian translocation occurs when an entire chromosome moves and fixes itself to another chromosome's centromere. In [Figure 4](#), we show an example image of del20q chromosomes from our dataset.

Deletion, duplication, mutation, and trisomy are causes of cancer and neurological disorders such as epilepsy, Down syndrome, and autism spectrum disorder (ASD) syndrome. Neurological disorders are typically studied using electroencephalogram (EEG), ultrasonography, and magnetic resonance imaging (MRI). However, these techniques are usually applied after the onset of symptoms. These methods effectively monitor brain function once they are developed and visible. Genetic predispositions during the early developmental stages can be identified through chromosome analysis, which can help in the early diagnosis of such diseases. For this, fetal samples were collected and analyzed by karyotyping. This could help to identify any anomaly in chromosomes at the early stage of development,

such as neurological disorders, before symptoms manifest. This way, karyotyping offers a more proactive approach to treatment and management.

1.1 Related work

Genetic diseases are mainly identified by karyotyping, but there are some diseases that different imaging techniques can identify. Methods commonly used for the detection of neurological disorders such as epilepsy often rely on EEG signals and various imaging techniques such as MRI. Machine learning has been used to automate the classification process of these techniques. Similar to multidomain feature fusion and selection approach proposed by [Kong et al. \(2024\)](#), it uses advanced signal processing and machine-learning techniques to optimize feature extraction and classification.



Machine learning (ML) has transformed healthcare by offering practical applications that have enhanced diagnosis, treatment, monitoring, and decision-making across various clinical domains. From the early detection of diseases to personalized treatment planning, automated reporting, and predictive analytics, ML models support healthcare practitioners in delivering more accurate, efficient, and scalable clinical solutions. This section outlines the key practical applications of ML in clinical workflows across different areas of healthcare, showcasing its versatility and impact beyond specialized fields like cytogenetics. For example, AI models are used in medical imaging to review X-rays, MRIs, and Computed Tomography scans to identify fractures, tumors, and organ failures as efficiently and accurately as possible.

[Ibrahim et al. \(2024\)](#) explored how deep learning using a pre-trained AlexNet model can help classify chest X-ray (CXR) images into four categories: COVID-19 pneumonia, non-COVID-19 viral pneumonia, bacterial pneumonia, and routine. [Ahmad et al. \(2024\)](#) introduced a computer-aided diagnosis (CAD) system for detecting breast cancer by combining deep learning and computer vision techniques. [Islam et al. \(2024\)](#) introduce BrainNet, a deep learning method for accurately classifying brain tumors using MRI images.

[Montobbio et al. \(2024\)](#) emphasized the potential and challenges of computational modeling and machine learning approaches for diagnosing and treating neurological disorders. Their insights, particularly in disease diagnosis, classification, and personalized therapeutic strategies, highlight the promising applications of these techniques. All of them used EEG and MRI images. [Duarte et al. \(2024\)](#) used flair images and machine learning for segmentation tasks. Alzheimer's disease (AD) was also diagnosed by [Slimi et al. \(2024\)](#) using machine learning on MRI images, and [Li and Zhong \(2024\)](#) explored the integration of deep learning in neuroscience, highlighting key trends and identifying major research hotspots in the field. Therefore, machine learning has been widely used for diagnosing such diseases but with different images adopted from different imaging techniques, as discussed earlier.

Anomaly detection by karyographic analysis is a common technique used to identify any numerical or structural abnormalities in human chromosomes. The conventional method

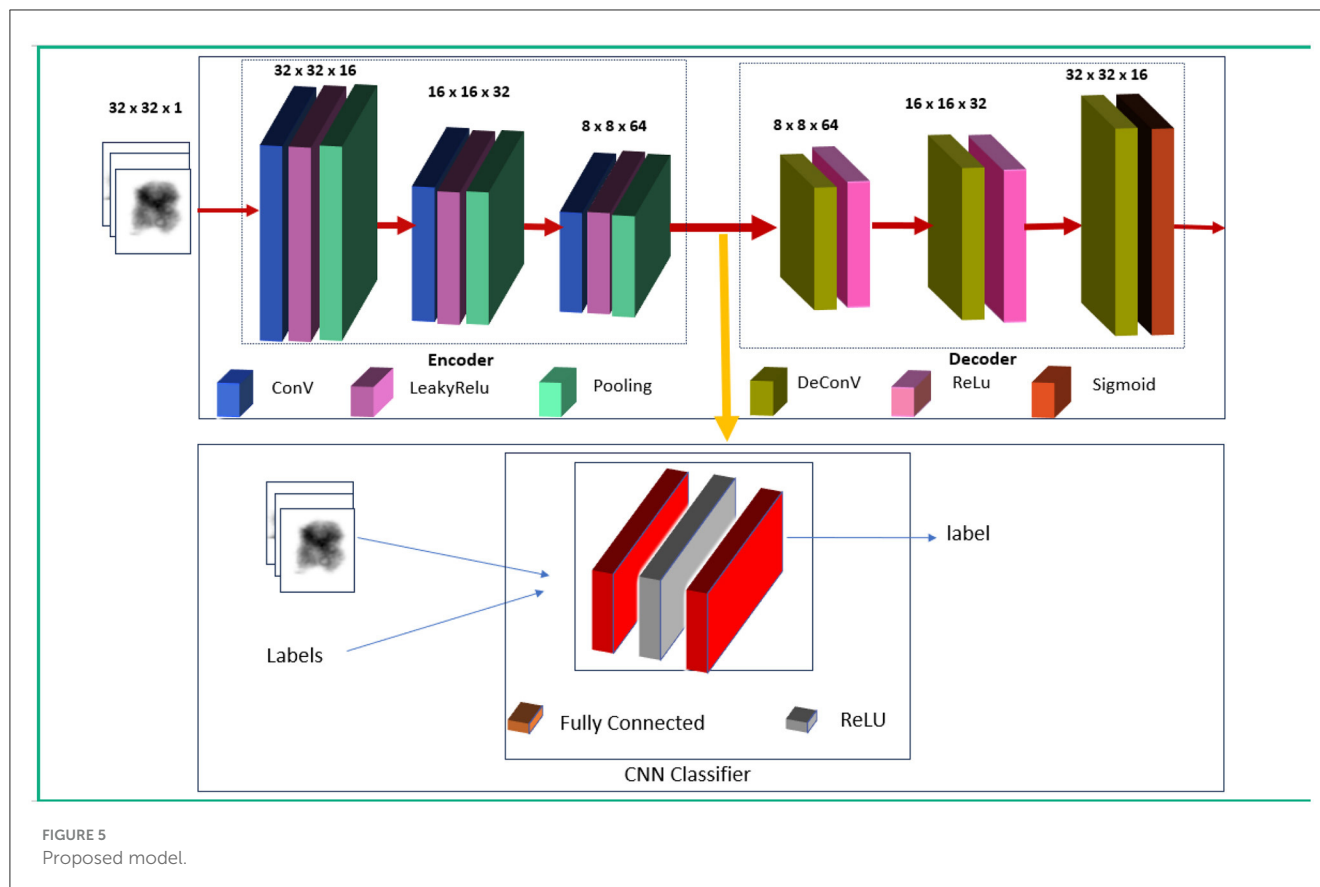
for classifying chromosomes in most cytogenetic laboratories involves manual work by skilled experts. This procedure is time-consuming and requires significant effort from experienced operators, making it expensive. Experts commonly examine microscopic chromosome images in the conventional analysis of chromosomal anomalies, relying on their experience and expertise in detecting abnormalities that may lead to genetic disorders, congenital disabilities, or even cancer ([Britto and Ravindran, 2007](#)). The analysis of chromosome morphology involves a sequence of procedures, including selecting metaphase chromosome images. This encompasses the segmentation of individual chromosomes ([Poletti et al., 2012](#)), the classification of chromosomes ([Madian et al., 2018](#)), and the detection of chromosomal anomalies ([Park et al., 2019](#)). Significant efforts are being made to investigate how machine learning can improve pathological diagnosis. Deep learning technologies have experienced widespread adoption in recent years. The efficacy of these methods lies in their robust capacity for automatic feature extraction and learning from images, making them well-suited for the development of automated image analysis systems.

In medicine, artificial intelligence (AI) is being implemented, although some challenges exist. For example, the availability of labeled data is often limited, and labeling itself is challenging because of a lack of domain knowledge. Medical images containing anomalies are increasingly being analyzed using artificial intelligence. Aberrations, alternatively termed abnormalities, anomalies, or outliers, are often challenges in anomaly detection. The increasing popularity of deep learning-based anomaly detection algorithms is also facilitated by advancements in computational power and availability of big data.

Detecting aberrations poses a persistent challenge, particularly in the case of clonal chromosomal abnormalities in hematological malignancies. These abnormalities are characterized by their high complexity, diversity, and occasional rarity ([Fang et al., 2023](#)). To date, deep learning methods have been applied for detecting chromosomal abnormalities; however, challenges have arisen regarding data availability. Deep learning models rely heavily on data, and when it comes to the analysis of chromosomal aberrations, two primary issues emerge: privacy concerns and a limited amount of available data. [Yan et al. \(2019\)](#) employed ResNet to detect translocations between chromosomes 9 and 22 using only 200 individual karyotypes. [Li et al. \(2020\)](#) used generative adversarial network to detect anomalies in chromosome images using 320 images per class.

In this study, we attempted to automate the steps involved in detecting abnormal chromosomes in karyograms. Our approach involves feeding individual chromosomes into the model and identifying abnormal chromosomes. The primary contributions of this study are as follows:

1. We designed a hybrid deep learning model to identify abnormal chromosomes for genetic disorder identification.
2. We utilized unsupervised and supervised machine learning techniques to obtain the best results for classification.
3. We used a structural similarity index measure to distinguish the different parts of the anomalous chromosome from the normal one.



4. We performed template matching to identify the translocated part of the abnormal chromosome.
5. We aimed to identify the most common structural abnormalities in neurological disorders by comparing the abnormal and normal chromosomes.

The remainder of this study is organized as follows:

Section II elaborates the proposed model for aberration detection for individual chromosomes. Section III describes the experiments and evaluation of model performance. In Section IV, we discuss the proposed method and its results. Finally, Section V concludes the study.

2 Materials and methods

2.1 Proposed approach

Our approach is Hybrid, combining both supervised and unsupervised methods. In this way, we are taking advantage of the small amount of labeled data available for anomaly detection. **Supervised learning** is a branch of machine learning, in which a model is trained using a labeled dataset. **Unsupervised learning** is a category of machine learning, in which an algorithm provides input data without specific instructions for processing it. This helps the model capture the underlying structure and variations in data.

The proposed system comprises of three major stages, as shown in Figure 5. The first stage involves training the autoencoder with unlabeled data. This is validated with both normal and abnormal data. The input to this stage is the individual chromosome extracted from the karyograms without labels. Chromosomes in the karyograms were arranged in classes. Therefore, we used karyogram singlets to determine whether the results were normal or abnormal. In the second stage, the encoder was utilized as a feature extractor. The extracted compressed features were fed into the CNN classifier as the input. Next, the CNN classifier is trained on the extracted features and labeled data. Finally, the encoder and classifier are trained using labeled data to fully leverage the encoder's ability to generalize from unlabeled data, enhancing its performance in classifying chromosomes.

2.2 Dataset

Images of chromosomes were used as a dataset that was manually annotated and verified by expert cytogeneticists. The dataset was divided into karyograms from which the individual chromosomes were extracted. In this study, we used images of singleton chromosomes for classification. Each chromosome was thoroughly inspected and annotated, and the final dataset of the individual chromosomes was verified by experts. The dataset comprises 234,259 individual chromosomes, of which 216,433 were normal chromosome images and 17,828 were abnormal chromosome images. This dataset included 7,412 chromosome

TABLE 1 Summary of the dataset used for “training,” “validation,” and “testing,” with “normal” and “abnormal” chromosome breakdown.

Dataset type	Number of chromosome images	Normal	Abnormal
Training images (encoder)	140,000	140,000	—
Training images (encoder + classifier)	65,000	50,000	15,000
Validation images (encoder)	12,112	10,912	1,200
Validation images (encoder + classifier)	12,100	10,900	1,200
Test images	5,047	4,621	428

images with translocation abnormalities and 10,416 chromosome images with deletion abnormalities. This ensures a comprehensive representation of the two anomalous categories. A total of 140,000 unlabeled normal chromosome images from all 24 classes were used to train the encoder, and 12,112 images including normal and abnormal chromosome images were used for validation purposes. The encoder and classifier were trained using 65,000 labeled chromosome images, of which 50,000 were normal chromosome images and 15,000 were abnormal chromosome images. To validate the encoder and classifier, we used 12,100 labeled chromosome images, including 1200 abnormal chromosome images and 10,900 normal chromosome images. A total of 5,047 chromosome were tested, including 426 abnormal chromosome images. [Table 1](#) summarizes the distribution of the dataset.

Deletion, addition, and translocation are the primary chromosomal anomalies. If the quality of an image is not good, then it is not easy to detect anomalies accurately, and banding patterns are the core to identify structural abnormalities; if the banding pattern is unclear, it is difficult to identify anomalies in the chromosome. Another problem that hurdles chromosomal anomalies is whether the chromosome is straight or curved. To avoid this, we selected straight and good-quality chromosome images for our approach.

2.3 Proposed method

We employed both supervised and unsupervised learning methods to develop a model for detecting chromosomal anomalies. The key steps of our approach are as follows:

2.3.1 Unsupervised training using autoencoder

It involves autoencoder training with normal data to capture normal chromosome features.

An **autoencoder (AE)** represents an unsupervised machine learning approach utilizes a multilayered feed-forward neural network ([Albahar and Binsawad, 2020](#)). Information is input into the input layer and then passed through a series of hidden layers, making AE a straightforward feed-forward network. Each layer contains a variable number of nodes or neurons responsible for

processing the input and generating the output. These nodes are distributed across different layers, each connected to all nodes in the preceding layers. The input and output layers both possess an identical number of nodes, denoted as “n,” because of the symmetric structure of the autoencoder, which aims to reconstruct the input on the output side. The predictions generated at each node, facilitated by the activation functions, are transmitted to consecutive layers. An autoencoder comprises two primary stages: Encoder and Decoder ([Tan et al., 2019](#)). We utilized this part because the autoencoder is trained solely on standard chromosome images without labels. This phase aims to help the encoder learn the typical patterns and structures found in the standard chromosome images. As the encoder model only sees normal data, it specializes in understanding and encoding these standard patterns into a compressed, lower-dimensional latent space representation. The decoder part attempts to reconstruct the input image from the latent-space representation, allowing the AE model to learn a good feature for the extraction process. For generalization, we validated it using abnormal and normal unlabeled chromosome data.

2.3.2 Feature extraction from trained encoder

Once the AE is trained, the encoder extracts features from normal and abnormal chromosome images. The encoder provided feature representations for each image fed into the classifier. The features extracted from the encoder contain latent representations of the input chromosome images. These features are compressed and abstract forms of the original images, capturing the essential characteristics of the chromosomes while discarding less critical details. These features contain information, such as chromosome patterns, shapes, and structures.

2.3.3 Training the (encoder + CNN classifier) with extracted features (supervised learning)

The features extracted by the AE encoder are then passed to the CNN classifier, which learns to classify images based on the encoder’s output. This step uses the labeled data to train the classifier. The CNN classifier learns to distinguish between normal and abnormal chromosomes based on the features extracted from the encoder and is trained with the standard and abnormal labeled images while keeping the encoder weights fixed (frozen).

2.3.4 Fine-tuning of encoder and classifier

In this case, the encoder’s weights are unfrozen, and the encoder and classifier are fine-tuned using the labeled data. The last two layers of the encoder are fine-tuned. Training only the last two layers is computationally efficient and preserves the robust pretrained knowledge of the encoder’s initial layers. This step is also impactful, because these layers represent higher-level abstract features of the input data. These features are closer to the final compressed representation and contain critical semantic information, making them crucial for adapting the model to new tasks or datasets. Fine-tuning these layers allows the model to adjust the high-level features to the new dataset without drastically altering the generalized low-level feature extraction learned earlier. Focusing on these layers helps us to reduce the risk of overfitting,

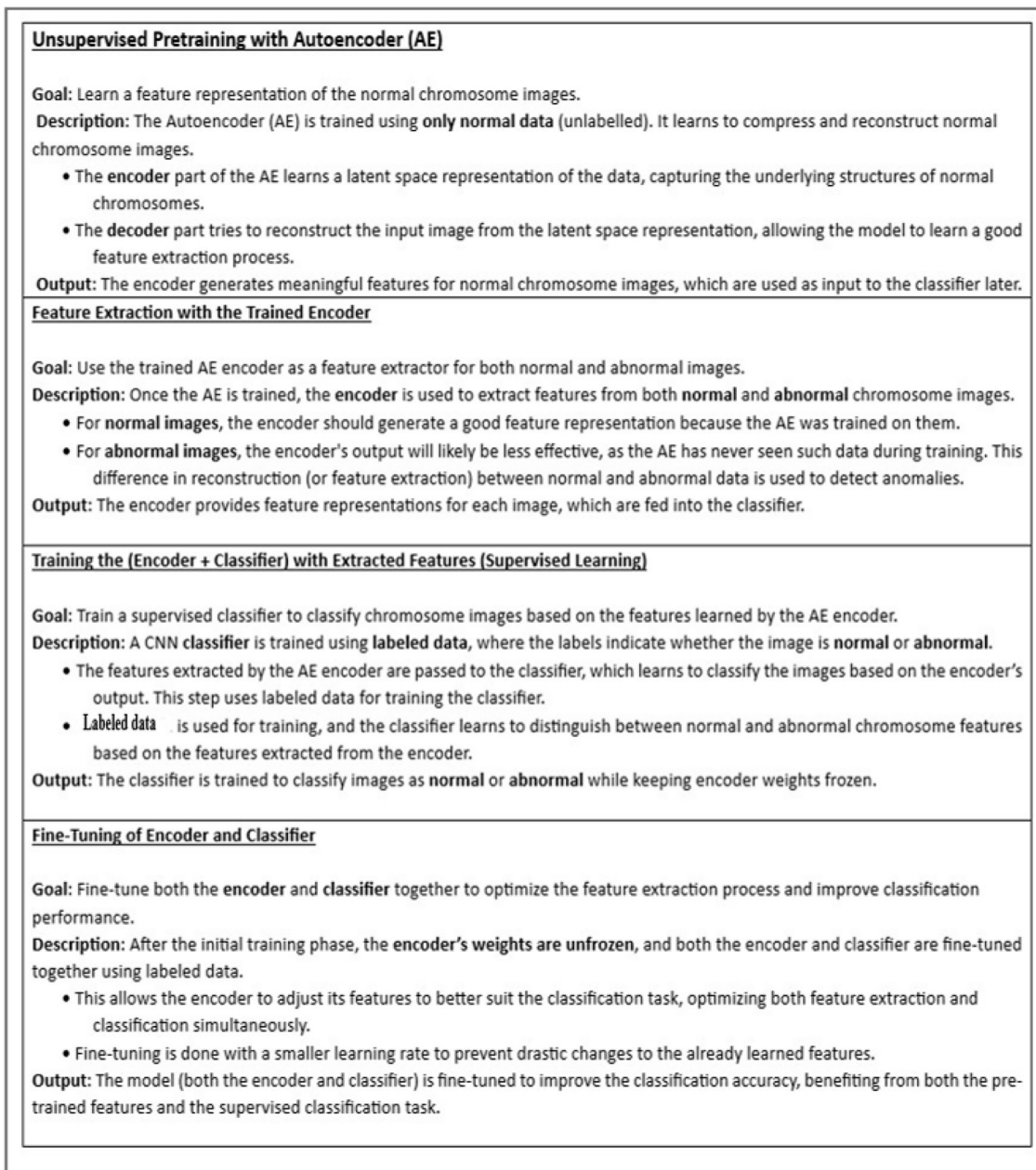


FIGURE 6
Flow of the proposed approach.

as they retain generalized features, which is beneficial as our dataset is small.

This step helps the encoder adjust its features to suit the classification task better. Simultaneously, the classifier learns to effectively map these extracted features to the desired classes (normal and abnormal chromosomes). By jointly optimizing both the encoder and classifier, the model can better capture discriminative features, improving overall classification accuracy. Finally, the model was validated using normal and abnormal chromosome images. The steps of our approach are shown in Figure 6.

The **encoder** plays a crucial role in our hybrid model, serving as the foundation for feature extraction and anomaly detection, enabling our approach to detect chromosomal abnormalities effectively. Its role can be broken down into several key functions:

1: Unsupervised feature extraction: The encoder is initially trained on unlabeled data, which then learns a compressed representation of chromosome images through an unsupervised approach. It then extracts meaningful latent features to capture essential chromosomal characteristics, such as patterns, shape, and structure. These features highlight important chromosome variations and anomalies, which are often difficult to detect using conventional methods.

2: Data compression and dimensionality: The encoder effectively performed dimensionality reduction by converting input chromosome images into a low-dimensional space. When non-essential information was discarded, only significant characteristics were preserved. This abstraction enriches the classifier by directing the implementation of the most essential features of the chromosomes, and enhances the general efficiency of the model.

3: Enhancing supervised learning of the CNN classifier: This extracted features are then given to the CNN classifier, which is trained on labeled data to differentiate normal chromosome patterns and abnormal patterns. The encoder output serves as a rich input representation, enabling the classifier to perform better by learning more discriminative patterns from these high-level informative features.

4: Fine-tuning for task optimization: In last stage, the encoder and CNN classifier are jointly fine-tuned with labeled data, enabling the encoder to refine its feature extraction process to suit the specific requirements of the classification task.

Therefore, this joint fine-tuning guarantees feature learning and classification in the best manner, thereby minimizing the generation of incorrect chromosomal anomaly detection models. It is worth noting that the encoder is a key component of the proposed hybrid model. It encompasses unsupervised anomaly detection to a supervised form of classification, allowing the system to deliver more accurate, scalable, and generalizable solutions to automate karyogram analysis.

2.3.5 Anomaly detection

Once the hybrid model classifies chromosomes as abnormal, structural anomalies can be detected. For this purpose, we used SSIM and pattern matching to identify chromosomal abnormalities. The **SSIM** is a computer vision technique that identifies the differences between two images. It helps to identify the differences between chromosomes in cases of structural abnormalities, such as deletions, additions, and translocations. In the case of deletion or addition, the difference is clear; however, for translocation, we used the template matching technique. We first find the different parts from the normal with the help of SSIM. We also had to identify the translocated portion. For this purpose, we used pattern matching to find the translocated part. Pattern matching is a Computer Vision (CV) technique in which regions are located within an image that corresponds to the template. In this way, we successfully identified an anomalous part in chromosomes. Our main focus was to identify the structural abnormalities involving deletion and translocation in the chromosome structures.

a) Structural similarity index measure

SSIM was used to assess the quality of images by examining the structural details of two images (James et al., 2023).

$$\text{SSIM} = \frac{(2\mu_x\mu_y)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (1)$$

Where in Equation 1:

μ_x, μ_y : Mean intensities of two images.

$\sigma_x^2 + \sigma_y^2$: Variances of intensities.

σ_{xy} : Covariance of intensities.

c_1, c_2 : Constants for avoiding instability when the denominator is close to zero.

Figure 7 shows the implication of SSIM. b) Template matching:

Template matching is a machine-vision technique used to locate regions within an image corresponding to the template. A template is a predefined image or part of the image used to match the part in the main image. This process is performed by moving the template over the image. The similarity between the main image and the template image was calculated. Open CV template matching was then performed. The template image slides over the main image and the patch of the main image is compared with the template image.

3 Experiments and results

This section outlines the experimental setup, performance metrics, and the results of the proposed model.

3.1 Experimental setup

In this study, a CNN autoencoder and a CNN classifier were combined as models for classification tasks. Both models were trained using Python software. The Spyder platform (v. 5.4.3) was used for the training, validation, and testing of the model. The Spyder platform was implemented using PyTorch framework (v. 11.8 with torch version 2.3.0), and the experiments were conducted on UBUNTU 18.04, deployed with an NVIDIA RTX 1080 Ti.

3.2 Parameter setting and preprocessing

3.2.1 Preprocessing of data

The images were preprocessed before being provided to the model as an input. Some of the images were large and some were small. The large images were compressed, and the small images were padded to obtain 32×32 dimensional images. This step was performed to maintain the uniformity of the images. We also normalized the images by scaling the pixel values to between 0 and 1.

3.2.2 Parameters setting

• Encoder

The encoder in our model consisted of three convolution layers with the following filter configurations: 16, 32, and 64. Each layer employs the Leaky ReLU activation function to enhance the learning of non-linear relationships and prevent vanishing gradient issues. The architecture progressively extracts hierarchical features from chromosome images, thereby capturing low- and high-level chromosomal patterns. We selected a batch size of 20 for training and 10 for validation. The optimizer was Adam, who had a learning rate of 0.001 and was trained for 50 epochs. We trained the

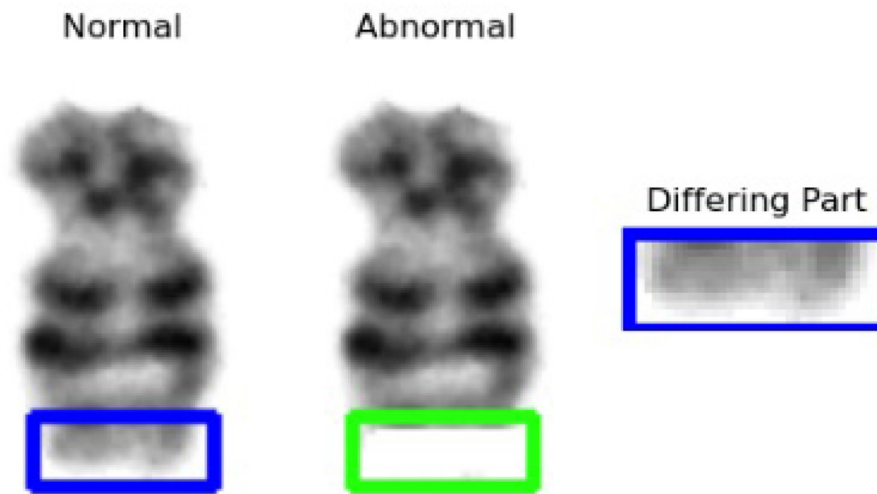


FIGURE 7
Structural similarity index measure result.

model for 100 epochs earlier; therefore, the model could learn sufficient information in 50 epochs, so we stopped it at 50 epochs. The training and validation loss plots are shown in Figure 8A.

The loss function was MSE (Mean Square Error), as expressed in Equation 2.

Mean squared error:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

Where

n : number of data points.

y_i : the actual value for the i^{th} data point.

$\hat{y}_i$: predicted value of the i^{th} data point.

This provides the mean of the squared discrepancies between the actual and predicted values, offering a metric for the overall accuracy of the prediction. For Normal data samples, the reconstruction error is typically low, whereas for anomalous data, the values tend to be higher and exceed a specific threshold.

- Decoder

The decoder consists of three deconvolutional networks (deConvNets) with filter values (64,32,16) with ReLu.

- CNN classifier

We selected a batch size of 20 for training and 10 for validation. Adam was used as the optimizer, with a learning rate of 0.0001, and was trained for 20 epochs. The loss function was CrossEntropyLoss. Figure 8B shows the training and validation losses.

- Encoder + classifier

We selected a batch size of 20 for training and 10 for validation. The optimizer was Adam with the learning rate 0.0001 and was trained for 20 epochs. The loss function was weighted CrossEntropyLoss to effectively address the class

imbalance, and the training and validation plots are shown in Figure 8C.

3.3 Model training

We implemented several strategies throughout the training pipeline to ensure the model's robustness and to mitigate overfitting and bias. The performance was continuously monitored on a separate validation set, and early stopping was applied based on validation loss trends to prevent overfitting, as well as regularization techniques such as the dropout layer. Data scaling was performed as an added data preprocessing technique, as it helped scale the input feature pixel values and achieve a stable convergence rate. To improve generalization, features were learned by passing both labeled and unlabeled data to the autoencoder before proceeding to the supervised classification component. Although the transformations used during data preprocessing did not include aggressive augmentation strategies such as flipping or cropping, we resized the chromosome images to a standard size of 32×32 . Normalization was also applied to standardize the dataset's intensity range, ensuring sample consistency and minimizing noise.

Because the dataset was imbalanced, where abnormal chromosome samples were significantly fewer than the standard, steps were taken to prevent biased learning. Although the autoencoder was initially trained solely on standard samples to extract robust latent representations, the subsequent classifier was trained on normal and abnormal samples. For the evaluation, the test set comprised normal and abnormal chromosomes for a fair comparison of the model. Specifically, evaluation measures such as precision, recall, and F1-score for each class label were presented to measure the model's ability to identify deviations. Combined with this detailed evaluation and validation-based approach to monitoring during the training process, overfitting and accurate outcomes were significantly reduced.

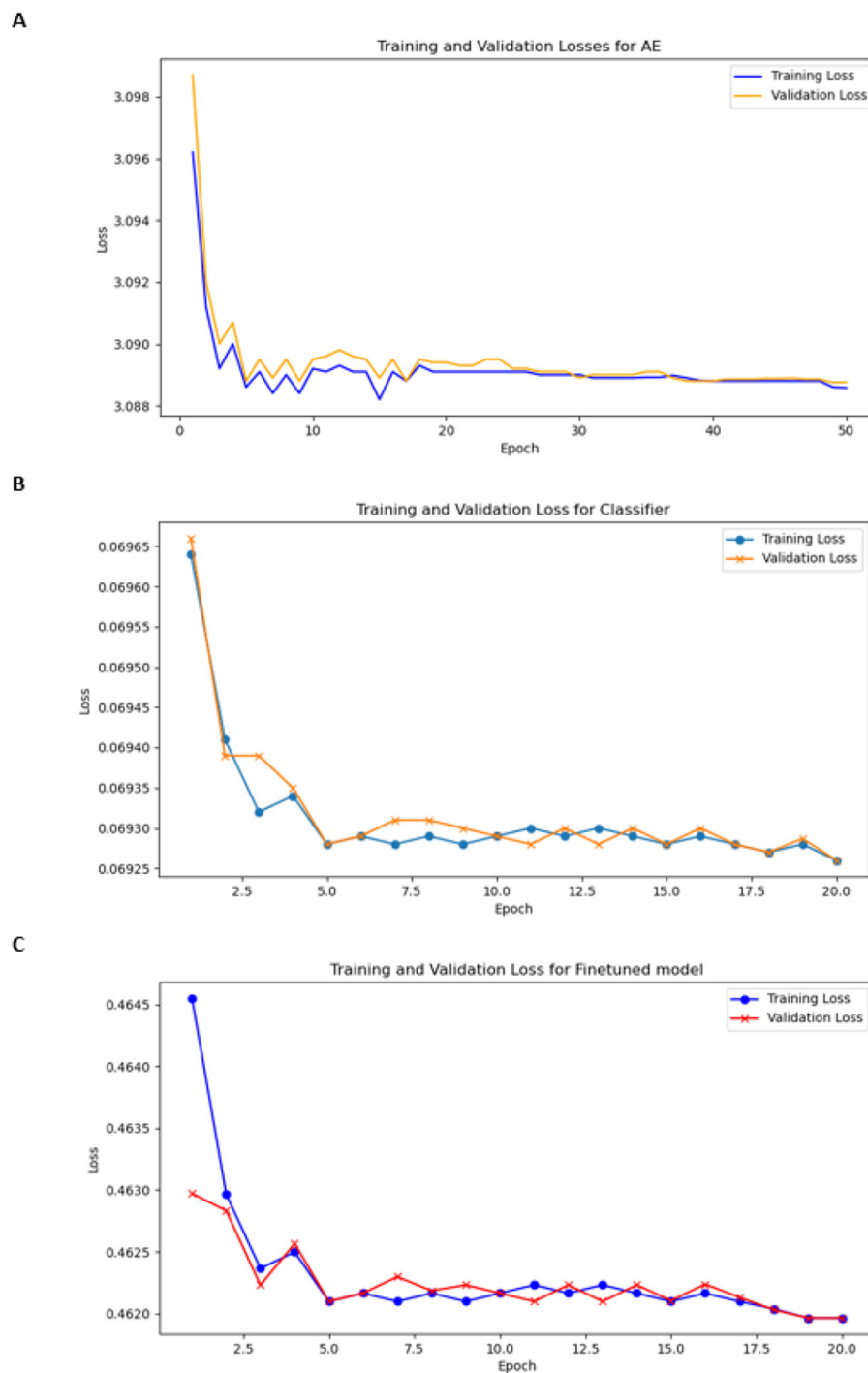


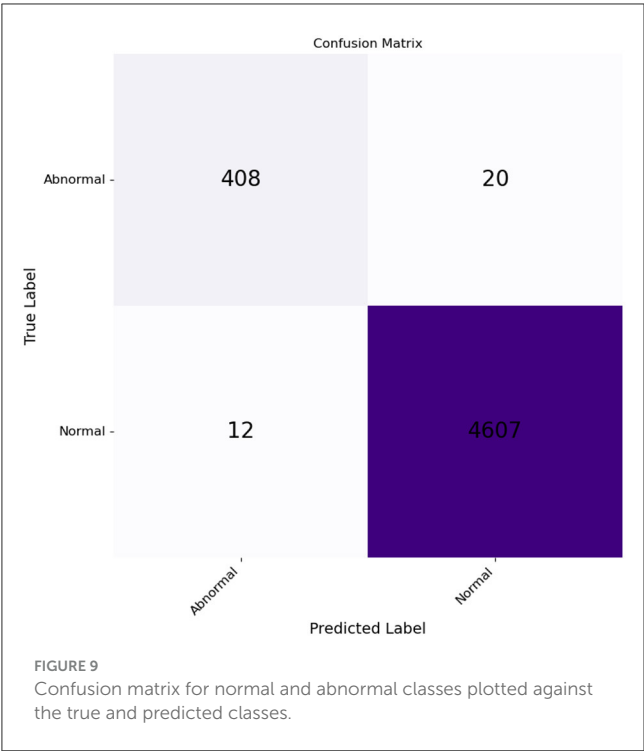
FIGURE 8
Training and validation losses (A) for encoder, (B) for classifier, and (C) for finetuned model.

3.4 Performance metrics

Four performance metrics were used for the evaluation. Accuracy was determined by dividing the number of correctly predicted cases by the total number of cases. A high accuracy value indicated that the model is made accurate predictions. Specifically, accuracy is calculated as the sum of true positives (TP) and true

negatives (TN) divided by the total sum of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), as shown in Equation 3.

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



Precision: Equation 4 measures the number of correct results out of all predicted positive results. It is calculated by dividing the number of true positives (TP) by the sum of true positives (TP) and false positives (FP).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

Recall: This is also known as sensitivity or the true positive rate, which is the ratio of correctly predicted positive results to the total positive cases. It is calculated by dividing the number of true positives by the sum of the true positives and false negatives, as given in Equation 5.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5}$$

F1 Score: The F1 score is the harmonic mean of precision and recall, providing a single metric that balances both. The Equation 6 helps calculate F1 score.

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

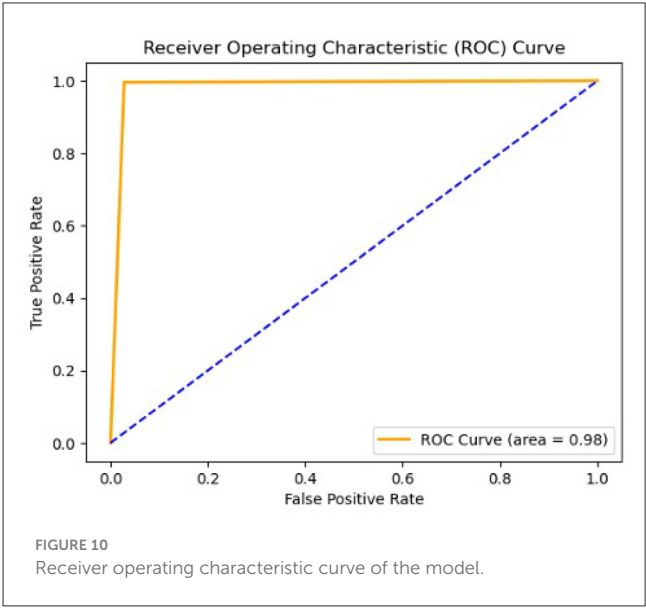
3.5 Results

The confusion matrix in Figure 9 shows the results. From the 428 input images of chromosomes, 408 were correctly classified as abnormal, and 20 were classified as normal. Of the 4,619 images of chromosomes, 4,607 were classified as normal and 12 were classified as abnormal, but were identified incorrectly as abnormal.

The evaluation metrics accuracy, precision, recall, and F1 score are summarized in Table 2 for our model.

TABLE 2 Model performance metrics.

Metrics	Normal class %	Abnormal class %
Accuracy	99.37	99.37
Precision	99.57	95.32
Recall	99.74	97.14
F1 score	99.65	96.22



A Receiver Operating Characteristic (ROC) curve was also generated to evaluate the performance of our model in predicting the probabilities of outcomes, distinguishing between normal and anomalous chromosome images, as shown in Figure 10. This curve was plotted against the true positive rate (TPR) and false positive rate (FPR). The area under the curve (AUC) was used to assess the level of discrimination between classes. Figure 10, with the value of AUC = 0.97, shows that our model is effectively distinguished between normal and abnormal chromosomes.

In the dataset, only the translocations between chromosomes 9 and 22 were identified. Therefore, a pattern-matching technique was applied to detect abnormalities. As shown in Figure 11A, two abnormalities were observed in the karyograms: one on chromosome 9b and the other on chromosome 22b. Both 9b and 22b were identified as translocated chromosomes. In the first step, the two chromosomes were found to be abnormal. Subsequently, the type of abnormality was identified by comparing chromosomes 9 and 22 with their corresponding normal reference chromosomes. Differences between 9b and 22b were also observed. Different parts of chromosome 9 were identified using SSIM, as shown in Figure 11B. The same process was performed on chromosome 22, and different parts are shown in Figure 11C. In the final step, the template-matching technique was applied to locate the translocated parts. Figure 11D shows that part of chromosome 22 was located on chromosome 9, while Figure 11E shows the translocated part of chromosome 9 on chromosome 22. This

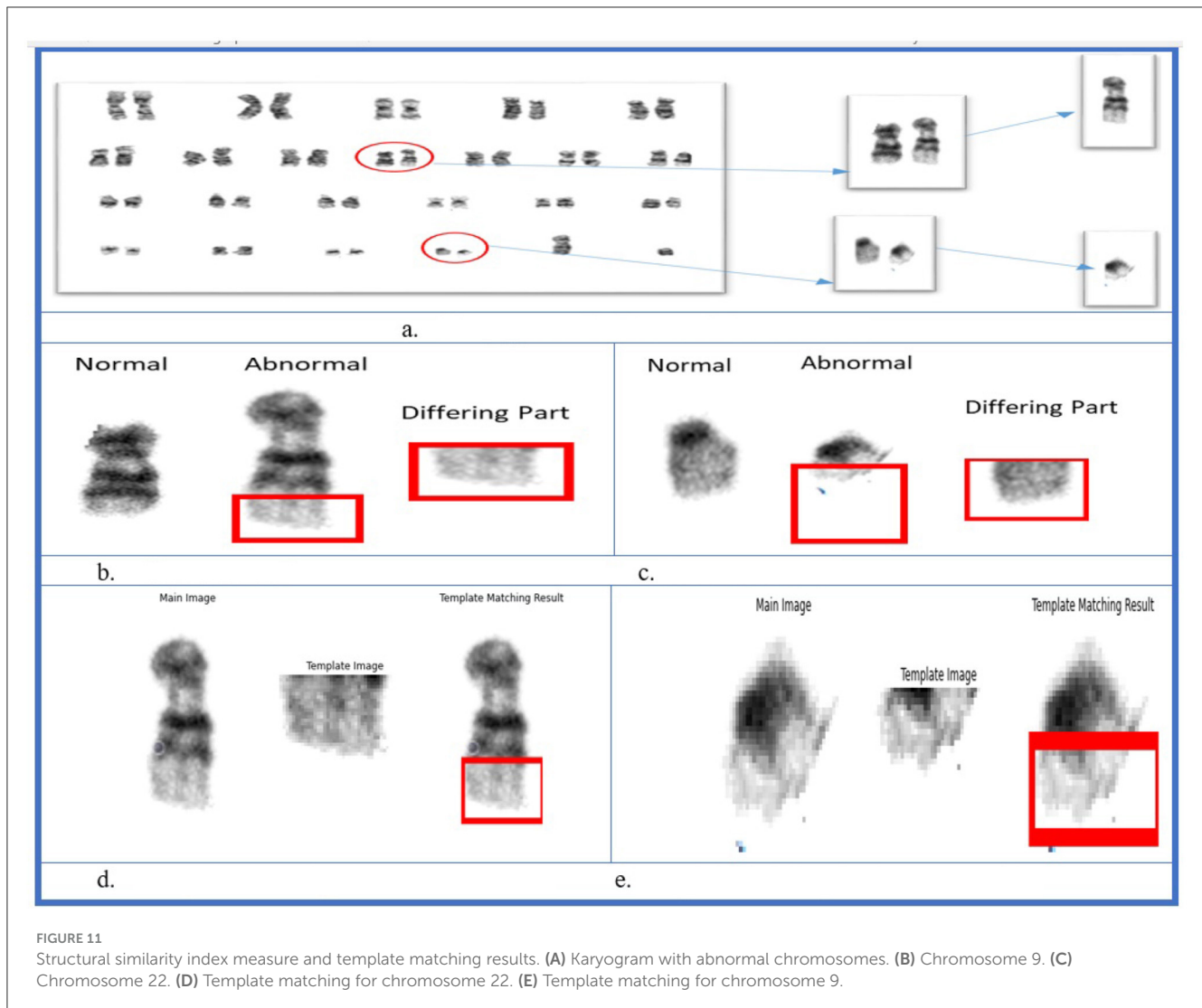


FIGURE 11
Structural similarity index measure and template matching results. (A) Karyogram with abnormal chromosomes. (B) Chromosome 9. (C) Chromosome 22. (D) Template matching for chromosome 22. (E) Template matching for chromosome 9.

approach enabled the identification of deleted or translocated parts of abnormal chromosomes.

4 Discussion

Cytogenetics is a branch of genetics that attempts to explain the relationship between human chromosomes and their genetic makeup and functions. Furthermore, it examines into the health and evolutionary implications arising from the architectural distortions of the chromosome. Cancer and other related abnormalities related to genetic diseases or neurological disorders are diagnosed after samples have been analyzed in laboratories. These methods are employed to search for and evaluate their effects, particularly on neurological disorders, in the health and developmental aspects of humans. This basic method of karyotyping is complex and requires a considerable amount of knowledge in the domain and time. Automated karyotyping enhances the speed and efficiency of chromosomal analysis, allowing for quicker identification of abnormalities. It reduces human involvement, addresses the challenges of manual analysis,

and reduces the scarcity of large datasets. The major limitation observed is the absence of datasets because deep learning methods are data-intensive, and data related to abnormalities are much more complex and not easy to understand by every one.

Chromosomal analysis, when performed during fetal development, offers the unique advantage of detecting genetic abnormalities before the onset of clinical symptoms. This is crucial for disorders such as Down syndrome, autism, intellectual disabilities, edwards syndrome, cri-du-chat syndrome, mosaic Turner syndrome, and other underdevelopment disorders that have a strong genetic component. The earlier a disorder is detected, the earlier medical interventions, lifestyle adjustments, and support mechanisms can be implemented. Moreover, prenatal testing can allow families to prepare mentally and emotionally, while also making informed decisions about pregnancy, care, and future management of the child's health.

Disorders such as Down syndrome and other underdevelopment disorders are primarily identified and studied using techniques such as EEG, MRI, and other imaging technologies. However, these methods are only applied when a child or person shows signs of neurological disorder. For example,

they can be identified when a child is already experiencing developmental delays or cognitive impairment. These technologies help monitor the brain's electrical activity and neural function. However, these studies do not offer predictive insights into the genetic basis of these conditions, particularly during the early stages of development. Visualizing chromosomes at an early stage allows the early detection of chromosomal abnormalities during fetal development. Anomalies such as deletions, duplications, or translocations that cause neurological disorders can be identified by analyzing fetal cell chromosomes.

With the advancement of deep learning models, including unsupervised and supervised approaches, it is now possible to automate and scale the analysis of chromosomal images of fetal or later blood or bone marrow samples. This automated analytical approach is more accurate and efficient. We introduce a hybrid model approach that utilizes unsupervised learning and supervised techniques. This hybrid model can efficiently process genetic data to quickly identify anomalies and provide more precise diagnoses. This facilitates the identification of structural abnormalities that are often associated with neurological disorders.

Our objective was to achieve the automatic detection of any structural chromosomal abnormality without the necessity for training for each distinct abnormality with labels. Our approach is beneficial because labeled examples are scarce, especially for rare anomalies. Prior CV and ML studies have addressed various challenges related to chromosomes (Boddupally and Thuraka, 2023), including segmentation, and Saleh et al. (2019) proposed Unet for chromosome segmentation. Fan et al. (2024) proposed DaCSeg for segmentation of chromosomes. Kang et al. (2024) proposed the model UC-Det model for counting chromosomes. Classifications: Qin et al. (2019) designed Verifocal-net for chromosome classification. Chang et al. (2024) proposed a DL model that uses attention to classify chromosomes. Wu et al. (2018) used GANs for the augmentation of chromosomes. Uzolas et al. (2022) used GANs for chromosome generation. Al-Kharraz et al. (2020) used YOLOV2 and VGG19 to identify the numerical aberrations. Wang et al. (2010) detected translocation in chromosomes using an adaptive matching technique. Kao et al. (2023) proposed 3 step process for identifying individual and clustered chromosomes. Cox et al. (2022) provided a supervised technique to identify abnormal chromosomes using Residual CNN. Bechar et al. (2023) used a supervised Siamese Network to classify chromosomes. Among the various studies mentioned previously, the prevalent approach involves the application of traditional supervised learning methods on relatively small datasets.

4.1 Significance of proposed approach

The proposed model integrates supervised and unsupervised learning techniques, leveraging the strengths of both approaches to improve the performance and robustness of automated chromosome classification.

4.1.1 Supervised learning

Supervised learning uses labeled data to train models with the objective to learn a mapping between input features and their

corresponding output labels. There are some **advantages** like: with a sufficient amount of labeled data this approach excels at learning discriminative patterns and distinguishing between normal and abnormal chromosomes with high accuracy. Supervised learning excels in tasks such as classification with high accuracy, particularly when labeled data are abundant. However, this is limited by the challenge of acquiring large labeled datasets in clinical settings. Supervised models also have some **disadvantages**, such as their dependence on large amounts of labeled data. Obtaining a large amount of labeled data requires significant time and expertise, which is a limitation in the clinical environment. and a model trained solely on limited labeled data reduces the generalization ability for unseen abnormal cases.

4.1.2 Unsupervised learning

Unsupervised learning aims to identify structures inherent in data without using labeled learning information. The merits of unsupervised learning include that it works with large amounts of data that are not labeled and is easier to access than labeled data. It excels at discovering hidden patterns and relationships that can work well for feature extraction and feature space dimensionality reduction thereby enhancing the computational performance and generalization across diverse data. However, this method has some limitations. It has no direct relation to the target outputs, which makes it unsuitable for tasks involving exact quantitative predictions without further processing. However, the extracted features are more complex to analyze, and comparing their performance without a labeled dataset is challenging.

Unsupervised learning extracts meaningful features without relying on the labeled data. In our approach, an autoencoder is used for feature extraction, providing compressed representations of chromosome images. Unsupervised learning also has some **advantages** over supervised learning, such as the unsupervised approach enables to utilize a large number of unlabeled chromosome images that are more readily available and cost-effective to acquire. **Robust feature extraction:** The encoder captures essential structural and morphological information about chromosomes, making it possible to detect subtle patterns that are difficult to capture using supervised methods alone. **Better generalization:** Because the encoder was trained and validated on a large dataset, it can generalize better across different variations and imaging conditions. Like supervised models, they also have some **Disadvantages**: as: **Indirect labels:** While unsupervised models are good at feature extraction, they do not directly map to class labels and require subsequent integration with a supervised classifier. **Interpretability challenges:** Understanding the exact features extracted by the encoder can sometimes be less interpretable than supervised models, making it harder to explain specific clinical findings.

4.1.3 Hybrid approach

By combining supervised and unsupervised techniques, our model leveraged the strengths of both approaches. Supervised learning excels in tasks that require labeled data, particularly in distinguishing between normal and abnormal chromosomes. However, acquiring large labeled datasets, particularly for rare anomalies, can be challenging. In contrast, unsupervised learning

TABLE 3 Comparison of our hybrid approach with existing approaches for karyogram analysis.

Aspect	Existing models	Our approach (hybrid)
Model architecture	CNN, Fully Connected Networks	AutoEncoder (Unsupervised) + CNN Classifier (Supervised)
Approach type	Mostly supervised	Hybrid (Supervised + Unsupervised)
Dataset diversity	Often limited to normal or simple anomalies	Comprehensive dataset with deletion and translocation structural abnormalities
Model generalizability	Poor generalizability on rare anomalies	Better generalizability as trained on unlabeled data

can handle large amounts of unlabeled data and is effective for feature extraction and pattern discovery. However, it lacks direct connections to target outputs, making it less suitable for classification tasks.

To address these limitations, our hybrid approach integrates the advantages of both methods. The unsupervised encoder extracts meaningful features, whereas the supervised classifier refines these features for the accurate classification of normal and abnormal chromosomes. This combination allowed us to harness the power of unsupervised learning for handling large unlabeled datasets and the precision of supervised learning for effective classification.

Table 3 compares our method with other methods, emphasizing the differences in the learning patterns. Our hybrid approach uses an autoencoder (AE) trained on unlabeled data for feature extraction, followed by a supervised classifier for the final classification task. Because normal data are often more abundant and easier to obtain than abnormal data, an autoencoder uses normal data to extract features. This eliminates the need for labels thereby allowing the autoencoder to autonomously identify valuable features from the dataset. The classifier then focuses on the most relevant features provided by the autoencoder thereby enhancing the classification performance. In addition, as the autoencoder is trained on unlabeled data, its reliance on labeled samples decreases, which is particularly beneficial when labeled anomalous data are scarce or costly.

We trained, validated, and tested our model using a large data set that is not publicly available. The dataset contains not only normal chromosomes, but also abnormal chromosomes. After intensive training and validation, we tested our model on test data comprised of 5,047 images, including 428 abnormal images. Our model achieved an AUC value of 0.98, demonstrating its ability to distinguish between normal and abnormal chromosomes effectively. Our model outperforms identifying abnormal chromosomes from normal chromosomes using hybrid unsupervised and supervised deep learning. Compared to existing methods, as shown in Table 4, our hybrid approach achieved an accuracy of 99.3%, surpassing the DeepResidual model by Yan et al. (2019), which reached 97.5%, and the DNN model by Kang et al. (2024), which achieved 99.2% accuracy.

Our approach comprises two distinct steps: first, detection of anomalous chromosomes, and second, identification of specific

TABLE 4 Comparison with previous models.

References	Model	Approach	Accuracy
Yan et al. (2019)	DeepResidual	Supervised	97.5%.
Kang et al. (2024)	DNN	Supervised	99.2 %
Our approach	AutoEncoder + CNN classifier	Hybrid	99.3 %

abnormalities within these chromosomes. The initial step was executed by employing normal images. We validated our AE model using a dataset containing both abnormal and normal samples. This demonstrates how our model is better than the others in detecting aberrant chromosomes; hence, we demonstrate our efficiency and precision in the hybrid mode.

After determining whether the chromosome is normal or abnormal, the following step seeks to determine a particular abnormality. Several methods in computer vision can detect abnormalities in chromosomes. Our approach involves aligning a normal chromosome with a counterpart chromosome to determine the area of the anomaly. Chromosomes are usually compared with normal chromosomes or ideograms to check for subtractive or translocation presence. To perform this task, we used the SSIM and pattern-matching methods. We compared the normal chromosomes instead of ideograms.

SSIM helps to identify the differences between the two images. We compared normal and abnormal images and identified different parts in cases of deletions and translocations. However, we first identified the difference between the normal tissue and different parts of the translocation. We also had to identify the translocated part and used pattern matching to find the translocated part. In this way, we successfully identified aberrations in the chromosomes. Our primary focus was identifying structural abnormalities involving deletion and translocation in chromosome structures.

We presented an approach for identifying structural aberrations in individual chromosomes extracted from karyograms. The methodology relies on analyzing banding patterns to detect and characterize these abnormalities. Substantial effort has been made to explore the integration of machine learning into pathology diagnoses. We presented a hybrid approach comprising both unsupervised and supervised learning that proved advantageous, particularly when dealing with a limited number of anomalous images. Gathering anomalous datasets in the medical field is inherently challenging. Our model was uniquely trained, validated, and tested on a large dataset, one of the first of its kind for this task, thereby significantly enhancing the robustness of anomaly detection and demonstrating its effectiveness in identifying chromosomal abnormalities.

In real-world scenarios, time constraints often lead to the standard practice of analyzing only a few meta-phase cells per specimen despite the availability of hundreds of cells. Despite this restricted analysis, challenges persist regarding the cost and turnaround time for diagnosis. This task is perfectly tailored for deep learning because of the complexity of expert analysis, which implies the use of visualization and the expected common mean of a sample set with its genus of origin. In addition, when applied to the initial assessment of chromosomal abnormalities for conditions

such as epilepsy and Down syndrome, we expect that our model will provide prognostic advice for more effective patient management. The prediction of these disorders through the identification of genetic markers contributes to early intervention, which will help reduce the impact of the disorders on development as a result of early diagnosis and management. Recognizing neurological disorders at a preliminary stage significantly boosted genetic anomaly detection and preventive diagnostics in our model.

Here, we present a new methodology for a hybrid model to resolve the issue of automated chromosome anomaly detection, which is an important paradigm of cytogenetic analysis. Our study innovates by combining supervised and unsupervised learning frameworks, which enhances the detection accuracy and offers significant improvements over other methods. Our approach maximizes the value of the available data by utilizing unlabeled chromosome images during the feature extraction phase while still using labeled data for supervised classification. This strategy overcomes the limitations of imbalanced datasets, where obtaining many labeled abnormal chromosome images is difficult in clinical and research settings.

The method we propose for karyogram analysis is expected to greatly enhance the diagnostic process, allowing for the faster identification of potential genetic issues. Implementation of our model in clinical decision support systems can help cytogeneticists and practitioners obtain automatic and confident classification results, thus increasing diagnostic accuracy and reducing time.

4.2 Challenges in real-world adoption and limitations

In clinical contexts, it is essential to protect patient data. Simultaneously, datasets are usually associated with restrictions regarding the availability of information, which can be a problem for training and validation. In addition, the images of chromosomes in the model may not be consistent with those of other laboratories and imaging equipment, which might cause a difference. To handle such variations, robust domain adaptation techniques are necessary. Moreover, integrating our model seamlessly into existing laboratory software and clinical workflows requires technical compatibility and collaboration with various healthcare information technology (IT) systems. However, the proposed system has certain limitations. It is mainly used to work with straight chromosomes, but it is useful with curved chromosomes that are first straightened. However, refining the straightening process may enhance the outcomes. Although the model has shown promising results on the custom dataset used, it lacks validation on external datasets that contain similar complex structural chromosomal abnormalities, which could be an area for improvement. Despite its high classification accuracy, the model has significant computational demands, particularly during training. The concept of the model is defined by multiple convolutional layers that contribute to numerous parameters and significant GPU memory and processing power. This may restrict its application in the real world, especially in areas where resources are scarce, such as small clinical environments. However, a trade-off

between computational inputs and model efficacy exists in the process. The inference times for such applications depend on the specific hardware, and in large-scale clinical trials, selecting the best hardware resources and their integration solutions is crucial. In our opinion, we can eliminate all of these problems with the help of further development of interdisciplinary cooperation, additional model refinements, and numerous clinical trials that will allow us to implement the proposed method in various clinics successfully.

5 Conclusion

Our study strongly emphasizes that reliable detection of anomalous data is important in medical applications, primarily in genetic diagnosis by karyotyping. Identification of anomalies in medical data is considered a task in computing science and is important for patient care and treatment. Therefore, developing robust methodologies such as the automated approach presented here is vital for ensuring the accuracy of diagnostic procedures. Our hybrid model, which combines an unsupervised encoder trained with unlabeled normal data and a supervised CNN classifier trained on labeled normal and abnormal chromosome data, is a powerful approach to karyogram analysis. Thus, by training the encoder with data that are not labeled as normal or abnormal and validating the model with normal and abnormal data, we ensure that we obtain the best of both worlds from a model where all the relevant features are captured. The encoder also learns the basic features that help enhance the task of chromosome classification; separating normal and abnormal chromosomes is performed accurately.

Our model was trained and validated using a large dataset, and eliminated false or misleading anomalies. Furthermore, we identified the anomalous chromosomes in detail using CV methods, SSIM, and template matching. Thus, the combined use of appropriate methodologies strengthened our approach and increased the accuracy of the results. After evaluating the test data, we found that precision, recall, F1 score, and accuracy were all impressive, with a total accuracy of 99.37% for both normal and abnormal classes, and F1 scores for both normal and abnormal classes were 99.65% and 96.22%, respectively. These results demonstrate that the model effectively classifies normal and abnormal chromosomes. In addition, the model achieved an AUC of 0.98, demonstrating its effectiveness in classifying normal and abnormal chromosomes.

Our study addresses the demand for automation in genetic disorder assessment and underscores the transformative potential of interdisciplinary approaches in healthcare and neurological computations. Future study will involve working with research laboratories and hospitals to obtain data on various imaging sources, lighting conditions, and types of chromosomal abnormalities. Moreover, while the system is currently designed to detect only structural anomalies, future study plans will incorporate numerical anomaly detection. We also plan to integrate explainable AI (XAI) to visually discuss the prediction results so that cytogeneticists and doctors can use this information efficiently for further case analysis.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: the dataset used in this study is proprietary and belongs to NUST. Due to ongoing research and confidentiality agreements, the data is not publicly available. However, interested researchers may contact the author for potential collaboration or for inquiries regarding data access within the constraints of institutional guidelines. Requests to access these datasets should be directed to tabassum.sumairaawan@gmail.com.

Ethics statement

The studies involving humans were approved by the Ethical Review Committee (ERC)/IRB National University of Sciences and Technology Islamabad. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

ST: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing, Funding acquisition. MK: Project administration, Supervision, Writing – review & editing, Resources. JI: Resources, Supervision, Writing – review & editing. AW: Writing – review & editing, Investigation. MI: Writing – review & editing.

References

- Ahmad, J., Akram, S., Jaffar, A., Ali, Z., Bhatti, S. M., Ahmad, A., et al. (2024). Deep learning empowered breast cancer diagnosis: advancements in detection and classification. *PLoS ONE* 19:e0304757. doi: 10.1371/journal.pone.0304757
- Albahar, M., and Binsawad, M. (2020). Deep autoencoders and feedforward networks based on a new regularization for anomaly detection. *Secur. Commun. Netw.* 2020:7086367. doi: 10.1155/2020/7086367
- Al-Kharraz, M., Elrefaie, L., and Fadel, M. (2020). Automated system for chromosome karyotyping to recognize the most common numerical abnormalities using deep learning. *IEEE Access* 8, 157727–157747. doi: 10.1109/ACCESS.2020.3019937
- Bechar, M., Guyader, J., El Bouz, M., Douet-Guilbert, N., Al Falou, A., and Troadec, M. (2023). Highly performing automatic detection of structural chromosomal abnormalities using siamese architecture. *J. Mol. Biol.* 435:168045. doi: 10.1016/j.jmb.2023.168045
- Boddupally, K., and Thuraka, E. (2023). Artificial intelligence for prenatal chromosome analysis. *Clin. Chim. Acta* 552:117669. doi: 10.1016/j.cca.2023.117669
- Britto, A., and Ravindran, G. (2007). A review of cytogenetics and its automation. *J. Med. Sci.* 7, 1–18. doi: 10.3923/jms.2007.1.18
- Chang, L., Wu, K., Gu, C., and Chen, C. (2024). A novel end-to-end chromosome classification approach using deep neural network with triple attention mechanism. *Biomed. Signal Process. Control* 91:105930. doi: 10.1016/j.bspc.2023.105930
- Cox, A., Park, C., Koduru, P., Wilson, K., Weinberg, O., Chen, W., et al. (2022). Automated classification of cytogenetic abnormalities in hematolymphoid neoplasms. *Bioinformatics* 38, 1420–1426. doi: 10.1093/bioinformatics/btab822
- Duarte, K., Sidhu, A., Barros, M., Gobbi, D., McCreary, C., Saad, F., et al. (2024). Multi-stage semi-supervised learning enhances white matter hyperintensity segmentation. *Front. Comput. Neurosci.* 18:1487877. doi: 10.3389/fncom.2024.1487877
- Fan, X., Liu, H., Zheng, H., Zhai, J., and Zhang, L. (2024). Dacseg: divide and conquer for accurate overlapping chromosome instance segmentation in metaphase cell images. *Biomed. Signal Process. Control* 89:105880. doi: 10.1016/j.bspc.2023.105880
- Fang, M., Shamsi, Z., Bryant, D., Wilson, J., Qu, X., Dubey, A., et al. (2023). Automated deep aberration detection from chromosome karyotype images. *Cancer Genet.* 278, 13. doi: 10.1016/j.cancergen.2023.08.050
- Gersen, S. L. (2013). *The Principles of Clinical Cytogenetics*. Cham: Springer. doi: 10.1007/978-1-4419-1688-4
- Ibrahim, A. U., Ozsoz, M., Serte, S., Al-Turjman, F., and Yakoi, P. S. (2024). Pneumonia classification using deep learning from chest x-ray images during covid-19. *Cogn. Comput.* 4, 1–13. doi: 10.1007/s12559-020-09787-5
- Institute, N. H. G. R. (2023). *Chromosome Abnormalities Fact Sheet*. Available at: <https://www.genome.gov/about-genomics/fact-sheets/Chromosomes-Fact-Sheet>
- Islam, M. M., Talukder, M. A., Uddin, M. A., Akhter, A., and Khalid, M. (2024). Brainnet: Precision brain tumor classification with optimized efficientnet architecture. *Int. J. Intell. Syst.* 2024:3583612. doi: 10.1155/2024/3583612
- James, G., Witten, D., Hastie, T., Tibshirani, R., and Taylor, J. (2023). *An Introduction to Statistical Learning: With Applications in Python*. New York: Springer Nature. doi: 10.1007/978-3-031-38747-0
- Kang, S., Han, J., Lee, I., Joo, H., Chung, Y., and Yang, S. (2024). Chromosome analysis method based on deep learning: counting chromosomes and detecting abnormal chromosomes. *Biomed. Signal Process. Control* 91:105891. doi: 10.1016/j.bspc.2023.105891
- Kao, E., Hsieh, Y., Ke, C., Lin, W., and Yang, F. (2023). Automated identification of single and clustered chromosomes for metaphase image analysis. *Heliyon* 9:e17917. doi: 10.1016/j.heliyon.2023.e16408

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by National University of Sciences and Technology, Islamabad, Pakistan.

Acknowledgments

Authors acknowledge the research support provided by the National University of Sciences and Technology, Islamabad, Pakistan.

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.

- Kong, G., Ma, S., Zhao, W., Wang, H., Fu, Q., and Wang, J. (2024). A novel method for optimizing epilepsy detection features through multi-domain feature fusion and selection. *Front. Comput. Neurosci.* 18:1416838. doi: 10.3389/fncom.2024.1416838
- Li, Y., and Zhong, Z. (2024). Decoding the application of deep learning in neuroscience: a bibliometric analysis. *Front. Comput. Neurosci.* 18:1402689. doi: 10.3389/fncom.2024.1402689
- Li, Z., Zhao, G., Yin, A., Wang, T., Chen, H., Guo, L., et al. (2020). "Cs-ganomaly: a supervised anomaly detection approach with ancillary classifier GANS for chromosome images," in *Proceedings of the 2020 IEEE 3rd International Conference of Safe Production and Informatization (IICSPI)* (IEEE), 492–499. doi: 10.1109/IICSPI51290.2020.9332331
- Madian, N., Jayanthi, K., and Suresh, S. (2018). Analysis of human chromosome images: application towards an automated chromosome classification. *Int. J. Imaging Syst. Technol.* 28, 235–245. doi: 10.1002/ima.22274
- Montobbio, N., Maffulli, R., and Abrol, A. (2024). Computational modeling and machine learning methods in neurodevelopment and neurodegeneration: from basic research to clinical applications. *Front. Comput. Neurosci.* 18:1514220. doi: 10.3389/fncom.2024.1514220
- Natarajan, A. (2002). Chromosome aberrations: past, present and future. *Mutat. Res. Fundam. Molec. Mechan. Mutag.* 504, 3–16. doi: 10.1016/S0027-5107(02)00075-1
- Park, H., Chun, S., Shim, J., Oh, J., Cho, E., Hwang, H., et al. (2019). Detection of chromosome structural variation by targeted next-generation sequencing and a deep learning application. *Sci. Rep.* 9:3644. doi: 10.1038/s41598-019-40364-5
- Poletti, E., Zappelli, F., Ruggeri, A., and Grisan, E. (2012). A review of thresholding strategies applied to human chromosome segmentation. *Comput. Methods Programs Biomed.* 108, 679–688. doi: 10.1016/j.cmpb.2011.12.003
- Qin, Y., Wen, J., Zheng, H., Huang, X., Yang, J., Song, N., et al. (2019). Varifocal-net: a chromosome classification approach using deep convolutional networks. *IEEE Trans. Med. Imaging* 38, 2569–2581. doi: 10.1109/TMI.2019.2905841
- Rosenfeld, J. A., and Patel, A. (2017). Chromosomal microarrays: understanding genetics of neurodevelopmental disorders and congenital anomalies. *J. Pediatr. Genet.* 6, 42–50. doi: 10.1055/s-0036-1584306
- Saleh, H., Saad, N., and Isa, N. (2019). Overlapping chromosome segmentation using u-net: convolutional networks with test time augmentation. *Procedia Comput. Sci.* 159, 524–533. doi: 10.1016/j.procs.2019.09.207
- Slimi, H., Balti, A., Abid, S., and Sayadi, M. (2024). A combinatorial deep learning method for Alzheimer's disease classification-based merging pretrained networks. *Front. Comput. Neurosci.* 18:1444019. doi: 10.3389/fncom.2024.1444019
- Tan, Y., Jin, B., Nettekoven, A., Chen, Y., Yue, Y., Topcu, U., et al. (2019). "An encoder-decoder based approach for anomaly detection with application in additive manufacturing," in *2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA)* (IEEE), 1008–1015. doi: 10.1109/ICMLA.2019.00171
- Uzolas, L., Rico, J., Coupé, P., SanMiguel, J., and Cserey, G. (2022). Deep anomaly generation: an image translation approach of synthesizing abnormal banded chromosome images. *IEEE Access* 10, 59090–59098. doi: 10.1109/ACCESS.2022.3178786
- Wang, X., Zheng, B., Li, S., Mulvihill, J., Chen, X., and Liu, H. (2010). Automated identification of abnormal metaphase chromosome cells for the detection of chronic myeloid leukemia using microscopic images. *J. Biomed. Opt.* 15:046026. doi: 10.1117/1.3476336
- Wu, Y., Yue, Y., Tan, X., Wang, W., and Lu, T. (2018). "End-to-end chromosome karyotyping with data augmentation using GAN," in *2018 25th IEEE International Conference on Image Processing (ICIP)* (IEEE), 2456–2460. doi: 10.1109/ICIP.2018.451041
- Yan, J., Tucci, E., and Jaffe, N. (2019). Detection of t (9; 22) chromosome translocation using deep residual neural network. *J. Comput. Commun.* 7:102. doi: 10.4236/jcc.2019.712010



OPEN ACCESS

EDITED BY

Saad Arif,
King Faisal University, Saudi Arabia

REVIEWED BY

Shahid Latif,
University of the West of England,
United Kingdom
Muhammad Zubair,
University of Central Punjab, Pakistan
Naeem Iqbal,
Queen's University Belfast, United Kingdom

*CORRESPONDENCE

Hafza Ayesha Siddiqua
✉ hafizaasha@yahoo.com
Laishuan Wang
✉ laishuanwang@163.com
Wei Chen
✉ wei.chenbme@sydney.edu.au

RECEIVED 06 October 2024

ACCEPTED 07 January 2025

PUBLISHED 31 January 2025

CITATION

Siddiqua HA, Qureshi MF, Khurshid A, Xu Y,
Wang L, Abbasi SF, Chen C and Chen W
(2025) EEG electrode setup optimization using
feature extraction techniques for neonatal
sleep state classification.
Front. Comput. Neurosci. 19:1506869.
doi: 10.3389/fncom.2025.1506869

COPYRIGHT

© 2025 Siddiqua, Qureshi, Khurshid, Xu, Wang,
Abbasi, Chen and Chen. 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.

EEG electrode setup optimization using feature extraction techniques for neonatal sleep state classification

Hafza Ayesha Siddiqua^{1*}, Muhammad Farrukh Qureshi²,
Arsalan Khurshid³, Yan Xu⁴, Laishuan Wang^{5*},
Saadullah Farooq Abbasi⁶, Chen Chen⁷ and Wei Chen^{8*}

¹Center for Intelligent Medical Electronics, Department of Electronic Engineering, School of Information Science and Technology, Fudan University, Shanghai, China, ²Department of Electrical Engineering, Namal University Mianwali, Mianwali, Pakistan, ³Department of Electrical Engineering, Engineering Institute of Technology, Melbourne, VIC, Australia, ⁴Department of Neurology, Children's Hospital of Fudan University, National Children's Medical-Center, Shanghai, China, ⁵Department of Neonatology, Children's Hospital of Fudan University, Shanghai, China, ⁶Department of Electronic, Electrical and Systems Engineering, University of Birmingham, Birmingham, United Kingdom, ⁷Human Phenome Institute, Fudan University, Shanghai, China, ⁸School of Biomedical Engineering, The University of Sydney, Sydney, NSW, Australia

An optimal arrangement of electrodes during data collection is essential for gaining a deeper understanding of neonatal sleep and assessing cognitive health in order to reduce technical complexity and reduce skin irritation risks. Using electroencephalography (EEG) data, a long-short-term memory (LSTM) classifier categorizes neonatal sleep states. A 16,803 30-second segment was collected from 64 infants between 36 and 43 weeks of age at Fudan University Children's Hospital to train and test the proposed model. To enhance the performance of an LSTM-based classification model, 94 linear and nonlinear features in the time and frequency domains with three novel features (Detrended Fluctuation Analysis (DFA), Lyapunov exponent, and multiscale fluctuation entropy) are extracted. An imbalance between classes is solved using the SMOTE technique. In addition, the most significant features are identified and prioritized using principal component analysis (PCA). In comparison to other single channels, the C3 channel has an accuracy value of $80.75\% \pm 0.82\%$, with a kappa value of 0.76. Classification accuracy for four left-side electrodes is higher ($82.71\% \pm 0.88\%$) than for four right-side electrodes ($81.14\% \pm 0.77\%$), while kappa values are respectively 0.78 and 0.76. Study results suggest that specific EEG channels play an important role in determining sleep stage classification, as well as suggesting optimal electrode configuration. Moreover, this research can be used to improve neonatal care by monitoring sleep, which can allow early detection of sleep disorders. As a result, this study captures information effectively using a single channel, reducing computing load and maintaining performance at the same time. With the incorporation of time and frequency-domain linear and nonlinear features into sleep staging, newborn sleep dynamics and irregularities can be better understood.

KEYWORDS

EEG, sleep analysis, neonatal sleep state classification, principal component analysis, SMOTE, LSTM

1 Introduction

Sleep is a natural, repetitive period of rest and unconsciousness that is required for the healthy functioning of both the body and the mind (Baker, 1985). During sleep, the body undergoes a series of stages, and each stage offers a distinct benefit and influences numerous physiological and psychological functions, including memory consolidation, cognitive function, mood regulation, and physical ability restoration (Song et al., 2024). As a general rule, sleep involves a reduction in consciousness and awareness of the environment, a reduction in voluntary muscle contraction, a decrease in metabolism, and a reversible and periodic state (Arif et al., 2021). As a result of inadequate sleep, cognitive function can be impaired, the immune system weakens, and the risk of chronic diseases increases. These diseases, including obesity, diabetes, heart disease, and hypertension can increase (Khan S. et al., 2020; Killick et al., 2022; Parish, 2009; Pan et al., 2024; Chen and Zhu, 2024). The recommended amount of sleep for adults is between 7–9 h per night (Baker, 1985). Neonates, however, have shorter sleep cycles, making them more susceptible to unpredictable sleep patterns. It is common for infants to sleep approximately 16–17 h per day, but the duration varies depending on the individual.

Just like adults, neonates also go through various sleep stages (Newson, 2017). In neonates, there are two main stages of sleep: Active Sleep (AS) and Quiet Sleep (QS). The infant is in AS state when he or she has rapid eye movements, involuntary breathing, and a rapid heart rate. During this state of sleep, babies are able to move, express their facial expressions, and are even capable of sucking. The development of the brain and the learning process of the infant are directly related to AS. During QS, babies' hearts beat slower, their breathing is regular, and they do not move very much. Physical development and growth are strongly influenced by QS. In addition to the AS and QS stages, infants also experience a third transitional stage in their sleep cycle, which combines both the AS and QS stages. There are two main differences between Active Sleep 1 (AS1) and Active Sleep 2 (AS2). The main difference is how much the brain is active and how much the eyes move. In AS1, highly irregular brain waves and frequent changes in the eye movements are characterized, however, in AS2, the eye movements are less frequent and the brain activity is more regular. As an alternative, QS can be divided into two categories, one of which is Quiet Sleep 1 and the other is Quiet Sleep 2. There is a significant difference between QS1 and QS2, as the movements and brain waves differ significantly. In QS1, there is increased activity, with abnormal brain activity and body movements. As opposed to this, QS2 is a quieter state in which the brain is more active regularly and the body is less active.

1.1 Main motivation of the proposed approach

The primary objective of this study is to evaluate the potential for differentiating neonatal sleep into five states using single-channel and multi-channel EEG data. To identify the best electrode configuration and minimize technical difficulties and potential irritation of the skin that may occur during the collection of EEG

data for neonates, data collected from single-channel EEG is being used. The LSTM algorithm is used to classify an infant's sleep into five stages by using various EEG features including three novel features (Detrended Fluctuation Analysis (DFA), Lyapunov exponent, and multiscale fluctuation entropy).

1.2 Main contributions

There are five main parts to this study, and they are outlined below:

1. Extraction of multiple linear and non-linear features in the time and frequency domains.
2. As a non-linear state-of-the-art approach for EEG-based neonatal sleep staging, Detrended Fluctuation Analysis (DFA), Multiscale Fluctuation Entropy (MFE), and Lyapunov exponent are taken into account.
3. To address class imbalance, the SMOTE technique is used to balance the dataset.
4. PCA-based feature normalization and selection.
5. Using both one channel at a time as well as different combinations of multiple channels at the same time to classify five different sleep states.

In addition, the study examines the optimal configuration of EEG electrodes for five-state classification, including how many electrodes to use and where they should be placed. To reduce complexity, skin irritation risk, and cost in neonatal sleep studies, this study evaluated sleep stage classification accuracy using various electrode setups.

This article is structured as follows: Section 2 reviews relevant literature; Section 3 presents the methodology that has been proposed and its findings based on the proposed methodology; and A discussion of the proposed work's findings and limitations is provided in Section 4. In Section 5, the proposed study's conclusions are presented.

2 Related work

Human sleep behavior was first studied using electroencephalography (EEG) in Loomis et al. (1937). With the advent of deep and machine learning algorithms, there are a number of algorithms that have been developed in order to categorize adult sleep patterns (Lajnef et al., 2015; Xiao et al., 2013; Fonseca et al., 2016; Gudmundsson et al., 2005; Turnbull et al., 2001; De Wel et al., 2017; Dereymaeker et al., 2017; Koolen et al., 2017; Pillay et al., 2018; Ansari et al., 2020; Fraiwan and Lweesy, 2017). Pillay et al. (2018) developed a model based on multichannel EEG recordings to automatically classify a person's sleep using Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs) and the Cohen's Kappa of the model was 0.62, which was higher than the Cohen's Kappa of a GMMs. A CNN was also used to classify sleep stages 2 and 4 (Ansari et al., 2020). Wake states were not included in these techniques. In Awais et al. (2020), developed using pre-trained CNNs to extract features to classify neonatal sleep and wake. According to this study, a model that has been pre-trained was inadequate for categorizing sleep and

wake in neonates with high accuracy. In [Awais et al. \(2021\)](#), the authors combine deep convolutional neural networks (DCNN) with self-learning models to classify infant sleep and waking states based on video facial expressions. EEG video data could be classified accurately at $93.8 \pm 2.2\%$ and F1-scores were 0.93 ± 0.3 . It is worth mentioning that video EEG data can contain infant's faces and voices, creating privacy issues as a result.

A study conducted in 2021 by authors in [Lee et al. \(2021\)](#) with IR-UWB radar to classify non-contact sleep and wake in infants found an accuracy of 75.2%. According to another study that classified quiet sleep based on EEG data, the value of Kappa was 0.77 ± 0.01 for eight-channels and 0.75 ± 0.01 for single bipolar-channel ([Ansari et al., 2021](#)). According to a study conducted by Abbasi et al., a MLP neural network algorithm developed for binary classification of neonatal sleep has been tested and the value of Kappa has been determined to be 62.5%, and the accuracy has been determined to be 82.5% using the algorithm ([Abbasi et al., 2020](#)). A three-state classification of the same dataset was performed in 2022 using bagging and stacking ensemble methods with an accuracy of 81.99% ([Abbasi et al., 2022](#)). By using publicly available single-channel EEG datasets, [Yu et al. \(2022\)](#) classified neonate's sleep patterns into W, N1, N2, and N3. The multi-resolution attention sleep network (MRASleepNet) module was tested to classify sleep patterns. A feature extraction module, a multi-resolution analysis module, and a gated MLP module were all included in the algorithm. Through an adaptive boosting (AdaBoost) classifier, [Arasteh et al. \(2023\)](#) classified AS and QS with 81% accuracy achieved through cross-validation of tenfold. The AutoML-based Random Forest estimator obtained an accuracy rate of 84.78% and a kappa rate of 69.63% for prediction of neonatal sleep and wake states in [Siddiqi et al. \(2023\)](#). According to [Ansari et al. \(2018\)](#), an 18-layer CNN is used to detect neonatal QS sleep stages with multichannel EEG data. A Multi-Scale Hierarchical Neural Network (MS-HNN) has been developed in [Zhu et al. \(2023\)](#) Using two, four, and eight channels to automatically classify neonatal sleep states. Features including temporal information were extracted using multi-scale convolutional neural networks (MSCNN). They attained an accuracy of 75.4% using single-channel classification and 76.5% using a combination of eight channels for three-stage classification. [Supratak et al. \(2017\)](#) performed classification of sleep states in newborn with DeepSleepNet and attained 69.8% accuracy. In [Eldele et al. \(2021\)](#), authors proposes AttenSleep, a deep learning approach based on attention for sleep stage classification. Instead of using RNNs, AttenSleep uses multi-head attention (MHA) to identify the chronological relationship among different stages of neonatal sleep. Using multi-branch CNN and reached classification accuracy of 74.27% with single channel and 75.36% with four channel EEG, Hafza et al. proposed three-state EEG-based neonatal sleep state classification ([Siddiqi et al., 2024](#)). The authors incorporated 74 features in the time and frequency domains.

As a result of limited classifications, privacy concerns, long training times, and poor accuracy, existing approaches for recognizing infant sleep stages have significant limitations. Without taking into account awake, it is challenging to classify newborn sleep accurately. Non-linear features which aren't typically included in current sleep staging methodologies for neonates include DFA, MFE, and the Lyapunov Exponent. Further, these methods

require multichannel EEG data, which disrupts the skin and causes discomfort, highlighting the need for methods that are non-invasive. To effectively differentiate between the five-state sleep patterns in newborns, it is crucial to develop a dependable and privacy-conscious strategy that ensures high accuracy while minimizing any potential negative consequences.

3 Materials and methods

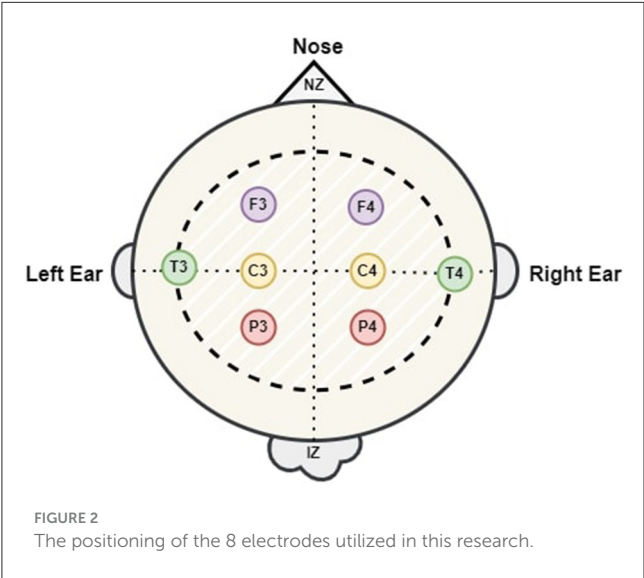
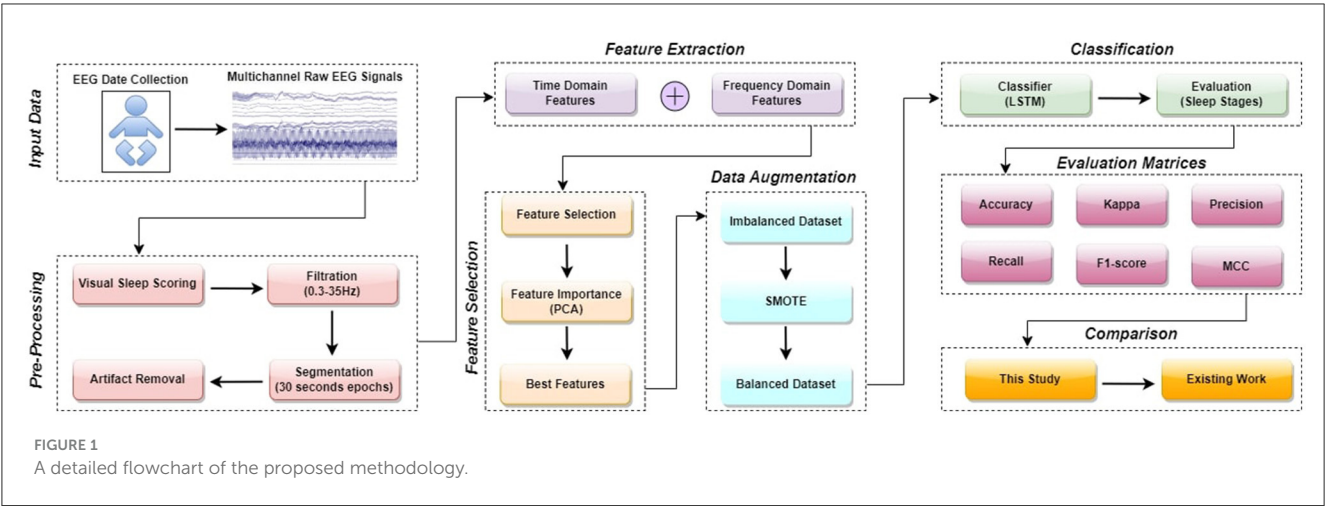
An LSTM model for the categorization of neonate's sleep into five distinct states is introduced in this article. In this section, a step-wise overview of the proposed design is provided. The sequential flowchart of the proposed methodology is illustrated in [Figure 1](#). The process can be further explained by following these steps:

3.1 EEG dataset

EEG data was obtained from 64 neonates admitted to the neonatal intensive care unit (NICU) at Children's Hospital of Fudan University (CHFU), located in China. This work has obtained approval from the Research Ethics Committee of Children's Hospital of Fudan University, with the assigned Approval No. (2017) 89. The proposed model was tested and trained using these EEG recordings. The data was collected during observations of neonates at various time points. A full 10-20 electrode installation system comprises the following 17 electrodes: "FP1," "FP2," "F3," "F4," "F7," "F8," "C3," "C4," "P3," "P4," "T3," "T4," "T5," "T6," "O1," "O2," and "Cz." Every letter is associated with a distinct region or lobe of the brain. The letters FP, F, T, P, O, and C represent the prefrontal, frontal, temporal, parietal, occipital, and central regions of the brain. Throughout this time frame, we have witnessed a multitude of sleep patterns. The study included EEG recordings from eight specific channels: "C3," "C4," "F3," "F4," "P3," "P4," "T3," and "T4." The NicoletOne multi-channel EEG equipment was utilized for the purpose of recording of the EEG data at a sampling rate of 500 Hz. The NicoletOne EEG devices have lightweight electrode caps that securely fasten scalp electrodes, ensuring accurate signal capture. The NicoletOne EEG device enables the acquisition of high-quality EEG signals with a high sampling rate of up to 2 kHz and a broad frequency range spanning from 0.053 to 500 Hz. [Figure 2](#) illustrates the locations of the eight electrodes used in this study, in accordance with the 10–20 system recommendations. Nz represents the foundation of the nose, whereas Iz indicates the protuberance.

3.2 Visual sleep scoring of EEG dataset

The EEG segments were visually classified by experienced neurologists from Fudan children hospital Shanghai, based on five main categories: Wakefulness, AS1, AS2, QS1, and QS2. When classifying sleep states, non-cognitive features were employed in conjunction with the EEG. In addition, the experts took into account NICU videos when conducting the annotating procedure. [Table 1](#) provides comprehensive details regarding the dataset ([Siddiqi et al., 2024](#)).



3.3 EEG dataset pre-processing

Distortion and artifacts during recording have an impact on the quality and reliability of the EEG data. The EEG data was recorded at a sampling rate of 500 Hz. These EEG recordings underwent a pre-processing phase to eliminate noise and artifacts. The pre-processing involves the following steps:

1. An FIR (Finite Impulse Response) filter was employed to eliminate undesired signals from EEG recordings within the frequency range of 0.3 to 35 Hz (High Pass = 0.3 Hz and Low Pass = 35 Hz).
2. The EEG signals that have been processed by a filter are now divided into segments of 30 seconds each.
3. Following the process of segmentation, a label given by experienced neurologists is issued to each epoch. The five-state classification assigns W as the first state, AS1 as the second state, QS1 as the third state, QS2 as the fourth state, and AS2 as the fifth state.

TABLE 1 A detailed description of the dataset (Siddiqua et al., 2024).

Variable/category	Descriptions
Sampling frequency	500 Hz
Number of channels	8
Number of subjects	64
Number of epochs	16,803
Gestational age	38.3 ± 1.8 (wk+d)
Post-menstrual age	40.5 ± 1.7 (wk+d)
Gender	32 males and 32 females
Sleep time	1.44 ± 0.57 h
Wake time	0.71 ± 0.57 h
Weight	3.3 ± 0.6 kg
Reason for admittance	Septicemia, Hyperbilirubinemia, and etc.

4. Artifacts and noise were introduced into the EEG recordings during the recording and processing stages. Consequently, following the pre-processing stage, there are a total of 16,803 epochs available for the testing and training over the channels “C3,” “C4,” “F3,” “F4,” “P3,” “P4,” “T3,” and “T4.”

3.4 Feature extraction

The extraction of features from the EEG signals is essential for categorization. Since it aids in distinguishing among various sleep stages or events by analyzing patterns and characteristics. Interpreting EEG data can be difficult because of the fact that there are so many signals that change over time produced as a result of electrical activity in the brain. This work utilizes linear and non-linear feature extraction techniques to decrease the number of dimensions of the data that need to be analyzed and identify relevant characteristics of the data that can be employed for categorization purposes, such as frequency and time distributions (Gosala et al., 2023; Khan J. S. et al., 2020). Overall, 94 linear

and non-linear features were retrieved from each channel utilizing various procedures, which include:

3.4.1 Time domain features

- *Statistical features of the EEG signal and its first and second derivatives*: To study and summarize the main statistical features of the EEG signal as well as its derivatives, it would be helpful to pull out features in the time domain of the dataset in order to group newborn's sleep stages (Siddiqua et al., 2023). The extraction of time-domain features is a valuable as well as practical approach to evaluating EEG data, serving both clinical and research applications. Initially, the signal's nine statistical characteristics (mean, median, standard deviation, minimum, maximum, kurtosis, skewness, variance, and range) are computed. Subsequently, an identical collection of five statistics is computed for both the first derivative of the signals obtained from the EEG as well as the second derivative.
- *Detrended fluctuation analysis (DFA)*: It is a non-linear feature, computed to measure if EEG signals are correlated at either long or short ranges or if they are self-similar. It also quantifies the extent to which the fluctuations of a signal, after being combined and detrended at various epochs, diverge from a linear pattern (Lal et al., 2023). The DFA, or Detrended Fluctuation Analysis, is a mathematical measure that quantifies the scaling exponent characterizing the connection between the amplitude of fluctuations and the corresponding time scales. The equation for calculating the DFA is as follows:

$$F(n) = \sqrt{\frac{\sum [Y(i) - y(i)]^2}{n}} \quad (1)$$

The fluctuations are represented by $F(n)$ for window size n , the integrated or cumulative profiles of the EEG data are represented by $Y(i)$, and the regression line is represented by $y(i)$. To calculate the DFA, this study uses the `nolds.dfa()` function from the `nolds` library. The EEG signal data are converted into NumPy arrays and the DFA is calculated. Conversely, lower values of DFA imply less reliable correlations or less predictable patterns, whereas high values of DFA show better correlations over long distances or similarity to itself, and this implies that a signal is more structured and easier to predict. By utilizing the various DFA characteristics, individuals can acquire a deeper understanding of what is going on within the signal as well as its intricacy. These characteristics have the potential to be advantageous in a range of different applications, such as the evaluation of signals, statistical analysis of time series, and biological studies as well.

- *Lyapunov exponent*: The Lyapunov exponent, a nonlinear feature, measures the responsiveness of a dynamical system to its initial circumstances (Cao et al., 2023). EEG feature extraction is a valuable tool for understanding the predictability and stability of brain processes. The Rosenstein approach is employed to calculate the Lyapunov

exponent based on EEG data. The algorithm involves defining parameters for data embedding, initializing tangent vectors, and performing Jacobian matrix calculations. The QR decomposition is used to orthogonalize the tangent vectors, which are then normalized to quantify the system's sensitivity to perturbations. Logarithms of Jacobians divided by tangent vectors and iterations determine the Lyapunov exponent. The Lyapunov exponent (λ) is given by:

$$\lambda = \frac{1}{N-1} \sum_{n=1}^{N-1} x \quad (2)$$

where,

$$x = \log \left(\frac{d(n+1)}{d(n)} \right) \quad (3)$$

This sum is taken over time steps from $n = 1$ to $N - 1$, where N is the total number of time steps. The variable x describes the relative changes in distances between nearby trajectories in the dynamical system, which is used in calculating the Lyapunov Exponent to characterize the behavior and predictability of the dynamical system. The term x represents the logarithm between $d(n+1)$ and $d(n)$, which are the distances of the perturbed trajectory at time $n+1$ and n , respectively. The Lyapunov exponent values not only offer insight into the classification of sleep stages in EEG analysis but also provide information about how complex neonatal sleep dynamics are and the extent to which they can be predicted.

- *Multiscale fluctuation entropy (MFE)*: Within the scope of the present study, MFE values have been computed for every epoch of EEG data in order to measure the degree of complexity as well as the irregularity of the signal (Wan et al., 2023). The standard deviation is calculated segment by segment using a scaling factor. The procedure entails multiple sequential stages. The EEG signal is divided into segments according to the scale factor. The variation of each segment is determined by comparing the standard deviation of each segment to the mean of each segment and then calculating the average of the standard deviations of each segment. There is a formula known as the Shannon entropy formula, which is employed to calculate entropy for an ensuing string of fluctuations. Mathematically, MFE can be written as:

$$MFE = \frac{1}{K} \sum_{k=1}^K H_k \quad (4)$$

In this case, Shannon entropy at each scale is represented by H_k , and total number of scales is represented by K . The objective of this work is to obtain a deeper understanding of the complexities and inconsistencies of neural activity at different levels by calculating MFE values. It specifically aids in the study of EEG data, which offers vital insights into underlying brain activity through the examination of fluctuation and complexity patterns.

3.4.2 Frequency domain features

Frequency domain features play a crucial role in the interpretation of EEG signals, since they are necessary for the diagnosis of neurological illnesses and for monitoring the brain's activity during the performance of cognitive functions. This research computed the subsequent features in the frequency domain:

- *Identification of central tendency features using EEG band's spectral features:* The spectral analysis of the four frequency bands (delta, theta, alpha, and beta) in an EEG signal can be utilized for determining central tendency attributes such as mean, median, mode, variance, standard deviation, kurtosis, skewness, minima, and maxima (Siddiqa et al., 2023). The central tendency of a dataset can be defined as the tendency of a dataset to accumulate around the average value or center of the dataset. A measure of the central tendency can offer insights into the common or predominant values found in a dataset. They have the ability to depict and provide a concise overview of data distributions. In order to compute central tendency characteristics based on spectral statistics, it was first necessary to determine the power spectral density (PSD) of the EEG data that was initially determined (Arif et al., 2023). Using Welch's method, PSD is calculated by segmenting the EEG signal into overlapping windows, computing the Fourier transform for each segment, and averaging the spectra to estimate the PSD. A more detailed spectral analysis of the EEG signal can be obtained by using this method. As a default, the resolution parameter is set to none. By doing this, the function determines the segment length automatically based on the input data length. As a default, the behavior attempts to strike a reasonable balance between frequency resolution and computational efficiency. Subsequently, the PSD has been subdivided into distinct frequency ranges: delta (0.5–3 Hz), theta (4–7 Hz), alpha (8–12 Hz), and beta (13–30 Hz). Afterwards, a total of 32 features representing central tendency of each frequency band were computed using frequency band spectral statistics.
- *Norm power of four EEG bands:* The normalized power is calculated by dividing the power inside each frequency band by the integral of the overall power spectral density (PSD) across all frequencies. By normalizing the power levels, it ensures a justifiable comparison of power levels across various frequency bands, while taking into consideration the fluctuations in the total power spectrum of the EEG signal. The normalized power values are useful parameters for classifying infant sleep stages because they represent the relative contribution and distribution of brain activity in specific frequency ranges.
- *Average frequency of four EEG bands:* The average frequency of each of four EEG band is determined by multiplying the frequencies within the relevant frequency indices by their respective PSD values. Subsequently, those values are added together, and the outcome is divided by the total sum of the PSD values within the specified frequency range. This calculation yields a weighted average frequency that signifies the central point or the most prominent frequency within the particular range of frequencies under consideration. This

technique enables a numerical evaluation of the spectrum properties of the EEG data and offers a valuable understanding of the frequency distribution within each EEG band. It also aids in the classification of various sleep stages in neonates.

- *Maximum power of four EEG bands:* The maximum power of each EEG frequency band is determined by determining the frequency indices in the PSD that correspond to the specific frequency range of interest for each band. The indices are derived by comparing the frequency values with the lower and upper frequency limitations specified for each band. The highest PSD value within these specific frequency indices is subsequently obtained for each time point, resulting in the peak power level within the corresponding frequency range. In EEG signals, time points are discrete instances where the PSD can be estimated. This computation allows for the determination of the maximum intensity of brain activity within each distinct frequency band and offers vital insights into the prevailing power peaks found in the EEG signal.
- *EEG band ratios:* The power ratios between EEG frequency bands are calculated by dividing the normalized power of one band by the normalized power of another band. These ratios, such as the delta-theta ratio, alpha-beta ratio, delta-alpha ratio, theta-beta ratio, delta-beta ratio, and theta-alpha ratio, enable the evaluation of the proportional distribution of power and interactions among different frequency bands. The ratios are calculated using the normalized power values derived from the PSD analysis of the EEG data. The normalized power quantifies the relative impact of a particular frequency range in the complete power spectrum. The power ratios are obtained by dividing the normalized power of one band by the normalized power of another band. These ratios offer vital information into the equilibrium and supremacy of brain activity across various frequency ranges. Their contribution involves analyzing EEG data to characterize different sleep stages in newborns, providing insights on the relative importance of specific frequency components in the EEG spectrum.
- *Fast fourier transform (FFT):* By employing FFT, it is possible to examine the time-domain EEG signal by interpreting it into the frequency domain and analyzing its constituent frequency components. The input EEG data was subjected to a FFT to calculate its frequency spectrum. Subsequently, the 10 frequencies with the most significant FFT values were selected.

Consequently, all the above mentioned characteristics can be used to create automated sleep staging algorithms that have the potential to enhance the identification and treatment of infant's sleep disorders.

3.5 Feature importance and feature selection

In order to classify sleep states using EEG, we need to define what features in the frequency and time domains are the most informative. By using these techniques, we can distinguish sleep stages by using the most informative features. Using machine

learning models, you can achieve better performance and more accurate results by selecting and emphasizing features (Ilyas et al., 2020). In this research, Principal Component Analysis (PCA) is utilized to select and prioritize features. The PCA algorithm determines which of the principal components captures the greatest proportion of variance in a dataset by analyzing its variance (Wold et al., 1987). The explained variance ratio can be used to determine a subset of principal components can be selected to reduce the dimensionality of the data. High variances indicate that the number of features in the dataset captures as much information as possible. By preserving the variance in the dataset, information which is the most important and relevant to the data can be preserved, and at the same time, the least important data can be eliminated. As a result of the designed PCA, 95% of the variance in the EEG was explained by the most informative features. A small number of principal components account for 95% of the variance in the dataset. After scaling the dataset and performing PCA, we found that a few principal components captured most of the variance. Using the columns that have been selected and the variable that is being targeted, a new dataframe is generated based on how many principal components there are. As a result, the information relevant to the prediction of the variable that is being targeted remains, and at the same time, the data is reduced in dimensionality. In the original dataset, 94 features from preprocessed EEG data were extracted. The resulting dataframe is used to classify sleep into five states using an LSTM model. However, based on PCA results, a total of 21 features have been decided upon for further consideration.

3.6 Synthetic minority oversampling technique analysis

SMOTE is a widely utilized data augmentation approach employed to deal with class imbalance in machine learning (Fernández et al., 2018). It is especially efficient when handling datasets in which one class is considerably less represented than the other. The process involves generating artificial data points for the underrepresented category by interpolating between adjacent examples. The objective of this strategy is to create more synthetic instances that closely resemble the existing samples from the minority class, hence enhancing their presence in the dataset (Fernández et al., 2018). The creation of synthetic samples includes the subsequent steps:

1. Determine the instances belonging to the minority class: Initially, the dataset is examined to identify the instances that belong to the minority class.
2. Randomly choose an instance x_i from the specified minority class instances.
3. Locate the k nearest neighbors: The k nearest neighbors of the given instance are determined using a selected distance metric, such as Euclidean distance (Li et al., 2024).
4. Choose one of the k nearest neighbors at random: A single neighbor is selected at random from the k nearest neighbors.
5. Create a synthetic instance x_{new} : A novel synthetic instance is generated by interpolating between the selected instances and the chosen neighbor. This is achieved by employing a random selection process to choose a point located on the line segment that connects the two instances (Li et al., 2024).

$$x_{new} = x_i + (\hat{x}_i - x_i)\delta \quad (6)$$

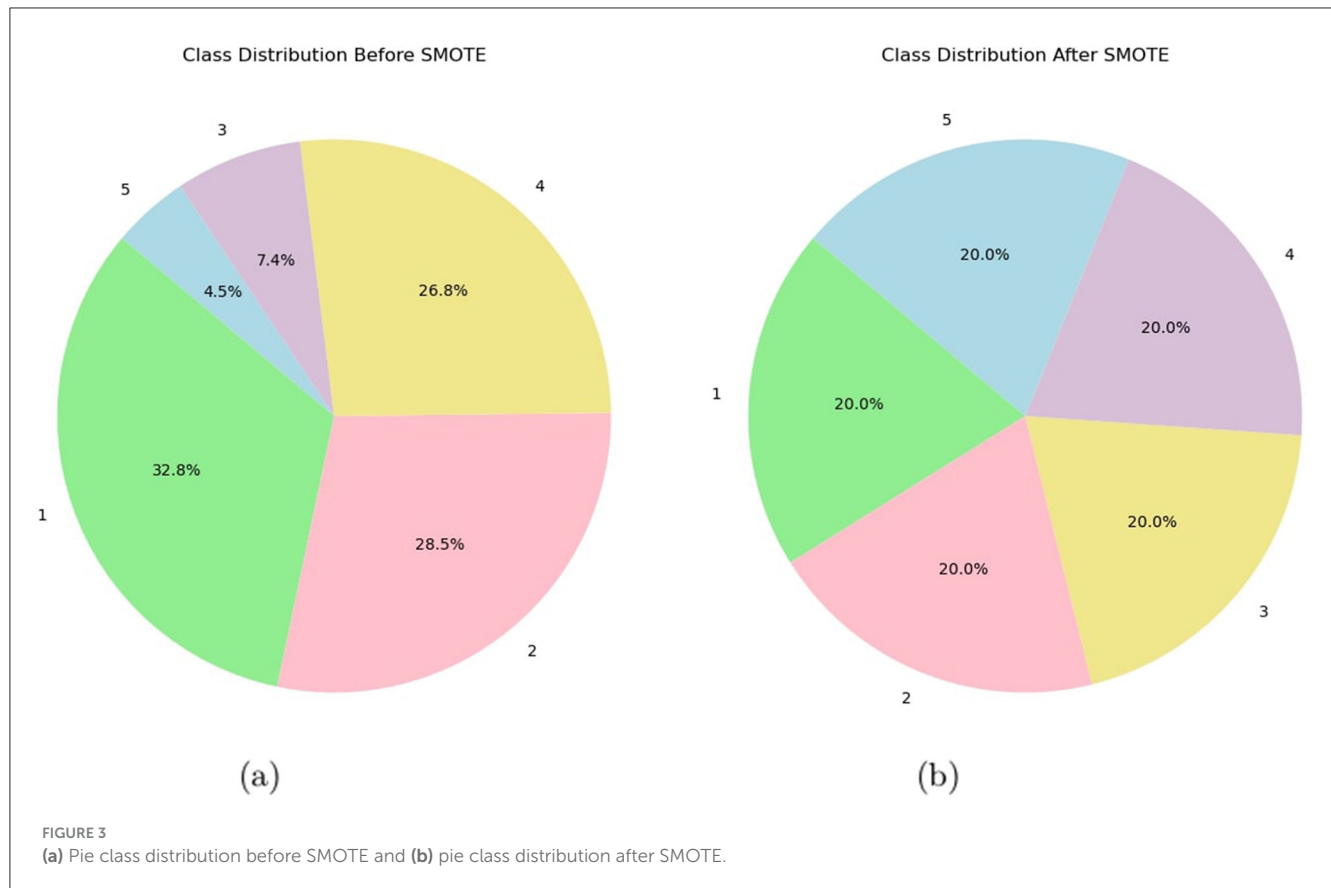
Interpolation between the x_i and $\hat{x}_i$ is controlled by δ , a value between 0 and 1. The value of δ specifies the extent of “smoothing” or “stretching.” The closer the synthetic samples are to the originals, the smaller the value of δ , and the farther they are from them, the larger the value.

6. Iterate the procedure: Steps 2 to 5 are iterated until the required extent of oversampling of the minority class is attained.

When applying SMOTE in the analysis of EEG features, the default delta value was used for oversampling, as specified by the SMOTE implementation. By defaulting the delta value, the implementation process becomes easier, ensuring a standard oversampling level without the need to tweak parameters manually, thereby making class imbalances easier to handle. The SMOTE algorithm is utilized in this specific study, employing the implementation provided by the scikit-learn module. The SMOTE function begins execution with a random state of 42. The effectiveness of the SMOTE technique is assessed by computing and presenting the counts of the resampled labels using a Pandas series. This analysis offers valuable information on the distribution of the balanced classes following the implementation of SMOTE. Figure 3 shows pie class distribution before and after SMOTE. This algorithm provides synthetic samples for the training set, improving the model’s generalization and prediction capabilities (Gamel et al., 2024). A more precise representation of the fundamental distribution of the data is provided by this approach, which lessens the challenges faced by imbalanced datasets. The proposed methodology thus eliminates class imbalances and improves the performance of the model by training it on a more representative and balanced dataset. Using SMOTE, data leakage was prevented and model evaluation was ensured in this research after the train-test split. As a result of applying SMOTE only to the training set, the test data was kept intact, enabling us to assess model performance accurately. By doing so, the test set remains intact, simulating real-world conditions and enhancing model generalization.

3.7 Long short-term memory

An LSTM (Long Short-Term Memory) model is a variant of a recurrent neural network (RNN) that addresses long-term dependencies in sequential data. When processing long sequences, traditional RNNs struggle to capture information from earlier time steps due to the vanishing gradient problem. It can process entire sequences of data, not just individual data points, due to its feedback connections, unlike traditional neural networks. As a result, it is very effective at identifying and predicting patterns in sequential data, such as time series, text, and speech. As a powerful tool for artificial intelligence and deep learning, LSTMs are enabling breakthroughs in a wide range of fields by capturing valuable insights from sequential data.



3.7.1 LSTM architecture

A LSTM network resolves the problem of vanishing gradients faced by RNN. At a high level, LSTM functions similarly to an RNN cell. **Figure 4** illustrates its internal workings. As shown in **Figure 4**, the LSTM network architecture is composed of three components, each of which performs a specific task. Based on the previous timestamp, the first component determines whether the information is relevant or not. Using the input in this cell, the second component tries to learn new information. Finally, in the third component of the cell, the current timestamp is passed on to the next timestamp. The single-time step of the LSTM is considered to be one cycle. Gates are three components of LSTM units. The flow of information between the memory cell and the LSM cell is controlled by them. The forget gate is the first gate, the input gate is the second gate, and the output gate is the last gate. LSTM units composed of these gates and memory cells are similar to layers of neurons in traditional feed-forward neural networks, with each neuron having a current state and a hidden layer. Following is the step-by-step explanation of each gate:

1. **Forget gate:** This gate determines which information from the previous cell state should be discarded. Using the sigmoid activation function, which squashes values between zero and one, the forget gate output (f_t) is calculated from the current input (x_t) and the previous hidden state (h_{t-1}).

A forget gate can be described mathematically as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (7)$$

In this equation, σ represents sigmoid function, W_f represents the forget gate's weight matrix, $[h_{t-1}, x_t]$ represents the concatenation of the previous hidden state with the current input, and b_f is the gate's bias term.

2. **Input gate:** As the input gate determines the amount of new information to be stored in the state of the cell, it takes into account both the current input and the previous hidden input (h_{t-1}). A sigmoid activation function is used to compute the input gate output (i_t).

Input gates are mathematically defined as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (8)$$

$$C_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (9)$$

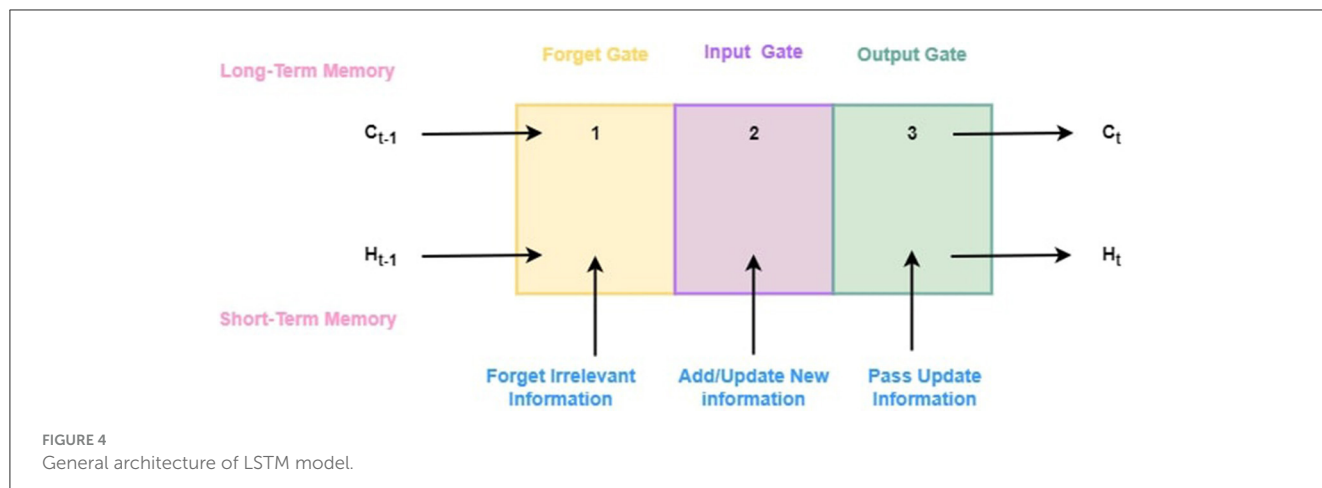
In this case, W_i and W_C stands for the weight matrices associated with the input gate, h_{t-1} and x_t stand for the previous hidden state and current input, while b_i and b_C stands for the bias terms associated with the gate.

3. **Output gate:** By comparing the current input (x_t) with the previous hidden state (h_{t-1}), it determines which parts of the cell state should be output. The output gate output (o_t) is determined by the sigmoid activation function.

An output gate's mathematical equation is as follows:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (10)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (11)$$



Here, W_o represents the weight matrix associated with the output gate, $[h_{t-1}, x_t]$ represents the concatenation of the previous hidden state and the current input, while b_o represents the output gate bias term and h_t shows the output for hidden state.

In an LSTM cell, the gate outputs (f_t , i_t , o_t) are important for controlling information flow. As a result, they determine which parts of the previous cell state should be forgotten, which new information should be added to the cell state, and which parts of the updated cell state should be hidden.

3.8 Proposed model architecture

The proposed LSTM model for neonatal sleep staging is presented in this subsection with detailed descriptions of the mathematical model, its architecture, and all parameters. In this paper, an eight-layer LSTM architecture has been proposed in order to represent the LSTM. Figure 5 provides a comprehensive depiction of the model's structure and offers in-depth insights into its individual layers. Sequentially stacking LSTM layers, this model consists of three layers with different regularization levels and units. There are 500 units in the first layer, and it returns sequences, while there are 250 units in the second layer, and it also returns sequences. LSTM layers are regularized using L2 regularization with a factor of 0.0001 to prevent overfitting. The third layer does not return sequences and has 100 units. Each LSTM layer is followed by a batch normalization layer for speed and stability. After two dense layers of 100 and 50 units, respectively, and ReLU activation, a final dense layer with a number of units corresponding to the classification task's classes is added, and class probability is output using softmax activation. Adam's optimizer, cross-entropy loss function, and accuracy metric are used to compile the model. During training, the model's states and parameters are reset, and with a batch size of 128 and an early stopping with a patience of 10 is implemented. Table 2 presents details about all other hyper-parameters used in proposed LSTM. Experimentation was conducted in order to select and tune all hyperparameters in order to optimize

performance and convergence during training. The model is trained and evaluated for one epoch using the data provided. Then the model's performance on the validation set is evaluated after each epoch.

3.9 Performance assessment metrics

In order to test and evaluate the proposed scheme, different performance metrics are used, including confusion matrix, accuracy, Cohen's kappa, recall, precision, Mathew's co-relation coefficient, and F1-score. In this study, the classification model is examined based on these metrics to determine whether it can accurately identify EEG patterns.

- **Confusion matrix:** An analysis of a classification model's quality is conducted using a confusion matrix. In multi-class classification, confusion matrixes show the number of correct and incorrect predictions for each class as a tabular representation of the model's performance. Identifying specific types of classification errors helps to improve the model's accuracy for individual classes. It is possible to evaluate the model's performance across multiple classes by calculating metrics such as precision, recall, and F1-score.
- **Accuracy:** The accuracy of machine learning (ML) algorithms is commonly measured as a percentage of correctly classified measurements. The formula (Ali et al., 2020) can be used to calculate this percentage:

$$Acc = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (12)$$

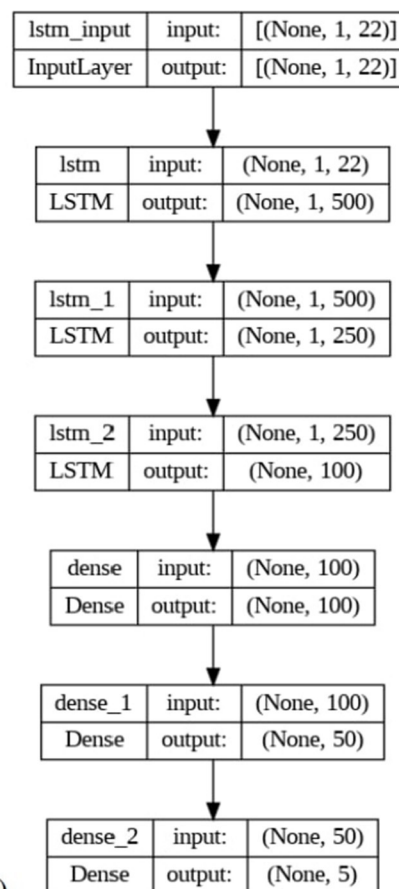
- **Cohen's Kappa:** The Cohen's Kappa is commonly used to estimate how well two raters agree. It is also used to determine the performance of classifiers. The confusion matrix cells are used to calculate it as follows (Chicco et al., 2021):

$$kappa = \frac{2(TP \cdot TN - FP \cdot FN)}{(TP + FP) \cdot (FP + TN) + (TP + FN) \cdot (FN + TN)} \quad (13)$$

Model: "sequential_10"

Layer (type)	Output Shape	Param #
lstm_30 (LSTM)	(None, 1, 500)	1046000
batch_normalization_27 (Batch Normalization)	(None, 1, 500)	2000
lstm_31 (LSTM)	(None, 1, 250)	751000
batch_normalization_28 (Batch Normalization)	(None, 1, 250)	1000
lstm_32 (LSTM)	(None, 100)	140400
batch_normalization_29 (Batch Normalization)	(None, 100)	400
dense_30 (Dense)	(None, 100)	10100
dense_31 (Dense)	(None, 50)	5050
dense_32 (Dense)	(None, 5)	255
Total params: 1956205 (7.46 MB)		
Trainable params: 1954505 (7.46 MB)		
Non-trainable params: 1700 (6.64 KB)		

(a)



(b)

FIGURE 5

(a) Detailed information about LSTM layers. (b) An overview of the model's architecture.

TABLE 2 Details about hyper-parameters.

Parameter	Value
Epochs	50
Batch size	128
Optimizer	Adam
Kernel regularization	L2
Learning rate	1×10^{-4}
Cross-validation k-folds	10
Loss function	Binary cross-entropy

When Kappa is -1 , it is the worst, and when it is $+1$, it is the best.

- **Recall:** Recall in machine learning refers to how well an algorithm can identify a class based on a set of sampled data. In mathematics, recall is expressed as [Shaukat et al. \(2020\)](#):

$$Rec = \frac{TP}{TP + FN} \quad (14)$$

- **Precision:** In order to determine a model's precision, it must be able to identify a significant number of relevant items. Accordingly, it can be written as follows ([Shaukat et al., 2020](#)):

$$Pre = \frac{TP}{TP + FP} \quad (15)$$

- **Matthews correlation coefficient (MCC):** MCC measures the difference between the predicted values and recorded values. The confusion matrix is used to calculate this ([Chicco et al., 2021](#)):

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \quad (16)$$

MCC value of -1 is the worst, while a value of $+1$ is the best.

- **F1-Score:** F1-score is the combination of recall and precision, making it a powerful metric. It is mathematically computed by [Shaukat et al. \(2020\)](#), and [Bing et al. \(2022\)](#):

$$F1_Score = \frac{2 \times Pre \times Rec}{Pre + Rec} \quad (17)$$

- *Accuracy line graph*: The accuracy line graph permits comparisons, thresholds, and determinations of the model's performance over a range of values. This graph displays accuracy values along the Y-axis and fold counts along the X-axis. On the graph, every data point represents an individual cross-fold's accuracy. As the number of folds increases, the line connecting the data points indicates a trend in accuracy.
- *Validation accuracy curve*: Validation accuracy curves for N-fold cross-validation show how accuracy changes over time for each of the N folds. One can visualize the model's performance across different subsets of data by plotting validation accuracy vs. training iterations or epochs. As well as providing valuable insights into the model's learning behavior, this visualization allows assessment of the model's stability and generalization ability.

4 Results

To evaluate the performance of the model, a 10-fold cross-validation procedure was used. The data sets were shuffled randomly beforehand to avoid bias. Ten subsets of data were used for this methodology, with one set serving as the testing set and the remaining nine sets serving as the training set. Thus, it was possible to assess the generalization performance of the model in a way that minimized the leakage between the training and testing phases. As a result of the rigorous methodology used in this study, the performance of the proposed model has been rigorously and unbiasedly evaluated. In this study, the F3-channel and C3-channel show the greatest confusion matrix values when it comes to single-channel EEG data. In [Figure 6](#), confusion matrices for the combinations of channels on the left and right sides and all single channels are shown. [Tables 3, 4](#) present the analytically computed values for each performance assessment metric. For the combinations of channels on the left and right sides and all single channels, a line graph showing the level of accuracy can be seen in [Figure 7](#). The accuracy values are displayed on the Y-axis in [Figure 7](#). In [Figure 7](#), accuracy line graphs represent model performance during 10 cross-folds. Lastly, [Figure 8](#) illustrates validation accuracy curves for C3 single-channel and a combination of four left-side channels.

5 Discussion

Using an LSTM classifier, this study proposes a method of neonatal sleep staging based on single-channel and then four-channel EEG data. In order to determine which EEG channel is important in neonatal sleep staging and which channels are most appropriate for five-state classification, single-channel EEG data needs to be used to determine which channel and which side of the head should be used. After preprocessing the EEG data collected from 64 infants, 16,803 segments are left for testing and training of channels F3, F4, C3, C4, P3, P4, T3, and T4. EEG data is then processed for 94 linear and non-linear features. These features are divided into the time and frequency domains. A total of 27 statistical parameters were included in the analysis for the time domain, including mean, median, standard deviation,

minima, maxima, range, skewness, and kurtosis. The data was further processed to extract nonlinear features such as Detrended Fluctuation Analysis (DFA), Lyapunov exponents, and Multiscale Fluctuation Entropy. A FFT is used in order to extract frequency domain features by separating ten features based on their FFT values and then using spectral statistics to calculate 36 central tendency features for each frequency band in the first place. Through the capture of complex dynamics and irregularities in neonatal EEG signals, these features allow a better understanding of neonatal sleep patterns. By preserving 95% of the variance of the data, we reduced the dimensionality and retained the most informative features by applying Principal Component Analysis (PCA). The Synthetic Minority Oversampling Technique (SMOTE) is also applied for data augmentation to address the imbalanced nature of the dataset. By using this technique, we were able to improve the classification model by balancing the classes.

A description of the proposed LSTM has already been provided in [Section 3](#). A model is used to classify sleep states using 94 features that are obtained from each channel of the EEG signals. Four channels on the left side and four channels on the right side are combined in order to determine the neonate's sleep states. [Figure 5](#) shows the proposed LSTM in its entirety. The description of all layers and their types, as well as their parameters, can also be found in [Figure 5](#). It has been tried many times to get the best performance from the model by testing kernel regularization, unit number, and activation function in the real world. A final choice was made by considering how to balance the complexity of the model with the generalizability of the models after testing a variety of combinations and assessing the effectiveness of each combination. The performance evaluation step involved a 10-fold cross-validation procedure. This methodology used ten subsets of data, nine of them as training sets and one as a test set. In order to eliminate bias in the data sets, the data sets were shuffled prior to the analysis at random. Thus, the generalization performance of the proposed model could be assessed without leaking information between the phases of training and testing of the model. This unbiased evaluation method was used to rigorously and unbiasedly evaluate the performance of the proposed method. In this study, accuracy and other matrices values are expressed as Mean $\pm$ SD. Using the mean, one can see how accurate the experiments are, while the standard deviation indicates how uncertain or variable the accuracy measurements are. Averaging the individual accuracy values obtained from multiple trials yielded the mean accuracy, whereas the standard deviation measures how far the accuracy measurements are from the mean. By presenting the accuracy results in this way, we can gain insight into both their central tendency and their variability. In [Tables 3, 4](#), data from single channel and four channel EEGs for five-state neonate sleep classification is used. In single-channel five-state classification, the F3, F4, C3, and C4 channels achieve maximum mean accuracy and kappa. For the F3 channel, the accuracy and the kappa values are $80.41 \pm 0.94\%$ and 76%, respectively. For the F4, these values are $80.52 \pm 1.14\%$ and 76%. For the C3 channel, these values are 80.75 ± 0.82 and 76%, respectively. For the C4, these values are 80.40 ± 1.13 and 76%, respectively. There is also evidence to suggest that by combining four left-side channels (F3, C3, P3, and T3), the highest mean accuracy and kappa values can be achieved, with accuracy and kappa values of 82.71 ± 0.88 and 78%, respectively.

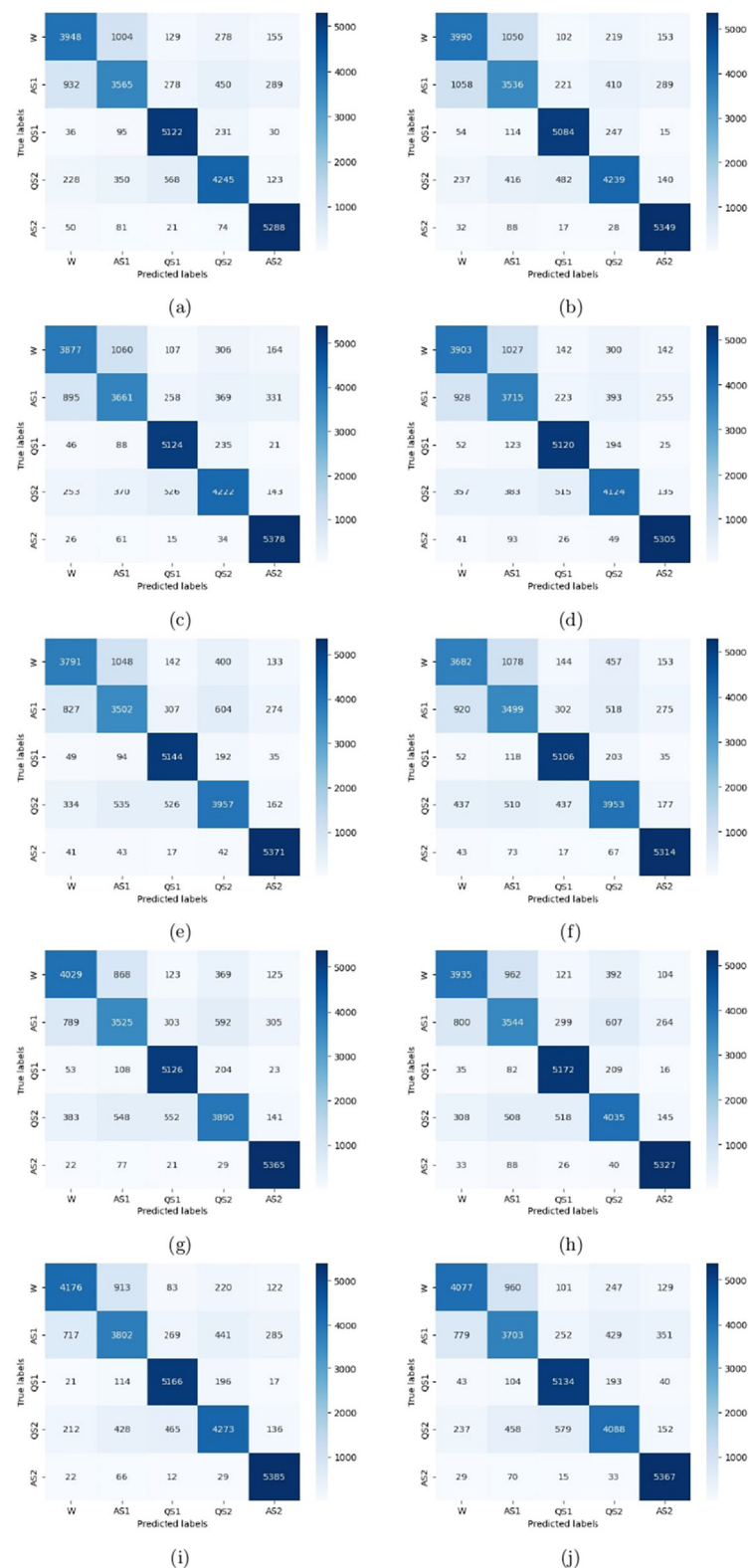


FIGURE 6
Confusion matrices for the channels: (a) F3, (b) F4, (c) C3, (d) C4, (e) P3, (f) P4, (g) T3, (h) T4, (i) Left side, and (j) Right side.

Right side electrode combinations (F4, C4, P4, and T4) have values of 81.14 ± 0.77 and 76%, respectively. In addition, accuracy line curves and confusion matrices for five states are also shown in

Figure 7 in order to visualize the model's performance and learning progress. As shown in the above Table 3, for the classification of the five-state sleep stage of newborns, channels P3, P4, T3, and T4

TABLE 3 Single-channel EEG classification results for five states.

Channel	Acc (%)	Kappa	Rec (%)	Pre (%)	MCC (%)	F1_Sco (%)
F3	80.41 ± 0.94	0.76	80.41 ± 0.94	80.08 ± 1.06	76 ± 0.01	80.01 ± 1.02
F4	80.52 ± 1.14	0.76	80.52 ± 1.14	80.20 ± 1.20	76 ± 0.01	80.24 ± 1.21
C3	80.75 ± 0.82	0.76	80.75 ± 0.82	80.39 ± 0.91	76 ± 0.01	80.41 ± 0.89
C4	80.40 ± 1.13	0.76	80.40 ± 1.12	80.17 ± 1.17	76 ± 0.01	80.15 ± 1.15
P3	78.94 ± 0.72	0.74	78.94 ± 0.72	78.53 ± 0.78	74 ± 0.01	78.54 ± 0.76
P4	78.18 ± 0.58	0.73	78.18 ± 0.58	77.72 ± 0.59	73 ± 0.01	77.83 ± 0.58
T3	79.56 ± 0.58	0.74	79.56 ± 0.57	79.16 ± 0.72	75 ± 0.01	79.18 ± 0.64
T4	79.84 ± 0.67	0.75	79.84 ± 0.67	79.46 ± 0.71	75 ± 0.01	79.51 ± 0.71

TABLE 4 Four-channel EEG classification results for five states.

Channel	Acc (%)	Kappa	Rec (%)	Pre (%)	MCC (%)	F1_Sco (%)
Four-channel (Left)	82.71 ± 0.88	0.78	82.71 ± 0.88	82.47 ± 0.94	78 ± 0.01	82.46 ± 0.92
Four-channel (Right)	81.14 ± 0.77	0.76	81.14 ± 0.77	80.87 ± 0.83	77 ± 0.01	80.83 ± 0.85

Left side channels: F3, C3, P3, T3 & Right side channels: F4, C4, P4, T4.

are far less helpful than channels P3 and P4 in determining the sleep stage. However, F3, F4, C3, and C4 perform well. When there are four channels, left-side channels perform better than right-side channels. Even with fewer channels, performance is still favorable when the parameters relating to performance are compared with those presented in Tables 3, 4. It has been shown that sleep analysis can enhance the care of neonates and enable them to be monitored effectively in order to detect sleep-related abnormalities, such as sleep disorders, early in order to treat them early.

Comparisons of existing and proposed methods are presented in Table 5. This article and Zhu et al. (2023) refer to the same dataset, ensuring consistency and comparability in evaluating the models listed in Table 5. Most of the models in this Table have been evaluated on this dataset by Zhu et al. (2023), and the results obtained are also reflected in that Table. The proposed study uses datasets that are several times larger than those used in Ansari et al. (2020) and Ansari et al. (2018). On the basis of this dataset, these models were found to be underfitting. For adult sleep, Supratak et al. (2017) and Eldele et al. (2021) are presented. Taking into account the difference in sleep patterns between infants and adults, these models are prone to convergence problems and overfitting. Therefore, it is hard to transfer an adult sleep staging model directly to neonate data because this causes convergence problems and overfitting. The model needs to be modified to reflect the neonate's sleep characteristics. A serial recurrent neural network (RNN) is used as part of the TIL module in the model architecture in Zhu et al. (2023), which results in a lengthy training time and inefficient training.

Based on the experiments, limitations and future directions should be identified. Using only EEG signals as inputs to the proposed scheme in this paper is the primary objective of this paper, which is to assess its feasibility and reliability. In this study, electrooculography (EOG), electromyography (EMG), and electrocardiography (ECG) were not used. However, they could be used in the future to assess neonatal sleep with various input signals. Further improvement could be accomplished by

using Transformer (Vaswani et al., 2017) rather than CNN to learn. Additionally, all subjects were randomly divided into a set of training subjects and a set of test subjects in this study. Future research can increase the accuracy of the classification of neonatal sleep stages by incorporating an independent set of subjects in the training and testing phases. As a result, the performance of MFE in the context of sleep staging should be compared to Multiscale Dispersion Entropy and Multiscale Fluctuation Dispersion Entropy. A number of studies have shown that these methods are better at detecting meaningful patterns (Zandbagleh et al., 2023; Chakraborty et al., 2021). In addition to potential overfitting from the Multi-Branch CNN, its limited capacity for hierarchical temporal learning may have made it difficult to capture long-range EEG signal dependencies. Further, its inefficiency in learning sequential patterns and its sensitivity to signal variability could have adversely impacted generalization and contextual understanding. In comparison to 1D CNNs, LSTM models generally perform better when dealing with time series data. LSTM networks, on the other hand, yield more accurate results by retaining long-term dependencies, interpreting context over sequences, and capturing fine-scale changes in EEG data, making them more suitable for effectively identifying five distinct sleep states. With the integration and evaluation of these techniques, future research can enhance sleep staging algorithms.

6 Conclusion

Using an LSTM classifier that takes into account features in the time and frequency domains, this study proposes an efficient and accurate classification of neonatal sleep states based on EEG, using single and multi-channel EEG data. A combination of Detrended Fluctuation Analysis (DFA), Multiscale Fluctuation Entropy, and Lyapunov Exponents is used to analyze the data in this study. PCA is used to select features. With the use of both single-channel as well as multiple-channel EEG data, it achieves favorable and comparable

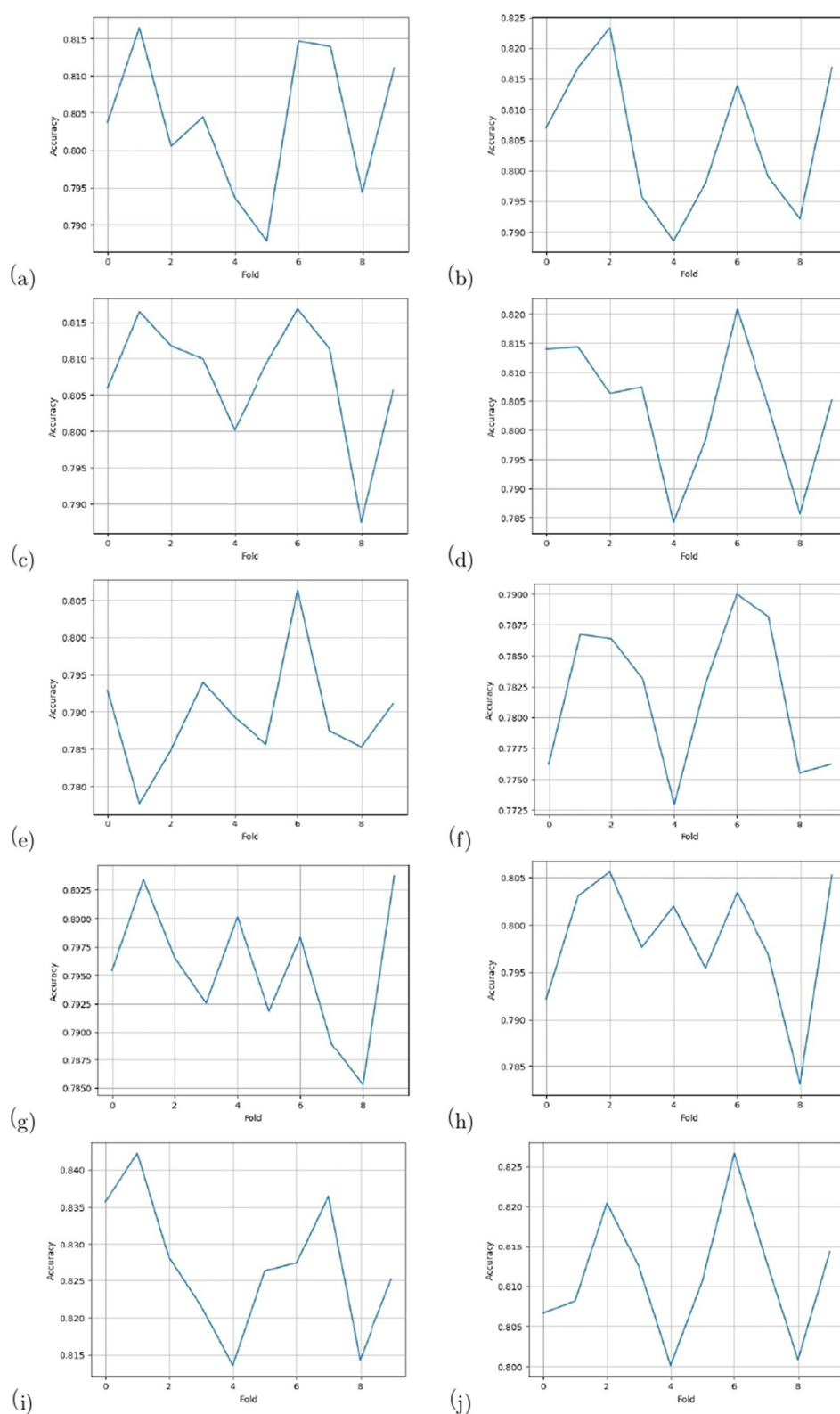


FIGURE 7

Accuracy line graphs for the channels: (a) F3, (b) F4, (c) C3, (d) C4, (e) P3, (f) P4, (g) T3, (h) T4, (i) Left side, and (j) Right side.

results. The number and placement of channels play a critical role in the optimal electrode configuration for the assessment of neonatal sleep stages and the most effective channels in five states.

Using a variety of electrode setups, the purpose of this study was to evaluate the accuracy of sleep stage classification for neonatal sleep studies in order to reduce complexity and cost. The frontal

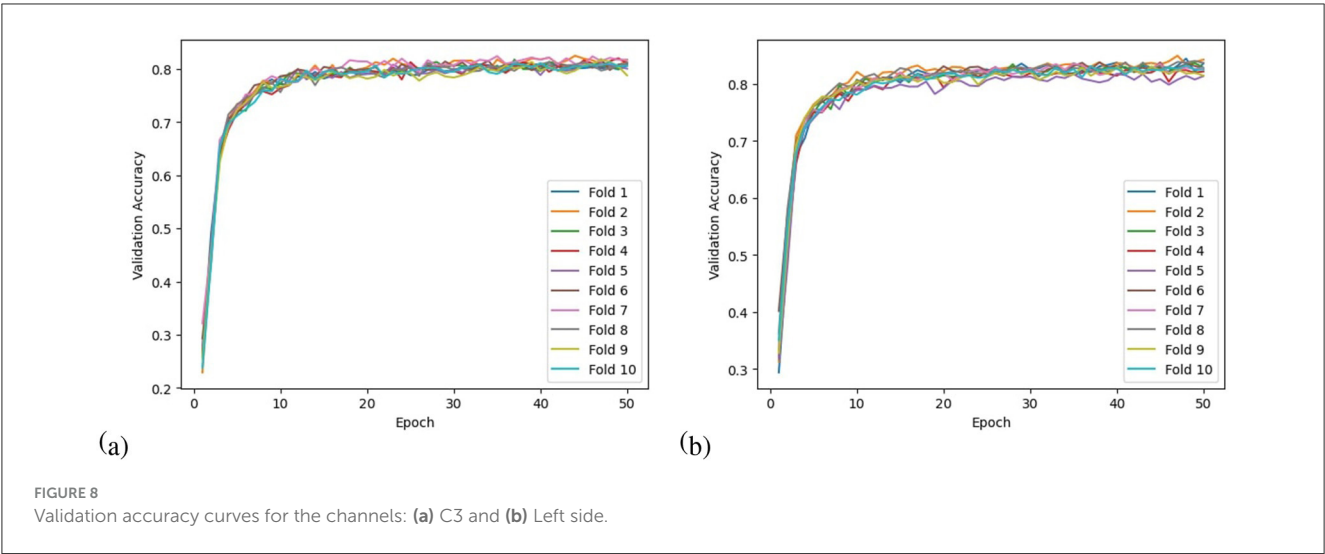


TABLE 5 Comparison of existing and proposed methods.

References	Algorithms	No. of channels	Accuracy	Kappa
Ansari et al. (2020)	Conv-2d	8	52.3%	0.41
Ansari et al. (2018)	Conv-2d	8	53.5%	0.48
Zhu et al. (2023)	MS-HNN	1	75.4%	0.72
Zhu et al. (2023)	MS-CNN	1	69.3%	0.65
Supratak et al. (2017)	DeepSleepNet	2	69.8%	0.64
Eldele et al. (2021)	AttnSleep	1	68.0%	0.65
Siddiqa et al. (2024)	Multi-Branch CNN	1	74.27%	0.61
Siddiqa et al. (2024)	Multi-Branch CNN	4	75.36%	0.63
This study	LSTM	1	80.75% ± 0.82%	0.76
This study	LSTM	4	82.71% ± 0.88%	0.78

and central EEG channels worked better independently or jointly, based on the results. In the future, neonate sleep staging can be simplified, comfort levels can be increased, and data analysis can be sped up by reducing the number of channels. Through sleep analysis, it is possible to detect sleep-related abnormalities, such as sleep disorders, early, allowing for more effective neonate care and monitoring of sleep. Also, the experimental results suggest that the proposed approach captures information effectively within a single channel, reducing computing load by reducing the number of channels, while maintaining good performance. Furthermore, including linear and non-linear features in the time and frequency domains of neonatal sleep staging can improve

accuracy and provide insights into newborn sleep dynamics and irregularities.

Data availability statement

The datasets presented in this article are not readily available because due to privacy concerns. Requests to access the datasets should be directed to hafizaasha@yahoo.com.

Ethics statement

This work has obtained approval from the Research Ethics Committee of Children’s Hospital of Fudan University, with the assigned Approval No. (2017). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants’ legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

HS: Conceptualization, Methodology, Resources, Software, Visualization, Writing – original draft. MQ: Methodology, Software, Writing – review & editing. AK: Visualization, Writing – original draft. YX: Data curation, Resources, Writing – review & editing. LW: Data curation, Resources, Writing – review & editing. SA: Visualization, Writing – review & editing. CC: Formal analysis, Writing – review & editing. WC: Formal analysis, Investigation, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We thank the Research Ethics Committee of Childrens Hospital of Fudan University [Approval No. (2017) 89] and all the subjects involved in this study.

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.

The handling editor SA declared a past co-authorship with the author MQ.

References

- Abbasi, S. F., Ahmad, J., Tahir, A., Awais, M., Chen, C., Irfan, M., et al. (2020). EEG-based neonatal sleep-wake classification using multilayer perceptron neural network. *IEEE Access* 8, 183025–183034. doi: 10.1109/ACCESS.2020.3028182
- Abbasi, S. F., Jamil, H., and Chen, W. (2022). EEG-based neonatal sleep stage classification using ensemble learning. *Comput. Mater. Contin* 70, 4619–4633. doi: 10.32604/cmc.2022.020318
- Ali, A., Shaukat, S., Tayyab, M., Khan, M. A., Khan, J. S., Ahmad, J., et al. (2020). “Network intrusion detection leveraging machine learning and feature selection,” in *2020 IEEE 17th International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET)* (IEEE), 49–53. doi: 10.1109/HONET50430.2020.9322813
- Ansari, A. H., De Wel, O., Lavanga, M., Caicedo, A., Dereymaeker, A., Jansen, K., et al. (2018). Quiet sleep detection in preterm infants using deep convolutional neural networks. *J. Neural Eng.* 15:066006. doi: 10.1088/1741-2552/aadc1f
- Ansari, A. H., De Wel, O., Pillay, K., Dereymaeker, A., Jansen, K., Van Huffel, S., et al. (2020). A convolutional neural network outperforming state-of-the-art sleep staging algorithms for both preterm and term infants. *J. Neural Eng.* 17:016028. doi: 10.1088/1741-2552/ab5469
- Ansari, A. H., Pillay, K., Dereymaeker, A., Jansen, K., Van Huffel, S., Naulaers, G., et al. (2021). A deep shared multi-scale inception network enables accurate neonatal quiet sleep detection with limited eeg channels. *IEEE J. Biomed. Health Inform.* 26, 1023–1033. doi: 10.1109/JBHI.2021.3101117
- Arasteh, E., de Groot, E. R., Van den Ende, D., Alderliesten, T., Long, X., de Goederen, R., et al. (2023). Unobtrusive cot side sleep stage classification in preterm infants using ultra-wideband radar. *Front. Sleep* 2:1150962. doi: 10.3389/fnrsle.2023.1150962
- Arif, S., Khan, M. J., Naseer, N., Hong, K.-S., Sajid, H., and Ayaz, Y. (2021). Vector phase analysis approach for sleep stage classification: a functional near-infrared spectroscopy-based passive brain-computer interface. *Front. Hum. Neurosci.* 15:658444. doi: 10.3389/fnhum.2021.658444
- Arif, S., Munawar, S., and Ali, H. (2023). Driving drowsiness detection using spectral signatures of EEG-based neurophysiology. *Front. Physiol.* 14:1153268. doi: 10.3389/fphys.2023.1153268
- Awais, M., Long, X., Yin, B., Abbasi, S. F., Akbarzadeh, S., Lu, C., et al. (2021). A hybrid DCNN-SVM model for classifying neonatal sleep and wake states based on facial expressions in video. *IEEE J. Biomed. Health Inform.* 25, 1441–1449. doi: 10.1109/JBHI.2021.3073632
- Awais, M., Long, X., Yin, B., Chen, C., Akbarzadeh, S., Abbasi, S. F., et al. (2020). Can pre-trained convolutional neural networks be directly used as a feature extractor for video-based neonatal sleep and wake classification? *BMC Res. Notes* 13, 1–6. doi: 10.1186/s13104-020-05343-4
- Baker, T. L. (1985). Introduction to sleep and sleep disorders. *Med. Clin. North Am.* 69, 1123–1152. doi: 10.1016/S0025-7125(16)30979-8
- Bing, P., Liu, Y., Liu, W., Zhou, J., and Zhu, L. (2022). Electrocardiogram classification using TSST-based spectrogram and convit. *Front. Cardio. Med.* 9:983543. doi: 10.3389/fcvm.2022.983543
- Cao, T., Lian, Z., Du, H., Shen, J., Fan, Y., and Lyu, J. (2023). A sleep staging model for the sleep environment control based on machine learning. *Build. Simul.* 16, 1409–1423. doi: 10.1007/s12273-023-1049-6
- Chakraborty, M., Mitra, D., et al. (2021). Automated detection of epileptic seizures using multiscale and refined composite multiscale dispersion entropy. *Chaos, Solitons Fractals* 146:110939. doi: 10.1016/j.chaos.2021.110939
- Chen, B., and Zhu, X. (2024). Impacts of different noise environments on polysomnographic monitoring parameters and sleep quality among hospitalized patients. *Noise Health* 26, 114–119. doi: 10.4103/nah.nah_12_24
- Chicco, D., Warrens, M. J., and Jurman, G. (2021). The matthews correlation coefficient (mcc) is more informative than cohen's kappa and brier score in binary classification assessment. *IEEE Access* 9, 78368–78381. doi: 10.1109/ACCESS.2021.3084050
- De Wel, O., Lavanga, M., Dorado, A. C., Jansen, K., Dereymaeker, A., Naulaers, G., et al. (2017). Complexity analysis of neonatal eeg using multiscale entropy: applications in brain maturation and sleep stage classification. *Entropy* 19:516. doi: 10.3390/e19100516
- Dereymaeker, A., Pillay, K., Vervisch, J., Van Huffel, S., Naulaers, G., Jansen, K., et al. (2017). An automated quiet sleep detection approach in preterm infants as a gateway to assess brain maturation. *Int. J. Neural Syst.* 27:1750023. doi: 10.1142/S012906571750023X
- Eldele, E., Chen, Z., Liu, C., Wu, M., Kwok, C.-K., Li, X., et al. (2021). An attention-based deep learning approach for sleep stage classification with single-channel eeg. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 809–818. doi: 10.1109/TNSRE.2021.3076234
- Fernández, A., Garcia, S., Herrera, F., and Chawla, N. V. (2018). Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *J. Artif. Intell. Res.* 61, 863–905. doi: 10.1613/jair.1.11192
- Fonseca, P., Den Teuling, N., Long, X., and Aarts, R. M. (2016). Cardiorespiratory sleep stage detection using conditional random fields. *IEEE J. Biomed. Health Inform.* 21, 956–966. doi: 10.1109/JBHI.2016.2550104
- Fraiwan, L., and Lweesy, K. (2017). “Neonatal sleep state identification using deep learning autoencoders,” in *2017 IEEE 13th International Colloquium on Signal Processing its Applications (CSPA)* (IEEE), 228–231. doi: 10.1109/CSPA.2017.8064956
- Gamel, S. A., Ghoneim, S. S., and Sultan, Y. A. (2024). Improving the accuracy of diagnostic predictions for power transformers by employing a hybrid approach combining smote and dnn. *Comput. Electr. Eng.* 117:109232. doi: 10.1016/j.compeleceng.2024.109232
- Gosala, B., Kapgate, P. D., Jain, P., Chaurasia, R. N., and Gupta, M. (2023). Wavelet transforms for feature engineering in eeg data processing: an application on schizophrenia. *Biomed. Signal Process. Control* 85:104811. doi: 10.1016/j.bspc.2023.104811
- Gudmundsson, S., Runarsson, T. P., and Sigurdsson, S. (2005). “Automatic sleep staging using support vector machines with posterior probability estimates,” in *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)* (IEEE), 366–372. doi: 10.1109/CIMCA.2005.1631496
- Ilyas, M., Ahmad, J., Lawson, A., Khan, J. S., Tahir, A., Adeel, A., et al. (2020). “Height prediction for growth hormone deficiency treatment planning using deep learning,” in *Advances in Brain Inspired Cognitive Systems: 10th International Conference, BICS 2019, Guangzhou, China, July 13–14, 2019, Proceedings 10* (Springer), 76–85. doi: 10.1007/978-3-030-39431-8_8

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

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.

- Khan, J. S., Tahir, A., Ahmad, J., Shah, S. A., Abbasi, Q. H., Russell, G., et al. (2020). "5G-fog: Freezing of gait identification in multi-class softmax neural network exploiting 5G spectrum," in *Science and Information Conference* (Springer), 26–36. doi: 10.1007/978-3-030-52243-8_3
- Khan, S., Malik, B. H., Gupta, D., Rutkofsky, I., and Rutkofsky, I. H. (2020). The role of circadian misalignment due to insomnia, lack of sleep, and shift work in increasing the risk of cardiac diseases: a systematic review. *Cureus* 12:e6616. doi: 10.7759/cureus.6616
- Killick, R., Stranks, L., and Hoyos, C. M. (2022). Sleep deficiency and cardiometabolic disease. *Clin. Chest Med.* 43, 319–336. doi: 10.1016/j.ccm.2022.02.011
- Koolen, N., Oberdorfer, L., Rona, Z., Giordano, V., Werther, T., Klebermass-Schrehof, K., et al. (2017). Automated classification of neonatal sleep states using eeg. *Clin. Neurophysiol.* 128, 1100–1108. doi: 10.1016/j.clinph.2017.02.025
- Lajnef, T., Chaibi, S., Ruby, P., Aguera, P.-E., Eichenlaub, J.-B., Samet, M., et al. (2015). Learning machines and sleeping brains: automatic sleep stage classification using decision-tree multi-class support vector machines. *J. Neurosci. Methods* 250, 94–105. doi: 10.1016/j.jneumeth.2015.01.022
- Lal, U., Mathavu Vasanthana, S., and Hoblidar, A. (2023). Temporal feature extraction and machine learning for classification of sleep stages using telemetry polysomnography. *Brain Sci.* 13:1201. doi: 10.3390/brainsci13081201
- Lee, W. H., Kim, S. H., Na, J. Y., Lim, Y.-H., Cho, S. H., Cho, S. H., et al. (2021). Non-contact sleep/wake monitoring using impulse-radio ultrawideband radar in neonates. *Front. Pediatr.* 9:782623. doi: 10.3389/fped.2021.782623
- Li, K., Wang, Z., Zhou, Y., and Li, S. (2024). Lung adenocarcinoma identification based on hybrid feature selections and attentional convolutional neural networks. *Mathem. Biosci. Eng.* 21, 2991–3015. doi: 10.3934/mbe.2024133
- Loomis, A. L., Harvey, E. N., and Hobart, G. A. (1937). Cerebral states during sleep, as studied by human brain potentials. *J. Exp. Psychol.* 21:127. doi: 10.1037/h0057431
- Newson, J. J. (2017). *The EEG from infancy to adulthood* (Tech. Rep). Washington, DC: Sapien Labs.
- Pan, H., Li, Z., Fu, Y., Qin, X., and Hu, J. (2024). Reconstructing visual stimulus representation from EEG signals based on deep visual representation model. *IEEE Trans. Hum. Mach. Syst.* 54, 711–722. doi: 10.1109/THMS.2024.3407875
- Parish, J. M. (2009). Sleep-related problems in common medical conditions. *Chest* 135, 563–572. doi: 10.1378/chest.08-0934
- Pillay, K., Dereymaeker, A., Jansen, K., Naulaers, G., Van Huffel, S., and De Vos, M. (2018). Automated eeg sleep staging in the term-age baby using a generative modelling approach. *J. Neural Eng.* 15:036004. doi: 10.1088/1741-2552/aaab73
- Shaukat, S., Ali, A., Batool, A., Alqahtani, F., Khan, J. S., Ahmad, J., et al. (2020). Intrusion detection and attack classification leveraging machine learning technique," in *2020 14th International Conference on Innovations in Information Technology (IIT)* (IEEE), 198–202. doi: 10.1109/IIT50501.2020.9299093
- Siddiqua, H. A., Irfan, M., Abbasi, S. F., and Chen, W. (2023). Electroencephalography (EEG) based neonatal sleep staging and detection using various classification algorithms. *Comput. Mater. Continua* 77, 1759–1778. doi: 10.32604/cmc.2023.041970
- Siddiqua, H. A., Tang, Z., Xu, Y., Wang, L., Irfan, M., Abbasi, S. F., et al. (2024). Single-channel eeg data analysis using a multi-branch cnn for neonatal sleep staging. *IEEE Access* 12, 29910–29925. doi: 10.1109/ACCESS.2024.3365570
- Song, Z., Jiang, Z., Zhang, Z., Wang, Y., Chen, Y., Tang, X., et al. (2024). Evolving brain network dynamics in early childhood: insights from modular graph metrics. *Neuroimage* 297:120740. doi: 10.1016/j.neuroimage.2024.120740
- Supratak, A., Dong, H., Wu, C., and Guo, Y. (2017). Deepsleepnet: a model for automatic sleep stage scoring based on raw single-channel EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1998–2008. doi: 10.1109/TNSRE.2017.2721116
- Turnbull, J., Loparo, K., Johnson, M., and Scher, M. (2001). Automated detection of tracé alternant during sleep in healthy full-term neonates using discrete wavelet transform. *Clin. Neurophysiol.* 112, 1893–1900. doi: 10.1016/S1388-2457(01)00641-1
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). "Attention is all you need," in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA, 30.
- Wan, W., Gao, Z., Zhang, Q., Gu, Z., Chang, C., Peng, C.-K., et al. (2023). Resting state eeg complexity as a predictor of cognitive performance. *Phys. Stat. Mech. Applic.* 624:128952. doi: 10.1016/j.physa.2023.128952
- Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometr. Intell. Labor. Syst.* 2, 37–52. doi: 10.1016/0169-7439(87)80084-9
- Xiao, M., Yan, H., Song, J., Yang, Y., and Yang, X. (2013). Sleep stages classification based on heart rate variability and random forest. *Biomed. Signal Process. Control* 8, 624–633. doi: 10.1016/j.bspc.2013.06.001
- Yu, R., Zhou, Z., Wu, S., Gao, X., and Bin, G. (2022). Mrasleepnet: a multi-resolution attention network for sleep stage classification using single-channel EEG. *J. Neural Eng.* 19:066025. doi: 10.1088/1741-2552/aca2de
- Zandbagleh, A., Azami, H., Mirzakuchaki, S., Daliri, M. R., Sanei, S., and Premkumar, P. (2023). Multiscale fluctuation dispersion entropy of EEG as a physiological biomarker of schizotypy. *IEEE Access* 11, 110124–110135. doi: 10.1109/ACCESS.2023.3321868
- Zhu, H., Xu, Y., Shen, N., Wu, Y., Wang, L., Chen, C., et al. (2023). MS-HNN: multi-scale hierarchical neural network with squeeze and excitation block for neonatal sleep staging using a single-channel EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* 31, 2195–2204. doi: 10.1109/TNSRE.2023.3266876



OPEN ACCESS

EDITED BY

Muhammad Zia Ur Rehman,
Aalborg University, Denmark

REVIEWED BY

Muhammad Farrukh Qureshi,
Namal College, Pakistan
Saadullah Farooq Abbasi,
University of Birmingham, United Kingdom

*CORRESPONDENCE

Zhe Li,
✉ kenan_zhentanc@yeah.net

RECEIVED 20 December 2024

ACCEPTED 24 March 2025

PUBLISHED 07 April 2025

CITATION

Li Z (2025) Mamba with split-based pyramidal convolution and Kolmogorov-Arnold network-channel-spatial attention for electroencephalogram classification.
Front. Sens. 6:1548729.
doi: 10.3389/fsens.2025.1548729

COPYRIGHT

© 2025 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.

Mamba with split-based pyramidal convolution and Kolmogorov-Arnold network-channel-spatial attention for electroencephalogram classification

Zhe Li^{1,2*}

¹Hankou University, Jiangxiazhen, China, ²Wuhan University of Engineering Science, Wuhan, China

Deep learning is widely used in brain electrical signal studies, among which the brain-computer interface is an important direction. Deep learning can effectively improve the performance of BCI machines, which is of great medical and commercial value. This paper introduces an efficient deep learning model for classifying brain electrical signals based on a Mamba structure enhanced with split-based pyramidal convolution (PySPConv) and Kolmogorov-Arnold network (KAN)-channel-spatial attention (KSA) mechanisms. Incorporating KANs into the attention module of the proposed KSA-Mamba-PySPConv model better approximates the sample function while obtaining local network features. PySPConv, on the other hand, swiftly and efficiently extracts multi-scale fusion features from input data. This integration allows the model to reinforce feature extraction at each layer in Mamba's structure. The model achieves a 96.76% accuracy on the eegmmidb dataset and demonstrates state-of-the-art performance across metrics such as the F1 score, precision, and recall. KSA-Mamba-PySPConv promises to be an effective tool in electroencephalogram classification in brain-computer interface systems.

KEYWORDS

mamba, Kolmogorov-Arnold network, electroencephalogram, deep learning, BCI

1 Introduction

Electroencephalogram (EEG) has been a hotspot for medical, computer science, and commercial research, and it is often used to diagnose brain diseases and to study human mental activities. Brain-computer interfacing is a promising technology for scientists and engineers, which converts human EEG signals into programs that machines can recognize. Deep learning is an important tool in brain electricity studies and has been employed by many scholars, as exemplified below.

The DeepConvnet model, employed by Schirrmeister et al., achieved a classification accuracy of 76.7% on the PhysioNet EEG Motor Movement/Imagery Database (eegmmidb). The model utilized multiple layers of conventional convolutional neural network (CNN)

convolutions to construct deep learning networks, achieving 63.7% and 83.2% accuracy on the BCI-2a and BCI-2b datasets, respectively (Schirrmester et al., 2017). The EEGnet model, adopted by Lawhern et al., achieved an accuracy of 79.3% on the eegmmidb dataset (Lawhern et al., 2018). This model utilized depth-separable convolutions to construct multi-layer deep neural networks, thereby facilitating the segregation of channels and regions while reducing the parameter count. Jia et al. developed a multi-branch multi-scale CNN (MMCNN) (Jia et al., 2021) that decoded the original EEG signal without filtering or other pre-processing techniques. It also successfully characterized information in various frequency bands and thus determined the optimal convolution scale. Roots et al. introduced a multi-branch two-dimensional (2D) CNN that employed distinct hyperparameter values for each branch, resulting in accuracies of 84.1% and 83.8% when applied to the eegmmidb dataset for performing and imagining motor actions, respectively (Roots et al., 2020). Chowdhury et al. developed an EEGNet Fusion V2 model that enhanced the extracted features via diverse filters, which yielded a spectrum of features. Subsequently, these features were integrated into the fusion layer to generate more intricate features (Chowdhury et al., 2023). To identify spectral features and improve the decoding of motor imagery electroencephalogram (MI-EEG), Li et al. employed a novel time-spectrum squeezed-excitation feature fusion network with multi-stage wavelet convolutions in parallel for multi-spectral convolution block capture (Li et al., 2021). Hou et al. combined bidirectional long and short-term memory (BiLSTM), attention mechanisms, and a graph convolutional neural network (GCN) to enhance the decoding performance. This was achieved by leveraging the feature topology estimated from the comprehensive data set to accurately identify the human body's intention to move from the raw EEG signals (Hou et al., 2022). Steady-state visual evoked potential (SSVEP) represents one of the most frequently utilized control signals in brain-computer interface systems. In an interdisciplinary classification scenario, Chen et al. proposed an SSVEP classification model based on the highly effective deep learning Transformer structure, which fully exploited harmonic information and established a methodology based on filter bank technology (Chen et al., 2022). Luo et al. employed a shallow mirror Transformer comprising a multi-head self-attentive layer with a global receptive field to detect and utilize discriminative segments across input EEG trials. They also constructed mirror EEG signals and mirror network structures based on integrated learning to improve classification accuracy (Luo et al., 2023). Keutayeva and Abibullaev developed a hybrid model that fused a CNN with a visual Transformer for decoding motion image EEG signals. The CNN was employed to extract local features, whereas the Transformer was utilized to perceive global dependencies. The model demonstrated 80.44% and 74.73% accuracy on the BCI-2a and BCI-2b datasets, respectively, which represented a significant improvement over previous models (Keutayeva and Abibullaev, 2023).

In recent years, there has been a notable increase in the popularity of Kolmogorov-Arnold networks (KANs) as an alternative to the multi-layer perceptron (MLP) (Vaca-Rubio et al., 2024). KANs utilize the Kolmogorov-Arnold representation

theorem, which enables the activation functions of a neural network to be executed on edges. This facilitates the “learning” of the activation functions and enhances the model performance. KANs lack linear weights; each weight parameter is replaced by a univariate function parameterized as a spline. Smaller KANs can be visualized intuitively and achieve comparable or superior accuracies in data fitting and partial differential equation (PDE) solutions compared to larger MLPs.

The Mamba model addresses the limited efficiency of Transformers in long sequence processing by combining linear layers, gating, and selective structured state space models (Gu et al., 2023); its core is a selectivity mechanism that efficiently compresses and filters contextual information. The hardware algorithm significantly improves computational speed by scanning rather than convolving.

However, the mamba model is very limited to handle local features, and the obtained features have a large redundancy. Therefore, we can use some new methods to improve the local feature extraction ability of mamba, using the attention module to screen the reinforcement main features. In light of the studies above, we propose a novel deep learning model integrating a Mamba backbone splicing a split-based pyramidal convolution (PySPConv) module and a KAN-channel-spatial attention (KSA) mechanism. The model is designated as KSA-Mamba-PySPConv, and its objective is to leverage the Mamba and KAN architectures to enhance model classification capabilities and reduce resource costs.

It includes the literature review, model methodology, experimental design, discussion of experiments, and conclusion. The primary contributions of this paper are as follows:

- 1) We propose the novel KSA mechanism, which incorporates a KAN network into the attention mechanism. This integration aims to enhance the module's feature extraction capabilities, leveraging the fitting approximation capacity of KANs.
- 2) We employ the novel PySPConv scheme to replace the standard convolutions in Mamba, aiming to address the limitations of Mamba's local feature extraction capability while minimizing computational overhead.
- 3) We conduct experiments using the proposed KSA-Mamba-PySPConv model and multiple existing models on eegmmidb to compare their accuracy, F1 score, and recall.

2 Related work

2.1 KANs

MLPs use a multi-layer linear function plus a nonlinear activation function to model and approximate the input-output relationships of a sample, which consists of a large trainable data matrix, as expressed by $f(\mathbf{x}) \approx \sum_{i=1}^{N(c)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$. In this formula, $\mathbf{x}$ is the input data, $\mathbf{w}_i$ is the weight of $\mathbf{x}$, b_i is the bias, σ is activation function, a_i is the ratio of Scale coefficient. MLPs consume a lot of memory and computational resources in complex tasks and are prone to overfitting. In recent studies, KANs have been shown

to outperform MLPs in accuracy and interpretability, with smaller KANs achieving accuracy comparable to or better than larger MLPs in data fitting and PDE solving. Besides, KANs can be represented intuitively as the summation of multiple spline functions and, therefore, have stronger interpretability. The functional relationship of KANs can be expressed by

$$f(\mathbf{x}) = \sum_{q=1}^Q \Phi_q \left(\sum_{p=1}^P \phi_{q,p}(x_p) \right).$$

Here, $f(\mathbf{x})$ is the multivariate function to be represented; Φ_q denotes a combinatorial function that can be learned and is typically used at higher network levels; $\phi_{q,p}$ is the learnable unitary function corresponding to an activation function on the network edges, generally parameterized as a spline function; x_p is the p th component of the input vector; Q and P are the number of combinatorial and unitary functions, respectively.

KANs can normally achieve comparable or better performance than wider MLPs with fewer parameters. However, KANs have more parameters than MLPs for the same depth and width. The training process of KANs is much more complex than that of conventional neural networks, and its training speed is 10 times slower than that of MLPs. In practice, the resources consumed by KANs are huge and difficult to implement in high-dimensional spaces. Therefore, leveraging the advantages of KANs while compensating for their shortcomings is a question worth studying.

2.2 Mamba and transformers

Transformers capture global features more efficiently than CNNs, though with n^2 computational complexity, n is the length of sequence. In contrast, Mamba exhibits linear complexity and can address the memory consumption issue of Transformers when processing long sequences. Moreover, Mamba adopts hardware-aware parallel algorithms to optimize graphics processing unit (GPU) memory usage and improve the design of the state space model (SSM) architecture, which achieves higher efficiency. Mamba also performs selective processing of input information, which means it can focus on specific information in the input sequence. As a result, Mamba is five times faster than Transformers in inference (predicting or generating texts), and its performance can match that of a Transformer twice its size in certain areas. However, Mamba uses complex S6 and MLP components, making the model complicated and less interpretable. Moreover, it is weak for local feature extraction of sequences. Improving the local feature extraction capability and simplifying the complexity of Mamba are research areas worthy of investigation.

2.3 Pyramidal convolution and split-based convolution

Pyramidal Convolution (PyConv) utilizes a pyramid structure with different kernel sizes and depths to capture details on various

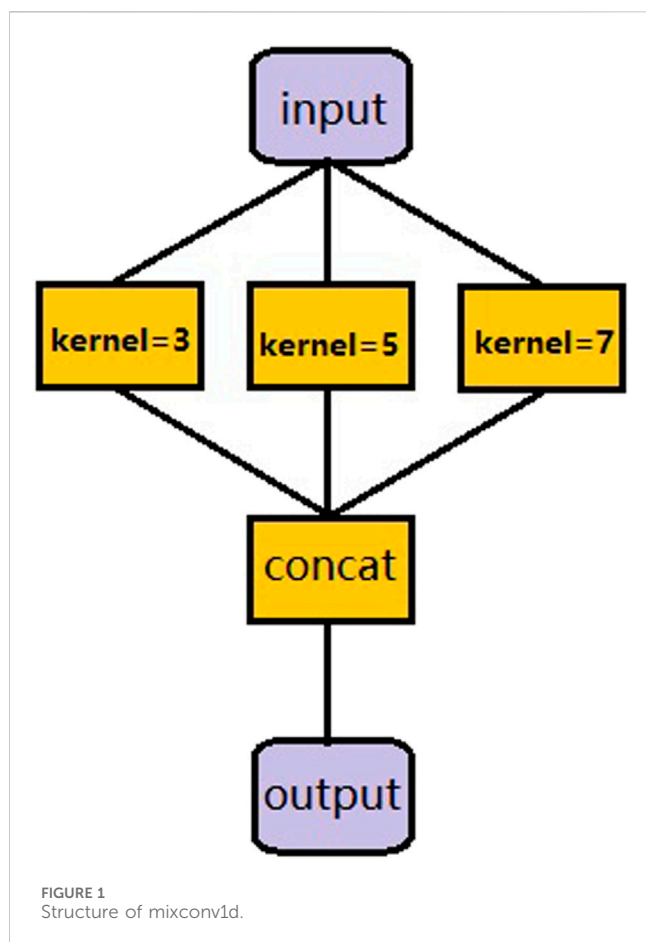
levels (Duta et al., 2021). The PyConv architecture has multiple levels of kernels, gradually increasing kernel size from the bottom (level 1) to the top (level n) while reducing depth. This approach aims to capture diverse scale information at different layers. In PyConv, the different kernel types complement each other to enhance the network's recognition capabilities. Smaller kernels excel at focusing on fine details, capturing information about small objects or specific regions; larger kernels gather a more robust overview of larger objects or contextual information. The PyConv architecture exhibits parameter and computational resource requirements comparable to conventional convolution while benefiting from its ability to leverage multi-threaded parallel processing. This configuration makes PyConv exceptionally efficient. PyConv's 50-layer network outperformed a baseline ResNet with 152 layers in recognition performance while reducing the number of parameters by 2.39 times, computational complexity by 2.52 times, and layer count by over three times (Duta et al., 2021).

Split-based convolution (SPConv) splits the input feature map into a representative part and an uncertain redundant part (Zhang et al., 2020). The representative part is processed with relatively heavy computation to extract intrinsic information, while the uncertain redundant part uses lightweight operations to handle tiny details. The SPConv architecture employs a $k = 3$ convolutional layer to extract essential information and a lightweight $k = 1$ convolutional layer to supplement fine-grained hidden details. The final step of the process involves merging the extracted features from the two parts using a parameter-free feature fusion module. Therefore, SPConv consistently outperforms baselines in accuracy and inference time while showing significant reductions in floating-point operations per second (FLOPs) and parameter counts. Experiments on Cifar10, ImageNet, and Microsoft Common Objects in Context (MS-COCO) datasets demonstrated that networks using SPConv achieved state-of-the-art (SOTA) performance in accuracy and inference speed at the GPU level. The parameter count for SPConv could also be reduced by 2.8 times while maintaining superior performance and inference speed (Zhang et al., 2020).

3 Methodology

3.1 KSA-seq attention

EEG involves multichannel one-dimensional (1D) data, for which the features of individual channel waveforms and the potential relationships between different channels must be considered. KAN has a stronger fitting ability than MLP, and it can obtain the waveform features of a single channel better. On the other hand, the attention mechanism of the lateral axis captures the feature relationships between different channels. We fuse the attention mechanisms in both directions to obtain a more comprehensive EEG feature relationship. We call this attention mechanism KSA, and the algorithmic steps for realizing the KSA attention mechanism are as follows in Algorithm 1:



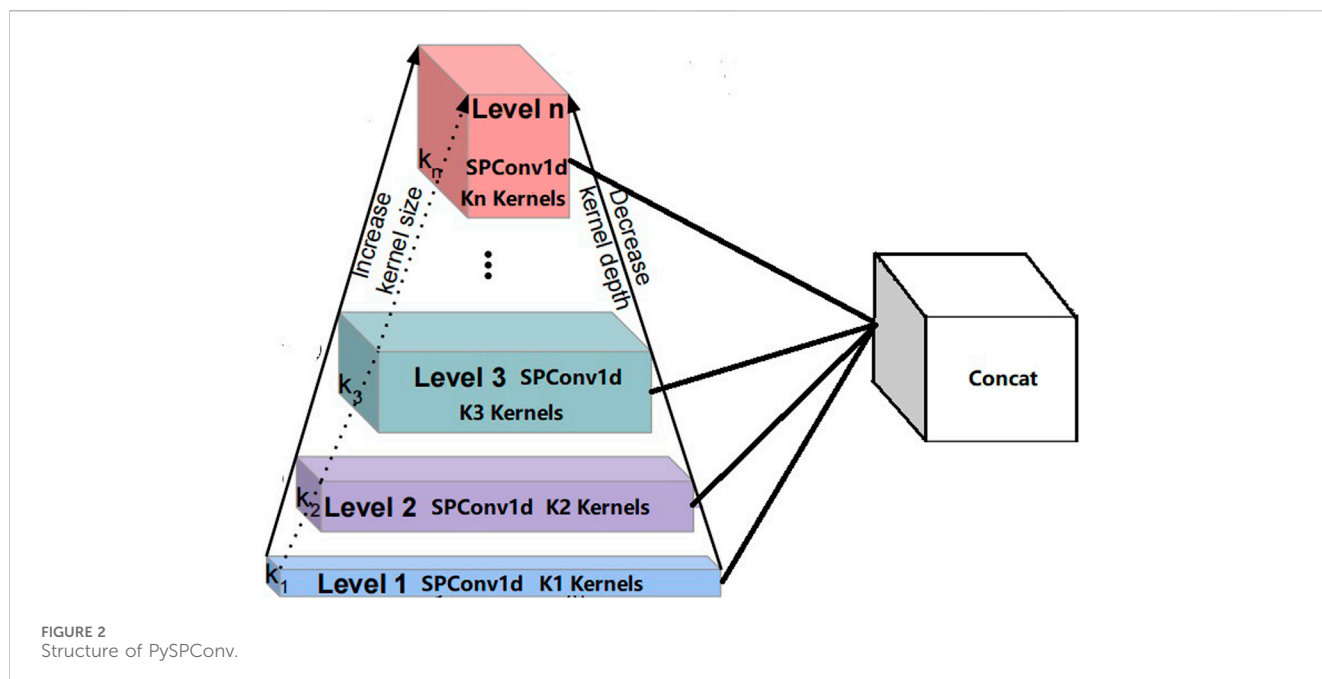
Input: $x: (B, C, L)$
 Output: $y: (B, C, L)$

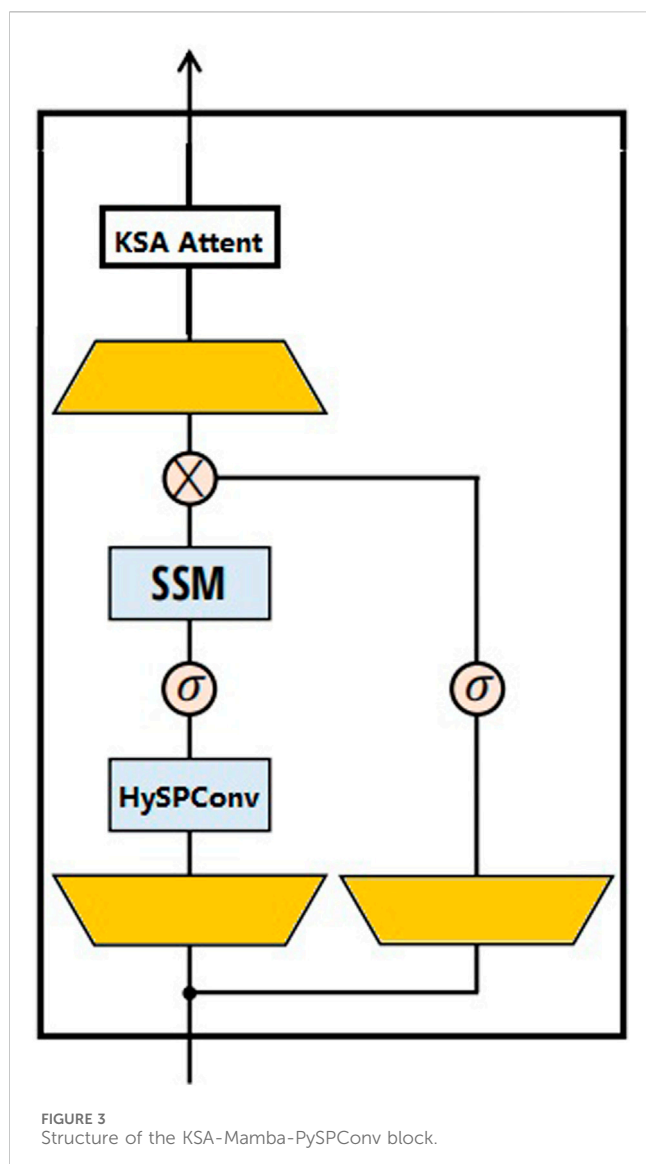
1. $r1: (B, C, 1) \leftarrow \text{AdaptiveAvgPool1d}(B, C, L)$
2. $r2: (B, C, 1) \leftarrow \text{AdaptiveMaxPool1d}(B, C, L)$
3. $r1': (B, C, 1) \leftarrow \text{KAN_Expand}(\text{KAN_Compress}(r1'))$
4. $r2': (B, C, 1) \leftarrow \text{KAN_Expand}(\text{KAN_Compress}(r2'))$
4. $m: (B, C, L) \leftarrow (r1' + r2') * x$
 ▷ KAN-seq attention is done
1. $m: (B, C, L)$
2. $n1: (B, 1, L) \leftarrow \text{Mean}(m, \text{dim} = 1)$
3. $n2: (B, 1, L) \leftarrow \text{Max}(m, \text{dim} = 1)$
4. $n: (B, 2, C) \leftarrow \text{Concat}(n1, n2, \text{dim} = 1)$
4. $\text{SpatialAtt}: (B, 1, C) \leftarrow \text{conv1d}(B, 2, C, \text{kernel} = 7)$
5. $\text{out}: (B, C, L) \leftarrow \text{SpatialAtt} * m$
 ▷ SP1D-seq attention is done

Return out

Algorithm 1. KAN-seq attention.

The algorithm describes the implementation of the attention mechanism and the main input and output data. KAN_Compress represents using the KAN-MLP architecture to compress channels, which is achieved by setting the number of output channels C_{out} of the KAN network to be $1/r$ of the number of input channels C_{in} (where r is the scaling ratio). KAN_Expand represents using the KAN-MLP architecture to expand channels, which is accomplished by configuring the number of output channels C_{out} to expand to r times the number of input channels C_{in} . We obtain channel-axis





attention by performing channel transformation with KAN-seq attention. Then, we use the 1D spatial attention algorithm to calculate the spatial sequence features, which can obtain the y-axis attention. By combining the operations of both sections, we obtain multi-dimensional attention for the input sequence data. We call this module KSA-seq attention.

3.2 PySPConv

Mamba is a novel selective structured SSM that can efficiently deal with long sequential data while maintaining linear time complexity. In Mamba modules, a conventional CNN is used to extract local features. However, CNN-extracted features are limited by the size of kernels and the number of layers, which means that CNN suffers from a lack of flexibility and restricted abilities. Therefore, a more capable local feature extractor needs to be used to obtain better local features.

We design a mixed convolutional model that uses different-sized CNN kernels to extract features of varying receptive fields. We utilize $k = 3$, $k = 5$, and $k = 7$ convolution operations to obtain features under different receptive fields. These features are then combined to form the final output. This module addresses the limitation of narrow receptive fields in conventional CNNs and enhances the quality of extracted features by incorporating multiple convolution sizes. We name this module “mixconv1d.” Its structure is represented in Figure 1, which consists of convolutions with $k = 3$, $k = 5$, and $k = 7$, as well as a feature concatenation compression module.

To obtain multi-scale features, we leverage the pyramid convolutional technique and modify it to enhance its performance and speed while minimizing computational complexity. Specifically, we replace the conventional convolution in the pyramid structure with a lightweight and efficient separable convolution. Here, we employ SPConv, which splits the features into representative and redundant parts, using $k = N$ and $k = 1$ convolutions, respectively. This approach captures the main features in the representative part and details features in the redundant part as supplementary information. This design makes the convolutions efficient and lightweight. Replacing the conventional convolutions in a pyramid convolution with SPConv, which uses different kernel sizes, can reduce computational burdens and enable efficient feature extraction at different scales. This novel convolution module is called PySPConv, whose architecture is shown in Figure 2. It is composed of one-dimensional SPConv convolutions with a kernel size of K_n and a feature concatenation module. We can independently configure the kernel sizes and the number of feature layers, which makes this convolutional module highly flexible to fit our needs.

In subsequent experiments, we will compare the performance of the two convolutional modules. We will test and evaluate various metrics to confirm the advantages of PySPConv in computational load and performance.

3.3 Mamba-PySPConv with KSA attention

We include this new PySPConv module in the Mamba structure and employ the KSA attention mechanism between each layer of Mamba blocks, which enables the model to filter out the important features in the input and improves the expressive power of the model. We also add the residual structure between different layers of Mamba so that the features obtained from shallow Mamba blocks can be fused into deeper Mamba blocks, which improves the model convergence, enriches the extracted features, and fully utilizes the features obtained from each block layer. This final model is called KSA-Mamba-PySPConv, and its block structure is shown in Figure 3.

4 Experiments and results

To evaluate the performance of the aforementioned models, we train and evaluate them on the eegmmidb dataset.

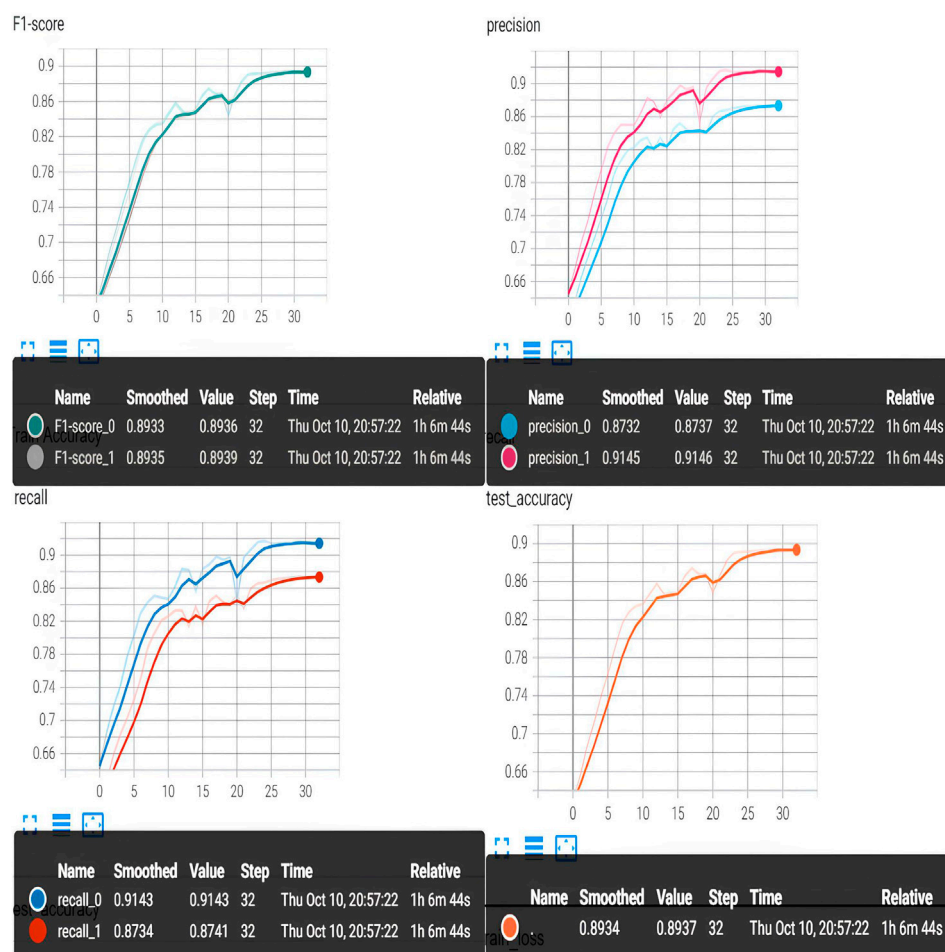


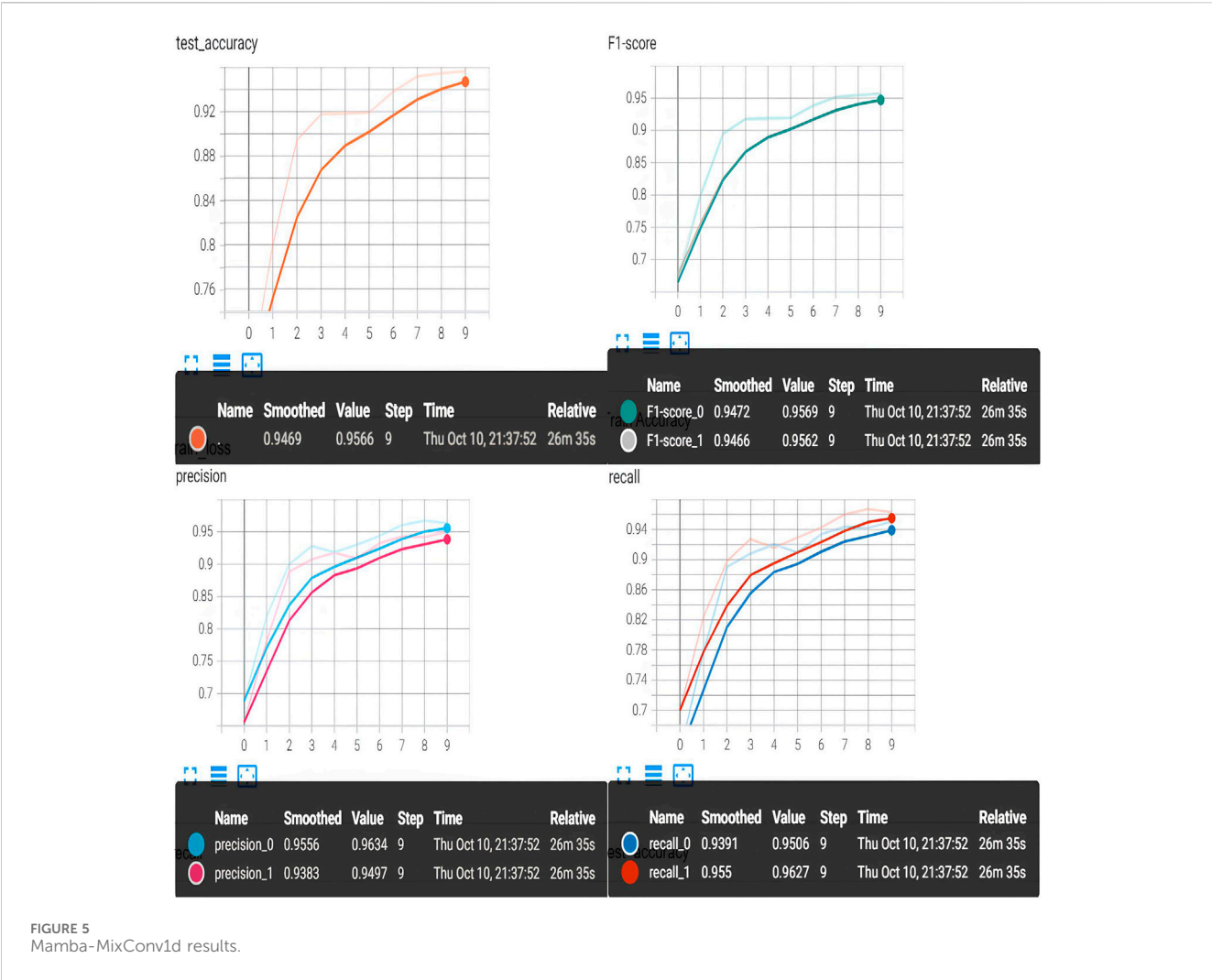
FIGURE 4
Original Mamba results.

4.1 Experiments

Dataset introduction. The eegmimdb dataset contains over 1,500 1-min and 2-min EEG recordings from 109 volunteers, which were obtained from subjects completing a series of motor/imagery tasks. Motor imagery or movement tasks were recorded as EEG signals from 64 channels positioned upon the subject's scalp. Each channel was annotated with three codes: T0, T1, and T2. T0 designates the rest period; T1 signifies the movement of the left hand in selected tasks; T2 denotes the movement of the right hand for certain tasks. Of the 109 participants, six individuals lacked sufficient data recordings and were excluded from the training experiment. All trials involved sustained and continuous movements of 4–4.1 s for execution and imagery tasks. To ensure consistent dataset representation, 4-s trial segments were extracted and clipped, removing any static states or extraneous non-experimental segments. The sampling rate was 160 Hz, and after each trial segment's clipping, 640 samples were obtained.

Experimental setup. We utilize the Magnetoencephalography and Electroencephalography (MNE) library to read raw general data format (GDF) files from the eegmimdb database. A 60 Hz bandpass filter is applied to remove power line interferences. A low-pass filter

with a cutoff frequency of 0.5 Hz is then used to suppress low-frequency noise. Finally, a bandpass filter ranging from 1 to 60 Hz is employed to attenuate high-frequency artifacts. The “T1” labels are converted into “0” labels, and the “T2” labels are converted into “1” labels. To maintain consistency in the dataset, the 640 continuous 4-s action data samples are divided into four non-overlapping windows of 160 samples each, which maintains the labeling of the original experiment. Datasets are divided into the motor task data, imagery task data, and data for both tasks. The data obtained is stored in a matrix format. We divide the processed EEG data matrix into a training set, test set and a validation set at a ratio of 7:2:1. The Adam optimizer is used to train the model, with an initial learning rate of 0.0001. Every 20 epochs during the training process, the learning rate is adjusted by 0.1 times its original value. The input sequence length of this Mamba model is 160, the state dimension is 256, and it has three layers. To shorten the training time, we use GPU servers and set a batch size of 24 for our training process. The configuration of the GPU server used in the experiment is as follows: the CPU model is AMD EPYC 9654, the graphics card is an RTX 4090 with 24 GB of video memory, and the system is equipped with 128 GB of RAM. It is also possible to conduct training on a laptop with more than 8 GB of video memory, although the process may be



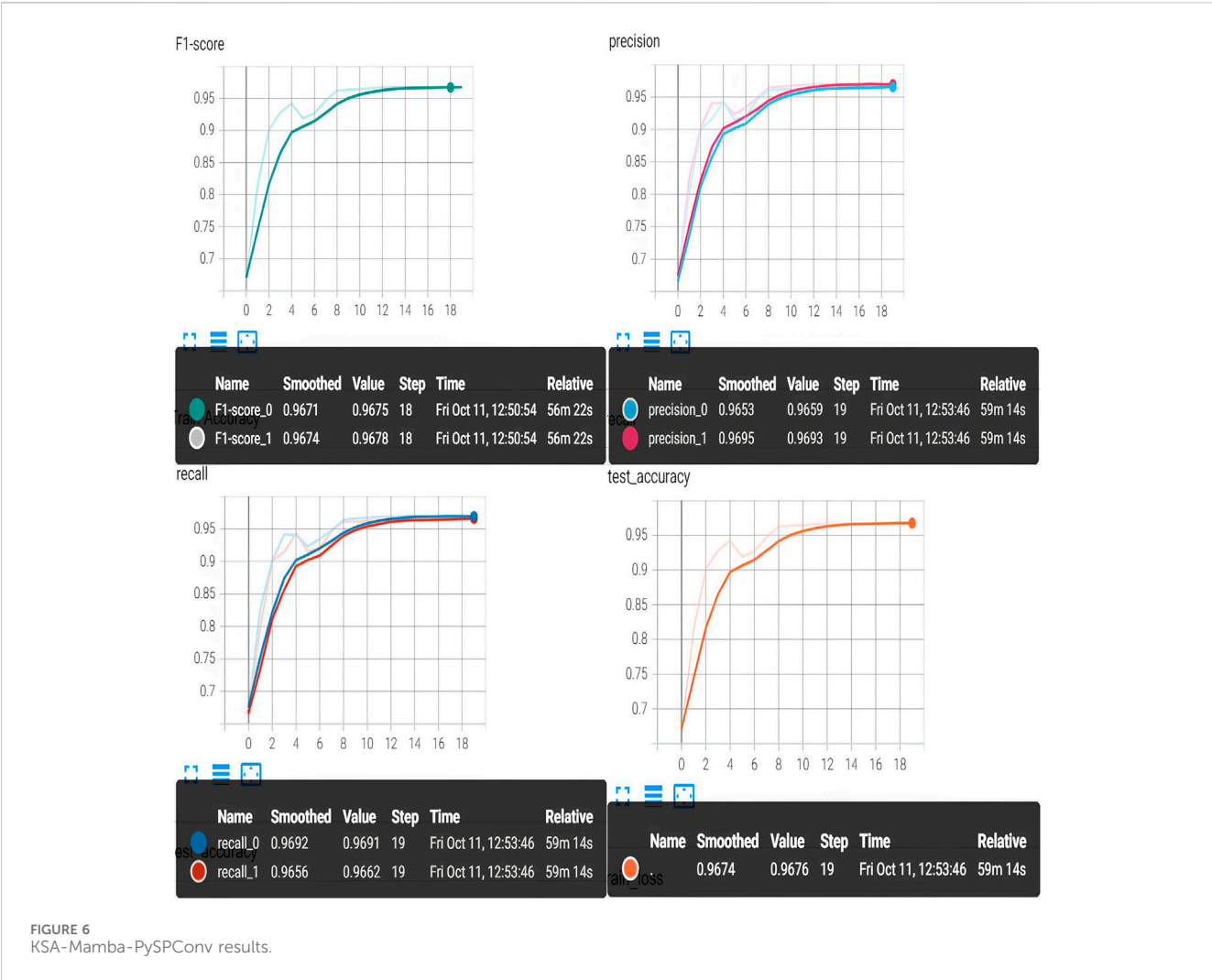
slower. We perform five sets of comparative experiments using different models to validate the model performance, namely, the original mamba model experiment, Mamba-mixconv1d model experiment, KSA-Mamba-PySPConv model experiment, executed motor task dataset experiment, and imagery motor task dataset experiment.

4.2 Results

Original mamba results. Figure 4 presents the experimental results of the original Mamba model on the brain-EEG motor imagery recognition task. The model employs a standard 1D convolution to extract local features, with a kernel size of 5. We compare the classification performance of the original Mamba model with EEGNet Fusion V2. The original Mamba model achieves significantly improved recognition performance, achieving an accuracy of 89.4%, and its precision, F1 score, and recall values all approach those of EEGNet Fusion V2, being higher than 85%. Furthermore, the original Mamba model requires fewer parameters compared to EEGNet Fusion V2. Increasing the size of the CNN kernel in the Mamba module yields improved model

performance, demonstrating that the Mamba structure is more effective in handling sequence data classification tasks compared to multi-layer deep neural networks (DNNs).

Mamba-MixConv1d results. To enhance the Mamba model's ability to extract local features, we replace the original 1D convolution with a MixConv1d convolutional module. Figure 5 presents the experimental results of the Mamba-MixConv1d model on the EEG motor imagery recognition task. We compare the classification performance of the Mamba-MixConv1d model with the original Mamba model. The Mamba-MixConv1d model achieves an accuracy of 95.6%, which is 6% higher than that of the original Mamba model. This improved model exhibits a precision and F1 score approaching 95%, and its recall is near 94%, showing that Mamba-MixConv1d significantly outperforms the original Mamba model. Experimental results conclusively demonstrate that the MixConv1d module possesses strong local feature extraction capabilities and solves the problem of insufficient local feature extraction in the original Mamba model. However, adding more convolutional branches and larger kernel sizes and performing additional fusion calculations result in substantial computational costs and memory consumption.

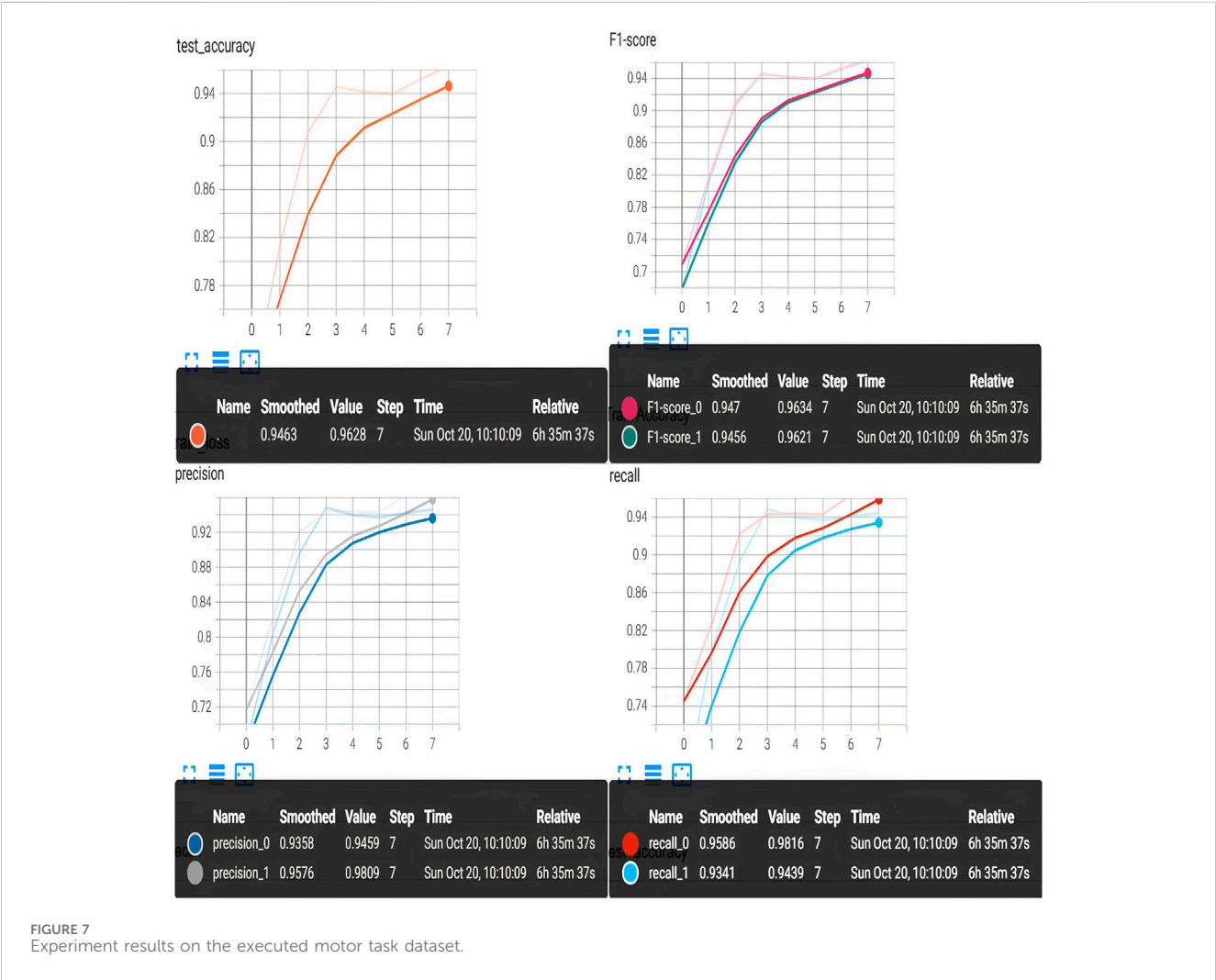


KSA-Mamba-PySPConv results. To reduce the computational load of the convolutional module, we replace MixConv1d with PySPConv. In addition, to improve the expressive power and feature quality of the Mamba structure output, we add the KSA attention module to enhance the channel and spatial features, capturing relationships between different parts of brain activity and detecting more complex data patterns. In Figure 6, we present the experimental results for the KSA-Mamba-PySPConv model on the EEG motor imagery recognition task. We compare its classification performance with that of the Mamba-MixConv1d model. The KSA-Mamba-PySPConv model achieves an accuracy rate of 96.76%, which is 1.76% higher than that of the Mamba-MixConv1d model. The precision, F1 score, and recall of the KSA-Mamba-PySPConv model are all above 96.5%, exceeding those of the Mamba-MixConv1d model by approximately 1%–2%. Experimental results demonstrate that KSA-Mamba-PySPConv possesses stronger local feature extraction capability and exhibits better overall performance than Mamba-MixConv1d. The statistical comparison of the parameters of the two convolutional modules finds that PySPConv has 21.1% fewer parameters compared to MixConv1d when configured with the same kernel size. PySPConv carries fewer redundant features and boasts higher

feature extraction efficiency and faster computation speed. Moreover, PySPConv allows for autonomous adjustment and configuration of the depth and kernel size of the convolutional layers while enabling the addition of more convolutional branches. Using the SPConv method significantly reduces the computational burden associated with adding branches and adjusting depths in PySPConv.

Executed motor task dataset experiment results. We test the performance of the KSA-Mamba-PySPConv model using the executed motor task dataset, and the results are shown in Figure 7. We compare the results to those from EEGNet Fusion V2. For the executed motor task, the KSA-Mamba-PySPConv model achieves an accuracy of 96.28%, 6.68% higher than that of EEGNet Fusion V2 (Chowdhury et al., 2023). The precision, F1 score, and recall of KSA-Mamba-PySPConv are all higher than those of EEGNet Fusion V2 by approximately 6.5%. This demonstrates that the KSA-Mamba-PySPConv model outperforms EEGNet Fusion V2 in executed motor movement tasks on eegmddb.

Imagery motor task dataset experiment results. We also test the performance of the KSA-Mamba-PySPConv model using the imagery motor task dataset, and the results are shown in Figure 8. In this domain, the KSA-Mamba-PySPConv model



achieves an accuracy of 96.33%, 6.73% higher than that of EEGNet Fusion V2. The precision, F1 score, and recall of KSA-Mamba-PySPConv are all higher than those of EEGNet Fusion V2 by approximately 6.5%. This demonstrates that the KSA-Mamba-PySPConv model outperforms EEGNet Fusion V2 in imagery motor tasks on eegmddb.

5 Discussion

This section discusses the performance differences between various models and their respective advantages and disadvantages. First, the experimental results for executed and imagery motor movement tasks are organized and compared for the models detailed in this study, as summarized in Table 1.

The original Mamba model demonstrates good performance in classifying heart rates, validating the model's effectiveness at handling 1D data. The use of MixConv1D in Mamba significantly improves classification accuracy, proving that MixConv1D enhances convolutional features through local

feature extraction and boosts overall model performance. When using PySPconv and KSA attention modules in conjunction with Mamba, compared to using MixConv1D alone, we observe improved classification accuracy. Furthermore, the parameters of PySPconv are 21.1% fewer than those of MixConv1D, which can be obtained by torchstat, highlighting the efficiency and lightweight nature of this module.

The results of some advanced deep learning models and our method for executed motor movement tasks are compared in Table 2.

As can be seen in Table 2, our new model exceeds previous research models in the performance metrics of the classification task. The KSA module helps the model better capture the nonlinear relationships in the data. PySPConv allows the model to process data sparsely, thereby reducing the number of parameters and computational complexity.

The results for imagery motor movement tasks are shown in Table 3. It can be seen that the proposed KSA-Mamba-PySPConv model also demonstrates excellent classification performance in imagery motor movement tasks, reaching the SOTA level.

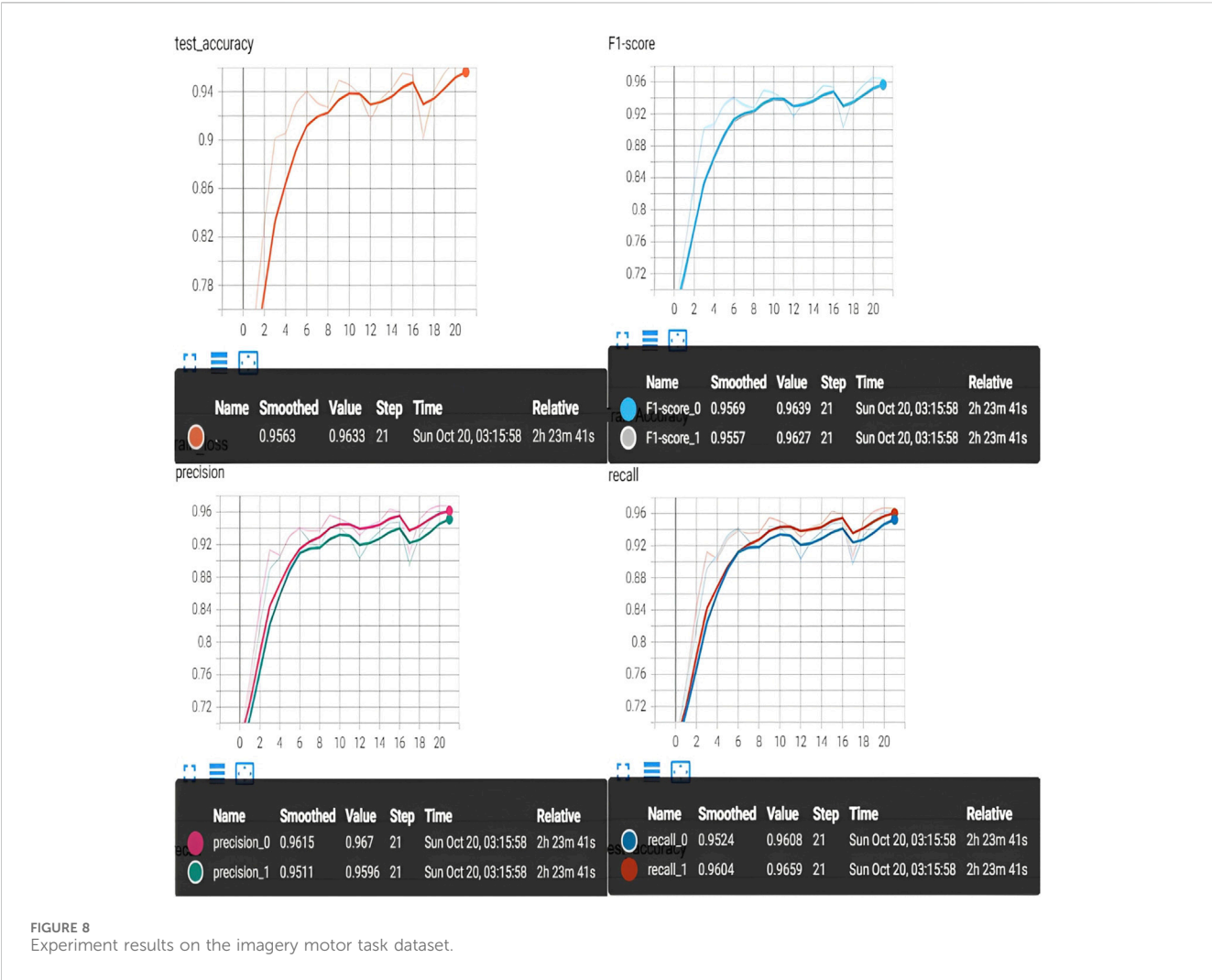


TABLE 1 Results for executed and imagery motor movement tasks.

Model	Accuracy %	Precision %		Recall %		F1 score %		Params	flops
		Left	Right	Left	Right	Left	Right		
Original Mamba	89.37	87.4	91.5	91.4	87.4	89.4	89.4	716.802k	2.369G
Mamba-MixConv1d	95.66	96.3	94.9	95.1	96.3	95.7	95.6	1513M	5.434G
KSA-Mamba-PySPConv	96.76	96.6	96.9	96.9	96.6	96.7	96.8	1193k	4.206G

6 Conclusion

In this paper, we propose a novel architecture called KSA-Mamba-PySPConv for the EEG imagery/motor movement classification tasks. The proposed scheme includes a KSA attention mechanism and a PySPConv module to enhance the features extracted from a single module layer. The KSA attention mechanism achieves enhanced and filtered channel and spatial features by integrating the KAN network with attention mechanisms. PySPConv utilizes different convolutional kernels to extract pyramid-like multi-scale features and employs split

operation and parameter-free feature fusion algorithms to achieve lightweight and efficient convolutions. These configurations enable KSA-Mamba-PySPConv to outperform conventional EEG classification models and achieve SOTA performance. The model exhibits excellent performance across different tasks on eegmimdb, proving its strong generalization capabilities. When deploying this model in a practical BCI system, we may need to consider the model's size and the consumption of computational resources. Therefore, techniques such as quantization and pruning might be employed for the deployment of the model. In the future, we will explore pruning algorithms and optimization methods to enhance the speed of this model.

TABLE 2 Results for executed motor movement tasks.

Model	Accuracy %	Precision %		Recall %		F1 score %		Params	Flops
		Left	Right	Left	Right	Left	Right		
DeepConvNet (Schirrmester et al., 2017)	76.6	76.2	77.1	77.3	76.0	76.7	76.5	97.302K	203.453M
ShallowConvNet (Schirrmester et al., 2017)	79.3	79.2	79.3	79.2	79.3	79.2	79.3	80B	580K
MMCNN (Jia et al., 2021)	81.4	82.2	80.6	80.5	82.3	81.3	81.4	-	-
EEGNet (Lawhern et al., 2018)	66.6	69.1	64.8	59.9	73.3	64.2	68.8	1.114K	132.251M
EEGNet Fusion (Roots et al., 2020)	84.1	84.2	84.5	83.8	83.9	84.0	84.2	17.682K	1.597G
EEGNet Fusion V2 (Chowdhury et al., 2023)	89.6	89.9	89.4	89.4	89.8	89.7	89.6	9.636M	16.546G
KSA-Mamba-PySPConv	96.28	94.59	98.09	98.16	94.39	96.34	96.21	1.193M	4.206G

TABLE 3 Results for imagery motor movement tasks.

Model	Accuracy %	Precision %		Recall %		F1 score %	
		Left	Right	Left	Right	Left	Right
DeepConvNet (Schirrmester et al., 2017)	76.2	76.5	75.9	76.0	76.4	76.3	76.1
ShallowConvNet (Schirrmester et al., 2017)	78.2	78.2	78.3	78.7	77.8	78.5	78.0
MMCNN (Jia et al., 2021)	81.6	81.7	81.5	81.9	81.2	81.8	81.3
EEGNet (Lawhern et al., 2018)	68.4	68.3	68.4	69.2	67.5	68.8	67.9
EEGNet Fusion (Roots et al., 2020)	83.8	85.0	83.3	82.9	84.8	83.9	84.0
EEGNet Fusion V2 (Chowdhury et al., 2023)	87.8	88.1	87.5	87.5	88.1	87.8	87.8
KSA-Mamba-PySPConv	96.33	96.7	95.96	96.08	96.59	96.39	96.27

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 in the article/supplementary material.

Author contributions

ZL: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that no financial support was received for the research 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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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

- Chen, J., Zhang, Y., Pan, Y., Xu, P., and Guan, C. (2022). A transformer-based deep neural network model for SSVEP classification.
- Chowdhury, R. R., Muhammad, Y., and Adeel, U. (2023). Enhancing cross-subject motor imagery classification in EEG-based brain-computer interfaces by using multi-branch CNN. *Sensors* 23 (18), 7908. doi:10.3390/s23187908
- Duta, I. C. (2021). *Pyramidal convolution: rethinking convolutional neural networks for visual recognition*. Cornell University - arXiv, Cornell University - arXiv.
- Gu, A., and T. D. Mamba (2023). Linear-time sequence modeling with selective state spaces.
- Hou, Y., Jia, S., Lun, X., Zhang, S., Chen, T., Wang, F., et al. (2022). Deep feature mining via the attention-based bidirectional long short term memory graph convolutional neural network for human motor imagery recognition. *Front. Bioeng. Biotechnol.* 9. doi:10.3389/fbioe.2021.706229
- Jia, Z., Lin, Y., Wang, J., Yang, K., Liu, T., and Zhang, X. (2021). “MMCNN: a multi-branch multi-scale convolutional neural network for motor imagery classification,” in *Proceedings of the machine learning and knowledge discovery in databases: European conference, ECML PKDD 2020, ghent, Belgium, 14–18 september 2020; proceedings, Part III* (Berlin, Germany: Springer), 736–751.
- Keutayeva, A., and Abibullaev, B. (2023). Exploring the potential of attention mechanism-based deep learning for robust subject-independent motor-imagery based BCIs. *IEEE Access* 11, 107562–107580. doi:10.1109/access.2023.3320561
- Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., and Lance, B. J. (2018). EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *J. Neural Eng.* 15, 056013. doi:10.1088/1741-2552/aace8c
- Li, Y., Guo, L., Liu, Y., Liu, J., and Meng, F. (2021). A temporal-spectral-based squeeze-and-excitation feature fusion network for motor imagery EEG decoding. *IEEE Trans. Neural Syst. Rehabilitation Eng.* 29, 1534–1545. doi:10.1109/tnsre.2021.3099908
- Luo, J., Wang, Y., Xia, S., Lu, N., Ren, X., Shi, Z., et al. (2023). A shallow mirror transformer for subject-independent motor imagery BCI. *Comput. Biol. Med. Sept* 164, 107254. doi:10.1016/j.combiomed.2023.107254
- Roots, K., Muhammad, Y., and Muhammad, N. (2020). Fusion convolutional neural network for cross-subject EEG motor imagery classification. *Computers* 9, 72. doi:10.3390/computers9030072
- Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., et al. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain Mapp.* 38, 5391–5420. doi:10.1002/hbm.23730
- Vaca-Rubio, C. J., Blanco, L., Pereira, R., and Caus, M. (2024). Kolmogorov-arnold networks (KANs) for time series Analysis.
- Zhang, Q., Jiang, Z., Lu, Q., Han, J., Zeng, Z., Gao, S., et al. (2020). “Split to Be slim: an overlooked redundancy in vanilla convolution,” in *Proceedings of the twenty-ninth international joint conference on artificial intelligence*. doi:10.24963/ijcai.2020/442



OPEN ACCESS

EDITED BY

Zohaib Mushtaq,
University of Sargodha, Pakistan

REVIEWED BY

Guoqiang Long,
Southeast University, China
Javed Khan,
Department of Software Engineering
University of Science and Technology,
Pakistan
Muhammad Owais,
Khalifa University, United Arab Emirates

*CORRESPONDENCE

Muhammad Hanif
✉ muhammad.hanif@oru.se

RECEIVED 01 February 2025

ACCEPTED 25 March 2025

PUBLISHED 16 April 2025

CITATION

Umair M, Ahmad J, Alasbali N, Saidani O,
Hanif M, Khattak AA and Khan MS (2025)
Decentralized EEG-based detection of major
depressive disorder via transformer
architectures and split learning.
Front. Comput. Neurosci. 19:1569828.
doi: 10.3389/fncom.2025.1569828

COPYRIGHT

© 2025 Umair, Ahmad, Alasbali, Saidani, Hanif,
Khattak 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.

Decentralized EEG-based detection of major depressive disorder via transformer architectures and split learning

Muhammad Umair¹, Jawad Ahmad², Nada Alasbali³,
Oumaima Saidani⁴, Muhammad Hanif^{5*}, Aizaz Ahmad Khattak⁶
and Muhammad Shahbaz Khan⁶

¹Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia, ²Cybersecurity Center, Prince Mohammad Bin Fahd University, Al-Khobar, Saudi Arabia, ³Department of Informatics and Computer Systems, College of Computer Science, King Khalid University, Abha, Saudi Arabia, ⁴Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, ⁵Department of Informatics, School of Business, Örebro Universitet, Örebro, Sweden, ⁶School of Computing, Engineering and the Built Environment, Edinburgh Napier University, Edinburgh, United Kingdom

Introduction: Major Depressive Disorder (MDD) remains a critical mental health concern, necessitating accurate detection. Traditional approaches to diagnosing MDD often rely on manual Electroencephalography (EEG) analysis to identify potential disorders. However, the inherent complexity of EEG signals along with the human error in interpreting these readings requires the need for more reliable, automated methods of detection.

Methods: This study utilizes EEG signals to classify MDD and healthy individuals through a combination of machine learning, deep learning, and split learning approaches. State of the art machine learning models i.e., Random Forest, Support Vector Machine, and Gradient Boosting are utilized, while deep learning models such as Transformers and Autoencoders are selected for their robust feature-extraction capabilities. Traditional methods for training machine learning and deep learning models raises data privacy concerns and require significant computational resources. To address these issues, the study applies a split learning framework. In this framework, an ensemble learning technique has been utilized that combines the best performing machine and deep learning models.

Results: Results demonstrate a commendable classification performance with certain ensemble methods, and a Transformer-Random Forest combination achieved 99% accuracy. In addition, to address data-sharing constraints, a split learning framework is implemented across three clients, yielding high accuracy (over 95%) while preserving privacy. The best client recorded 96.23% accuracy, underscoring the robustness of combining Transformers with Random Forest under resource-constrained conditions.

Discussion: These findings demonstrate that distributed deep learning pipelines can deliver precise MDD detection from EEG data without compromising data security. Proposed framework keeps data on local nodes and only exchanges intermediate representations. This approach meets institutional privacy requirements while providing robust classification outcomes.

KEYWORDS

split learning, transformers, autoencoder, EEG, major depressive disorder, smart diagnostic, neurological behavior

1 Introduction

The human body possess remarkable complexity, and the brain plays a pivotal role in cognitive and behavioral functions (Vohryzek et al., 2025). Maintaining a healthy brain is essential for optimal decision-making (Hagan et al., 2025). The human brain contains billions of neuron, which coordinate various neurological activities (Herculano-Houzel, 2009). Nonetheless, a range of disorders impact brain function, including Major Depressive Disorder (MDD) i.e., leading contributor to mental health challenges (Kreiviniene et al., 2025).

Early diagnosis of MDD is important for maintaining mental well-being, but current diagnostic methods rely on subjective clinical evaluations and self-reported symptoms prone to human error and inefficiency (Hagan et al., 2025; Kreiviniene et al., 2025). This underscores the need for a reliable diagnostic tool that assists clinicians in making accurate and timely decisions.

Electroencephalography (EEG) offers a promising approach for examining the neurophysiological underpinnings of mental health conditions (Perrottelli et al., 2021). It measures electrical brain activity with high temporal resolution and is non-invasive, cost-effective, and portable (Perrottelli et al., 2021). Previous studies have revealed changes in EEG patterns, such as power spectral density shifts and alterations in brain wavebands, among individuals with MDD (Liang et al., 2021). Although EEG signals contain valuable diagnostic information, extracting meaningful insights from these high-dimensional and noisy data remains a challenge.

Machine learning (ML) and deep learning (DL) techniques demonstrate potential for analyzing EEG signals (Subhani et al., 2017; Rahul et al., 2024; Umair et al., 2021; Diehl and Cook, 2015). DL models can automatically extract relevant patterns, aiding in differentiating healthy individuals from those affected by MDD (Subhani et al., 2017). However, traditional ML and DL training often occurs in centralized systems, which raises privacy risks and demand costly computational infrastructure (Umair et al., 2024; Rahul et al., 2024). Healthcare institutions also hesitate to share sensitive data, highlighting the need for decentralized methods that safeguard patient privacy (Umair et al., 2023).

Federated Learning (FL) has emerged as a key approach to decentralized training by enabling local model updates on client devices while aggregating models at a central server (McMahan et al., 2017). Although FL preserves data privacy, some clients may face resource constraints that hinder local training (Umair et al., 2023). However, a similar concept as FL i.e., split learning (SL) addresses this challenge by splitting the model architecture between clients and a central server, transferring only intermediate representations instead of raw data (Gupta and Raskar, 2018). This structure reduces the computational burden on resource-limited devices having on device training as well (Jia et al., 2024). In the context of EEG-based MDD diagnosis, SL can integrate distributed data from multiple healthcare providers without centralizing sensitive information, offering a scalable and reliable framework for developing effective diagnostic models.

This study explores the concept of SL in conjunction with various ML and DL models to classify MDD patients using an EEG dataset. Model selection is critical for robust classification, so

multiple ML classifiers including Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbors (KNN), and Gradient Boosting (GB) are utilized for their proven performance. In addition, advanced DL architectures such as Transformers and Autoencoders are employed to capture the complex, high-dimensional characteristics of EEG data. An ensemble learning principles is then implemented in a SL framework, with three clients chosen for comparative evaluation. Classification reports and confusion matrices serve as the primary metrics to assess the performance of these models. Thus, key contribution of this study is as follows:

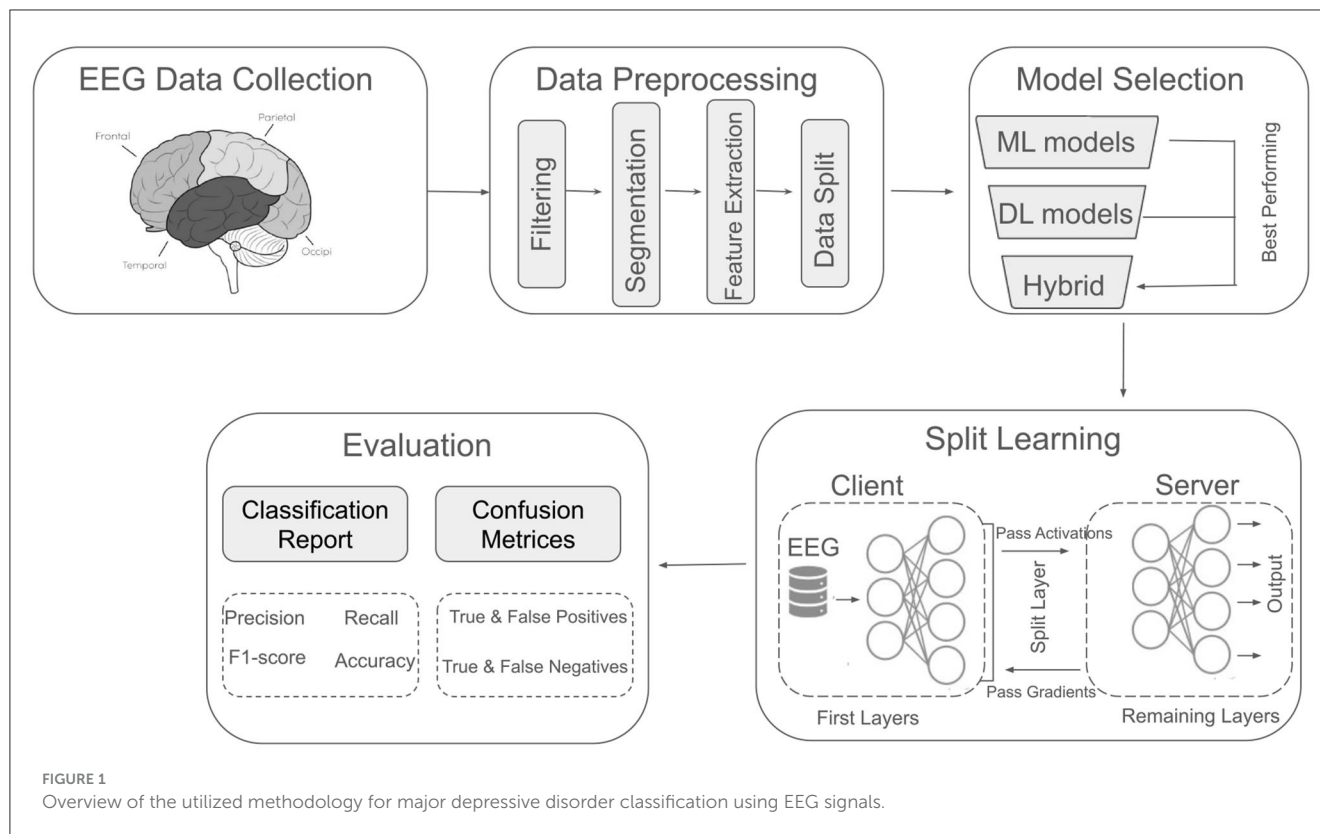
1. Split learning framework tailored for EEG-based MDD classification. And within this SL approach ML and DL models are utilized for EEG features extractions and classification.
2. Proposed a ensemble model tailored for MDD disorder classification through comprehensive performance metrics across three clients in SL settings.

This article is organized into five main sections. Section 1 provides the background and context of the study. Section 2 reviews related work and relevant literature. Section 3 details the methods and materials used in the experiments. Section 4 presents the obtained results and offers a comprehensive discussion. Finally, Section 5 concludes the study by summarizing the key findings.

2 Related work

Researchers have recently explored a range of ML and DL models for medical applications (Gour et al., 2023; Sultan et al., 2023; Owais et al., 2022) yielding promising results. However, as discussed in Section 1, the majority of these algorithms rely on centralized architectures that raise privacy concerns and limit their practical applicability. This section reviews recent studies that utilize ML and DL approaches for EEG-based analysis, as well as decentralized solutions aimed at safeguarding data privacy and promoting scalability.

Park et al. (2021) employed multiple ML models SVM, RF, and elastic net regression to classify six major psychiatric disorders and healthy controls using EEG features such as power spectrum density (PSD) and functional connectivity (FC). Their elastic net model achieved the highest accuracy across disorders, notably identifying schizophrenia with 93.83% accuracy using alpha PSD, anxiety disorders with 91.03% accuracy via whole-band PSD, and trauma and stress-related disorders with 91.21% accuracy from beta FC features. Rafiei et al. (2022) proposed a DL model based on a customized InceptionTime architecture for MDD detection, achieving 91.67% accuracy with full-channel EEG data and 87.5% after channel reduction. Rivera et al. (2022) conducted a systematic mapping of 46 primary studies that leveraged DL for EEG-based mental disorder diagnoses, revealing CNNs as the most common approach and epilepsy as the most frequently studied disorder. Wang et al. (2024) developed DiffMDD, a diffusion-based DL framework for diagnosing MDD, incorporating Forward Diffusion Noisy Training and Reverse Diffusion Data Augmentation to mitigate noise and data scarcity. Anik et al. (2024) introduced an 11-layer 1D-CNN for MDD classification, focusing on gamma band



EEG segments of 15-second epochs, and attained 99.60% accuracy, 100% sensitivity, and 99.21% specificity.

Earl et al. (2024) used an RF model on resting-state and emotionally charged EEG-based FC features, achieving classification accuracies of 92.3%, 94.9%, and 89.7%. Metin et al. (2024) combined 1D-CNN with LSTM and 2D-CNN to classify bipolar disorder, reporting a higher accuracy (95.91%) with the 2D-CNN compared to the 1D-CNN+LSTM (93%). de et al. (2024) proposed SLiTRANet, a transformer-based DL framework for MDD detection, achieving 99.92% accuracy, 99.90% sensitivity, and 99.95% specificity. Zhu et al. (2025) introduced MTNet, a transformer network integrating EEG and eye-tracking data for depression detection, obtaining 91.79% accuracy and highlighting the benefits of intermediate fusion. Ahmed et al. (2024) utilized an ensemble of transformer based models (vanilla BERT, BERTweet, ALBERT) to classify depression severity from social media posts, while Ilias et al. (2024) employed BERT and MentalBERT with extra-linguistic information for depression and stress detection. Sun et al. (2023) introduced TensorFormer, a multimodal transformer framework for sentiment analysis and depression detection, demonstrating performance enhancements on multiple datasets.

Decentralized learning approaches such as FL have also garnered attention. Zhang et al. (2023) proposed FedBrain for diagnosing brain disorders, integrating data augmentation, domain alignment, and personalized predictors to handle high-dimensional features and variable data distributions. FedBrain achieved 79% accuracy with privacy preservation through differential privacy and homomorphic encryption. Li et al. (2023) introduced CAFed,

an asynchronous federated CNN-based optimizer for detecting depression from social media data, improving communication efficiency, convergence rates, and privacy protection while surpassing FedAvg in non-convex problem settings.

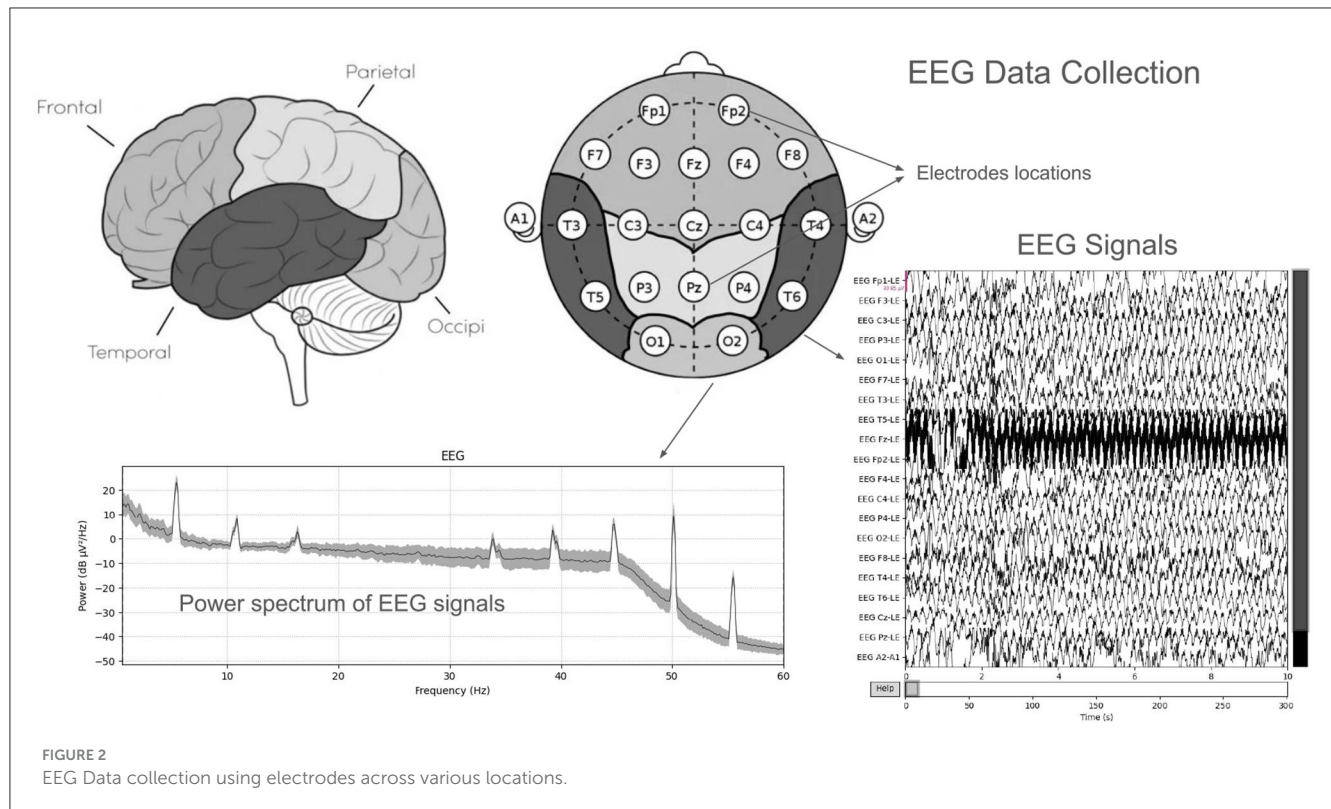
Although these studies demonstrate promising performance, their reliance on traditional ML and DL methods often involves centralized or FL-based architectures that either risk privacy or suffer from resource constraints. Therefore, this work adopts SL as a resource-sharing methodology to address these concerns, balancing privacy preservation with computational feasibility.

3 Materials and methods

This section describes the experimental procedures and methods employed in this study. Figure 1 presents an overview of the methodology, which comprises five key components: EEG data collection, data preprocessing, model selection, SL, and evaluation. Each component is discussed in detail in the subsequent subsections.

3.1 Data collection

A publicly available EEG dataset (Mumtaz, 2016) is used in this study, comprising of two groups: 33 MDD patients (mean age 40.33 ± 12.86) and 30 age-matched healthy controls (mean age 38.23 ± 15.64), recruited from the outpatient clinic at Hospital Universiti Sains Malaysia (HUSM) (Mumtaz et al., 2017). EEG data



were recorded under controlled conditions, with 5-min eyes-closed (EC) and eyes-open (EO) sessions, using a 19-channel system aligned with the international 10–20 standard and a linked-ear reference (Figure 2). The system applied a 0.5–70 Hz bandpass filter, a 50 Hz notch filter, and a sampling rate of 256 Hz, followed by referencing to an infinity reference for subsequent analyses (Mumtaz et al., 2017). Participants were instructed to avoid caffeine and other substances because caffeine intake can alter arousal states by inhibiting adenosine, thus introducing variability and potential noise into EEG recordings (Lesar et al., 2025; Zhu et al., 2024). MDD severity was assessed using the Beck Depression Inventory-II (BDI-II) and the Hospital Anxiety and Depression Scale (HADS) (Mumtaz et al., 2017). A sample shown in Figure 3 of a raw EEG signal recorded over 19 channels in a 10-second window, demonstrates the time-domain structure of brain activity. Accessed dataset (Mumtaz et al., 2017) contains the files structure in pdf format, thus, we utilized python library [i.e., mne (Gramfort et al., 2013)] in order to preprocess these EEG recording for our case, Figure 3 is basically the recordings of EEG sample that is preprocessed via MNE library. Each channel corresponds to a specific scalp location following the international 10–20 system (e.g., Fp1, F3, P3), allowing for regional analysis of cortical oscillations. Notable fluctuations in amplitude can be seen across channels, which may reflect ongoing cognitive or physiological processes, as well as potential artifacts (e.g., eye blinks or muscle movements). Similarly, in Figure 4, the power spectral density of the EEG signal, color-coded to highlight the standard frequency bands: Delta (0.5–4 Hz), Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), and Gamma (>30 Hz). The PSD curve represents the distribution of signal power across frequencies, with characteristic

peaks often observed in the Delta and Alpha ranges. Identifying the relative power in these frequency bands can reveal important information about the participant's mental state and the presence of any abnormal patterns indicative of neurological or psychiatric conditions.

3.2 Data preprocessing

3.2.1 Data loading

All EEG recordings were loaded in a standardized manner to ensure uniform data handling. A common input of 10 seconds of EEG recording was used from each sample was then applied across all channels to facilitate consistent inter-subject comparisons.

3.2.2 Filtering

Filtering is an essential step in EEG signal processing because raw signals often contain noise and artifacts in frequency ranges that are not relevant for subsequent analysis. To address this, we employed a bandpass filter to removes unwanted noise and keeping frequency components as well. A bandpass filter in the 0.5–60 Hz range was employed to suppress low-frequency drifts and high-frequency noise. Mathematically, it shown in Equation 1, where $x(t)$ denotes the raw EEG signal and filtered signal is denoted as $\tilde{x}(t)$

$$\tilde{x}(t) = \mathcal{F}^{-1} \left\{ \mathcal{F}\{x(t)\} \cdot H(\omega) \right\} \quad (1)$$

Here, $\mathcal{F}$ denotes the Fourier transform, and $H(\omega)$ is the ideal passband response for the specified frequency range.

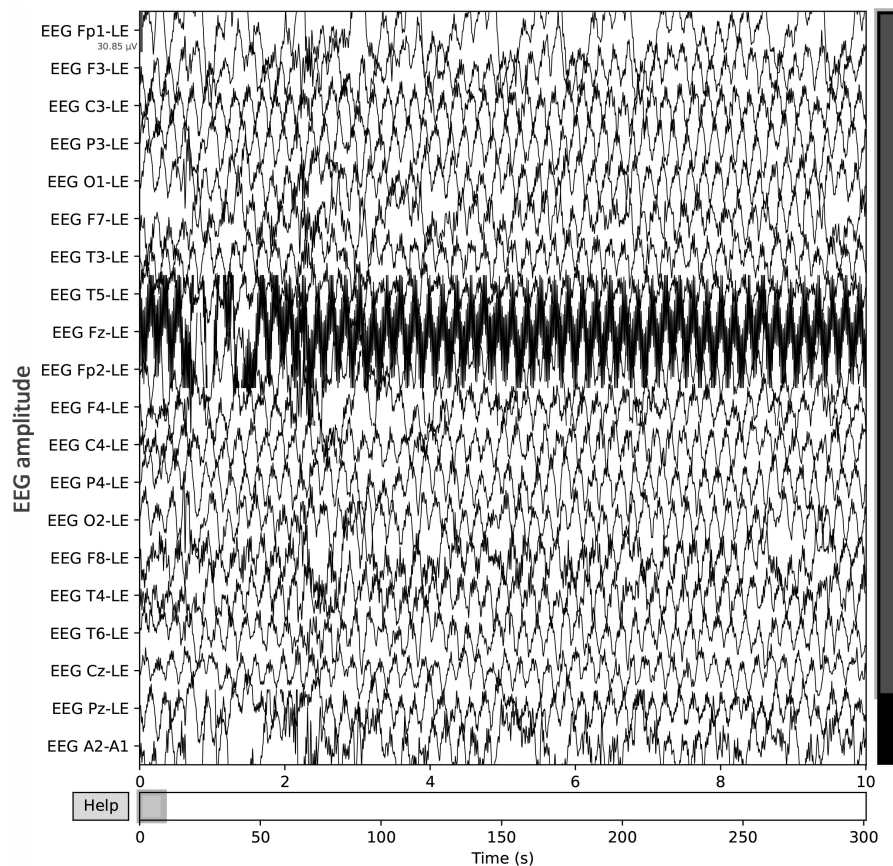


FIGURE 3

Raw EEG data, recorded over 19 channels in a 10-second window, demonstrates the time-domain structure of brain activity.

3.2.3 Epoch segmentation

We segmented the continuous EEG into fixed-length epochs of 5 seconds each, with a 1 second overlap between consecutive segments. This specific window length strikes a practical balance between capturing relevant EEG frequency components (e.g., alpha, beta, and gamma bands) and maintaining adequate temporal resolution for classification. Shorter windows (2–3 seconds) often fail to capture stable patterns, while substantially longer windows (e.g., 8–10 seconds) risk smoothing out important transient features. The 1 second overlap ensures continuity across segment boundaries and mitigates the loss of transitional information that can occur at strict epoch boundaries. Mathematically it is given in Equation 2.

$$E_i = \{x(t) \mid t \in [i \cdot \Delta, (i \cdot \Delta + \tau)]\} \quad (2)$$

where $\tau = 5$ seconds is the epoch length, and $\Delta = \tau - 1$ seconds denotes the shift applied between consecutive segments.

3.2.4 Feature extraction

Each epoch was transformed into a feature vector by computing a set of statistical descriptors that capture both amplitude variations and higher-order properties of the signal distribution. If x_n denotes the amplitude of the signal at time index n , and N is the number of

samples per epoch, the following examples (using Equations 3–eq6) illustrate key feature computations. Whereas, P2P in Equation 5 refers the peak to peak amplitude of the recorded EEG signal.

$$\mu = \frac{1}{N} \sum_{n=1}^N x_n \quad (3)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{n=1}^N (x_n - \mu)^2} \quad (4)$$

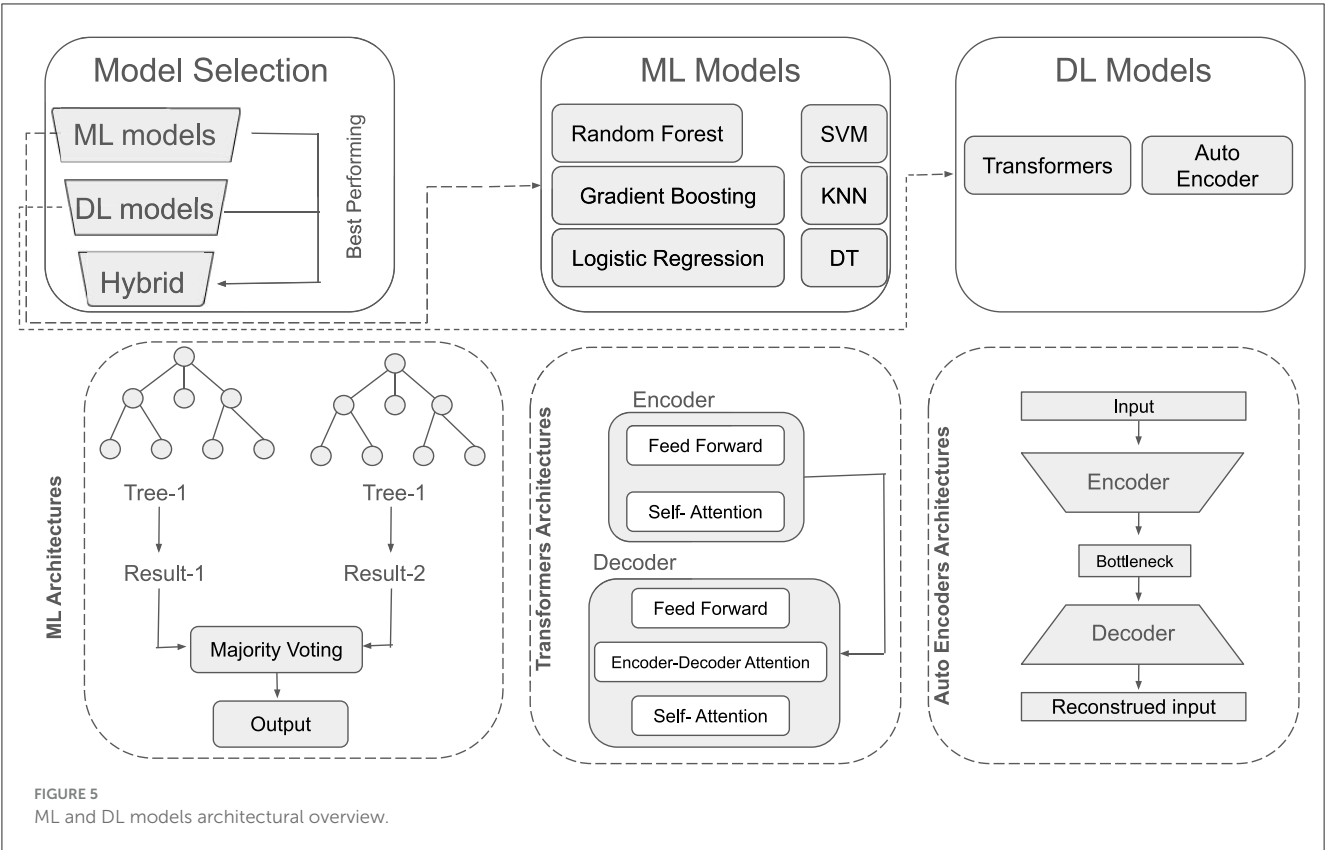
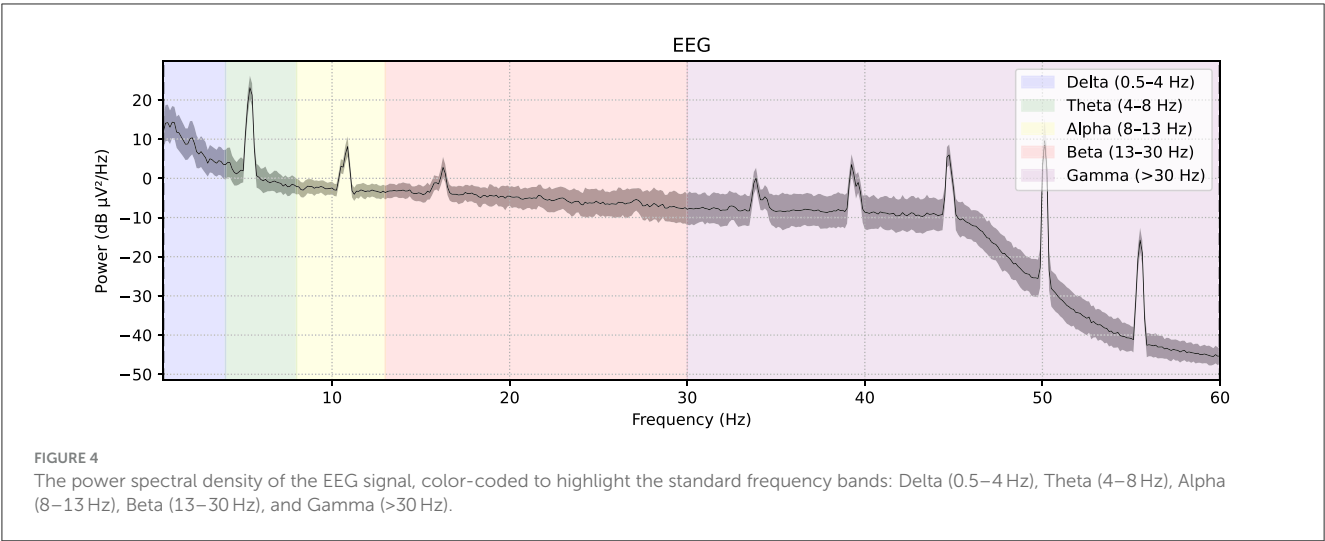
$$\text{P2P} = \max\{x_n\} - \min\{x_n\} \quad (5)$$

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{n=1}^N x_n^2} \quad (6)$$

Higher-order moments, including skewness and kurtosis, were also evaluated to account for asymmetry in the signal distribution.

3.2.5 Labeling

Each epoch was then assigned a class label based on the participant's diagnostic status (0 for healthy controls, 1 for MDD). The final output of this preprocessing pipeline was a feature matrix of size along with a corresponding label vector. This structured dataset was then used for the model training and evaluation.



3.3 ML and DL models

In this section, the architectures of utilized ML and DL models has been discussed. An overview of their architecture has been shown in [Figure 5](#).

3.3.1 Machine learning classifiers

Model selection plays a pivotal role in achieving robust classification performance. Consequently, the following tree-based and other conventional ML classifiers were employed: LR, RF, SVM,

DT, KNN, and GB. Each classifier offers distinct inductive biases and learning strategies that are used to capture diverse patterns in EEG-based features for distinguishing MDD patients from healthy controls.

Moreover, all hyperparameter settings (e.g., `n_estimators` = 100 for RF, `max_depth` = 10 for DT, `n_neighbors` = 7 for KNN) were determined via a grid search procedure. This involved systematically varying key parameters within predefined ranges and evaluating model performance through cross-validation on the training set. The final configurations were selected based on their classification report.

3.3.1.1 Decision tree

A DT recursively partitions the feature space by selecting optimal split points that maximize homogeneity in the resulting subsets. As given in Equation 7, D represent the training dataset and j be the index of a potential split on feature x_j . The split criterion can be based on information gain or the Gini index. For instance, using the Gini index G , the split s on feature x_j is chosen to minimize.

$$s^* = \arg \min_s \left[\frac{n_L}{n} G(D_L) + \frac{n_R}{n} G(D_R) \right], \quad (7)$$

where D_L and D_R are the left and right child partitions of D after the split s , n_L and n_R are the respective sizes of these partitions, and n is the total number of samples in D .

3.3.1.2 Random forest

RF constructs an ensemble of decision trees, each trained on a bootstrap sample of the original dataset. At each split node, a random subset of features is considered to enhance diversity among the trees. The model's prediction is obtained via majority voting (for classification) across all trees. Mathematically it is given in Equation 8.

$$\hat{y} = \text{mode}(\{h_t(\mathbf{x}) \mid t = 1, \dots, T\}), \quad (8)$$

where $h_t(\mathbf{x})$ denotes the prediction from the t -th tree and T is the total number of trees in the forest.

3.3.1.3 Gradient boosting

GB sequentially fits new weak learners (often decision trees) to the negative gradient of a specified loss function. As given in Equation 9, y_i denotes the true label of instance i , and let F_{m-1} be the ensemble model at iteration $(m - 1)$. A new base learner h_m is trained to approximate the negative gradient of the loss $\ell(y_i, F_{m-1}(\mathbf{x}_i))$. The ensemble is then updated as:

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \eta \cdot h_m(\mathbf{x}), \quad (9)$$

where η is the learning rate. This iterative procedure allows optimizer to correct the residual errors from the previous step, leading to improved performance over single-tree methods.

3.3.1.4 Logistic Regression

LR estimates the probability that a sample $\mathbf{x}$ belongs to the positive class (denoted by $y = 1$) using the sigmoid function. As given in Equation 10:

$$p(\mathbf{x}) = \sigma(\boldsymbol{\beta}^\top \mathbf{x} + \beta_0) = \frac{1}{1 + \exp(-(\boldsymbol{\beta}^\top \mathbf{x} + \beta_0))}, \quad (10)$$

where $\boldsymbol{\beta}$ is the weight vector, β_0 is the intercept, and $\sigma(\cdot)$ represents the sigmoid. A threshold (i.e., 0.5) is applied to $p(\mathbf{x})$ to determine class of the given input.

3.3.1.5 Support vector machine

SVM is a widely used supervised learning technique renowned for its effectiveness in high-dimensional spaces and robust generalization capabilities. The key principle of SVM lies in finding an optimal decision boundary (hyperplane) that maximizes the

margin between different classes, thus improving classification performance. In its linear form, SVM is given in Equation 11,

$$\underset{\mathbf{w}, b}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{w}\|^2 \quad \text{subject to} \quad y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \quad \forall i, \quad (11)$$

where $\mathbf{w}$ and b define the hyperplane, and $y_i \in \{-1, +1\}$ denotes class labels. Nonlinear decision boundaries can be learned via kernel functions.

3.3.1.6 K-Nearest Neighbors

K-Nearest Neighbors (KNN) is a simple yet effective non-parametric, instance-based learning method. It assigns a class to a query point $\mathbf{x}_q$ by considering the classes of its k nearest neighbors. The distance metric often used is the Euclidean distance (as given in Equation 12).

$$d(\mathbf{x}_q, \mathbf{x}_i) = \sqrt{\sum_{j=1}^M (x_{qj} - x_{ij})^2}, \quad (12)$$

where M is the number of features. The predicted class is determined by a majority vote among these k neighbors.

3.4 DL models

3.4.1 Transformer models

Transformer architectures have gained prominence for their capacity to capture long-range dependencies and context within sequential data, making them particularly appealing for EEG-based analysis. Unlike traditional recurrent networks, Transformers dispense with explicit recurrence and convolutional operations, relying instead on an attention mechanism. Mathematically (as shown in Equation 13), Q , K , and V denote the query, key, and value matrices, respectively, then a single-head attention module can be written as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right) V, \quad (13)$$

where d_k is the dimension of the key vectors, and softmax function normalizes the attention scores. Multi-head attention extends this formulation by employing several parallel attention mechanisms and concatenating their outputs to enrich the representational capacity (as given in Equation 14).

$$\text{MultiHead}(Q, K, V) = \parallel_{h=1}^H \text{Attention}(QW_h^Q, KW_h^K, VW_h^V) W_h^O, \quad (14)$$

where $\parallel$ denotes concatenation across H attention heads, and W_h^Q , W_h^K , W_h^V , and W_h^O are learned projection matrices.

In EEG analysis, input sequences can be framed as embeddings of multi-channel signals over time, enabling the transformer to learn context-dependent patterns relevant for mental health classification. Positional encodings are commonly added to the input embeddings to preserve temporal order. This attention-based approach often yields superior performance in capturing nuanced dependencies within EEG signals, especially for tasks such as MDD detection.

3.4.2 Autoencoders

Autoencoders are a family of neural network models designed to learn compressed representations (encodings) of the input data by minimizing reconstruction error. They consist of two main components i.e., Encoder and Decoder. Encoder, maps an input $\mathbf{x} \in \mathbb{R}^D$ to a latent code $\mathbf{z} \in \mathbb{R}^d$ (with $d < D$) as given in Equation 15,

$$\mathbf{z} = f_{\text{enc}}(\mathbf{x}). \quad (15)$$

Whereas, decoder reconstructs the original input from $\mathbf{z}$, producing $\hat{\mathbf{x}} \in \mathbb{R}^D$ as when in Equation 16:

$$\hat{\mathbf{x}} = f_{\text{dec}}(\mathbf{z}). \quad (16)$$

The model is typically optimized to minimize:

$$\mathcal{L} = \|\mathbf{x} - \hat{\mathbf{x}}\|^2, \quad (17)$$

another suitable measure of reconstruction fidelity. By constraining the latent dimension d , autoencoders learn salient features that represent the most informative aspects of the data. In EEG-based MDD detection, autoencoders can help denoise signals or extract meaningful representations that capture underlying neural patterns. These learned representations may then serve as inputs for downstream classifiers or be integrated into end-to-end DL pipelines for improved diagnostic accuracy.

3.5 Ensemble learning

Ensemble learning combines multiple base models to achieve improved predictive performance relative to any single constituent model. This approach capitalizes on the principle of “wisdom of the crowd,” where diverse model outputs are aggregated to form a final decision. A common strategy for building ensembles include:

3.5.1 Bagging

Bagging (Bootstrap Aggregating) trains each base learner on a different bootstrap sample (randomly drawn with replacement) of the original dataset. Let $\{\mathcal{D}_b\}_{b=1}^B$ be the collection of bootstrap samples, each used to train a distinct model $h_b(\mathbf{x})$. The final prediction is obtained by averaging or voting across the ensemble:

$$\hat{y}_{\text{bagging}} = \begin{cases} \text{majority}\{h_b(\mathbf{x})\}_{b=1}^B, & \text{classification} \\ \frac{1}{B} \sum_{b=1}^B h_b(\mathbf{x}), & \text{regression} \end{cases} \quad (18)$$

Bagging often reduces variance without substantially increasing bias, making it effective for high-variance models like decision trees.

Ensemble learning is particularly relevant for EEG-based MDD classification due to the high dimensionality and variability inherent in EEG signals. It is because of this reason, ensemble learning was utilized using best performing ML model and then best performing DL model. By integrating these models, ensembles have the potential to yield more reliable and generalizable predictions for clinical applications.

3.6 Split learning

SL offers a decentralized framework designed to address privacy and resource constraints, particularly relevant when clinical or EEG datasets cannot be shared in raw form. Unlike fully centralized methods, where all data must reside on a single server, SL divides a neural network into multiple segments to be trained collaboratively between clients and a central server. In this study, three clients are assumed, each holding a portion of the EEG data locally (as shown in Figure 6). After data preprocessing (Section 3.2), SL is implemented to enable model training without direct data exchange across clients.

3.6.1 Architectural overview

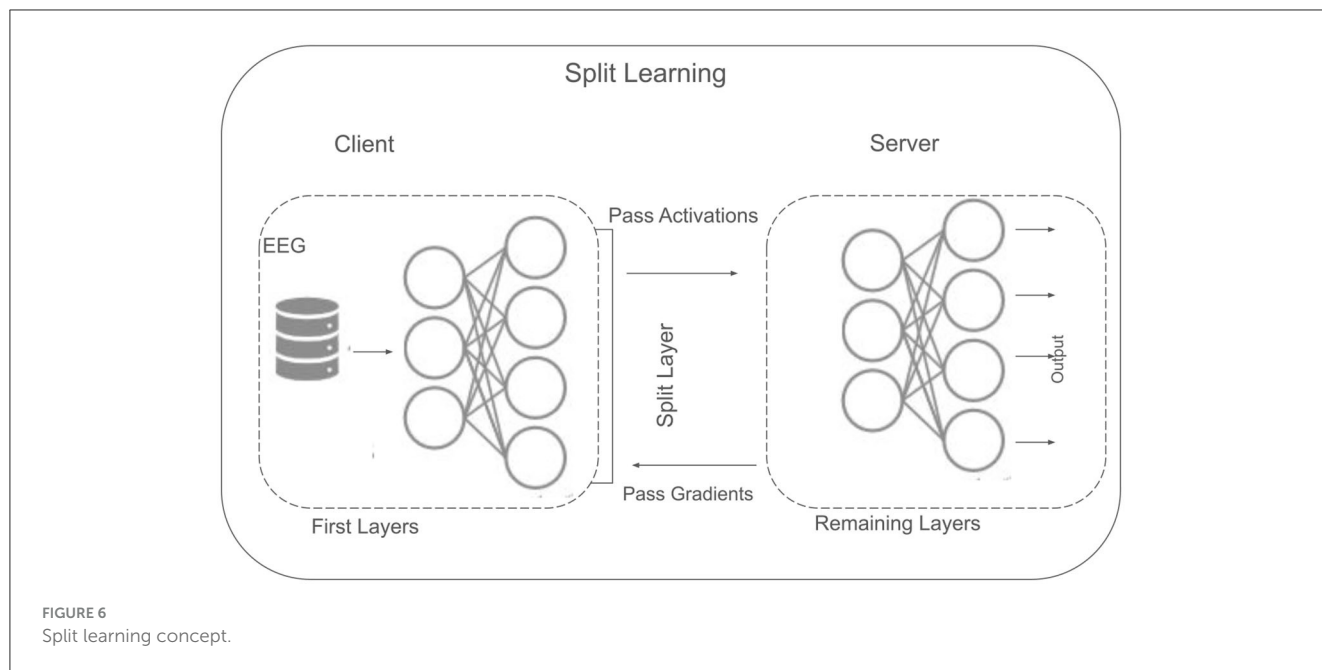
SL offers a collaborative training framework by partitioning a neural network between clients and a central server. This approach helps ensure that sensitive data like EEG signals remain local to each client, while still enabling the development of robust, shared models. In the context of MDD classification, SL architecture that we used is shown in Figure 6 that is particularly beneficial, as it enable data training while managing resources efficiently. Consider a neural network $f(\cdot)$ decomposed into two primary segments (as given in Equation 19). Where: f_{client} denotes the partial model residing on the client side, parametrized by θ , which transforms local data $\mathbf{x}$ into an intermediate representation $\mathbf{z}$. f_{server} denotes the remaining portion of the model, located on a central server and parametrized by ϕ . It processes the intermediate representation $\mathbf{z}$ to produce predictions (e.g., class probabilities). And, $\diamond$ symbolizes the functional concatenation of the two segments.

$$f(\mathbf{x}; \theta, \phi) = f_{\text{client}}(\mathbf{x}; \theta) \diamond f_{\text{server}}(\mathbf{z}; \phi) \quad (19)$$

Each client trains only f_{client} on its local dataset, while f_{server} is trained on the server side using the intermediate representations $\mathbf{z}$ received from the clients. This design ensures that raw EEG data never leaves the client's local environment. In utilized methodology for SL, each client i forwards only intermediate activations $\mathbf{z}$ derived from its local data to the server, which handles the remaining layers and calculates the global loss. The server's gradients are backpropagated to the clients, enabling local updates while preserving data privacy. This division of computational labor also alleviates resource constraints on client devices, as the heaviest computations can be offloaded to the server. This makes SL particularly applicable for MDD classification, where healthcare institutions typically hold proprietary EEG data. By sharing only intermediate features, SL mitigates privacy concerns and fosters collaborative model development, enabling a more inclusive and robust system for detecting and monitoring mental health conditions.

3.6.2 Algorithmic workflow for split learning

In this subsection, workflow of the utilized SL methodology has been described, as it starts with initialization, local processing and then toward clients processing and propagation, these steps are given as below:



3.6.2.1 Initialization

Each client C_i initializes its local model parameters θ_i , while the central server initializes its parameters ϕ . Data normalization or other preliminary setup is performed here.

3.6.2.2 Local preprocessing

Prior to training, each client cleans and preprocesses its local EEG data (e.g., filtering, artifact removal). This ensures high-quality input to the client-side model $f_{\text{client}}(\cdot; \theta_i)$.

3.6.2.3 Client forward pass

The client-side model f_{client} processes the local EEG data D_i to produce intermediate representations z_i . Because only z_i is shared, raw EEG data remains private.

3.6.2.4 Intermediate transmission

Clients transmit z_i to the central server. This step preserves data privacy, as the raw EEG signals never leave the local environment.

3.6.2.5 Server forward pass and loss computation

The central server processes all received activations $\{z_i\}$ using The server computes a global loss L by aggregating individual losses (e.g., cross-entropy) for each client's predictions y_i .

3.6.2.6 Backpropagation and parameter updates

Using the global loss L , the server performs backpropagation to update its parameters ϕ . By the chain rule, partial gradients are also computed and sent back to each client.

3.6.2.7 Client-side parameter updates

Upon receiving the relevant gradients, each client updates its local parameters θ_i . This allows clients to learn collaboratively without ever sharing raw EEG data.

3.6.2.8 Iteration and convergence

All previous steps (from local preprocessing to parameter updates) are repeated for multiple epochs. Once convergence is

reached, the final model consists of updated client-side parameters $\{\theta_i\}$ and server-side parameters ϕ .

3.6.2.9 Output

The trained SL model can be deployed for EEG classification. Each client retains its local model segment θ_i , while the server holds ϕ , ensuring continual privacy protection.

3.7 Evaluation metrics

Classification performance was evaluated using standard metrics derived from the confusion matrix in a binary classification setting (Healthy vs. MDD). Let TP (True Positive) be the number of MDD instances correctly classified, TN (True Negative) the number of Healthy instances correctly classified, FP (False Positive) the number of Healthy instances misclassified as MDD, and FN (False Negative) the number of MDD instances misclassified as Healthy. These values form the following 2×2 confusion matrix, from this matrix, the evaluation metrics are computed that are accuracy, precision, recall, and F1-score:

$$\begin{bmatrix} \text{TP} & \text{FP} \\ \text{FN} & \text{TN} \end{bmatrix}$$

3.7.1 Accuracy

Accuracy (Equation 20) measures the overall rate of correct predictions across all instances. It is the proportion of TP and TN from all predicted values by the model. It measures the proportion of instances that are correctly predicted out of the total number of predictions.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (20)$$

3.7.2 Precision

Precision is a crucial metric that quantifies the model's ability to correctly identify positive (MDD) cases among all predicted positives. As given in Equation 21, it is the ratio of TP to the sum of FP and TN

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (21)$$

3.7.3 Recall

Recall, sometimes referred to as sensitivity, measures the model's effectiveness at identifying all positive (MDD) instances in a dataset. Mathematically, as given in Equation 22, it is the ratio of TP to the sum of TN and FN.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (22)$$

3.7.4 F1-Score

The F1-Score provides a balanced assessment of a model's performance by combining both Precision and Recall into a single metric. Mathematically expressed in Equation 23, it is the harmonic mean of Precision and Recall. Unlike a simple arithmetic mean, the harmonic mean penalizes extreme values, ensuring both Precision and Recall share comparable significance in the final score.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (23)$$

3.7.5 Confusion matrix

Confusion matrix provides a visual overview of classification performance. It indicates how frequently the classifier confuses one class for the other, offering deeper insight into errors (FPs vs. FNs). For binary classification (Healthy vs. MDD), the matrix aids in diagnosing misclassification patterns and refining model strategies. All these metrics collectively form the classification report, enabling a comprehensive assessment of each model's performance in detecting MDD from EEG signals.

4 Results and discussion

In this section, we present a comprehensive evaluation of the proposed classification approaches for MDD detection. We analyze the performance of both ML and DL models, and additionally showcase an ensemble method that utilizes the SL framework. By assessing metrics such as accuracy, precision, recall, and F1-Score, we gain insight into each model's strengths and limitations.

4.1 ML models results

As discussed earlier, several ML models i.e., LR, RF, SVM, DT, KNN, and GB were utilized to classify MDD using EEG data. Table 1 presents their respective performances on the test set, along with best cross-validation (CV) scores and optimal hyperparameter configurations. The key findings for each model are summarized below.

4.1.1 LR model

Achieved a test accuracy of 92.41%, with F1-Scores of 0.9160 for the Healthy class and 0.9308 for the MDD class. Its best CV score was 0.8833. These results suggest that LR provides a stable generalization capability when distinguishing between Healthy and MDD samples. The best hyperparameter setting at C: 0.1 indicates a preference for regularization to control overfitting in high-dimensional EEG feature spaces.

4.1.2 RF model

Achieved a test accuracy of 100%, outperforming other ML models. Its best CV score was 0.9138. The selected hyperparameter (number of estimators 100) enable an ensemble of sufficiently large and diverse trees. Due to its strong performance, RF was chosen for the ensemble approach with Deep Learning models, as shown in Figure 7.

4.1.3 SVM model

Achieved an accuracy of 98.74%, indicating a clear separation between the two classes. Its F1-Scores of 0.9865 (Healthy) and 0.9882 (MDD) reflect the model's effectiveness. The best CV score was 0.9182, achieved with hyperparameter (C: 10, kernel: rbf). This shows that SVM is suitable for handling EEG data with potentially complex class boundaries.

4.1.4 DT model

Achieved an accuracy of 97.75%. Its best CV score was 0.8740. By employing a moderately deep tree with max depth of 10, the DT model partitions the EEG feature space effectively. Although decision trees can overfit, this depth appears to balance training accuracy and generalization for the MDD classification task.

4.1.5 KNN model

Achieved an accuracy of 100%, similar to the RF model. Its best CV score was 0.8713. The chosen hyperparameters number of neighbors: 7, weights: distance uses distance-based weighting in separable EEG clusters. However, KNN can be computationally expensive at inference time and typically requires extensive parameter tuning for integration with DL pipelines, so it was not selected for the ensemble stage.

4.1.6 GB model

Achieved an accuracy of 99.35%, with a best CV score of 0.9184. It iteratively refined weak learners using a learning rate of 0.2 and 100 estimators. Its F1-Scores of 0.9931 (Healthy) and 0.9939 (MDD) indicate that boosting rounds improve classification by reducing both bias and variance.

Table 1 shows that all models attain high classification performance. RF and KNN reach 100% accuracy on the test set, while SVM, DT, LR, and GB also present strong results. The consistent F1-Scores reinforce the effectiveness of EEG features for detecting MDD.

TABLE 1 Performance of Various Machine Learning Models for MDD Classification.

Model	Best score (CV)	Best params	Accuracy	F1 (Healthy)	F1 (MDD)	TP	FP	FN	TN
Logistic regression	0.8833	{C: 0.1}	0.9241	0.9160	0.9308	9,382	1,022	374	7,609
Random forest	0.9138	{max_depth: None, n_estimators: 100}	1.0000	1.0000	1.0000	9,756	0	0	8,631
SVM	0.9182	{C: 10, kernel: rbf}	0.9874	0.9865	0.9882	9,699	175	57	8,456
Decision tree	0.8740	{max_depth: 10}	0.9775	0.9759	0.9790	9,606	263	150	8,368
K-Nearest Neighbors	0.8713	{n_neighbors: 7, weights: distance}	1.0000	1.0000	1.0000	9,756	0	0	8,631
Gradient Boosting	0.9184	{learning_rate: 0.2, n_estimators: 100}	0.9935	0.9931	0.9939	9,721	84	35	8,547

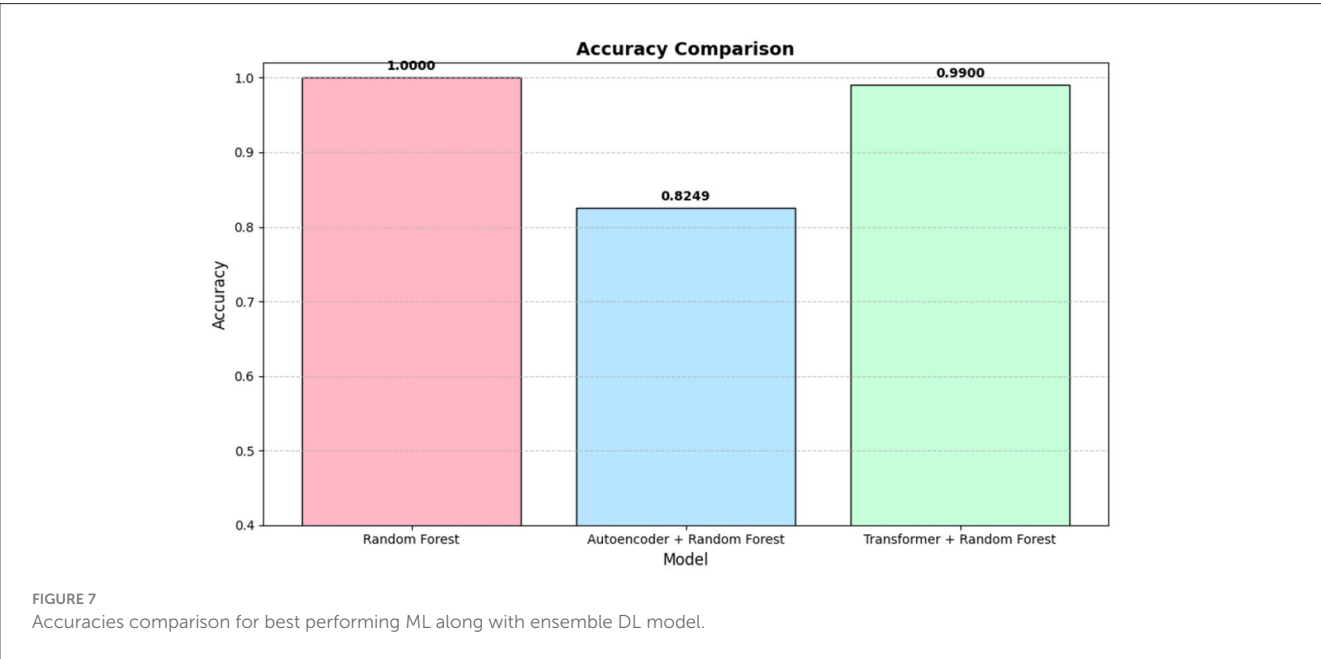


TABLE 2 4-Fold cross-validation accuracies for each classifier.

Model	Fold 1	Fold 2	Fold 3	Fold 4	Mean
Logistic regression	0.92	0.88	0.93	0.90	0.91
Random forest	1.00	0.99	0.96	1.00	0.98
SVM	0.95	0.94	0.98	0.97	0.96
Decision tree	0.87	0.88	0.90	0.86	0.88
K-Nearest Neighbors	0.99	0.96	0.95	0.98	0.97
Gradient boosting	0.93	0.94	0.95	0.92	0.93

4.1.7 K Fold cross validation results

Table 2 shows the accuracy for each classifier across four folds of cross-validation. The Mean column reports the average accuracy across all folds. By separating the data into distinct training/validation splits for each fold, we reduce the risk of overfitting and obtain a more realistic estimate of out-of-sample performance.

4.2 DL models performances along with ensemble learning

After training an autoencoder to learn compact EEG representations, multiple classifiers were evaluated on these latent features. Table 3 summarizes the results for both a baseline autoencoder-only ensemble and five conventional ML algorithms trained on autoencoder outputs. Each row reports the overall accuracy as well as precision, recall, and F1-scores for both classes (Healthy and MDD).

4.2.1 Discussion of autoencoder-based results

Table 3 demonstrates that using autoencoder-derived representations yield competitive performance across multiple classifiers. The baseline ensemble (first row) provides a moderate accuracy of 0.6884, indicating that unsupervised feature extraction alone captures some discriminative patterns.

RF and SVM show the highest accuracies (over 0.82), suggesting that tree-ensemble and margin-based methods effectively exploit these latent features. K-Nearest Neighbors and

TABLE 3 Classification performance on autoencoder and with ensemble autoencoder.

Method	Accuracy	Healthy			MDD		
		Precision	Recall	F1	Precision	Recall	F1
Autoencoder (baseline)	0.6884	0.7031	0.5817	0.6367	0.6791	0.7828	0.7273
Autoencoder + random forest	0.8249	0.9321	0.6761	0.7837	0.7696	0.9565	0.8529
SVM	0.8222	0.9061	0.6929	0.7853	0.7752	0.9365	0.8483
Autoencoder + decision tree	0.6833	0.6947	0.5800	0.6321	0.6759	0.7746	0.7219
Autoencoder + K-Nearest Neighbors	0.7692	0.7627	0.7375	0.7499	0.7745	0.7971	0.7857
Autoencoder + gradient boosting	0.7735	0.8187	0.6645	0.7336	0.7457	0.8699	0.8030

TABLE 4 Classification performance on transformer and ensemble models.

Method	Accuracy	Healthy			MDD		
		Precision	Recall	F1	Precision	Recall	F1
Transformer (baseline)	0.9000	0.9100	0.8800	0.8950	0.9000	0.9200	0.9100
Transformer + decision tree	0.8800	0.8900	0.8600	0.8750	0.8700	0.8900	0.8800
Transformer + K-Nearest Neighbors	0.9200	0.9100	0.9200	0.9150	0.9300	0.9200	0.9250
Transformer + SVM	0.9300	0.9400	0.9200	0.9300	0.9300	0.9400	0.9350
Transformer + gradient boosting	0.9500	0.9500	0.9400	0.9450	0.9400	0.9500	0.9450
Transformer + random forest	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900

Gradient Boosting also achieve an accuracies of approximately 0.77, while the single DT model exhibits lower performance (0.68) relative to ensemble approaches. RF high precision for Healthy (0.9321) and recall for MDD (0.9565) underline its balanced detection capabilities in this context.

4.2.2 Transformer-based classification

As transformer model is utilized to capture long-range dependencies in EEG signals, several classifiers were applied to the Transformer outputs for final predictions as well. Table 4 summarizes the results, including a standalone Transformer baseline and five conventional ML classifiers. The table reports overall accuracy, alongside precision, recall, and F1-scores for the two classes (Healthy vs. MDD). Their detailed results discussion has been given in Section 4.2.3.

4.2.3 Discussion of transformer-based results

In this subsection the results achieved for ensemble learning has been discussed, as we utilized transformers along with ML models and these has been given in Table 4 that shows the classification performance of the baseline Transformer model and its combinations with different ML classifiers. The standalone Transformer (Baseline) achieves an accuracy of 0.90, with 0.91 precision, 0.88 recall, and 0.895 F1 for the Healthy class, and 0.90 precision, 0.92 recall, and 0.91 F1 for the MDD class. These results indicate that the Transformer can extract features from EEG signals that help differentiate between Healthy and MDD instances.

Transformer + DT yields an accuracy of 0.88. For the Healthy class, it achieves 0.89 precision, 0.86 recall, and 0.875 F1, while for

the MDD class it attains 0.87 precision, 0.89 recall, and 0.88 F1. Even though this is lower than some other combinations, it still shows reasonable performance compared to traditional EEG-based methods.

Transformer + KNN reports an accuracy of 0.92. The Healthy class has 0.91 precision, 0.92 recall, and 0.915 F1, and the MDD class has 0.93 precision, 0.92 recall, and 0.925 F1. These numbers suggest that local distance-based methods can work well when applied to Transformer outputs.

Transformer + SVM achieves an accuracy of 0.93. For the Healthy class, precision, recall, and F1 are 0.94, 0.92, and 0.93, respectively, while for the MDD class they are 0.93, 0.94, and 0.935. This indicates that margin-based classification benefits from sequence-aware features extracted by the Transformer.

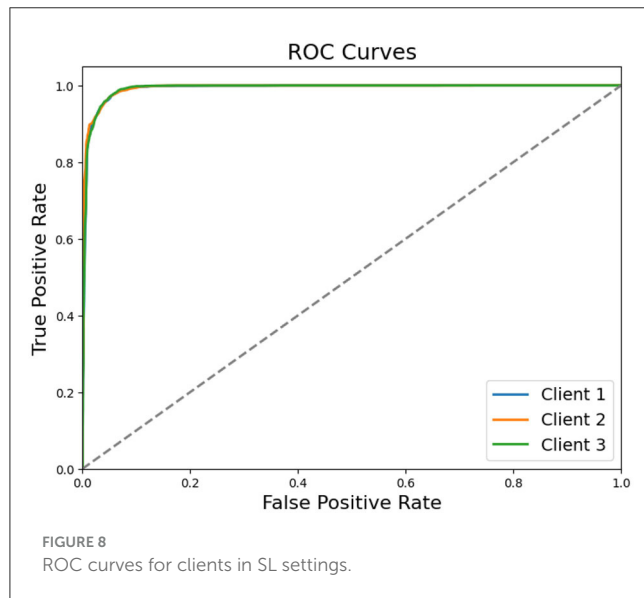
Transformer + GB attains an accuracy of 0.95. Its Healthy metrics are 0.95 precision, 0.94 recall, and 0.945 F1, and its MDD metrics are 0.94 precision, 0.95 recall, and 0.945 F1. This suggests that boosting rounds are effective at refining the latent representations provided by the Transformer.

Transformer + RF achieves the highest accuracy of 0.99. Precision, recall, and F1 for both Healthy and MDD classes are all 0.99, showing that the ensemble of decision trees makes good use of attention-based features.

Thus, combining the Transformer with robust classification algorithms enhances performance compared to the baseline. The best results come from pairing the Transformer with RF, followed by GB, SVM, KNN, and DT. These findings illustrate that attention-based feature extraction can improve EEG-based MDD classification when integrated with well-chosen ML methods.

TABLE 5 Performance of split learning across three clients.

Client	Accuracy	Precision	Recall	F1-Score	Healthy		MDD	
					Correct	Misclass.	Correct	Misclass.
Client 1	0.9574	0.9577	0.9574	0.9574	2,744	172	3,124	89
Client 2	0.9623	0.9625	0.9623	0.9623	2,679	148	3,219	83
Client 3	0.9543	0.9549	0.9543	0.9543	2,691	197	3,158	83



4.3 Split learning results

SL framework was implemented across three clients, each training local Transformer-based encoders whose latent representations were subsequently processed by a RF classifier on the server side. Table 5 shows the key performance metrics (Accuracy, Precision, Recall, and F1-Score) for each client, alongside the main confusion matrix values (correct vs. misclassified instances of Healthy and MDD). The average inference time per client was measured at 2.0866 seconds.

To quantify inference time, we define the total inference time for a single sample on the i -th client as given in Equation 24:

$$T_{\text{inference}}^{(i)} = T_{\text{local}}^{(i)} + T_{\text{transfer}}^{(i)} + T_{\text{server}}, \quad (24)$$

where $T_{\text{local}}^{(i)}$ is the local forward pass time through the Transformer on client i , $T_{\text{transfer}}^{(i)}$ is the latency for transmitting the latent representation to the server, and T_{server} is the server-side classification time using the RF model. The average inference time $T_{\text{inference}}$ across all k clients mathematically is given in Equation 25.

$$T_{\text{inference}} = \frac{1}{k} \sum_{i=1}^k T_{\text{inference}}^{(i)}. \quad (25)$$

Thus got an average $T_{\text{inference}}$ of 2.0866 seconds. This end-to-end metric reflects the time from when an EEG sample arrives at the client to when the final classification outcome is returned, including both local and server-side computations.

4.3.1 Discussion of split learning results

Table 5 illustrates that all three clients attain high classification accuracy, exceeding 95%. Client 2 achieves the best overall accuracy of 0.9623, closely followed by Client 1 (0.9574) and Client 3 (0.9543). Precision and Recall remain closely aligned for each client, reflecting a balanced ability to detect both Healthy and MDD classes. Confusion matrix counts indicate that relatively few Healthy samples are misclassified as MDD and vice versa. ROC curve shown in Figure 8 also reflects that each client achieved higher true positive rate showing their ability and reliability.

These findings shows that a SL approach, utilized with a transformer architecture for local feature extraction and RF model for final classification, can maintain robust performance while preserving data privacy. Additionally, the measured average inference time of 2.0866 seconds per client suggests that this collaborative framework is computationally feasible for real-world EEG based mental health applications.

While these performance metrics are promising, practical deployment on devices with limited compute capabilities (e.g., mobile EEG headsets, embedded healthcare systems) demands additional optimization. Because SL partitions the model into client-side and server-side segments, heavier computations—such as the Transformer's attention blocks—are executed on the server, reducing on-device resource usage. Future work will involve benchmarking these strategies across diverse hardware platforms to quantify improvements in latency, memory use, and power efficiency.

5 Conclusion

This work presented an effective methodology for major depressive disorder classification by integrating advanced EEG feature extraction, ensemble models, and split learning to safeguard privacy. In conventional centralized experiments, RF, KNN, and GB achieved commendable performance, while a Transformer-RF ensemble model achieved 99% accuracy. Autoencoder-based feature learning provided notable results, illustrating that unsupervised approaches can be profitably combined with supervised classifiers. Crucially, the split learning implementation validated the feasibility of decentralizing training:

three distinct clients each achieved over 95% accuracy, with minimal performance trade-offs relative to centralized schemes. By maintaining data on local nodes and exchanging only intermediate representations, the framework supported institutional privacy requirements while offering robust classification outcomes. Future investigations may include refining model architectures for improved efficiency, exploring additional neurophysiological data modalities, and extending the approach to multi-disorder classification scenarios, thereby broadening the applicability of privacy-preserving, high-performance EEG analytics in clinical settings.

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. The datasets analyzed and utilized for this study can be found at DOI: [10.6084/m9.figshare.4244171.v2](https://doi.org/10.6084/m9.figshare.4244171.v2).

Ethics statement

The studies involving humans were approved by the ethics committee, Hospital Universiti Sains Malaysia (HUSM), Malaysia. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

MU: Methodology, Software, Writing – original draft. JA: Investigation, Validation, Writing – review & editing. NA: Formal analysis, Funding acquisition, Project administration, Writing – review & editing. OS: Formal analysis, Funding acquisition, Project administration, Writing – review & editing. MH: Investigation, Validation, Writing – review & editing, Funding acquisition. AK: Investigation, Validation, Writing – review & editing. MK: Conceptualization, Methodology, Supervision, Writing – original draft.

References

- Ahmed, T., Ivan, S., Munir, A., and Ahmed, S. (2024). Decoding depression: analyzing social network insights for depression severity assessment with transformers and explainable AI. *Nat. Lang. Proc. J.* 7:100079. doi: 10.1016/j.nlp.2024.100079
- Anik, I. A., Kamal, A. H. M., Kabir, M. A., Uddin, S., and Moni, M. A. (2024). A robust deep-learning model to detect major depressive disorder utilizing EEG signals. *IEEE Trans. Artif. Intell.* 5, 4938–4947. doi: 10.1109/TAI.2024.3394792
- de, S., Singh, A., Tiwari, V., Patel, H., Vivekananda, G. N., and Singh Rajput, D. (2024). Slitrant: an EEG-based automated diagnosis framework for major depressive disorder monitoring using a novel LGCN and transformer-based hybrid deep learning approach. *IEEE Access* 12, 173109–173126. doi: 10.1109/ACCESS.2024.3493140
- Diehl, P. U., and Cook, M. (2015). Unsupervised learning of digit recognition using spike-timing-dependent plasticity. *Front. Comput. Neurosci.* 9:99. doi: 10.3389/fncom.2015.00099
- Earl, E. H., Goyal, M., Mishra, S., Kannan, B., Mishra, A., Chowdhury, N., et al. (2024). EEG based functional connectivity in resting and emotional states may identify major depressive disorder using machine learning. *Clin. Neurophysiol.* 164, 130–137. doi: 10.1016/j.clinph.2024.05.017
- Gour, N., Hassan, T., Owais, M., Ganapathi, I. I., Khanna, P., Seghier, M. L., et al. (2023). Transformers for autonomous recognition of psychiatric dysfunction via raw and imbalanced EEG signals. *Brain Inform.* 10:25. doi: 10.1186/s40708-023-00201-y

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work is funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R760), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through a Small Group Research Project under grant number RGP1/405/44.

Acknowledgments

This work is funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R760), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through a Small Group Research Project under grant number RGP1/405/44.

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.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

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.

- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2013). MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7:267. doi: 10.3389/fnins.2013.00267
- Gupta, O., and Raskar, R. (2018). Distributed learning of deep neural network over multiple agents. *J. Netw. Comput. Applic.* 116, 1–8. doi: 10.1016/j.jnca.2018.05.003
- Hagan, A. T., Xu, L., Klugah-Brown, B., Li, J., Jiang, X., and Kendrick, K. M. (2025). The pharmacodynamic modulation effect of oxytocin on resting state functional connectivity network topology. *Front. Pharmacol.* 15:1460513. doi: 10.3389/fphar.2024.1460513
- Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. *Front. Hum. Neurosci.* 3:31. doi: 10.3389/neuro.09.031.2009
- Ilias, L., Mouzakitis, S., and Askounis, D. (2024). Calibration of transformer-based models for identifying stress and depression in social media. *IEEE Trans. Comput. Soc. Syst.* 11, 1979–1990. doi: 10.1109/TCSS.2023.3283009
- Jia, Y., Liu, B., Zhang, X., Dai, F., Khan, A., Qi, L., et al. (2024). Model pruning-enabled federated split learning for resource-constrained devices in artificial intelligence empowered edge computing environment. *ACM Trans. Sen. Netw.* doi: 10.1145/3687478
- Kreiviniene, B., Šaltyt, E., Vaisiauskė, L., and Mačiulskytė, S. (2025). Therapeutic effect of proprioceptive dolphin assisted activities on health-related quality of life and muscle tension, biomechanical and viscoelastic properties in major depressive disorder adults: case analysis. *Front. Hum. Neurosci.* 18:1487293. doi: 10.3389/fnhum.2024.1487293
- Lesar, M., Sajovic, J., Novaković, D., Primožič, M., Vetrh, E., Sajovic, M., et al. (2025). The complexity of caffeine's effects on regular coffee consumers. *Heliyon* 11:e41471. doi: 10.1016/j.heliyon.2024.e41471
- Li, J., Jiang, M., Qin, Y., Zhang, R., and Ling, S. H. (2023). Intelligent depression detection with asynchronous federated optimization. *Complex Intell. Syst.* 9, 115–131. doi: 10.1007/s40747-022-00729-2
- Liang, A., Zhao, S., Song, J., Zhang, Y., Zhang, Y., Niu, X., et al. (2021). Treatment effect of exercise intervention for female college students with depression: analysis of electroencephalogram microstates and power spectrum. *Sustainability* 13:6822. doi: 10.3390/su13126822
- McMahan, B., Moore, E., Ramage, D., Hampson, S., and Arcas, B. A. (2017). "Communication-efficient learning of deep networks from decentralized data, in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, volume 54 of Proceedings of Machine Learning Research*, eds. A. Singh, and J. Zhu (PMLR), 1273–1282.
- Metin, B., Uyulan, C., Ergüzel, T. T., Farhad, S., Cifci, E., Türk, Ö., et al. (2024). The deep learning method differentiates patients with bipolar disorder from controls with high accuracy using EEG data. *Clin. EEG Neurosci.* 55, 167–175. doi: 10.1177/15500594221137234
- Mumtaz, W. (2016). *MDD Patients and Healthy Controls EEG Data (New)*. doi: 10.6084/m9.figshare.4244171.v2
- Mumtaz, W., Xia, L., Ali, S. S. A., Yasin, M. A. M., Hussain, M., and Malik, A. S. (2017). Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (mdd). *Biomed. Signal Process. Control* 31, 108–115. doi: 10.1016/j.bspc.2016.07.006
- Owais, M., Sultan, H., Baek, N. R., Lee, Y. W., Usman, M., Nguyen, D. T., et al. (2022). Deep 3D volumetric model genesis for efficient screening of lung infection using chest CT scans. *Mathematics* 10:4160. doi: 10.3390/math10214160
- Park, S. M., Jeong, B., Oh, D. Y., Choi, C.-H., Jung, H. Y., Lee, J.-Y., et al. (2021). Identification of major psychiatric disorders from resting-state electroencephalography using a machine learning approach. *Front. Psychiatry* 12:707581. doi: 10.3389/fpsyt.2021.707581
- Perrotelli, A., Giordano, G. M., Brando, F., Giuliani, L., and Mucci, A. (2021). EEG-based measures in at-risk mental state and early stages of schizophrenia: a systematic review. *Front. Psychiatry* 12:653642. doi: 10.3389/fpsyt.2021.653642
- Rafiei, A., Zahedifar, R., Sitaula, C., and Marzbanrad, F. (2022). Automated detection of major depressive disorder with EEG signals: a time series classification using deep learning. *IEEE Access* 10, 73804–73817. doi: 10.1109/ACCESS.2022.3190502
- Rahul, J., Sharma, D., Sharma, L. D., Nanda, U., and Sarkar, A. K. (2024). A systematic review of EEG based automated schizophrenia classification through machine learning and deep learning. *Front. Hum. Neurosci.* 18:1347082. doi: 10.3389/fnhum.2024.1347082
- Rivera, M. J., Teruel, M. A., Mate, A., and Trujillo, J. (2022). Diagnosis and prognosis of mental disorders by means of EEG and deep learning: a systematic mapping study. *Artif. Intell. Rev.* 55, 1209–1251. doi: 10.1007/s10462-021-09986-y
- Subhani, A. R., Mumtaz, W., Saad, M. N. B. M., Kamel, N., and Malik, A. S. (2017). Machine learning framework for the detection of mental stress at multiple levels. *IEEE Access* 5, 13545–13556. doi: 10.1109/ACCESS.2017.2723622
- Sultan, H., Owais, M., Nam, S. H., Haider, A., Akram, R., Usman, M., et al. (2023). MDFU-net: Multiscale dilated features up-sampling network for accurate segmentation of tumor from heterogeneous brain data. *J. King Saud Univ. Comput. Inf. Sci.* 35:101560. doi: 10.1016/j.jksuci.2023.101560
- Sun, H., Chen, Y.-W., and Lin, L. (2023). Tensorformer: a tensor-based multimodal transformer for multimodal sentiment analysis and depression detection. *IEEE Trans. Affect. Comput.* 14, 2776–2786. doi: 10.1109/TAFFC.2022.3233070
- Umair, M., Khan, M. S., Ahmed, F., Baothman, F., Alqahtani, F., Alian, M., et al. (2021). Detection of covid-19 using transfer learning and grad-cam visualization on indigenously collected x-ray dataset. *Sensors* 21:5813. doi: 10.3390/s21175813
- Umair, M., Tan, W.-H., and Foo, Y.-L. (2023). "Challenges in federated learning for resource-constrained IoT environments: energy efficiency, privacy, and statistical heterogeneity, in *2023 IEEE 8th International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, 1–6. doi: 10.1109/ICRAIE59459.2023.10468189
- Umair, M., Tan, W.-H., and Foo, Y.-L. (2024). "Dynamic federated learning aggregation for enhanced intrusion detection in IoT attacks, in *2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, 524–529. doi: 10.1109/ICAIIIC60209.2024.10463247
- Vohryzek, J., Sanz-Perl, Y., Kringelbach, M. L., and Deco, G. (2025). Human brain dynamics are shaped by rare long-range connections over and above cortical geometry. *Proc. Nat. Acad. Sci.* 122:e2415102122. doi: 10.1073/pnas.2415102122
- Wang, Y., Zhao, S., Jiang, H., Li, S., Luo, B., Li, T., et al. (2024). Diffmdd: a diffusion-based deep learning framework for MDD diagnosis using EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* 32, 728–738. doi: 10.1109/TNSRE.2024.3360465
- Zhang, C., Meng, X., Liu, Q., Wu, S., Wang, L., and Ning, H. (2023). Fedbrain: a robust multi-site brain network analysis framework based on federated learning for brain disease diagnosis. *Neurocomputing* 559:126791. doi: 10.1016/j.neucom.2023.126791
- Zhu, F., Zhang, J., Dang, R., Hu, B., and Wang, Q. (2025). Mtnet: multimodal transformer network for mild depression detection through fusion of EEG and eye tracking. *Biomed. Signal Process. Control* 100:106996. doi: 10.1016/j.bspc.2024.106996
- Zhu, Y., Ma, J., Li, Y., Gu, M., Feng, X., Shao, Y., et al. (2024). Adenosine-dependent arousal induced by astrocytes in a brainstem circuit. *Adv. Sci.* 11:2407706. doi: 10.1002/advs.202407706



OPEN ACCESS

EDITED BY
Saad Arif,
King Faisal University, Saudi Arabia

REVIEWED BY
Farman Ali,
Sungkyunkwan University, Republic of Korea
Zeba Idrees,
University of Alberta, Canada

*CORRESPONDENCE
Abdullah Aziz
✉ abdullah.aziz@umu.se

RECEIVED 11 March 2025
ACCEPTED 14 May 2025
PUBLISHED 06 June 2025

CITATION
Asiri F, Al Malwi W, Zhukabayeva T, Nafea I,
Aziz A, Gazem NA and Qayyum A (2025)
Enhancing medical image privacy in IoT with
bit-plane level encryption using chaotic map.
Front. Comput. Neurosci. 19:1591972.
doi: 10.3389/fncom.2025.1591972

COPYRIGHT
© 2025 Asiri, Al Malwi, Zhukabayeva, Nafea,
Aziz, Gazem and Qayyum. 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.

Enhancing medical image privacy in IoT with bit-plane level encryption using chaotic map

Fatima Asiri¹, Wajdan Al Malwi¹, Tamara Zhukabayeva²,
Ibtehal Nafea³, Abdullah Aziz^{4*}, Nadhmi A. Gazem⁵ and
Abdullah Qayyum⁶

¹Informatics and Computer Systems Department, College of Computer Science, King Khalid University, Abha, Saudi Arabia, ²Department of Information Systems, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, ³College of Computer Science and Engineering, Taibah University, Medina, Saudi Arabia, ⁴High Performance Computing Centre North, Umeå University, Umeå, Västerbotten, Sweden, ⁵Department of Information Systems, College of Business Administration-Yanbu, Taibah University, Medina, Saudi Arabia, ⁶6G Lab, School of Engineering and Informatics University of Sussex, Brighton, United Kingdom

Introduction: Preserving privacy is a critical concern in medical imaging, especially in resource limited settings like smart devices connected to the IoT. To address this, a novel encryption method for medical images that operates at the bit plane level, tailored for IoT environments, is developed.

Methods: The approach initializes by processing the original image through the Secure Hash Algorithm (SHA) to derive the initial conditions for the Chen chaotic map. Using the Chen chaotic system, three random number vectors are generated. The first two vectors are employed to shuffle each bit plane of the plaintext image, rearranging rows and columns. The third vector is used to create a random matrix, which further diffuses the permuted bit planes. Finally, the bit planes are combined to produce the ciphertext image. For further security enhancement, this ciphertext is embedded into a carrier image, resulting in a visually secured output.

Results: To evaluate the effectiveness of our algorithm, various tests are conducted, including correlation coefficient analysis ($CC < 0$ or negative), histogram analysis, key space [$(10^{90})^8$] and sensitivity assessments, entropy evaluation [$E(S) > 7.98$], and occlusion analysis.

Conclusion: Extensive evaluations have proven that the designed scheme exhibits a high degree of resilience to attacks, making it particularly suitable for small IoT devices with limited processing power and memory.

KEYWORDS

image encryption, Chen chaotic map, chaos, meaningful encryption, bit-level encryption, IoT

1 Introduction

The Internet of Things (IoT) connects devices and objects via the Internet, whether wirelessly or wired. In recent years, the concept has become increasingly popular as it is used for various purposes, including business development, transportation, education, and communication. The hyper-connectivity created by the IoTs enables individuals and organizations to communicate seamlessly from a distance (Porras et al., 2018). IoT has been widely embraced in a wide range of industries, including e-health, manufacturing, smart cities, agribusiness, and home automation. According to Cisco, Internet-connected gadgets will number approximately 500 billion by 2030 (Aman et al., 2020). As IoT advances exponentially, medical imaging and data have become more widely used, and are therefore need to be secured before being shared.

Medical images have become increasingly important in diagnosing and treating illnesses. The visuals are used directly by doctors during the evaluation and therapy of patients (Ismail et al., 2018). For medical applications, securing the transmission and storage of medical images has become increasingly important due to their containment of private information (Ye and Huang, 2015; Dridi et al., 2016; Al-Haj et al., 2015; Cao et al., 2017; Khan et al., 2018; Hu et al., 2024; Chu et al., 2024; Belazi et al., 2019). A number of academics have therefore focused on developing methods to secure images in IoT applications. The authors in Ye and Huang (2015) utilized logistic and Arnold chaotic maps to design an autoblocking and Electrocardiography (ECG) signal-based medical image encryption scheme. ECG signals and the Wolf algorithm calculates initial conditions for the chaotic system. A key characteristic of this cryptoarchitecture is that it performs autoblocking diffusion only during the encryption phase of the process, in contrast to traditional cryptoarchitectures. A new chaos and neural network-based medical image encryption scheme has been presented in Dridi et al. (2016). Plaintext image pixels are XORed with a generated key. The weight and bias values for neural networks have been computed using the Logistic map. By using this technique, medical images can be made more secure with a simpler algorithm than current ones. Using Strong cryptographic functions with internal symmetric keys and hash codes, the author designed an encryption scheme for medical images that ensures confidentiality, authenticity, and integrity (Al-Haj et al., 2015). With the whirlpool hash function and Galois counter mode, advanced encryption standards are used to secure confidentiality and authenticity, while digital signature algorithms employ elliptic curves to secure integrity and authenticity. The edge map-based medical encryption scheme has been presented in Cao et al. (2017). It consists of three main steps: (a) extraction of bit planes, (b) generating random numbers, and (c) permutations. The source image can be any type of image and distinct edge maps can be produced by varying edge detection approaches and thresholds, depending on the source image type. An Intertwining Logistic map and Deoxyribonucleic acid (DNA) are utilized by Khan et al. to protect medical images (Khan et al., 2018). A DNA sequence is passed through SHA-512 in order to calculate the chaotic system's initial condition. Plaintext pixel correlations are broken down through shuffling. In addition to XORing, an affine transformation is also applied to diffuse the shuffled pixels. A two-round medical encryption scheme is designed by Belazi et al. by combining chaos and DNA (Belazi et al., 2019). During each round, six steps are performed, namely block permutation, pixel substituting, DNA encoding, bit substitution, DNA decoding, and bit diffusion.

As Internet-related technologies continue to grow exponentially, new technologies, energy, or modifications are added daily. Applications and systems that use the Internet of Things benefit greatly from the recent advancements in wireless technology from 1G to 5G (Hasan et al., 2021). In recent years, high-quality medical care has become increasingly important as a result of population growth, urbanization, and the COVID-19 pandemic (Trujillo-Toledo et al., 2023). In medical diagnostics, X-rays, Computer Tomography Scans (CT scans), nuclear medicine imaging, and ultrasounds are modern imaging techniques. Thus, these high-resolution diagnostic images need to be secured before being exchanged. Recently, cyber attacks could make healthy

patients appear sick and vice versa. Therefore, cyber-security threats will increase alarmingly in the area of medical image communication. It is therefore increasingly important to have fast and secure cyber-security systems regarding the diagnosis of medical images (Kester et al., 2015). The Internet of Medical Things (IoMT) can provide many advantages to hospitals and healthcare organizations. However, they need to ensure that the right policies and protocols are in place to tackle the security challenges posed by IoMT. Researchers are curious about the potential security and privacy issues associated with this concept, particularly when bandwidth and frequency are high. Therefore, it is essential to design a robust medical image encryption scheme to guarantee the safe and trustworthy transmission and receipt of patients' symptomatic data through IoT. Double permutation techniques are used in Hasan et al. (2021) to design a lightweight, efficient encryption algorithm to protect healthcare images. In this method, plaintext images are broken down into blocks and encrypted. A chaotic encryption technique, based on the Message Queuing Telemetry Transport (MQTT) protocol, is proposed in this research for enhancing security and secrecy when transmitting medical images over the Healthcare Internet of Things (H-IoT) network (Trujillo-Toledo et al., 2023). Initially, chaotic maps are enhanced and applied to encrypt plaintext pixels through diffusion. The designed scheme efficiency is confirmed via a number of tests. The designed embedded medical cryptosystems transmit real-time medical images over the Internet and WiFi, thus enhancing real-time medical image security. Using multiple chaotic maps, the authors propose Multiple Map Chaos Based Image Encryption (MMCBIE) scheme, a novel method that encrypts images in the IoT environment (Jain et al., 2024). Unlike existing schemes, MMCBIE combines multiple chaotic maps, like Henon and 2D Logistic chaotic maps in a unique combination. According to security assessments and cryptanalysis, MMCBIE possesses high-level security properties, making it a superior method of image encryption. Hanchate and Anandan (2024) presented a hybrid scheme that combines Adaptive Elliptic Curve Cryptography (AECC) and Logistic mapping to encrypt medical images for the IoT. As a first step, the image is encrypted using the AECC technique, then again encrypted using the logistic map-based DNA sequence algorithm for greater security. The diffused DNA matrix is then decoded to produce the cipher image. The plain image determines the rules for encoding and decoding DNA as well as the key matrix. Liu et al. (2024) utilize compressive sensing (CS) and chaotic systems to design an encryption scheme for IoT scenarios to ensure security and efficiency. A chaotic laser system generates Masuemet matrices with complex phase space. The measurement matrices are further enhanced through the use of cyclic matrix methods. The image reconstruction quality is further improved using segmented linear thresholding. Further, large images are compressed block-wise in order to reduce storage space and improve reconstruction efficiency. The authors in Nadhan and Jacob (2024) investigated how a cryptography-based network might be able to encode medical images, as well as how deep learning could be used to ensure that the images are transmitted safely. Various image representations have been mapped using the ResNet-50 architecture. As a result of the extensive empirical setup and the security analysis, the suggested method is likely to provide unprecedented levels of

security. An IWT-based DNA encoding scheme is proposed to encrypt medical images within the Healthcare IoT (Lai and Hua, 2025). Random sequences were generated using a 3D hyperchaotic map. In addition to IWT, a novel diffusion algorithm masks critical information by generating approximation components. Bit-level permutations further enhance encryption complexity. The scheme further uses the DNA shuffle technique and encrypts the permuted images using a DNA-encoding technique to enhance security.

Most traditional image encryption algorithms convert plain images into noise-like ciphers, making them easily detectable and vulnerable to attack during transmission or storage. Visual security should be considered when designing an image encryption method to avoid hackers' attention. Therefore, to avoid the eavesdropper's attention, meaningful image encryption algorithms must be developed that may generate visually meaningful ciphertext images. Image encryption algorithms that provide a visual sense of meaning have attracted considerable research attention (Khan et al., 2024, 2020; Gan et al., 2024; Sathananthavathi et al., 2024; Zhang et al., 2024). A bit plane image encryption scheme was designed by Khan et al. (2024) using hash function and chaos theory. A SHA-512 hash algorithm is used to compute the key for the chaotic map. The chaotic random vectors are used to shuffle the plaintext image pixels row- and column-wise, while the random matrix is used for XOR-based diffusion. By embedding the noise-like ciphered text within a host image, a visually secure ciphertext image has been generated. The authors in Khan et al. (2020), presented a chaotic visual selective image encryption scheme. The key for the scheme has been derived from the DNA and plaintext image. The system key space is increased by using three different chaotic 1D maps. The original image is divided into blocks of varying sizes. Blocks with correlation coefficients above a predefined threshold are XORed with random matrices. The diffused blocks are then permuted to break the correlation between pixel values. As a final step, the ciphertext is encapsulated in a carrier image to create a visually secure ciphertext image.

Contribution

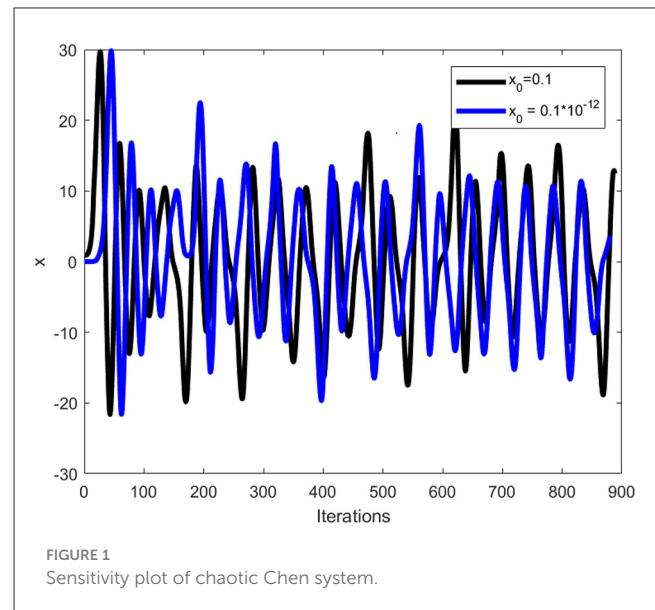
- The enhanced medical image encryption scheme has confusion and diffusion characteristics, making it ideal for the IoT environment.
- This scheme resists classical attacks due to its reliance on plaintext images as keys.
- To avoid attackers' attention, ciphered images are embedded in carrier images to produce visually secure images.

The remaining article is organized as follows: Section 2 discusses the preliminaries; Section 3 outlines the proposed methodology; the result analysis of the proposed work is provided in Section 4. Conclusion is provided in Section 5.

2 Preliminaries

2.1 Chaotic Chen system

Using simple state feedback, Chen developed a new 3D chaotic system in 1999 [1]. Similarly to the Lorenz system, Chen's



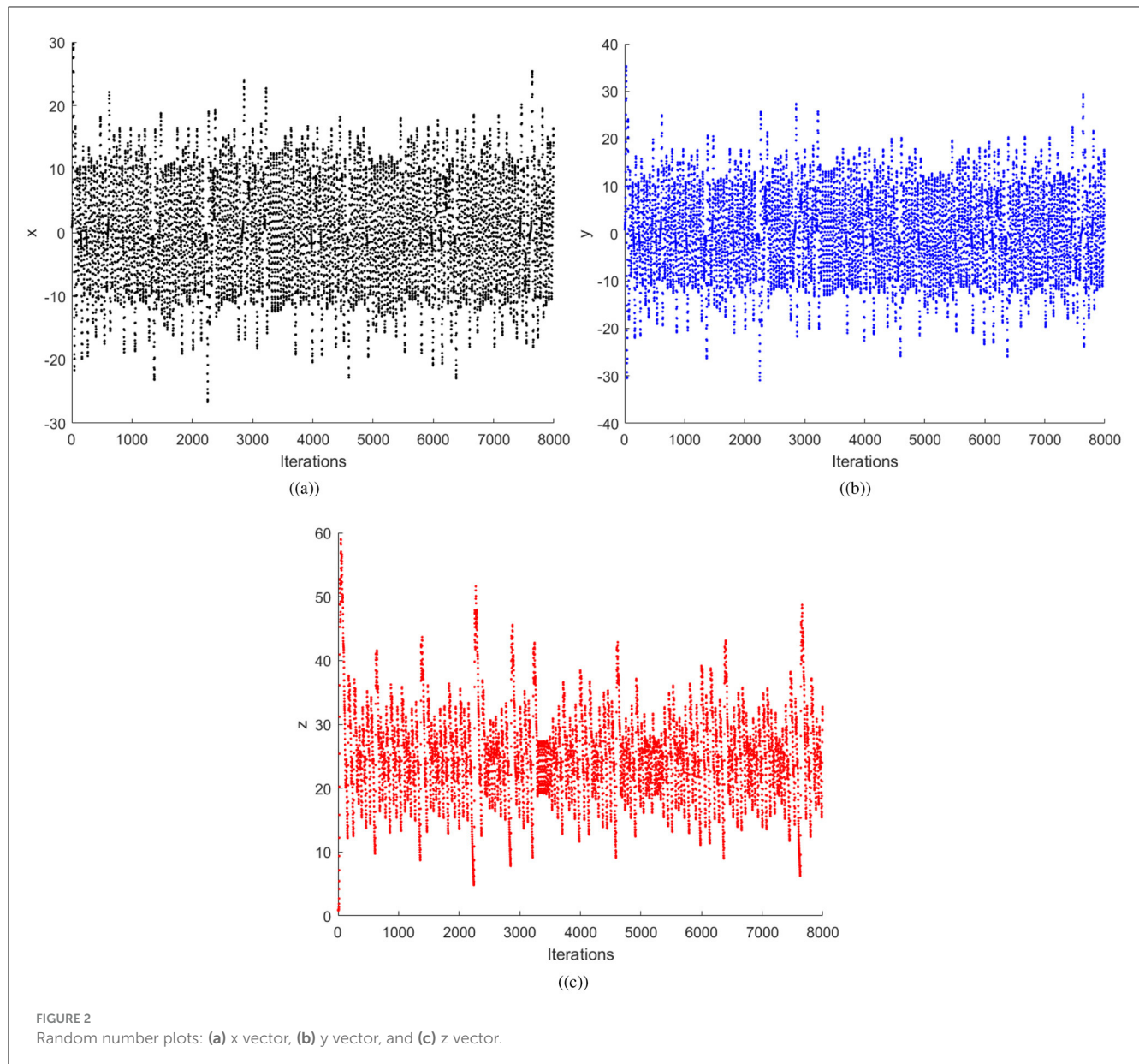
second and third equations contain cross-product terms. From a topological point of view, the Lorenz and Chen systems have different structures. Mathematically, the system can be written as Qi et al. (2005):

$$\begin{aligned}\dot{x} &= a * (y - x), \\ \dot{y} &= ((c - a) * x) + (c * y) - (x * z), \\ \dot{z} &= (x * y) - (b * z).\end{aligned}\quad (1)$$

where x , y , and z are the variables indicating the state of the system, and a , b , and c are the parameters. It has been proven that the Chen system has chaotic behavior for parameter values being $\alpha > 0.82$ and $a = 35$, $b = 3$, and $c = 28$. In the proposed scheme, the random numbers will be computed using the $\alpha = 0.9$ value. In order to illustrate Chen system sensitivity, the chaotic system is iterated twice with $x_0 = 0.01$ and $x_0 = 0.01 \times 10^{-12}$. Thus, one can confirm that both the sets of random numbers in Figure 1 are different. Further, Figure 2 shows three sets of 8,000 random numbers generated through the Chen chaotic system. Therefore, one can conclude that the chaotic system is extremely sensitive and produces different random numbers with small changes in the initial condition or control parameter.

2.2 SHA-512

In 2002, the National Security Agency (NSA) developed a cryptographic hashing algorithm named Secure Hash Algorithm 2 (SHA-2) (Wang et al., 2021). Compared to its predecessors SHA-0 and SHA-1, SHA-2 provides a more robust solution. SHA-512 is the most secure and efficient hash function in the SHA-2 family (Bhonge et al., 2020). Based on an arbitrary message length, it computes a 512-bit hash value by splitting the data into blocks of 1024 bits and passing the data through the module, consisting of 80 rounds. In our proposed scheme, SHA-512 is used to generate eight 512-bit hash values for eight plaintext bit planes, respectively. The



hash values are used to generate the initial conditions of the chaotic system.

3 Proposed methodology

To divert the attention of an attacker, visually secure encryption facilitates the transfer of private information over an insecure channel. This process embeds the ciphertext image into a carrier or host image to produce visually secure ciphertext images. **Figure 3** illustrates the general workflow of an image encryption scheme while **Figure 4** demonstrates the step-by-step flow chart of the proposed meaningful privacy preservation of medical images in IoT environments. An end-to-end encryption method has been developed that enables medical images to be transmitted over the Internet using any H-IoT device with enhanced security and confidentiality. The proposed scheme is comprised of the following

steps:

Step: 1 Let the original plaintext medical image with dimensions $m \times n$ can be represented as M and its constituent bit planes can be represented as:

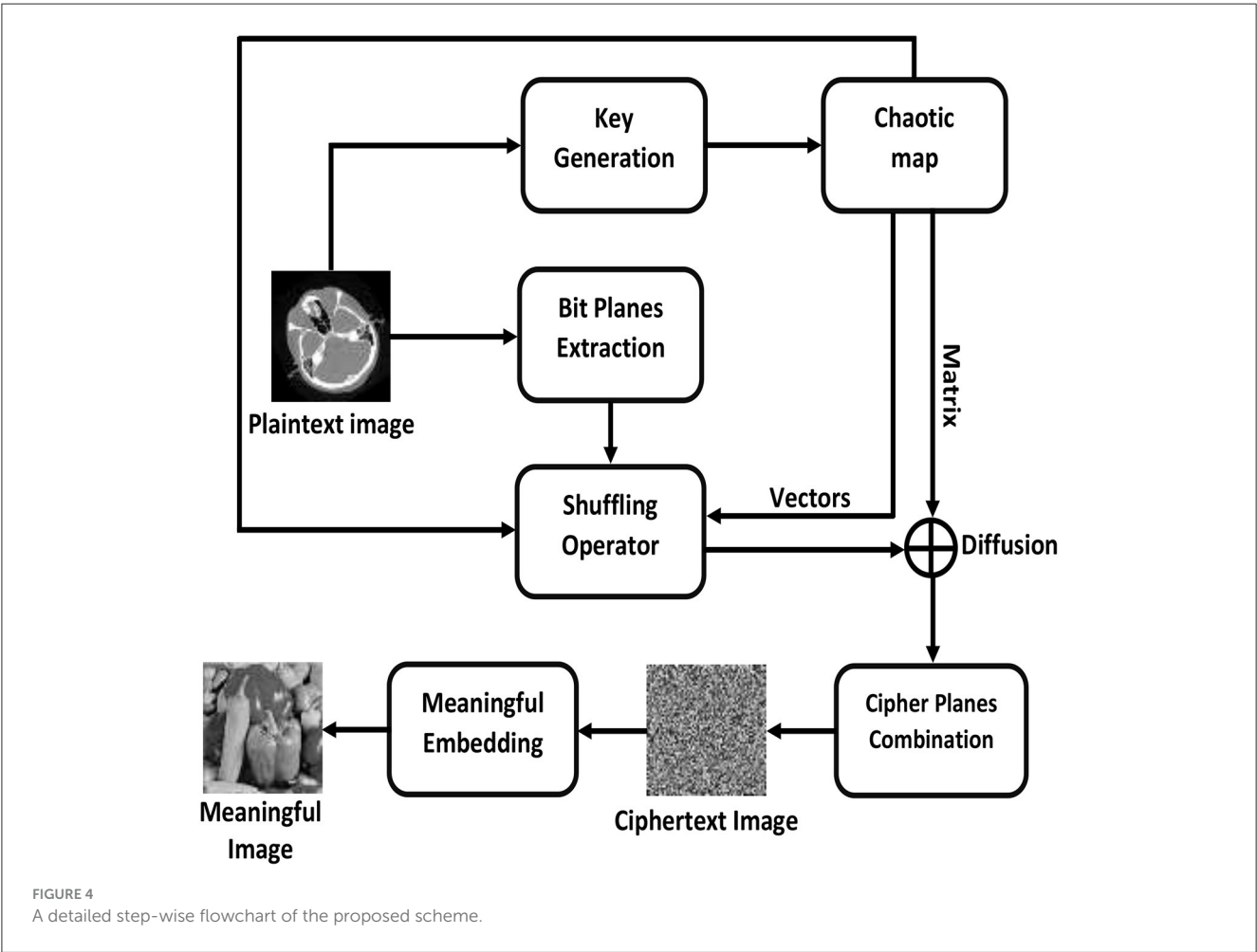
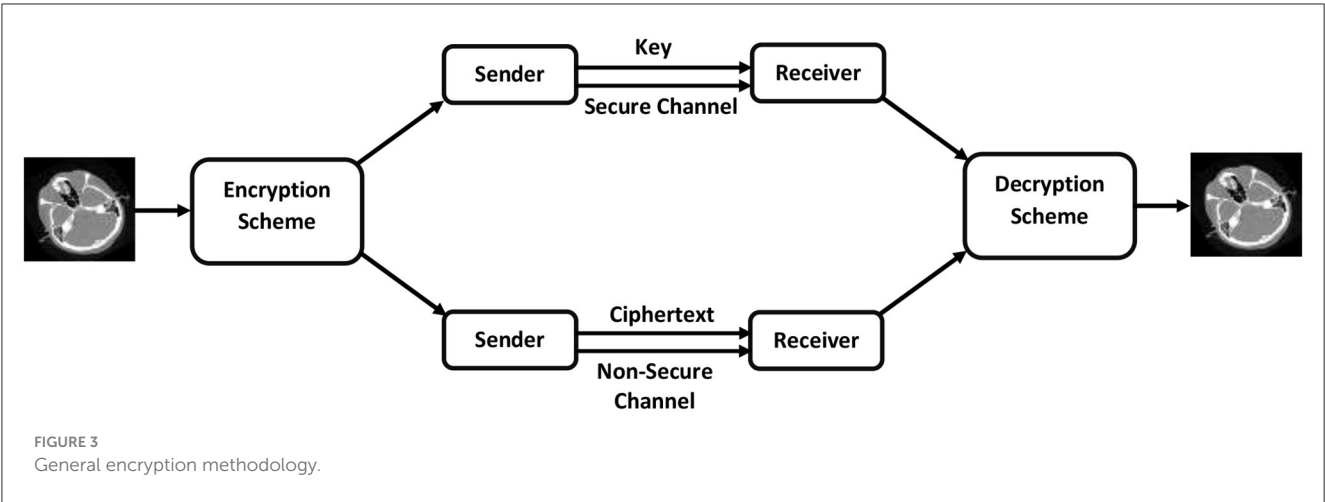
$$M = [M_1, M_2, M_3, \dots, M_8]. \quad (2)$$

Step: 2 To determine the initial conditions for the Chen chaotic system and to ensure the integrity and non-repudiation of the image data, each of these planes is cryptographically hashed utilizing SHA-512.

$$H_1 = \text{SHA} - 512(M_1). \quad (3)$$

Step: 3 For numerical interpretation purposes, the computed H_1 value is converted to a decimal value.

$$N = \text{bi2de}(H_1). \quad (4)$$



Step: 4 Now, the initial conditions can be calculated as follows:

$$x_0, y_0 \text{ and } z_0 = \frac{N}{2^{48}}. \tag{5}$$

Step: 5 The chaotic Chen system is iterated to generate three random vectors x, y , and z .

Step: 6 For each generated random vector x, y , and z , the $Mod256$ is applied to bring the values within the range of 0 and 255.

$$x, y \text{ and } z = \text{mod}((x, y, \text{ and } z) \times 10^{14}, 256). \tag{6}$$

Step: 7 The vectors x and y are utilized to permute the plaintext medical image M row- and column-wise, respectively.

$$R_M = x(M),$$

$$C_M = y(R_M). \quad (7)$$

Step: 8 The vector z is rearranged in a matrix and XORed bitwise with the permuted image to generate the final medical bit-plane ciphertext.

$$C_M = z \oplus C_M. \quad (8)$$

Step: 9 Steps 2 through 8 must be repeated eight times to encrypt each layer.

Step: 10 Combine all ciphertext planes to produce the final ciphertext or encrypted medical image.

$$C = [C_{M1}, C_{M2}, C_{M3}, \dots, C_{M8}]. \quad (9)$$

Step: 11 The carrier image C_C is passed through the Lifting Wavelet Transformation (LWT).

$$[LL, LH, HL, HH] = LWT(C_C) \quad (10)$$

Step: 12 The ciphertext image C is divided into 4 Most Significant Bits (MSBs) and 4 Least Significant Bits (LSBs). Now, the HL and HH blocks of C_C are replaced by the MSBs and LSBs.

Finally, the Inverse Lifting Wavelet Transformation (ILWT) was used to generate a visually meaningful medical image V_M . As the final visually meaningful medical image V_M contains values greater than 255 and less than 0, it is scaled by a min-max normalization function to keep them between 0 and 255.

$$V_M = ILWT[LL, LH, MSBs, LSBs] \quad (11)$$

Decryption can be accomplished by reversing all of the above steps in reverse order.

4 Results

This section presents simulations to illustrate the effectiveness and robustness of the proposed scheme. Our analysis in this section demonstrates that the IoT encryption scheme developed for medical images is robust against different security attacks. **Figure 5** shows the encryption outcomes of the designed scheme for cthead and chest images of size 128×128 . The ciphered images in **Figures 5c, g** are noise-like images, so they are encapsulated inside a carrier image (Pepper image of size 256×256) to generate a visually secure medical image. Further, correlation analyses, histogram analyses, entropy analyses, key sensitivity, key space analyses, robustness analyses, etc, are performed to demonstrate the strength of the developed medical image encryption scheme for IoT against statistical attacks, brute force attacks, noise attacks, and classical attacks.

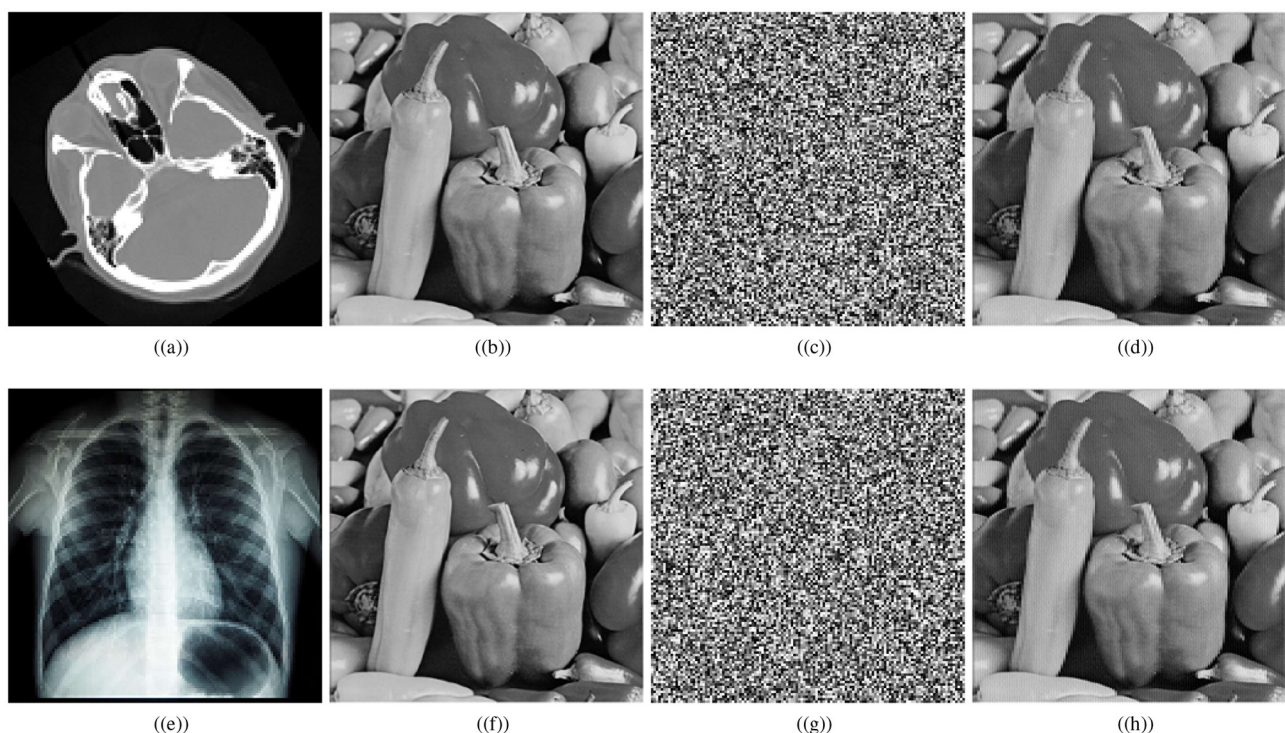


FIGURE 5

Encryption results: (a) plaintext cthead image, (b) carrier plaintext image, (c) ciphertext cthead image, (d) visually secure image, (e) plaintext chest image, (f) carrier plaintext image, (g) ciphertext chest image, (h) visually secure image.

4.1 Correlation analysis

Correlation analysis quantifies the relationship between image pixel values. Original plaintext medical images show a close association between neighboring pixels. An encrypted image is secured against pixel relation analysis attacks or statistical attacks when effective cryptographic techniques are applied to reduce the relationship between pixels. A ciphertext image with a lower correlation between adjacent pixels shows a better cryptographic technique. Mathematically, the correlation coefficient can be calculated as follows (Khan and Ahmad, 2019):

$$C.C(x, y) = \frac{\frac{1}{N} \sum_{n=1}^N (x_n - E(x))(y_n - E(y))}{\sigma_x \times \sigma_y} \quad (12)$$

where

$$\sigma_x = \sqrt{\text{Var}(x)}$$

$$\sigma_y = \sqrt{\text{Var}(y)}$$

$$\text{Var}(x) = \frac{1}{N} \sum_{n=1}^N (x_n - E(x))^2$$

$$\text{Var}(y) = \frac{1}{N} \sum_{n=1}^N (y_n - E(y))^2$$

The variables N indicate the total number of pixels while Var , σ , and E calculate the variance, standard deviation, and expected operator, respectively. Table 1 summarizes the computed correlation coefficient values for the proposed medical image encryption scheme. Almost all the encrypted images have a $C.C$ value of zero or less than 0. Meanwhile, the carrier or host image and the visually secure image have $C.C$ values near 1. Thus, embedding the ciphertext medical image does not significantly alter the carrier image. Figures 6a–c illustrates the 5,000 adjacent pixels

correlation distribution of the original plaintext cthead medical image in three distinct directions, i.e., horizontal (h), vertical (v), and diagonal (d). Figures 6d–f shows the 5,000 adjacent pixels correlation distribution of the corresponding ciphertext image. Therefore, it can be concluded from Figures 6a–c that neighboring pixels are closely associated in the original plaintext medical image. Furthermore, Figures 6d–f confirms that this association breaks down within the ciphered image's pixels, and the correlation among the pixels is totally different. Additionally, Figure 7 shows a strong association between neighbors pixels in the carrier image and the visually secured image, indicating that the visually secured image's pixels are not significantly changed.

4.2 Histogram analysis

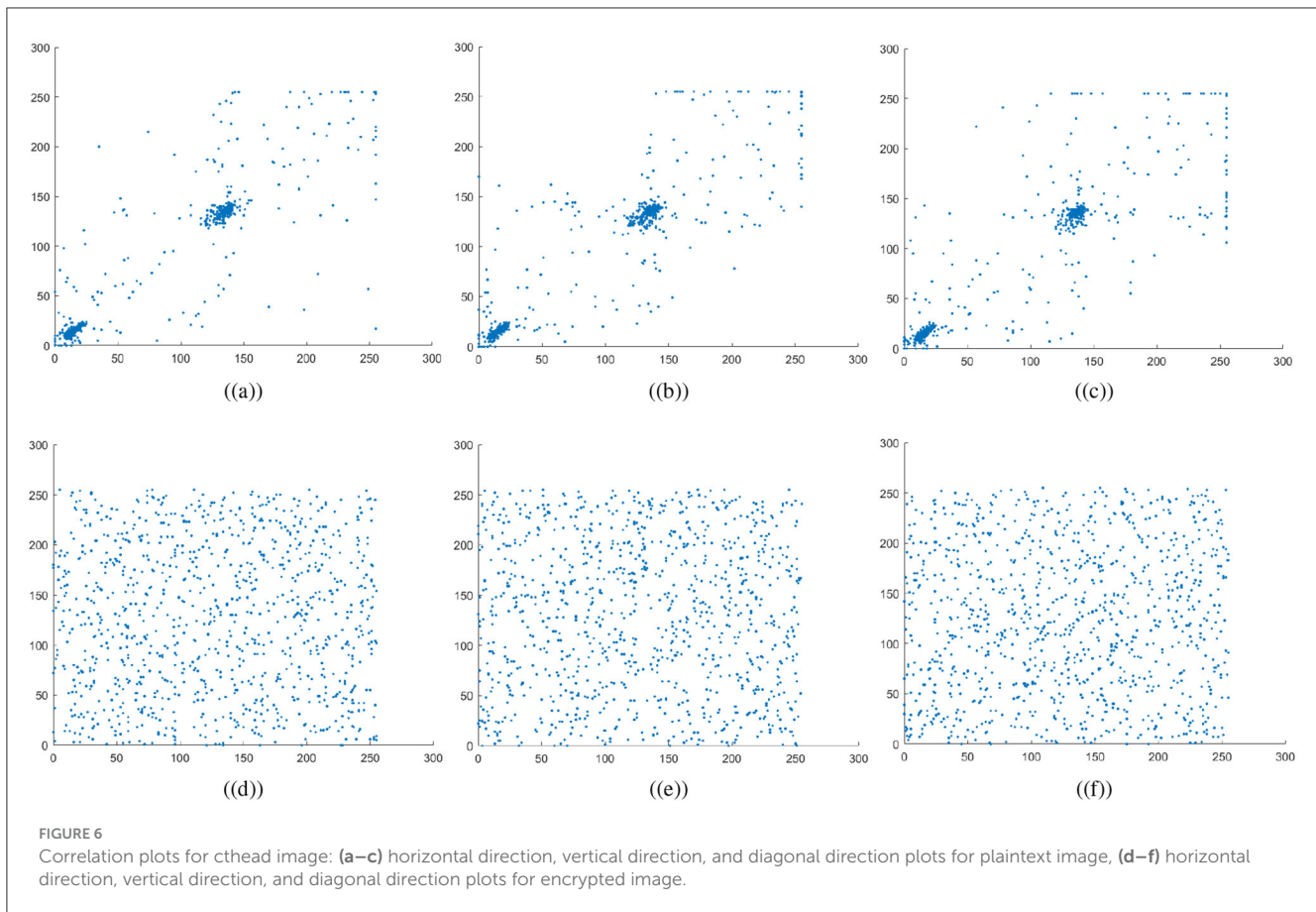
Image histograms are statistical plots, plotting the intensity of pixels against the pixel count in a digital image. Mathematically, it can be computed as follows (Singh and Kumar, 2025).

$$H(x_i) = m_i \quad (13)$$

where m_i represents the multiplicity of x_i intensity number. Histogram analysis helps to determine whether pixel intensities are distributed evenly throughout the encryption process. An encryption scheme's robustness against statistical attack can be assessed by ensuring that the encrypted image's histogram is uniform, making it impossible to use statistical analysis to guess the original image's structure (Khan and Ahmad, 2019). Figure 8 shows the histograms of the original and cipher images. Figure 8 confirms the non-uniformity of the histograms for the original cthead and chest images; that is, some pixel intensities may be dominant depending on the contents of the image. In contrast, the cipher images' histograms are uniformly distributed. As a result, the encryption process scrambles pixel values such that no feature of the plaintext image can be identified. Because of the histogram's uniformity, the proposed medical image encryption for IoT is

TABLE 1 Computed correlation coefficient values.

Image	Direction	Plaintext	Ciphertext	Carrier	Visually secure
Cthead	h	0.9480	0.0097	0.9472	0.9343
	v	0.9577	-0.0062	0.9594	0.9585
	d	0.9224	-0.0499	0.9297	0.9227
Chest	h	0.9768	-0.0384	0.9472	0.9368
	v	0.9628	-0.0258	0.9594	0.9368
	d	0.9340	-0.0380	0.9297	0.9055
Medani et al. (2025)	h	0.9173	-0.0598	-	-
	v	0.8868	0.0386	-	-
	d	0.7851	0.0239	-	-
Kumar and Sharma (2024)	h	0.7586	-0.0075	-	-
	v	0.8665	-0.0071	-	-
	d	0.7261	0.0041	-	-



highly resistant to statistical attacks. Histograms of carrier images and visually secured images appear to be nearly identical. Thus, the attacker will not be able to determine that the carrier image is embedded with an encrypted image, as the embedding is not producing significant changes in the host image.

4.3 Key space

In an encryption algorithm, key space refers to all possible secret keys and different parameters. The authors in [Alvarez and Li \(2006\)](#) illustrate how key space size influences the strength of image ciphering techniques. It is essential that the key space be sufficiently large and must exceed 2^{100} to withstand brute force attacks. The proposed meaningful privacy preserving of medical images in IoT environment utilizes the Chen chaotic system, with state variables x , y , and z and control parameters a , b , and c . Each of these parameters has a floating precision of 10^{15} . Further, the map is iterated 8 times for each bit plane. Therefore, the key space of the designed scheme can be computed as follows:

$$K = (10^{15} \times 10^{15} \times 10^{15} \times 10^{15} \times 10^{15} \times 10^{15})^8$$

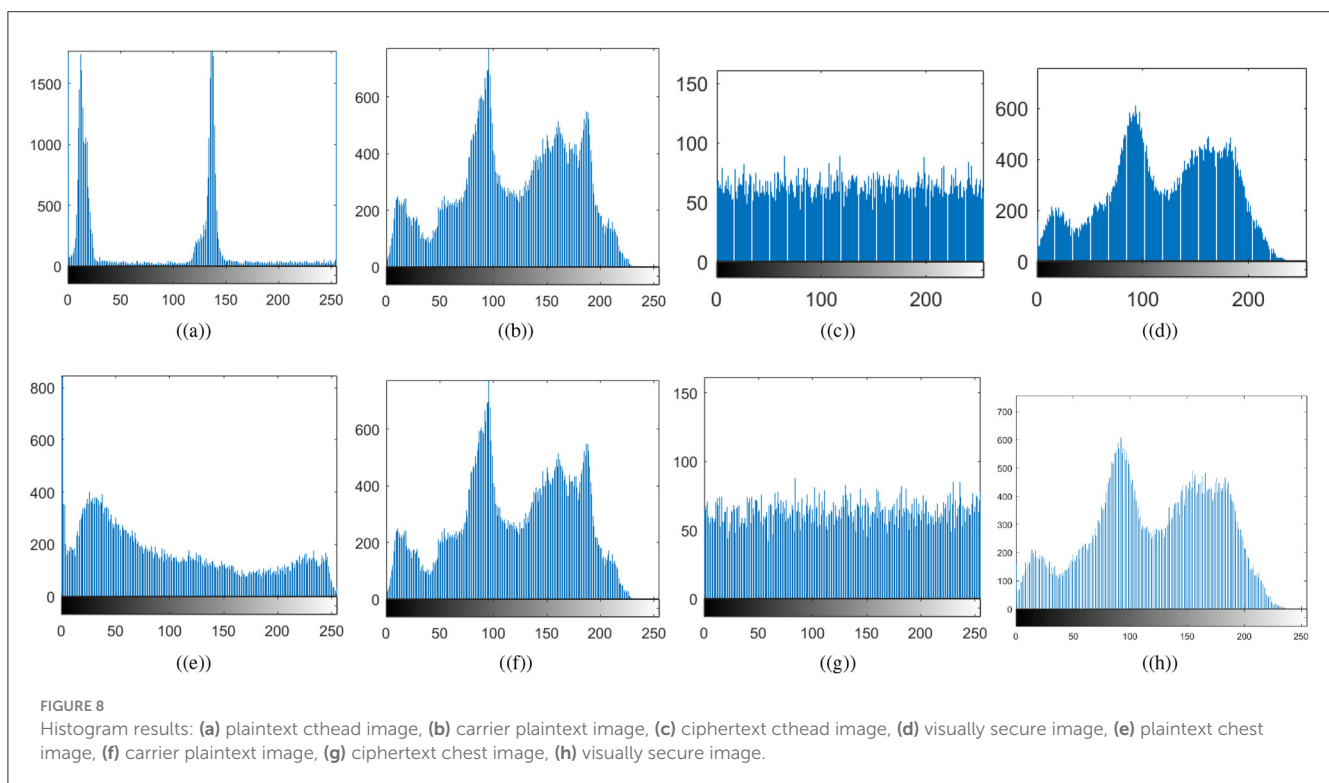
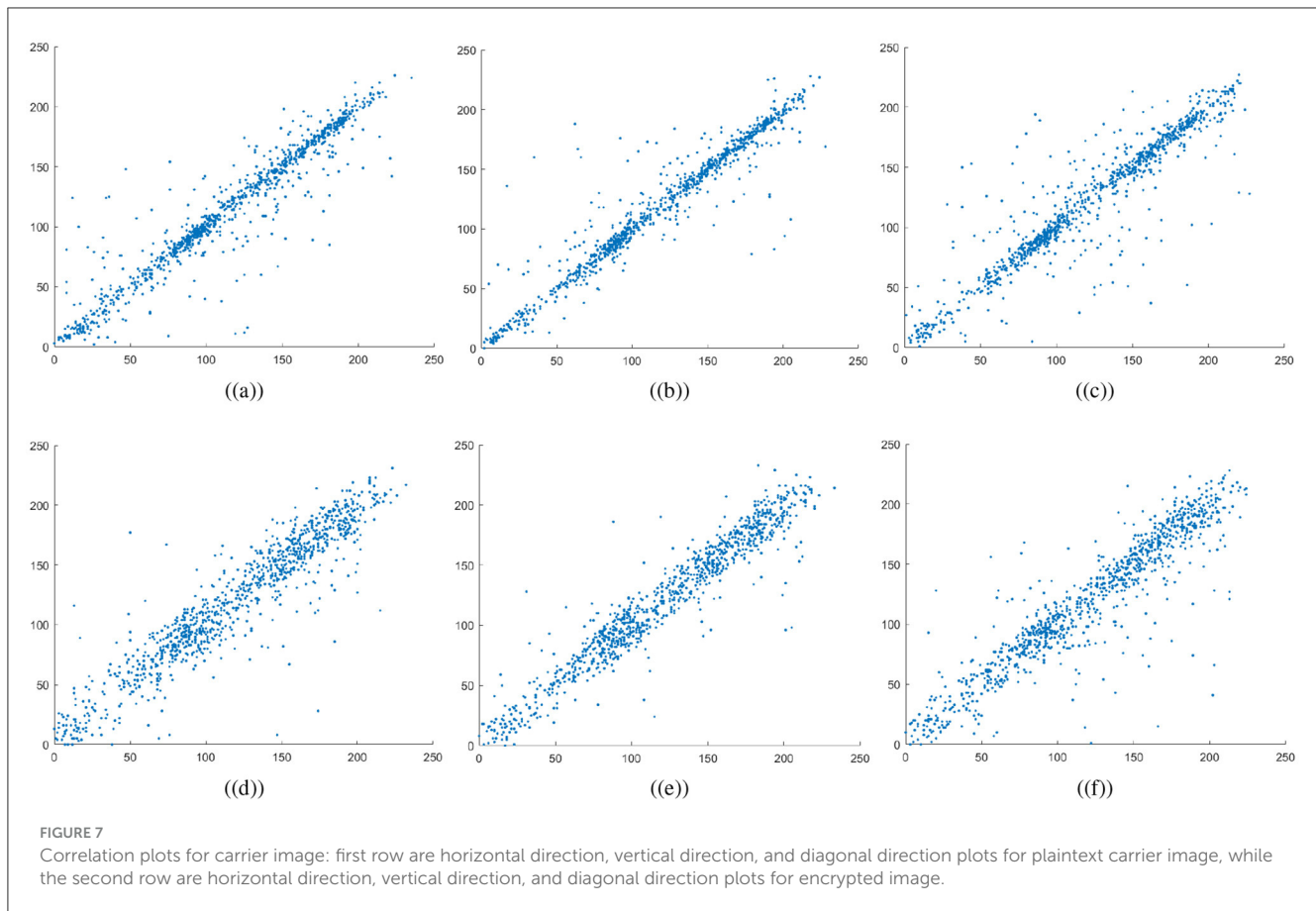
$$K = (10^{90})^8 >> 2^{100} \quad (14)$$

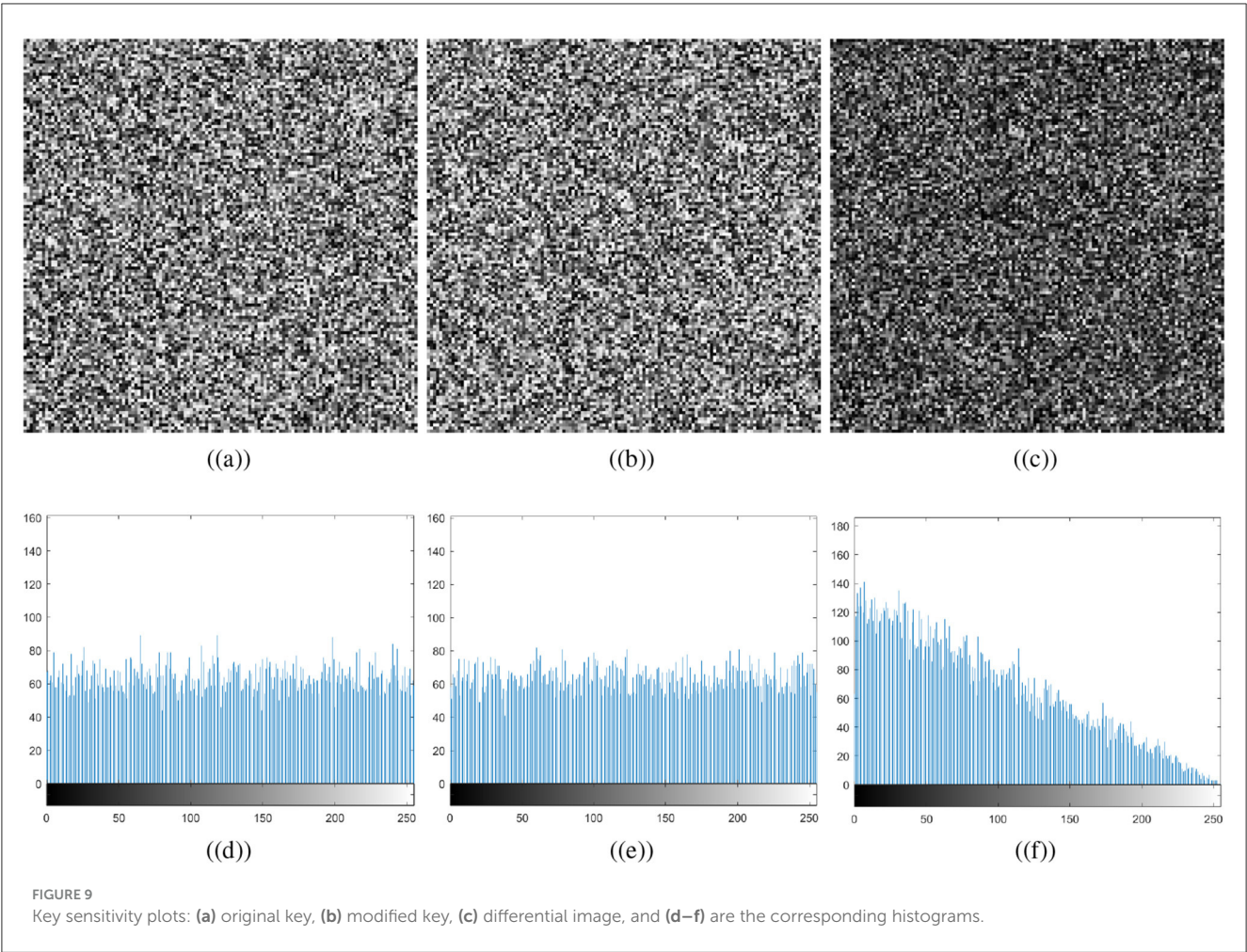
Furthermore, the key space computed in [Kanwal et al. \(2024\)](#) and [Medani et al. \(2025\)](#) is 2^{282} and 2^{598} , respectively. Thus, one

can conclude that the key space of the presented medical image encryption scheme is sufficiently large to resist a brute force attack significantly.

4.4 Key sensitivity

A good image encryption technique should be able to detect subtle changes in secret keys and parameters, resulting in decoded data that is different from plain image data. The proposed medial image encryption is extremely sensitive to the control parameters and initial conditions. Let's make a small change of 10^{-12} in one of the initial conditions or the control parameters, i.e., x_0 of the Chen chaotic system. As a result, the chaotic system will generate different random numbers. [Figure 1](#) shows the different number generation for a small modification in the initial conditions. [Figure 9](#) illustrates the resultant images after decrypting the ciphertext cthead image with the same and modified keys. A differential image of the two resultant images is shown in [Figure 9c](#). A small change to the initial conditions or control parameters of the Chen chaotic system will fail the decryption process, resulting in a completely different image for the attacker. The differential image demonstrates that both resultant images are different and lack any recognizable information related to the plaintext cthead image. It can therefore be concluded that the proposed meaningful medical image encryption scheme is exceptionally sensitive to even minor changes in the chaotic system control parameters.





4.5 Entropy analysis

Entropy analysis is usually used to assess image encryption’s robustness against entropy attacks. Mathematically, entropy of a data source can be computed as follows (Singh and Kumar, 2025):

$$E(S) = \sum_{k=0}^{256-1} \left(\frac{1}{256}\right) \log_2\left(\frac{1}{\frac{1}{256}}\right) \tag{15}$$

To resist the entropy attack, the entropy value of the encrypted images should be close to 8. Table 2 summarizes the computed entropy values for the proposed medical image encryption scheme. Thus, one can confirm that the entropy value of the ciphertext medical image is approximately equal to 8. The designed technique is robust against entropy attacks without exposing sensitive information.

4.6 Differential attack analysis

To measure the effectiveness and reliability of image encryption algorithms against differential attacks, it is important to determine

TABLE 2 Computed entropy values.

Image	Plaintext	Ciphertext	Carrier	Visually secure
Cthead	5.6763	7.9987	7.6110	7.6485
Chest	7.4040	7.9982	7.6110	7.6498
Medani et al. (2025)	7.6414	7.9998	-	-
Kumar and Sharma (2024)	7.3579	7.9987	-	-

TABLE 3 Number of pixels change rate and unified average change intensity computed values.

Image	NPCR	UACI
Cthead	99.6755%	33.5105%
Chest	99.6867%	33.5241%
Medani et al. (2025)	99.6653%	33.5328%
Kumar and Sharma (2024)	99.5800%	33.1800%

the Number of Pixels Change Rate (NPCR) as well as the Unified Average Change Intensity (UACI). These two matrices can be

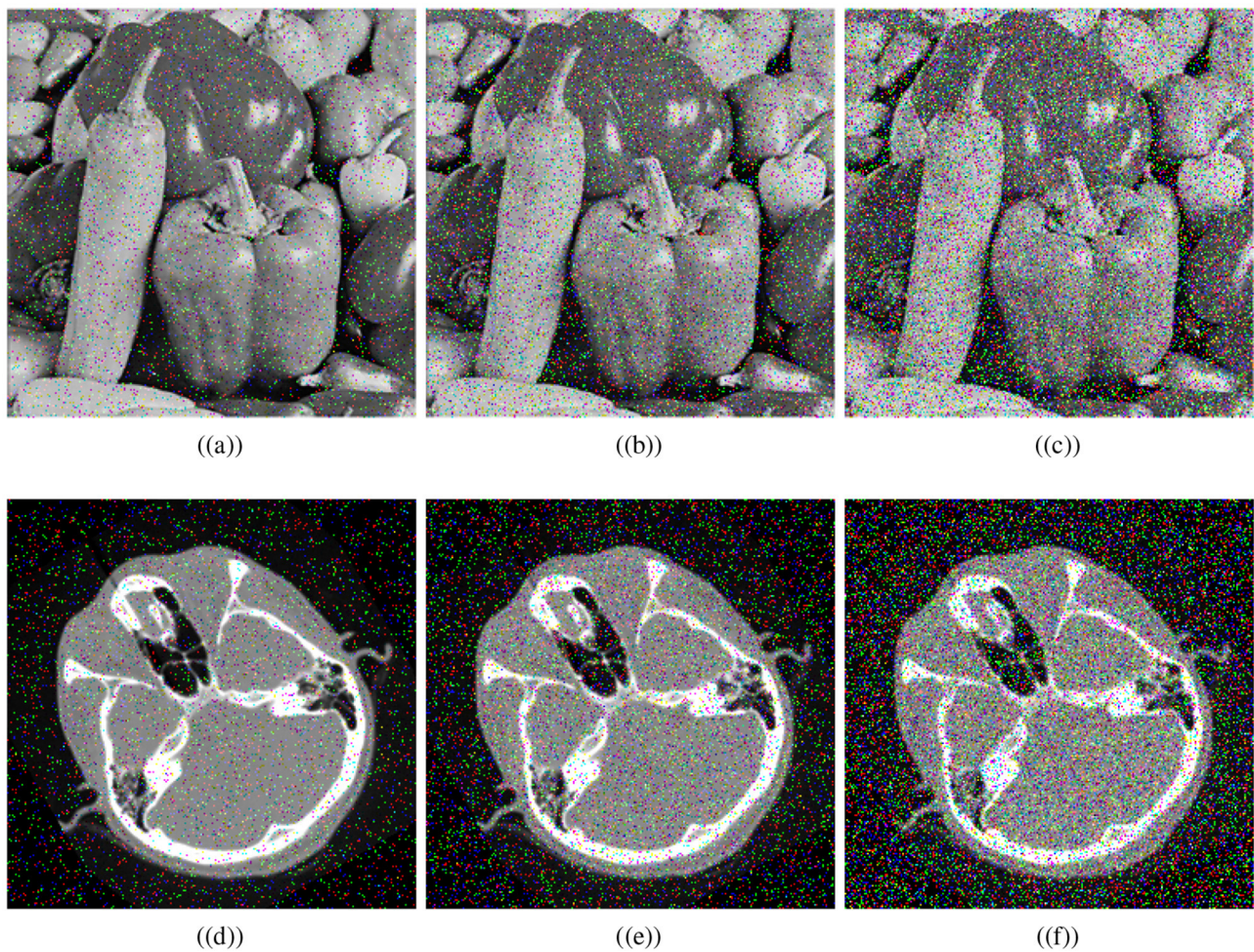


FIGURE 10
Noise results: salt and pepper noise ratios; (a) 5%, (b) 10%, (c) 20%, and (d–f) corresponding decrypted images.

mathematically defined as follows (Liu et al., 2024):

$$D(x, y) = \begin{cases} 1, & \text{if } C_1(x, y) \neq C_2(x, y) \\ 0, & \text{if } C_1(x, y) = C_2(x, y) \end{cases} \quad (16)$$

$$NPCR = \sum_{x,y} \frac{D(x, y)}{N} \times 100\% \quad (17)$$

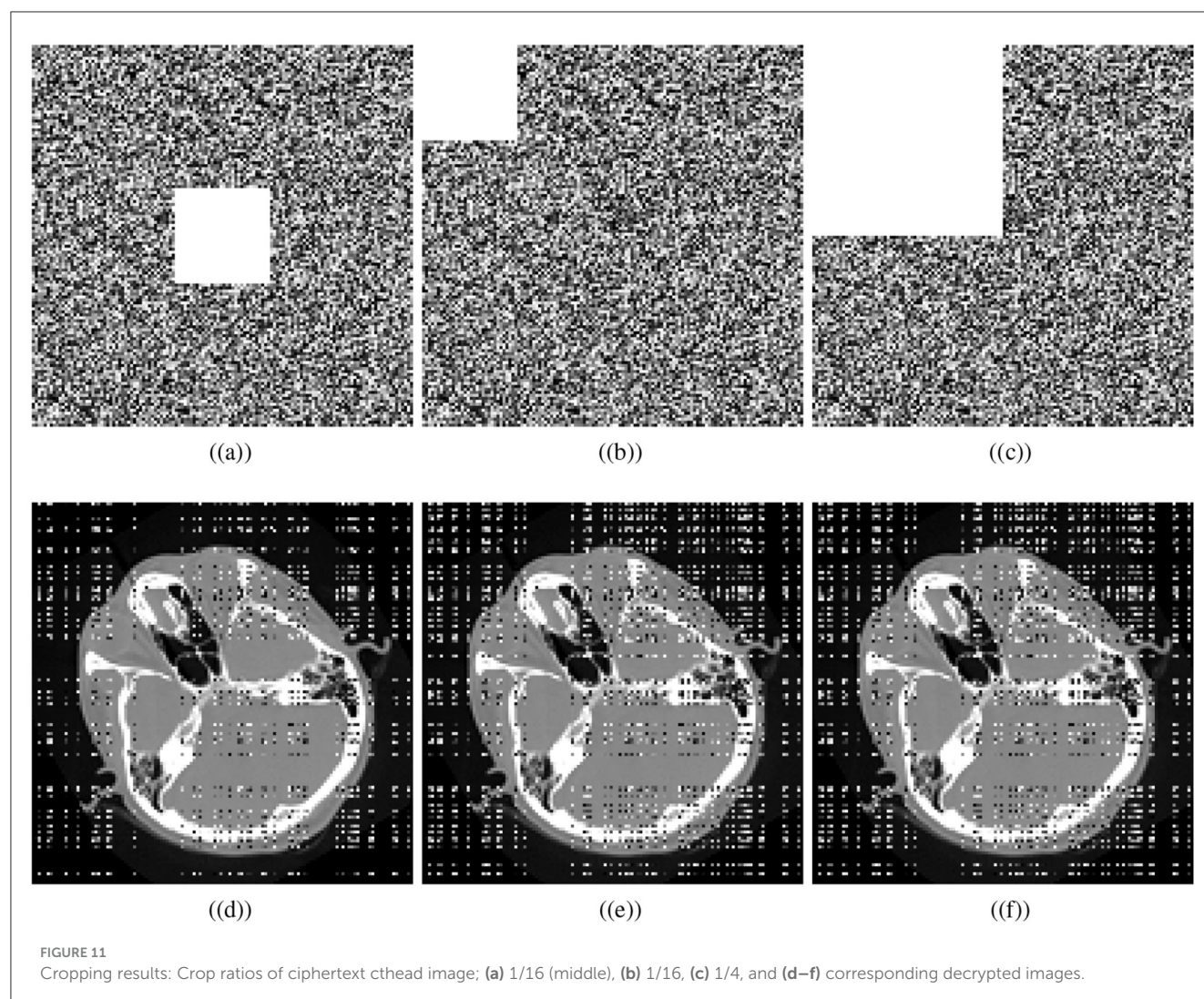
$$UACI = \frac{1}{N} \sum_{x,y} \frac{|C_1(x, y) - C_2(x, y)|}{255} \times 100 \quad (18)$$

where N shows the total number of pixel values and C_1 represents the first encrypted image generated without any change in the original plaintext image while C_2 represents the encrypted image generated after altering just one pixel in the original image. When comparing two images that have been encrypted, the UACI test measures the difference in pixel intensity, whereas the NPCR test measures how frequently the pixels are changed in the plaintext.

The calculated NPCR and UACI values for the designed medical image security scheme are illustrated in Table 3. Therefore, the values $UACI > 33\%$ and $NPCR > 99\%$ confirm that the proposed strategy is resilient to differential attack.

4.7 Noise attack analysis

It has become increasingly important to analyze noise attacks when data is transmitted over open networks due to the presence of noise during transmission. Therefore, the proposed algorithm's effectiveness is determined by comparing the decryption of encrypted images under different noise intensities. Figure 10 shows the recovered images after adding salt and pepper noise of (5%, 10%, and 20%) intensities to the visually secured image. Thus, one can see that the proposed medical encryption scheme can decrypt the noise-polluted ciphertext image, illustrating the robustness of the scheme.



4.8 Occlusion attack analysis

Various factors can cause data to be lost during image transmission over a network. The purpose of occlusion analysis is to determine whether or not an image encryption scheme can recover a plaintext image from a ciphertext image that has been occluded. Different-sized portions of the encrypted image are cropped and decrypted. This analysis can provide insight into how the encryption scheme scrambles plaintext images. Generally, the better the scrambling effect, the more likely the algorithm is to reconstruct the visual characteristics of the plaintext image, even if some part of it has been lost. We cropped the ciphertext image and visually secured image with the ratios 1/16 (middle), 1/16, and 1/4. Decryption is performed utilizing the presented scheme. Figure 11 shows the cropped ciphertext images and the corresponding decipher images while Figure 12 illustrates the cropped visually secured images and the corresponding decipher images. The visual results clearly demonstrate that the proposed scheme strongly deciphers the cropped images without causing any noticeable distortion.

4.9 Resilience against classical attacks

Classical attack analysis focuses on identifying and analyzing various types of attacks (known plaintext, chosen plaintext, ciphertext only, and chosen ciphertext) against encrypted images. The key of the chaotic maps is computed based on the plaintext hash value. It is used to determine initial conditions and control parameters. Because of the dependence on plaintext images, the proposed enhanced medical image privacy in IoT with bit-plane level encryption using a chaotic map avoids the classical attacks cited above. Therefore, all random vectors and matrices are determined by plain image bit planes. When a single pixel is changed in the plaintext image, the keystream changes. This will result in a completely different ciphertext image.

4.10 Complexity analysis

Time complexity is a metric used to estimate the running time of an encryption algorithm and generally determine the scheme's

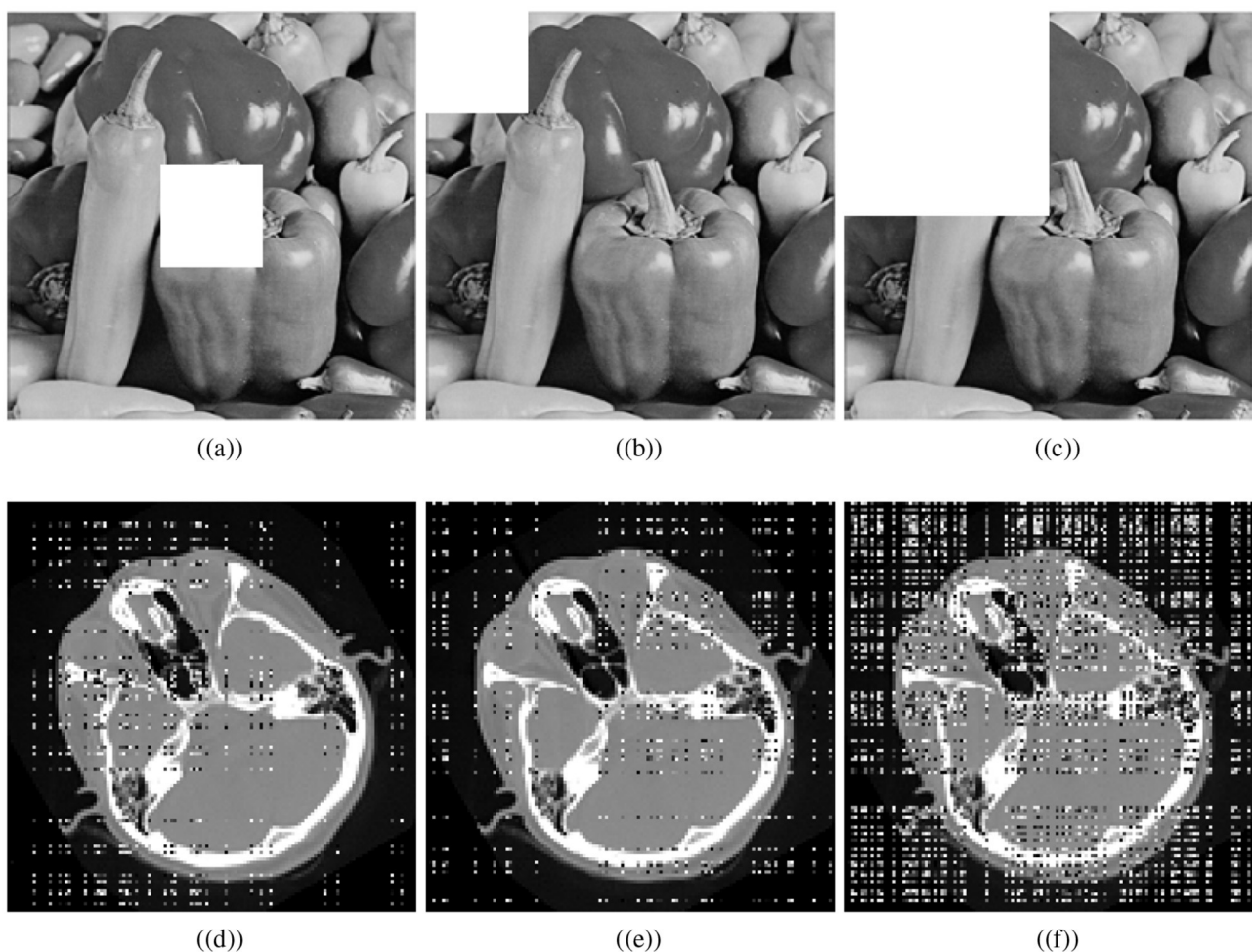


FIGURE 12
Cropping results: Crop ratios of visually meaningful image; (a) 1/16 (middle), (b) 1/16, (c) 1/4, and (d–f) corresponding decrypted images.

feasibility. A good algorithm needs to have a short running time. The encryption and decryption results are performed on MATLAB 2018a with Microsoft Windows 10, 4 GB of memory, and a 1 GHz CPU. The cthead and chest images have a size of 128×128 , while the carrier image has a size of 256×256 . The proposed scheme takes 0.85s to generate the ciphertext image and 0.62s to produce the visually meaningful ciphertext. Thus, the proposed scheme takes 1.47s to generate the final meaningful ciphertext. The image encryption scheme presented in Kumar and Sharma (2024) takes 0.85s while the scheme discussed in Medani et al. (2025) takes 4.57s to produce the final encrypted images. The designed scheme takes less time than the scheme presented in Medani et al. (2025) and more time than the scheme discussed in Kumar and Sharma (2024).

5 Conclusion

This paper presents a novel and robust medical image encryption scheme for resource-constrained devices. Due to simplicity and exceptional performance in terms of unpredictability, the proposed scheme utilizes 3D Chen chaotic system. The simplicity and excellent performance make the

Chen chaotic map an excellent choice for lightweight encryption applications. The designed meaningful bit-plane-level medical image encryption scheme for IoT leverages the pixel scrambling and diffusion characteristics to effectively break pixel relationships, thus, enhancing encryption efficiency and security. To enhance security, the plaintext bit planes are hashed using the Secure Hash Algorithm (SHA-512) to compute the initial conditions of the chaotic map. This dependency on the plaintext images makes the designed scheme resilient against classical attacks (known-plaintext, chosen-plaintext, ciphertext-only, and chosen-ciphertext). As a result, three random vectors for permutation and XOR diffusion are generated. A permutation and XOR operation are applied to each bit-plane to produce a ciphertext plane. After combining the ciphertext bit-planes, the visually secured ciphertext image is now generated by embedding the ciphered image within the carrier image. Extensive evaluations have proven that the designed scheme exhibits a high degree of resilience to attacks, making it particularly suitable for small IoT devices with limited processing power and memory. Computational complexity could be a possible limitation of the designed scheme, as image sizes increase, the encryption process could take longer.

Data availability statement

Interested researchers may contact the author for potential collaboration or for inquiries regarding data access within the constraints of institutional guidelines.

Author contributions

FA: Formal analysis, Methodology, Conceptualization, Writing – original draft, Software, Data curation. WA: Conceptualization, Formal analysis, Methodology, Writing – original draft. TZ: Formal analysis, Methodology, Validation, Writing – review & editing. IN: Formal analysis, Methodology, Visualization, Writing – original draft. AA: Investigation, Supervision, Validation, Visualization, Writing – review & editing, Project administration. NG: Formal analysis, Investigation, Project administration, Supervision, Writing – review & editing. AQ: Formal analysis, Writing – review & editing, Methodology, Resources, Software.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Group Project under grant number (RGP.2/245/46).

References

- Al-Hajj, A., Abandah, G., and Hussein, N. (2015). Crypto-based algorithms for secured medical image transmission. *IET Inf. Secur.* 9, 365–373. doi: 10.1049/iet-ifs.2014.0245
- Alvarez, G., and Li, S. (2006). Some basic cryptographic requirements for chaos-based cryptosystems. *Int. J. Bifurc. Chaos* 16, 2129–2151. doi: 10.1142/S0218127406015970
- Aman, A. H. M., Yadegaridehkordi, E., Attarbashi, Z. S., Hassan, R., and Park, Y.-J. (2020). A survey on trend and classification of internet of things reviews. *IEEE Access* 8, 111763–111782. doi: 10.1109/ACCESS.2020.3002932
- Belazi, A., Talha, M., Kharbech, S., and Xiang, W. (2019). Novel medical image encryption scheme based on chaos and dna encoding. *IEEE Access* 7, 36667–36681. doi: 10.1109/ACCESS.2019.2906292
- Bhonge, H. N., Ambat, M. K., and Chandavarkar, B. (2020). “An experimental evaluation of sha-512 for different modes of operation,” in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (IEEE), 1–6. doi: 10.1109/ICCCNT49239.2020.9225559
- Cao, W., Zhou, Y., Chen, C. P., and Xia, L. (2017). Medical image encryption using edge maps. *Signal Proc.* 132, 96–109. doi: 10.1016/j.sigpro.2016.10.003
- Chu, L., Su, Y., Zan, X., Lin, W., Yao, X., Xu, P., et al. (2024). A deniable encryption method for modulation-based DNA storage. *Comput. Life Sci.* 16, 872–881. doi: 10.1007/s12539-024-00648-5
- Dridi, M., Hajjaji, M. A., Bouallegue, B., and Mtibaa, A. (2016). Cryptography of medical images based on a combination between chaotic and neural network. *IET Image Process.* 10, 830–839. doi: 10.1049/iet-ipr.2015.0868
- Gan, Z., Xiong, B., Pang, Z., Chai, X., Jiang, D., and He, X. (2024). A visually secure image encryption scheme using newly designed 1d sinusoidal chaotic map and p-tensor product compressive sensing. *Nonlinear Dyn.* 112, 2979–3001. doi: 10.1007/s11071-023-09203-1
- Hanchate, R., and Anandan, R. (2024). Medical image encryption using hybrid adaptive elliptic curve cryptography and logistic map-based DNA sequence in IoT environment. *IETE J. Res.* 70, 5734–5749. doi: 10.1080/03772063.2023.2268578
- Hasan, M. K., Islam, S., Sulaiman, R., Khan, S., Hashim, A.-H. A., Habib, S., et al. (2021). Lightweight encryption technique to enhance medical image security on internet of medical things applications. *IEEE Access* 9, 47731–47742. doi: 10.1109/ACCESS.2021.3061710
- Hu, J., Jiang, H., Chen, S., Zhang, Q., Xiao, Z., Liu, D., et al. (2024). Wishield: privacy against wi-fi human tracking. *IEEE J. Select. Areas Commun.* 42, 2970–2984. doi: 10.1109/JISAC.2024.3414597
- Ismail, S. M., Said, L. A., Radwan, A. G., Madian, A. H., and Abu-Elyazeed, M. F. (2018). Generalized double-humped logistic map-based medical image encryption. *J. Adv. Res.* 10, 85–98. doi: 10.1016/j.jare.2018.01.009
- Jain, K., Titus, B., Krishnan, P., Sudevan, S., Prabhu, P., and Alluhaidan, A. S. (2024). A lightweight multi-chaos-based image encryption scheme for IoT networks. *IEEE Access* 12, 62118–62148. doi: 10.1109/ACCESS.2024.3377665
- Kanwal, S., Inam, S., Nawaz, Z., Hajje, F., Alfraihi, H., and Ibrahim, M. (2024). Securing blockchain-enabled smart health care image encryption framework using tinkerbelle map. *Alexandria Eng. J.* 107, 711–729. doi: 10.1016/j.aej.2024.08.115
- Kester, Q.-A., Nana, L., Pascu, A. C., Gire, S., Eghan, J. M., and Quaynor, N. N. (2015). A cryptographic technique for security of medical images in health information systems. *Procedia Comput. Sci.* 58, 538–543. doi: 10.1016/j.procs.2015.08.070
- Khan, J. S., and Ahmad, J. (2019). Chaos based efficient selective image encryption. *Multidimens. Syst. Signal Process.* 30, 943–961. doi: 10.1007/s11045-018-0589-x
- Khan, J. S., Ahmad, J., Abbasi, S. F., and Kayhan, S. K. (2018). “DNA sequence based medical image encryption scheme,” in *2018 10th Computer Science and Electronic Engineering (CEECE)* (IEEE), 24–29. doi: 10.1109/CEECE.2018.8674221
- Khan, J. S., Ahmed, S. S., Haq, F. U., Tariq, N., Ahmad, J., and Ashraf, M. (2024). “A robust visually secure bit plane image encryption using chaotic map and sha-512,” in *2024 3rd International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE)* (IEEE), 1–6. doi: 10.1109/ETECTE63967.2024.10823709

Acknowledgments

The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Group Project under grant number (RGP.2/245/46).

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.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

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.

- Khan, J. S., Boulila, W., Ahmad, J., Rubaiee, S., Rehman, A. U., Alroobaea, R., et al. (2020). Dna and plaintext dependent chaotic visual selective image encryption. *IEEE Access* 8, 159732–159744. doi: 10.1109/ACCESS.2020.3020917
- Kumar, S., and Sharma, D. (2024). A chaotic based image encryption scheme using elliptic curve cryptography and genetic algorithm. *Artif. Intell. Rev.* 57:87. doi: 10.1007/s10462-024-10719-0
- Lai, Q., and Hua, H. (2025). Secure medical image encryption scheme for healthcare IoT using novel hyperchaotic map and DNA cubes. *Expert Syst. Appl.* 264:125854. doi: 10.1016/j.eswa.2024.125854
- Liu, W., Wang, H., Chen, Y., and Sun, K. (2024). A new image encryption scheme based on block compressive sensing and chaotic laser system for IoT. *Eur. Phys. J. Plus* 139:473. doi: 10.1140/epjp/s13360-024-05221-z
- Medani, M., Said, Y., Othman, N. A., Yuldashev, F., Kchaou, M., Aldawood, F. K., et al. (2025). Chaos-based novel watermarked satellite image encryption scheme. *Comput. Model. Eng. Sci.* 143, 1049–1070. doi: 10.32604/cmesci.2025.063405
- Nadhan, A. S., and Jacob, I. J. (2024). Enhancing healthcare security in the digital era: Safeguarding medical images with lightweight cryptographic techniques in iot healthcare applications. *Biomed. Signal Process. Control* 88:105511. doi: 10.1016/j.bspc.2023.105511
- Porras, J., Pänkäläinen, J., Knutas, A., and Khakurel, J. (2018). “Security in the internet of things-a systematic mapping study,” in *Proceedings of the 51st Hawaii International Conference on System Sciences*. doi: 10.24251/HICSS.2018.473
- Qi, G., Chen, G., Du, S., Chen, Z., and Yuan, Z. (2005). Analysis of a new chaotic system. *Physica A* 352, 295–308. doi: 10.1016/j.physa.2004.12.040
- Sathananthavathi, V., Ganesh Kumar, K., and Sathish Kumar, M. (2024). Secure visual communication with advanced cryptographic and image processing techniques. *Multimed. Tools Appl.* 83, 45367–45389. doi: 10.1007/s11042-023-17224-6
- Singh, D., and Kumar, S. (2025). Image authentication and encryption algorithm based on RSA cryptosystem and chaotic maps. *Expert Syst. Applic.* 274:126883. doi: 10.1016/j.eswa.2025.126883
- Trujillo-Toledo, D. A., López-Bonilla, O. R., García-Guerrero, E. E., Esqueda-Elizondo, J. J., Cárdenas-Valdez, J. R., Tamayo-Pérez, U. J., et al. (2023). Real-time medical image encryption for H-IoT applications using improved sequences from chaotic maps. *Integration* 90, 131–145. doi: 10.1016/j.vlsi.2023.01.008
- Wang, K., Wu, X., Wang, H., Kan, H., and Kurths, J. (2021). New color image cryptosystem via sha-512 and hybrid domain. *Multimed. Tools Appl.* 80, 18875–18899. doi: 10.1007/s11042-021-10511-0
- Ye, G., and Huang, X. (2015). An image encryption algorithm based on autoblocking and electrocardiography. *IEEE MultiMedia* 23, 64–71. doi: 10.1109/MMUL.2015.72
- Zhang, D., Yan, C., Duan, Y., Liang, S., Wu, J., and Li, T. (2024). A fast visually meaningful image encryption algorithm based on compressive sensing and joint diffusion and scrambling. *Multim. Tools Applic.* 83, 70693–70725. doi: 10.1007/s11042-024-18343-4



OPEN ACCESS

EDITED BY

Muhammad Zia Ur Rehman,
Riphah International University, Pakistan

REVIEWED BY

Guillem Domènech-Gil,
Linköping University, Sweden
Jawad Ahmad,
Prince Mohammad bin Fahd University, Saudi
Arabia

*CORRESPONDENCE

Hiroki Sato,
✉ hiroki@shibaura-it.ac.jp

RECEIVED 24 March 2025

ACCEPTED 03 June 2025

PUBLISHED 19 June 2025

CITATION

Asahi R, Toriyama S, Minagawa Y, Yoshimoto S
and Sato H (2025) Classification of power grip
and precision grip in children using an EIT-
based tactile sensor.
Front. Sens. 6:1598903.
doi: 10.3389/fsens.2025.1598903

COPYRIGHT

© 2025 Asahi, Toriyama, Minagawa, Yoshimoto
and Sato. 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.

Classification of power grip and precision grip in children using an EIT-based tactile sensor

Ryunosuke Asahi¹, Shumpei Toriyama¹, Yasuyo Minagawa²,
Shunsuke Yoshimoto³ and Hiroki Sato^{1*}

¹Graduate School of Engineering and Science, Shibaura Institute of Technology, Saitama, Japan,

²Department of Psychology, Faculty of Letters, Keio University, Tokyo, Japan, ³Graduate School of
Engineering, Osaka University, Suita, Japan

Quantitative monitoring and measurement of hand motion in children are crucial to support healthy development. Electrical impedance tomography-based tactile sensors, also known as tomographic tactile sensors, provide a promising approach for grasp classification. Our previous study in adults and children demonstrated the feasibility of pinch classification using a cylindrical device equipped with the tomographic tactile sensor. In this study, we developed a new sensing device to classify the power grip and precision grip in children. In order to address concerns that children might lick or swing the device, a cylindrical sensing device was integrated sensor and measurement circuit, incorporated a protective layer for enhanced safety. Seventeen children participated in an experiment to evaluate the feasibility of the grasp classification. The classification features were voltage vectors and reconstructed images obtained from the sensor, and two machine learning methods were used as the classifiers. The average classification accuracy exceeded 85% for both feature types, surpassing the chance level of 50%. These results demonstrate that the basic grasp patterns in children can be accurately classified using a tomographic tactile sensor. This study provides new insights into the future application of grasp motion classification in children.

KEYWORDS

children, classification, electrical impedance tomography, neuro-developmental Engineering, power grip, precision grip, tactile sensors

1 Introduction

Humans have uniquely developed manual dexterity and built an advanced civilization using tools through motions, such as gripping a hammer and pinching nails. Napier divided these grasps into two basic definitions: power grip and precision grip (Napier, 1956). The power grip is defined as a motion that involves grasping an object with the palm and thumb and corresponds to the motion of gripping a hammer. This grasping is observed in healthy infants at 25 weeks post-pregnancy (Allen and Capute, 1986). On the other hand, precision grip is defined as a motion in which an object is pinched between the thumb and other fingers, corresponds to the motion of pinching nails. A study analyzing grasping patterns through video coding in infants between 2 and 22 weeks of age reported that the pre-precision grip was first performed at 2.74 months of age and the precision grip was first performed at 5.97 months of age (Wallace and Whishaw, 2003).

These grasping motions are related to various aspects of the infant and child development. For example, some studies have suggested that fine motor skills (FMS)

TABLE 1 Conventional classification method of infants' and children's hand motions.

Reference	Method	Classification of hand motion	Does not interfere with grasping	Angle of view	Degree of freedom of shape	Identification of contact area	Division by Xue et al. (2019)
Wallace and Whishaw (2003)	Video coding	Intra-rater and inter-rater reliability were 90% and 74%, respectively, with four grasp patterns	No	Limited	High	Partially possible	5) vision-based capturing
Campolo et al. (2008)	Hemispherical sensing devices using force sensors and kinematic sensor	Not reported	No	Not limited	Unclear	Not reported	2) attached force-based capturing
Boschi and Frère (2013)	Sensing devices using limit switches, micro switches	The agreement from the physical therapists and the system was 86.6% for five different movements	No	Not limited	Low	Partially possible	2) attached force-based capturing
Del Maestro et al. (2011) Serio S et al. (2013) (Serio et al., 2013)	Sensing device using air pressure sensor	Not reported	No	Not limited	Low	Impossible	2) attached force-based capturing
Rocha et al. (2016)	Cylindrical sensing device digital camera, a special convex mirror, and IMU sensor	A preliminary hand posture evaluation was reported, though visually performed, suggesting the possibility of future application	No	The image quality degrades at the vertex part of the convex mirror when converted to panoramic format	Unclear	Possible	2) attached force-based capturing & 5) vision-based capturing
Schröer et al. (2021)	Optical motion capture	Recording of hand motion (reaching)	Possible interference	Not limited	Unclear	Impossible	4) optical markers-based capturing
Owada et al. (2022)	Data glove	Classified eight grasps with an accuracy of 98.75% in a study of adults (Pratap et al., 2024)	Possible interference	Not limited	High	Partially possible	1) data glove-based capturing
Udayagiri et al. (2024)	Optical force sensors	Classified four actions of adults with an accuracy of approximately 100%	No	Not limited	High	Not reported	2) attached force-based capturing
Battraw et al. (2024)	Surface electromyography	Nine participants with unilateral congenital below-elbow deficiency were classified into 11 hand movements, with a maximum accuracy of 95.37% using KNN.	Possible interference	Not limited	Unclear	Impossible	3) surface electromyography-based capturing
Our study	Tomographic tactile sensor based on resistive coupling	The maximum average accuracies classed by power grip or precision grip was 88.5%	No	Not limited	High	Possible	2) attached force-based capturing

associated with precision grip are related to early numerical skills, early counting, and conceptual counting knowledge ([Barrocas et al., 2020](#); [Fischer et al., 2018](#); [Suggate et al., 2017](#)). Another study reported that children who were trained in FMS through intervention improved not only their pegboard test scores but also their mathematical performance compared to a control group that read books ([Asakawa et al., 2019](#)). Moreover, some studies have indicated a potential correlation between the FMS

and reading and writing abilities ([Lê et al., 2023](#); [Suggate et al., 2023](#)). These findings suggest that children's hand dexterity is related to the development of academic abilities such as mathematical, reading, and writing skills.

In addition, information on children's grasping skill development is beneficial from a medical perspective. A delay in the development of infant FMS has been suggested as a useful indicator for early diagnosis of developmental disorders. Autism

spectrum disorder (ASD), which is characterized by poor communication, strong interest, preoccupation, and obsessive behavior, is usually diagnosed at 3 years of age. However, some prospective studies have reported that infants at high risk for ASD, who were later diagnosed with the condition exhibited developmental delays in fine motor skills compared to typically developing infants (Choi et al., 2018; Landa and Garrett-Mayer, 2006). It has also been reported that children of ages five to ten with attention deficit hyperactivity disorder (ADHD), which is characterized by inattention, hyperactivity, and impulsiveness, have significantly delayed motor development in all domains of the Motor Development Scale, including the FMS, compared to typically developing children (Neto et al., 2015).

In this context, quantitative techniques for monitoring and measuring hand motion in infants and children are important, because they are believed to contribute to healthy development. According to a review by Xue et al., human hand motion analysis can be classified into five methods: 1) data glove-based capturing; 2) attached force-based capturing; 3) surface electromyography-based capturing; 4) optical markers-based capturing; and 5) vision-based capturing (Xue et al., 2019). Based on the review, conventional devices for children listed in Table 1. Although these methods have unique advantages, they have unavoidable limitations. For example, the methods of 1) data glove-based capturing, 3) surface electromyography-based capturing, and 4) optical marker-based capturing require sensors to be attached. This could lead to a decrease in the children's attention and interfere with their grasp. In addition, the method of 5) vision-based capturing has some limitations, such as a limited angle of view and privacy. Owing to these limitations, 2) attached force-based capturing is considered a powerful method. However, the conventional method limited the degree of freedom of shape and identification of contact area. To overcome these limitations, we focused on a *tomographic tactile sensor based on resistive coupling*, which is a sensing technology with extended flexibility, shape versatility, and designability compared with electrical impedance tomography (EIT)-based tactile sensors (Yoshimoto et al., 2024; 2020). This technology is based on the principles of EIT-based tactile sensors (Kato et al., 2007; Nagakubo et al., 2007; Silveira-Tawil et al., 2015).

Park et al. reported the superior discriminability of touch modalities using a tomographic tactile sensor (Park et al., 2021). Additionally, we developed a small peg-based device and demonstrated that six types of pinching in adults could be classified with an accuracy exceeding 80% (Asahi et al., 2024b). In children, we have reported a classification study on the same six types of pinching. The results showed an accuracy of approximately 60%, which was lower than that of adults, revealing limitations and challenges in classifying children's hand motion (Asahi et al., 2024a).

Children's hands differ from those of adults in terms of size, grip strength, and dexterity (Bear-Lehman et al., 2002). Consequently, results obtained from adult participants may not be directly applicable to children. Moreover, protective measures for the device and an extension of its swing range are necessary, as children may lick or swing the device. Considering these factors, as a first step toward developing a hand motion analysis system for children, we focused on the fundamental classification of power grip and precision grip defined by Napier (1956), along with the implementation of protective measures and an extended swing range. Thus, demonstrating the ability to classify power and

precision grips in this study represents an essential step toward more comprehensive and generalized grasp classifications. Our findings may contribute to the development of educational toys and diagnostic systems for assessing developmental disabilities.

2 Materials and methods

2.1 Sensing device overview

2.1.1 Development of sensing devices

Regarding the development of sensing device, we first decided the design requirements of the device. The previous devices had measurement circuits outside the device. This limits their swing range and portability. To overcome these limitations, we developed a new cylindrical sensing device. It could contain the measurement circuit (Figure 1). The device height and diameter were 85 mm and 40 mm, respectively. The sensor consists of five layers: protective, driving, insulating, detection, and electrode layers (Figure 2a). The protective layer was added because the children torn through the drive layer during the preliminary experiments. The protective layer consisted of a 1-mm-thick yellow felt cloth. The driving layer consisted of a conductive silicone sheet (EC-20BH, Shin-Etsu Chemical Co. Ltd.). This layer was connected to a 3.3 V DC voltage source. The DC voltage source used a 3.3 V pin microcontroller (ESP32-DevKitC, Espressif Systems). The insulating layer was a glass fiber sheet (13-7127, KLASS). The detection layer was composed of a conductive sheet (ZC-85, ENGINEER) with a surface resistance of 10 k Ω /sq. When the driving layer contacts with the detection layer, the electrical circuit is closed. As a result, current flows to the electrodes through the detection layer (Figure 2b). The electrode and detection layer were bonded using a conductive epoxy (CW2400, CircuitWorks). The electrode layer was an original flexible printed circuit board with 16 electrodes and a diameter of 2 mm. One of these electrodes was used as the ground condition and the other was used as the measurement electrode. This operation was repeated until all electrodes were used in all conditions. Thus, 256 voltage data points (16 grounding conditions $\times$ 16 electrodes) are obtained. The multiplexer (MUX) used was CD74HC4067 (Texas Instruments). The measurement period was 0.25 s per frame.

2.1.2 Reconstruction

A tomographic tactile sensor requires a solver to reconstruct pressure distribution from the measured voltage vector. This solver addresses an ill-posed problem because the output dimension (reconstructed image) is larger than the input dimension (measured voltage vector). Therefore, we used the linear reconstruction method of Tikhonov regularization with two-dimensional finite element method (FEM) model, based on our previous studies (Asahi et al., 2024b; Yoshimoto et al., 2020). The hyperparameter of the Tikhonov regularization was set to 5000. These reconstruction processes were performed using MATLAB 2023b (MathWorks Inc.).

2.2 Experiment methods

2.2.1 Participant information

In this study, participants were required to meet two criteria: (1) the ability to perform both power grip and precision grip, and (2) an

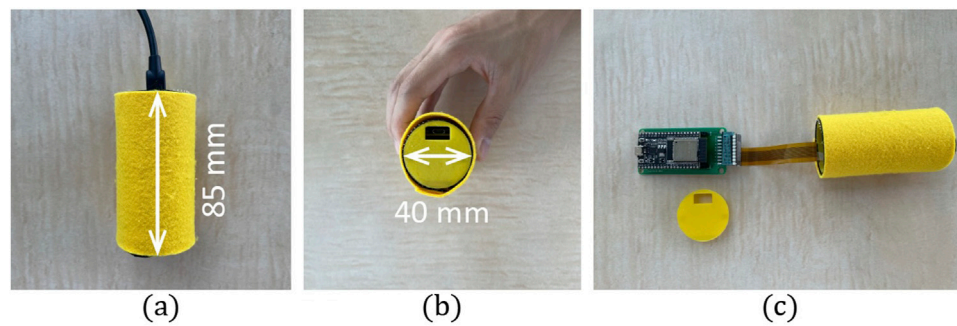


FIGURE 1
Cylindrical sensing device. (a) Height: 85 mm (b) Diameter: 40 mm. (c) Device containing the measurement circuit.

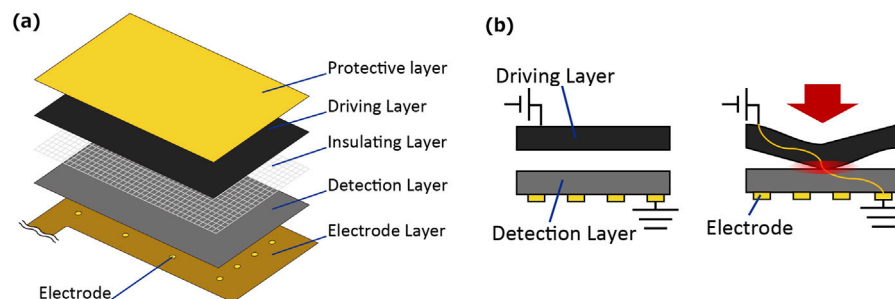


FIGURE 2
(a) Layer of tomographic tactile sensors based on resistive coupling. The protective layer was a yellow felt cloth. This layer was used only in experiment I. A conductive silicone sheet connected to a DC voltage source (3.3 V) was used as the driving layer. The insulating layer was a glass fiber sheet. The detection layer was a conductive sheet. The electrode layer was a flexible printed circuit board. (b) Schematic illustration of contact between driving layer and detection layer and current flow.

age at which they could sufficiently understand verbal instructions. Regarding (1) the ability to perform power grip and precision grip, as mentioned in the introduction, primitive grasping has been reported to emerge between 2 and 22 weeks, while precision grip develops at 5.97 months. Regarding (2) the ability to sufficiently understand verbal instructions, B. Buckley have reported that three-year-old children are capable of communicating using language. Additionally, four-year-old children can focus on and follow verbal instructions even without explicit cues, such as being called by name (Buckley, 2003). Based on these considerations, this study targeted four-year-old children as participants. The participants of this study were 17 children (4.43 ± 0.30 years old, 8 boys, 9 girls). In order to ensure that the participants had adequate communication skills and no developmental disorders, we administered the KINDER INFANT DEVELOPMENT SCALE questionnaire type C prior to the experiment (Hassanein, 1982). None of the participants had any serious disease or disorder. In the analysis, 11 children (5 boys and 6 girls) were included, excluding those who stopped the measurement halfway through because they could not listen to the experimenter's instructions or did not want to participate (4 participants), those whose actual dominant hand seemed to differ from that reported by their parents (1 participant), and those who grasped without placing their palm on the object during the power grip (1 participant). Evaluation of the modified Japanese version of the FLANDERS handedness questionnaire

(Okubo et al., 2014) indicated 10 right-handed children and one left-handed child.

This study was approved by the Ethics Committees of Shibaura Institute of Technology and Keio University. The experiment was conducted only when informed consent was obtained from the parents of the participating children.

2.2.2 Measurement method

The participants practiced freely grasping the device, without external assistance. Voltage measurements were also performed during the hardware and software testing. The participants were then instructed to perform either a power grip or precision grip, with the order being counterbalanced. Each grip was measured ten times. However, if the hand was released during the measurement process or if the grasp force was not applied (i.e., the object was grasped only by the frictional force of the fingers), the grasp was excluded from the analysis, and an additional measurement was performed. Ten times per grasp were measured, that is, 100 measurement frames ($10 \text{ measurements} \times 10 \text{ times}$) were obtained for each grasping category.

2.2.3 Classification method

The measured voltage vectors and reconstructed images were used as features to classify the power grip and precision grip. For classification using the measured voltage vectors, 256 data points

TABLE 2 CNN classification network architecture.

No	Layer	Description
1	Input 2D	Each classification has different inputs
2	Convolutional 2D	32 3×3 convolutions with stride 1
3	Batch Normalization	Batch Normalization
4	ReLU	ReLU
5	Max Pooling 2D	2 $\times$ 2 Max Pooling
6	Convolutional 2D	64 3×3 convolutions with stride 1
7	Batch Normalization	Batch Normalization
8	ReLU	ReLU
9	Max Pooling 2D	2 $\times$ 2 Max Pooling
10	Convolutional 2D	128 3×3 convolutions with stride 1
11	Batch Normalization	Batch Normalization
12	ReLU	ReLU
13	Fully Connected	256 fully connected
14	ReLU	ReLU
15	Fully Connected	2 fully connected
16	Softmax	Softmax
17	Classification	Output layer

were used (16 electrodes $\times$ 16 measurements). For classification using the reconstructed images, the FEM values normalized from 0% to 100% were used. The size of FEM model was $43 \times 66 \times 1$.

In this study, we considered more practical applications, such as educational toys or diagnostic systems for developmental disabilities. For such applications, we conducted a comparative analysis using the k-Nearest Neighbors (KNN) algorithm (Bansal et al., 2022), which is computationally efficient and easy to implement, and the Convolutional Neural Network (CNN), which has been reported to achieve high classification accuracy (Park et al., 2021). In KNN, we used the MATLAB function (fitcknn) with a k value of 1. The network architecture of the CNN consisted of 17 layers, as listed in Table 2. The input layer was configured to input each feature with dimensions of $16 \times 16 \times 1$ for classification using the measured voltage vectors, and $43 \times 66 \times 1$ for classification using the reconstructed images. Subsequently, a three-step convolution was performed. The convolution layers were organized with filter sizes of $3 \times 3 \times 32$, $3 \times 3 \times 64$, and $3 \times 3 \times 128$, in that order. In each convolution layer, the ReLU was applied as the activation function after batch normalization. In steps one and two, a 2×2 max pooling layer is utilized, resulting in downsampling with a stride of two. In step three, a fully connected layer was used to classify the data into two classes. The Softmax function was applied to the output layer, resulting in a final classification into two classes: power grip and precision grip. Stochastic gradient descent was employed for training with an initial learning rate of 0.001. In addition, the learning rate was configured to be reduced by 95% after ten epochs. The maximum number of epochs was set to 36, and the data were randomized at the beginning of each epoch.

Cross-validation was performed to validate the classification accuracy of these two types of features and classification methods.

Verification was performed for each participant, with one grasp (10 frames) as the test data and the remaining grasps (190 frames) as the training data, and was repeated until all grasps were the test data.

For the evaluation of classification results, the following accuracy was calculated for each participant as Equation 1:

$$\text{Accuracy} = \frac{T_{\text{Pow}} + T_{\text{Pre}}}{T_{\text{Pow}} + T_{\text{Pre}} + F_{\text{Pow}} + F_{\text{Pre}}} \quad (1)$$

where, initial character *T* or *F* indicates whether the class predicted by the classifier matches the true class or not. The characters *Pow* and *Pre* indicate power grip and precision grip, respectively. Thus, T_{Pow} is the matching case of the power grip as predicted class by the classifier and the power grip as true class. Subsequently, to evaluate the classification performance in different grasps, a confusion matrix was calculated for each grasp. Precision (Prec), recall, and F-measure were then calculated for each grasp as follow Equations 2–4:

$$\text{Prec}_{\text{Pow or Pre}} = \frac{T_{\text{Pow or Pre}}}{T_{\text{Pow or Pre}} + F_{\text{Pow or Pre}}} \quad (2)$$

$$\text{Recall}_{\text{Pow or Pre}} = \frac{T_{\text{Pow or Pre}}}{T_{\text{Pow or Pre}} + F_{\text{Pre or Pow}}} \quad (3)$$

$$F\text{-measure}_{\text{Pow or Pre}} = \frac{2 \times \text{Prec} \times \text{Recall}}{\text{Prec} + \text{Recall}} \quad (4)$$

These index values ranged from 0 to 1. In addition, t-Distributed Stochastic Neighbor Embedding (t-SNE) has been used to confirm the distribution of features and clustering trends (Van Der Maaten and Hinton, 2008). The classification features were compressed into a two-dimension map by t-SNE. The distances between points in the t-SNE plot reflect similarity relationships in the original high-dimensional space. Additionally, clearly separated clusters indicate natural groupings based on differences in classification features.

3 Results

We classified the basic grip classifications—power grip and precision grip—in children (Figure 3a). The average accuracy values are shown in Figure 4a. When classified using the measured voltage vector (Figure 3b), the average classification accuracy using KNN was 86.8%. The highest and lowest accuracies for the participants were 95.5% and 79.0%, respectively. The average accuracy obtained using the CNN was 88.5%, and the highest and lowest accuracies were 95.0% and 75.5%, respectively. When classifying using the reconstructed image as a feature (Figure 3c), the average accuracy was 85.7% using KNN. The highest and lowest accuracies are 94.5% and 72.5%, respectively. In the classification using CNN with the reconstructed image as the feature, the average accuracy was 87.9%. The highest and lowest accuracies are 99.0% and 77.0%, respectively. All average classification accuracies exceeded the chance level (50%).

The confusion matrixes shows that the classification results were better for the precision grip than for the power grip for all classification methods (Figure 4b). As shown in Table 3, Prec was higher for power grip across all classifiers. However, Recall was higher for the precision grip in all cases. Similarly, the F-measure was also higher for the precision grip across all classifiers. Figure 5 presents the t-SNE plot of the participant who achieved the highest accuracy. The results

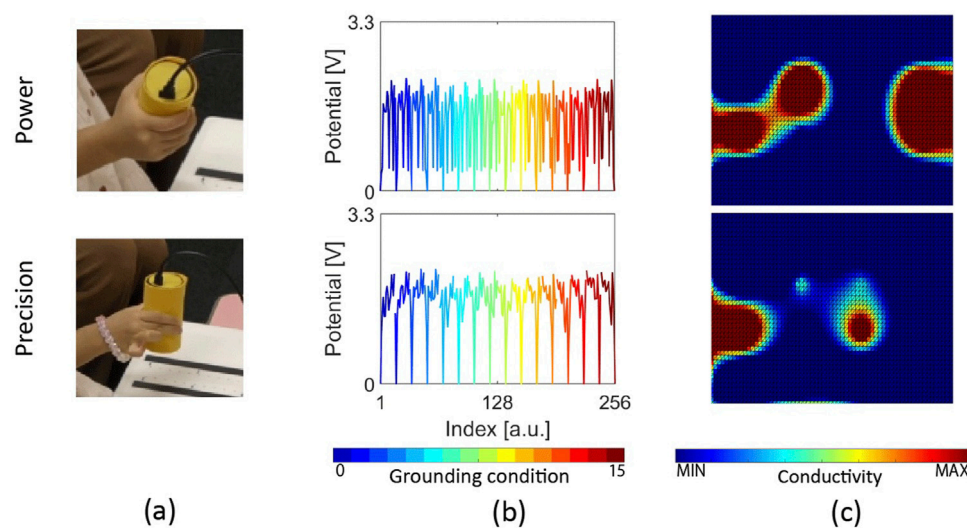


FIGURE 3 Representative images. The upper section corresponds to the power grip, and the lower section corresponds to the precision grip. (a) Images captured for each grasp. (b) Measured voltage vector. (c) Reconstructed 2D image.

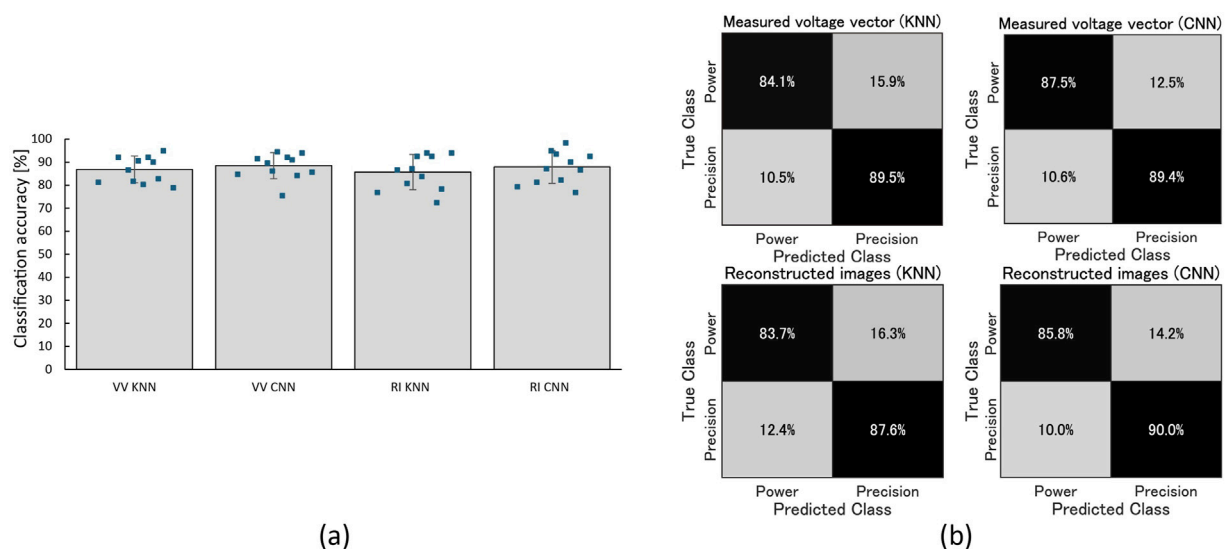


FIGURE 4 (a) Mean and standard deviation of classification accuracy. The average classification accuracies of the measured voltage vector (VV) with KNN and CNN classifications were 86.8% and 88.5%, respectively. The average classification accuracies of the reconstructed images (RI) with KNN and CNN classifications were 85.7% and 87.9%, respectively. (b) Cross-validation of each classification. The closer the color of the diagonal cell is to black, the higher the classification accuracy.

indicated that clustering was achieved based on feature labels. However, some instances of the power grip appear to be located within the precision grip cluster.

4 Discussion

The purpose of this study is to demonstrate the feasibility of classifying power grip and precision grip in children using a

tomographic tactile sensor based on resistive coupling, as a first step towards the development of a hand motion analysis system for children. In order to avoid accidents and damage due to unexpected behavior of children, the sensor device and measurement circuit were integrated into a single unit and a felt fabric layer was added. As a result, there was no damage to the device during the experiment. Using this device, the results of power and precision grip classification showed an average classification accuracy of over 85%, higher than the chance level of 50%.

TABLE 3 Classification index.

Accuracy [%]		VV KNN	VV CNN	RI KNN	RI CNN
		86.8	88.5	85.7	87.9
Prec	power	0.889	0.892	0.871	0.896
	precision	0.849	0.877	0.843	0.864
Recall	power	0.841	0.875	0.837	0.858
	precision	0.895	0.894	0.876	0.900
F-measure	power	0.864	0.883	0.854	0.876
	precision	0.871	0.886	0.859	0.881

4.1 Sensing device and measurements on children

Children's behavior is differed from adults, they may lick or throw the device. In fact, during a preliminary experiment, one child scratched the driving layer with their fingernails, causing damage. To address such risks, the device must be designed to minimize potential breakage factors. As one protective measure, we introduced a felt fabric with cushioning properties as a protective layer. This layer serves to prevent licking and damage to the driving layer. Such felt fabric is suitable as a protective material for children due to its safe composition, durability, availability in various colors, and flexibility. Regarding the safety of the device itself, the felt fabric and PLA material used in this study are non-toxic and hypoallergenic, ensuring no safety concerns.

Additionally, to prevent damage from being thrown or mishandled, the device needed to cover a broader swing range. To achieve this, the sensor and measurement circuit were integrated into a sensing device. Furthermore, a 2-meter cable was used for communication between the PC and the measurement circuit. As a result of these design improvements, no participants damaged the device during the experiment. Furthermore, wireless measurement is an effective approach to improving operational range and portability. In this study, the computer and sensing device were connected via a cable.

However, as reported by Yoshimoto et al., wireless measurement is feasible, and its implementation is expected to eliminate limitations in the measurement environment (Yoshimoto et al., 2020).

Finally, regarding the experimental protocol, grasping motions were instructed verbally in this experiment. However, some participants treated the device as a cup, mimicking toasting or pretending to drink from it. This suggests that a role-play-based protocol may be more suitable for future studies.

4.2 Classification method

Reconstructed images and measured voltage vectors were used as classification features, and both CNN- and KNN-based methods achieved an average accuracy exceeding 85%, which is higher than the 50% chance level. These findings indicate that classification using a tomographic tactile sensor can achieve a high classification performance and reproducibility. The difference in classification accuracy among all classifications was 2.8%. Thus, it demonstrated a comparable classification accuracy across all methods.

Analysis of the confusion matrix revealed that the precision grip was classified more accurately than the power grip. Although the Prec for the power grip was higher than that for the precision grip across all methods, the recall and F-measure for the precision grip were higher (Table 3). These results suggest that while all classification methods correctly identified the precision grip, the power grip was frequently misclassified as the precision grip. Moreover, as illustrated in the t-SNE plots (Figure 5), some plots in the power grip were mixed in the precision grip cluster. This indicates that the classification error did not depend on the classifier but rather on the potential classification features.

One potential factor differentiating the power grip from the precision grip is the contact area. Visual assessments indicated that the contact area of the power grip was larger than that of the precision grip (Figure 3c). In the power grip, opposition is generated by the thumb, other fingers, and the palm; however, in the precision grip, opposition was generated by the thumb and other fingers (Figure 3a). This difference is considered to be the cause of the difference in the

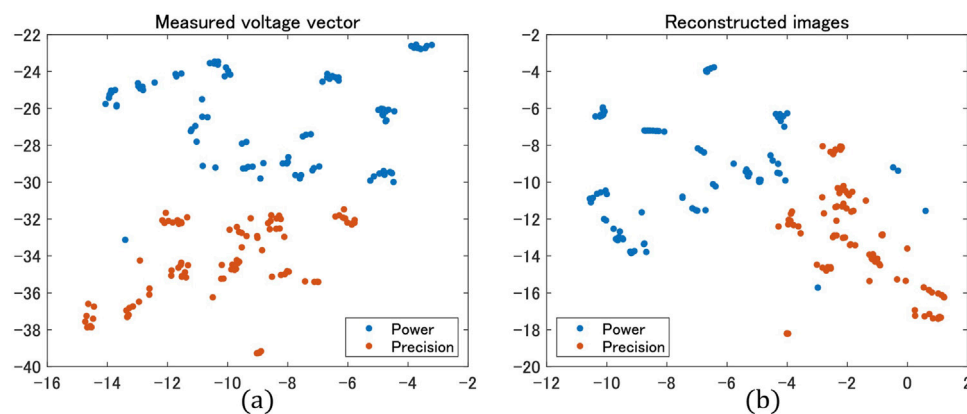


FIGURE 5 T-SNE plots with measured voltage vectors and reconstructed images for participants who achieved the highest classification accuracy. The perplexity was set to 50, and the learning rate was set to 750. The input classification features were normalized. (a) t-SNE plot using the measured voltage vector (b) t-SNE plot obtained using the reconstructed images.

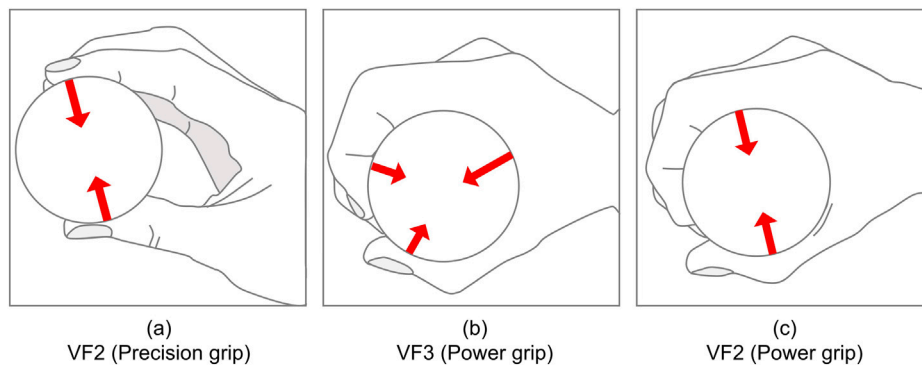


FIGURE 6
Illustration of Virtual fingers. (a) Precision grip in VF2. (b) Power grip in VF3. (c) Power grip in VF2.

contact area. Another distinguishing factor is the number of virtual fingers (VF), which indicates the number of primary force vectors generated during grasping. The precision grip is characterized by VF2 due to its pinching action. In contrast, the power grip can be characterized by either VF2 or VF3, as it involves both the fingers and the palm (Figure 6). These differences in contact area and VF may contribute to the misclassification of the power grip relative to the precision grip. It is also possible that the power grip in t-SNE is the cause of some mixing with the precision grip cluster (Figure 5).

A previous study examining the agreement between experienced physical therapists and devices for analyzing the locations of gripping (the power grip) and functional pinching (the precision grip) reported an agreement of 86.6% (Boschi and Frère, 2013). Although direct comparisons could not be made owing to the differences in the experimental conditions, the results of this study demonstrated comparable accuracy.

4.3 Limitations and future prospects

There are two main limitations in this study. First is that the device size was fixed at a height of 85 mm and a diameter of 40 mm. Customizing the device to accommodate individual hand sizes could enhance its ease of grasping, and improve classification accuracy. Second is the grasp types used for classification. In this study, the basic categories of the power grip and precision grip were classified. For practical applications, a more detailed classification of grasping may be necessary.

Based on the results of this study, there are three prospects for future research: The first is to develop a sensing device with a system that provides humorous feedback stimuli. Feedback systems encourage children to take action (Boschi and Frère, 2013). The feedback system that uses the grasp classification system developed in this study may contribute to rehabilitation and intervention. The second is to identify the specific fingers contacting the sensor. This identification system might be a useful alternative to the video coding. The video coding typically requires a lot of time and effort. Previous studies have reported that it took 3 hours to code a 10-min video (Wallace and Whishaw, 2003). Replacing video coding with sensor-based analysis could reduce the analysis time. To achieve this, a large amount of data on the children must be collected. We believe that making the device toy-shaped will help keep children's attention and enable the measurement

of a large amount of data. Lastly, we propose the potential application of this system as a diagnostic support tool for developmental disorders. Previous studies have reported that children at high risk for ASD often exhibit delayed development of fine motor skills compared to typically developing children. Accordingly, if the present system can be employed to assess fine motor skills in both typically developing children and high-risk ASD children, it may contribute to early diagnostic support for ASD. To examine this feasibility, future research should aim to measure and compare grasping behaviors between these two groups.

5 Conclusion

In this study, we demonstrated the feasibility of classifying power grip and precision grip in children using a tomographic tactile sensor based on resistive coupling. To address concerns that children might lick or swing the device, we developed a medium-sized cylindrical sensing device with an integrated sensor and measurement circuit, incorporating a protective layer for enhanced safety. These design considerations ensured that no damage occurred to the device during the experiment.

Using the device, machine learning-based classification of children's grasps demonstrated that power grip and precision grip could be classified with an accuracy exceeding 85%, above the chance level of 50%. These grip types are among the major categories in the GRASP taxonomy, which defines 33 distinct grasp classifications. Therefore, the findings of this study represent a foundational step toward classifying a broader range of grasp types and establishing a comprehensive grasp classification system.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by The biotechnology ethics committee of Shibaura Institute of

Technology (24-009), and the ethics committee of Keio University, Faculty of Literature (240290000). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

RA: Conceptualization, Writing – original draft, Writing – review and editing, Software, Investigation, Formal Analysis, Data curation, Visualization, Methodology. ST: Writing – review and editing, Investigation. YM: Resources, Writing – review and editing. SY: Writing – review and editing. HS: Project administration, Funding acquisition, Writing – review and editing, Conceptualization, Supervision, Investigation.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by JSPS KAKENHI (Grant Number 23K25652).

Acknowledgments

We would like to express our gratitude to the children and their parents who participated in this study. We also wish to thank Hina Furukawa from the Shibaura Institute of

Technology, Hina Katayama from Keio University, and the staff members of Keio University for their support during the experiments.

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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

- Allen, M. C., and Capute, A. J. (1986). The evolution of primitive reflexes in extremely premature infants. *Pediatr. Res.* 20, 1284–1289. doi:10.1203/00006450-198612000-00018
- Asahi, R., Yoshimoto, S., Fujita, T., Toriyama, S., Shimada, Y., Itakura, S., et al. (2024a). "Classification of pinching action in children using a tomographic tactile sensor," in *2024 IEEE international symposium on medical measurements and applications, MeMeA 2024 - proceedings*. doi:10.1109/MeMeA60663.2024.10596747
- Asahi, R., Yoshimoto, S., and Sato, H. (2024b). Development of pinching motion classification method using EIT-based tactile sensor. *IEEE Access* 12, 62089–62098. doi:10.1109/ACCESS.2024.3395271
- Asakawa, A., Murakami, T., and Sugimura, S. (2019). Effect of fine motor skills training on arithmetical ability in children. *Eur. J. Dev. Psychol.* 16, 290–301. doi:10.1080/17405629.2017.1385454
- Bansal, M., Goyal, A., and Choudhary, A. (2022). A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. *Decis. Anal. J.* 3, 100071. doi:10.1016/j.dajour.2022.100071
- Barrocas, R., Roesch, S., Gawrilow, C., and Moeller, K. (2020). Putting a finger on numerical development – reviewing the contributions of kindergarten finger gnosis and fine motor skills to numerical abilities. *Front. Psychol.* 11, 1012. doi:10.3389/fpsyg.2020.101012
- Battraw, M. A., Fitzgerald, J., Winslow, E. J., James, M. A., Bagley, A. M., Joiner, W. M., et al. (2024). Surface electromyography evaluation for decoding hand motor intent in children with congenital upper limb deficiency. *Sci. Rep.* 14, 31741. doi:10.1038/s41598-024-82519-z
- Bear-Lehman, J., Kafko, M., Mah, L., Mosquera, L., and Reilly, B. (2002). An exploratory look at hand strength and hand size among preschoolers. *J. Hand Ther.* 15, 340–346. doi:10.1016/S0894-1130(02)80005-9
- Boschi, S. R. M. S., and Frère, A. F. (2013). Grip and pinch capability assessment system for children. *Med. Eng. Phys.* 35, 626–635. doi:10.1016/j.medengphy.2012.07.008
- Buckley, B. (2003). *Children's Communication Skills: From Birth to Five Years*. Routledge. doi:10.4324/9780203865750
- Campolo, D., Taffoni, F., Schiavone, G., Laschi, C., Keller, F., and Guglielmelli, E. (2008). "A novel technological approach towards the early diagnosis of neurodevelopmental disorders," in *Proceedings of the 30th annual international conference of the IEEE engineering in medicine and biology society, EMBS'08 - "personalized healthcare through technology"*. doi:10.1109/iembs.2008.4650306
- Choi, B., Leech, K. A., Tager-Flusberg, H., and Nelson, C. A. (2018). Development of fine motor skills is associated with expressive language outcomes in infants at high and low risk for autism spectrum disorder. *J. Neurodev. Disord.* 10, 14. doi:10.1186/s11689-018-9231-3
- Del Maestro, M., Cecchi, F., Serio, S. M., Laschi, C., and Dario, P. (2011). Sensing device for measuring infants' grasping actions. *Sens. Actuators A Phys.* 165, 155–163. doi:10.1016/j.sna.2010.08.016
- Feix, T., Romero, J., Schmiedmayer, H. B., Dollar, A. M., and Kragic, D. (2016). The GRASP taxonomy of human grasp types. *IEEE Trans. Hum. Mach. Syst.* 46, 66–77. doi:10.1109/THMS.2015.2470657
- Fischer, U., Suggate, S. P., Schmir, J., and Stoeger, H. (2018). Counting on fine motor skills: links between preschool finger dexterity and numerical skills. *Dev. Sci.* 21, e12623. doi:10.1111/desc.12623
- Hassanein, R. S. (1982). The KIDS Chart: A Simple, Reliable Infant Development Screening Tool. *Am. J. Dis. Child* 136, 1012. doi:10.1001/archpedi.1982.03970470041012
- Kato, Y., Mukai, T., Hayakawa, T., and Shibata, T. (2007). "Tactile sensor without wire and sensing element in the tactile region based on EIT method," in *Proceedings of IEEE sensors*. doi:10.1109/ICSENS.2007.4388519
- Landa, R., and Garrett-Mayer, E. (2006). Development in infants with autism spectrum disorders: a prospective study. *J. Child. Psychol. Psychiatry* 47, 629–638. doi:10.1111/j.1469-7610.2006.01531.x

- Lê, M., Quémart, P., Potocki, A., Gimenes, M., Chesnet, D., and Lambert, E. (2023). Improving literacy development with fine motor skills training: a digital game-based intervention in fourth grade. *Cogn. Dev.* 67, 101363. doi:10.1016/j.cogdev.2023.101363
- Nagakubo, A., Alirezai, H., and Kuniyoshi, Y. (2007). "A deformable and deformation sensitive tactile distribution sensor," in *2007 IEEE international conference on robotics and biomimetics, ROBIO*. doi:10.1109/ROBIO.2007.4522352
- Napier, J. R. (1956). The prehensile movements of the human hand. *J. Bone Jt. Surg. Br.* 38, 902–913. doi:10.1302/0301-620x.38b4.902
- Neto, F. R., Goulardins, J. B., Rigoli, D., Piek, J. P., and de Oliveira, J. A. (2015). Motor development of children with attention deficit hyperactivity disorder. *Rev. Bras. Psiquiatr.* 37, 228–234. doi:10.1590/1516-4446-2014-1533
- Okubo, M., Suzuki, H., and Nicholls, M. E. R. (2014). A Japanese version of the FLANDERS handedness questionnaire. *Jpn. J. Psychol.* 85, 474–481. doi:10.4992/jjpsy.85.13235
- Owada, S., Sugimura, H., and Isshiki, M. (2022). "Toddler's hand motion acquisition with hand-made data glove," in *LifeTech 2022 - 2022 IEEE 4th global conference on life sciences and technologies*. doi:10.1109/LifeTech53646.2022.9754849
- Park, H., Park, K., Mo, S., and Kim, J. (2021). Deep neural network based electrical impedance tomographic sensing methodology for large-area robotic tactile sensing. *IEEE Trans. Robotics* 37, 1570–1583. doi:10.1109/TRO.2021.3060342
- Pratap, S., Narayan, J., Hatta, Y., Ito, K., and Hazarika, S. M. (2024). Glove-net: enhancing grasp classification with multisensory data and deep learning approach. *Sensors* 24, 4378. doi:10.3390/s24134378
- Rocha, A. C. P., Tudella, E., Pedro, L. M., Appel, V. C. R., Da Silva, L. G. P., and Caurin, G. A. de P. (2016). A novel device for grasping assessment during functional tasks: preliminary results. *Front. Bioeng. Biotechnol.* 4, 16. doi:10.3389/fbioe.2016.00016
- Schröer, L., Cooper, R. P., and Mareschal, D. (2021). Science with Duplo: multilevel goal management in preschoolers' toy house constructions. *J. Exp. Child. Psychol.* 206, 105067. doi:10.1016/j.jecp.2020.105067
- Serio, S. M., Cecchi, F., Assaf, T., Laschi, C., and Dario, P. (2013). Design and development of a sensorized wireless toy for measuring infants' manual actions. *IEEE Trans. Neural Syst. Rehabilitation Eng.* 21, 444–453. doi:10.1109/TNSRE.2013.2255066
- Silvera-Tawil, D., Rye, D., Soleimani, M., and Velonaki, M. (2015). Electrical impedance tomography for artificial sensitive robotic skin: a review. *IEEE Sens. J.* 15, 2001–2016. doi:10.1109/JSEN.2014.2375346
- Suggate, S., Stoeger, H., and Fischer, U. (2017). Finger-based numerical skills link fine motor skills to numerical development in preschoolers. *Percept. Mot. Ski.* 124, 1085–1106. doi:10.1177/0031512517727405
- Suggate, S. P., Karle, V. L., Kipfelsberger, T., and Stoeger, H. (2023). The effect of fine motor skills, handwriting, and typing on reading development. *J. Exp. Child. Psychol.* 232, 105674. doi:10.1016/j.jecp.2023.105674
- Udayagiri, R., Yin, J., Cai, X., Townsend, W., Trivedi, V., Shende, R., et al. (2024). Towards an AI-driven soft toy for automatically detecting and classifying infant-toy interactions using optical force sensors. *Front. Robot. AI* 11, 1325296. doi:10.3389/frobt.2024.1325296
- Van Der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9.
- Wallace, P. S., and Whishaw, I. Q. (2003). Independent digit movements and precision grip patterns in 1-5-month-old human infants: hand-babbling, including vacuous then self-directed hand and digit movements, precedes targeted reaching. *Neuropsychologia* 41, 1912–1918. doi:10.1016/S0028-3932(03)00128-3
- Xue, Y., Ju, Z., Xiang, K., Chen, J., and Liu, H. (2019). Multimodal human hand motion sensing and analysis-A review. *IEEE Trans. Cogn. Dev. Syst.* 11, 162–175. doi:10.1109/TCDS.2018.2800167
- Yoshimoto, S., Kuroda, Y., and Oshiro, O. (2020). Tomographic approach for universal tactile imaging with electromechanically coupled conductors. *IEEE Trans. Industrial Electron.* 67, 627–636. doi:10.1109/TIE.2018.2879296
- Yoshimoto, S., Sakamoto, K., Takeda, R., and Yamamoto, A. (2024). Design of a high-performance tomographic tactile sensor by manipulating the detector conductivity. *IEEE Trans. Industrial Electron.* 71, 16783–16791. doi:10.1109/TIE.2024.3384613

Frontiers in Computational Neuroscience

Fosters interaction between theoretical and experimental neuroscience

Part of the world's most cited neuroscience series, this journal promotes theoretical modeling of brain function, building key communication between theoretical and experimental neuroscience.

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

