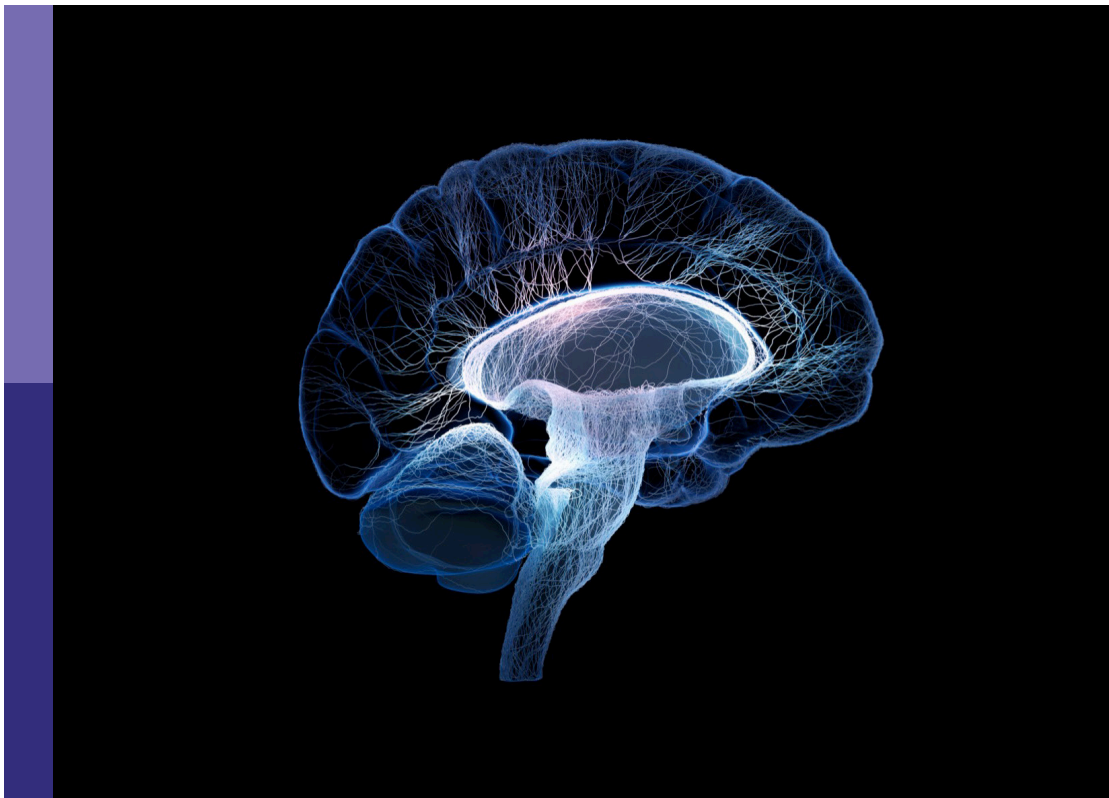
frontiers | Research Topics

The application of artificial intelligence in brain-computer interface and neural system rehabilitation

Edited by
Fangzhou Xu, Dong Ming, Tzyy-Ping Jung, Peng Xu and Minpeng Xu

Published in
Frontiers in Neuroscience



ISSN 1664-8714
ISBN 978-2-8325-3902-6
DOI 10.3389/978-2-8325-3902-6

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

The application of artificial intelligence in brain-computer interface and neural system rehabilitation

Topic editors

Fangzhou Xu — Qilu University of Technology, China
Dong Ming — Tianjin University, China
Tzyy-Ping Jung — University of California, San Diego, United States
Peng Xu — University of Electronic Science and Technology of China, China
Minpeng Xu — Tianjin University, China

Citation

Xu, F., Ming, D., Jung, T.-P., Xu, P., Xu, M., eds. (2023). *The application of artificial intelligence in brain-computer interface and neural system rehabilitation*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3902-6

Table of contents

frontiers | Frontiers in Neuroscience

TYPE Editorial
PUBLISHED 27 October 2023
DOI 10.3389/fnins.2023.1290961

OPEN ACCESS

EDITED AND REVIEWED BY
Michela Chiappalone,
University of Genoa, Italy

*CORRESPONDENCE
Fangzhou Xu
xfz@qlu.edu.cn

RECEIVED 08 September 2023
ACCEPTED 05 October 2023
PUBLISHED 27 October 2023

CITATION
Xu F, Ming D, Jung T-P, Xu P and Xu M (2023) Editorial: The application of artificial intelligence in brain-computer interface and neural system rehabilitation.
Front. Neurosci. 17:1290961.
doi: 10.3389/fnins.2023.1290961

Editorial: The application of artificial intelligence in brain-computer interface and neural system rehabilitation

Fangzhou Xu[1]*, Dong Ming[2], Tzyy-Ping Jung[3], Peng Xu[4] and Minpeng Xu[2]

[1]Qilu University of Technology, Jinan, China, [2]Tianjin University, Tianjin, China, [3]University of California, San Diego, San Diego, CA, United States, [4]University of Electronic Science and Technology of China, Chengdu, China

KEYWORDS

brain-computer interface, pattern recognition, artificial intelligence, brain function, neural computing

Editorial on the Research Topic
The application of artificial intelligence in brain-computer interface and neural system rehabilitation

With the rapid growth of the global population, diseases related to motor dysfunction, such as stroke and spinal cord injury (SCI), are becoming an increasing challenge to public health. These diseases may lead to a series of functional declines such as cognitive impairment and emotional instability in patients, which seriously affects people's quality of life, and even endangers people's life and health, imposing heavy burdens on patients, families, and society. However, the pathogenesis of motor dysfunction-related diseases is complex, we currently lack effective and objective clinical diagnosis and intervention strategies. Artificial intelligence (AI) technology is developing rapidly, furthermore, it is attracting more attention from researchers and medical staff around the world in the brain-computer interface (BCI) and the clinical rehabilitation of motor dysfunction.

This Research Topic provides 19 papers on the application of AI techniques in the diagnosis and intervention of neurological diseases. The aim is to discover new algorithms, models, systems and applications that will facilitate the intersection of AI and neuroscience as well as promote the use of artificial intelligence in clinical medicine.

Innovations in classification recognition models and feature extraction methods improve performance on BCI systems. Li et al. proposes partial maximum correntropy regression (PMCR), a robust implementation of partial least squares regression (PLSR) using the maximum correntropy criterion. PMCR achieves better prediction and decoding performance compared to existing methods in noisy, inter-correlated, and high-dimensional decoding tasks. It minimizes neurophysiological pattern deterioration and improves electrocorticography decoding robustness for BCIs. Gao D-R. et al. introduces a novel unsupervised domain adaptation (UDA) approach. Effective data augmentation techniques are also explored. Experimental results demonstrate the superiority of the proposed method over state-of-the-art UDA methods in accuracy and MF1-Score. Gao D. et al. proposes a log-Mel spectrogram and convolution recurrent neural network (CRNN) model for fatigue detection using electroencephalogram (EEG) signals. Experimental results demonstrate

accurate and stable detection performance, outperforming existing methods. This approach has potential for enhancing driver safety and accident prevention. Du et al. presents a single-trial P300 classification algorithm based on multi-person data fusion convolutional neural networks (CNN) to improve the efficiency and accuracy of P300 EEG signal classification. The algorithm outperforms single-person CNN classification and achieves higher accuracy with smaller models and fewer parameters. Zhang et al. focuses on the recognition of motor imagery (MI) EEG signals for the right upper limb and proposes a multi-branch fusion convolutional neural network (MF-CNN) that combines raw EEG signals and two-dimensional time-frequency maps. Compared to single CNN branch algorithms, MF-CNN shows improved decoding accuracy and has potential applications in motor function rehabilitation training after stroke. Wang et al. proposes an approach that combines an improved lasso with relief-f for feature extraction and selection in EEG signals. The method effectively extracts wavelet packet entropy features and topological features of brain function network, leading to high classification accuracy. Experimental results on two public EEG datasets demonstrate the effectiveness of this approach, and the average classification accuracy aboves 90%. This technology has potential applications in MI-BCI medical, rehabilitation, and other fields. Chen et al. proposes a layered spindle detection algorithm that combines the Morlet wavelet, root mean square (RMS) method, and an improved k-means algorithm to improve the accuracy and speed of detecting spindles during sleep. The algorithm demonstrates better performance stability, achieving higher precision, recall, specificity, accuracy, and F1-score compared to other methods. It provides an effective tool for automatic spindle detection and improves detection efficiency. Li et al. proposes a coherence-based graph convolutional network (C-GCN) method for extracting features and functional connectivity information from EEG signals in a MI-BCI. The C-GCN method achieves reliable and stable classification performance, with a maximum accuracy of 96.85%. The analysis of EEG data from SCI patients and healthy subjects provides an effective theoretical basis for the rehabilitation treatment of SCI patients. Xu et al. proposes a method for analyzing EEG signals during sleep stages using phase-locked value (PLV) to construct a functional connection network and investigates brain interaction. The α frequency band (8–13Hz) achieves the best classification effect with an accuracy of 92.59%. The proposed algorithm enhances sleep staging performance and promotes the development of EEG sleep staging systems. Tang et al. presents a modified sequential backward floating search (SBFS) approach for channel selection in MI-BCIs. The proposed method improves the time complexity of SBFS by selecting symmetrical channel pairs based on the EEG channel map. Experimental results on four BCI datasets demonstrate that the SBFS method achieves higher classification accuracy compared to using all channels or conventional MI channels, outperforming state-of-the-art selection methods. Du et al. proposes a dual attentive fusion model (DAFM) for EEG-based BCI classification. Experimental results on four datasets demonstrate that the proposed method outperforms state-of-the-art approaches, highlighting the effectiveness of the DAFM in enhancing feature expression. Yang et al. explores effective EEG features for recognizing different valence emotions. First-order difference, second-order difference, high-frequency power, and high-frequency differential entropy features performe well in emotion recognition. These findings provide valuable guidance for EEG-based emotion recognition feature extraction and selection. Wen et al. proposes future directions for cross-task EEG signal analysis research, including increasing sample size, exploring feature extraction and classification simultaneously, subdividing tasks, and investigating cross-task regression models. Conducting research in these areas can advance cross-task EEG analysis to a higher level.

In addition, the Research Topic is innovative in the motion paradigm of BCIs. Liu et al. proposes movement intention encoding paradigm based on sequential finger movement, showing potential for expanding the instruction set of MI-BCIs. Offline and online experiments are conducted, demonstrating the feasibility of the proposed paradigm. Xiao et al. presents a novel v-BCI paradigm using weak and small stimuli to achieve nine instructions, and demonstrates higher information transfer rate and feasibility for widespread use. Huang et al. presents a wireless group-synchronized neural recording system for real-time multi-subject BCI analysis, achieves high signal correlation, low noise, and high information transfer rate. Bai et al. presents a hybrid BCI system combining P300 and steady-state visually evoked potential (SSVEP) for improved spelling accuracy and speed. The implemented BCI achieves 94.29% accuracy and 28.64 bit/min information transfer rate (ITR) in online tests. Offline calibration tests demonstrate an accuracy of 96.86%.

Moreover, ther e are studies to find physiological phenomena in patients EEG. Zhu et al. explores effects of repetitive transcranial magnetic stimulation (rTMS) on functional connectivity in chronic insomnia disorder (CID) patients. Findings indicate potential biomarkers for predicting clinical outcomes and suggest rTMS can improve symptoms and optimize treatment. Promising evidence for future clinical trials.

Finally, this Research Topic also includes a human-computer interaction discourse paper. Zhao et al. suggest combining cognitive psychology with AI to develop computers capable of recognizing emotions, understanding human feelings, achieving dialogue and empathy. Three examples highlight potential and importance of AI in understanding human mental states:face attraction, affective computing, and music emotion.

In summary, there are large varieties among the included studies in this Research Topic. This Research Topic emphasizes the importance of AI system combined with cognitive psychology to the development of AI, and introduces potential and application value of AI in understanding and identifying human psychological state. These research results provide useful reference and guidance for the development of BCI technology, sleep problem diagnosis and management, emotion recognition, motor function rehabilitation and other fields.

Author contributions

FX: Writing—review & editing. DM: Build the article framework. T-PJ: Polish and revise article. PX: Collect the accepted articles. MX: Summarize the accepted articles. All authors contributed to the article and approved the submitted version.

Funding

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

Conflict of interest

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

Publisher's note

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

frontiers | Frontiers in Neuroscience

TYPE Review
PUBLISHED 06 October 2022
DOI 10.3389/fnins.2022.1024316

OPEN ACCESS

EDITED BY
Fangzhou Xu,
Qilu University of Technology, China

REVIEWED BY
Bao Ge,
Shaanxi Normal University, China
Cai Wen,
McMaster University, Canada
Lei Min,
Beijing University of Posts and Telecommunications (BUPT), China

*CORRESPONDENCE
Jian Jia
jiajian@nwu.edu.cn

SPECIALTY SECTION
This article was submitted to Neuroprosthetics, a section of the journal Frontiers in Neuroscience

RECEIVED 21 August 2022
ACCEPTED 13 September 2022
PUBLISHED 06 October 2022

CITATION
Zhao J, Wu M, Zhou L, Wang X and Jia J (2022) Cognitive psychology-based artificial intelligence review.
Front. Neurosci. 16:1024316.
doi: 10.3389/fnins.2022.1024316

Cognitive psychology-based artificial intelligence review

Jian Zhao[1], Mengqing Wu[1], Liyun Zhou[1], Xuezhu Wang[1] and Jian Jia[2,3]*

[1]School of Information Science and Technology, Northwest University, Xi'an, China, [2]Medical Big Data Research Center, Northwest University, Xi'an, China, [3]School of Mathematics, Northwest University, Xi'an, China

Most of the current development of artificial intelligence is based on brain cognition, however, this replication of biology cannot simulate the subjective emotional and mental state changes of human beings. Due to the imperfections of existing artificial intelligence, this manuscript summarizes and clarifies that artificial intelligence system combined with cognitive psychology is the research direction of artificial intelligence. It aims to promote the development of artificial intelligence and give computers human advanced cognitive abilities, so that computers can recognize emotions, understand human feelings, and eventually achieve dialog and empathy with humans and other artificial intelligence. This paper emphasizes the development potential and importance of artificial intelligence to understand, possess and discriminate human mental states, and argues its application value with three typical application examples of human–computer interaction: face attraction, affective computing, and music emotion, which is conducive to the further and higher level of artificial intelligence research.

KEYWORDS

cognitive psychology, artificial intelligence, cognitive theory, behavioral science, human–computer interaction

Introduction

At present, in the development of artificial intelligence (AI), the scientific community is mostly based on brain cognition research (Nadji-Tehrani and Eslami, 2020), which is to reproduce the real physiological activities of our human brain through computer software. This replication of the biology of the human brain cannot well simulate the subjective psychological changes (Zador, 2019). For example, in terms of memory, human memory forgetting is non-active, and the more we want to forget the more memorable it becomes, while machine forgetting is an active deletion, which deviates from our psychological expectations. In the process of promoting the progress of artificial intelligence, psychology and its derived philosophy of mind play an important role directly or indirectly, can be considered as one of the fundamental

supporting theories of AI. For example: The current reinforcement learning theory in AI is inspired by the behaviorist theory in psychology, i.e., how an organism gradually develops expectations of stimuli in response to rewarding or punishing stimuli given by the environment, resulting in habitual behavior that yields maximum benefit. The current challenges faced by the artificial intelligence community – the emotional response of artificial intelligence machines, decision making in ambiguous states also need to rely on breakthroughs in the corresponding fields of psychology. Psychology and its derived philosophy of mind can be considered as one of the fundamental support theories for artificial intelligence (Miller, 2019). Cognitive psychology is mainly a psychological science that studies the advanced mental processes of human cognition, including the degree of thinking, deciding, reasoning, motivation and emotion. The most important feature that distinguishes humans from machines is that humans process external input by feeding back different attitudes toward things through our already internalized knowledge units about the external world, stimulating different subjective emotional orientations such as satisfaction, dissatisfaction, love, dislike and so on. These labeled emotional traits are generated by human cognitive psychology. By measuring subjective emotional changes, the internal knowledge structure is updated and the artificial intelligence machine is guided to re-learn, so that human attitudes, preferences and other subjective emotional experiences are given in AI (Kriegeskorte and Douglas, 2018; Pradhan et al., 2020).

Research on artificial intelligence is still in the developmental stage in terms of simulating human memory, attention, perception, knowledge representation, emotions, intentions, desires, and other aspects (Shi and Li, 2018). As the existing AI is not perfect, the AI system combined with cognitive psychology is the research direction of AI: Promote the development of artificial intelligence, endow the computer with the ability to simulate the advanced cognition of human beings, and carry out learning and thinking, so that computers can recognize emotions, understand human feelings, and finally achieve dialog and empathy with humans and other AI.

In terms of existing research results and methods, artificial intelligence combines new theories and methods such as psychology, brain science and computer science to conduct artificial intelligence machine simulation on people's psychological activities, reproduce people's psychology, integrate and promote each other, and jointly create more universal and autonomous artificial intelligence, which can better realize human–computer interaction (Yang et al., 2018) and further improve the level of social intelligence. At the same time, with the development of psychology, the scope of research and the choice of research objects are more extensive and universal, making artificial intelligence products have the conditions for rapid penetration into the field of psychology, resulting in research products such as facial expression-based emotion recognition system, public opinion analysis based on big data analysis technology, intelligent medical image grading or diagnosis, suicide early warning system and intelligent surveillance management system, which in turn promotes the development of psychology and shortens the research cycle of psychology (Branch, 2019).

The review of artificial intelligence based on cognitive psychology at this stage is not comprehensive enough. This manuscript does the following: (a) introduce the current situation and progress of artificial intelligence research on cognitive psychology in recent years; (b) analyze the experimental data on the application examples of cognitive psychology in artificial intelligence; (c) summarize and outlook the related development trend.

Research status

Research related to artificial intelligence in cognitive psychology is trending in recent years. In the mid-1980s, the term "Kansei Engineeirng" was introduced in the Japanese science and technology community (Ali et al., 2020). They interpret sensibility as human psychological characteristics, study people's perceptual needs with engineering methods, and then conduct in-depth research on people's perceptual information, and the scope of their research is the human psychological perceptual activities.

Professor Wang Zhiliang of University of Science and Technology Beijing proposed the concept of "artificial psychology" on this basis: The artificial psychological theory is to use the method of information science to realize the more comprehensive content of people's psychological activities. He broadened the range of psychological characteristics involved in "Kansei Engineeirng," including low-level psychological activities and high-level processes of psychological activities. It is the reflection of human brain on objective reality, which makes artificial psychology have a new meaning and broader content.

Minsky, one of the founders of artificial intelligence, proposed the theory of "society of mind" in his 1985 monograph "The Society of Mind" (Auxier, 2006), which attempts to combine the approaches of developmental psychology, dynamic psychology and cognitive psychology with the ideas of artificial intelligence and computational theory. Since then, the research on endowing the computer with emotional ability and enabling the computer to understand and express emotions has set off an upsurge in the computer field.

In 1978, deepmind team put forward the theory of mind (Rabinowitz et al., 2018). In a broad sense, it refers to the ability of human beings to understand the psychological state of themselves and others, including expectations, beliefs and intentions, and to predict and explain other people's behaviors

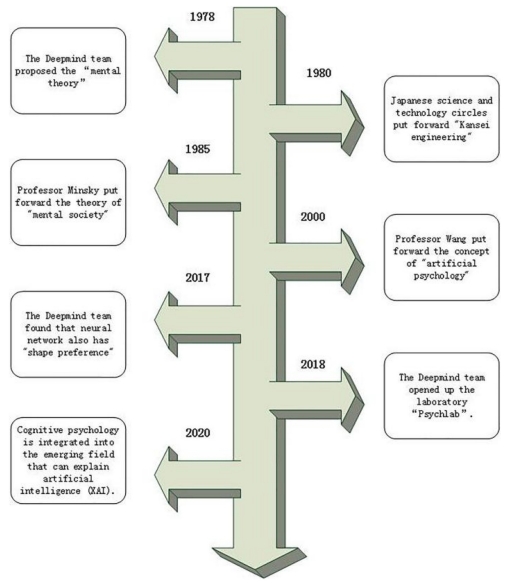


FIGURE 1
The evolution of artificial intelligence in cognitive psychology.

based on this. In 2017, in the case study of deepmind team, the research team selected "shape preference" as the entry point for detecting neural networks. It found that, like human beings, the network's perception of shape exceeded its preference for color and material, which proved that neural networks also have "shape preference" (Ritter et al., 2017). In 2018, the Deepmind team open sourced the simulation psychology laboratory Psychlab, which uses knowledge in cognitive psychology and other fields to study the behavior of artificial agents in controlled environments, thereby simulating human behavior (Leibo et al., 2018).

In 2020, Taylor incorporated cognitive psychology into the emerging field of explainable artificial intelligence (XAI) with the aim of improving the interpretability, fairness, and transparency of machine learning. **Figure 1** shows the evolution of AI in cognitive psychology (Taylor and Taylor, 2021).

Example of cognitive psychological artificial intelligence applications

Cognitive psychology has been very instructive for the development of AI, and current AI design makes extensive reference to human cognitive models. The process of human mental activity is simulated in various aspects such as attention, encoding, and memory. Cognitive psychological artificial intelligence has been researched in many fields. In this manuscript, we study the basic contents and latest progress of psychology and brain science, and systematically analyze and summarize three typical application scenarios: face attraction, affective computing, and music emotion. These examples guide the learning of AI through the higher mental processes of human cognition, including subjective mental orientations such as thinking and emotion. Artificial intelligence is trained to recognize emotions, understand human feelings, and replicate the human psyche, which in turn accelerates research in cognitive psychology.

Face attraction

Different aesthetic judgments of human faces are one of the most common manifestations of human visual psychology, which is an important source of social emotion generation and plays a role in human social interaction and communication (Han et al., 2020). In daily life, most people think that beauty is a subjective feeling, however, scientists have broken the long-held belief that beauty lacks objectivity and found a high degree of consistency in human perception of facial beauty across race, age, gender, social class, and cultural background. This observation also suggests that face attractiveness reflects to some extent general human psychological commonalities.

SCUT-FBP5500, a database for face attractiveness prediction, was collected and released by the Human–Computer Interaction Laboratory of South China University of Technology. The dataset has 5,500 face frontal photos with different attributes (male/female, age and so on) and different feature labels including face feature point coordinates, face value score (1∼5), face value score distribution and so on. These mental preference features were experimentally used as training data to form mental state embeddings. Then different computer models (AlexNet, ResNet-18, ResNeXt-50) were used for classification, regression and ranking to form a deep learning-based face attractiveness template (Huang, 2017). Evaluate the benchmark according to various measurement indicators, including Pearson correlation coefficient (PC), maximum absolute error (MAE) and root mean square error (RMSE) evaluation model. We used the five-fold method to analyze the performance of the face attractiveness templates under different computer models, and found that the Pearson correlation coefficient was above 0.85, the maximum absolute error was around 0.25, and the root mean square error was between 0.3 and 0.4 (Liang et al., 2018).

Elham Vahdati proposes and evaluates a face facial attractiveness prediction method using facial parts as well as a multi-task learning scheme. First, face attractiveness prediction is performed using a deep convolutional neural network (CNN) pre-trained on a massive face dataset to automatically learn advanced face representations. Next, the deep model is extended to other facial attribute recognition tasks using a multi-task learning scheme to learn the best shared features for three related tasks (such as facial beauty assessment, gender

recognition, and race recognition). To further improve the accuracy of the attractiveness computation, specific regions of the face image (such as left eye, nose, and mouth) as well as the entire face are fed into a multi-stream CNN (such as three dual-stream networks). Each dual-stream network uses partial features of the face and the full face as input. Extensive experiments were conducted on the SCUT-FBP5500 benchmark dataset, with a significant improvement in accuracy (Vahdati and Suen, 2021).

Irina Lebedeva, Fangli Ying learned a large number of aesthetic preferences shared by many people during the meta-training process. The model is then used on new individuals with a small sample of rated images in the meta-testing phase. These experiments were conducted on a facial beauty dataset that included faces of different races, genders, and age groups and were scored by hundreds of volunteers with different social and cultural backgrounds. The results show that the proposed method is effective in learning individual beauty preferences from a limited number of annotated images and outperforms existing techniques for predicting facial beauty in terms of quantitative comparisons (Lebedeva et al., 2022).

We summarize the theoretical concepts of artificial intelligence based on cognitive psychology, and do relevant research on this basis. Since the database of face attractiveness needs to be characterized by large samples, diversity and universality, in 2016, we built a Chinese face database containing different ethnicities of different genders. In 2017, considering that the contour structure, geometric features and texture features of faces change with age, in order to study the impact of different face features on the evaluation of face attractiveness under different age groups, we built a middle-aged and elderly face database. In 2018, we used migration learning to migrate the face feature point templates of face recognition to the construction of face attractiveness face templates, and constructed a geometric feature-based face attractiveness evaluation model. In 2019, we established a face database of Chinese males in different eras, and studied the aesthetic characteristics and trends of Chinese males from the perspective of era development. An 81-point face feature point template for face attractiveness analysis was also proposed through feature vector analysis of face image quantification and light model. In 2020, a comprehensive facial attractiveness evaluation system was proposed considering the combined effects of face structure features, facial structure features, and skin texture features on face attractiveness scores, and the experimental results are shown in **Table 1**, when these three features are integrated with each other, the Pearson correlation coefficient reached the highest value of 0.806 (Zhao et al., 2019a,b,c; Zhao et al., 2020).

Through years of research at the intersection of artificial intelligence + face attractiveness, it is shown that although it may be difficult to establish a clear, interpretable and accepted set of rules to define face attractiveness. However, it is possible to explore the relationship between ordinary faces and attractive faces, and the qualitative study of face aesthetic preferences can be described quantitatively by artificial intelligence. The results highly fit contemporary aesthetic standards, demonstrating that it is feasible for computers to simulate advanced human cognitive abilities to recognize emotions and understand human feelings, and that the development of artificial intelligence based on cognitive psychology has potential and significance.

Affective computing

Emotion is a psychological state of positive or negative attitude toward external things and objective reality, and can be defined as a group of psychological phenomena expressed in the form of emotions, feelings or passions. Emotions not only refer to human emotions, but also refer to all human sensory, physical, psychological and spiritual feelings. Damasio found in his research that due to the defect of the channel between the cerebral cortex (Cortex: control of logical reasoning) and the limbic system (Limbic System: control of emotion), his "patients" despite having normal or even supernormal rational thinking and logical reasoning. However, their decision-making ability has encountered serious obstacles (Bechara et al., 2000), proving that human intelligence is not only manifested in normal rational thinking and logical reasoning abilities, but also in rich emotional abilities.

More than 40 years ago, Nobel Laureate Herbert Simon emphasized in cognitive psychology that problem solving should incorporate the influence of emotions (Simon, 1987). As one of the founders of artificial intelligence, Professor Marvin Minsky of the Massachusetts Institute of technology of the United States first proposed the ability to make computers have emotion. In his monograph the society of mind, he emphasized that emotion is an indispensable and important ability for machines to achieve intelligence. The concept of affective computing was first introduced by Picard (1995), when she stated that "affective computing is computing that can measure and analyze and influence emotions in response to human outward expressions" (Picard, 2003). This opened up a new field of computer science, with the idea that computers should have emotions and be able to recognize and express them as humans do, thus making human–computer interaction more natural.

As an important means of interpersonal communication, emotion conveys the information of emotional state and explains complex psychological activities and behavioral motives through physiological indicators such as human language text, intonation volume change, facial expression, action posture and brain wave.

In, Ekman (1972) an American professor of psychology, proposed a method for the expression of facial emotions (Facial Motor Coding System FACS) (Buhari et al., 2020).

TABLE 1 Performance of face attractiveness prediction with different features.

Predictive performance	F	S	LBP	F × S	F × LBP	S × LBP	F × S × LBP
LR	0.502	0.616	0.658	0.683	0.654	0.637	0.722
KNN	0.619	0.672	0.694	0.753	0.771	0.782	0.794
SVM-LIN	0.649	0.738	0.712	0.768	0.732	0.724	0.797
SVM-RBF	0.702	0.713	0.741	0.763	0.754	0.781	0.806

TABLE 2 Emotion formula.

Expression	Formula of AU
Happiness	AU6 + AU12
Sadness	AU1 + AU4 + AU15
Surprise	AU1 + AU2 + AU5 + AU26
Fear	AU1 + AU2 + AU4 + AU5 + AU7 + AU20 + AU26
Anger	AU4 + AU5 + AU7 + AU23
Disgust	AU9 + AU15 + AU16
Contempt	AU12 + AU14

By the combination of different coding and motor units, complex expression changes can be formed on the face. Facial motion coding system FACS can analyze emotions using deep region and multi-label learning (DRML) architecture, using feedforward functions to induce important facial regions, and able to learn weights to capture structural information of the face. The resulting network is end-to-end trainable and converges faster than alternative models with better learning of AU relationships (Zhao et al., 2016). The corresponding emotion computation formula can be derived based on the facial motion encoding, as **Table 2** shown.

In the process of human information interaction, speech is the most common way for people to communicate. As the most basic audiovisual signal, speech cannot only identify different vocalists, but also effectively distinguish different emotional states. International research on emotional speech focuses on the analysis of acoustic features of emotions, such as rhythm, sound source, resonance peaks and spectrum and so on (Albanie et al., 2018). In recent years, deep learning has been widely studied and has many applications in speech emotion computation. Dongdong Li proposed a bidirectional long short-term memory network with directed self-attention (BLSTM-DSA). Long Short Term Memory (LSTM) neural networks can learn long-term dependencies from learned local features. In addition, Bi-directional Long Short-Term Memory(Bi-LSTM) can make the structure more robust through the direction mechanism, and the direction analysis can better identify the hidden emotions in sentences. Also, the autocorrelation of speech frames can be used to deal with the problem of missing information, thus introducing a self-attention mechanism in Speech Emotion Recognition (SER). When evaluated on the Interactive Emotional Binary Motion Capture (IEMOCAP) database and the Berlin Emotional Speech Database (EMO-DB), BLSTM-DSA achieves a recognition rate of over 70% for each algorithm on the speech emotion recognition task (Li et al., 2021).

Human posture often carries emotional information during interaction. Researchers have combined human posture with artificial intelligence to quantitatively assess the external representation of a person's mental state in the face of different situations through a series of movement and body information capture devices. For example, the intelligent seat is applied to the driver's seat of the vehicle to dynamically monitor the emotional state of the driver and give timely warnings. Some scientists in Italy also conduct automatic emotional analysis on office staff through a series of posture analysis to design a more comfortable office environment.

Electroencephalographic(EEG) is a graph obtained by amplifying and recording the spontaneous biological potential of the brain from the scalp through precise electronic instruments. It has been widely used in the field of emotion recognition. The DEAP dataset used to study human emotional states (Luo et al., 2020), recording EEG and peripheral physiological signals from 32 participants watching 40 one-minute long music video clips. Participants rated each video according to arousal, potency, like/dislike, dominance, and familiarity. Correlations between EEG signal frequencies and participants' ratings were investigated by emotional label retrieval, and decision fusion was performed on classification results from different modalities. The experiments obtained an average recognition rate of up to 84.2% and up to 98% by identifying a single emotional state, while for two, three and four emotions, the average recognition rate was up to 90.2, 84.2, and 80.9%, respectively. **Table 3** shows the validated classification accuracy of the DEAP dataset based on different recognition models (Khateeb et al., 2021).

Our research group has also carried out relevant research on multimodal affective computing, and has a patent for automatic diagnosis of depression based on speech and facial expression: By combining facial gesture features, we propose a new double dictionary idea with gesture robustness. In 2016, feature extraction and evaluation of depressed speech were performed, and in the following year, we proposed to use the change of expression of depressed patients as one of the evaluation indicators to determine whether they suffer from depression as well. **Figures 2** and **3** shows the data.

TABLE 3 Classification accuracy of deap dataset based on different recognition models.

Stimulus	Classifier	Emotions	Subjects	Accuracy
Video	GELM	4	32	69.67
Audio	MLP	4	30	78.11
Video	Nearest neighbour	4	32	73.62
Video	Domain-adaptation	5	14	39.05
Video	SVM	Valence-dominance	10	63.04
Video	K-NN	2	30	69.50

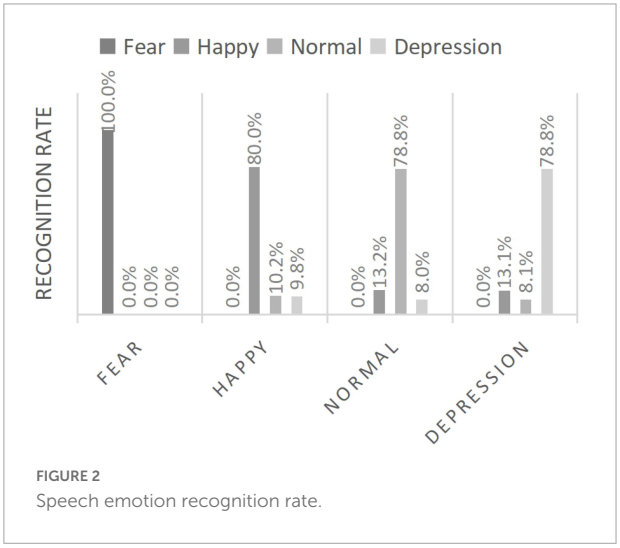


FIGURE 2
Speech emotion recognition rate.

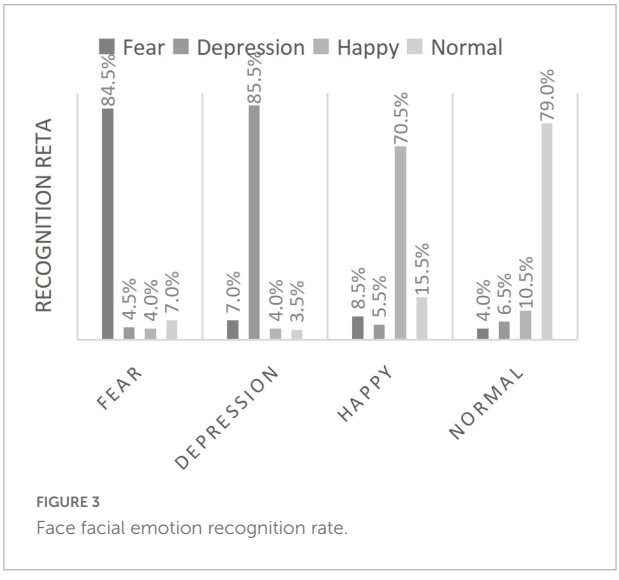


FIGURE 3
Face facial emotion recognition rate.

In 2018, a new automatic depression assistant discrimination algorithm integrating speech and facial expression was proposed. Firstly, the signal enhancement was performed for depressed speech; the fundamental frequency and the first three resonance peaks features were extracted by the inverse spectral method, and the energy, short-time average amplitude and Mel-Frequency Ceptral Coefficients(MFCC) features were extracted; the speech recognition model and the facial expression recognition model were established to assist in judging whether a person has depression; finally, the Adaboost algorithm based on back propagation(BP) neural network was proposed and validated in a practical situation for an automatic depression-assisted detection system. As Table 4 shown, the recognition rate of the depression detection algorithm based on fused speech and facial emotion reached 81.14%. The development of artificial intelligence provides a more objective judgment basis for the diagnosis of depression in psychological medical health, which has cutting-edge and application value (Zhao et al., 2019d).

TABLE 4 The integration of voice and facial expression recognition rate.

	Speech recognition results (%)	Facial expression recognition results (%)	After fusion recognition results (%)
Before the speech signal enhancement	62.4	85.5	75.75
Enhanced speech signal	78.8	85.5	82.29

Affective computing is a combination of computational science with physiology science, psychological science, cognitive science and other disciplines. Based on the common cognition and knowledge structure of human on different emotional expressions, it studies the emotions in the process of human-human interaction and human–computer interaction, and guides the design of artificial intelligence with emotion recognition and feedback functions, understands human emotional intentions and makes appropriate responses to achieve human–computer emotional interaction.

Music emotion

Extensive research on musical emotions suggests that music can trigger emotional activity in listeners. Scientists believe that when a person is in a beautiful and pleasant musical environment, the body secretes an active substance that is beneficial to health and helps eliminate psychological factors that cause tension, anxiety, depression and other adverse psychological states (Rahman et al., 2021). People's preference for different kinds of music is not without rules, after psychological cognition and data test, there is a precise music signal α value can measure the ear-pleasant degree. The closer the music signal α is to the value 1, the better it sounds. The value of α also can be obtained by artificial intelligence (Banerjee et al., 2016). This shows that people's psychological state toward

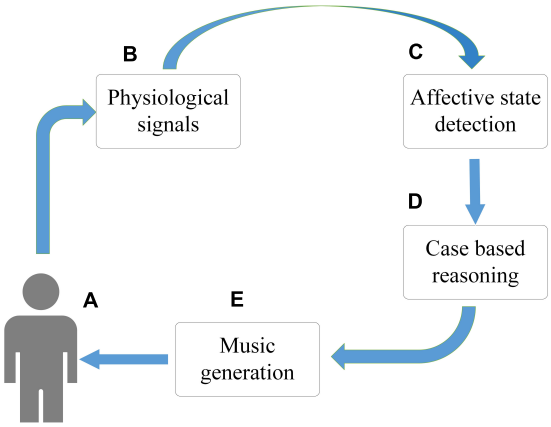


FIGURE 4
The proposed affective brain-computer music interface (aBCMI). The system consists of five key elements: **(A)**. The user of the system **(B)**. The user's physiological signal acquisition module (including the electroencephalogram (EEG), electrocardiogram (ECG) and respiration rate) **(C)**. An emotional state detection system for identifying a current emotional state that a user is experiencing **(D)**. A case-based reasoning system that determines how a user moves from his current emotional state to a new target emotional state **(E)**. The music generator is used to play music for the user. The case-based reasoning system identifies the most appropriate emotional trajectory and moves them to the target emotional state.

music can be judged by machines, and further research can be based on this law to simulate good-sounding music in line with public aesthetics and realize the interaction between emotions and machines.

As Figure 4, a team of researchers from the University of Reading and the University of Plymouth in the UK developed and evaluated an affective brain-computer music interface (aBCMI) for detecting a user's current emotional state and attempting to modulate it by playing music generated by a music composition system based on specific emotional goals.

The affective state detection method achieved statistically significant online single-trial classification accuracy in classifying user potency in seven-eighths of participants and in classifying user arousal in three-eighths of participants. The mean accuracy for affective state detection was 53.96% (chemotaxis) and 53.80% (arousal) (Daly et al., 2016). The experimental data also demonstrate that the aBCMI system is able to detect the emotional states of most of the participants and generate music based on their emotional states to achieve "happy" and "calm" mental states. By visualizing abstract mental states, extracting features from changes in emotional states, and quantifying different emotions in different musical environments, the aBCMI system can effectively characterize and provide feedback to regulate current emotional states, realizing the combination of psychology and artificial intelligence.

Musical emotion regulation aims to record physiological indicators from users with a signal acquisition component in order to capture the cognitive and physiological processes associated with their current affective state. Features are extracted from the physiological signals that most likely correspond to changes in the user's affective state. Then the case-based reasoning system is used to determine the best method to transfer them to the target emotional state, so as to move the user to the target emotional state.

Dapeng Li and Xiaoguang Liu have also combined incremental music teaching methods to assist therapy. The combination of contextual teaching and artificial intelligence attention theory makes the assisted treatment system more targeted. The design of treatment content more fully takes into account the patient's actual situation. When designing the music teaching-assisted treatment context, the physician will fully consider various factors of the patient, from the perspective of mobilizing the patient's interest in the music learning work, to achieve the full activity of brain neurons and more fully access the pathological information around the lesion to promote autoimmunity and subsequent treatment (Li and Liu, 2022).

The evocation of musical emotions is based on functional connections between sensory, emotional and cognitive areas of the brain, including subcortical reward networks common to humans and other animals, such as the nucleus accumbens, amygdala and dopaminergic systems, as well as the evolutionary end of the cerebral cortex with complex cognitive functions. Musical emotions regulate the activity of almost all limbic and paralimbic structures of the brain. Music can induce different emotions, and we can also use music emotions to guide the development of artificial intelligence. Further research is expected in such fields as music generation, education, medical treatment and so on.

Summary and outlook

Through systematic analysis and application examples, this manuscript points out that the artificial intelligence system combined with cognitive psychology is the development direction of artificial intelligence: to promote the development of artificial intelligence, to give computers the ability to simulate human's advanced cognition, and to learn and think, so that computers can recognize emotions and understand human feelings, and finally realize dialog and empathy with human beings and other artificial intelligence. Artificial intelligence with human psychological cognition cannot only simulate the rational thinking of "brain," but also reproduce the perceptual thinking of "heart," and can realize the emotional interaction between people and machines, machines and machines, similar to human communication.

Nowadays, the theory of artificial intelligence based on cognitive psychology also has imperfections: due to

the differences in race, region and growth environment, the evaluation criteria for each subject are not completely consistent, and the random sampling difference is even greater Moreover, mental activities are generally ambiguous and chaotic.

The future interdisciplinary combination of AI and psychology will focus on the following aspects: big data medical, human–computer interaction, brain-computer interface, general artificial intelligence and so on. Through the combination of cognitive science in psychology and AI, breakthroughs in many aspects will be achieved based on multimodal data and extraction of high-dimensional data. The two accomplish each other, complementing each other and developing together.

This manuscript provides a research direction for the development of artificial intelligence to simulate machines with human emotions and to realize human–computer interaction. It has the characteristics of cutting-edge science, which is not only of great theoretical significance, but also has good development potential and application prospects. It is hoped that it can provide research basis for follow-up researchers.

Author contributions

JZ formulated the research manuscript idea, provided substantial edits to the manuscript and final draft, and aided in the interpretation of the manuscript. MW wrote the main body of the manuscript, participated in revisions, and submitted the final manuscript. LZ contributed to the formulation of the research manuscript idea, provided substantial edits to the manuscript and the final draft, and aided in the interpretation of the manuscript. XW and JJ participated in the conception of the idea and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by National Natural Science Foundation of China: 12071369 and Key Research and Development Program of Shaanxi (No. 2019ZDLSF02-09-02).

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

References

Albanie, S., Nagrani, A., Vedaldi, A., and Zisserman, A. (2018). "Emotion recognition in speech using cross-modal transfer in the wild," in *Proceedings of the 26th ACM international conference on multimedia* (New York, NY: Association for Computing Machinery), 292–301. doi: 10.1145/3240508.3240578

Ali, S., Wang, G., and Riaz, S. (2020). Aspect based sentiment analysis of ridesharing platform reviews for kansei engineering. *IEEE Access* 8, 173186–173196. doi: 10.1109/ACCESS.2020.3025823

Auxier, R. E. (2006). *The pluralist: An editorial statement. The pluralist.* Champaign, IL: University of Illinois Press, v–viii.

Banerjee, A., Sanyal, S., Patranabis, A., Banerjee, K., Guhathakurta, T., Sengupta, R., et al. (2016). Study on brain dynamics by non linear analysis of music induced EEG signals. *Phys. A Stat. Mech. Appl.* 444, 110–120. doi: 10.1016/j.physa.2015.10.030

Bechara, A., Damasio, H., and Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cereb. Cortex* 10, 295–307. doi: 10.1093/cercor/10.3.295

Branch, B. (2019). Artificial intelligence applications and psychology: An overview. *Neuropsychopharmacol. Hung.* 21, 119–126.

Buhari, A. M., Ooi, C. P., Baskaran, V. M., Phan, R. C., Wong, K., and Tan, W. H. (2020). Facs-based graph features for real-time micro-expression recognition. *J. Imaging* 6:130. doi: 10.3390/jimaging6120130

Daly, I., Williams, D., Kirke, A., Weaver, J., Malik, A., Hwang, F., et al. (2016). Affective brain–computer music interfacing. *J. Neural Eng.* 13:046022.

Han, S., Liu, S., Li, Y., Li, W., Wang, X., Gan, Y., et al. (2020). Why do you attract me but not others? Retrieval of person knowledge and its generalization bring diverse judgments of facial attractiveness. *Soc. Neurosci.* 15, 505–515. doi: 10.1080/17470919.2020.1787223

Huang, C. (2017). "Combining convolutional neural networks for emotion recognition," in *Proceedings of the 2017 IEEE MIT undergraduate research technology conference (URTC)* (Cambridge, MA: IEEE), 1–4. doi: 10.1109/URTC.2017.8284175

Khateeb, M., Anwar, S. M., and Alnowami, M. (2021). Multi-domain feature fusion for emotion classification using DEAP dataset. *IEEE Access* 9, 12134–12142. doi: 10.1109/ACCESS.2021.3051281

Kriegeskorte, N., and Douglas, P. K. (2018). Cognitive computational neuroscience. *Nat. Neurosci.* 21, 1148–1160. doi: 10.1038/s41593-018-0210-5

Lebedeva, I., Ying, F., and Guo, Y. (2022). Personalized facial beauty assessment: A meta-learning approach. *Vis. Comput.* 1–13. doi: 10.1007/s00371-021-02387-w

Leibo, J. Z., d'Autume, C. D. M., Zoran, D., Amos, D., Beattie, C., Anderson, K., et al. (2018). Psychlab: A psychology laboratory for deep reinforcement learning agents. *arXiv* [Preprint]. arXiv:1801.08116,

Li, D., and Liu, X. (2022). Design of an incremental music Teaching and assisted therapy system based on artificial intelligence attention mechanism. *Occup. Ther. Int.* 2022:7117986. doi: 10.1155/2022/7117986

Li, D., Liu, J., Yang, Z., Sun, L., and Wang, Z. (2021). Speech emotion recognition using recurrent neural networks with directional self-attention. *Expert Syst. Appl.* 173:114683. doi: 10.1016/j.eswa.2021.114683

Liang, L., Lin, L., Jin, L., Xie, D., and Li, M. (2018). "SCUT-FBP5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction," in *Proceedings of the 2018 24th international conference on pattern recognition (ICPR)* (Beijing: IEEE), 1598–1603. doi: 10.1109/ICPR.2018.8546038

Luo, Y., Fu, Q., Xie, J., Qin, Y., Wu, G., Liu, J., et al. (2020). EEG-based emotion classification using spiking neural networks. *IEEE Access* 8, 46007–46016. doi: 10.1109/ACCESS.2020.2978163

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.* 267, 1–38. doi: 10.1016/j.artint.2018.07.007

Nadji-Tehrani, M., and Eslami, A. (2020). A brain-inspired framework for evolutionary artificial general intelligence. *IEEE Trans. Neural Netw. Learn. Syst.* 31, 5257–5271. doi: 10.1109/TNNLS.2020.2965567

Picard, R. W. (2003). Affective computing: Challenges. *Int. J. Hum. Comput. Stud.* 59, 55–64. doi: 10.1016/S1071-5819(03)00052-1

Pradhan, N., Singh, A. S., and Singh, A. (2020). Cognitive computing: Architecture, technologies and intelligent applications. *Mach. Learn. Cogn. Comput. Mob. Commun. Wirel. Netw.* 3, 25–50. doi: 10.1002/9781119640554.ch2

Rabinowitz, N., Perbet, F., Song, F., Zhang, C., Eslami, S. A., and Botvinick, M. (2018). "Machine theory of mind," in *Proceedings of the international conference on machine learning* (Orlando, FL: PMLR), 4218–4227.

Rahman, J. S., Gedeon, T., Caldwell, S., Jones, R., and Jin, Z. (2021). Towards effective music therapy for mental health care using machine learning tools: Human affective reasoning and music genres. *J. Artif. Intell. Soft Comput. Res.* 11, 5–20. doi: 10.2478/jaiscr-2021-0001

Ritter, S., Barrett, D. G., Santoro, A., and Botvinick, M. M. (2017). "Cognitive psychology for deep neural networks: A shape bias case study," in *Proceedings of the international conference on machine learning* (Cancun: PMLR), 2940–2949.

Shi, Y., and Li, C. (2018). "Exploration of computer emotion decision based on artificial intelligence," in *Proceedings of the 2018 international conference on virtual reality and intelligent systems (ICVRIS)* (Hunan: IEEE), 293–295. doi: 10.1109/ICVRIS.2018.00078

Simon, H. A. (1987). Making management decisions: The role of intuition and emotion. *Acad. Manag. Perspect.* 1, 57–64. doi: 10.5465/ame.1987.4275905

Taylor, J. E. T., and Taylor, G. W. (2021). Artificial cognition: How experimental psychology can help generate explainable artificial intelligence. *Psychon. Bull. Rev.* 28, 454–475. doi: 10.3758/s13423-020-01825-5

Vahdati, E., and Suen, C. Y. (2021). Facial beauty prediction from facial parts using multi-task and multi-stream convolutional neural networks. *Int. J. Pattern Recognit. Artif. Intell.* 35:2160002. doi: 10.1142/S0218001421600028

Yang, G. Z., Dario, P., and Kragic, D. (2018). Social robotics—trust, learning, and social interaction. *Sci. Rob.* 3:eaau8839. doi: 10.1126/scirobotics.aau8839

Zador, A. M. (2019). A critique of pure learning and what artificial neural networks can learn from animal brains. *Nat. Commun.* 10, 1–7. doi: 10.1038/s41467-019-11786-6

Zhao, J., Cao, M., Xie, X., Zhang, M., and Wang, L. (2019a). Data-driven facial attractiveness of Chinese male with epoch characteristics. *IEEE Access* 7, 10956–10966. doi: 10.1109/ACCESS.2019.2892137

Zhao, J., Deng, F., Jia, J., Wu, C., Li, H., Shi, Y., et al. (2019b). A new face feature point matrix based on geometric features and illumination models for facial attraction analysis. *Discrete Contin. Dyn. Syst. S* 12, 1065–1072. doi: 10.3934/dcdss.2019073

Zhao, J., Su, W., Jia, J., Zhang, C., and Lu, T. (2019c). Research on depression detection algorithm combine acoustic rhythm with sparse face recognition. *Cluster Comput.* 22, 7873–7884. doi: 10.1007/s10586-017-1469-0

Zhao, J., Zhang, M., He, C., and Zuo, K. (2019d). Data-driven research on the matching degree of eyes, eyebrows and face shapes. *Front. Psychol.* 10:1466. doi: 10.3389/fpsyg.2019.0146

Zhao, J., Zhang, M., He, C., Xie, X., and Li, J. (2020). A novel facial attractiveness evaluation system based on face shape, facial structure features and skin. *Cogn. Neurodynamics* 14, 643–656. doi: 10.1007/s11571-020-09591-9

Zhao, K., Chu, W. S., and Zhang, H. (2016). "Deep region and multi-label learning for facial action unit detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Las Vegas, NV: IEEE), 3391–3399. doi: 10.1109/CVPR.2015.7298833

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 17 October 2022
DOI 10.3389/fnins.2022.1010951

OPEN ACCESS

EDITED BY
Peng Xu,
University of Electronic Science and Technology of China, China

REVIEWED BY
Ming Meng,
Hangzhou Dianzi University, China
Liming Zhao,
Shanghai Jiao Tong University, China
Yanrong Hao,
Lanzhou University, China

*CORRESPONDENCE
Bin Yan
ybspace@hotmail.com

SPECIALTY SECTION
This article was submitted to Neuroprosthetics, a section of the journal Frontiers in Neuroscience

RECEIVED 03 August 2022
ACCEPTED 05 September 2022
PUBLISHED 17 October 2022

CITATION
Yang K, Tong L, Zeng Y, Lu R, Zhang R, Gao Y and Yan B (2022) Exploration of effective electroencephalography features for the recognition of different valence emotions. *Front. Neurosci.* 16:1010951. doi: 10.3389/fnins.2022.1010951

Exploration of effective electroencephalography features for the recognition of different valence emotions

Kai Yang, Li Tong, Ying Zeng, Runnan Lu, Rongkai Zhang, Yuanlong Gao and Bin Yan*

Henan Key Laboratory of Imaging and Intelligent Processing, People's Liberation Army (PLA), Strategy Support Force Information Engineering University, Zhengzhou, China

Recent studies have shown that the recognition and monitoring of different valence emotions can effectively avoid the occurrence of human errors due to the decline in cognitive ability. The quality of features directly affects emotion recognition results, so this manuscript explores the effective electroencephalography (EEG) features for the recognition of different valence emotions. First, 110 EEG features were extracted from the time domain, frequency domain, time-frequency domain, spatial domain, and brain network, including all the current mainly used features. Then, the classification performance, computing time, and important electrodes of each feature were systematically compared and analyzed on the self-built dataset involving 40 subjects and the public dataset DEAP. The experimental results show that the first-order difference, second-order difference, high-frequency power, and high-frequency differential entropy features perform better in the recognition of different valence emotions. Also, the time-domain features, especially the first-order difference features and second-order difference features, have less computing time, so they are suitable for real-time emotion recognition applications. Besides, the features extracted from the frontal, temporal, and occipital lobes are more effective than others for the recognition of different valence emotions. Especially, when the number of electrodes is reduced by 3/4, the classification accuracy of using features from 16 electrodes located in these brain regions is 91.8%, which is only about 2% lower than that of using all electrodes. The study results can provide an important reference for feature extraction and selection in emotion recognition based on EEG.

KEYWORDS

emotion recognition, emotion valence, EEG, feature extraction, feature selection

Introduction

Emotions play an important role in our daily life because they can affect people's work efficiency, decision-making, memory et al. Compared with neutral emotions, positive and negative emotions tend to decline our cognitive ability (Qian et al., 2015; Wang and Liu, 2020). If people's cognitive abilities for special jobs are affected by their emotions, there may be serious consequences. Therefore, if the emotions that lead to cognitive decline can be effectively identified and timely warned, most of the adverse consequences caused by cognitive decline can be avoided.

Almost everyone experienced the change from a positive emotional state to a negative emotional state at some point. According to the continuous emotion model proposed by Lang et al. (2001), the change of positive emotion and negative emotion indicates the variations of emotion valence. Previous studies have pointed out that different valence emotions have different effects on cognition performance. For example, although an extremely negative emotional state may reduce cognitive abilities significantly, a moderately negative emotional state can enhance alertness and responsiveness (Chen et al., 2013). Besides, people's memory and judgment abilities tend to be weakened in negative valence emotional states, which makes it easier to make irrational decisions. Compared with negative emotional states, most positive emotional states are usually harmless and may even improve people's cognitive abilities. Blair et al. (2007) investigated the interaction between positive, neutral, and negative valence emotions and goal-directed processing tasks. They found that positive and negative valence stimuli have a greater impact on goal-directed tasks than neutral valence stimuli. Based on event-related potentials (ERP) and functional magnetic resonance imaging (fMRI) research, Li et al. (2006) found that negative valence emotion has a greater impact on spatial working memory than on verbal working memory. Yuan et al. (2007) explored people's sensitivity to valence differences in emotional stimuli by using different valence pictures with no significant difference in arousal as stimulus materials, and they found that people are more sensitive to negative valence pictures. Meng et al. explored the influence of attention on human sensitivity to valence differences in emotional stimuli. They found that the ERP amplitude of extremely negative valence pictures was greater than that of moderately negative and neutral pictures within 150–250, 250–350, and 350–450 ms after the pictures were presented (Meng et al., 2009). As described above, people have different sensitivities to emotions with different valence, and different valence emotions have different effects on people's cognitive performance. However, it is not enough to claim that emotions with different valence have an impact on people's cognition. When people are in these emotional states, if early warning can be given to ask them to stop working, serious consequences may be avoided. The premise of early warning is to accurately classify different valence emotions.

Electroencephalography has the advantages of non-invasiveness, high time resolution, and good portability, and it has been widely used in emotion recognition research (Hu et al., 2019). Feature extraction is a key step in emotion recognition based on EEG. The quality of features will directly affect the accuracy of emotion recognition (Wang and Wang, 2021). EEG features in emotion recognition can be mainly divided into time domain (statistical features), frequency domain, time-frequency domain, spatial asymmetry, and brain network features (Li et al., 2016; Gonuguntla et al., 2020). Time-domain features mainly include first-order difference, second-order difference, fractal dimension, sample entropy, approximate entropy, and standard deviation (Lan et al., 2016). Frequency-domain features mainly include power spectral density and power (An et al., 2021). Time-frequency domain features are mainly features extracted based on the discrete Fourier transform or Hilbert Huang transform (Khare and Bajaj, 2021). Differential entropy is the most used time-frequency feature, which has achieved the highest classification accuracy in multiple studies (Nie et al., 2011; Zheng et al., 2019). Spatial asymmetry feature refers to the difference or ratio of features from left and right hemisphere electrodes (Zheng and Lu, 2015). Additionally, the brain network feature including the connection between electrodes is a new feature in recent years, which is increasingly used in emotion recognition and has achieved good results. Li et al. (2019) proposed to fuse local features extracted from a single electrode with brain network features containing global information, which improved the performance of emotion classification. Our previous study also compared the performance of EEG network features of different frequency bands, and the results showed that the high gamma band brain network features were more closely related to emotion (Yang et al., 2020). Wu et al. (2022) pointed out that the brain network features representing the relationship between different electrodes have better classification performance than the differential entropy extracted from a single electrode. Though there are currently many types of features used in emotion recognition, there is no agreement upon which features are most appropriate. Since the computational complexity of multiple feature extraction is high, and the extraction of some features requires rich experience and professional knowledge, only a few studies compared the performance of different features. For example, Jenke et al. (2014) compared multiple features on a self-recorded dataset of 16 subjects and five emotions. Li et al. explored two set of features for cross-subject emotion recognition, and the Hjorth parameter of mobility in the beta rhythm achieved the best mean recognition. Moreover, using multiple electrodes will need more preparation time and lead to unfriendly user experience. Many studies have been conducted on the electrode selection in emotion recognition. For instance, Zheng and Lu (2015) explored the four most

important electrodes on the SEED dataset, and they found FT7, FT8, T7, and T8 are most important electrodes. Li et al. (2018) explored the most important 10, 14, and 18 electrodes on the DEAP dataset, and most of these electrodes are distributed in the frontal brain region.

Previous studies have investigated feature extraction and selection in emotion recognition, but a major limitation of these studies is that they did not explore the effective features for the recognition of different valence emotions. Meanwhile, most studies rely on a different and usually small dataset. This work aims to systematically explore effective features for the recognition of different valence emotions. According to previous studies, arousal has a nonspecific effect on valence (Johnson, 1995; Carretié et al., 1997). Firstly, this study selected five types of valence pictures from the Chinese Affective Picture System (CAPS) (Bai et al., 2005), including extremely negative, moderately negative, neutral, moderately positive, and extremely positive, which have no significant difference in arousal. Then, 110 features commonly used in other papers are extracted from five feature domains of time domain, frequency domain, time-frequency domain, spatial asymmetry, and brain network. Finally, the classification performance, calculation time, and important few electrodes of the features are systematically compared and analyzed on the self-built big dataset of 40 subjects and the public dataset DEAP (Koelstra, 2012).

Materials and methods

Participants

This experiment includes 40 healthy subjects (20 women) of native college students, aged 18–28 years old (average age 22 years old), one of whom is double handed, and the rest are right-handed. All subjects had no mental illness, did not take drugs that affected their mental state, and all subjects with normal mental state tested by Baker depression scale (Jackson-Koku, 2016) and Baker anxiety scale (Wilson et al., 1999), and all subjects' with normal vision or corrected normal vision. Before the experiment, each subject was informed of the content and the purpose of the experiment and signed the informed consent. After the experiment, subjects received a certain amount experimental fee.

Experiment procedure

Emotion pictures are divided into five categories according to valence: extremely negative (EN), moderately negative (MN), neutral, moderately positive (MP), and extremely positive (EP), and each category contains 30 pictures. The mean value and standard deviation of valence degrees of the different categories pictures are EN = 1.87/0.35, MN = 3.56/0.54, neutral = 5.6/0.49, MP = 6.28/0.17, and EP = 6.81/0.16, and the arousal degrees are EN = 5.54/0.16, MN = 5.5/0.2, neutral = 5.54/0.28, MP = 5.49/0.19, and EP = 5.57/0.18. There are significant differences between the valence degrees of the five categories of pictures ($P < 0.01$), and there was no significant difference ($P > 0.05$) between the arousal degrees.

The numbers in the digital picture are 2, 3, 4, 5, 6, 7, 8, and the number of numbers varies from 3 to 6. The content of the digital picture is 3*3, and numbers or "*" appear randomly at 9 positions. The left and right numbers are not adjacent in the picture, e.g., the left and right numbers will not be "2" and "3." There are two types of digital pictures: consistent and inconsistent. The consistent situation is that there are more numbers with large values or fewer numbers with small values, while the inconsistent situation is that there are more numbers with small values or fewer numbers with large values.

Each trial begins with a white "+" for 2–4 s in the center of the screen with black background, then presents an emotion picture for 2000 ms, then presents a digital picture for 1000 ms, and then presents the valence and arousal rating pictures. When the subjects see the digital picture, they need to press the key quickly and accurately to determine which side of the numbers on the left and right is larger. If the number on the left is larger, the subjects should press the alphabet "Q" on the keyboard with the left index finger; if the number on the right is larger, they should press the number "0" on the keyboard with the right index finger, and the key response should be made within 1000 ms. The digital picture will disappear once the subjects press the keyboard, and the valence rating picture will be represented. If the key response is not made after 1000 ms, the valence rating picture will also be represented. Valence and arousal ratings are achieved by pressing keys 1–9 on the keyboard (Morris, 1995). The subjects take a 2-min break between blocks to eliminate the emotional impact of the previous block on the next block and alleviate the subjects' mental fatigue. Five blocks are presented randomly, and 30 trials in each block are presented in random order. The experimental paradigm was conducted by Tobii Pro Lab software, and the subjects' key response values and time were recorded.

Data acquisition and preprocessing

This experiment was carried out in a professional laboratory with electromagnetic shielding condition and suitable temperature and light. In the experiment, the subjects sat on a chair with adjustable height facing the screen, and their eyes were about 65 cm away from the screen. EEG signals were recoded with 64 channel G.HIamp system. During the experiment, the impedance of each electrode was kept below 10 KΩ, the electrodes were located according to the international 10–20 standard system. Electrode AFz was used as

the ground electrode, electrode Fz and right earlobe were used as references, and the number of effective electrodes was 62. The EEG data sampling rate is 600 Hz. Online 0.1–100 Hz band-pass filtering and 48–52 Hz notch filtering are conducted during EEG data acquisition. Meanwhile, eye gaze data is collected by the Tobii Nano device with a sampling rate of 60 Hz. All these multi-modal data compose the Emotion-Stroop dataset (ESD). This paper only uses EEG data. In the pre-processing procedure, the collected original EEG signals are segmented first, and the data of 500 ms before picture presentation and the data of 2000 ms during picture presentation are extracted. The subsequent preprocessing mainly includes 0.1–80 Hz filtering and performing the blind-source analysis algorithm Fast-ICA (HyvRinen, 1999) to remove electrooculography (EOG) artifacts, average reference, and baseline correction.

Feature extraction

This manuscript summarizes the commonly used EEG features and extraction methods in recent years. Jenke et al. have extensively studied feature extraction (Jenke et al., 2014), and our study supplements some recently developed important features on their basis. The features are roughly divided into five feature domains: time domain, frequency domain, time-frequency domain, spatial asymmetry, and brain network features. In this paper, the total number of EEG electrodes is denoted as ch, the number of time points per electrode is denoted as N, and the EEG data of a certain electrode at a certain time is denoted as x (n). The specific extraction methods of each feature are as follows.

Time-domain features

The time-domain features extracted in this paper include standard deviation, first-order difference, second-order difference, normalized first-order difference, normalized second-order difference, fractal dimension, sample entropy, and approximate entropy. The specific calculation methods are as follows.

- Standard deviation (Std)

$$\delta_{ch} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(x(n) - \frac{1}{N}\sum_{n=1}^{N}x(n)\right)^2} \tag{1}$$

- First-order difference (Fir-dif)

$$first\_diff_{ch} = \frac{1}{N-1}\sum_{n=1}^{N-1}|x(n+1) - x(n)| \tag{2}$$

- Normalized first-order difference (N-fir-dif)

$$Nor\_first\_diff_{ch} = \frac{first\_diff_{ch}}{\delta_{ch}} \tag{3}$$

- Second-order difference (Sec-dif)

$$\sec ond\_diff_{ch} = \frac{1}{N-2}\sum_{n=1}^{N-2}|x(n+2) - x(n)| \tag{4}$$

- Normalized second-order difference (N-sec-dif)

$$Nor\_\sec ond\_diff_{ch} = \frac{\sec ond\_diff_{ch}}{\delta_{ch}} \tag{5}$$

- Fractal Dimension (FD)

Fractal dimension (FD) is a non-linear feature used to measure the complexity of EEG signals. The commonly used calculation methods of fractal dimension are box dimension fractal and Higuchi fractal. In this manuscript, the Higuchi fractal is used to calculate the fractal dimension (Lan et al., 2016), and the specific calculation process is as follows:

Let the initial sequential EEG signal be $X(1), X(2), \ldots, X(N)$. The EEG signal sequence is sampled at every k points as follows:

$$\begin{aligned} &X_k^m : X(m), X(m+k), \ldots, X(m + [\tfrac{N-m}{k}] \cdot k) \\ &m = 1, 2, 3, \ldots, k \end{aligned} \tag{6}$$

where m is the initial time of sampling, and k is the time interval of sampling.

Define m sampling points as $L_k(m)$:

$$L_k(m) = \frac{1}{k} \cdot \left[\frac{\left(\sum_{i=1}^{\lfloor \frac{N-m}{k} \rfloor} |X(m+ik) - X(m+(i-1))k|\right)(N-1)}{\lfloor \frac{N-m}{k} \rfloor k}\right] \tag{7}$$

Denote the mean value of all the sampling points in $L_k(m)$ as $L(k)$. FD is inversely proportional to $L(k)$ as follows:

$$FD = -\lim_{k\to\infty}\frac{\log\langle L(k)\rangle}{\log k} \tag{8}$$

- Approximate Entropy (ApEn)

Approximate entropy (ApEn) reflects the possibility of new information in time series. The larger the approximate entropy,

the more complex the time series. Denote an integer as m and a real number as r. Then, an m-dimensional vector $x(1)$,$x(2)$, $\ldots x(\mathrm{N}-\mathrm{m}+1)$ can be constructed from the original EEG signal, where $x(\mathrm{i}) = [\mathrm{X}(1), \mathrm{X}(2), \ldots \mathrm{X}(\mathrm{i}+\mathrm{M}-1))]$ counts the number of vectors that meet the following conditions:

$$C_i^m(r) = (number\ of \quad X(j)\ such\ that\ d[X(i), X(j)] \leq r)/ \\ (N-m+1) \tag{9}$$

where $d[X(i), X(j)] = \max\left|X(i) - X(j)\right|$,

$$\Phi^m(r) = (N-m+1)^{-1} \sum_{i=1}^{N-m+1} log(C_i^m(r)) \tag{10}$$

Then, the approximate entropy ApEn is defined as:

$$ApEn = \Phi^m(r) - \Phi^{m+1}(r) \tag{11}$$

- Sample Entropy (SamEn)

Sample entropy (SamEn) is improved based on approximate entropy by eliminating the problem of approximate entropy self-matching, which is equivalent to optimizing the approximate entropy. In the calculation $d[X(i), X(j)] = \max\left|X(i) - X(j)\right|$, $i \neq j$

$$C^m(r) = (N-m+1)^{-1} \sum_{i=1}^{N-m+1} (C_i^m(r)) \tag{12}$$

Then, SamEn is defined as:

$$SampEn(m, r) = \lim_{N\to\infty}[-\ln \frac{C^{m+1}(r)}{C^m(r)}] \tag{13}$$

when n is a finite number. SamEn can be further expressed as:

$$SampEn(m, r, N) = \ln C^m(r) - \ln C^{m+1}(r) \tag{14}$$

Frequency domain features

- Power Spectral Density (PSD)

Power spectral density (PSD) is commonly used to measure the frequency-domain information features of EEG signals. In this manuscript, power spectral density uses the p-welch method to calculate the frequency band power spectral density:

$$psd = \sum_{fl}^{fu} P(f)/(fu - fl) \tag{15}$$

where $P(f)$ is the power spectral density at the frequency; fl is and fu are the lowest and highest frequency of the band of interest, respectively.

- Power (P)

Band power is based on short-time Fourier transform (STFT),

$$\mathrm{STFT}_{x,\gamma}(n, f) = \int_{-\infty}^{+\infty} x(\tau)\,\gamma^*(n-\tau)\,e^{-j2\pi f\tau} d\tau = \\ \int_{-\infty}^{+\infty} x(\tau)\,\gamma_{n,f}^* e^{-j2\pi f\tau} \tag{16}$$

$$power = \sum_{fl}^{fu} \left|S(n, f)\right|^2 \tag{17}$$

The power in six frequency bands is calculated for ESD data, including delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), gamma (30–50 Hz), and high gamma (50–80 Hz). Meanwhile, the power in four frequency bands is calculated for the DEAP dataset, including theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–50 Hz).

Time frequency domain features

- Different Entropy (DE)

The differential entropy (DE) feature is the most used feature at present, and its calculation is based on STFT. It calculates the differential entropy in six frequency bands of delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), gamma (30–50 Hz), and high gamma (50–80 Hz) of the ESD datasets, and it calculates the differential entropy in four frequency bands of theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30- 50 Hz) of the DEAP datasets:

$$DE = \log\left(\sum_{fl}^{fu} \left|STFT(n, f)\right|^2\right) \tag{18}$$

Brain network features

The brain network connection matrix takes electrodes as network nodes to calculate the relationship between the data between electrodes, and the network can be mainly divided into a directed network and an undirected network. Here, only three commonly used undirected networks are considered for calculation: Pearson correlation (Pea), coherence (Coh), and phase lock value (PLV). Then, the clustering coefficient (CC), characteristic path length (CPL), and local efficiency (Le) are calculated based on the network connection matrix, and the global efficiency (Ge) is characterized by four commonly used network attributes (Van Straaten and Stam, 2013).

- Pearson correlation coefficient (Pea)

$$pearson = \frac{\sum_{i=1}^{N}(X(i)-\overline{X})(Y(i)-\overline{Y})}{\sqrt{\sum_{i=1}^{N}(X(i)-\overline{X})^2}\sqrt{\sum_{i=1}^{N}(Y(i)-\overline{Y})^2}} \quad (19)$$

where $X(i)$, $Y(i)$ indicating the EEG value of two electrodes at time i.

- Coherence (Coh)

$$C_{XY}(f) = \frac{|P_{XY}(f)|^2}{P_{XX}(f)\,P_{YY}(f)} \quad (20)$$

- Phase lock value (PLV)

$$PLV = \left|\frac{1}{ch}\sum_{i=0}^{N-1} e^{i(\Phi_x(it)-\Phi_y(it))}\right| \quad (21)$$

Where ch is the total number of EEG electrodes, t is the sampling period, $\Phi_x(it)$ and $\Phi_y(it)$ is the instantaneous phase of two electrodes $x(t)$ and $y(t)$ at time point i.

The CC describes the tightness and clustering characteristic of nodes in the brain network; the CPL is used to measure the connectivity degree of the network, and it represents the average length of the shortest path between any two nodes in the network; the Le is used to measure the information interaction ability in the local network, and the Ge describes the information transmission efficiency of the whole brain network. The definitions of the four network properties are given below, where n represents the number of nodes; Θ represents a node set; w_{ij} indicates the network connection value between nodes i and j; d_{ij} represents the shortest path length between nodes i and j.

$$CC = \frac{\sum_{j,h\in\Theta}\left(w_{ij}w_{ih}w_{jh}\right)}{\sum_{j\in\Theta}w_{ij}\left(\sum_{j\in\Theta}w_{ij}-1\right)}^{1/3} \quad (22)$$

$$CPL = \frac{1}{n}\sum_{i\in\Theta}\frac{\sum_{i\in\Theta,j\neq i}d_{ij}}{n-1} \quad (23)$$

$$Le = \frac{\sum_{j,h\in\Theta,j\neq i}\left(w_{ij}w_{ih}[d_{jh}(\Theta_i)]^{-1}\right)}{\sum_{j\in\Theta}w_{ij}\left(\sum_{j\in\Theta}w_{ij}-1\right)}^{1/3} \quad (24)$$

$$Ge = \frac{1}{n}\sum_{i\in\Theta}\frac{\sum_{j\in\Theta,j\neq i}\left(d_{ij}\right)^{-1}}{n-1} \quad (25)$$

Spatial asymmetry features

Spatial asymmetry features are based on the asymmetry characteristic of the brain reported in previous studies, mainly including differential asymmetry (DA) and rational asymmetry (RA) (Duan et al., 2012). The DA features represent the subtraction values of the features from the left and right hemisphere electrodes, and the RA features represent the ratio of the features from the left and right hemisphere electrodes. DA and RA are defined as follows:

- Differential Asymmetry (DA)

$$DA = fea_l - fea_r \quad (26)$$

- Rational Asymmetry (RA)

$$RA = fea_l/fea_r \quad (27)$$

Where fea_l and fea_r represent the features extracted by the symmetrical position electrodes of the left and right hemispheres, respectively. The asymmetry features of ESD data include 27 pairs of electrodes, and the DEAP dataset includes 17 pairs of electrodes.

Feature selection

The purpose of feature selection is to find the key electrodes for the recognition of different valence emotions and provide a foundation for using a few electrodes for emotion recognition in practical applications.

Currently, the commonly used feature selection methods mainly include Relief (Jia et al., 2013), min redundancy max relevance (mRMR) (Ding and Peng, 2005), and forward floating search (Bhadra and Bandyopadhyay, 2021). Among them, mRMR is the most famous feature selection algorithm and has been applied in many emotion recognition studies. mRMR exploits mutual information to characterize the performance of feature subsets. This study also used the mRMR feature selection algorithm to explore the key electrodes for emotion recognition. This study attempted to find the most important 1, 4, 8, and 16 electrodes for the recognition of different valence emotions. First, the most important 1, 4, 8, and 16 electrodes in each subject's classification are determined based on mRMR. Then, the frequency of each electrode in 40 subjects is counted, the most selected electrodes are taken as key electrodes, and the electrodes' location in the brain regions is also analyzed.

Classification settings

Classification is to match features with emotions to obtain classification accuracy. Classifiers can be roughly divided into two categories. The first category is the current popular

classifiers based on the deep neural network (DNN). These classifiers are mainly the convolution neural network (CNN) (Santamaria-Vazquez et al., 2020), recurrent neural network (RNN) (Lukosevicius and Jaeger, 2009), long and short-term memory (LSTM) (Peng et al., 2016), and graph revolutionary neural network (GCNN) (Zhong et al., 2020). In recent years, most of these DNN classifiers have achieved excellent classification results, and most of these DNN classifiers have re-extracted the input features (Chen et al., 2019). However, due to the "black box" characteristic of DNNs, the significance of re-extracted features cannot be clearly explained (Zhou et al., 2016). The other category is the traditional shallow classifiers, such as support vector machine (SVM) (Chen et al., 2020), k-nearest neighbor (KNN) (Keller et al., 2012), linear discriminate analysis (LDA) (Saeedreza et al., 2009), extremely learning machine (ELM) (Huang et al., 2006), random forest (RF) (Liaw and Wiener, 2002), naive Bayes (NB) (Rish, 2001), discriminant analysis classifier (DAC) (Alkan and Günay, 2012), and boosting (Sun et al., 2007). Usually, the hyper-parameters in the DNN need to be tuned (He et al., 2019), and this procedure will change input features to unknown features (Saha and Fels, 2019), which makes it difficult to objectively compare the performance of different features. Most of the shallow classifiers do not need to tune complex parameters. Therefore, this paper used the six commonly used shallow classifiers (SVM, KNN, RF, NB, DAC, and boosting) to compare the classification performance of different features. Meanwhile, LibSVM (Chen et al., 2020) was used with a linear kernel, and the parameter was set to "−s 0 −t 0."

Two experiments were conducted on the classification of the ESD dataset. The first is a two-category experiment that separates EN valence emotions from other valence emotions. In this experiment, 30 samples of EN valence emotions of each subject are regarded as one category, and 120 samples of MN, MP, EP, and neutral emotions are mixed as another category. Because it is a mixture of 4 emotions, each subject is sampled 4 times, 5-fold cross-validation is used for classification, and the average accuracy of 4*5 times classification is taken as the final accuracy of each subject. The second experiment is to classify five types of valence emotions: EN, MN, MP, EP, and neutral. Five-fold cross-validation is used in the classification, and the average accuracy five times classification is used as the final classification accuracy of each subject.

The DEAP dataset contains multimodal data such as EEG, galvanic skin response and respiratory rate during 32 subjects watching 40 1-min music videos with different valence and arousal. The EEG signals are collected from 32 active electrodes arranged according to the 10–20 international system. Only EEG signals are used in this study. To determine the effective features of different valence emotion classification in the DEAP dataset, the samples with arousal ratings in the range of 3.5–6 of each subject are selected, and then these selected samples are divided into high valence emotion samples with valence ratings greater than 5 and low valence emotion samples with valence ratings less than 5. Since the two types of samples of some subjects are unbalanced, this study calculates the sample number of two categories and then randomly selects the same number of samples as the fewer samples category from the category with more samples. According to experimental experience and previous research, emotion does not occur immediately after the stimulus is presented, so only the last 30 s of data induced by each video in the DEAP dataset were used in the experiment, and the data were divided into 5 s by non-overlapping segmentation. The preprocessing method of the EEG data in the DEAP dataset is consistent with that in the original dataset.

Results

Behavior data

The mean reaction time (RT) of 40 subjects under five different valence emotions is presented in **Table 1**. Compared with neutral valence emotion, the other four valence emotions all cause the subjects to react more slowly. This indicates that different valence emotions can affect the subjects' reaction ability. When subjects are under an extremely negative valence emotion, the reaction time is longer than that under other valence emotions ($p < 0.05$), and the reaction time is the shortest when the subjects are under natural emotions. Subjects respond to trials involving MP and EP more slowly than to natural trials, but it does not reach significance ($P = 0.06$). So, different valence emotions can affect cognitive performance.

Classification performance of different features

The classification performance of each feature for different valence emotions was first compared on the ESD dataset and DEAP dataset. Then, the highest classification accuracy of the time domain, frequency domain, time-frequency domain, DA, RA, and brain network features by using the classifiers of KNN, RF, SVM, DAC, Bayes, and boosting were presented,

TABLE 1 Mean reaction time under different valence emotions.

Emotion	RT (ms)	Significance
EN	617.8	**
MN	599.6	*
Natural	585.0	–
MP	596.4	*
EP	581.7	–

**$P < 0.05$; *, in the edge of significance; -, no significance.

respectively. Two-class and five-class classification experiments were conducted on the ESD dataset, and high and low valence two-class classification experiment was carried out on the DEAP dataset.

Classification performance on the ESD dataset

As shown in Figure 1, all features achieved a classification accuracy of no less than 70% in recognizing the EN emotion from other valence emotions, and the highest classification accuracy of 93.7% is achieved with the second-order difference by the SVM classifier. According to the classification results, time-domain features, first-order difference features, and second-order difference features perform better. The classification accuracy of first-order difference features under the KNN classifier is 90.3%, and the classification accuracy of second-order difference features under the SVM classifier is 93.7%. High gamma (HG) band power achieved the highest classification accuracy of 90% under the SVM classifier among frequency-domain features. The high gamma band DE obtained the highest accuracy in the time-frequency domain, and the classification accuracy under the SVM classifier is 92.1%. For brain network features, the CPL of the high gamma band coherence network performed the best and achieved a classification accuracy of 71.4% under KNN. For spatial asymmetric features, the DA feature extracted from second-order difference features obtained the highest classification accuracy of 89.7% under the SVM classifier, and the RA features extracted from first-order difference features performed the best and achieved a classification accuracy of 88.8% under the SVM classifier. The results show that the performance of first-order difference and second-order difference features in the time domain can achieve higher accuracy than that of power and differential entropy, which are mostly used in previous studies. In addition, it can be found that frequency-domain features, time-frequency domain features, and coherence network features in higher frequency bands performed better than those in lower frequency bands, and the high gamma band features achieved the highest classification accuracy. By comparing the performance of different classifiers, it can be found that the SVM classifier performed better in binary classification on ESD datasets.

As shown in Figure 2, for the classification of five-class valence emotions on ESD datasets, the highest classification accuracy of 89.9% is achieved by second-order difference features under the SVM classifier and by FD features under the DAC classifier. Among time-domain features, second-order difference features and FD features achieved the best performance. In the frequency domain, high gamma band power performed best, and the classification accuracy achieved by the SVM classifier was 84.6%. Among time-frequency features, the high gamma band DE achieved the highest classification accuracy of 87.5% under the SVM classifier. The high gamma band CPL of the coherence network performed the best among all brain network properties, and the classification accuracy based on the DAC classifier is 60.6%. For spatial asymmetric features, both the DA features and RA features extracted from FD features achieved the highest classification accuracy of 85.7% under the DAC classifier. When classifying five kinds of valence emotions, it was also found that the accuracy of features in higher frequency bands is higher than that in lower frequency bands, and the high gamma band feature obtained the highest classification accuracy. Comparing the performance of different classifiers, it can be found that SVM and DAC classifiers have better classification performance in the classification of five-class valence emotions on ESD datasets.

Classification performance on the DEAP dataset

As shown in Figure 3, on the DEAP dataset, the classification accuracy of various features for high and low valence emotions is not less than 50%, and the first-order differential features achieved the highest classification accuracy of 69.4% under the SVM classifier. Among the time-domain features, the first-order differential features achieved the highest classification accuracy of 69.4% under the SVM classifier. In frequency-domain features, the classification performance of gamma-band power features is the best, and the classification accuracy is 66.3% under the RF classifier. In the time-frequency domain, the gamma band differential entropy obtained the highest classification accuracy of 67.6% under the SVM classifier. Among the brain network properties, the gamma band local efficiency extracted from the coherence network achieved the highest classification accuracy of 62.7% under the RF classifier. Among the DA features, the gamma band DE obtained the highest classification accuracy of 66.4% under the DAC classifier. The RA features extracted from gamma band DE features performed the best and achieved the highest classification accuracy of 66.8% under the DAC classifier. On the DEAP dataset, it was also found that the features in higher frequency bands performed better than those in lower frequency bands, and the gamma band features achieved the highest classification accuracy, which is consistent with the results on the ESD dataset. Comparing the performance of different classifiers, it can be found that SVM and DAC classifiers have better classification performance.

To show the classification performance of different features visually, this study selected one subject's data and adopted the t-SNE algorithm to map high-dimensional features to two-dimensional space and compare the distribution of features within and between classes. On the ESD dataset, six types of features were selected, namely, first-order difference, second-order difference, FD, high gamma power, high gamma differential entropy, and high gamma band CPL of the coherence network, to visualize by t-SNE. On the DEAP dataset, first-order difference, second-order difference, FD, gamma-band power, gamma band DE, and gamma band

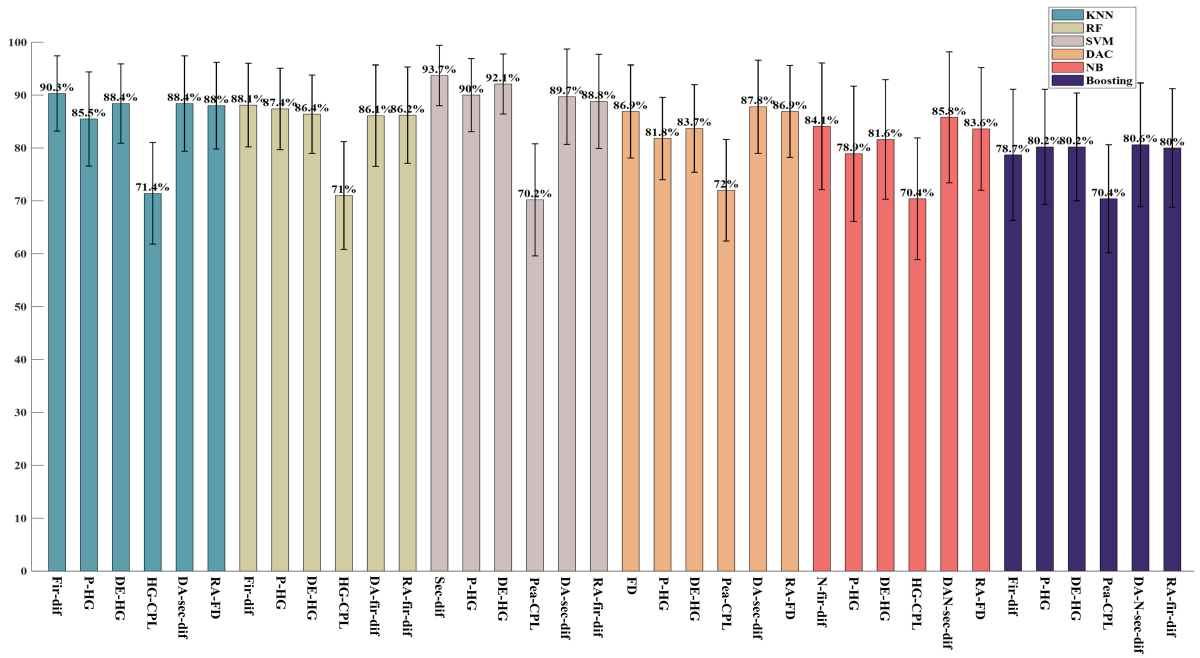


FIGURE 1
The highest two-class classification accuracy in each feature domain on the ESD. The block which including six bars of same color is the results of same classifier and from left to right, it is the classification results by classifiers KNN, RF, SVM, DAC, NB, and Boosting, respectively.

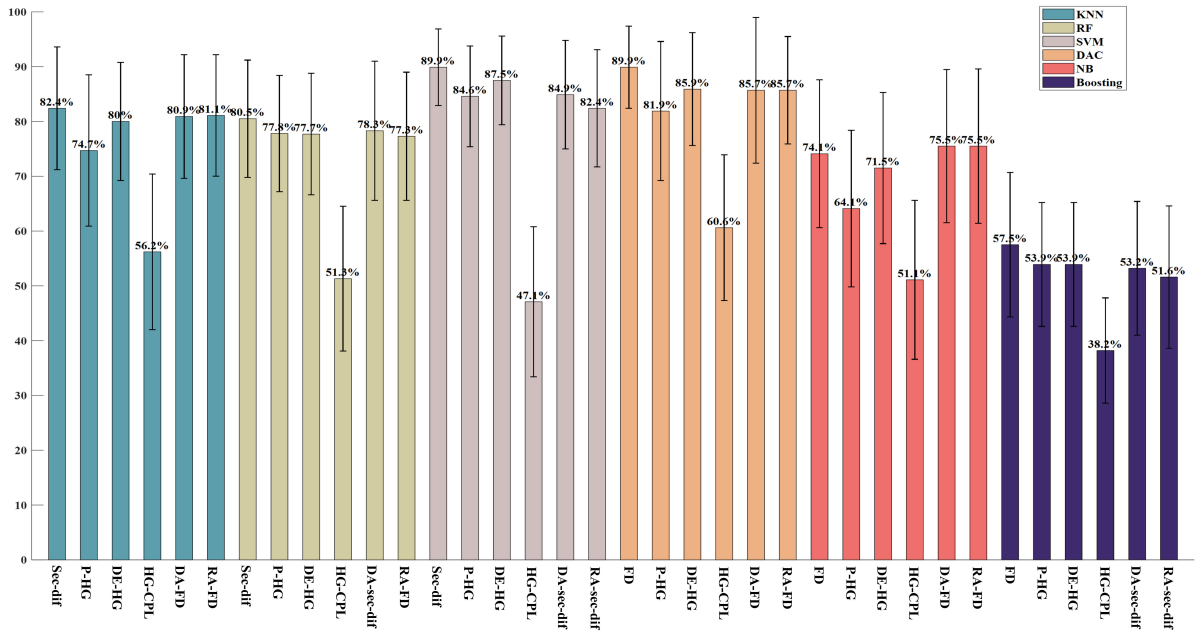


FIGURE 2
The highest five-class classification accuracy in each feature domain on the ESD. The block which including six bars of same color is the results of same classifier and from left to right, it is the classification results by classifiers KNN, RF, SVM, DAC, NB, and Boosting, respectively.

Le of the coherence network were selected to visualize by t-SNE (Donahue et al., 2013). **Figures 4**, **5** show the feature distribution maps of the ESD and DEAP dataset. The feature distribution map shown in **Figure 4** indicates that the second-order difference, FD, high gamma power, and high gamma DE, which obtained higher classification accuracy on the ESD

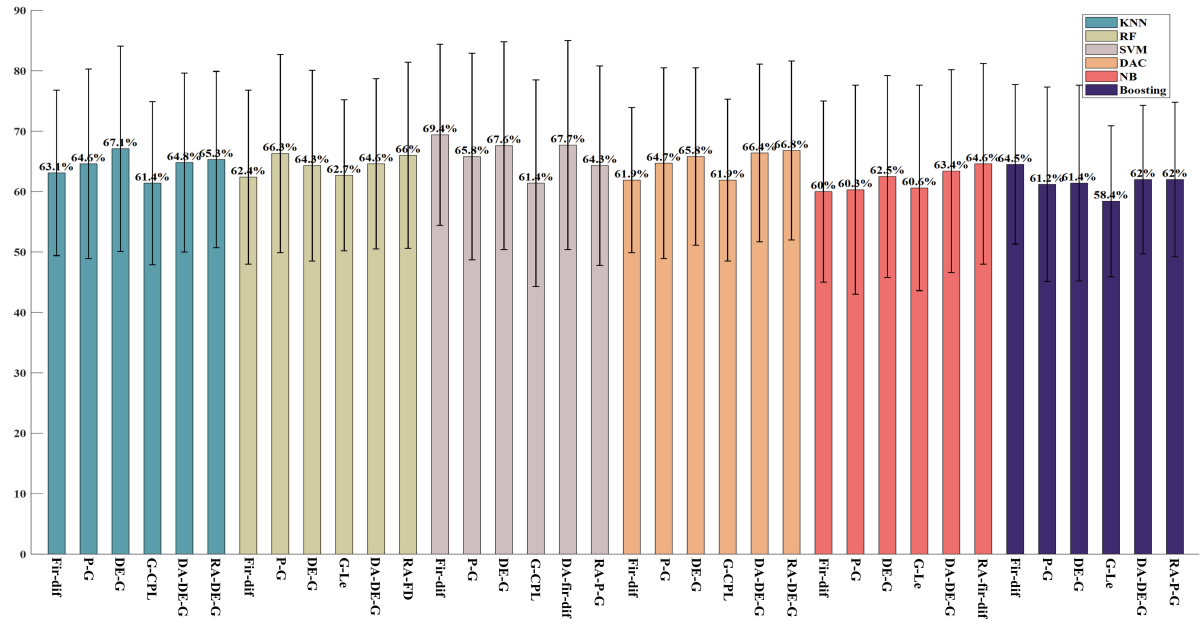


FIGURE 3
The highest classification accuracy for high and low valence emotion in each feature domain on the DEAP. The block which including six bars of same color is the results of same classifier and from left to right, it is the classification results by classifiers KNN, RF, SVM, DAC, NB, and Boosting, respectively.

dataset have a small distance within the class and a large distance between classes. From Figure 5, it can be seen that the first-order difference, second-order difference, and gamma DE on the DEAP dataset have a small distance within the class and a large distance between classes. The feature visualization results explain why the features performed better in the classification of different valence emotions, i.e., features with a small intra-class distance and a large inter-class distance can achieve higher accuracy.

Important electrodes

From the classification results on both the ESD dataset and the DEAP dataset, it can be found that the features with better classification performance are the first-order difference, the second-order difference in the time domain, the high-frequency band power, and the high-frequency band DE features. Based on these four features, this study adopted the mRMR algorithm to reduce the feature dimension and find the most important electrodes on the ESD dataset and the DEAP dataset, respectively. On the ESD dataset, the most important 1-dimension, 4-dimension, 8-dimension, and 16-dimension features were selected from all 62 dimensions of each feature, and on the DEAP dataset, the features were selected from all 32 dimensions of each feature.

Figures 6, 7 show the most important 1, 4, 8, and 16 electrodes of the first-order difference, second-order difference, high-frequency band power, and high-frequency band DE on the ESD dataset and the DEAP dataset, respectively. The blue rotundities in the figure represent the selected electrodes. From the results in the table, it can be seen that the electrodes from the prefrontal and temporal lobes are important for selecting the most important 1-dimension and 4-dimension features, and the features from the electrodes distributed in the occipital lobe are also selected when choosing more features.

According to the above classification performance comparison results, the first-order difference, second-order difference, high-frequency band energy, and high-frequency band DE show better classification performance. Then, this study investigated the classification performance of the four features extracted from the above-mentioned most important 1, 4, 8, 16 electrodes and all electrodes. Figures 8, 9 show the classification results of the features extracted from different numbers of electrodes on the ESD dataset and DEAP dataset, respectively. It can be seen from the results that the more electrodes are used, the higher the classification accuracy is. On the ESD dataset, when the number of electrodes is reduced by 3/4, i.e., using 16 electrodes located in the frontal lobe, temporal lobe, and occipital lobe, the classification accuracy is only 2% lower than that using all 62 electrodes. On the DEAP dataset, when only 1/2 of all electrodes are used, the gamma band DE even achieved 0.8% higher accuracy than that of using all 32 electrodes, and the accuracies of the first-order difference, second-order difference, and gamma band power decreased.

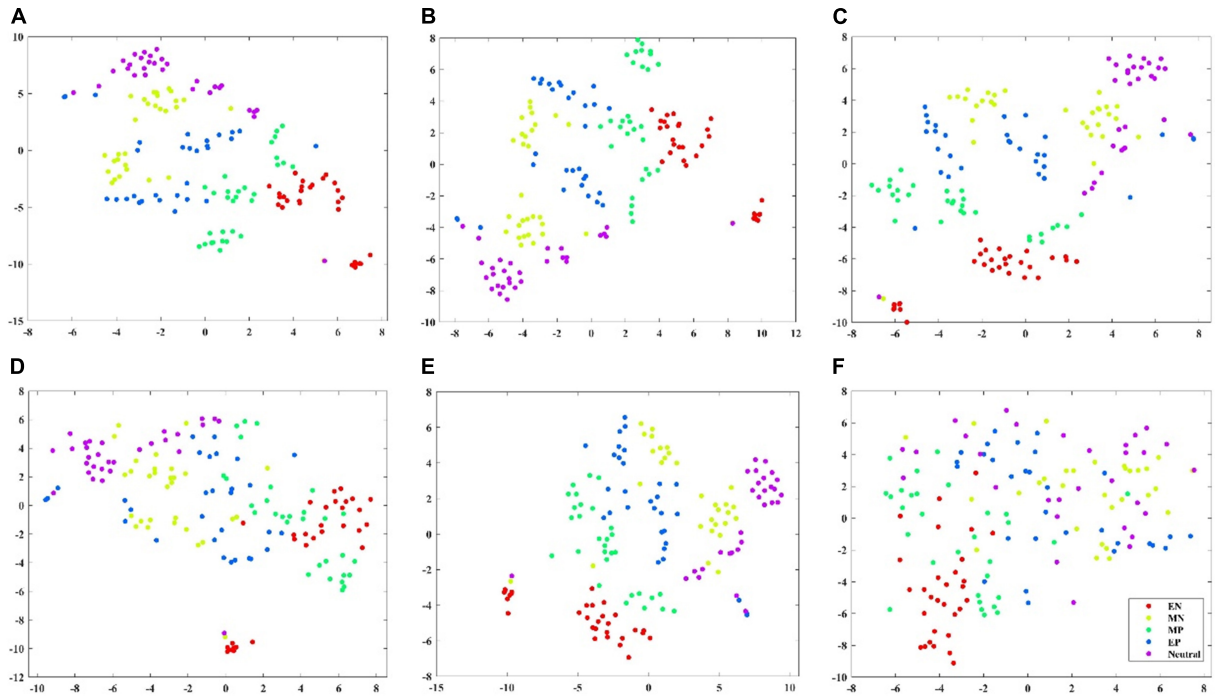


FIGURE 4
Feature visualization map on ESD dataset. **(A)** Second-order difference; **(B)** first-order difference; **(C)** fractal dimension; **(D)** high gamma band power; **(E)** high gamma band differential entropy; **(F)** high gamma band CPL of coherence network.

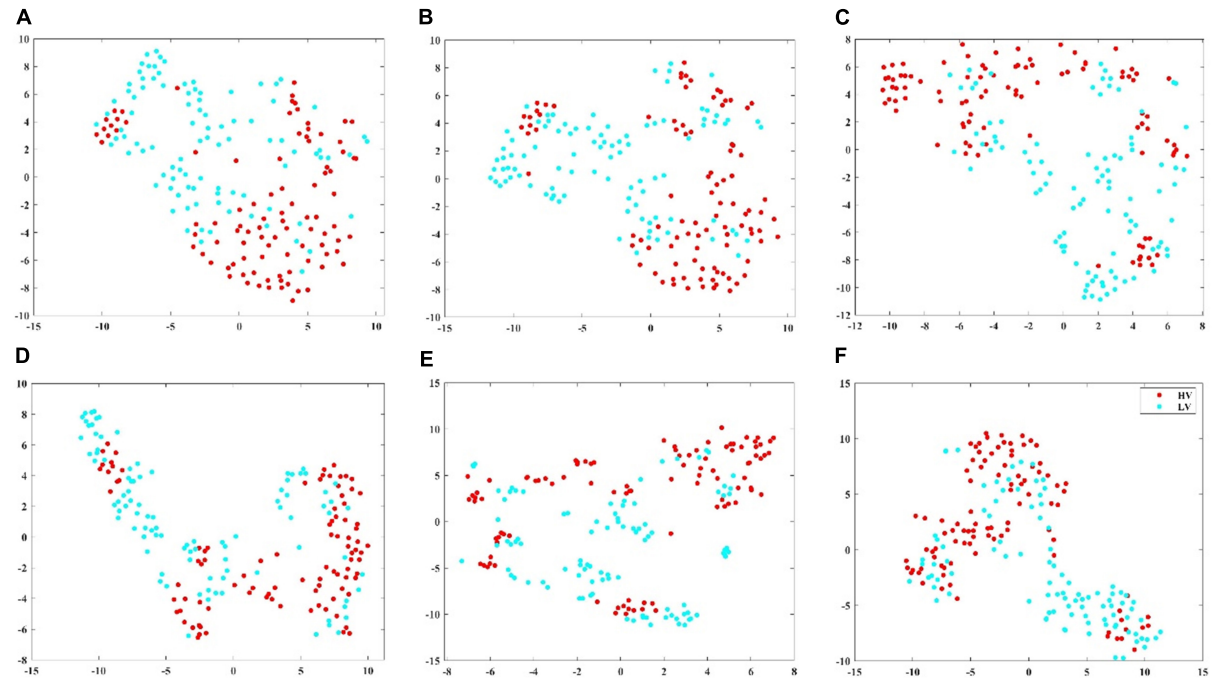


FIGURE 5
Feature visualization map on DEAP dataset. **(A)** Second-order difference; **(B)** first-order difference; **(C)** fractal dimension; **(D)** gamma band power; **(E)** gamma band differential entropy; **(F)** gamma band CPL of coherence network.

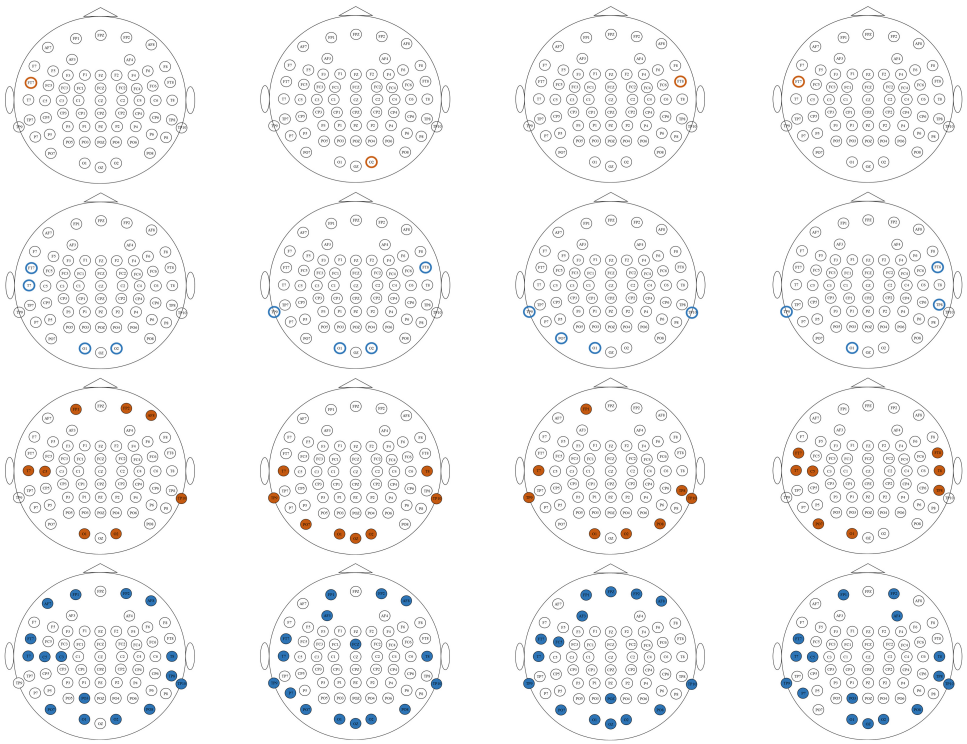


FIGURE 6
The distribution of the most important1, 4, 8, and 16 electrodes on the ESD dataset. Brown circles mark the most important 1 electrode, blue circles mark the most important 4 electrodes, brown solid circles mark the most important 8 electrodes, and blue solid circles mark the most important 16 electrodes. From top to bottom, each row corresponds to first-order difference, second-order difference, high gamma band power, and high gamma band DE, respectively.

Computing time

In practical applications, the feature extraction time should be as short as possible. In this paper, the computing time of features in different feature domains was compared. The computer used in this experiment is the AMAX server equipped with two Intel (R) Xeon (R) Gold 5120 2.20GHz CPUs, 256 GB RAM, and two NVIDIA Titan RTX GPUs, and running the 64-bit windows10 operating system. The data processing of this study was conducted on MATLAB 2018a. The feature extraction time was compared based on one sample 62*1200 (62 indicates the electrode number, 1200 indicates the data length, and the sampling rate is 600 Hz) in the ESD dataset, and the MATLAB commands "tic" and "toc" were used to record the computing time of different features. The calculation times of each feature in different feature domains are presented in **Table 2**.

It can be seen from **Table 2** that the features whose calculation time is less than 0.1 s are FD, first-order difference, second-order difference, normalized first-order difference, normalized second-order difference, standard deviation, CC, CPL, and Ge extracted from the Pearson correlation network. The brain network features extracted from the coherence network need a long computing time, and the computing time is more than 50 s.

Discussion

Classification performance of features

Through the classification results on the ESD dataset and DEAP dataset, it can be found that the four features, namely the first-order difference, second-order difference, high-frequency band power, and high-frequency band DE performed better for the classification of different valence emotions. Specifically, among the four features, the classification performance of time-domain first-order difference and second-order difference under the SVM classifier achieved higher accuracy than that of the most used band power and differential entropy. Lan et al. (2016) also reported that the time-domain features have better classification performance than frequency features. Meanwhile, the feature visualization results show that the first-order difference and second-order difference have a large inter-class distance and a small intra-class distance while characterizing different valence emotions.

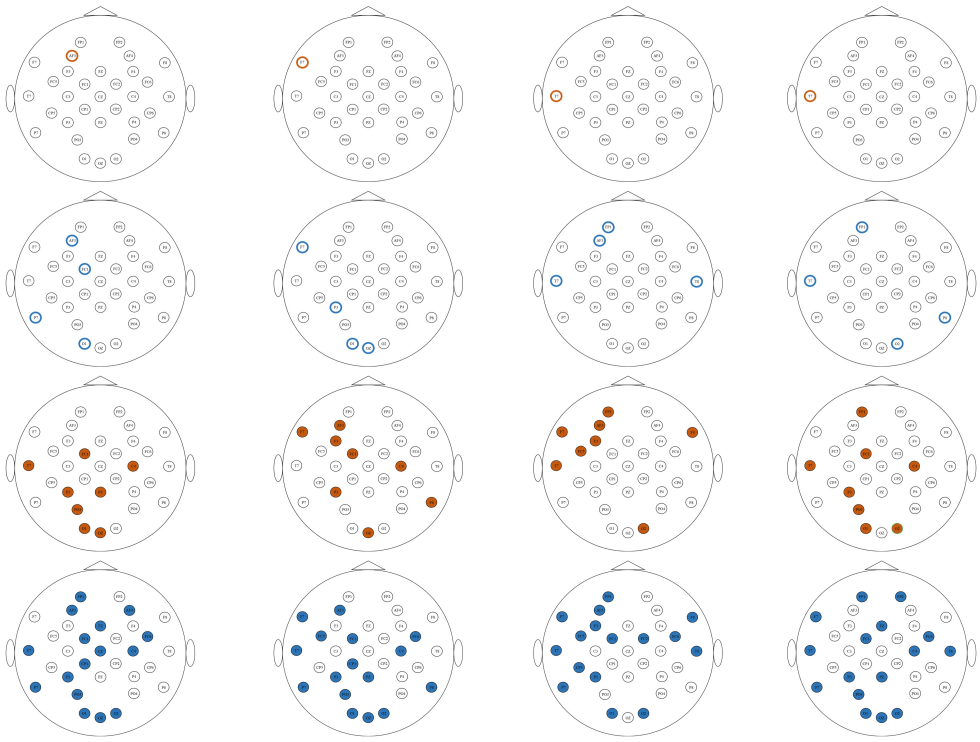


FIGURE 7
The distribution of the most important 1, 4, 8, and 16 electrodes on the DEAP dataset. Brown circles mark the most important 1 electrode, blue circles mark the most important 4 electrodes, brown solid circles mark the most important 8 electrodes, and blue solid circles mark the most important 16 electrodes. From top to bottom, each row corresponds to first-order difference, second-order difference, high gamma band power, and high gamma band DE, respectively.

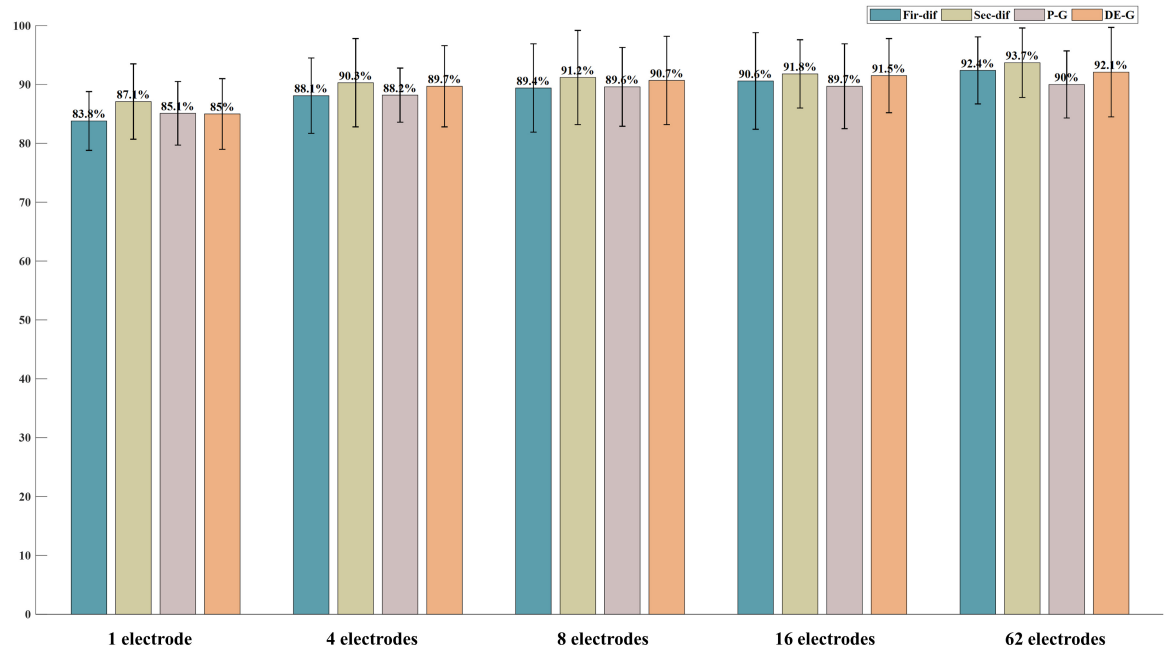


FIGURE 8
The classification results of features from 1, 4, 8, 16, and all 62 electrodes on the ESD dataset.

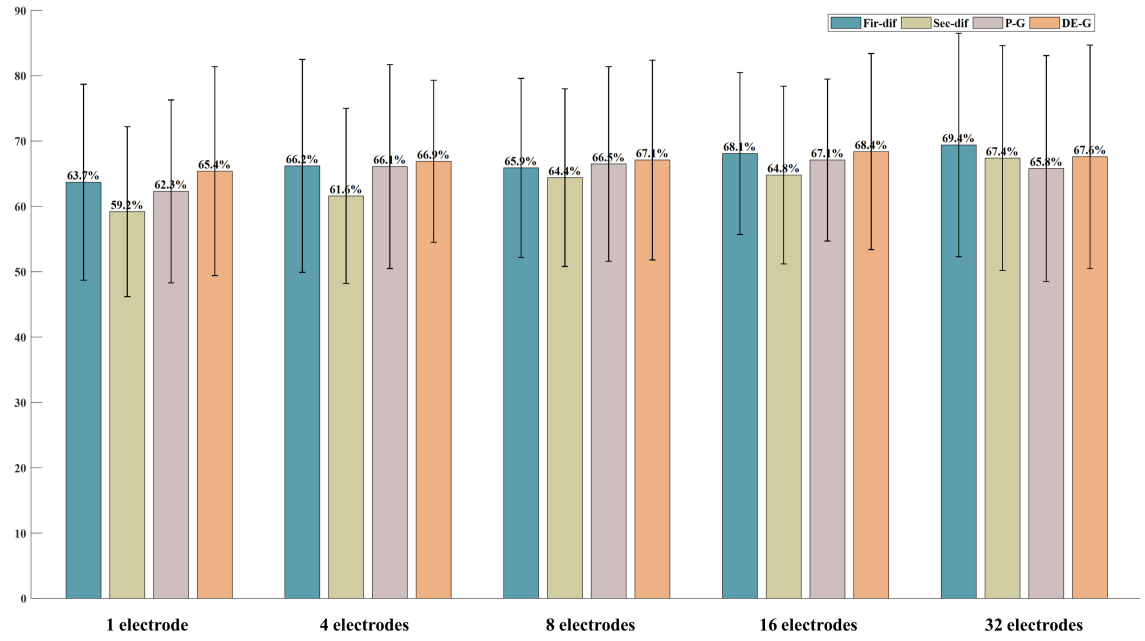


FIGURE 9
The classification results of features from 1, 4, 8, 16, and all 32 electrodes on the DEAP dataset.

From the above results, the feature that performed the best in each feature domain while classifying different valence emotions can be found. Among time-domain features, the first-order difference and second-order difference achieved higher accuracy than other features. In the frequency domain, the high-frequency band power performed better than that in low-frequency bands on both ESD and DEAP datasets. Moreover, on the ESD dataset, the high gamma band power achieved the highest accuracy among all frequency-domain features, and on the DEAP dataset, the gamma band power obtained the highest accuracy. For the time-frequency domain, high-frequency band DE features perform better than those of the low-frequency band, and the high gamma band and gamma band DE achieved the highest accuracy on the ESD and DEAP dataset, respectively. Both frequency domain and time-frequency domain features show that a higher frequency band feature can achieve higher classification accuracy in the classification of different valence emotions. Many previous studies have also reported that high-frequency EEG features have a better performance in emotion recognition (Zheng and Lu, 2015; Zhuang et al., 2018). Our previous studies also proved the effectiveness of high-frequency features in emotion recognition (Yang et al., 2020). This study also explored the performance of brain network features that are widely used in recent years' research. Three brain network calculation methods were used in this study, and then four network attributes were extracted as features. The results showed that the network feature calculated based on the Pearson correlation network had better performance, and the classification performance of CPL shows that it is more effective in characterizing different valence emotions. For DA and RA features, it can be found that the performance of RA features is slightly better than that of DA features. In addition, the performance of spatial asymmetric features is related to the original feature, i.e., if the classification performance of the original feature is good, the classification performance of the corresponding asymmetric feature is also good. This may be because both RA and DA features are both simple linear transformations of the original features. Through the classification results of features in different domains, it can be found that when classifying different valence emotions, the commonly used frequency domain and time-frequency domain features characterizing the rhythm features of EEG should be considered, and more attention should be given to the time-domain features that representing the time-varying information of EEG signals. Overall, when classifying different valence emotions with a first-order difference, second-order difference, high-frequency band power, and high-frequency band DE can achieve better classification results than other features.

Important electrodes

Figures 6, 7 show the most selected electrodes and their distributions in the classification of different valence emotions. The most important features are extracted from similar electrodes of different features, and the results on ESD and DEAP datasets are consistent. The features are mainly extracted from the electrodes distributed at the frontal lobe, occipital lobe,

TABLE 2 Calculation time of different features (s).

Time domain	ApEn	SamEn	FD	Fir-dif	Sec-dif	N-fir-dif	N-sec-dif	Std
	3.053	**6.650**	**0.029**	**0.003**	**0.004**	**0.006**	**0.006**	**0.005**
Frequency/time-frequency	PSD	Power	DE					
	0.602	0.208	0.118					
PLV	CC	CPL	Le	Ge				
	0.464	0.442	2.926	0.435				
Pearson	CC	CPL	Le	Ge				
	0.027	0.058	2.577	0.043				
Coherence	CC	CPL	Le	Ge				
	54.338	58.726	56.025	52.621				

and temporal lobe. Meanwhile, according to previous research, the frontal lobe is the brain region that executes high-level cognitive functions, including emotion processing and memory, and the occipital lobe mainly processes visual information related to emotions. Zheng et al. (2019) also pointed out that the key brain regions for emotion recognition included the frontal lobe, temporal lobe, and occipital lobe. Shuang et al. (2018) also found that the electrodes of the frontal lobe, temporal lobe, occipital lobe, and other brain regions are more important in exploring key electrodes for emotion classification (Liu et al., 2018). According to the results in this paper and previous studies, it can be found that different valence emotions have stable EEG patterns, and the prefrontal lobe, occipital lobe, and temporal lobe play an important role in characterizing different valence emotions. Additionally, the classification performance of different dimensions features was compared in this study, and it was found that the classification accuracy decreases with the reduction in the number of feature dimensions. On the ESD dataset, when the number of electrodes is reduced by 3/4, the classification accuracy is only about 2% lower than that of using all 62 electrodes. On the DEAP dataset, the accuracy of extracting the gamma band DE from only half of all electrodes is higher than that of using all electrodes. These classification results indicate that it is feasible to recognize different valence emotions based on a few electrodes, which can reduce computing complexity and is more convenient in actual applications. Therefore, the designing of an EEG acquisition device with a few electrodes or classifying different valence emotions based on a few electrodes can refer to the electrodes located in the frontal lobe, temporal lobe, and occipital lobe.

Calculation time of features

Feature computing time is also very important for emotion recognition, especially in online emotion recognition because it affects the result feedback of emotion recognition. This study compared the calculation time of different features and presented the features with less calculation time, which can be used as a reference for other studies. Meanwhile, it was found that most time-domain features can be extracted in a short time. Especially, the first-order difference and second-order difference features have low computational complexity, and they are suitable for real-time emotion recognition situations.

Comparison of classifiers

In this study, six commonly used shallow classifiers are used. As shown in the **Figures 1–3**, in each figure every block which including six bars of same color is the classification results of each classifier, and the same sequence location of each block is the result of same feature domain. By comparing the results of the same sequence location of each block in same figure, it can be seen that when feature is fixed, classifiers has different influence on classification results. The comparison results show that the SVM classifier has better performance in two-class classification tasks on both the ESD dataset and DEAP dataset, and the highest classification accuracy of different feature domains is mostly achieved by the SVM classifier. In the classification of five valence emotions, both SVM and DAC classifiers can obtain excellent results, and for some features, the DAC classifier may obtain better results than the SVM classifier. By comparing each block including six bars of same color, it can be seen, when classifier is fixed, the classification is decided by feature. And among all the features, the first-order difference, the second-order difference, the high-frequency band power and the high-frequency band differential entropy performed better. Generally, if we want to achieve the highest classification accuracy, we not only need to select feature but also need to select classifiers, the optimal combination of classifier and feature is required.

Limitations

The limitation of this study is that it only explored the features for different valence emotions, but the effective features

for different arousal emotions were ignored. Meanwhile, the combination of different features can provide complementary information and may contribute to better classification performance, so feature fusion methods will be explored in our future work.

Conclusion

This manuscript systematically evaluated the performance of 110 features extracted from the time domain, frequency domain, time-frequency domain, spatial domain, and brain network on our self-built ESD dataset of 40 subjects and the public dataset DEAP. Meanwhile, the classification performance, computing time, and important electrodes of each feature were systematically analyzed and compared. From the experimental results, it can be seen that the first-order difference, second-order difference, high-frequency power, and high-frequency DE features outperform other features for the recognition of different valence emotions. Also, most time-domain features have less computing time than other features, which are more suitable for online emotion recognition. Besides, the electrodes in the frontal lobe, temporal lobe, and occipital lobe are more important for the recognition of different valence emotions, and when the number of electrodes is reduced by 3/4, the classification accuracy of features from 16 electrodes located in these brain regions is 91.8%, which is only about 2% lower than that of using all electrodes. In addition, the SVM classifier outperforms other shallow classifiers used in this study, and most features can obtain the highest accuracy with SVM. In the future, we will explore effective feature fusion methods in emotion recognition.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

KY was mainly responsible for the research design, data collection, data analysis, and manuscript writing of this study. LT was mainly responsible for research design and data analysis. YZ was mainly responsible for experiment design, data analysis, and manuscript writing. RZ was mainly responsible for data analysis and classification algorithm. RL was mainly responsible for production of figures and document retrieval. YG was mainly responsible for manuscript writing and document retrieval. BY was mainly responsible for research design and manuscript writing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Major Projects of Technological Innovation 2030 of China under Grant: 2022ZD0208500.

Acknowledgments

The authors would like to thank all the subjects who participated in the experiment and the review, as well as MJEditor (www.mjeditor.com) for providing English editing services during the preparation of this manuscript.

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

Alkan, A., and Günay, M. (2012). Identification of EMG signals using discriminant analysis and SVM classifier. *Expert Syst. Appl.* 39, 44–47. doi: 10.1016/j.eswa.2011.06.043

An, Y., Hu, S., Duan, X., Zhao, L., Xie, C., and Zhao, Y. (2021). Electroencephalogram emotion recognition based on 3D feature fusion and convolutional autoencoder. *Front. Comput. Neurosci.* 15:743426. doi: 10.3389/fncom.2021.743426

Bai, L., Ma, H., Huang, Y. X., and Luo, Y. J. (2005). The development of native Chinese affective picture system-A pretest in 46 college students. *Chin. Ment. Health J.* 19, 719–722.

Bhadra, T., and Bandyopadhyay, S. (2021). Supervised feature selection using integration of densest subgraph finding with floating forward-backward search. *Inf. Sci.* 566, 1–18. doi: 10.1016/j.ins.2021.02.034

Blair, K. S., Smith, B. W., Mitchell, D., Morton, J., Vythilingam, M., Pessoa, L., et al. (2007). Modulation of emotion by cognition and cognition by emotion. *NeuroImage* 35, 430–440. doi: 10.1016/j.neuroimage.2006.11.048

Carretié, L., Iglesias, J., and GarcA, T. (1997). A study on the emotional-processing of visual stimuli through event-related potentials. *Brain Cogn.* 34, 207–217. doi: 10.1006/brcg.1997.0895

Chen, C., Wang, Q. H., Liu, X. H., and Liu, F. (2013). The study of relationship between negative emotion and cognitive executive function. *Prog. Mod. Biomed.* 13, 1149–1153.

Chen, J., Wang, L., Jia, X., and Zhang, P. (2019). "EEG-based emotion recognition using deep convolutional neural network," in *Proceedings of the 2019 IEEE 8th Data Driven Control and Learning Systems Conference (DDCLS)* (Dali: IEEE).

Chen, T., Ju, S., Ren, F., Fan, M., and Gu, Y. (2020). EEG emotion recognition model based on the LIBSVM classifier. *Measurement* 164, 1–13. doi: 10.1016/j.measurement.2020.108047

Ding, C., and Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *J. Bioinformatics Comput. Biol.* 3, 185–205. doi: 10.1142/S0219720005001004

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., and Darrell, T. (2013). "DeCAF: A deep convolutional activation feature for generic visual recognition," in *Proceedings of the 31st international conference on machine learning*. (Tianjin).

Duan, R. N., Wang, X. W., and Lu, B. L. (2012). "EEG-based emotion recognition in listening music by using support vector machine and linear dynamic system," in *Proceedings of the 19th international conference on neural information processing* (Berlin: Springer), 468–475. doi: 10.1007/978-3-642-34478-7_57

Gonuguntla, V., Veluvolu, K. C., and Kim, J. H. (2020). *Recognition of event-associated brain functional networks in EEG for brain network based applications*. Iowa, IA: IEEE, 271–274. doi: 10.1109/ISBI45749.2020.9098708

He, T., Zhang, Z., Zhang, H., Zhang, Z., and Li, M. (2019). "Bag of tricks for image classification with convolutional neural networks," in *Proceedings of the 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (Long Beach, CA). doi: 10.1109/CVPR.2019.00065

Hu, X., Chen, J., Wang, F., and Zhang, D. (2019). Ten challenges for EEG-based affective computing. *Brain Sci. Adv.* 5, 1–20. doi: 10.26599/BSA.2019.9050005

Huang, G. B., Zhu, Q. Y., and Siew, C. K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing* 70, 489–501. doi: 10.1016/j.neucom.2005.12.126

HyvRinen, A. (1999). The fixed-point algorithm and maximum likelihood estimation for independent component analysis. *Neural Process. Lett.* 10, 1–5. doi: 10.1023/A:1018647011077

Jackson-Koku, G. (2016). Beck depression inventory. *Occup. Med.* 18, 174–175. doi: 10.1093/occmed/kqv087

Jenke, R., Peer, A., and Buss, M. (2014). Feature extraction and selection for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* 5, 327–339. doi: 10.1109/TAFFC.2014.2339834

Jia, J., Yang, N., Zhang, C., Yue, A., Yang, J., and Zhu, D. (2013). Object-oriented feature selection of high spatial resolution images using an improved relief algorithm. *Math. Comput. Model.* 58, 619–626. doi: 10.1016/j.mcm.2011.10.045

Johnson, R. (1995). On the neural generators of the P300: Evidence from temporal lobectomy patients. *Electroencephalogr. Clin. Neurophysiol.* 44, 110–129.

Keller, J. M., Gray, M. R., and Givens, J. A. (2012). A fuzzy K-nearest neighbor algorithm. *IEEE Trans. Syst. Man Cybernet.* 15, 580–585. doi: 10.1109/TSMC.1985.6313426

Khare, S. K., and Bajaj, V. (2021). Time-frequency representation and convolutional neural network-based emotion recognition. *IEEE Trans. Neural Netw. Learn. Syst.* 32, 2901–2909. doi: 10.1109/TNNLS.2020.3008938

Koelstra, S. (2012). DEAP: A database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* 3, 18–31. doi: 10.1109/T-AFFC.2011.15

Lan, Z., Sourina, O., Wang, L., and Liu, Y. (2016). Real-time EEG-based emotion monitoring using stable features. *Vis. Comput.* 32, 347–358. doi: 10.1007/s00371-015-1183-y

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (2001). *International affective picture system (IAPS); instruction manual and effective ratings*. Technical Report A-4. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Li, F., Bei, C., He, L., Tao, Z., Fei, W., Yi, J., et al. (2016). The time-varying networks in P300: A task-evoked EEG study. *IEEE Trans. Neural Syst. Rehabil. Eng.* 24, 725–733. doi: 10.1109/TNSRE.2016.2523678

Li, M., Xu, H., and Liu, X. (2018). Emotion recognition from multichannel EEG signals using K-nearest neighbor classification. *Technol. Health Care* 26, 509–519. doi: 10.3233/THC-174836

Li, P. Y., Liu, H., Si, Y., Li, C., Li, F., Zhu, X., et al. (2019). EEG based emotion recognition by combining functional connectivity network and local activations. *IEEE Trans. Biomed. Eng.* 66, 2869–2881. doi: 10.1109/TBME.2019.2897651

Li, X., Li, X., Luo, Y. J., and Luo, Y. (2006). Differential influences of negative emotion on spatial and verbal working memory: Evidence from event-related potential and source current density analysis. *Neuroreport* 17, 1555–1559. doi: 10.1097/01.wnr.0000234744.50442.2b

Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R News* 2, 18–22.

Liu, S., Chen, L., Guo, D., Liu, X., Sheng, Y., Ke, Y., et al. (2018). Incorporation of multiple-days information to improve the generalization of eeg-based emotion recognition over time. *Front. Hum. Neurosci.* 12, 267–276. doi: 10.3389/fnhum.2018.00267

Lukosevicius, M., and Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.* 3, 127–149. doi: 10.1016/j.cosrev.2009.03.005

Meng, X., Yuan, J., and Hong, L. (2009). Automatic processing of valence differences in emotionally negative stimuli: Evidence from an ERP study. *Neurosci. Lett.* 464, 228–232. doi: 10.1016/j.neulet.2009.08.064

Morris, J. D. (1995). Observations: SAM: The selfAssessment Manikin; an efficient cross-cultural measurement of emotional response. *J. Adv. Res.* 35, 63–68.

Nie, D., Wang, X. W., Shi, L. C., and Lu, B. L. (2011). "EEG-based emotion recognition during watching movies," in *Proceedings of the 2011 5th international IEEE/EMBS conference on neural engineering (NER)*. (Cancun), 667–670. doi: 10.1109/NER.2011.5910636

Peng, Z., Wei, S., Tian, J., Qi, Z., and Bo, X. (2016). "Attention-based bidirectional long short-term memory networks for relation classification," in *Proceedings of the 54th annual meeting of the association for computational linguistics*, Vol. 2. (Berlin: Association for Computational Linguistics).

Qian, Y., Tan, J., and Chen, A. (2015). The stages of information processing and emotional stimuli processing. *J. Psychol. Sci.* 38, 801–806.

Rish, I. (2001). An empirical study of the naive bayes classifier. *J. Univers. Comput. Sci.* 1:127.

Saeedreza, E., Rezazadeh, S. A., Moussavi, S. Z., and Ali, S. (2009). A new pattern recognition technique in non destructive testing by the use of linear discriminate analysis. *Mod. Appl. Sci.* 3, 118–126. doi: 10.5539/mas.v3n5p118

Saha, P., and Fels, S. (2019). Hierarchical deep feature learning for decoding imagined speech from EEG. *Proc. AAAI Conf. Artif. Intell.* 33, 10019–10020. doi: 10.1609/aaai.v33i01.330110019

Santamaria-Vazquez, E., Martinez-Cagigal, V., Vaquerizo-Villar, F., and Hornero, R. (2020). EEG-Inception: A novel deep convolutional neural network for assistive ERP-based brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 2773–2782. doi: 10.1109/TNSRE.2020.3048106

Shuang, L., Chen, L., Guo, D., Liu, X., and Yue, S. (2018). Incorporation of multiple-days information to improve the generalization of EEG-based emotion recognition over time. *Front. Hum. Neurosci.* 12:267. doi: 10.3389/fnhum.2018.00267

Sun, Y., Kamel, M. S., Wong, A., and Yang, W. (2007). Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognit.* 40, 3358–3378. doi: 10.1016/j.patcog.2007.04.009

Van Straaten, E. C. W., and Stam, C. J. (2013). Structure out of chaos: Functional brain network analysis with EEG, MEG, and functional MRI. *Eur. Neuropsychopharmacol.* 23, 7–18. doi: 10.1016/j.euroneuro.2012.10.010

Wang, H., and Liu, L. (2020). Experimental investigation about effect of emotion state on people's thermal comfort. *Energy Build.* 211:109789. doi: 10.1016/j.enbuild.2020.109789

Wang, J., and Wang, M. (2021). Review of the emotional feature extraction and classification using EEG signals. *Cogn. Rob.* 1, 29–40. doi: 10.1016/j.cogr.2021.04.001

Wilson, K. A., Chambless, D. L., and Beurs, E. D. (1999). "Beck anxiety inventory," in *The use of psychological testing for treatment planning and outcomes assessment*, ed. M. E. Maruish (Hillsdale, NJ: Lawrence Erlbaum Associates Publishers), 971–992.

Wu, X., Zheng, W. L., Li, Z., and Lu, B. L. (2022). Investigating EEG-based functional connectivity patterns for multimodal emotion recognition. *J. Neural Eng.* 19:016012. doi: 10.1088/1741-2552/ac49a7

Yang, K., Tong, L., Shu, J., Zhuang, N., and Zeng, Y. (2020). High gamma band EEG closely related to emotion: Evidence from functional network. *Front. Hum. Neurosci.* 14:89. doi: 10.3389/fnhum.2020.00089

Yuan, J., Zhang, Q., Chen, A., Li, H., Wang, Q., Zhuang, Z., et al. (2007). Are we sensitive to valence differences in emotionally negative stimuli? Electrophysiological evidence from an ERP study. *Neuropsychologia* 45, 2764–2771. doi: 10.1016/j.neuropsychologia.2007.04.018

Zheng, W. L., and Lu, B. L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Ment. Dev.* 7, 162–175. doi: 10.1109/TAMD.2015.2431497

Zheng, W.-L., Liu, W., Lu, Y., Cichocki, A., and Lv, B.-L. (2019). EmotionMeter: A multimodal framework for recognizing human emotions. *IEEE Trans. Cybern.* 49, 1110–1122. doi: 10.1109/TCYB.2018.2797176

Zhong, P., Wang, D., and Miao, C. (2020). EEG-based emotion recognition using regularized graph neural networks. *IEEE Trans. Affect. Comput.* 1, 1–12. doi: 10.1109/TAFFC.2020.2994159

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). "Learning deep features for discriminative localization," in *Proceedings of the 2016 IEEE conference on computer vision and pattern recognition CVPR* (Las Vegas, NV: IEEE). doi: 10.1109/CVPR.2016.319

Zhuang, N., Ying, Z., Kai, Y., Chi, Z., Tong, L., and Yan, B. (2018). Investigating patterns for self-induced emotion recognition from EEG signals. *Sensors* 18, 841–862. doi: 10.3390/s18030841

frontiers | Frontiers in Neuroscience

TYPE Methods
PUBLISHED 21 October 2022
DOI 10.3389/fnins.2022.1045851

OPEN ACCESS

EDITED BY
Peng Xu,
University of Electronic Science and Technology of China, China

REVIEWED BY
Jing Jin,
East China University of Science and Technology, China
Yu Sun,
Zhejiang University, China

*CORRESPONDENCE
Badong Chen
chenbd@mail.xjtu.edu.cn

SPECIALTY SECTION
This article was submitted to Neuroprosthetics, a section of the journal Frontiers in Neuroscience

RECEIVED 16 September 2022
ACCEPTED 05 October 2022
PUBLISHED 21 October 2022

CITATION
Tang C, Gao T, Li Y and Chen B (2022) EEG channel selection based on sequential backward floating search for motor imagery classification. *Front. Neurosci.* 16:1045851. doi: 10.3389/fnins.2022.1045851

EEG channel selection based on sequential backward floating search for motor imagery classification

Chao Tang[1], Tianyi Gao[1], Yuanhao Li[2] and Badong Chen[1]*

[1]Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China, [2]Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Japan

Brain-computer interfaces (BCIs) based on motor imagery (MI) utilizing multi-channel electroencephalogram (EEG) data are commonly used to improve motor function of people with motor disabilities. EEG channel selection can enhance MI classification accuracy by selecting informative channels, accordingly reducing redundant information. The sequential backward floating search (SBFS) approach has been considered as one of the best feature selection methods. In this paper, SBFS is first implemented to select the optimal EEG channels in MI-BCI. Further, to reduce the time complexity of SBFS, the modified SBFS is proposed and applied to left and right hand MI tasks. In the modified SBFS, based on the map of EEG channels at the scalp, the symmetrical channels are selected as channel pairs and acceleration is thus realized by removing or adding multiple channels in each iteration. Extensive experiments were conducted on four public BCI datasets. Experimental results show that the SBFS achieves significantly higher classification accuracy ($p < 0.001$) than using all channels and conventional MI channels (i.e., C3, C4, and Cz). Moreover, the proposed method outperforms the state-of-the-art selection methods.

KEYWORDS

electroencephalogram (EEG), channel selection, sequential backward floating search (SBFS), motor imagery (MI), brain-computer interface (BCI)

1. Introduction

Brain-computer interface (BCI) refers to a complete system that processes signals from human brain to control different communication devices (Gao et al., 2021). With the advantages of non-invasiveness, portability, low cost, and high temporal resolution, electroencephalogram (EEG) is widely used in BCI systems (Padfield et al., 2019). Potential signals that are commonly used in EEG-based BCI system mainly include P300 evoked potentials (Picton, 1992; Li et al., 2010), steady state visually evoked potentials (SSVEP) (Wang et al., 2008; Zhang et al., 2018), and event-related desynchronization/synchronization (ERD/ERS) (Pfurtscheller and Da Silva, 1999; Pfurtscheller and Neuper, 2006).

Compared with stimuli-based BCI, the potential signals for motor imagery (MI) (Ang and Guan, 2016; Yang et al., 2020) can be easier acquired without external stimulus. MI tasks can bring out cortical rhythm amplitude suppression (ERD) and enhancement (ERS) over primary sensorimotor areas (Taniguchi et al., 2000; Neuper et al., 2005). According to ERD/ERS phenomenon, the corresponding imagery category can be determined. Therefore, it is of great significance to select brain area with active neural activities as signal sources to improve the quality of EEG signals. Excessive channels not only deteriorate the portability of BCI system, but also increase the difficulty of signal analysis (Handiru and Prasad, 2016). Selecting appropriate EEG channels for different subjects can improve the performance of MI-based BCI system.

According to the prior knowledge of neurology, C3, C4, and Cz electrodes and their surrounding channels contain the most information related to MI, for which these specific channels are commonly selected. Although the experience-dependent artificial channel selection is easy for preparation and implementation, it could be not sufficient for each subject. The popular channel selection schemes (Alotaiby et al., 2015) for MI can be mainly divided as embedded techniques, filtering techniques (Baig et al., 2020), wrapper techniques, etc. Embedded techniques integrate the channel selection processes with the model training process, such as recursive channel estimation with the training of support vector machine (SVM) (Lal et al., 2004; Schröder et al., 2005). Filtering techniques are usually based on EEG signal statistics such as common spatial pattern (CSP) filter coefficients (Tam et al., 2011) and specific criteria such as mutual information (Ang et al., 2012). Wrapper techniques typically adopt wrapper approachs with complete (Kamrunnahar et al., 2009), random (Wei and Wang, 2011) or sequential (Qiu et al., 2016) search strategies for subset channel selection (Arvaneh et al., 2010). In addition, neural network genetic method (Yang et al., 2012) and bispectrum-based method (Jin et al., 2020) were investigated for EEG channel selection. Recently, neurophysiological approaches based on correlation (Jin et al., 2019) and Granger causality (Varsehi and Firoozabadi, 2021) have also been used in MI channel selection. However, the EEG channel selection methods of existing studies have either shown unsatisfactory performance or can only be used for specific types of data (Varsehi and Firoozabadi, 2021).

Sequential backward floating search (SBFS) is a well-known feature selection method which has been used to process various physiological signals (Tork et al., 2013; Karnaukh et al., 2018; Ahirwal, 2021) and to perform body state assessments (Dreißig et al., 2020). In this paper, SBFS is utilized in EEG channel selection for MI-based BCI. The main contributions of this paper are as follows:

1) To the best of our knowledge, this is the first time SBFS has been utilized for EEG channel selection.

2) The modified SBFS was proposed and applied to left and right hand MI tasks to reduce the time complexity of SBFS.

3) Extensive experiments were conducted on four datasets to confirm the effectiveness of the proposed method.

The remainder of this paper is detailed as follows. Section 2 describes the data used in this paper and the proposed methods. Section 3 presents the results. The discussion is provided in Section 4, and finally we conclude the paper in Section 5.

2. Materials and methods

2.1. Datasets

In this work, four common public datasets were used to evaluate the proposed methods. All EEG data were collected from the subjects' brain through acquisition equipments rather than artificially generated.

1) BCI Competition IV-dataset 1: This dataset recorded 59 channels of EEG signals from 7 healthy subjects (Tangermann et al., 2012). We only used the data collected from subject a, b, f, and g, since the other data were artificially generated. Each subject participated in two classes (from the three classes left hand, right hand, and foot) of MI tasks. Each data included two runs, where each run contained 100 trials. In these two runs, arrows pointing left, right or down were displayed on the screen for visual cues. Cues were shown for a period of 4 s, during which the subjects were asked to perform the MI task. After and before the task, there were 2 s of blank and 2 s of display with a fixation cross presented in the center of the screen. Namely, the fixation cross was superimposed on the cues for 6 s. Each trial for the EEG data acquisition is illustrated in Figure 1A. The EEG signals were downsampled to 100 Hz.

2) BCI Competition III-dataset IIIa: The dataset was recorded from 3 subjects (k3, k6, and l1) in 60 channels with a sampling rate of 250 Hz (Blankertz et al., 2006). The subjects performed imagery left hand, right hand, foot or tongue movements according to a cue of random order. When a trial began, the first 2 s were quiet black-screen. Then an acoustic stimulus and a cross "+" were presented at t = 2 s. From t = 3 s an arrow pointing to left, right, up or down was shown for 1 s. In the meantime, the subjects imagined the movement corresponding to the arrow until t = 7 s. Each trial for the EEG data acquisition is shown in Figure 1B. The number of trials per class was 90 or 60 for different subjects. We only use the left and right hand MI trials in this study.

3) BCI Competition III-dataset IVa: The dataset was recorded from 5 healthy subjects (aa, al, av, aw, and ay) (Blankertz et al., 2006). The subjects performed one of the left hand, right hand and right foot MI within 3.5 s of the occurrence of the visual cues. Target cues were presented at random intervals (1.75–2.25 s), during which subjects could relax. Each trial for the EEG data acquisition is presented in Figure 1C. Each subject participated in 280 trials. The EEG signals were recorded with 118 channels and were downsampled at 100 Hz.

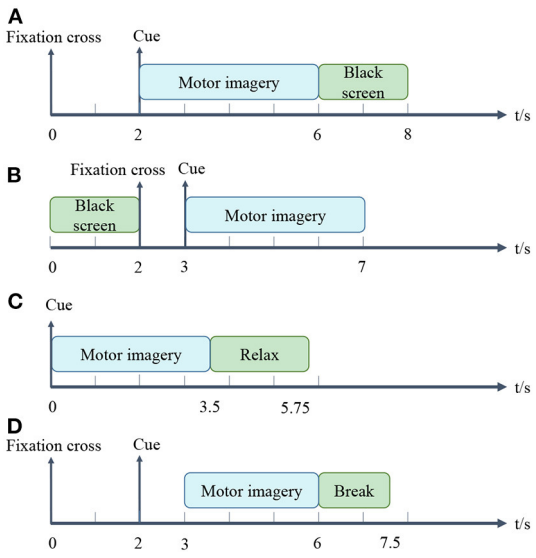


FIGURE 1
Time interval of one trial. **(A–D)** is for dataset 1), 2), 3) and 4), respectively.

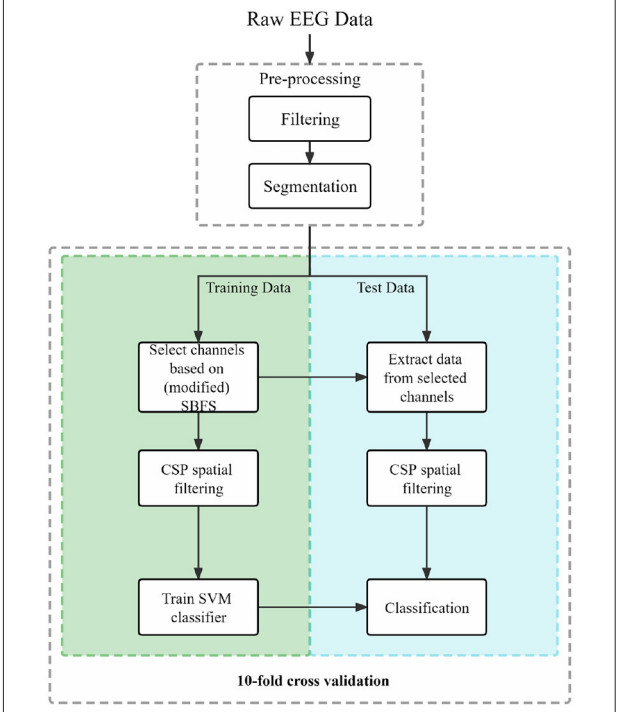


FIGURE 2
Block diagram of the proposed framework.

4) BCI competition IV-dataset 2a: The dataset recorded the EEG data of 9 subjects (A01–A09) who participated in the 4-class (left hand, right hand, both feet, and tongue) MI experiments (Tangermann et al., 2012). Raw data were collected at 22 channels and 250 Hz sampling rate. Each subject's data were recorded in 2 sessions, each session consisted of 6 runs, and each run contained 48 trials. i.e., each session was composed of 288 trials in 4 classes, and each class contained 72 trials. We only classified the trials of left hand and right hand in Session 1. The timeline for a trial is about 7.5 s, as detailed in Figure 1D. At the beginning (t = 0 s), a cross "+" appeared on the black screen. After 2 s (t = 2 s), an arrow pointing either to the left, right, down or up appeared and stayed on the screen for 1.25 s. The subjects performed the desired MI tasks until t = 6 s. After a short break, the screen went black again.

2.2. Data preprocessing

The acquired EEG data were refined in the preliminary analysis prior to channel selection, feature extraction, and classification. A portion of Figure 2 shows the preprocessing procedure.

1) Filtering: A third-order butterworth filter was applied to raw EEG data in the filtering part. The EEG data from each trial were filtered between 8 and 30 Hz.

2) Segmentation: The filtered EEG data were segmented by extracting data segments related to event types. MI events mainly consist of two intervals: MI and other states (rest or black screen). For dataset 1), 2), 3) and 4), we used fixed time windows of 2–6 s, 3–6 s, 0–3 s, and 3–6 s, respectively. More details exhibit in Figure 1.

2.3. Channel selection

2.3.1. SBFS-based channel selection

The purpose of channel selection is to identify important channels and remove redundant and irrelevant channels. The SBFS starts with a complete set, which is based on a top-down approach (Pudil et al., 1994). We investigated the SBFS method for EEG channel selection in MI classification. In this study, Y denotes the entire channel set. X_k denotes the subset of channels containing k channels. $J(X_k)$ denotes the classification performance of a subset X_k. The SBFS algorithm for channel selection is given in Algorithm 1.

The advantage of applying the update strategy to SBFS in EEG channel selection is the possibility to increase the value of optimal accuracy or decrease the number of channels of optimal accuracy. It is described as follows: the SBFS algorithm pursues the maximum accuracy under the current number of channels, and the intermediate (Inclusion) process of the later channel selection may result in the increase of the accuracy of the previous number of channels. Our update strategy is to replace with the maximum accuracy each time.

Input: the set of all channels, $Y = \{y_1, y_2, ..., y_d\}$

- The SBFS algorithm takes the entire channel as input.

Output: $X_k = \{x_j | j = 1, 2, ..., k; x_j \in Y\}$, where $k = (0, 1, 2, ..., d)$

- SBFS returns a subset of channels; the number of selected channels k, where $k < d$.

Initialization: $X_k = Y, k = d$

- We initialize the algorithm with the given channel set such that $k = d$.

Step 1(Exclusion):

$x^- = arg \ \ max \ \ J(X_k - x), where \ \ x \in X_k$

$X_{k-1} = X_k - x^-$

$k = k - 1$

Go to Step 2

- In step 1, we remove a channel x^- from the channel subset X_k.
- x^- is the channel that maximizes our criterion function upon removal, that is, the channel which is associated with the optimal classification performance if it is removed from X_k.

Step 2(Conditional Inclusion):

$x^+ = arg \ \ max \ \ J(X_k + x), where \ \ x \in Y - X_k$

$if \ \ J(X_k + x) > J(X_{k+1})$:

$X_{k+1} = X_k + x^+$

$k = k + 1$

Go to Step 1

- In step 2, we search for channels that would improve the classification performance when added back to the channel subset. If such channels exist, we add the channel x^+ that maximizes the performance improvement. If $k = n$ or an improvement cannot be made (i.e., such channel x^+ cannot be found), go back to step 1; else, repeat the current step.

Termination: $k = 2$

- The channel subset of size k contains the desired number of channels 2.

Algorithm 1. EEG channel selection using SBFS.

2.3.2. Reducing time complexity: Modified SBFS

Since SBFS is a search method, it makes sense to speed up the search process without compromising accuracy. Considering that mu (8–13 Hz bands) and beta (14–30 Hz bands) ERD/ERS phenomenon are elicited during imagined hand movements (Ramoser et al., 2000), depending on the location of the channels in the cerebral cortex, left-right symmetrical channels can be treated as a channel pair. As is shown in Figure 3, channels of the same color are considered as a channel pair. For example, red C3 and C4, blue CP1 and CP2, and green FC1 and FC2 are channel pairs, respectively. Both of them are left-right symmetrical with respect to the straight lines of CZ and CPZ. Thus, the whole set contains fewer channel pairs and the modified SBFS can remove or add multiple channels at a time. The main differences between SBFS and modified SBFS methods are shown in Figure 4. One can observe that the modified SBFS contains far fewer channel pairs than before. The time spent on searching process can be greatly reduced.

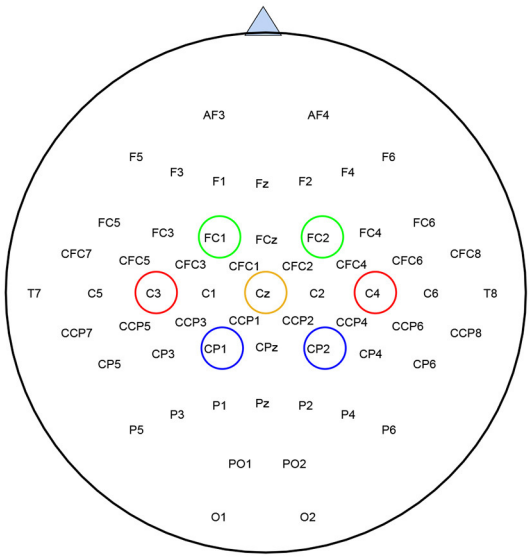


FIGURE 3
Location of EEG electrodes used for data acquisition, taking BCI competition IV-dataset 1 as an example. Channels of the same color are treated as a channel pair for selection.

2.4. Feature extraction

CSP (Ramoser et al., 2000) is an efficient feature extraction algorithm for binary classification tasks, which has been extensively used in MI-based BCI (Dong et al., 2017; Chen B. et al., 2018). It finds a spatial filter to maximize the differences in variance between two classes of multi-channel EEG data. Let $\mathbf{C}_a$ and $\mathbf{C}_b$ be the normalized covariance matrices of the two classes of EEG signals which are averaged over trials. The composite spatial covariance matrix is $\mathbf{C}_c = \mathbf{C}_a + \mathbf{C}_b$. Decomposing $\mathbf{C}_c$, we can obtain $\mathbf{C}_c = \mathbf{U}\lambda\mathbf{U}^{\mathrm{T}}$, where λ is the eigenvalue and $\mathbf{U}$ is the eigenvector. And the whitening transformation is

$$\mathbf{P} = \sqrt{\lambda^{-1}}\mathbf{U}^{\mathrm{T}} \tag{1}$$

Then the covariance matrices $\mathbf{C}_a$ and $\mathbf{C}_b$ can be transformed into:

$$\mathbf{S}_a = \mathbf{P}\mathbf{C}_a\mathbf{P}^{\mathrm{T}} \tag{2}$$

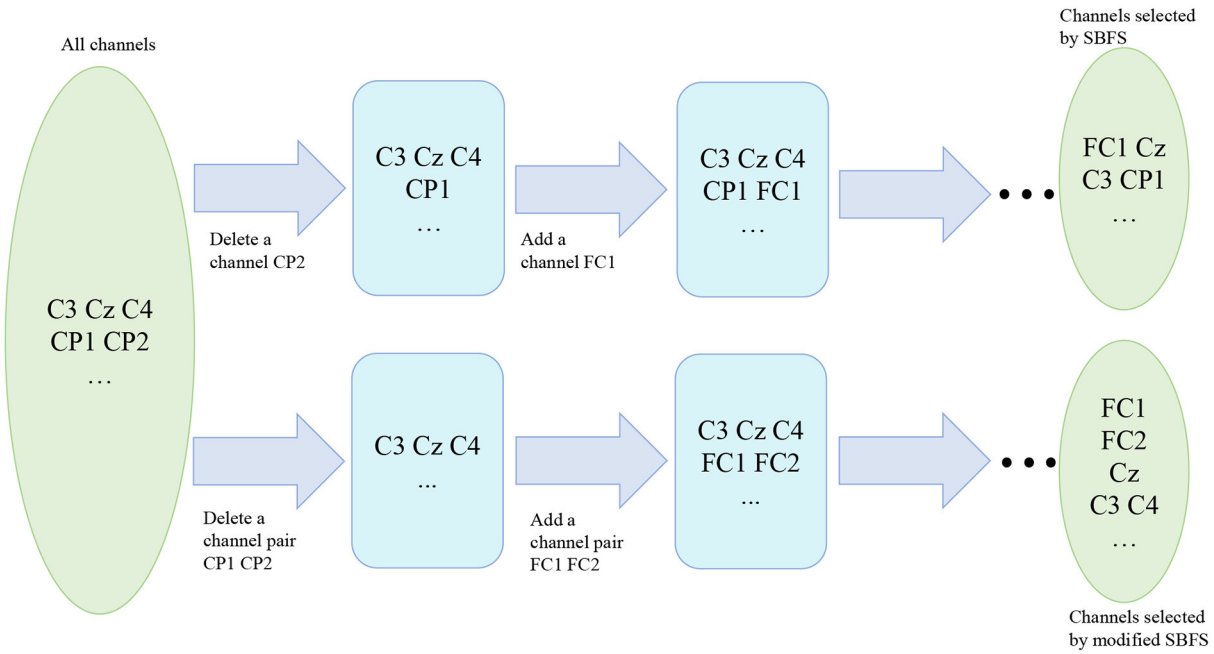


FIGURE 4
Search chart of the SBFS and the modified SBFS.

$$\mathbf{S}_b = \mathbf{P}\mathbf{C}_b\mathbf{P}^{\mathrm{T}} \tag{3}$$

It can be seen that $\mathbf{S}_a$ and $\mathbf{S}_b$ have the same eigenvector. Then we have $\mathbf{S}_a + \mathbf{S}_b = \mathbf{I}$. Any orthonormal matrix $\mathbf{V}$ satisfies the following

$$\mathbf{V}^{\mathrm{T}}(\mathbf{S}_a + \mathbf{S}_b)\mathbf{V} = \mathbf{I} \tag{4}$$

Using the orthonormal matrix $\mathbf{V}$, $\mathbf{S}_a$ and $\mathbf{S}_b$ can be decomposed as follows:

$$\mathbf{S}_a = \mathbf{V}\Lambda_a\mathbf{V}^{\mathrm{T}} \tag{5}$$

$$\mathbf{S}_b = \mathbf{V}\Lambda_b\mathbf{V}^{\mathrm{T}} \tag{6}$$

and $\Lambda_a + \Lambda_b = \mathbf{I}$. The projection matrix is

$$\mathbf{W} = \mathbf{P}^{\mathrm{T}}\mathbf{V} \tag{7}$$

$\mathbf{W}$ is called the CSP weight matrix. The optimal features can be obtained in the least square case. Finally, the vector of the features is expressed as:

$$f = log(var(\mathbf{W}\mathbf{x}(t))) \tag{8}$$

where, $\mathbf{x}(t)$ is EEG data.

2.5. Classification

SVM theory was proposed by Vapnik (1999). The core idea of SVM is to separate the data from the two classes by finding a hyperplane with the largest possible margin. As one of the most commonly used BCI-based MI classifiers, SVM (Subasi and Gursoy, 2010; Qin et al., 2019) can effectively solve the classification problem of two classes of EEG data. In this study, we used an SVM with a radial basis function kernel to classify MI tasks after feature extraction. The separation of training and test data is realized by using 10-fold cross validation in the classification part.

2.6. Framework overview

Firstly, the raw MI EEG data of each subject were preprocessed. The SBFS and the modified SBFS method were applied to the training data to obtain the selected channels. The CSP spatial domain filter was applied for training data to acquire weight matrices. Finally, the SVM classifier was trained and the classification performance with 10-fold cross validation was obtained. The block diagram of our proposed framework is shown in Figure 2.

TABLE 1 Comparison of classification accuracy (%) with different methods on datasets 1), 2), 3).

Subject	Methods		
	All channels	C3C4Cz	SBFS
a	52.0	43.5	**74.5**
b	44.0	51.0	**64.5**
f	49.0	46.5	**63.5**
g	50.0	63.0	**68.0**
Mean ± std	48.8 ± 3.4	49.3 ± 5.5	**67.6** ± 5.0
k3	45.6	53.3	**78.9**
k6	46.7	41.7	**76.7**
l1	50.0	46.7	**80.0**
Mean ± std	47.4 ± 2.3	47.2 ± 5.8	**78.5** ± 1.7
aa	62.1	58.8	**84.4**
al	62.3	74.1	**95.0**
av	47.5	60.0	**75.0**
aw	75.6	68.0	**100**
ay	67.5	45.0	**100**
Mean ± std	63.0 ± 10.3	61.2 ± 11.0	**90.9** ± 10.9
P-value	<0.0001	<0.0001	–

"–" Denotes the missing values. The optimal classification accuracies in each row are in bold.

TABLE 2 Comparison of classification accuracy (%) with different methods on 14 subjects which participate in left and right hand MI tasks.

Subject	Methods		
	All channels	SBFS	Modified SBFS
b	44.0	**64.5**	62.5
g	50.0	68.0	**69.0**
k3	45.6	**78.9**	74.4
k6	46.7	**76.7**	71.7
l1	50.0	80.0	**81.7**
A01	45.7	**70.7**	67.9
A02	51.4	**66.4**	65.7
A03	55.0	77.9	**79.3**
A04	46.4	62.9	**63.6**
A05	51.4	65.0	**67.9**
A06	49.3	**63.6**	61.4
A07	57.9	**69.3**	65.6
A08	60.7	**87.1**	**87.1**
A09	48.6	**71.4**	70.7
Mean ± std	50.2 ± 4.8	**71.6** ± 7.4	70.6 ± 7.6
P-value	<0.0001	0.1522	–

"–" Denotes the missing values. The optimal classification accuracies in each row are in bold.

3. Results

3.1. Classification accuracy and significance

The classification accuracies of all subjects from datasets 1), 2), 3) using different methods are shown in Table 1. The classification accuracies of 14 subjects which participated left and right hand MI using different methods are shown in Table 2. The optimal classification accuracy for each subject and mean are in bold. The last row of Tables 1, 2 gives the paired *t*-test results of the SBFS or the modified SBFS with the current column method. The C3C4Cz method indicates that only EEG data from these 3 channels are used in the classification of MI. From Table 1, for each single subject, the highest classification accuracy was obtained with SBFS. In particular, subjects aw and ay achieved 100% classification accuracy. Compared with all channels, the average performance improvement of the SBFS method in datasets 1), 2), 3) was 18.8, 31.1, and 27.9%, respectively. Meanwhile, the SBFS method improved by 18.3, 31.3, and 29.7%, respectively, compared with the C3C4Cz method. From Table 2, the average accuracy of the SBFS and the modified SBFS is improved by 21.4 and 20.4%, respectively. It is shown that the accuracy of the SBFS is significantly better than all channels and conventional MI channels ($p < 0.0001$). There is no significant difference ($p = 0.1522$) between the SBFS and the modified SBFS.

3.2. Number of selected channels

The number of selected channels with optimal classification accuracy is shown in Table 3. From Table 3, the number of channels selected by the SBFS shows a substantial decrease compared to all channels. In terms of averages, this is specifically shown as 23 vs. 59, 12 vs. 60, 25 vs. 118, and 10 vs. 22. Overall, the number of selected channels ranges from one-fifth to one-half of the total number of usable channels. The number of channels selected by the modified SBFS method is similar to that of the SBFS.

3.3. Computation time

In order to compare the computation time between the SBFS method and the modified SBFS method, the results of 14 subjects who participated in a left and right hand MI task from datasets 1), 2), 4) were used. The two algorithms were implemented and tested using MATLAB 2019[1] configured on Windows 10 professional operating system and the experiments

1 MATLAB is a commercial mathematical software from MathWorks, Inc. in the U.S. 2019a is the version number.

TABLE 3 Comparison of the number of channels selected at the highest accuracy.

Subject	Methods		
	All channels	SBFS	Modified SBFS
a	59	10	–
b	59	45	14
f	59	4	–
g	59	33	17
Average	59	23	–
k3	60	10	14
k6	60	19	20
l1	60	6	20
Average	60	12*	18
aa	118	40	–
al	118	31	–
av	118	37	–
aw	118	6	–
ay	118	13	–
Average	118	25*	–
A01	22	7	9
A02	22	4	8
A03	22	13	17
A04	22	3	14
A05	22	7	13
A06	22	14	6
A07	22	19	19
A08	22	8	10
A09	22	15	9
Average	22	10	12*

"*" Denotes that the original number is not an integer. "–" Denotes the missing values.

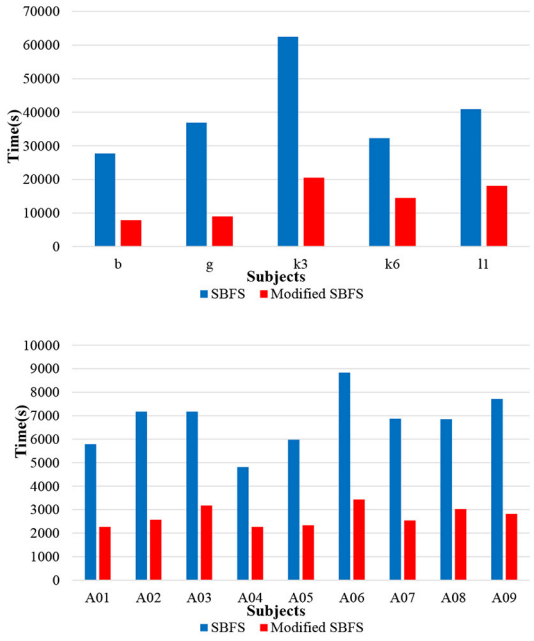


FIGURE 5
Computation time comparisons between SBFS and modified SBFS.

were performed on an Intel (R) Core (TM) i5-8265U CPU @ 1.60GHz, 8.00 GB RAM computer. As can be seen from Figure 5, the modified SBFS method for channel selection is faster than the SBFS method. It is precisely because more than one channel were added or deleted each time that the number of iterations was reduced, which greatly saved the time.

The computation time of the method is affected by several factors, such as the number of all channels, the number of trials, the length of trials, software and hardware configuration etc. For example, using a parallel for loop (parfor) in Matlab to speed up the algorithm, the iterations of the parfor loop can run in parallel on multiple cores of the target hardware (our computer has 4 cores), and speed of the algorithm obtained by testing is shown to be about 4 times faster. During the initial sessions (also called the training sessions) in BCI experiments, the desired parameters are adjusted offline according to the signals collected from different subjects. Thus, the running speed of the channel selection part is not a concern, even for the SBFS method. In short, users have the flexibility to choose the SBFS or the modified SBFS methods according to specific practical situations.

3.4. Comparison with other selection methods

We compared SBFS-based EEG channel selection method with other algorithms in this field. For fairness of comparison, the data preprocessing, feature extraction and classifier were used identically.

CSP-rank (Tam et al., 2011) is a channel selection method in MI-based BCI using CSP. The method is based on the sorting of CSP filters. To be specific, we first rank the absolute values of the filter coefficients in each filter respectively, and then take the electrodes with the next largest coefficients from the two spatial filters in turn.

Improved sequential floating forward selection (ISFFS) (Qiu et al., 2016) combines the practical distribution of channels and an intelligent selection algorithm to select EEG channels.

TABLE 4 Comparison of classification accuracy (%) with different channel selection methods on BCI competition IV-dataset 1.

Subject	Methods			
	CSP-rank	ISFFS	CCS	SBFS
a	62.0	64.5	57.5	**74.5**
b	57.5	61.5	58.0	**64.5**
f	53.5	63.5	55.0	**63.5**
g	61.5	63.5	63.0	**68.0**
Mean	58.6	63.3	58.4	**67.6**

The optimal classification accuracies in each row are in bold.

TABLE 5 Comparison of the number of selected channels with different channel selection methods on BCI competition IV-dataset 1.

Subject	Methods			
	CSP-rank	ISFFS	CCS	SBFS
a	53	41	42	10
b	27	11	21	45
f	57	8	38	4
g	7	24	25	33
Average	36	21	32*	23

"*" Denotes that the original number is not an integer.

Correlation-based channel selection (CCS) (Jin et al., 2019) assumes that there is a high correlation between task-related channels, then the relevant channels are selected.

Tables 4, 5 presents the classification accuracy and the number of channels at optimal accuracy for different methods, respectively. The optimal classification accuracy for each subject and mean are in bold. The SBFS method achieved the best classification accuracy for both single subjects and mean values. The SBFS and the ISFFS methods are similar in the number of selected channels, less than the CSP-rank and the CCS method.

4. Discussion

4.1. Maps of the selected channels

We used MATLAB 2019 (see text footnote 1) with the EEGLAB toolbox (Delorme and Makeig, 2004) to plot topographic maps of subjects g, k3, and av from each of the datasets 1), 2), 3), as shown in Figure 6. The map of the channels selected by SBFS is shown in Figure 7.

The location of channels was compared with topographic maps. On the whole, the channels selected by SBFS were consistent with the corresponding ERD phenomena for all channels. Channel C3, C4, and Cz or their surrounding channels located in the motor area of the brain were selected multiple times. For dataset 1), Channel C4 and Cz were selected twice. Also, some channels (C2, C5, CFC3, and CCP2) around C3 and C4 were selected multiple times. Figure 6 shows that subject g in the right hand MI task, the ERD phenomena occurred

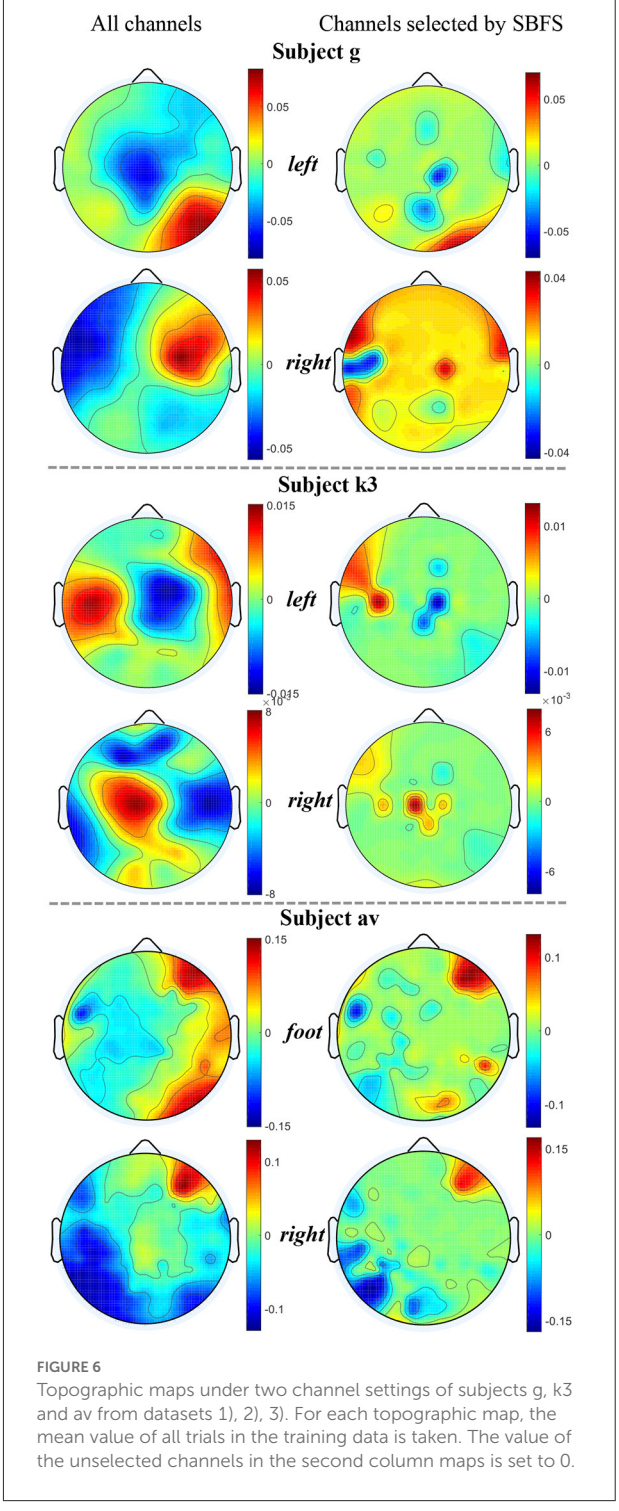


FIGURE 6
Topographic maps under two channel settings of subjects g, k3 and av from datasets 1), 2), 3). For each topographic map, the mean value of all trials in the training data is taken. The value of the unselected channels in the second column maps is set to 0.

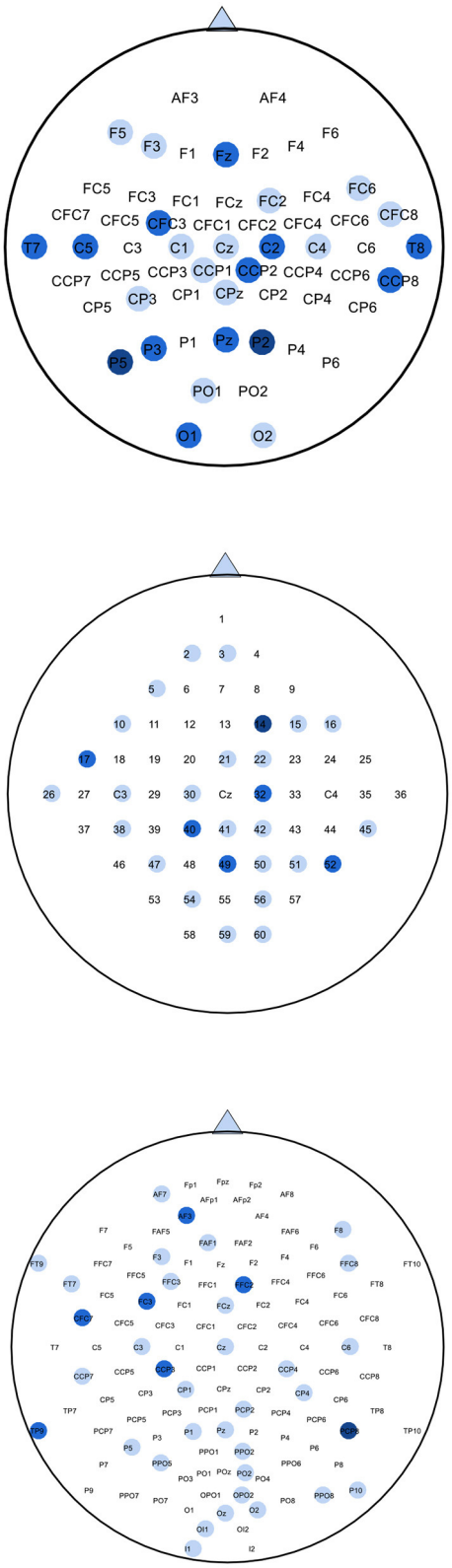


FIGURE 7
Maps of selected channels by SBFS for datasets 1), 2), 3). The blue circles represent selected channels. Darker colors indicate more selections. For dataset 1), channels selected more than twice from four subjects are colored in the top map. For dataset 2), channels selected more than once from three subjects are colored in the middle map. For dataset 3), channels selected more than twice from five subjects are colored in the bottom map.

mainly in the left cerebral cortex. For dataset 2), Channel C3 was selected. Channel 32 and 40, located around channel Cz, were also selected twice. In Figure 6, the ERD phenomena of subject k3 mainly occurred in the right cerebral cortex during left hand MI task. For dataset 3), all 5 subjects performed foot and right hand MI tasks. Channel C3 and Cz were selected twice. They were surrounded by channel CCP3 and CCP4 which were also selected multiple times. As shown in Figure 6, the ERD phenomena appears in the left cerebral cortex when subject av was performing a right hand MI task.

The SBFS method selected channels successively from bottom (i.e., serial number is larger) to top (i.e., serial number is smaller). This may lead to channels which are not related to MI being selected as well. The irregular channel positions resulted from an evaluation criterion by using cross-validation accuracy method. Many selected channels were located in the posterior part of the cerebral cortex. In the case of equal classification accuracy, the channels with larger ordinal numbers, i.e., the backward position, were retained, and the channels with smaller ordinal numbers, i.e., the forward position, were removed.

The topographic maps plotted using the EEGLAB toolbox under all channels and channels selected by the modified SBFS for subject k6 are shown in Figure 8, respectively. From Figure 8, it can also be seen that the topographic map of the channels selected by the modified SBFS was basically consistent with that of all channels. Specifically, clear ERD phenomenon can be observed in the left hand and right hand topographic maps under channels selected by modified SBFS.

4.2. Parameter sensitivity analysis

Figure 9 plots the variation of classification accuracy with the number of selected channels by SBFS. With the increase of the number of selected channels, the overall trend of classification accuracy increases, then decreases. This may be due to the fact that there are few channels containing available information at the beginning and the initial accuracy is low. As the number of channels increases, the useful information keeps increasing and the accuracy is improved. With the sustainable increase of the number of channels, the number of redundant information channels increases as well, leading to the decrease of accuracy. Specifically, the change in accuracy with the number of channels is different for each subject. For example, the

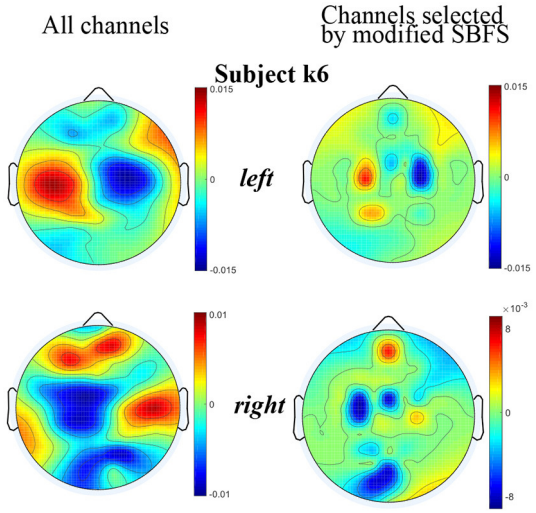


FIGURE 8
Topographic maps under two channel settings of subjects k6. For each topographic map, the mean value of all trials in the training data is taken. The value of the unselected channels in the maps is set to 0.

classification accuracies of subjects a, b, k6 and av have a slight decrease initially.

4.3. Future works

The SBFS method obviously outperforms other competing channel selection algorithms. In the following study, the proposed framework will be tested on datasets containing more subjects, such as the Physionet dataset (Goldberger et al., 2000), to strengthen sufficient persuasiveness. In addition, the SBFS can be combined with other features to further improve the classification performance in MI-based BCI. For example, for small sample data, it might be combined with regularized CSP feature (Lu et al., 2010).

In this paper, we studied the SBFS method for channel selection of MI tasks. The idea of a feature selection algorithm combined with task-related effective features can be extended to other tasks, e.g., for the SSVEP task, a combination of the SBFS and canonical correlation analysis (Lin et al., 2006) can be used, and for the rapid serial visual presentation task (Xu et al., 2021), a combination of the SBFS and hierarchical discriminant component analysis (Parra et al., 2007) can be used. Moreover, the proposed symmetrical strategy which is with respect to the optimization time can also be extended to other tasks, such as EEG-based emotion recognition (Chen T. et al., 2018; Gao et al., 2020; Tang et al., 2022).

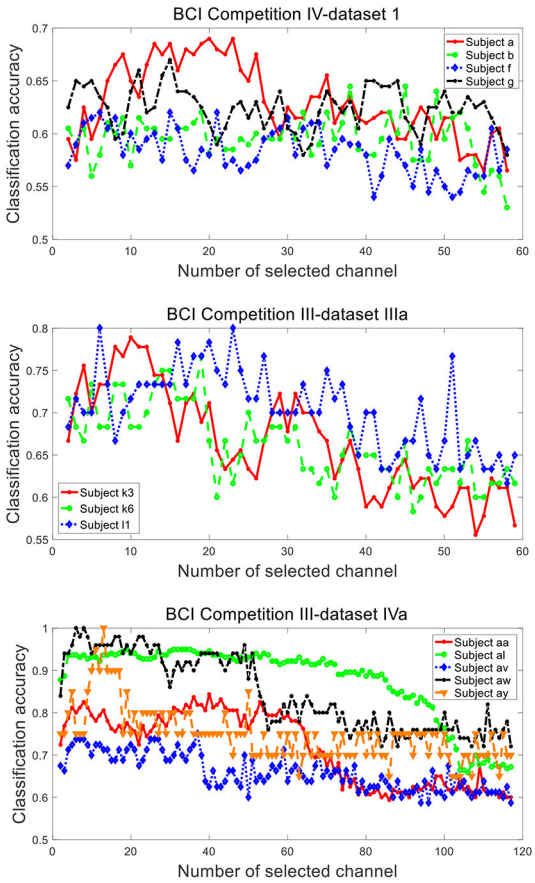


FIGURE 9
Classification accuracy of subjects from datasets 1), 2), 3) under different number of selected channels by SBFS.

Note that, ERD and ERS phenomenon are found not only in EEG but also in magnetoencephalography (MEG) recordings. As another non-invasive physiological signal, MEG-based BCI often involves more sensors. The existing MEG instrument based on superconducting quantum interference device technology is typically composed of 275 (gradiometer) or 306 (204 gradiometer and 120 magnetometer) sensors. Although a large number of sensors can provide higher spatial-temporal resolution for evaluating brain activity patterns, not all sensors can significantly improve classification accuracy. In addition, a larger number of channels implies a greater computation complexity. Recently, Roy et al. (2019, 2020) assessd the effect of channel selection using intelligent algorithms on MEG decoding of MI for the first time. Therefore, the application of the proposed method to MEG data can be explored in the future.

5. Conclusion

In this paper, the SBFS method is first applied to EEG channel selection, combining CSP features and an SVM classifier to form a new decoding framework. In order to reduce the time complexity of SBFS, the modified SBFS method is proposed, in which symmetrical channel pairs are removed or added in each iteration depending on the location of EEG channels at the scalp. Experimental results show that the proposed method can significantly improve the classification accuracy while reducing the number of EEG channels. The study provides a new approach to improve the reliability of future BCI systems.

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/s.

Author contributions

CT: data collection, methodology, and writing—original draft. TG: data collation and analysis. YL: writing—reviewing. BC: reviewing and supervision. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China with Grant Nos. (U21A20485 and 61976175).

Acknowledgments

We thank Ms. Mei Feng for her linguistic assistance during the preparation of this manuscript.

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

References

Ahirwal, M. K. (2021). "Analysis and identification of eeg features for mental stress," in *Evolution in Computational Intelligence*, eds V. Bhateja (Singapore: Springer), 201–209.

Alotaiby, T., Abd El-Samie, F. E., Alshebeili, S. A., and Ahmad, I. (2015). A review of channel selection algorithms for eeg signal processing. *EURASIP J. Adv. Signal Process* 2015, 1–21. doi: 10.1186/s13634-015-0251-9

Ang, K. K., Chin, Z. Y., Zhang, H., and Guan, C. (2012). Mutual information-based selection of optimal spatial-temporal patterns for single-trial eeg-based bcis. *Pattern Recognit*. 45, 2137–2144. doi: 10.1016/j.patcog.2011.04.018

Ang, K. K., and Guan, C. (2016). Eeg-based strategies to detect motor imagery for control and rehabilitation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 392–401. doi: 10.1109/TNSRE.2016.2646763

Arvaneh, M., Guan, C., Ang, K. K., and Quek, H. C. (2010). "EEG channel selection using decision tree in brain-computer interface," in *Proceedings of the Second APSIPA Annual Summit and Conference* (Singapore), 225–230.

Baig, M. Z., Aslam, N., and Shum, H. P. (2020). Filtering techniques for channel selection in motor imagery eeg applications: a survey. *Artif. Intell. Rev*. 53, 1207–1232. doi: 10.1007/s10462-019-09694-8

Blankertz, B., Muller, K.-R., Krusienski, D. J., Schalk, G., Wolpaw, J. R., Schlogl, A., et al. (2006). The bci competition III: validating alternative approaches to actual bci problems. *IEEE Trans. Neural Syst. Rehabil. Eng.* 14, 153–159. doi: 10.1109/TNSRE.2006.875642

Chen, B., Li, Y., Dong, J., Lu, N., and Qin, J. (2018). Common spatial patterns based on the quantized minimum error entropy criterion. *IEEE Trans. Syst. Man Cybern. Syst*. 50, 4557–4568. doi: 10.1109/TSMC.2018.2855106

Chen, T., Ju, S., Yuan, X., Elhoseny, M., Ren, F., Fan, M., et al. (2018). Emotion recognition using empirical mode decomposition and approximation entropy. *Comput. Electr. Eng.* 72, 383–392. doi: 10.1016/j.compeleceng.2018.09.022

Delorme, A., and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Dong, J., Chen, B., Lu, N., Wang, H., and Zheng, N. (2017). "Correntropy induced metric based common spatial patterns," in *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)* (Tokyo: IEEE), 1–6.

Dreißig, M., Baccour, M. H., Schäck, T., and Kasneci, E. (2020). "Driver drowsiness classification based on eye blink and head movement features using the k-NN algorithm," in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)* (Canberra, ACT: IEEE), 889–896.

Gao, X., Wang, Y., Chen, X., and Gao, S. (2021). Interface, interaction, and intelligence in generalized brain-computer interfaces. *Trends Cogn. Sci*. 25, 671–684. doi: 10.1016/j.tics.2021.04.003

Gao, Y., Wang, X., Potter, T., Zhang, J., and Zhang, Y. (2020). Single-trial eeg emotion recognition using granger causality/transfer entropy analysis. *J. Neurosci. Methods* 346, 108904. doi: 10.1016/j.jneumeth.2020.108904

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

Handiru, V. S., and Prasad, V. A. (2016). Optimized bi-objective eeg channel selection and cross-subject generalization with brain-computer interfaces. *IEEE Trans. Hum. Mach. Syst*. 46, 777–786. doi: 10.1109/THMS.2016.2573827

Jin, J., Liu, C., Daly, I., Miao, Y., Li, S., Wang, X., et al. (2020). Bispectrum-based channel selection for motor imagery based brain-computer interfacing. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 2153–2163. doi: 10.1109/TNSRE.2020.3020975

Jin, J., Miao, Y., Daly, I., Zuo, C., Hu, D., and Cichocki, A. (2019). Correlation-based channel selection and regularized feature optimization for mi-based bci. *Neural Netw.* 118, 262–270. doi: 10.1016/j.neunet.2019.07.008

Kamrunnahar, M., Dias, N., and Schiff, S. (2009). Optimization of electrode channels in brain computer interfaces. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2009, 6477–64780. doi: 10.1109/IEMBS.2009.5333585

Karnaukh, O., Karplyuk, Y., and Nikitiuk, N. (2018). "Evaluation of machine learning techniques for ECG t-wave alternans," in *2018 IEEE 38th International Conference on Electronics and Nanotechnology (ELNANO)* (Kyiv: IEEE), 346–350.

Lal, T. N., Schroder, M., Hinterberger, T., Weston, J., Bogdan, M., Birbaumer, N., et al. (2004). Support vector channel selection in bci. *IEEE Trans. Biomed. Eng.* 51, 1003–1010. doi: 10.1109/TBME.2004.827827

Li, Y., Long, J., Yu, T., Yu, Z., Wang, C., Zhang, H., et al. (2010). An eeg-based bci system for 2-d cursor control by combining mu/beta rhythm and p300 potential. *IEEE Trans. Biomed. Eng.* 57, 2495–2505. doi: 10.1109/TBME.2010.2055564

Lin, Z., Zhang, C., Wu, W., and Gao, X. (2006). Frequency recognition based on canonical correlation analysis for ssvep-based bcis. *IEEE Trans. Biomed. Eng.* 53, 2610–2614. doi: 10.1109/TBME.2006.886577

Lu, H., Eng, H.-L., Guan, C., Plataniotis, K. N., and Venetsanopoulos, A. N. (2010). Regularized common spatial pattern with aggregation for eeg classification in small-sample setting. *IEEE Trans. Biomed. Eng.* 57, 2936–2946. doi: 10.1109/TBME.2010.2082540

Neuper, C., Scherer, R., Reiner, M., and Pfurtscheller, G. (2005). Imagery of motor actions: differential effects of kinesthetic and visual-motor mode of imagery in single-trial eeg. *Cognit. Brain Res.* 25, 668–677. doi: 10.1016/j.cogbrainres.2005.08.014

Padfield, N., Zabalza, J., Zhao, H., Masero, V., and Ren, J. (2019). Eeg-based brain-computer interfaces using motor-imagery: techniques and challenges. *Sensors* 19, 1423. doi: 10.3390/s19061423

Parra, L. C., Christoforou, C., Gerson, A. C., Dyrholm, M., Luo, A., Wagner, M., et al. (2007). Spatiotemporal linear decoding of brain state. *IEEE Signal Process. Mag.* 25, 107–115. doi: 10.1109/MSP.2008.4408447

Pfurtscheller, G., and Da Silva, F. L. (1999). Event-related eeg/meg synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/S1388-2457(99)00141-8

Pfurtscheller, G., and Neuper, C. (2006). Future prospects of erd/ers in the context of brain-computer interface (bci) developments. *Prog. Brain Res.* 159, 433–437. doi: 10.1016/S0079-6123(06)59028-4

Picton, T. W. (1992). The p300 wave of the human event-related potential. *J. Clin. Neurophysiol.* 9, 456–479. doi: 10.1097/00004691-199210000-00002

Pudil, P., Novovičová, J., and Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognit. Lett.* 15, 1119–1125. doi: 10.1016/0167-8655(94)90127-9

Qin, X., Zheng, Y., and Chen, B. (2019). "Extract eeg features by combining power spectral density and correntropy spectral density," in *2019 Chinese Automation Congress (CAC)* (Hangzhou: IEEE), 2455–2459.

Qiu, Z., Jin, J., Lam, H.-K., Zhang, Y., Wang, X., and Cichocki, A. (2016). Improved sffs method for channel selection in motor imagery based BCI. *Neurocomputing* 207, 519–527. doi: 10.1016/j.neucom.2016.05.035

Ramoser, H., Muller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial eeg during imagined hand movement. *IEEE Trans. Rehabil. Eng.* 8, 441–446. doi: 10.1109/86.895946

Roy, S., Rathee, D., Chowdhury, A., McCreadie, K., and Prasad, G. (2020). Assessing impact of channel selectionon decoding of motor and cognitive imagery from meg data. *J. Neural Eng.* 17, 056037. doi: 10.1088/1741-2552/abbd21

Roy, S., Rathee, D., McCreadie, K., and Prasad, G. (2019). "Channel selection improves meg-based brain-computer interface," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)* (San Francisco, CA: IEEE), 295–298.

Schröder, M., Lal, T. N., Hinterberger, T., Bogdan, M., Hill, N. J., Birbaumer, N., et al. (2005). Robust eeg channel selection across subjects for brain-computer interfaces. *EURASIP J. Adv. Signal Process.* 2005, 1–10. doi: 10.1155/ASP.2005.3103

Subasi, A., and Gursoy, M. I. (2010). Eeg signal classification using pca, ica, lda and support vector machines. *Expert. Syst. Appl.* 37, 8659–8666. doi: 10.1016/j.eswa.2010.06.065

Tam, W.-K., Ke, Z., and Tong, K.-Y. (2011). "Performance of common spatial pattern under a smaller set of eeg electrodes in brain-computer interface on chronic stroke patients: a multi-session dataset study," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Boston, MA: IEEE), 6344–6347.

Tang, C., Li, Y., and Chen, B. (2022). Comparison of cross-subject eeg emotion recognition algorithms in the bci controlled robot contest in world robot contest 2021. *Brain Sci. Adv.* 8, 142–152. doi: 10.26599/BSA.2022.9050013

Tangermann, M., Müller, K.-R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., et al. (2012). Review of the BCI competition IV. *Front. Neurosci.* 6, 55. doi: 10.3389/fnins.2012.00055

Taniguchi, M., Kato, A., Fujita, N., Hirata, M., Tanaka, H., Kihara, T., et al. (2000). Movement-related desynchronization of the cerebral cortex studied with spatially filtered magnetoencephalography. *Neuroimage* 12, 298–306. doi: 10.1006/nimg.2000.0611

Tork, S., Yazdchi, M. R., and Karimizadeh, A. (2013). "How health related to quality of sleep and breathing: how to associate sleep disorder breathing and extracted respiratory signal," in *2013 21st Iranian Conference on Electrical Engineering (ICEE)* (Mashhad: IEEE), 1–6.

Vapnik, V. (1999). *The Nature of Statistical Learning Theory*. New York, NY: Springer Science & Business Media.

Varsehi, H., and Firoozabadi, S. M. P. (2021). An eeg channel selection method for motor imagery based brain-computer interface and neurofeedback using granger causality. *Neural Netw.* 133, 193–206. doi: 10.1016/j.neunet.2020.11.002

Wang, Y., Gao, X., Hong, B., Jia, C., and Gao, S. (2008). Brain-computer interfaces based on visual evoked potentials. *IEEE Eng. Med. Biol. Mag.* 27, 64–71. doi: 10.1109/MEMB.2008.923958

Wei, Q., and Wang, Y. (2011). "Binary multi-objective particle swarm optimization for channel selection in motor imagery based brain-computer interfaces," in *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, *Vol. 2* (Shanghai: IEEE), 667–670.

Xu, M., Chen, Y., Wang, D., Wang, Y., Zhang, L., and Wei, X. (2021). Multi-objective optimization approach for channel selection and cross-subject generalization in rsvp-based bcis. *J. Neural Eng.* 18, 046076. doi: 10.1088/1741-2552/ac0489

Yang, J., Singh, H., Hines, E. L., Schlaghecken, F., Iliescu, D. D., Leeson, M. S., et al. (2012). Channel selection and classification of electroencephalogram signals: an artificial neural network and genetic algorithm-based approach. *Artif. Intell. Med.* 55, 117–126. doi: 10.1016/j.artmed.2012.02.001

Yang, Q., Zhang, X., and Chen, B. (2020). "Mi3dnet: a compact cnn for motor imagery eeg classification with visualizable dense layer parameters," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Montreal, QC: IEEE), 510–513.

Zhang, Y., Guo, D., Li, F., Yin, E., Zhang, Y., Li, P., et al. (2018). Correlated component analysis for enhancing the performance of ssvep-based brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 948–956. doi: 10.1109/TNSRE.2018.2826541

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 23 November 2022
DOI 10.3389/fnins.2022.1044631

OPEN ACCESS

EDITED BY
Peng Xu,
University of Electronic Science and Technology of China, China

REVIEWED BY
Liangtian He,
Anhui University, China
Min Li,
Nanjing University of Science and Technology, China

*CORRESPONDENCE
Jian Huang
jackyhuang_cuit@126.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Neuroprosthetics, a section of the journal Frontiers in Neuroscience

RECEIVED 14 September 2022
ACCEPTED 26 October 2022
PUBLISHED 23 November 2022

CITATION
Du Y, Huang J, Huang X, Shi K and Zhou N (2022) Dual attentive fusion for EEG-based brain-computer interfaces. *Front. Neurosci.* 16:1044631. doi: 10.3389/fnins.2022.1044631

Dual attentive fusion for EEG-based brain-computer interfaces

Yuanhua Du[1†], Jian Huang[1*†], Xiuyu Huang[2], Kaibo Shi[3] and Nan Zhou[3]

[1]College of Applied Mathematics, Chengdu University of Information Technology, Chengdu, China, [2]Centre for Smart Health, The Hong Kong Polytechnic University, Hong Kong, Hong Kong SAR, China, [3]School of Electronic Information and Electronic Engineering, Chengdu University, Chengdu, China

The classification based on Electroencephalogram (EEG) is a challenging task in the brain-computer interface (BCI) field due to data with a low signal-to-noise ratio. Most current deep learning based studies in this challenge focus on designing a desired convolutional neural network (CNN) to learn and classify the raw EEG signals. However, only CNN itself may not capture the highly discriminative patterns of EEG due to a lack of exploration of attentive spatial and temporal dynamics. To improve information utilization, this study proposes a Dual Attentive Fusion Model (DAFM) for the EEG-based BCI. DAFM is employed to capture the spatial and temporal information by modeling the interdependencies between the features from the EEG signals. To our best knowledge, our method is the first to fuse spatial and temporal dimensions in an interactive attention module. This module improves the expression ability of the extracted features. Extensive experiments implemented on four publicly available datasets demonstrate that our method outperforms state-of-the-art methods. Meanwhile, this work also indicates the effectiveness of Dual Attentive Fusion Module.

KEYWORDS

brain-computer interface, electroencephalography, P300, motor imagery, dual attentive fusion

1. Introduction

Brain-computer interface (BCI) is a system that aims to establish a non-muscular communication pathway between humans and external devices *via* brain signals (Wolpaw et al., 2002). With the advances in information and computer science, various BCI paradigms have been developed and employed in many applications (Leeb et al., 2007; Dal Seno et al., 2010). The motor imagery (MI) paradigm attracts significant interest from researchers. It is the process of imagining movement in a certain body part rather than actually moving it. This technology can help patients with movement disorders manipulate external equipment such as artificial arms or wheelchairs (Leeb et al., 2007). The P300 event-related potential (ERP) is also another important paradigm in BCI. It is an evoked positive peak at around 300 ms after the occurrence of a low-probability stimulus. This type of BCI has been utilized

to assist individuals with severe neuromuscular diseases to spell characters by using brain waves (Dal Seno et al., 2010). Currently, electroencephalography (EEG) is one of the most widely used techniques for brain signal acquisition in BCIs due to its low cost, safety, and easy operation. The core of EEG-based BCI is to analyze EEG signals for the purpose of understanding human intentions. Therefore, improving the performance of EEG-based BCIs is very important for the future development of BCIs.

Generally, the classification based EEG first extracts discriminative features from EEG signals and adopts classifiers to classify the extracted features. However, it is not easy to deploy these processes due to the low signal-to-noise ratio (SNR) of EEG signals. Many previous methods rely on feature engineering and traditional machine learning approaches. For example, Rakotomamonjy and Guigue (2008) used a P300 detection model based on 896 hand-crafted features and an ensemble of SVMs classifiers. Fazli et al. (2009) proposed to combine feature extraction from the common spatial pattern (CSP) of the EEG signals and linear discriminant analysis (LDA) to classify the extracted features. The method developed by Li et al. (2006) alternatively used independent component analysis (ICA) to remove eye artifacts and selected a subset of electrodes prior to the classification made by support vector machine (SVM). Duan et al. (2013) first combined an SVM and K-nearest neighbor (KNN) to extract and classify features from multi-channel EEG data for emotion recognition. In Liu et al. (2005), principal component analysis (PCA) and T-weight value sums were applied for P300 classification. Although these attempts have achieved partial improvements in performance, all these methods only learn the features that the researchers focus on while ignoring other important features due to the limited abilities of hand-crafted features.

In addition, considerable effort has also been devoted to developing deep learning (DL) based methods for EEG signal classification (Zhang Y. et al., 2020; Huang et al., 2021), and they have demonstrated superior performance over conventional machine learning methods. Especially due to the temporal dynamics of EEG signals, recurrent neural network (RNN) based methods have been extensively applied to filter and classify EEG signals (Alhagry et al., 2017; Ma et al., 2018; Michielli et al., 2019). Alhagry et al. (2017) used an LSTM-RNN to learn and classify EEG signals for emotion recognition. Ma et al. (2018) proposed a pure RNNs-based parallel method to encode spatial and temporal information of raw EEG signals for motor imagery classification. Michielli et al. (2019) introduced a novel cascaded RNN architecture based on long short-term memory (LSTM) blocks for automated sleep stage classification.

Apart from RNN, convolutional neural network (CNN) has been popularly used for analyzing EEG signals and has gained much attention in recent years (Lawhern et al., 2018; Sakhavi et al., 2018; Shan et al., 2018; Yang et al., 2018a; Wu et al., 2019; Ding et al., 2021). Lawhern et al. (2018) presented a compact neural network named EEGNet, which can extract spatial and temporal features simultaneously. Wu et al. (2019) proposed a parallel multi-scale filter bank CNN architecture, generating temporal, and spatial features for MI classification. Ding et al. (2021) proposed TSception, a multi-scale CNN that learns discriminative in the time and channel dimensions to recognize the BCI's user emotion. Convolutional recurrent neural network (C-RNN) (Yang et al., 2018b; Zhang et al., 2018) was applied in EEG-based BCI and attained satisfactory performance. For example, Zhang et al. (2018) introduced cascade and parallel C-RNN models for human intention recognition and effectively learned the spatial-temporal representations of raw EEG signals. All these studies show the information in spatial and temporal dimensions carrying important information for BCI classification tasks. However, previous architectures handle the information of the EEG signal in temporal and spatial dimensions in either separate or subsequent manner without interaction.

Corresponding to this gap, we propose a simple but effective Dual Attentive Fusion Model (DAFM) for the EEG signal classification tasks. It leverages an interacting mechanism, which fuses spatial and temporal attention with a simple operation to generate the spatial-temporal pattern of the EEG signals. The main contributions of this paper can be summarized as follows.

1. The proposed model uses an interactive attention module, which can take both the spatial and temporal dimensions into consideration, and it successfully derives distinguishable features from EEG signals.
2. The proposed method is extensively evaluated on four widely used BCI datasets regarding both motor imagery (MI) and P300 tasks. Results exhibit that our approach has superior performance to state-of-the-art and baseline methods.

The remaining of this paper is organized as follows. Related works are described in Section 2. Section 3 presents the proposed method. Section 4 provides the datasets used in this paper, implementation details of our experiments and experimental results. Finally, Section 5 concludes this study.

2. Related work

2.1. Convolutional neural network

In recent years, deep learning, especially Convolutional Neural Networks (CNNs), has gained substantial interest in the computer vision field (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014; He et al., 2016). In 2012, Krizhevsky et al. (2012) proposed the AlexNet, which used a large, deep convolutional neural network to classify images in the ImageNet dataset and achieved considerably better results than the previous state-of-the-art methods. Convolutional

frameworks have become an essential medium in vision-related fields. VGGNet proposed by Simonyan and Zisserman (2014) has good transfer learning ability. Since then, the 3 × 3 convolution has become the standard configuration of the subsequent convolutional neural network structures. In 2015, He et al. (2016) noticed the gradient vanishing problem caused by the deepening of the network, and proposed the ResNet, which got rid of the troubles of the deep network and made the network depth reach astonishing 152 layers. Recently various CNN-based models are increasingly being used for EEG-based BCI and gain excellent performance. Sakhavi et al. (2018) introduced a new temporal representation of the data and used a CNN architecture for MI classification. Shan et al. (2018) proposed a novel and simple CNN, which only used a convolutional layer, to effectively learn feature representations from both temporal and spatial information for accurate P300 detection. Yang et al. (2018a) proposed to combine features of signals from different frequency bands and used a continuous convolutional neural network to make predictions.

2.2. Attention mechanism

It could be said that the attention mechanism has become one of the hottest topics in the deep learning field. The attention mechanism, which can selectively amplify valuable features and suppress useless features based on global information, has been employed in diverse domains. Fu et al. (2019) proposed a novel Dual Attention Network (DANet) to capture feature dependencies in the spatial and channel dimensions for scene segmentation. Huang et al. (2019) proposed a novel Criss-Cross Network (CCNet) to capture full-image contextual information adaptively in a more efficient way for semantic segmentation. Chen et al. (2019) proposed an Attentive but Diverse Network (ABD-Net) to integrate attention mechanism into ABD-Net, containing Channel Attention Module, and Position Attention Module for person re-identification. The attention mechanism is also used to transform the input into a more discriminative representation in the brain-computer interface field. Kim and Choi (2020) combined an attention mechanism and a long short-term memory network to assign weights to different emotional states based on importance and improved emotion recognition accuracy. Tao et al. (2020) proposed an attention-based convolutional recurrent neural network (ACRNN), which integrated the channel-wise attention into CNN to extract spatial information and extended self-attention into RNN to extract temporal information. Zhang D. et al. (2020) proposed a Graph-based Convolutional Recurrent Attention Model (G-CRAM) to explore EEG features across different subjects for motor imagery classification. Graph structure was employed to enhance the discriminative ability of EEG channels in this model.

3. Methods

In this section, we first present an overall framework of our network, which contains two modules. Then, we describe the details of Dual Attentive Fusion Module. Finally, we introduce the Feature Classification Module. The overall architecture of our model is shown in Figure 1.

3.1. Overview

Raw EEG signals contain spatial relationship among different channels and temporal dependency among different time points, which play an important role in feature classification. However, many studies (Kim and Choi, 2020; Tao et al., 2020; Zhang D. et al., 2020) suggest that features generated by traditional machine learning methods could not extract this information well. In 2017, the Transformer proposed by Vaswani et al. (2017) raised much attention in the natural language processing field. A transformer model is based on the self-attention module, which effectively focuses on the distinct features by assigning attention score to each feature and aggregating these scores. More and more work has introduced attention mechanism into the computer science field and achieved comparable performance. Recent work has focused on designing proper attention modules to adaptively explore attentive dynamics of EEG signals and focus on the most valuable information in brain-computer interface fields. Inspired by it, we propose a Dual Attentive Fusion Module which can take spatial and temporal attention into consideration in an interactive module. Our method can turn raw EEG signals into more discriminative features. More importantly, our method improves the accuracy of EEG signal classification. First, a filtering process is conducted on all EEG signals by implementing bandpass filter. Then, the proposed attention module is used to recode the EEG signals considering the spatial and temporal dimensions together. Finally, the features are fed into a convolutional neural network to make classification, and classification accuracy is considered as the final evaluation metric. The Dual Attentive Fusion Module is illustrated in Figure 2.

3.2. Dual attentive fusion module

As illustrated in Figure 1, an EEG signal **A**, is denoted as $\mathbf{A} \in \mathbb{R}^{H \times W}$, where H is the number of electrodes and W is the number of time points. We first feed the data into a convolution layer to generate a new feature map **B**, where **B** belongs to $\mathbb{R}^{H \times W \times C}$ and C = 1 denotes the number of feature map.

Then, to learn the spatial features of multi-channel EEG and explore the temporal features of different time points, we employ a self-attention mechanism in the EEG signals. In the

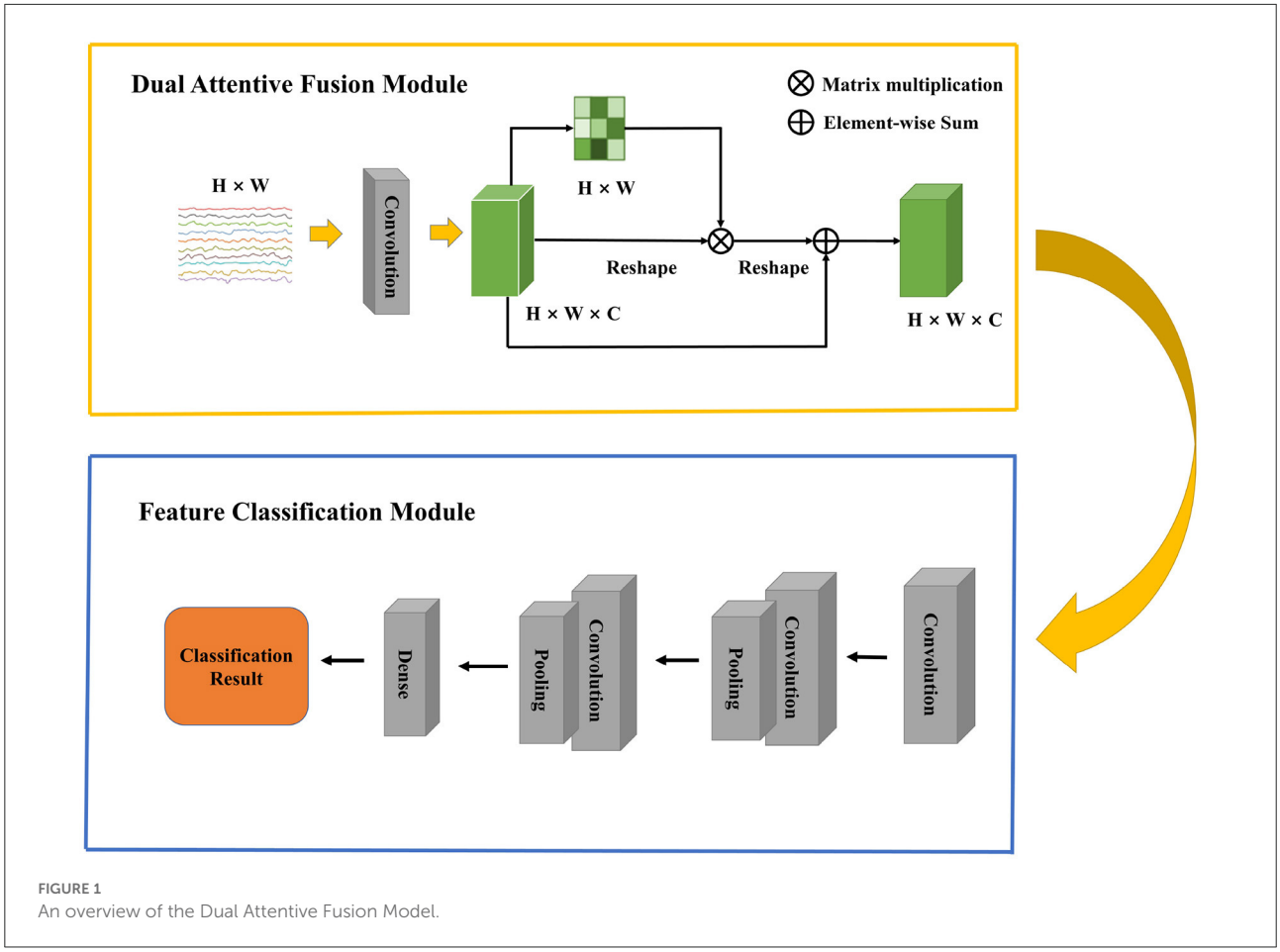


FIGURE 1
An overview of the Dual Attentive Fusion Model.

spatial dimension, a self-attention operation turns channels into a probability distribution as weights and recodes the EEG signals based on the weights. In this way, an important feature would gain a higher weight than less important features. Therefore, we compress **B** to feature map **C** by a convolution layer, where **C** belong to $\mathbb{R}^{H\times 1\times C}$. A softmax activation function is applied to **C** to obtain the attention map of the spatial dimension:

$$a_1 = \frac{exp(C_i^T v_i)}{\sum exp(C_i^T v_i)}, \tag{1}$$

The attention vector $v_i \in \mathbb{R}^H$ is randomly initialized and tuned by the above function during the training procedure. The softmax function makes sure the sum of weights is 1. The more similar feature representation of the two channels devotes to a more significant correlation between them.

Next, in the temporal dimension, to extract more discriminative temporal information, we also employ a self-attention operation to obtain a feature representation by perceiving global temporal features and assigning the weights according to the similarity of time points. Thus, a convolution layer is applied to compress **B** as **D**, which belongs to $\mathbb{R}^{1\times W\times C}$. A softmax activation function is also applied to **D** to obtain the attention map of the temporal dimension:

$$a_2 = \frac{exp(D_i^T w_i)}{\sum exp(D_i^T w_i)}, \tag{2}$$

The attention vector $w_i \in \mathbb{R}^W$ is randomly initialized and tuned by the above function during the training procedure. This attention map will focus on specific time points that are distinct from others. To enable matrix multiplication between a_1 and a_2, we reshape a_1 and a_2 as to $\mathbb{R}^{H\times C}$ and $\mathbb{R}^{C\times W}$, respectively.

Finally, a matrix multiplication is employed to obtain the spatial-temporal attention map $a \in \mathbb{R}^{H\times W}$ as:

$$a = a_1^T \cdot a_2, \tag{3}$$

where a_1 and a_2 are spatial and temporal attention map of EEG signals, respectively. A dual attentive fusion feature representation is further generated by considering the spatial-temporal attention map as weights to recode EEG signals. Thus, a is reshaped to $\mathbb{R}^{H\times W\times C}$, and we perform an element-wise matrix operation between a and **B**. The dual attentive fusion

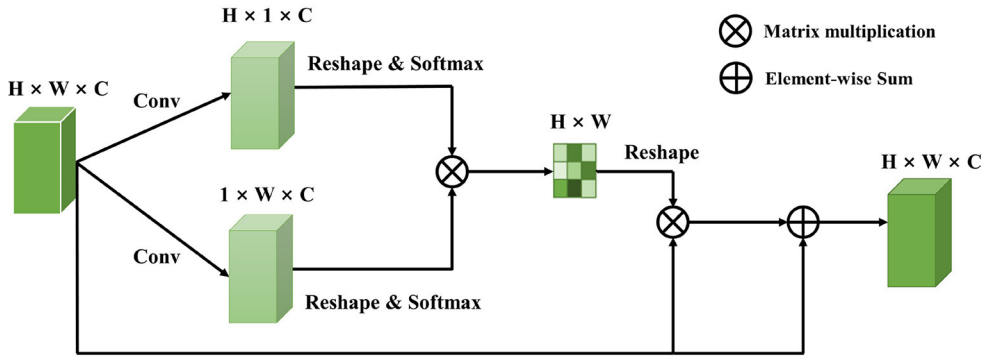


FIGURE 2
The detail of Dual Attentive Fusion Module.

feature extracted by the Dual Attentive Fusion Module can be expressed as follows:

$$E = \sum (a_i \cdot B_i), \tag{4}$$

where $\mathbf{E} \in \mathbb{R}^{H \times W \times C}$, a_i denotes the spatial-temporal attention map, and **B** represents the preprocessed EEG signals. In addition, a residual block (He et al., 2016) is applied between **E** and **B** to obtain the final dual attentive fusion feature as follows:

$$Z = W \cdot E + B, \tag{5}$$

where W belongs to a learnable parameter, which is randomly initialized and is gradually updated during the training procedure. Equation (5) shows that the final feature of EEG signals is a weighted sum of the spatial-temporal features and original features.

3.3. Feature classification module

In this module, we employ a CNN, which is inherited the architecture of the EEGNet (Lawhern et al., 2018), to classify the features extracted from the previous module. A 2D convolution layer with a kernel size of (1, K_1) is first applied to **Z** to capture temporal information in each electrode. Then, a depthwise convolution layer with a kernel size of (H, 1) is used for spatial feature extraction. An average pooling operation is followed to generate a coarser feature. Next, the separableConv2D with a kernel size of (1, K_2) is used to obtain deeper temporal patterns across all electrodes. An average pooling operation is also followed to reduce dimension. It is worth noting that batch normalization (Ioffe and Szegedy, 2015) and exponential linear unit (Clevert et al., 2015) are followed by some convolution operations for feature standardization and nonlinear transformation. Finally, the deep feature extracted by CNN is flattened as a vector by a flatten layer. For binary classification, the output of dense layer is forwarded into a sigmoid function. For multi-class classification, the output of dense layer is forwarded into a softmax function. The final prediction is based on conditional probability, which is calculated by the loss function. The loss value guides the gradient descent and the backpropagation for the whole neural network. The structure of Feature Classification Module and its parameters are shown in Tables 1, 2.

4. Experiments and results

In this section, we first describe the benchmark datasets used in this paper. Then, we demonstrate the model implementation details. Finally, we present the experimental results obtained by our method and other comparable approaches.

4.1. Dataset description

In our experiment, we use four public BCI competition datasets to evaluate the effectiveness of the proposed method. Among them, BCI Competition IV-2a (Tangermann et al., 2012) and BCI Competition IV-2b (Tangermann et al., 2012) are used for motor imagery classification. BCI Competition II Dataset IIb (Blankertz, 2010) and BCI Competition III Dataset II (Blankertz et al., 2008) are used for P300 detection. The detailed information of the four datasets is shown as follows.

4.1.1. The BCI competition IV-2a dataset

The BCI competition IV-2a dataset, provided by Graz University, contains EEG signals from nine healthy subjects(A01-A09) and two sessions on different days for each subject. Each session consists of 288 trials of four different MI classes: imagining the movement of the left hand, the right hand, the feet, and the tongue. The signals are recorded by

TABLE 1 Architecture of feature classification module.

Layer	Input	Filter	Kernel	Output
Conv2D	(H, W, 1)	F1	(1, K_1)	(H, W, F_1)
BatchNorm	(H, W, F_1)			(H, W, F_1)
DepthwiseConv2D	(H, W, F_1)	F_1*D	(H, 1)	(1, W, F_1*D)
BatchNorm	(1, W, F_1*D)			(1, W, F_1*D)
ELU activation	(1, W, F_1*D)			(1, W, F_1*D)
AveragePooling2D	(1, W, F_1*D)		(1, P_1)	(1, W/P_1, F_1*D)
SeparableConv2D	(1, W/P_1, F_1*D)	F_2	(1, K_2)	(1, W/P_1, F_2)
BatchNorm	(1, W/P_1, F_2)			(1, W/P_1, F_2)
ELU activation	(1, W/P_1, F_2)			(1, W/P_1, F_2)
AveragePooling2D	(1, W/P_1, F_2)		(1, P_2)	(1, W/(P_1*P_2), F_2)
Flatten	(1, W/(P_1*P_2), F_2)			(W*F_2)/(P_1*P_2)
Dense	(W*F_2)/(P_1*P_2)			N

TABLE 2 Hyperparameter setting.

Hyperparameter	II	III	IV-2a	IV-2b
H	10	10	22	3
W	144	144	1,000	1,000
F_1	8	8	8	8
F_2	16	16	16	16
K_1	72	72	64	64
K_2	16	16	16	16
P_1	4	4	4	4
P_2	8	8	8	8
D	2	2	2	2
N	2	2	4	2

22 electrodes at 250 Hz sampling frequency and bandpass filtered between 0.5 and 100 Hz. In this paper, as the same data division in the competition, we use the 288 trials of the first session as training and the 288 trials of the second session as testing. In each trial, we only use a 4 s temporal segment in our model, each sample can be represented as a 2D-matrix of 22 $\times$ 1,000, in which 22 represents the number of electrodes and 1,000 represents the number of sample points.

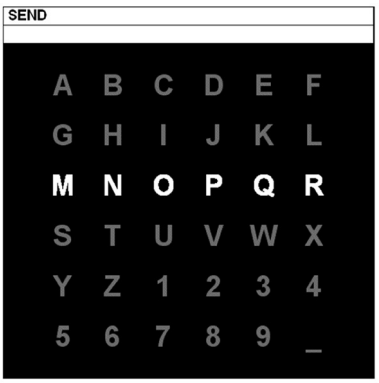


FIGURE 3
P300 speller paradigm.

4.1.2. The BCI competition IV-2b dataset

The BCI competition IV-2b dataset is also collected from nine healthy people (B01–B09) at a sample rate of 250 Hz but only recorded from three electrodes placed at positions C3, Cz, and C4. For each subject, 720 trials from two MI tasks, including left-hand and right-hand movement imagination, are performed. There are five sessions for each individual. The first three sessions are for training, and the remaining two are for testing as the same data division in the competition.

In this paper, The 4 s temporal segment of each trial is used as a sample, which can be represented as a 2D-matrix of 3 $\times$ 1,000.

4.1.3. BCI competition II—dataset IIb and BCI competition III—dataset II

Both datasets are offered by Wadsworth Center, New York State Department of Health. BCI Competition II—Dataset IIb is composed of a single subject data collected in three sessions containing 42 training and 31 testing characters. In BCI Competition III—Dataset II, there are two subjects: Subject A and Subject B. For each subject, the EEG signals are divided into a training set (85 characters) and a testing set (100 characters). In the experiments, the subject was presented with a 6 × 6 matrix of characters shown in Figure 3. In 1988, Farwell and Donchin developed this type of P300 speller paradigm (Farwell and Donchin, 1988). The user was asked to concentrate on the characters of a given word (one character at one time). All six rows and six columns randomly and successively intensified at 5.7 Hz. One row and one column out of these 12 intensive flashings contained the desired character. The sets of 12 intensifications were repeated 15 times for each character. The EEG data were bandpass filtered between 0.1 and 60 Hz and digitized at 240 Hz from 64 channels. In this paper, we choose 10 electrodes, including Fz, Cz, Pz, Oz, C3, C4, P3, P4, PO7, and PO8, in which the P300 signals are mainly generated. Due to a positive response around 300 ms after the onset of the stimulus in P300 ERP, we extract a time window of 600 ms after intensification onset as the input for each trial. With the collected frequency of 240 Hz, a trial can be denoted as a 10×144 data matrix.

4.2. Implementation details

In the motor imagery classification experiment, the model is implemented with the Keras framework and trained on Google online platform (Colab). The Adam optimizer (Kingma and Ba, 2014) with a learning rate of 0.001 is employed to minimize the cross-entropy loss function. The mini-batch size is set to

TABLE 3 Classification accuracies (%) obtained with the dataset BCI competition IV-2a.

Methods	Subject									*Average ± SD*
	A01	**A02**	**A03**	**A04**	**A05**	**A06**	**A07**	**A08**	**A09**	
FBCSP	76.00	56.50	81.25	61.00	55.00	45.25	82.75	81.25	70.75	67.75 ± 13.73
CCSP	84.72	52.78	80.90	59.38	54.51	49.31	88.54	71.88	56.60	66.50 ± 15.13
BOTDA	80.43	55.83	80.90	57.64	55.39	62.79	70.23	81.92	80.68	69.38 ± 11.95
EEGNet	85.76	61.46	88.54	67.01	55.90	52.08	89.58	83.33	**86.81**	74.50 ± 15.23
ConNet	76.39	55.21	89.24	74.65	56.94	54.17	92.71	77.08	76.39	72.53 ± 14.24
DEI	81.85	53.71	81.25	66.67	57.97	63.72	84.48	79.70	79.92	72.14 ± 11.66
DRDA	83.19	55.14	87.43	75.28	62.29	57.15	86.18	83.61	82.00	74.70 ± 12.96
DAJAN	86.46	68.75	93.06	**85.42**	72.57	63.54	**95.49**	**85.76**	83.68	81.52 ± 10.94
FTF	83.27	57.24	91.94	66.67	76.45	**66.51**	86.28	83.39	82.58	77.15 ± 11.34
DAFM	**86.83**	**72.43**	**96.70**	74.56	**81.52**	64.65	91.69	85.60	84.84	**82.09 ± 10.02**

Highest values are highlighted in boldface.

TABLE 4 Classification accuracies (%) obtained with the dataset BCI competition IV-2b.

Methods	Subject									*Average ± SD*
	B01	**B02**	**B03**	**B04**	**B05**	**B06**	**B07**	**B08**	**B09**	
FBCSP	70.00	60.36	60.94	97.50	93.12	80.63	78.13	92.50	86.88	80.01 ± 13.85
CCSP	63.75	56.79	50.00	93.44	65.63	81.25	72.81	87.81	82.81	72.70 ± 14.72
BOTDA	61.40	55.92	54.78	88.93	92.67	73.71	71.98	86.35	79.18	73.88 ± 14.18
EEGNet	68.44	57.86	61.25	90.63	80.94	63.13	84.38	93.13	83.13	75.88 ± 13.33
ConNet	76.56	50.00	51.56	96.88	93.13	85.31	83.75	91.56	85.62	79.37 ± 17.25
DEI	70.18	62.04	71.74	90.23	86.08	75.70	89.66	87.39	85.71	79.86 ± 10.17
DRDA	81.37	62.86	63.63	95.94	93.56	**88.19**	85.00	95.25	**90.00**	83.98 ± 12.67
DAJAN	**83.44**	58.57	59.06	98.13	96.56	84.38	86.25	92.81	87.81	83.00 ± 14.64
FTF	78.07	68.16	73.04	96.74	95.24	84.86	92.67	92.17	85.71	85.18 ± 10.17
DAFM	70.18	**71.84**	**89.56**	**99.02**	**100.00**	73.71	**94.40**	**95.65**	88.98	**87.04 ± 11.95**

Highest values are highlighted in boldface.

16, dropout regularization is 0.2, and the epoch is 1,000. Batch normalization is adopted to get better performance.

In the P300 detection experiment, the model is constructed with the Keras framework on Google online platform (Colab) and trained from scratch. The training procedure is performed by minimizing the binary cross-entropy loss function. It is guided by Stochastic Gradient Descent with Adam optimizer (Kingma and Ba, 2014). The learning rate is set as 0.001. The batch size is set to be 150, and the epoch is 300. Dropout regularization with 0.5 is applied in our model. Batch shuffling is implemented for better generalization.

4.3. Results on motor imagery datasets

4.3.1. Comparison results

In order to evaluate the effectiveness of our proposed method, we compare it with other state-of-the-art methods, including FBCSP (Ang et al., 2012), CCSP (Kang et al., 2009), BOTDA (Peterson et al., 2021), EEGNet (Lawhern et al., 2018), ConNet (Zhang Y. et al., 2020), DEI (Zhang C. et al., 2021), DRDA (Zhao et al., 2020), DAJAN (Hong et al., 2021), and FTF (Zhang K. et al., 2021). Tables 3, 4 show the classification accuracies of each subject and the average accuracies of different methods on BCI IV-2a and IV-2b datasets, respectively.

We observe that the proposed method achieves the highest average classification accuracies of 82.09 and 87.04% on BCI IV-2a and IV-2b datasets, respectively. Regarding the experimental results of every subject, our method achieves accuracy above 70% except the A06 subject on both datasets. The best classification accuracy is obtained at the A03 and B05 subjects on BCI IV-2a and IV-2b datasets, respectively. Moreover, the standard deviation (SD) of our method is lower than that of other approaches on the BCI IV-2a dataset. On the BCI IV-2b dataset, the SD of our method is lower than that of other approaches except DEI and FTF. Generally, our method achieves the best results and has good stability on both MI datasets. The main reason that our method outperforms traditional methods is its nonlinear modeling ability which is the advantage of deep learning methods. Our method also has superior performance over other deep learning methods due to our proposed DAFM. Compared with the simple CNN models such as EEGNet and ConNet without dual attention mechanism, the proposed module improves the performance of the model by selectively amplifying valuable features and suppressing useless features based on the data-driven attentive scores.

To evaluate the capacity of our method, we perform the classification experiments on BCI IV-2a and IV-2b datasets under both without DAFM and with DAFM, respectively. The classification accuracies on both datasets are shown in Figure 4.

On the BCI IV-2a dataset, DAFM has different influence on the classification accuracy for all subjects. The classification accuracies of eight subjects improve. Only the performance

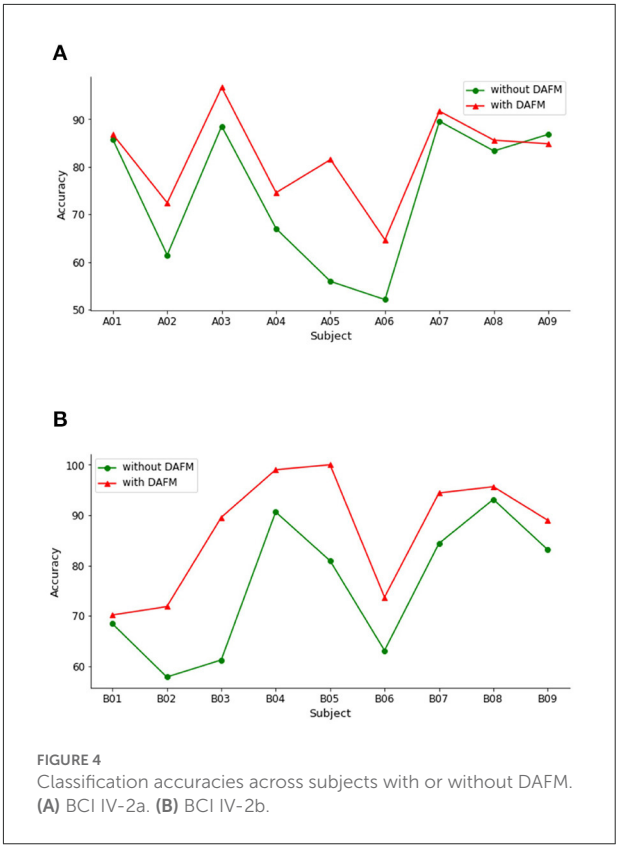


FIGURE 4
Classification accuracies across subjects with or without DAFM. **(A)** BCI IV-2a. **(B)** BCI IV-2b.

on subject A09 slightly decreases. As shown in Figure 4B, the performance of DAFM has a better performance across all subjects. These encouraging findings show that the DAFM is beneficial to MI classification and generate more discriminative feature regardless different individuals.

4.3.2. Result of the confusion matrices

In this part, we use confusion matrices to show the predictive outcome of our method in each class. Confusion matrices on BCI IV-2a and IV-2b datasets are presented in Figures 5, 6, respectively. The vertical axis represents the true label, and the horizontal one represents the predicted label. We randomly select two subjects on the BCI IV-2a dataset (i.e., A03 and A04) and BCI IV-2b dataset (i.e., B01 and B02).

First, on the BCI IV-2a dataset, for subject A03, we could observe that the left hand, right hand and foot are easier to be recognized than the tongue. By comparison between Figures 5A,B, in which Figure 5A does not use our method, we discover the classification accuracies of the four categories have improved significantly, which demonstrates that our proposed attention module is beneficial to the MI classification. Moreover, the gap between different classes has narrowed. For subject A04, it is obvious that the right hand and foot are easier to be recognized than the left hand and tongue. We could

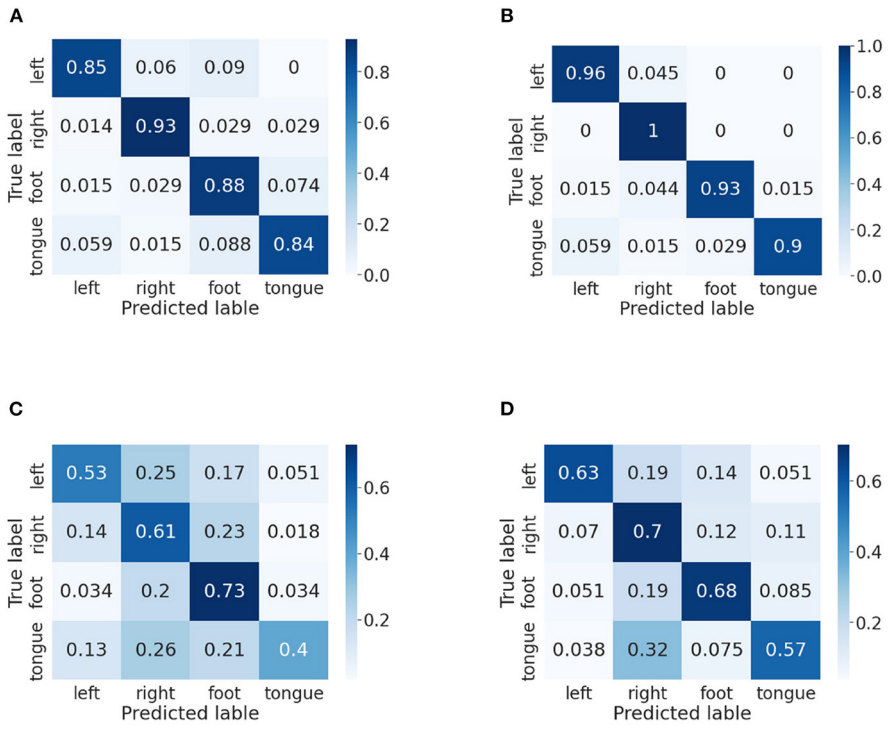


FIGURE 5
Confusion matrices of BCI competition IV-2a datasets. **(A)** A03 without our method. **(B)** A03 with our method. **(C)** A04 without our method. **(D)** A04 with our method.

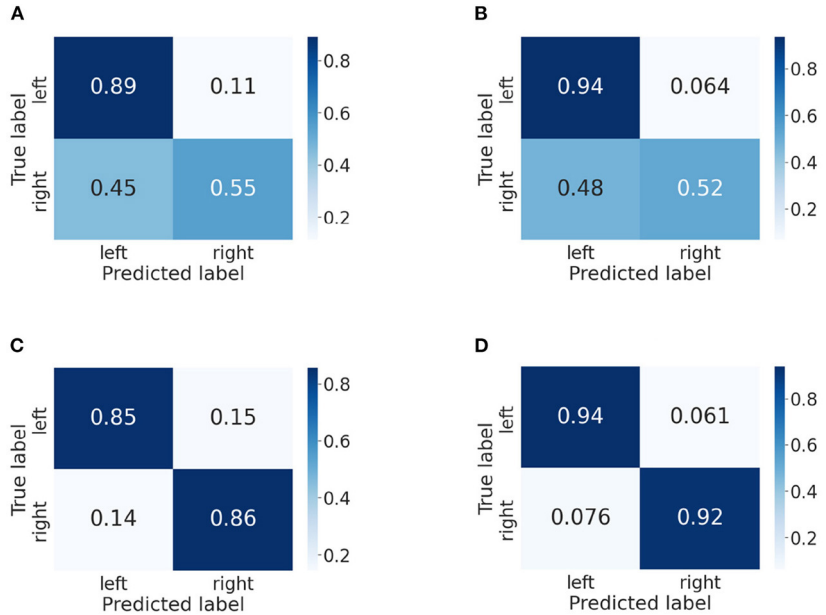


FIGURE 6
Confusion matrices of BCI competition IV-2b datasets. **(A)** B01 without our method. **(B)** B01 with our method. **(C)** B02 without our method. **(D)** B02 with our method.

TABLE 5 Classification accuracy (%) on BCI competition II dataset.

Method	Accuracy
CNN1	89.70
CNN3	87.54
CNNR	89.52
BN3	88.26
OCLNN	87.37
EEGNet	91.49
DeepConvNet	91.49
ShallowConvNet	88.62
Ours	**93.64**

Highest values are highlighted in boldface.

TABLE 6 Classification accuracy (%) on BCI competition III dataset.

Method	Subject		Average accuracy
	A	B	
CNN1	85.25	89.08	87.17
CNN3	83.92	86.92	85.42
CNNR	84.83	89.17	87.00
BN3	84.67	90.33	87.50
OCLNN	85.33	90.58	87.96
EEGNet	86.92	91.75	89.34
DeepConvNet	87.00	90.50	88.75
ShallowConvNet	83.50	86.50	85.00
Ours	**87.50**	**92.50**	**90.00**

Highest values are highlighted in boldface.

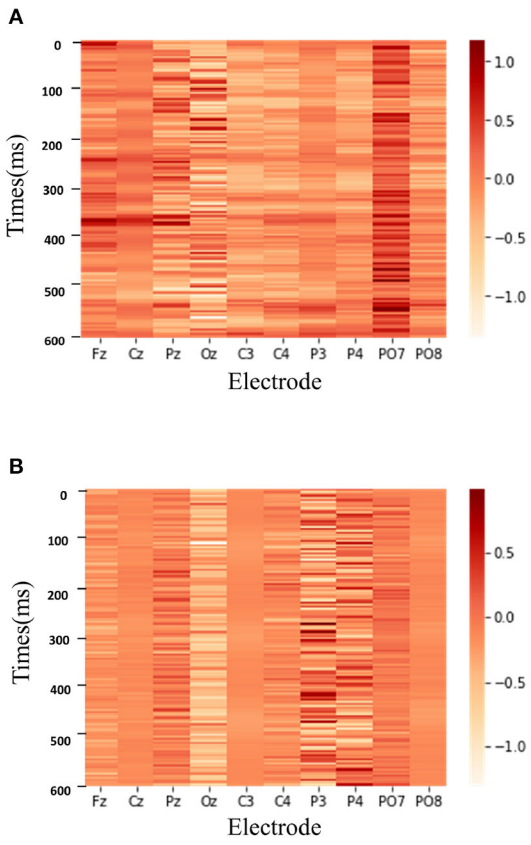


FIGURE 7
The two classes heatmap result of DAFM on the BCI Competition II dataset. **(A)** P300 signal. **(B)** Non-P300 signal.

discover that the proposed method increases the classification rate between each class except the foot. However, without using our method, other categories are easily misclassified as foot, and by using our method, the misclassification rate has decreased a lot.

Then, we analyze the confusion matrices of the BCI IV-2b dataset, which has two classes. For subject B01, we can find that the left hand is much easier to be recognized than the right hand. Except this, we can see that the classification effect of the left hand significant improves though the right hand's classification rate decreases slightly. The comparison of Figures 6C,D indicates that our method improves the classification performance for subject B02 and reduces the misclassification rate of both classes.

4.4. Results on P300 datasets

4.4.1. Comparison results

We perform a series of experiments on the BCI Competition II dataset and BCI Competition III dataset to further validate the effectiveness of our method. We compare the classification accuracies for our method with other state-of-the-art methods, including CNN1 (Cecotti and Graser, 2010), CNN3 (Cecotti and Graser, 2010), CNNR (Manor and Geva, 2015), BN3 (Liu et al., 2018), OCLNN (Shan et al., 2018), EEGNet (Lawhern et al., 2018), DeepConvNet (Zhang Y. et al., 2020), and ShallowConvNet (Zhang Y. et al., 2020). The experimental results on both datasets are shown in Tables 5, 6, respectively.

We observe that the proposed method improves clearly compared to other approaches, with around 2.15% better than the second-best method on the BCI Competition II dataset. Table 6 shows that our method outperforms all the comparable methods, obtaining an average accuracy of 90.00% on the BCI Competition III dataset. Thus, the proposed method can achieve the best performance on both datasets. The experimental results demonstrate that DAFM provides a more accurate classification outcome for P300 detection task.

To better demonstrate the role of the proposed method, we also exhibit the weighed features learned by our method.

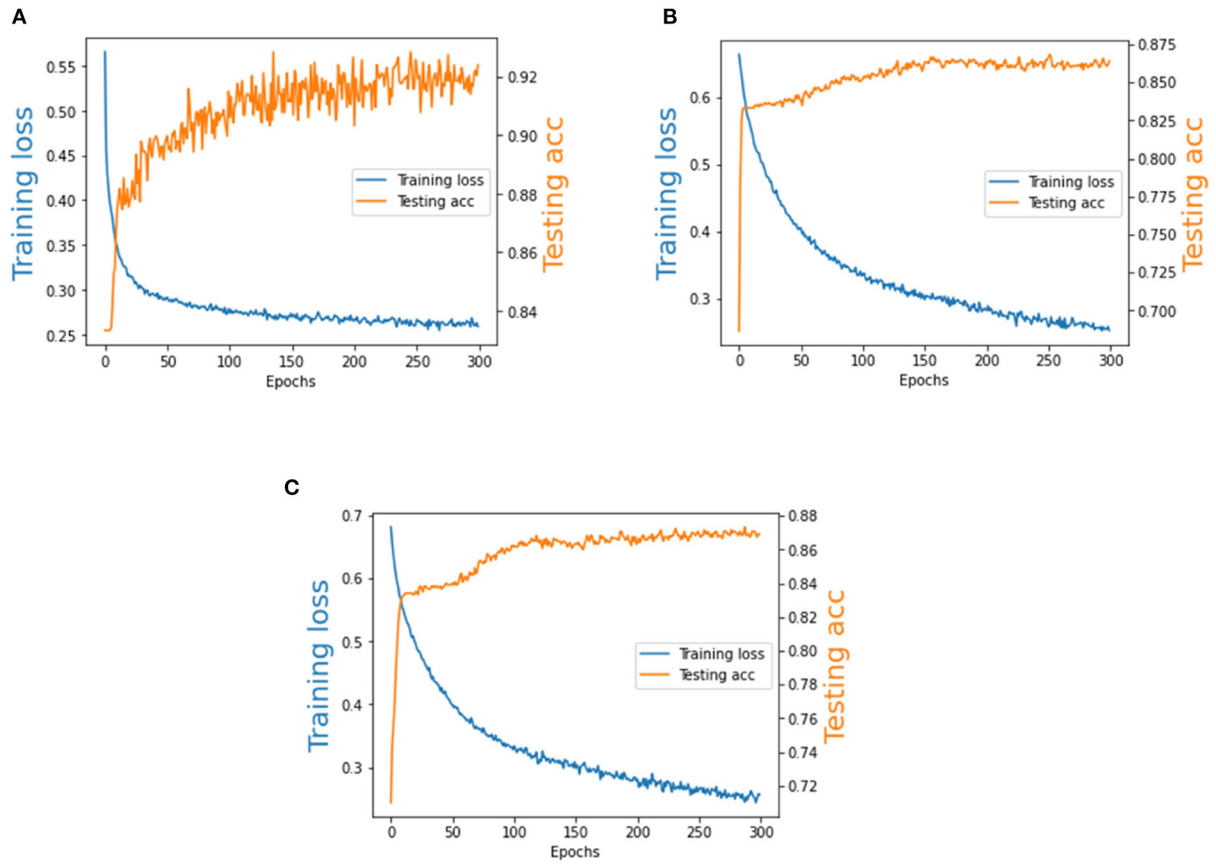


FIGURE 8
The training loss and testing accuracy of BCI Competition II dataset and BCI Competition III dataset. **(A)** BCI Competition II dataset. **(B)** Subject A of BCI Competition III dataset. **(C)** Subject B of BCI Competition III dataset.

Figure 7 shows the two classes heatmap result of DAFM on the BCI Competition II dataset.

As shown in Figure 7, DAFM focuses on different ranges of EEG signals for P300 detection. The deeper the color is, the more attention the model pays to the corresponding part of EEG signals. When recognizing the P300 signal, the model has a high degree of attention around 300 ms time points due to a positive peak appearing after 300 ms of the stimulus in the P300 signal. In contrast, the feature map of the non-P300 signals has a more scattered appearance over time. It is illustrated that the proposed attention module can automatically learn the priority of different temporal points, which contributes to better performance.

4.4.2. The training loss and testing accuracy of our method

We analyze the training loss and the testing accuracy of our method on the BCI Competition II dataset and BCI Competition III dataset. As is shown in Figure 8, the number of training epochs is 300. It can be observed that the testing accuracy increases quickly during the first 50 epochs and the training loss is generally stable after training about 100 epochs. Therefore, Our model exhibits a stable performance during the training procedure, and we observe that it converges quickly.

5. Conclusion

This study proposes a novel DAFM framework to effectively extract discriminative features from the EEG signals for different EEG-based classification tasks. It leverages an interactive attention module to generate the informative spatial-temporal features. The experimental results, conducted on four widely-used datasets, demonstrate that our method achieves superior performance to state-of-the-art and baseline methods. Our ablation experiments also confirm the effectiveness of our method. In summary, our method could be regarded as a potential approach to improve the performance of EEG-based BCI systems.

Due to a large amount of noise and artifacts in EEG signals, the proposed method can alleviate the interference of noise to a certain extent by focusing on useful information and ignoring useless information, but it cannot eliminate them. In the future, we plan to explore the more stable patterns of EEG signals using attention mechanism. Meanwhile, the proposed attention module will be extended to other tasks, such as image classification, semantic segmentation, etc.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: http://www.bbci.de/competition/.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

JH conceptualized the study, performed the majority of the experiments and analyses, made the figures, and wrote the first draft of the manuscript. XH, YD, KS, and NZ performed some experiments, updated the figures, performed the statistics, and edited the manuscript. All authors approved the submitted version.

Funding

The work in this paper was supported in part by National Nature Science Foundation of China (No. 11901063).

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

References

Alhagry, S., Fahmy, A. A., and El-Khoribi, R. A. (2017). Emotion recognition based on EEG using LSTM recurrent neural network. *Int. J. Adv. Comput. Sci. Appl.* 8, 355–358. doi: 10.14569/IJACSA.2017.081046

Ang, K. K., Chin, Z. Y., Wang, C., Guan, C., and Zhang, H. (2012). Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Front. Neurosci.* 6:39. doi: 10.3389/fnins.2012.00039

Blankertz, B., et al. (2008). *BCI Competition III Webpage*. Available online at: http://ida. first. fraunhofer. de/projects/bci/competition

Blankertz, B. (2010). *BCI Competition II (2003)-p300 Speller Dataset Webpage*. Berlin.

Cecotti, H., and Graser, A. (2010). Convolutional neural networks for p300 detection with application to brain-computer interfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 433–445. doi: 10.1109/TPAMI.2010.125

Chen, T., Ding, S., Xie, J., Yuan, Y., Chen, W., Yang, Y., et al. (2019). "ABD-net: attentive but diverse person re-identification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (Texas, TX), 8351–8361. doi: 10.1109/ICCV.2019.00844

Clevert, D.-A., Unterthiner, T., and Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (ELUS). *arXiv[Preprint].arXiv:1511.07289*. doi: 10.48550/arXiv.1511.07289

Dal Seno, B., Matteucci, M., and Mainardi, L. (2010). Online detection of p300 and error potentials in a bci speller. *Comput. Intell. Neurosci.* 2010:307254. doi: 10.1155/2010/307254

Ding, Y., Robinson, N., Zhang, S., Zeng, Q., and Guan, C. (2021). Tsception: Capturing temporal dynamics and spatial asymmetry from EEG for emotion recognition. *arXiv[Preprint].arXiv:2104.02935*. doi: 10.1109/TAFFC.2022.3169001

Duan, R.-N., Zhu, J.-Y., and Lu, B.-L. (2013). "Differential entropy feature for EEG-based emotion classification," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)* (Shanghai), 81–84doi: 10.1109/NER.2013.6695876

Farwell, L. A., and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr. Clin. Neurophysiol.* 70, 510–523. doi: 10.1016/0013-4694(88)90149-6

Fazli, S., Grozea, C., Danóczy, M., Blankertz, B., Popescu, F., and Müller, K.-R. (2009). Subject independent EEG-based BCI decoding. *Adv. Neural Inform. Process. Syst.* 22, 1305–1312. doi: 10.1016/j.neunet.2009.06.003

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., et al. (2019). "Dual attention network for scene segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Beijing), 3146–3154. doi: 10.1109/CVPR.2019.00326

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. doi: 10.1109/CVPR.2016.90

Hong, X., Zheng, Q., Liu, L., Chen, P., Ma, K., Gao, Z., et al. (2021). Dynamic joint domain adaptation network for motor imagery classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 556–565. doi: 10.1109/TNSRE.2021.3059166

Huang, X., Zhou, N., and Choi, K.-S. (2021). A generalizable and discriminative learning method for deep eeg-based motor imagery classification. *Front. Neurosci.* 15:79. doi: 10.3389/fnins.2021.760979

Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., and Liu, W. (2019). "CCNet: criss-cross attention for semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (Wuhan), 603–612. doi: 10.1109/ICCV.2019.00069

Ioffe, S., and Szegedy, C. (2015). "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning* (California, CA), 448–456.

Kang, H., Nam, Y., and Choi, S. (2009). Composite common spatial pattern for subject-to-subject transfer. *IEEE Signal Process. Lett.* 16, 683–686. doi: 10.1109/LSP.2009.2022557

Kim, Y., and Choi, A. (2020). EEG-based emotion classification using long short-term memory network with attention mechanism. *Sensors* 20:6727. doi: 10.3390/s20236727

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv[Preprint].arXiv:1412.6980*. doi: 10.48550/arXiv.1412.6980

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (Toronto), 25.

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

Leeb, R., Friedman, D., Müller-Putz, G. R., Scherer, R., Slater, M., and Pfurtscheller, G. (2007). Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with a tetraplegic. *Comput. Intell. Neurosci.* 2007:79642. doi: 10.1155/2007/79642

Li, Y., Ma, Z., Lu, W., and Li, Y. (2006). Automatic removal of the eye blink artifact from EEG using an ICA-based template matching approach. *Physiol. Measure.* 27:425. doi: 10.1088/0967-3334/27/4/008

Liu, M., Wu, W., Gu, Z., Yu, Z., Qi, F., and Li, Y. (2018). Deep learning based on batch normalization for p300 signal detection. *Neurocomputing* 275, 288–297. doi: 10.1016/j.neucom.2017.08.039

Liu, Y., Zhou, Z., Hu, D., and Dong, G. (2005). "T-weighted approach for neural information processing in p300 based brain-computer interface," in *2005 International Conference on Neural Networks and Brain, Vol. 3* (Changsha), 1535–1539.

Ma, X., Qiu, S., Du, C., Xing, J., and He, H. (2018). "Improving EEG-based motor imagery classification *via* spatial and temporal recurrent neural networks," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Beijing), 1903–1906. doi: 10.1109/EMBC.2018.8512590

Manor, R., and Geva, A. B. (2015). Convolutional neural network for multi-category rapid serial visual presentation BCI. *Front. Comput. Neurosci.* 9:146. doi: 10.3389/fncom.2015.00146

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

Peterson, V., Nieto, N., Wyser, D., Lambercy, O., Gassert, R., Milone, D. H., et al. (2021). Transfer learning based on optimal transport for motor imagery brain-computer interfaces. *IEEE Trans. Biomed. Eng.* 69, 807–817. doi: 10.1109/TBME.2021.3105912

Rakotomamonjy, A., and Guigue, V. (2008). BCI competition III: dataset II-ensemble of SVMs for BCI P300 speller. *IEEE Trans. Biomed. Eng.* 55, 1147–1154. doi: 10.1109/TBME.2008.915728

Sakhavi, S., Guan, C., and Yan, S. (2018). Learning temporal information for brain-computer interface using convolutional neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* 29, 5619–5629. doi: 10.1109/TNNLS.2018.2789927

Shan, H., Liu, Y., and Stefanov, T. P. (2018). "A simple convolutional neural network for accurate P300 detection and character spelling in brain computer interface," in *IJCAI* (Leiden), 1604–1610. doi: 10.24963/ijcai.2018/222

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

Tangermann, M., Müller, K.-R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., et al. (2012). Review of the BCI competition IV. *Front. Neurosci.* 2012:55. doi: 10.3389/fnins.2012.00055

Tao, W., Li, C., Song, R., Cheng, J., Liu, Y., Wan, F., and Chen, X. (2020). Eeg-based emotion recognition *via* channel-wise attention and self-attention. *IEEE Trans. Affect. Comput.* doi: 10.1109/TAFFC.2020.3025777

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). "Attention is all you need," in *Advances in Neural Information Processing Systems, 30* (California, CA). 1–1.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3

Wu, H., Niu, Y., Li, F., Li, Y., Fu, B., Shi, G., et al. (2019). A parallel multiscale filter bank convolutional neural networks for motor imagery EEG classification. *Front. Neurosci.* 13:1275. doi: 10.3389/fnins.2019.01275

Yang, Y., Wu, Q., Fu, Y., and Chen, X. (2018a). "Continuous convolutional neural network with 3D input for EEG-based emotion recognition," in *International Conference on Neural Information Processing* (Xiamen: Springer), 433–443. doi: 10.1007/978-3-030-04239-4_39

Yang, Y., Wu, Q., Qiu, M., Wang, Y., and Chen, X. (2018b). "Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network," in *2018 International Joint Conference on Neural Networks (IJCNN)* (Xiamen), 1–7. doi: 10.1109/IJCNN.2018.8489331

Zhang, C., Kim, Y.-K., and Eskandarian, A. (2021). EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification. *J. Neural Eng.* 18:046014. doi: 10.1088/1741-2552/abed81

Zhang, D., Chen, K., Jian, D., and Yao, L. (2020). Motor imagery classification *via* temporal attention cues of graph embedded EEG signals. *IEEE J. Biomed. Health Informatics* 24, 2570–2579. doi: 10.1109/JBHI.2020.2967128

Zhang, D., Yao, L., Zhang, X., Wang, S., Chen, W., Boots, R., et al. (2018). "Cascade and parallel convolutional recurrent neural networks on EEG-based intention recognition for brain computer interface," in *Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32* (Sydney). doi: 10.1609/aaai.v32i1.11496

Zhang, K., Robinson, N., Lee, S.-W., and Guan, C. (2021). Adaptive transfer learning for EEG motor imagery classification with deep convolutional neural network. *Neural Netw.* 136, 1–10. doi: 10.1016/j.neunet.2020.12.013

Zhang, Y., Zhou, Z., Bai, H., Liu, W., and Wang, L. (2020). Seizure classification from EEG signals using an online selective transfer tsk fuzzy classifier with joint distribution adaption and manifold regularization. *Front. Neurosci.* 14:496. doi: 10.3389/fnins.2020.00496

Zhao, H., Zheng, Q., Ma, K., Li, H., and Zheng, Y. (2020). Deep representation-based domain adaptation for nonstationary EEG classification. *IEEE Trans. Neural Netw. Learn. Syst.* 32, 535–545. doi: 10.1109/TNNLS.2020.3010780

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 13 January 2023
DOI 10.3389/fnins.2022.1097660

OPEN ACCESS

EDITED BY
Ernest N. Kamavuako,
King's College London, United Kingdom

REVIEWED BY
Jing Jin,
East China University of Science and Technology, China
Sun Hao,
East China University of Science and Technology, China
Banghua Yang,
Shanghai University, China

*CORRESPONDENCE
Chao Feng
✉ cfeng@qlu.edu.cn
Jiancai Leng
✉ jiancaileng@qlu.edu.cn
Yang Zhang
✉ zhangyang982003@163.com
Fangzhou Xu
✉ xfz@qlu.edu.cn

SPECIALTY SECTION
This article was submitted to Neuroprosthetics, a section of the journal Frontiers in Neuroscience

RECEIVED 14 November 2022
ACCEPTED 28 December 2022
PUBLISHED 13 January 2023

CITATION
Li H, Liu M, Yu X, Zhu J, Wang C, Chen X, Feng C, Leng J, Zhang Y and Xu F (2023) Coherence based graph convolution network for motor imagery-induced EEG after spinal cord injury.
Front. Neurosci. 16:1097660.
doi: 10.3389/fnins.2022.1097660

Coherence based graph convolution network for motor imagery-induced EEG after spinal cord injury

Han Li[1], Ming Liu[1], Xin Yu[1], JianQun Zhu[1], Chongfeng Wang[1], Xinyi Chen[1], Chao Feng[1]*, Jiancai Leng[1]*, Yang Zhang[2]* and Fangzhou Xu[1]*

[1]International School for Optoelectronic Engineering, Qilu University of Technology, Shandong Academy of Sciences, Jinan, China, [2]Rehabilitation Center, Qilu Hospital of Shandong University, Jinan, China

Background: Spinal cord injury (SCI) may lead to impaired motor function, autonomic nervous system dysfunction, and other dysfunctions. Brain-computer Interface (BCI) system based on motor imagery (MI) can provide more scientific and effective treatment solutions for SCI patients.

Methods: According to the interaction between brain regions, a coherence-based graph convolutional network (C-GCN) method is proposed to extract the temporal-frequency-spatial features and functional connectivity information of EEG signals. The proposed algorithm constructs multi-channel EEG features based on coherence networks as graphical signals and then classifies MI tasks. Different from the traditional graphical convolutional neural network (GCN), the C-GCN method uses the coherence network of EEG signals to determine MI-related functional connections, which are used to represent the intrinsic connections between EEG channels in different rhythms and different MI tasks. EEG data of SCI patients and healthy subjects have been analyzed, where healthy subjects served as the control group.

Results: The experimental results show that the C-GCN method can achieve the best classification performance with certain reliability and stability, the highest classification accuracy is 96.85%.

Conclusion: The proposed framework can provide an effective theoretical basis for the rehabilitation treatment of SCI patients.

KEYWORDS

electroencephalogram, motor imagery, brain-computer interface, coherence-based graph convolutional network, spinal cord injury

1. Introduction

Spinal cord injury (SCI) is a catastrophic disease, which can lead to the loss of motor and sensory functions. In severe cases, it can lead to the interruption of some routes connecting the brain and limbs. Many SCI patients experience chronic pain that is difficult to treat (Jensen et al., 2005; Cardenas and Jensen, 2006). There are more than 3.7 million SCI patients in China, with an annual incidence of about 90,000 new patients per year, and the annual incidence rate is 17.9 to 60.2 people/million people. The motor dysfunction caused by SCI not only brings serious physical and psychological harm to the patients themselves but also imposes a huge economic burden on society and families (Aguilar et al., 2010). To reduce the harm caused by

SCI patients, researchers have explored the changes in the brain of SCI patients after chronic injury. Studies have shown that chronic pain associated with SCI may be related to changes in brain activity reflected in the electroencephalogram (EEG) (López-Larraz et al., 2012), the differences in EEG reflect some extent the experience of pain. At present, the EEG studies after SCI have the following aspects, the changes in event-related synchronization/desynchronization (ERD/ERS) after SCI, the changes in power spectrum occurring after SCI, the changes in network characteristics of brain networks after SCI, the changes in performance of brain-computer interface (BCI) systems after SCI.

In recent years, motor imagery (MI)-based BCI systems have become the focus of attention in the field of rehabilitation medicine, such as neuro-robotics and neuro-prosthetic device control (Iturrate et al., 2009; Millán et al., 2010; Escolano et al., 2011). The MI-BCI system aims to deliver MI task interventions for SCI patients and to assist in the formulation of rehabilitation programs to alleviate patient suffering. As shown in **Figure 1**, MI-based BCI systems mainly consist of four parts, signal acquisition, signal processing, application, and feedback, use brain signals to control external assistive devices (Collinger et al., 2014). Traditional SCI rehabilitation training, which lacks the active participation of the patient and the reconstruction of neural pathways is slow, is mainly based on the passive movement by a patient to achieve the recovery of muscle strength and the reconstruction of neural pathways. BCI technology for SCI rehabilitation takes into account the functional coupling between the patient's MI intentions and the actual motor effects. It is more in line with the theoretical requirements of neurological reconstruction and can promote faster and better motor recovery in SCI patients. In the study of MI-based BCI systems for SCI patients, Müller-Putz et al. (2014) achieved an average accuracy of 66.1% using common spatial pattern (CSP) algorithms and linear discriminant analysis (LDA) for classification. Xu et al. (2022b) proposed Modified Graph Convolutional Neural Network (M-GCN) method, which performs time-frequency processing of data by modified S-transform (MST) to improve decoding performance.

MI refers to the procedure of imagining limb actions without actual limb movements (Zhang et al., 2018). Related studies have shown that the sensorimotor cortical areas of the brain stimulated by MI are the same as those stimulated by actual limb movement (Azab et al., 2019). MI is considered to be a process involving multiple higher cognitive functions (Rimbert et al., 2019). The brain is a complex network, the information related to MI is both spatially independent and interconnected, and brain network correlation methods can investigate the functional mechanisms of MI. EEG coherence provides an important estimate of the functional interactions between neural systems in each frequency band and is often used to assess the functional connectivity of the human cortex (Srinivasan et al., 2007). Due to its targeting of functional mechanisms, more and more people have begun to pay attention to coherence networks and have used them to decode relevant cognitive functions. Benefiting from MI therapy, patients with cortical damage have better performance in functional recovery, and BCI investigators achieve higher classification accuracy (Weiskopf et al., 2004).

Hinton and Salakhutdinov (2006) published a paper in Science on the dimensionality reduction of data with neural networks, which attracted great attention. AlexNet performed brilliantly in the ImageNet image recognition competition (Krizhevsky et al., 2017), which started the boom of deep learning (DL). Relying on advances in various aspects such as large data volume, non-convex optimization, hardware computation, and network structure, DL methods represented by convolutional neural networks (CNN), recurrent neural networks (RNN), and generative adversarial networks (GAN) achieved excellent results in the processing of regular data such as images, audio, video, and text. Deep neural networks have achieved great success in data processing, more and more people have begun to apply them to BCI systems. In 2022, Roy (2022) proposed a multi-scale CNN (MS-CNN) model with intrinsic feature integration for motor image EEG subject classification in the BCI system. Supakar et al. (2022) used the recurrent neural networks-Long Short-Term Memory (CNN-LSTM) method to analyze EEG signal data to diagnose schizophrenia. Xu et al. (2022a) applied the deep convolution generative adversarial network (DCGAN) to rehabilitation-based BCI systems.

Deep learning frameworks generally have large models and many parameters. They need higher amounts of training and requirements for computing conditions. How to extend DL methods to irregular data structures is a current research hotspot in the field of neural networks. Data processing based on graph structures mainly involves the representation learning of graph nodes, classification of graph nodes, prediction of edges in graphs, classification of graphs, and so on. Irregular data represented by graphs, such as traffic flow networks with cities as nodes, molecular structure networks with various types of atoms as nodes, and EEG structure networks with electrodes as nodes, are playing an increasingly important role in the storage of data and the description of relationships between entities. To efficiently extract space features on this data structure, graph convolutional network (GCN) is proposed. Chen et al. (2020) proposed the E-GCN method to deeply mine the relationship between EEG channels and to use it for the detection of epileptic EEG signals. Zeng et al. (2020) proposed the hierarchical graph convolution (HGCN) network for classification tasks using topological relationships between each electrode, where power spectral density and continuous wavelet transform features from the raw EEG signal are used as frequency domain inputs.

The above methods only consider EEG channel location relationships and do not explore functional linkages. Considering the working mechanism for the division of labor and cooperation between brain regions, the spatial location relationships and functional linkages of EEG channels do not maintain their consistency (Wang et al., 2019). In this paper, the coherence network-based graph convolution (C-GCN) method is proposed to analyze MI-based EEG data, the main contributions are as follows,

(1) Due to the fact that traditional GCN can only analyze the spatial relationship of channels but not describe the connection of brain functions, the C-GCN algorithm is proposed to represent the temporal-frequency-spatial domain representation of EEG data.
(2) Compared with Support Vector Machine (SVM), CNN, EEGNet, RNN, LSTM, traditional GCN, M-GCN, Graph Attention Network (GAT), and ResGCN, the proposed C-GCN algorithm can obtain the best performance of 96.85% for two-class MI recognition.
(3) The coherence network of EEG data at different frequency bands from SCI patients and healthy subjects is used to perform functional analysis and to provide rehabilitation training guidance for SCI patients.

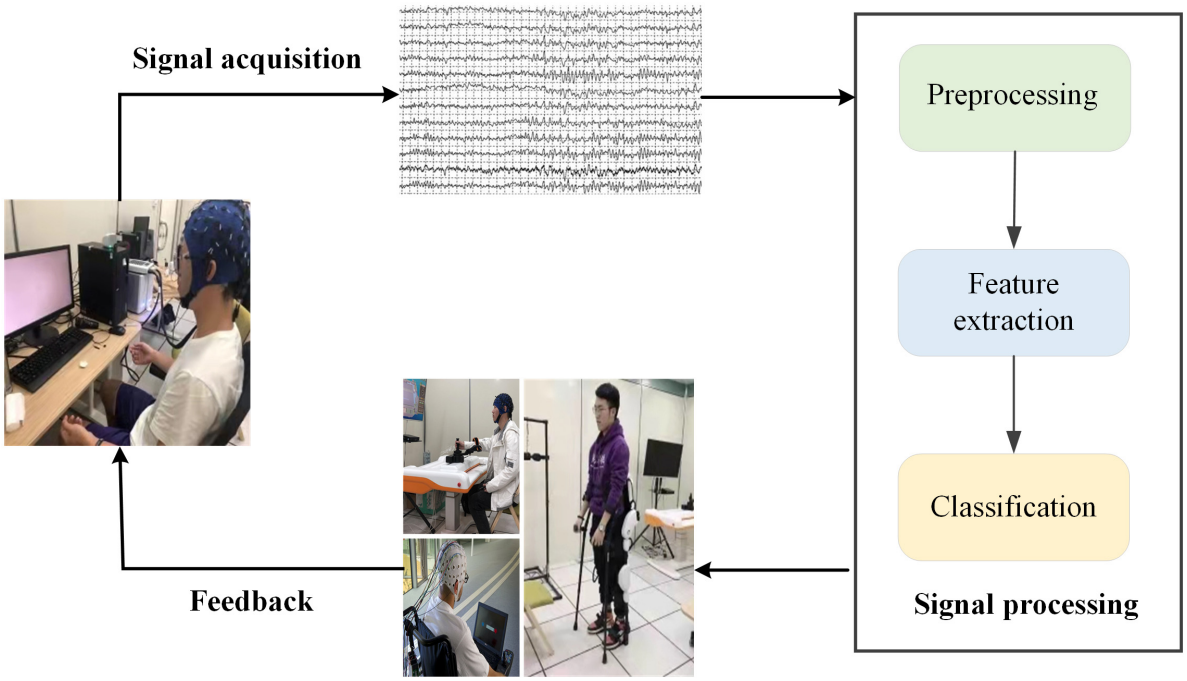


FIGURE 1
Block diagram of motor imagery (MI)-based brain-computer interface (BCI) system. The system consists of four main parts: signal acquisition, signal processing, application, and feedback.

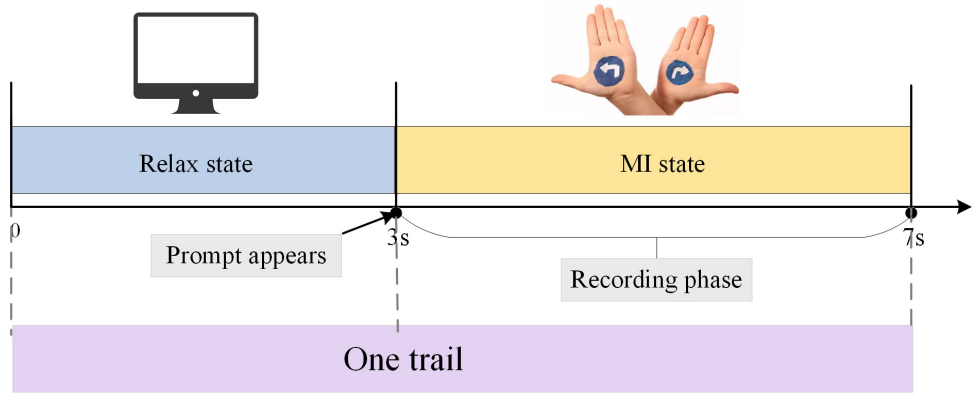


FIGURE 2
The procedure for one electroencephalogram (EEG) signal acquisition trial. The duration of one trial is 7 s, including 3 s of rest state and 4 s of MI state.

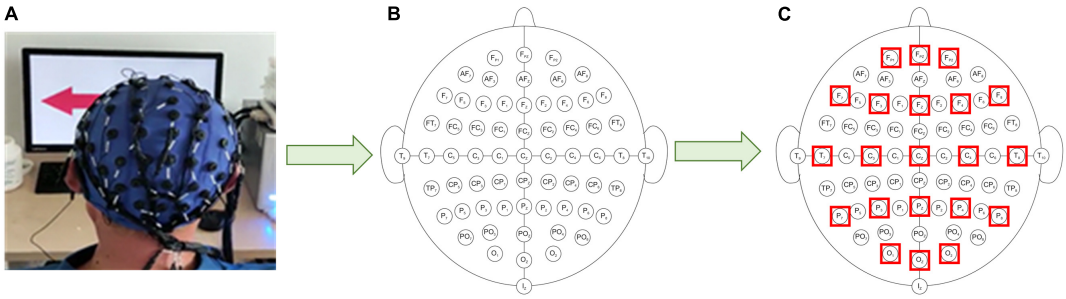


FIGURE 3
Schematic diagram of electrode position distribution. **(A)** Three-dimensional electrode position distribution. **(B)** Two-dimensional electrode position distribution. **(C)** Experimentally selected electrode position distribution.

The rest content is arranged as follows, section "2. Experimental data and experimental paradigm" introduces the experimental data. Section "3. Materials and methods" introduces the pre-processing work and the C-GCN model. Section "4. Results and discussions" shows the experimental results and discussions. Finally, section "4. Conclusion" summarizes the whole paper.

2. Experimental data and experimental paradigm

The EEG data used in the experiment were collected from the Department of Physical Medicine and Rehabilitation, Qilu Hospital, Shandong University, and the protocol of this experiment was approved by the Medical Ethics Committee of Qilu Medical College, Shandong University [No. KYLL-2020(KS)-475]. Before the experiment, all subjects signed an informed consent form and were free from habitual medication, alcohol consumption, and cognitive impairment. Experiments were carried out in a closed environment where subjects were undisturbed and attentive, E-Prime software was used for MI stimulation, and 64-lead EEG signal acquisition system was used to capture the subject's EEG signals. Twenty-five subjects were recruited for the experiment, including 18 SCI patients and 7 healthy subjects, the healthy subjects serving as controls in the experiment.

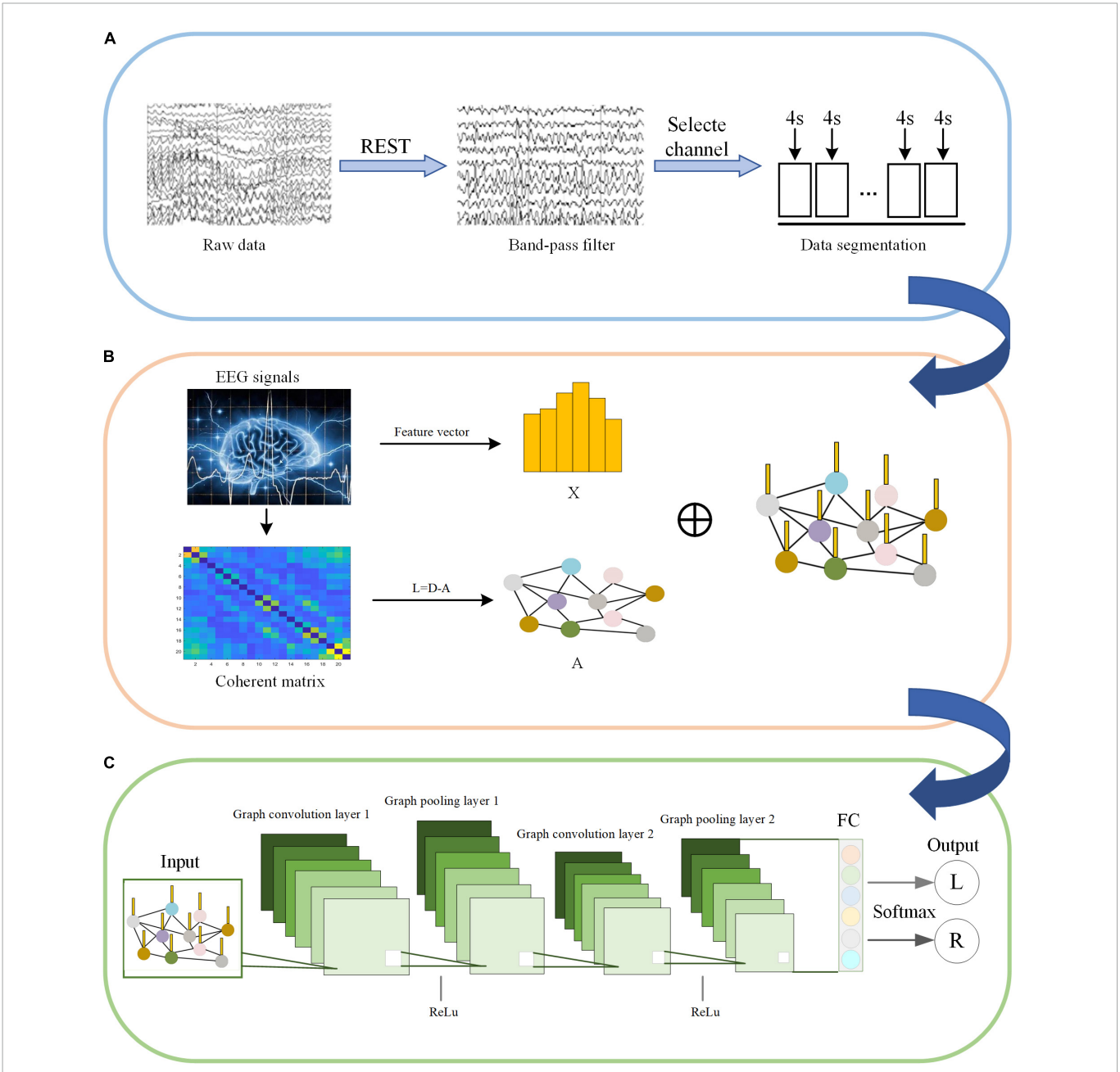


FIGURE 4
Motor imagery (MI) pattern recognition framework based on the coherence-based graph convolutional network (C-GCN) model. **(A)** Pre-processing of EEG. **(B)** Construction of graphical signals. **(C)** Specific GCN structure.

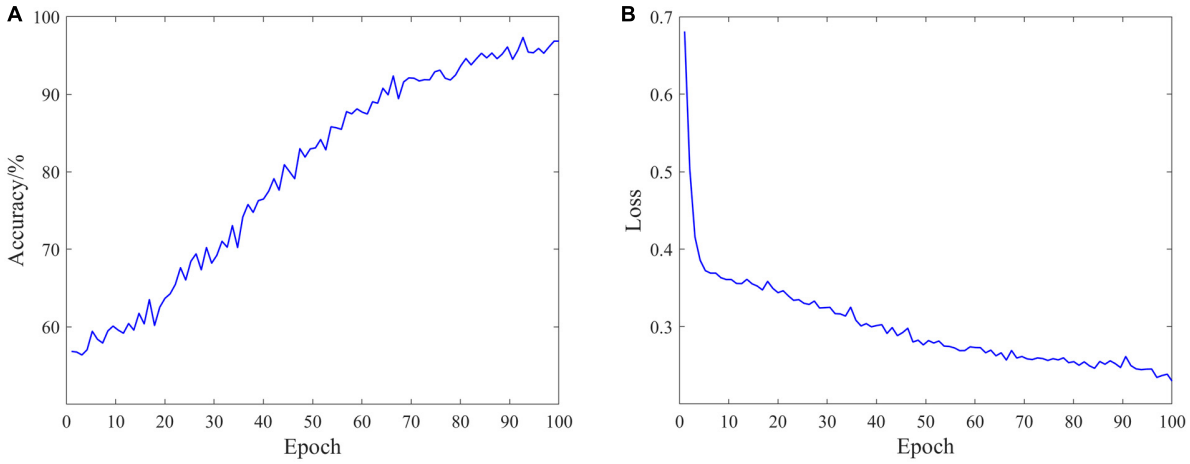


FIGURE 5
Classification performance of the coherence-based graph convolutional network (C-GCN) model. **(A)** Accuracy distribution of the C-GCN model. **(B)** Loss distribution of the C-GCN model.

During the MI experiment, the subjects sat in front of the instruction screen. Before each imaginary movement, the screen was blank, which was a rest period to prevent visually evoked potentials, that is, the time interval between two imaginary movements. After the left and right arrow prompts appeared, the MI tasks began. During this time, subjects began to imagine themselves performing left-hand or right-hand movements, and the duration of the imagined movements was 4 s. The experimental paradigm is shown in **Figure 2**. Each group of experiments consisted of 20 randomly occurring MI tasks, and each subject performed four groups of experiments with a 90 s rest period between every two groups. Each subject performed 80 trials, 40 each of the left-hand MI tasks and the right-hand MI tasks, and the emergence of the left-hand and right-hand MI tasks was randomized.

3. Materials and methods

3.1. Pre-processing

Electroencephalogram data are pre-processed before being fed into the C-GCN model. Zero-reference processing is performed by the reference electrode normalization technique (REST) (Xu et al., 2014) to obtain artifact-free data. EEG signals are filtered by a fifth-order Butterworth filter with 8–30 Hz to remove noise. Channel selection and data segmentation techniques are also involved in pre-processing procedure.

MI-related information is generally concentrated in specific frequency bands, therefore during the EEG data filtering process, the EEG data are divided into multiple data bands (μ rhythm, β rhythm, μ, and β rhythms). To reduce the effect of volume conduction between network nodes, 21 electrodes of 64 electrodes ("Fp1", "Fp2", "F7", "F3", "Fz", "F4", "F8", "T7", "Cz", "C4", "T8", "P7", "P3", "Pz", "P4", "P8", "O1", "O2", and "Oz") are selected for subsequent processing (Li et al., 2016), and the electrode positions select for the experiment are shown in **Figure 3**. Resting-state EEG shows spontaneous brain activity in the idle state, whereas MI-state EEG records the event-related activity during the desired task. One trial consists of 3 s resting state and 4 s MI state. The MI state of EEG is employed for the analysis of coherence network.

3.2. Coherence

The pre-processed EEG data are employed for coherence network construction. Coherence is the squared correlation coefficient (Zhang et al., 2015), which characterizes the connectivity between the brain network channels of the MI tasks. Coherence, which is the degree of linear correlation between two EEG signals $x(t)$ and $y(t)$ at specific frequencies, is used to measure the strength of the interaction between each pair of electrodes (Weiss and Mueller, 2003; Murias et al., 2007). High coherence between the two EEG electrodes indicates the contribution of synchronized neuronal oscillations to each electrode, indicating functional integration between neural populations. Low coherence indicates functional separation. The coherence coefficients of the EEG signals $x(t)$ and $y(t)$ are defined as,

$$C_{xy} = \frac{|P_{xy}(f)|^2}{P_{xx}(f)P_{yy}(f)} \quad (1)$$

where $P_{xy}(f)$ is the cross-spectral density of $x(t)$ and $y(t)$, $P_{xx}(f)$ and $P_{yy}(f)$ are the self-spectral densities of $x(t)$ and $y(t)$, respectively. $C_{xy}(f)$ is the coherence at frequency f.

After the coherence calculation, the coherence is averaged over the corresponding frequency band to obtain the final strength of the connection between the two nodes. The coherence coefficient takes values in the range of 0–1. When the coherence coefficient is closer to 1, the two signals are more coherent. The coherence network has 21 nodes due to the selected 21 channels of the subject. Therefore, the EEG coherence network is constructed by a 21 × 21 weighting matrix.

3.3. C-GCN

In traditional convolutional networks, convolution essentially uses a filter with shared parameters to extract spatial features by

TABLE 1 Parameters values and experimental results of the coherence-based graph convolutional network (C-GCN) model.

Label	Parameters	Value
1	Num_epochs	100
2	Batch_size	512
3	Regularization	0.001
4	Dropout	0.50
5	Learning_rate	0.01
6	Accuracy	96.85%
7	Loss	0.23
8	F1-Score	96.78%

computing a weighted sum of the central pixel and neighboring pixels. Convolution is an operation between signals on a regular grid. With the production of discrete data in the spatial domain, a graphical representation is proposed. The properties of graphs are studied using the eigenvalues and eigenvectors of the Laplacian matrix of the graph, extending DL techniques to the domain of graphs. The graph can be defined as,

$$G = (V, E, A) \tag{2}$$

where V is the set of nodes, E is the set of edges, and A is the adjacency matrix of the graph.

Let $v_i \in V$ denote a node and $e_{ij \notin E}$ denote an edge from v_i to v_j. The neighborhood of node v is defined as,

$$N(v) = \{u \in V \,|(v, u) \in E\} \tag{3}$$

The adjacency matrix A is the diagonal matrix $n \times n$. The Laplacian matrix of a graph is defined as,

$$L = D - A \tag{4}$$

where L is the Laplacian matrix and D is the degree matrix of the graph.

Since L is a symmetric matrix, it can be singular value decomposed (SVD) (Spielman, 2007), as follow,

$$L = U \wedge U^T \tag{5}$$

where $U = [u_0, \cdots, u_{N-1}] \in R_{N \times N}$ is the eigenvector matrix, $\wedge = diag([\lambda_0, \cdots, \lambda_N])$ is the diagonal matrix.

GCN can be divided into two types of convolution including spectral convolution and spatial domain convolution. Spectral convolution is to filter both the convolutional network and graphical signals into the Fourier domain and then process them. Spatial domain convolution is to connect the nodes of the graph in the spatial domain, implement a hierarchy, and then perform convolution.

The spectral convolution of the graph signal is defined as,

$$g_\theta * x = U g_\theta U^T x \tag{6}$$

where $x \in R^N$, the filter is defined as $g_\theta = diag(\theta)$, $\theta \in R^N$ is parameter in the Fourier domain. U is consist of the eigenvectors from the normalized Laplacian matrix, U is defined as,

$$L = I_N - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} = U \wedge U^T \tag{7}$$

where $\wedge$ is a diagonal matrix, which is consist of the eigenvalues of the Laplace matrix, U^T is the Fourier transform of the graph.

To locate the filter in space and reduce its computational complexity, the filter is approximated using a truncated expansion

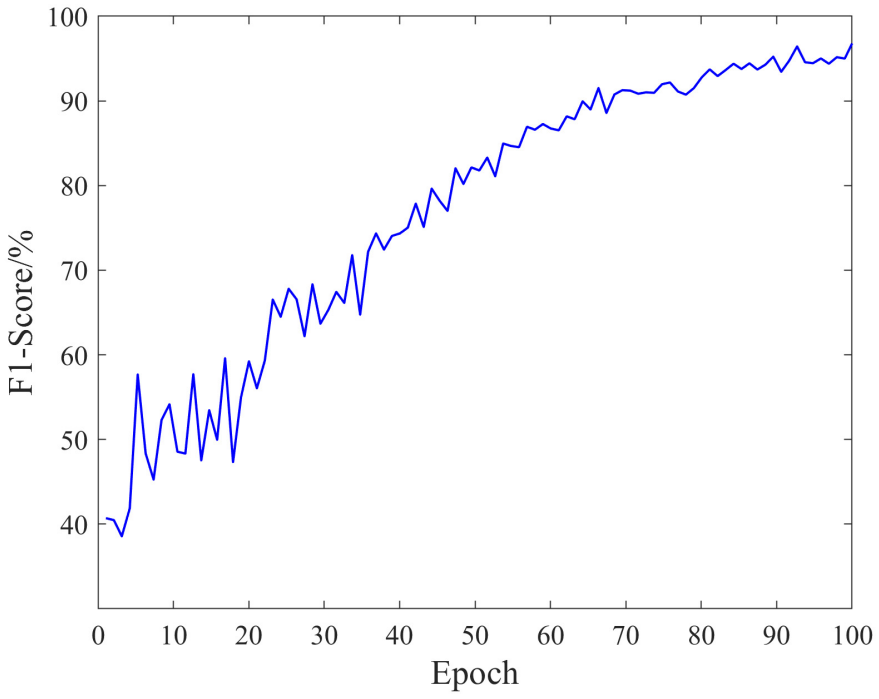


FIGURE 6
F1-Score of the coherence-based graph convolutional network (C-GCN) model.

TABLE 2 Classification accuracy of spinal cord injury (SCI) patients in the coherence-based graph convolutional network (C-GCN) model at different rhythms.

Subjects	Method	Accuracy%		
		μ rhythm	β rhythm	μ and β rhythms
SCI_1	C-GCN	96.45	96.68	97.33
SCI_2	C-GCN	96.63	93.75	97.58
SCI_3	C-GCN	95.15	97.23	96.75
SCI_4	C-GCN	95.33	96.50	97.15
SCI_5	C-GCN	97.43	96.65	98.08
SCI_6	C-GCN	96.38	96.25	96.50
SCI_7	C-GCN	96.65	98.00	97.25
SCI_8	C-GCN	96.50	96.33	96.88
SCI_9	C-GCN	96.00	97.75	98.23
SCI_10	C-GCN	97.15	98.00	97.78
SCI_11	C-GCN	96.75	97.78	97.75
SCI_12	C-GCN	93.88	92.15	98.50
SCI_13	C-GCN	96.23	97.63	97.00
SCI_14	C-GCN	91.50	92.78	93.63
SCI_15	C-GCN	96.78	97.50	97.25
SCI_16	C-GCN	90.33	92.00	95.08
SCI_17	C-GCN	93.45	96.43	97.63
SCI_18	C-GCN	95.78	97.08	97.78
Average	C-GCN	95.47	96.14	97.12
Standard deviation	C-GCN	1.91	1.96	1.13

of a K-order Chebyshev polynomial (Defferrard et al., 2016). The Chebyshev polynomial is defined as,

$$T_k(x) = 2xT_{K-1}(x) - T_{K-2}(x) \tag{8}$$

where $T_0(x) = 1$, $T_1(x) = x$. Then, the signal x is filtered by a k-domain filter y, which is defined as,

$$y = g_\theta(L) * x = \sum_{k=0}^{k} \theta_k T_k(\tilde{L}) x \tag{9}$$

where $\tilde{L} = 2L/\lambda_{max} - I_N\lambda_{max}$ represents the largest eigenvalue of L.

C-GCN is a model that combine coherence network with GCN. The framework consists of two main modules, including the construction of coherence-based graphical signals and pattern recognition for GCN. Before feeding into the C-GCN model, EEG data are first pre-processed as shown in Figure 4A. The input graphic signals of the C-GCN model integrate the temporal-frequency-spatial features from EEG data as shown in Figure 4B graphic signals of EEG data are implemented by formula (1) (9). After the graphic signals are constructed, EEG feature data is performed through two graph convolution layers, two Relu layers (Glorot et al., 2011), two graph pooling layers (Ouhmich et al., 2019), and one fully connected layer (Xu et al., 2022c) to complete the MI tasks classification as shown in Figure 4C.

The input of the C-GCN model is the pre-processed EEG time series. The X in Figure 4B represents the temporal-frequency features of EEG, the vertices in Figure 4B represent the EEG channels, and the edges connecting the vertices represent the coherence connectivity between electrodes. The purpose of performing graph convolution operations is to extract more discriminative EEG features. The graph convolutional layer is the core layer of C-GCN. To increase the non-linearity of the C-GCN model, the Relu function (Nair and Hinton, 2010) is used to mitigate the appearance of fitting problems. When the strength of the information is greater than a certain threshold, the valve is opened and the information is passed, otherwise, the valve is closed and the information is discarded. Graph pooling is a necessary module for GCN to perform classification, this module aggregates the previous results to obtain a smaller-scale representation of the graph. Graph pooling can be described as follows,

$$P : (G = (V, E, A), Y) \rightarrow (G' = (V', E', A'), Y') \tag{10}$$

where the array of graph G and corresponding node feature matrix Y are mapped to a smaller array of graph G′ and corresponding node feature matrix Y′. In a multilayer GCN, the operation of the pooling layer is correspondingly expressed as,

$$P : (G_l = (V_l, E_l, A_l), \mathrm{H}^{(l)}) \rightarrow (G_{l+1} = (V_{l+1}, E_{l+1}, A_{l+1}), \mathrm{H}^{(l+1)}) \tag{11}$$

Then, a fully connected (FC) output layer is employed for integrating global information from the graphs of the previous localization filters. Finally, the Softmax function (Han et al., 2018) is used for classification and recognition. In the C-GCN model, the cross-entropy loss function is used to optimize the network parameters, and the cross-entropy loss is expressed as follows,

$$Loss = -\sum_{x}(p(x)\log q(x) + (1 - p(x))\log(1q(x))) + \lambda K(w) \tag{12}$$

where $p(x)$ denotes the true value of the training data, $q(x)$ denotes the predicted value of the training data, $K(w)$ is used to evaluate the model complexity and $\lambda K(w)$ is aimed at preventing over fitting of the model. In summary, the EEG data based on the MI tasks are trained and tested in C-GCN to obtain classification recognition results. Algorithm 1 is a summary of the classification training steps of the C-GCN model.

Require: the pre-processed EEG signals, the class labels corresponding to the EEG signals, the numbers of Chebyshev polynomial order k;
Ensure: The desired model parameters of C-GCN;
for pre-processed EEG signals **do**
  $C_{xy} = \frac{|P_{xy}(f)|^2}{P_{xx}(f)P_{yy}(f)}$
end for
for graphical signals **do**
  $L = I_N - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$
  $\tilde{L} = 2L/\lambda_{max} - I_N\lambda_{max}$
  $T_k(\tilde{L})(k = 0, 1, \cdots, K-1)$
  $\sum_{K=0}^{K-1}\theta_k T_k(\tilde{L})x$
end for
Calculate the results of convolution after activation, FC layer, and Loss.

Algorithm 1. Training procedure of the C-GCN model.

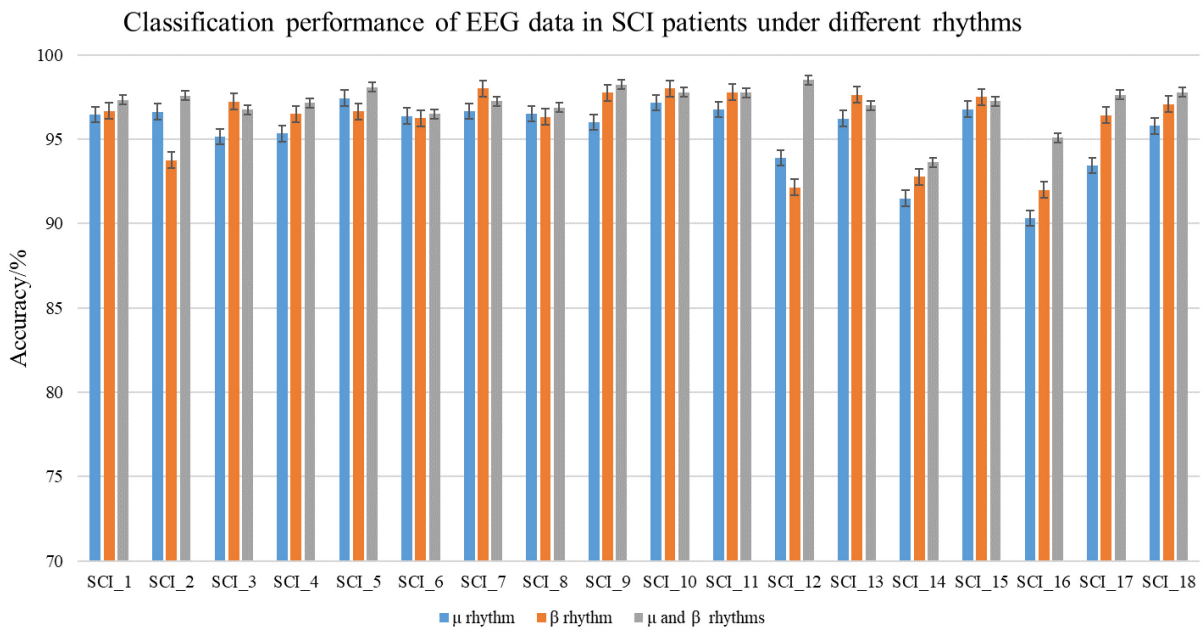


FIGURE 7
Classification performance of spinal cord injury electroencephalogram (SCI EEG) in the coherence-based graph convolutional network (C-GCN) model. This includes the classification accuracy under μ rhythm, β rhythm, μ and β rhythms.

4. Results and discussions

4.1. C-GCN

The classification performance is an important measure of data quality, and can also provide ideas for the rehabilitation of SCI patients. In the experiment, EEG data of 18 SCI patients have been trained and validated on the proposed C-GCN model with the cross-validation method. The 90% of the data set have been employed for training and the 10% of the data set have been used for validation. The parameters of the model and the experimental results are shown in Table 1. In the C-GCN model, the accuracy of SCI patients can achieve 96.85%. The accuracy and Loss are shown in Figure 5, the F1-Score is shown in Figure 6. The experimental results show that the C-GCN model has a high signal-to-noise ratio, good adaptability, and robustness to individual specificity. The model can guide the rehabilitation training and subsequent treatment of SCI patients.

To verify the difference of EEG data in MI tasks at different frequencies and the superiority of the C-GCN model, this paper also conducted experiments on μ rhythm, β rhythm, μ and β rhythms of 18 SCI patients respectively, the experimental results are shown in Table 2. In Figure 7, good classification accuracy can be obtained in μ rhythm, β rhythm, μ and β rhythms. There are slight differences in the classification results at different frequencies. In the vast majority of these cases, the classification accuracy of the MI tasks under μ and β rhythms is higher than that under μ or β rhythms alone, and the classification results of the MI tasks under β rhythms are higher than those under μ rhythms. This result indicates that the μ and β rhythms contains more information on MI in SCI patients, the MI information in the μ rhythm is less than that in the β rhythm. The information contained in the μ and β rhythms is more valuable for the rehabilitation research of SCI patients. The intra-individual classification accuracy of SCI patients is not significantly different from the overall classification accuracy. Table 2 indicates that the proposed C-GCN model has very strong adaptability and can mitigate the effects due to individual differences and the number of data samples.

In the experiment, healthy subjects as the control group. The classification accuracy of the EEG data under different rhythms in the C-GCN model for healthy subjects in Table 3. To observe the experimental results of the healthy subjects more visually, the classification results are presented in the form of histograms. In Figure 8, the experimental results of healthy subjects at different rhythms have the same regularity as those of SCI patients. The highest classification accuracy is obtained at the μ and β rhythms, followed by the second highest classification accuracy at the β rhythm and

TABLE 3 Classification accuracy of healthy subjects in the coherence-based graph convolutional network (C-GCN) model at different rhythms.

Subjects	Method	Accuracy%		
		μ rhythm	β rhythm	μ and β rhythms
Sub_1	C-GCN	88.33	90.78	94.63
Sub_2	C-GCN	94.00	96.33	97.50
Sub_3	C-GCN	93.88	95.43	95.75
Sub_4	C-GCN	96.23	97.67	98.38
Sub_5	C-GCN	96.67	97.38	97.63
Sub_6	C-GCN	91.68	94.50	95.25
Sub_7	C-GCN	96.15	96.88	97.63
Average	C-GCN	93.85	95.57	96.68
Standard deviation	C-GCN	2.78	2.21	1.34

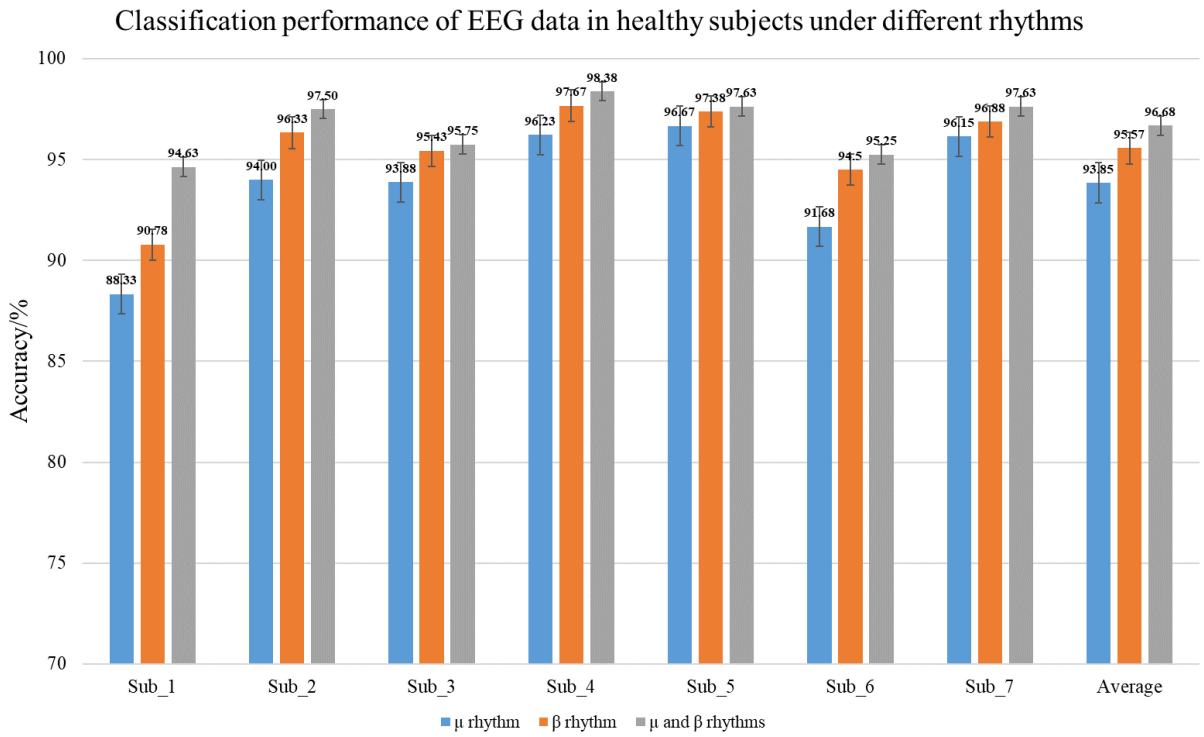


FIGURE 8
Classification performance of healthy subjects electroencephalogram (EEG) in the coherence-based graph convolutional network (C-GCN) model. This includes the classification accuracies under μ rhythm, β rhythm, μ and β rhythms.

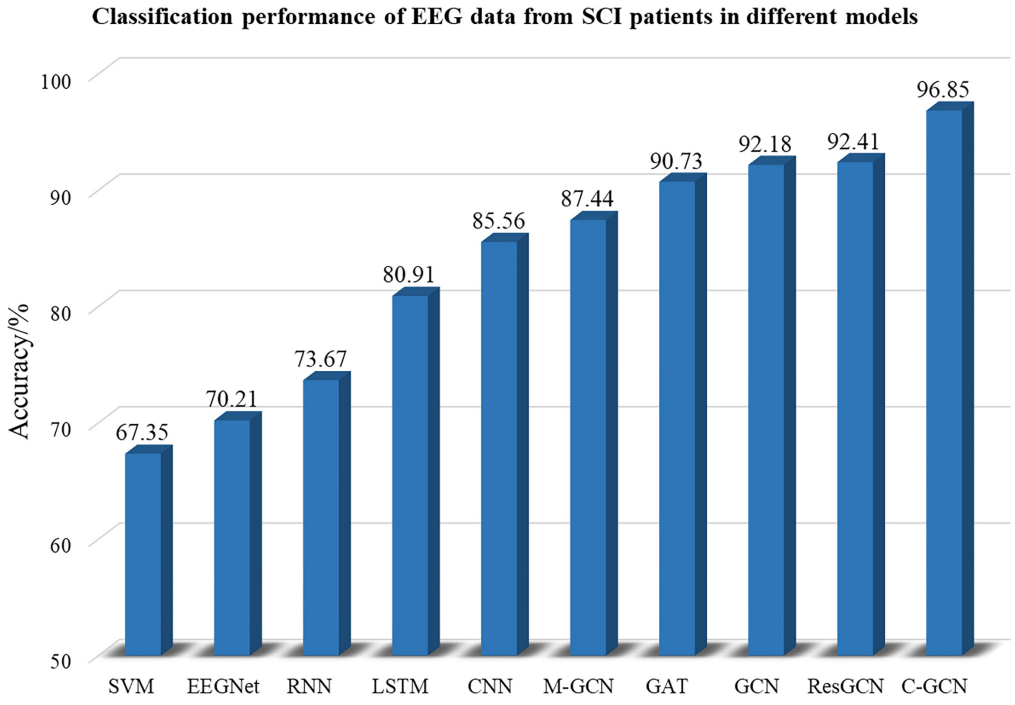


FIGURE 9
Classification performance of spinal cord injury electroencephalogram (SCI EEG) in different classification models. These include SVM, EEGNet, RNN, LSTM, CNN, GCN, M-GCN, GAT, ResGCN, and C-GCN models.

the lowest at the μ rhythm. Analysis of the mean accuracies revealed that the classification accuracy of SCI patients is slightly higher than that of healthy subjects at either rhythm and that the difference in accuracy between rhythms is lower in SCI patients than in healthy subjects. Combined with the self-assessment form of the subjects' EEG acquisition procedure and the SCI pathology analysis, it is found

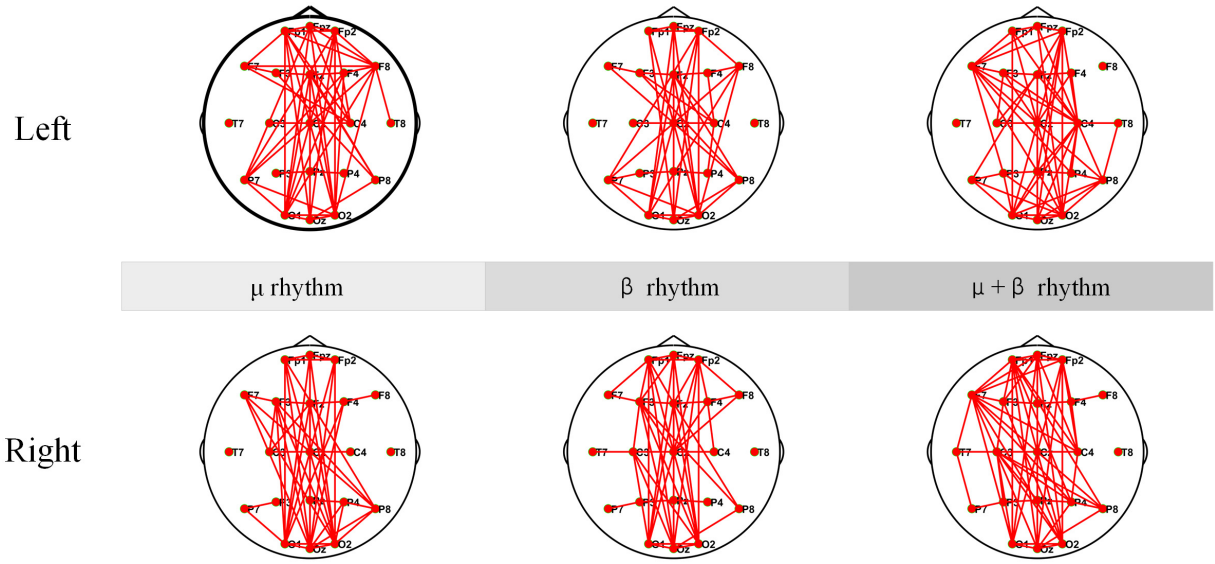


FIGURE 10
Coherence connectivity maps of motor imagery (MI) tasks in different frequency bands for spinal cord injury electroencephalogram (SCI EEG). These include left-hand and right-hand coherence connections under μ rhythm, β rhythm, μ and β rhythms.

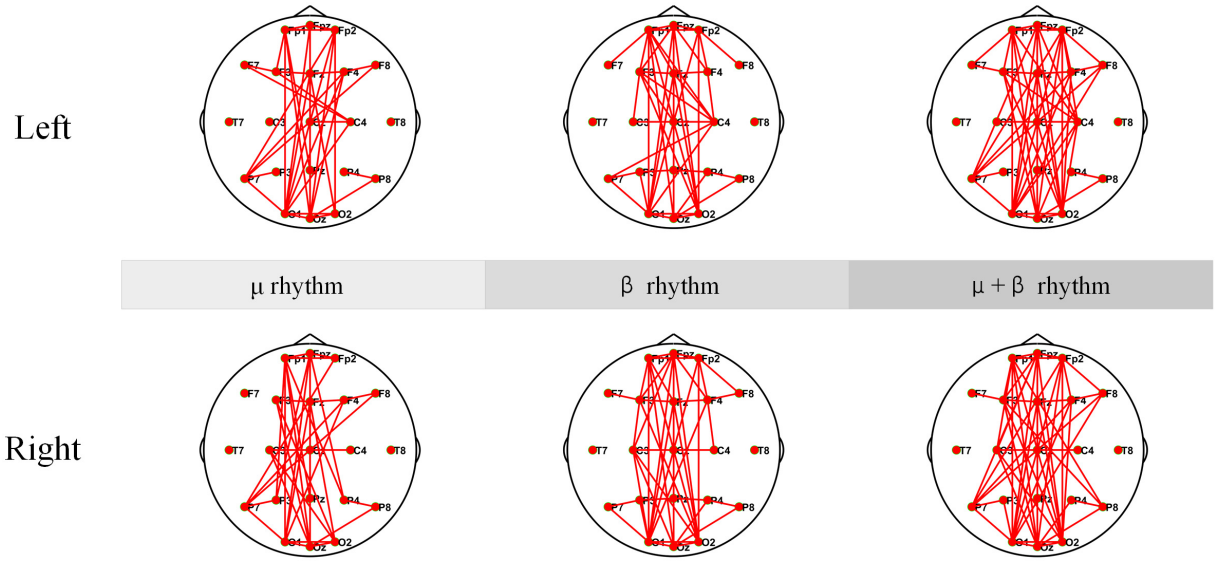


FIGURE 11
The Coherence connectivity maps of motor imagery (MI) tasks in different frequency bands for healthy subjects electroencephalogram (EEG). These include left-hand and right-hand coherence connections under μ rhythm, β rhythm, μ and β rhythms.

that SCI is more focused during the EEG acquisition experiment and the quality of the collected EEG data is higher. Whereas healthy subjects have more active minds and are more easily influenced by their surroundings.

To verify the high performance of the model, this paper compares SVM (Kaper et al., 2004), EEGNet (Lawhern et al., 2018), RNN (Patnaik et al., 2017), LSTM (Wang et al., 2018), CNN (Zhou et al., 2018), M-GCN (Xu et al., 2022b), GAT (Demir et al., 2022), traditional GCN, and ResGCN with C-GCN. 18 SCI patients' EEG data are involved in the training and testing of the above models. SVM is a traditional machine learning method whose classification performance may be affected by the extracted features and configuration parameters. EEGNet uses deep separable convolution to build EEG-specific models. RNNs memorize the previous information and apply it to the computation of the current output. LSTM is a special type of RNN that learns long-term dependent information. CNN can improve performance, but cannot effectively use the spatial information of EEG data. Traditional GCN only considers the spatial location relationships of channels, without considering the intrinsic connections of brain functions. M-GCN and ResGCN are improve patterns based on the traditional GCN. GAT, which helps to focus on the important information in the data, is a combination of graph neural network and attention layer. By selecting the optimal model parameters, the classification

performance of each model can achieve the highest level. The classification results of SCI patient EEG data for the MI tasks in the SVM, EEGNet, RNN, LSTM, CNN, GCN, and C-GCN models are shown in Figure 9. The highest classification accuracy of 96.85% can been obtained from the C-GCN model. Compare with other models, the C-GCN model is 29.50% higher than the lowest SVM model.

4.2. Coherence networks

Electroencephalogram coherence can generate network and functional integration information across brain regions. Any pair of EEG signals may be coherent in some frequency bands and incoherent in others. In the experiment, EEG coherence under μ rhythm, β rhythm, μ and β rhythms of SCI patients are analyzed separately, and the results are shown in Figure 10. According to the coherence network connectivity maps of SCI patients, it is found that SCI patients appear more obviously lateralized as well as long-range connections in the frontal-occipital lobe when performing left-hand and right-hand MI tasks. During performing the left-hand MI task, the connection between the frontal lobe (F8) and the parietal lobe (C4) is stronger under μ rhythm. The connection between F8 and C4 is weakened under β rhythm. In the μ and β rhythms, in addition to a strong connection at C4, the connection between the left frontal lobe (F7) and the right brain region becomes stronger. For right-handed subjects, the left brain also participates in processing relevant information during the execution of left-hand MI tasks to ensure task completion. When performing right-hand MI, the connectivity between the F3 and the C3 is enhanced under μ and β rhythms. In μ and β rhythms, the F3 and the C3 have a stronger connection. Meanwhile, the connectivity between the F7 and the parietal and occipital lobes is enhanced. In summary, C4 and its nearby electrode connectivity are enhanced when performing left-hand MI. C3 and its nearby electrode connectivity are enhanced when performing right-hand MI.

The coherence networks of healthy subject's EEG data under μ rhythm, β rhythm, μ and β rhythms have shown in Figure 11. Healthy subjects have shown more significant laterality when performing the MI tasks. When performing left-hand MI, the connectivity is stronger in the parietal lobe (C4), and some electrodes in the left brain (e.g., F7) are also stronger connected to the right brain. In particular, C4 connectivity is strongest within the μ and β rhythms, followed by the β rhythm, and the μ rhythm is weakest in comparison. When performing right-hand MI, the connectivity of the parietal lobe (C3) and P7 is enhanced in the μ rhythm. The C3 is enhanced and the P7 connection is weakened under β rhythm. The connectivity of the C3 connection is strongest and the P7 connection is also enhanced under μ and β rhythms. In summary, C4 and partial electrode connectivity in the left brain is enhanced during the left-hand MI. C3 connectivity is enhanced during the performance of right-hand MI.

The coherence network in Figures 10, 11 have shown that the connection density of each electrode in SCI patients is significantly higher than healthy subjects. The connections are mainly existed on the prefrontal and occipital lobes. The Fp1, Fpz, and Fp2 have stronger connectivity than the other electrodes under μ rhythm, β rhythm, μ and β rhythms. The F7 in the SCI patients show higher connectivity density than in the healthy subjects under μ and β rhythms. Between the prefrontal and occipital lobes, the SCI patients have significantly more long-range connections than the healthy subjects. It can be inferred that the motor functional areas and sensory functional areas are damaged after SCI. The long-range connections existing between the frontal and occipital lobes of SCI patients are blocked. The cortical functional reorganization, neural activity increases, and functional compensation occurs in related brain areas. The differences of the coherence network between SCI patients and healthy subjects can be used to evaluate SCI, it is important for the clinical rehabilitation of SCI.

5. Conclusion

Spinal cord injury brings a lot of inconvenience to patients' life. It is necessary to provide effective and scientific rehabilitation treatment methods. MI-based BCI system plays an increasingly important role in the rehabilitation treatment of SCI patients. The C-GCN model has been proposed to be applied for MI-based BCI system, which mainly consists of two parts, coherence network and GCN. The coherence network can analyze the intrinsic functional connectivity of the brain and fully exploit the relevant information between channels. GCN can connect the graphical information based on the functional connectivity of the brain to the fully connected layer and can learn the information of the surrounding nodes in the graphical signals. The C-GCN method combines the coherence network with GCN, retains the advantages of the two networks, and provides a guarantee for the classification and recognition of MI tasks in SCI patients. Specifically, the proposed algorithm uses a coherence matrix to characterize the relationship between channels, EEG features as graphical data and finally performs MI tasks classification recognition. The experiments are conducted in SCI patients and healthy subjects, the highest classification accuracy for the MI tasks in SCI patients is 96.85%, and the results are better than with six other classifiers. The average individual accuracy under μ rhythm for the MI pattern recognition in SCI patients is 95.47%, the average individual accuracy under β rhythm is 96.14%, and the average individual accuracy under μ and β rhythms is 97.12%. These experiments have proved that the C-GCN approach is reliable and effective. Furthermore, the C-GCN approach can provide a new strategy for the rehabilitation of SCI patients.

Data availability statement

The datasets presented in this article are not readily available because the article data involves ethics and cannot be disclosed. Requests to access the datasets should be directed to FX, xfz@qlu.edu.cn.

Ethics statement

The study protocols have been approved by the Medical Ethics Committee of Qilu Hospital, Cheeloo College of Medicine, Shandong University [No. KYLL-2020(KS)-475] on the Mar 25, 2020. The patients/participants provided their written informed consent to participate in this study.

Author contributions

FX and HL have contributed to the conception and design of the study. HL, ML, XY, YZ, JZ, CW, and XC have collected the data, processed the data, and performed the experiments. HL has drafted the manuscript. FX, CF, YZ, and JL have reviewed the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the Introduce Innovative Teams of 2021 "New High School 20 Items" Project, Grant No. 2021GXRC071, the Program for Youth Innovative Research Team in the University of Shandong Province in China, Grant No. 2019KJN010, the Natural Science Foundation of China, Grant Nos. 82172535 and 62271293, the Natural Science Foundation of Shandong Province, Grant Nos. ZR2019MA037, ZR2022MF289, and ZR202102200383, the Research Leader Program of Jinan Science and Technology Bureau, Grant No. 2019GXRC061, the Graduate Education and Teaching Reform Research Project of Qilu University of Technology in 2019, Grant No. YJG19007, the Graduate Education and Degree Site Construction and Development Projects of Qilu University of Technology in 2022, the Talent Training and Teaching Reform Project of Qilu University of Technology in 2022 under Grant No. P202204, the Fundamental Research Funds for the Central Universities under Grant No. 2022JC013, and the School-level Teaching and Research Projects of Qilu University of Technology in 2021 under Grant No. 2021yb08.

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

References

Aguilar, J., Humanes-Valera, D., Alonso-Calviño, E., Yague, J., Moxon, K., Oliviero, A., et al. (2010). Spinal cord injury immediately changes the state of the brain. *J Neurosci.* 30, 7528–7537. doi: 10.1523/JNEUROSCI.0379-10.2010

Azab, A., Mihaylova, L., Ang, K., and Arvaneh, M. (2019). Weighted transfer learning for improving motor imagery-based brain–computer interface. *IEEE Trans Neural Syst Rehabil Eng.* 27, 1352–1359. doi: 10.1109/TNSRE.2019.2923315

Cardenas, D., and Jensen, M. (2006). Treatments for chronic pain in persons with spinal cord injury: a survey study. *J Spinal Cord Med.* 29, 109–117. doi: 10.1080/10790268.2006.11753864

Chen, X., Zheng, Y., Niu, Y., and Li, C. (2020). "Epilepsy classification for mining deeper relationships between EEG channels based on GCN," in *Proceeding of the 2020 international conference on computer vision, image and deep learning (CVIDL)*, (IEEE), 701–706.

Collinger, J., Kryger, M., Barbara, R., Betler, T., Bowsher, K., Brown, E., et al. (2014). Collaborative approach in the development of high-performance brain–computer interfaces for a neuroprosthetic arm: translation from animal models to human control. *CTS Clin Transl Sci.* 7, 52–59. doi: 10.1111/cts.12086

Defferrard, M., Bresson, X., and Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. *Adv Neural Inf Process Syst.* 2016, 29.

Demir, A., Koike-Akino, T., Wang, Y., and Erdoğmuş, D. (2022). "EEG-GAT: graph attention networks for classification of electroencephalogram (EEG) signals," in *Proceeding of the 2022 44th annual international conference of the IEEE engineering in medicine & biology society (EMBC)*, (IEEE), 30–35. doi: 10.1109/EMBC48229.2022.9871984

Escolano, C., Antelis, J., and Minguez, J. (2011). A telepresence mobile robot controlled with a noninvasive brain–computer interface. *IEEE Trans Syst Man Cybern Part B Cybern.* 42, 793–804. doi: 10.1109/TSMCB.2011.2177968

Glorot, X., Bordes, A., and Bengio, Y. (2011). "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 315–323.

Han, Z., Liu, Z., Vong, C., Liu, Y., Bu, S., Han, J., et al. (2018). Deep spatiality: unsupervised learning of spatially-enhanced global and local 3D features by deep neural network with coupled softmax. *IEEE Trans Image Proc.* 27, 3049–3063. doi: 10.1109/TIP.2018.2816821

Hinton, G., and Salakhutdinov, R. (2006). Reducing the dimensionality of data with neural networks. *Science.* 313, 504–507. doi: 10.1126/science.1127647

Iturrate, I., Antelis, J., Kubler, A., and Minguez, J. (2009). A noninvasive brain-actuated wheelchair based on a P300 neurophysiological protocol and automated navigation. *IEEE Trans Robot.* 25, 614–627. doi: 10.1109/TRO.2009.2020347

Jensen, M., Hoffman, A., and Cardenas, D. (2005). Chronic pain in individuals with spinal cord injury: a survey and longitudinal study. *Spinal Cord.* 43, 704–712. doi: 10.1038/sj.sc.3101777

Kaper, M., Meinicke, P., Grossekathoefer, U., Lingner, T., and Ritter, H. (2004). BCI competition 2003-data set IIb: support vector machines for the P300 speller paradigm. *IEEE Trans Biomed Eng.* 51, 1073–1076. doi: 10.1109/TBME.2004.826698

Krizhevsky, A., Sutskever, I., and Hinton, G. (2017). Imagenet classification with deep convolutional neural networks. *Commun ACM.* 60, 84–90. doi: 10.1145/3065386

Lawhern, V., Solon, A., Waytowich, N., Gordon, S., Hung, C., and Lance, B. (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, F., Chen, B., Li, H., Zhang, T., Wang, F., Jiang, Y., et al. (2016). The time-varying networks in P300: a task-evoked EEG study. *IEEE Trans Neural Syst Rehabil Eng.* 24, 725–733. doi: 10.1109/TNSRE.2016.2523678

López-Larraz, E., Antelis, J., Montesano, L., Gil-Agudo, A., and Minguez, J. (2012). "Continuous decoding of motor attempt and motor imagery from EEG activity in spinal cord injury patients," in *Proceeding of the 2012 annual international conference of the IEEE engineering in medicine and biology society*, (IEEE), 1798–1801. doi: 10.1109/EMBC.2012.6346299

Millán, J., Rupp, R., Mueller-Putz, G., Murray-Smith, R., Giugliemma, C., Tangermann, T., et al. (2010). Combining brain–computer interfaces and assistive technologies: state-of-the-art and challenges. *Front Neurosci.* 4:161. doi: 10.3389/fnins.2010.00161

Müller-Putz, G., Daly, I., and Kaiser, V. (2014). Motor imagery-induced EEG patterns in individuals with spinal cord injury and their impact on brain–computer interface accuracy. *J Neural Eng.* 11, 035011. doi: 10.1088/1741-2560/11/3/035011

Murias, M., Swanson, J., and Srinivasan, R. (2007). Functional connectivity of frontal cortex in healthy and ADHD children reflected in EEG coherence. *Cereb Cortex.* 17, 1788–1799. doi: 10.1093/cercor/bhl089

Nair, V., and Hinton, G. (2010). "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the international conference on machine learning (ICML)* (Haifa: PMLR).

Ouhmich, F., Agnus, V., Noblet, V., Heitz, F., and Pessaux, P. (2019). Liver tissue segmentation in multiphase CT scans using cascaded convolutional neural networks. *Int J Comput Assist Radiol Surg.* 14, 1275–1284. doi: 10.1007/s11548-019-01989-z

Patnaik, S., Moharkar, L., and Chaudhari, A. (2017). "Deep RNN learning for EEG based functional brain state inference," in *Proceeding of the 2017 international conference on advances in computing, communication and control (ICAC3)*, (IEEE), 1–6.

Rimbert, S., Gayraud, N., Bougrain, L., Clerc, M., and Fleck, S. (2019). Can a subjective questionnaire be used as brain-computer interface performance predictor? *Front Hum Neurosci.* 12:529. doi: 10.3389/fnhum.2018.00529

Roy, A. (2022). An efficient multi-scale CNN model with intrinsic feature integration for motor imagery EEG subject classification in brain-machine interfaces. *Biomed Signal Process Control.* 74, 103496. doi: 10.1016/j.bspc.2022.103496

Spielman, D. (2007). "Spectral graph theory and its applications," in *Proceeding of the 48th annual IEEE symposium on foundations of computer science (FOCS'07)*, (IEEE), 29–38.

Srinivasan, R., Winter, W., Ding, J., and Nunez, P. (2007). EEG and MEG coherence: measures of functional connectivity at distinct spatial scales of neocortical dynamics. *J Neurosci Methods.* 166, 41–52. doi: 10.1016/j.jneumeth.2007.06.026

Supakar, R., Satvaya, P., and Chakrabarti, P. (2022). A deep learning based model using RNN-LSTM for the detection of schizophrenia from EEG data. *Comput Biol Med.* 2022, 106225. doi: 10.1016/j.compbiomed.2022.106225

Wang, P., Jiang, A., Liu, X., Shang, J., and Zhang, L. (2018). LSTM-based EEG classification in motor imagery tasks. *IEEE Trans. Neural Syst Rehabil Eng.* 26, 2086–2095. doi: 10.1109/TNSRE.2018.2876129

Wang, Z., Tong, Y., and Heng, X. (2019). Phase-locking value based graph convolutional neural networks for emotion recognition. *IEEE Access.* 7, 93711–93722. doi: 10.1109/ACCESS.2019.2927768

Weiskopf, N., Mathiak, K., Bock, S., Scharnowski, F., Veit, R., Grodd, W., et al. (2004). Principles of a brain-computer interface (BCI) based on real-time functional magnetic resonance imaging (fMRI). *IEEE Trans Biomed Eng.* 51, 966–970. doi: 10.1109/TBME.2004.827063

Weiss, S., and Mueller, H. (2003). The contribution of EEG coherence to the investigation of language. *Brain Lang.* 85, 325–343. doi: 10.1016/S0093-934X00067-1

Xu, F., Dong, G., Li, J., Yang, Q., Wang, L., Zhao, Y., et al. (2022a). Deep convolution generative adversarial network-based electroencephalogram data augmentation for post-stroke rehabilitation with motor imagery. *Int J Neural Syst.* 32, 2250039. doi: 10.1142/S0129065722500393

Xu, F., Li, J., Dong, G., Li, J., Chen, X., Zhu, J., et al. (2022b). EEG decoding method based on multi-feature information fusion for spinal cord injury. *Neural Netw.* 156, 135–151. doi: 10.1016/j.neunet.2022.09.016

Xu, F., Xu, X., Sun, Y., Li, J., Dong, G., Wang, Y., et al. (2022c). A framework for motor imagery with LSTM neural network. *Comput Meth Prog Biomed.* 218, 106692. doi: 10.1016/j.cmpb.2022.106692

Xu, P., Xiong, X., Xue, Q., Li, P., Zhang, R., Wang, Z., et al. (2014). Differentiating between psychogenic nonepileptic seizures and epilepsy based on common spatial pattern of weighted EEG resting networks. *IEEE Trans Biomed Eng.* 61, 1747–1755. doi: 10.1109/TBME.2014.2305159

Zeng, D., Huang, K., Xu, C., Shen, H., and Chen, Z. (2020). Hierarchy graph convolution network and tree classification for epileptic detection on electroencephalography signals. *IEEE Trans Cogn Dev Syst.* 13, 955–968. doi: 10.1109/TCDS.2020.3012278

Zhang, R., Yao, D., Valdés-Sosa, P., Li, F., Li, P., Zhang, T., et al. (2015). Efficient resting-state EEG network facilitates motor imagery performance. *J Neural Eng.* 12, 066024. doi: 10.1088/1741-2560/12/6/066024

Zhang, Y., Nam, C., Zhou, G., Jin, J., Wang, X., Cichocki, A., et al. (2018). Temporally constrained sparse group spatial patterns for motor imagery BCI. *IEEE Cybern.* 49, 3322–3332. doi: 10.1109/TCYB.2018.2841847

Zhou, M., Tian, C., Cao, R., Wang, B., Niu, Y., Hu, T., et al. (2018). Epileptic seizure detection based on EEG signals and CNN. *Front Neuroinform.* 12:95. doi: 10.3389/fninf.2018.00095

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 25 January 2023
DOI 10.3389/fnins.2022.1088116

OPEN ACCESS

EDITED BY
Ernest N. Kamavuako,
King's College London, United Kingdom

REVIEWED BY
Yajing Si,
Xinxiang Medical University, China
Lotfi Chaari,
Institut National Polytechnique de Toulouse, France

*CORRESPONDENCE
Fangzhou Xu
✉ xfz@qlu.edu.cn
Jiancai Leng
✉ jiancaileng@qlu.edu.cn
Jiyou Tang
✉ jytang001@sohu.com
Shanshan Lu
✉ lushanshan819@163.com

SPECIALTY SECTION
This article was submitted to Neuroprosthetics, a section of the journal Frontiers in Neuroscience

RECEIVED 03 November 2022
ACCEPTED 30 December 2022
PUBLISHED 25 January 2023

CITATION
Xu F, Zhao J, Liu M, Yu X, Wang C, Lou Y, Shi W, Liu Y, Gao L, Yang Q, Zhang B, Lu S, Tang J and Leng J (2023) Exploration of sleep function connection and classification strategies based on sub-period sleep stages. *Front. Neurosci.* 16:1088116. doi: 10.3389/fnins.2022.1088116

Exploration of sleep function connection and classification strategies based on sub-period sleep stages

Fangzhou Xu[1]*, Jinzhao Zhao[1], Ming Liu[1], Xin Yu[1], Chongfeng Wang[1], Yitai Lou[1], Weiyou Shi[1], Yanbing Liu[1], Licai Gao[1], Qingbo Yang[2], Baokun Zhang[3], Shanshan Lu[3,4]*, Jiyou Tang[3,4]* and Jiancai Leng[1]*

[1]International School for Optoelectronic Engineering, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China, [2]School of Mathematics and Statistics, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China, [3]Department of Neurology, Shandong Institute of Neuroimmunology, Shandong Key Laboratory of Rheumatic Disease and Translational Medicine, The First Affliated Hospital of Shandong First Medical University, Shandong Provincial Qianfoshan Hospital, Jinan, China, [4]Department of Neurology, Cheeloo College of Medicine, Shandong Qianfoshan Hospital, Shandong University, Jinan, Shandong, China

Background: As a medium for developing brain-computer interface systems, EEG signals are complex and difficult to identify due to their complexity, weakness, and differences between subjects. At present, most of the current research on sleep EEG signals are single-channel and dual-channel, ignoring the research on the relationship between different brain regions. Brain functional connectivity is considered to be closely related to brain activity and can be used to study the interaction relationship between brain areas.

Methods: Phase-locked value (PLV) is used to construct a functional connection network. The connection network is used to analyze the connection mechanism and brain interaction in different sleep stages. Firstly, the entire EEG signal is divided into multiple sub-periods. Secondly, Phase-locked value is used for feature extraction on the sub-periods. Thirdly, the PLV of multiple sub-periods is used for feature fusion. Fourthly, the classification performance optimization strategy is used to discuss the impact of different frequency bands on sleep stage classification performance and to find the optimal frequency band. Finally, the brain function network is constructed by using the average value of the fusion features to analyze the interaction of brain regions in different frequency bands during sleep stages.

Results: The experimental results have shown that when the number of sub-periods is 30, the α (8–13 Hz) frequency band has the best classification effect, The classification result after 10-fold cross-validation reaches 92.59%.

Conclusion: The proposed algorithm has good sleep staging performance, which can effectively promote the development and application of an EEG sleep staging system.

KEYWORDS

electroencephalography (EEG), sleep stage, classification, brain functional connectivity, phase-locked value (PLV)

1. Introduction

With the development of society, more and more people have the problem with sleep disorders, and how to diagnose and intervene in early sleep disorders is particularly important (Miyata et al., 2007; Younes, 2017). Although electroencephalography (EEG), electrocardiogram (ECG), electromyogram (EMG), and other physiological signals can be used for sleep staging (Yan et al., 2019), EEG signals contain more information and can better reflect the overall information. Sleep staging is the basis of sleep quality assessment. The traditional EEG sleep stage division still requires sleep experts to manually divide according to special brain waves and duration (Ronzhina et al., 2012). The procedure is time-consuming, laborious, and subject to subjective errors (Chapotot and Becq, 2010). The automatic sleep staging method plays an important role in the early diagnosis and intervention of sleep disorders.

In 1968, R&K (Wolpert, 1969) divided sleep into awake, rapid eye movement (REM), and non-rapid eye movement (NREM) stages, of which NREM is further subdivided into four stages: S1, S2, S3, and S4. Because the S3 and S4 stages are similar, the American Academy of Sleep Medicine (AASM) (Berry et al., 2012a) modified the R&K rule, which used N1, N2, and N3 to represent the different stages of NREM, and merged S3 and S4 into the N3 stage. Most studies interpret sleep stages sequentially according to the 30 s recording frame, and if the 30 s is divided into multiple segments, some unique features may be found. In the study of Diykh and Li (2016), the authors divided the EEG signal period of the 30 s into 75 sub-periods and then extracted 12 statistical features from each sub-period. This study achieved 92% classification of 6 sleep stages. Seo et al. (2020) used a convolutional neural network (CNN) to extract features from signal sub-bands and used bi-directional long short-term memory (BiLSTM) to learn the temporal context of representative features. The features learned from continuous signal sub-bands in this study can represent the temporal characteristics of EEG signals, but important sleep-related events may only appear in some special sub-bands (Weber et al., 2021), so it is also necessary to consider the temporal characteristics of EEG signals. Characteristics of learning different brain activities in signaling sub-bands. An et al. (2021) mapped multiple signal wavelets to the amplitude axis and the time axis, respectively, and extracted statistical classification features from the mapped feature information, The accuracy of the classification of sleep stages 5 and 6 sleep stages reached 89.18 and 88.42%. In summary, the key to EEG sleep staging is how to obtain effective classification features and find optimal features from EEG signals in multiple sub-periods.

To obtain effective classification features, researchers have proposed many traditional feature extraction methods, which are divided into the following four types: (1) time domain features. (2) frequency domain features. (3) time-frequency domain features. (4) nonlinear features. (Gunnarsdottir et al., 2018) extracted time-domain and frequency-domain features from PSG signals, using data from healthy people, and using a decision table classifier to classify the extracted attributes, with an overall classification accuracy of 80.70%. da Silveira et al. (2016) used discrete wavelet transform techniques to analyze the changes in sleep behavior in different frequency ranges, extracted skewness, kurtosis, and variance features from the corresponding input channels, and evaluated the ability of random forest classifiers to distinguish different sleep stages. Tests were carried out and the results showed an overall accuracy of 90%. Zhu et al. (2014) proposed a sleep stage classification method based on the time and frequency domain features of single-channel EEG signals. EEG signals were mapped onto visibility maps and level maps to detect gait-related movements, and the nine features extracted from the input signal were forwarded to the support vector machines (SVM) classifier that considers multiple sleep stages. The method achieved 87.50% accuracy for the two-state sleep stage classification problem. Tabar et al. (2021) used a bootstrapping method guided by mutual information to partition sleep stages into a low-dimensional feature space and used fewer features to classify sleep stages. In recent years, deep learning methods have been widely used in sleep stage classification. Seo et al. (2020) proposed a deep learning model intra- and inter-epoch temporal context network (IITNet), for learning intra and inter-epoch temporal context from raw single-channel EEG for automatic sleep scoring, this model has been tested on the Sleep-EDF, Montreal Archive of Sleep Studies (MASS), and Sleep Heart Health Study (SHHS) datasets and obtained the accuracies of 83.9, 87.2, and 86.7%. Mousavi et al. (2019) proposed a network architecture including 9 convolutional layers and 2 fully connected layers to extract features from raw EEG signals, this automatic identification method used single-channel EEG signals to classify 2–6 sleep-like stages. Khalili and Asl (2021) used a CNN to extract features and then employed a temporal convolutional neural network to extract temporal features from the feature vector extracted by CNN, respectively, in Sleep-EDF-2013 and Sleep-EDF-2018 two datasets got 85.39 and 82.46% classification accuracy.

Most of the existing feature extraction methods extract features from a single channel (Terzano et al., 2001; Tagliazucchi et al., 2013; Tagliazucchi and Laufs, 2014; Lv et al., 2015; Desjardins et al., 2017; Stevner et al., 2019; Fu et al., 2021), the calculation is also performed separately on a single channel. The amount of information obtained through a single channel does not fully characterize the changes in brain activities during sleep, making it difficult to explore sleep stage information from a global level. Current research mainly uses functional magnetic resonance imaging (fMRI) to analyze brain function in different brain regions. The fMRI has confirmed that each sleep stage is associated with specific functional connectivity patterns (Goldberger et al., 2000; Berry et al., 2012b; Brignol et al., 2012). Brain functional network is a relatively new measure to characterize the exchange of information between brain regions by calculating the temporal correlation or coherence between them (Baptista et al., 2010; Siettos and Starke, 2016; Rattenborg et al., 2020). EEG-based brain functional connectivity is increasingly being used in sleep studies to differentiate sleep disorders from healthy individuals (Liu et al., 2010; Gao et al., 2015; Guo et al., 2019). Functional connectivity is employed to explore synchronization mechanisms between different brain regions and sleep stage classification accuracy. At present, the common methods to analyze brain functional connections are phase-locked value (PLV) (Diykh et al., 2018), directional transfer function (DTF) (Brázdil et al., 2009), coherence (Bortel and Sovka, 2006), granger causality analysis (GCA) (Chen et al., 2019), and mutual information (MI) (Caballero-Gaudes et al., 2013). PLV is a classical method to construct a functional network. PLV is only sensitive to phase but not amplitude and is often used to measure the phase synchronization between two signals. Compared with other synchronization measures, PLV is easy to operate and can keep the same information level as more complex indicators. In this manuscript, PLV is used to construct brain networks in different sleep stages.

The functional network is used to analyze the brain interaction and connection mechanism in different sleep stages. In this

manuscript, the methods of multi-sub-periods and different frequency bands are proposed to decode EEG signals in different sleep stages. The main contributions are as follows,

(1) The PLV method is used to find the optimal frequency band and sub-period numbers. The experimental results have shown that when the number of sub-periods is 30, the α (8–13 Hz) frequency band has the best classification performance.

(2) The multiple sub-periods are used for feature fusion. The classification performance optimization strategy is used to obtain an accuracy of 92.59%.

(3) In the optimal frequency band, a functional connection network is constructed to explore the brain interaction and connection mechanism in different sleep stages. The specific process is shown in **Figure 1**.

The rest of this study is organized as follows. Sections "2. Materials" and "3. Methods" describe the materials and methods. Section "4. Experimental results and analysis" shows all results. Sections "5. Discussion" and "6. Conclusion" provide a discussion and a summary of future work, respectively.

2. Materials

This manuscript uses the Cyclic Alternating Pattern (CAP) sleep database (Diykh et al., 2016; Yüce and Yaslan, 2016), available on the PhysioNet website. There are 108 different subjects of sleep diseases and health in CAP database, including 92 sleep disorders subjects and 16 healthy subjects. The dataset includes at least 3 EEG channels, EOG, EMG, bilateral anterior tibial EMG, respiratory signal, and ECG. The sampling frequency is 512 Hz To better analyze the relationship between sleep state and brain regions, the calculation of functional brain network connectivity requires as many channels as possible. According to the international 10–20 system, FP2-F4, C4-P4, P4-O2, FP1- F3, C3-P3, P3-O1, and other 6 channels are analyzed. Because some subjects have no polysomnography data and reduce the influence of age on brain connectivity. Four healthy subjects and six nocturnal patients with frontal lobe epilepsy are used for analysis. These subjects are No. 3, No. 5, No. 10, and No. 11 in the healthy group and No. 3, No. 6, No. 11, No. 15, No. 16, and No. 21 in the nfle group. According to the latest sleep rules, S3 and S4 sleep stages are combined into the N3 sleep stage. The basic information on these subjects and epochs will be shown in **Table 1**.

3. Methods

3.1. Data preprocessing and channel selection

During human sleep, the stages of sleep change gradually, and there are no clear boundaries between different stages of sleep. This manuscript uses the data of the CAP database to segment the EEG into 30-s segments. For sleep stages, the adopted 30-s period comes from the R&K and AASM rules (Diykh et al., 2018), and related work has also revealed that 30-s period lengths are feasible for characterizing intrinsic brain activities (Phan et al., 2018; Zhou et al., 2020). The 30-s data is divided into 5-Sub (6 s), 10-Sub (3 s), 15-Sub (2 s), 20-Sub (1.5 s) and 30-Sub (1 s) methods. In addition, the data are filtered in five frequency bands, namely δ (0.5–4 Hz), θ (4–8 Hz), α (8–13 Hz), β (13–30 Hz), and γ (30–40 Hz).

In this manuscript, the adaptive channel selection algorithm in gradient boosting (GB) classifier has been proposed to achieve the optimal channel selection. The data of 12 channels in the CAP database have been connected in a one-to-one way. Different thresholds in the matrix have been set to find channels with good connectivity. Five different numbers of EEG channels, which include 12 channels, 10 channels, 8 channels, 6 channels, and 4 channels, have been selected for comparison. The 10-fold cross validation has been employed to verify the validation of the proposed algorithm.

3.2. Multi-subsegment strategy

It can be seen from **Figure 2** that the specific steps of the multi-sub-segment strategy are divided the 30-s signal into multiple consecutive signal sub-periods, the divided multi-sub-periods do not overlap. In addition, the divided sub-periods are divided into sub-periods of the same length according to the sample size of the 30-s signal, as shown in **Figure 2**.

Let L be a 30-s sleep EEG data sample with a length of (30*512). When the number of sub-periods is set to be long Ls, the length of the divided sub-periods satisfies the following constraints:

$$L = Ls * Ns \tag{1}$$

Ns is the number of sub-periods divided by the sleep period. L is the sub-period length. The sampling frequency in the CAP sleep database is 512 Hz, and the length of the data sample is 30 s, so the specific calculation of the length of the sub-period can be as follows,

$$Ls = L/Ns = (30 * 512)/Ns = 15360/Ns \tag{2}$$

According to the above formula, the number of sub-periods Ns plays an important role in the experiment. By setting an appropriate number of sub-periods, effective classification features can be obtained from the divided sub-periods, and at the same time, redundancy can be properly handled. In addition, the division of multiple sub-periods is also related to the computational complexity of feature extraction, the features extracted from consecutive sub-periods have time-series features for analyzing sleep signals. 30-s data has been divided into sub-period data. In general, 30-s data can be divided into 30 sub-periods. The sleep stages of sub-periods have been recognized by the proposed algorithm. 30 classification results can be obtained from the 30 sub-periods. The final classification result of a 30-s data can be obtained from the highest result from the sub-periods. The classification is at this stage, and different numbers of sub-segments are used for comparative experiments to improve the overall classification results.

3.3. Phase lock value

For network analysis of sleep signals, the construction of the brain network is the basis of research and is crucial for network analysis. In this manuscript, the processed data are used to construct the brain network of sleep signals. The nodes involved in the brain network refer to the electrodes used in the data acquisition procedure, the connection between the networks refers to the

10.3389/fnins.2022.1088116

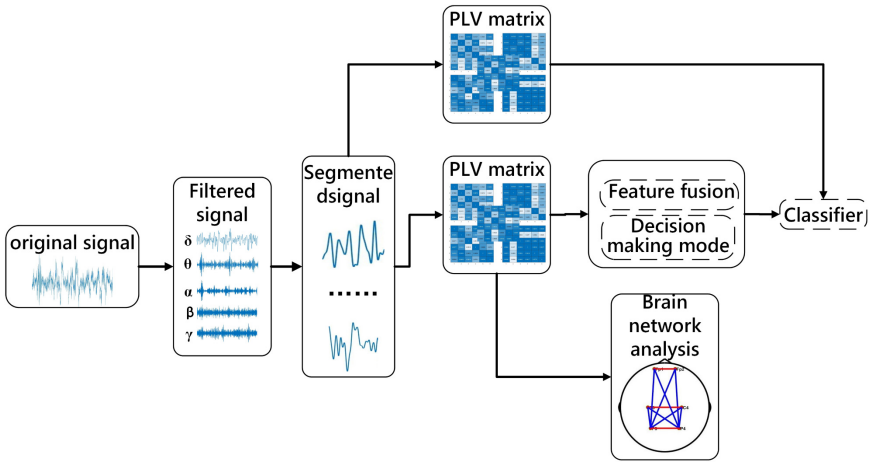


FIGURE 1
The flow diagram of the proposed framework.

TABLE 1 Basic information about each subject and period numbers of each stage.

	Condition	Age	N1	N2	N3	R	W	All stage	All time
Sub1	Healthy	35	49	347	279	188	136	999	499.5 min
Sub2	Healthy	35	49	413	303	232	10	1,007	503.5 min
Sub3	Healthy	23	2	261	308	215	67	853	426.5 min
Sub4	Healthy	28	6	266	344	380	56	1,052	526 min
Sub5	Nfle	29	72	419	209	261	136	1,097	548.5 min
Sub6	Nfle	32	37	323	236	190	24	810	405 min
Sub7	Nfle	31	28	320	366	279	27	1,020	510 min
Sub8	Nfle	29	19	417	274	227	97	1,034	517 min
Sub9	Nfle	30	9	398	131	152	109	799	399.5 min
Sub10	Nfle	27	31	209	257	254	78	801	400.5 min

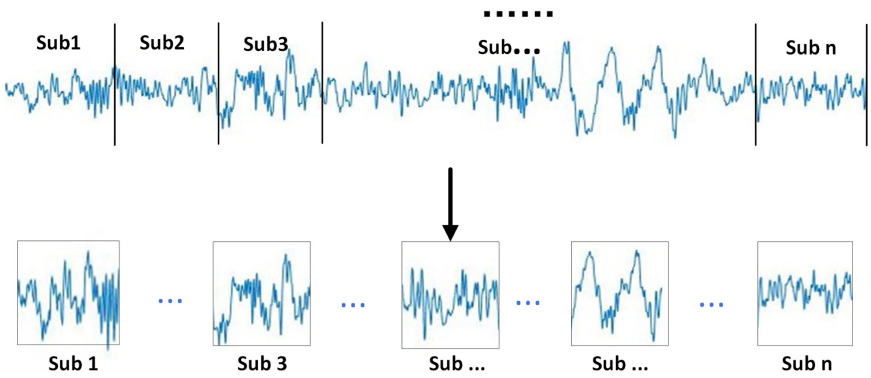


FIGURE 2
Sub-period division method direct PLV feature fusion decision choice.

functional connection between any two nodes (this study mainly considers the functional connection network). According to whether the flow of information between nodes is concerned, the constructed brain network can be divided into a directed network and an undirected network. There are many ways to build a network. For the construction of directed networks, the commonly used methods mainly include directional transfer function, granger causality, partial directional coherence, and so on. There are also many methods for constructing undirected networks, such as correlation, coherence, phase locking, and phase lag.

This manuscript adopts PLV to assess brain functional connectivity (Lachaux et al., 1999). PLV is widely used to measure the phase synchronization between each pair of electrodes. The reason is that PLV is only sensitive to the phase. Compared with

10.3389/fnins.2022.1088116

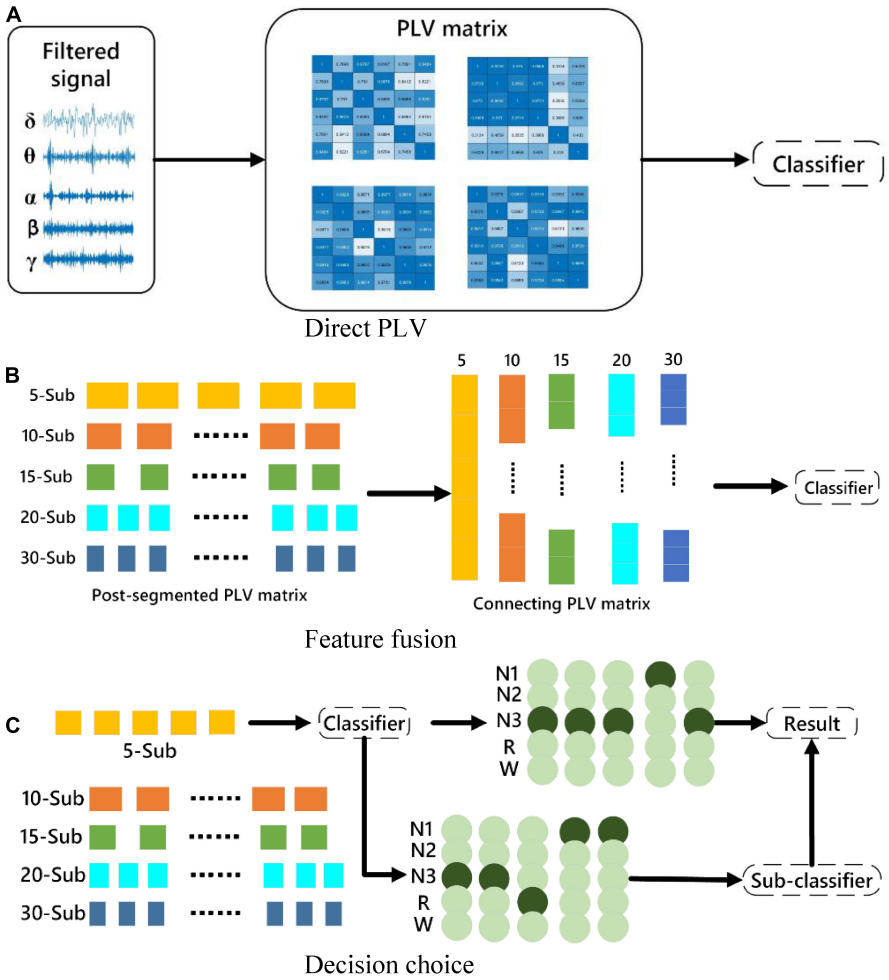


FIGURE 3
Schematic diagram of the classification strategy. **(A)** The original signal is directly subjected to PLV for classification. **(B)** Feature-level fusion for classification. **(C)** Multi-sub-segment classification for decision selection.

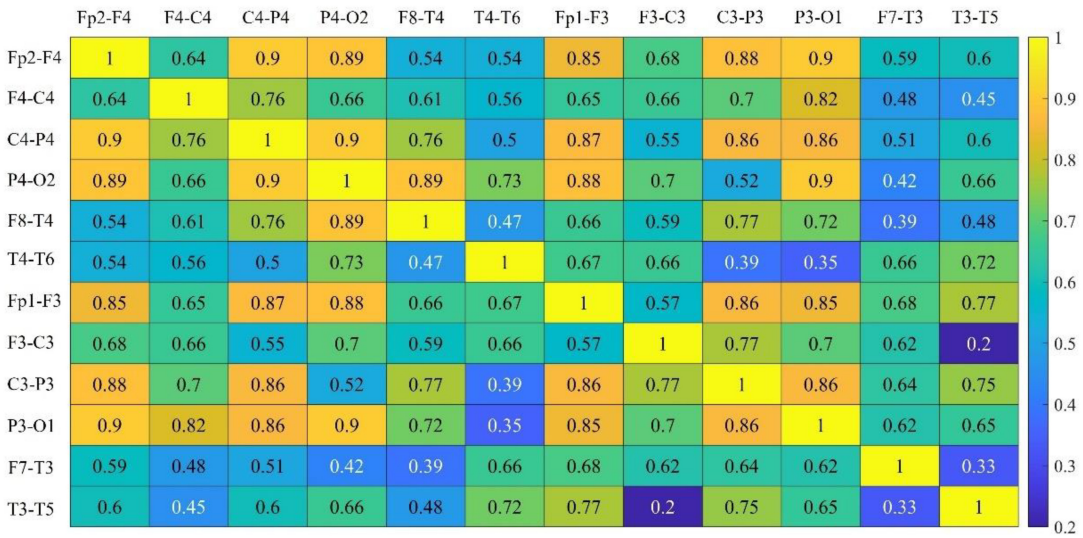


FIGURE 4
Communication strength of all channels.

TABLE 2 Selection of different channels and running time of every 100 events.

Channel selection	Acc (%)	Time/100 events
C4-P4, P4-O2, C3-P3, P3-O1	79.8	1.62 s
Fp2-F4, C4-P4, P4-O2, Fp1-F3, C3-P3, P3-O1	83.2	2.27 s
Fp2-F4, F4-C4, C4-P4, P4-O2, F8-T4, Fp1-F3, C3-P3, P3-O1	81.7	3.11 s
Fp2-F4, F4-C4, C4-P4, P4-O2, F8-T4, Fp1-F3, F3-C3, C3-P3, P3-O1, F7-T3	77.6	4.31 s
All channel	73.6	5.25 s

other methods, the PLV method is simple to operate. PLV is the comparison between channel *i*and channel *j*. A high PLV value indicates a strong coupling relationship between the pair of electrodes (Quiroga et al., 2010). Therefore, PLV is used to construct the corresponding brain network in this manuscript. The calculation formula of PLV is defined as follows,

$$PLV = \left| \frac{1}{N} \sum_{j=0}^{N-1} e^{i\Delta\Phi(t)} \right| \quad \Delta\phi(t) = \phi_x(j\Delta t) - \phi_y(j\Delta t) \tag{3}$$

Among them, N represents the number of samples of the sleep signal, and t represents the sampling period, the two-time series are $(x)\ t$ and $(y)\ t$, the two instantaneous phases are ϕ_x (t) and ϕ_y (t). Six EEG channels are used in the experiment, and a 6*6 PLV symmetric matrix has been obtained from each period, and each value in the matrix represented the coupling relationship of a pair of channels. In addition, brain network analysis between sleep stages is compared with the PLV matrix. The PLV matrices for each sleep stage are averaged and brain networks are constructed based on thresholds. The threshold is chosen from the maximum value that does not appear outlier in the network.

3.4. Fusion strategy

For the assessment of brain functional connectivity in different frequency bands and different numbers of sub-periods, three strategies of feature processing are used to classify sleep stages. The features mentioned here are the brain functional connectivity features extracted from the processed data. The three methods are, (1) Directly extract functional connectivity features from the 30-s EEG data. (2) Stack the functional connectivity features extracted from multiple sub-periods, and then input the features as a whole into the classifier. (3) Directly analyze the functional connectivity features of sub-periods Perform classification, and then take the mode of the classification result as the result of the entire 30-s segment. The specific description is shown in Figure 3. Divide the 30-s data into 5, 10, 15, 20 and 30 segments according to Section "3.1. Data preprocessing and channel selection." The proposed three methods are used to experiment with the segmented data.

Because method C may result in the same number of classifications for certain two categories, a sub-classifier is designed to re-extract the data with the same classification results, transfer it to the sub-classifier, perform binary classification, and use the binary classification result as the final classification result.

3.5. Classifier

This manuscript adopts a support vector machine (SVM) with a Gaussian kernel function, which is implemented in the LibSVM library (Cortes and Vapnik, 1995; Chang and Lin, 2000). The way to achieve multi-class classification is to use a one-to-one strategy. Classification performance is evaluated from sleep stage accuracy for three strategies of classification across frequency bands. 80% of the samples are used for model training, the remaining 20% are used as test data.

4. Experimental results and analysis

4.1. Channel selection and band comparison

The channel selection proposed in Section "3.2. Multi-subsegment strategy" has been tested. In Figure 4, the connectivity coefficients of double channels and all combinations have been separately calculated. In Table 2, the best performance has been obtained when six channels are selected.

To reduce algorithm complexity, irrelevant channels or noisy channels have been eliminated, the feature dimension has been reduced. The calculation burden have been reduced. The algorithm operation efficiency and algorithm performance have been improved.

To evaluate the effectiveness of the method proposed in this manuscript, feature extraction is performed directly on the 30 s EEG data using the feature extraction method described in Section "2. Materials." The results obtained by inputting the features into the classifier can demonstrate the classification performance of the PLV as a feature, and the classification results can be further compared with the classification results of the proposed multi-sub feature learning. In addition, the evaluation metrics used in this manuscript include accuracy (Acc), sensitivity (Recall), positive predictive value (Ppv), and F1 score (F1). Based on the experimental data in the second part, the five frequency ranges of δ (0.5–4 Hz), θ (4–8 Hz), α (8–13 Hz), β (13–30 Hz), and γ (30–40 Hz) are tested, respectively. A 5-category sleep staging task is tested. The specific classification results of these five frequency bands are shown in Table 3.

According to the sleep staging results in Table 3, the overall classification accuracy of the five frequency bands can be obtained as 59.4, 72, 77.8, 70.3, and 69.4%, respectively. It can be seen that the accuracy of the α (8–13 Hz) frequency band (Acc), sensitivity (Recall), positive predictive value (Ppv), F1 score (F1), and other evaluation indicators are better than other frequency bands. On the test set, the α (8–13 Hz) frequency band has the highest classification accuracy for EEG signals, which is 75%, and the δ (0.5–4 Hz) frequency band has the lowest classification accuracy for EEG signals, which is 59.4%. The specific situation is shown in Figure 5.

The classification performance of these five frequency bands on classification tasks is comprehensively analyzed. The classification results of N1 in the θ (4–8 Hz) frequency bands are significantly higher than other frequency bands. The classification accuracy is better than other frequency bands.

TABLE 3 The classification results of five types of sleep stages by directly extracting features from 5-band EEG.

		N1	N2	N3	W	*R*
δ (0.5–4 Hz)	Recall	0.8	0.67	0.5	0.52	0.47
	Ppv	0.72	0.56	0.51	0.66	0.5
	F1	0.76	0.61	0.5	0.58	0.48
	Acc (%)	59.4%				
θ (4–8 Hz)	Recall	0.95	0.75	0.53	0.63	0.73
	Ppv	0.9	0.66	0.73	0.76	0.59
	F1	0.93	0.70	0.62	0.69	0.65
	Acc (%)	72%				
α (8–13 Hz)	Recall	0.71	0.82	0.78	0.75	0.83
	Ppv	0.79	0.82	0.76	0.75	0.67
	F1	0.66	0.82	0.77	0.75	0.74
	Acc (%)	77.8%				
β (13–30 Hz)	Recall	0.55	0.87	0.65	0.63	0.8
	Ppv	0.61	0.70	0.78	0.79	0.65
	F1	0.58	0.78	0.71	0.70	0.72
	Acc (%)	70.3%				
γ (30–40 Hz)	Recall	0.6	0.83	0.52	0.73	0.77
	Ppv	0.9	0.71	0.61	0.76	0.57
	F1	0.72	0.77	0.56	0.75	0.65
	Acc (%)	69.4%				

4.2. Sub-period feature fusion

Based on the abovefive classification tasks for PLV to realize sleep signal, the features obtained by PLV in multiple sub-periods are subjected to feature fusion and are compared with the features extracted from the original signal. The raw EEG is divided into multiple consecutive signal sub-periods that do not overlap, while each piece of raw data uses a different wavelet number to obtain different classification performances. The number of sub-periods is 5, 10, 15, 20, and 30 for five experiments, the number of segments is not suitable for more than 30, because the duration of the K-complex wave and spindle wave needs to be greater than 0.5 s.

TABLE 4 Accuracy of classification of sleep stages by five frequency bands and five sub-periods under the sub-period feature fusion method.

	δ (0.5–4 Hz)	θ (4–8 Hz)	α (8–13 Hz)	β (13–30 Hz)	γ (30–40 Hz)
5-Sub	63.33%	69.04%	71.73%	70.59%	67.58%
10-Sub	71.67%	75.70%	78.77%	78.20%	72.97%
15-Sub	75.00%	80.59%	79.69%	76.89%	75.68%
20-Sub	78.33%	84.89%	87.00%	81.93%	83.10%
30-Sub	81.67%	86.81%	88.63%	83.82%	81.76%

The same classification model is used to divide the sleep stages into five classifications, the optimal feature set can be selected for the classification performance. According to the comparison results in Section "3.1. Data preprocessing and channel selection," the proposed algorithm uses the sleep EEG signal in the α (8–13 Hz) frequency band for experimental testing. The specific experimental results are shown in Table 4.

As shown in Table 4, the accuracy of the multi-segment feature fusion method is 10.83%. The multi-segment feature fusion method can obtain higher classification performance than the method of directly connecting the original signal through brain function. The accuracy of the 30-segment method is the highest, and the accuracy of sleep staging in the α-band reaches 88.63%.

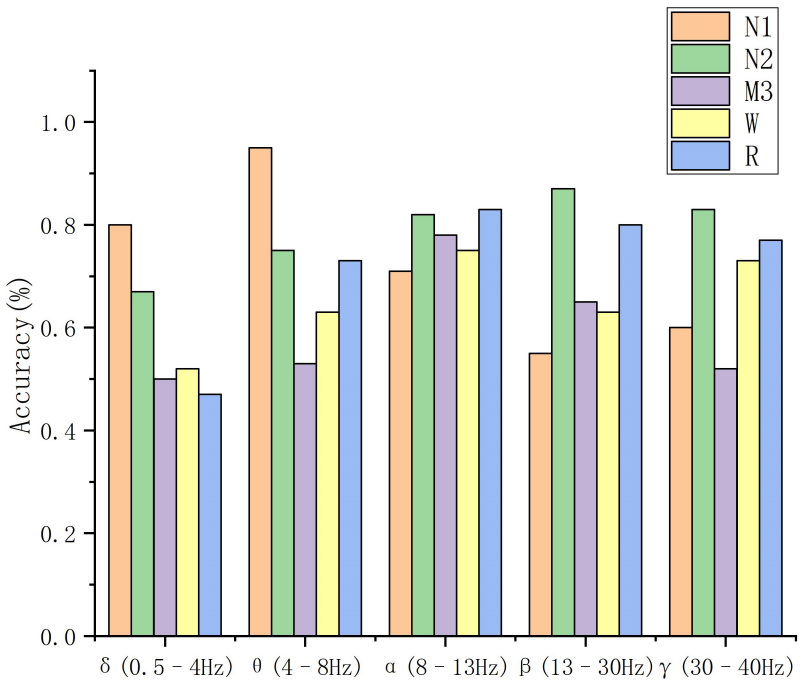


FIGURE 5
The results of the 5-band EEG direct feature extraction for 5 types of sleep stages.

TABLE 5 Accuracy of classification of sleep stages in five frequency bands and five sub-periods under sub-period decision-making sleep stage classification method.

	δ (0.5–4 Hz)	θ (4–8 Hz)	α (8–13 Hz)	β (13–30 Hz)	γ (30–40 Hz)
5-Sub	78.33%	84.59%	91.06%	92.02%	81.08%
10-Sub	83.33%	87.82%	91.42%	91.39%	89.19%
15-Sub	83.33%	92.00%	94.27%	92.65%	87.84%
20-Sub	85.00%	92.30%	94.84%	93.28%	90.54%
30-Sub	88.33%	93.04%	96.42%	94.12%	89.19%

4.3. Sub-period decision-making sleep stage classification

The sub-period features are extracted according to the optimal frequency band and the number of sub-periods is found in Section "3.2. Multi-subsegment strategy," the features extracted from sub-periods are directly classified, because the result of the classification is the result of the sub-period, and it cannot represent the category of the original EEG. The divided data are recombined in chronological order. In the combined classification results, the classification result with the most categories is found. As shown in Table 5, it still shows the best classification performance in the α band. The accuracy of 30 sub-period reaches 96.42%. During the experiment, there are cases where the probability of belonging to two or more sleep stages is the same. To solve this problem, the extracted features are input into the binary classifier and the performance of the binary classification is the best. Aiming at the problem of the same probability of sleep stages, the extracted features are re-input to the proposed algorithm to optimize the classification performance strategy and find the final classification result. Figure 6 shows the combination of 5 types of sleep stages, a total of 10 kinds of results, in which the abnormal value appears in the case of N1 Vs W, the classification accuracy is only 79.1%, the average classification accuracy of the remaining 9 combinations is 95.27%, multiple sub-classifiers has chosen to handle this situation. Table 5 shows the classification results of the functionally linked features for different sub-period numbers.

4.4. Contrast experiment

According to the optimal frequency band and the optimal number of sub-periods found in Section "4.3. Sub-period decision-making sleep stage classification," a number of comparative tests have been conducted. The comparison includes classification by using various classifiers. The classifiers used in the classification are LibSVM, GB, random forest, k-nearest neighbor (KNN), and CNN. The specific classification results are shown in Figure 7. It can be seen from Figure 7 that the classification effect of LibSVM is better than the other three classifiers as a whole. The accuracy of the method using CNN is low, probably because the extracted features are the features between channels, and the convolution kernel destroys the relative position relationship. Common brain network connection methods include PLV, DTF, coherence, GCA, and MI. The optimal frequency band and the optimal number of sub-periods are analyzed under different connection modes. The concrete results are shown in Figure 8, PLV connection method is better than other methods.

4.5. Cross-validation

According to the above experiments, direct PLV feature extraction, multi-sub-period feature fusion, and multi-sub-period decision-making methods are carried out, respectively, and finally, 30 sub-period features are found to have the best classification effect for decision-making. The EEG signals of different subjects have

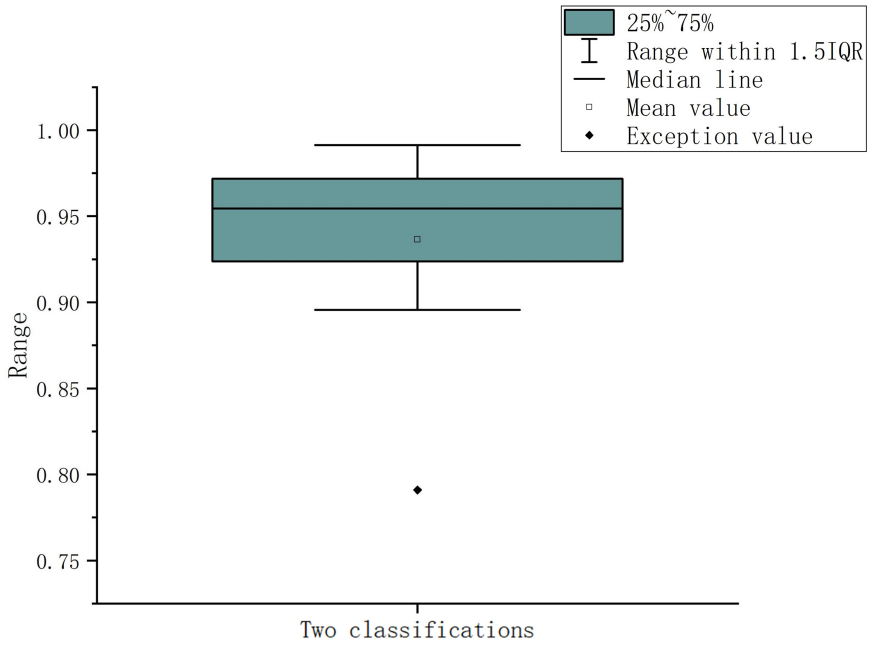


FIGURE 6
Pairwise classification results for sub-period data.

10.3389/fnins.2022.1088116

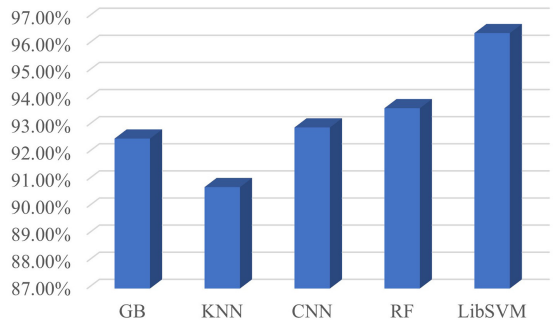


FIGURE 7
Compare different classifiers.

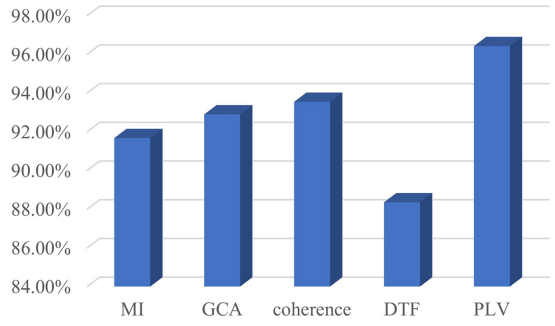


FIGURE 8
Compare different connection methods.

TABLE 6 Sleep classification results after cross-validation.

	N1	N2	N3	W	*R*
Recall	0.96	0.85	0.89	0.96	0.96
Ppv	0.96	0.92	0.86	0.93	0.96
F1	0.96	0.88	0.87	0.95	0.96
Acc (%)	92.59%				

individual differences, the generalization of the proposed algorithm cannot be guaranteed. Therefore, the method needs to be further evaluated for different subjects to test the performance of the method. To demonstrate the effectiveness of the proposed method, the 10-fold cross-validation experiments are conducted between subjects, selecting one subject as the test data and the other subjects as the training set. Since the training dataset and test dataset do not contain EEG samples of the same subjects, cross-validation between subjects can well reflect the generalization of the proposed in practical applications. The α (8–13 Hz) frequency band is used for testing with 30 sub-periods, the final cross-validation results are shown in Table 6. For the experimental results in Table 6, it is found that the final result for the 5-class sleep stage classification is 92.59%, although this result is lower than the result of the random division, the cross-validated method can overcome the randomness of the data.

4.6. Brain network analysis of PLV

The PLV values in different stages of the α frequency band of healthy people are combined in pairs. Figure 9 shows the comparison

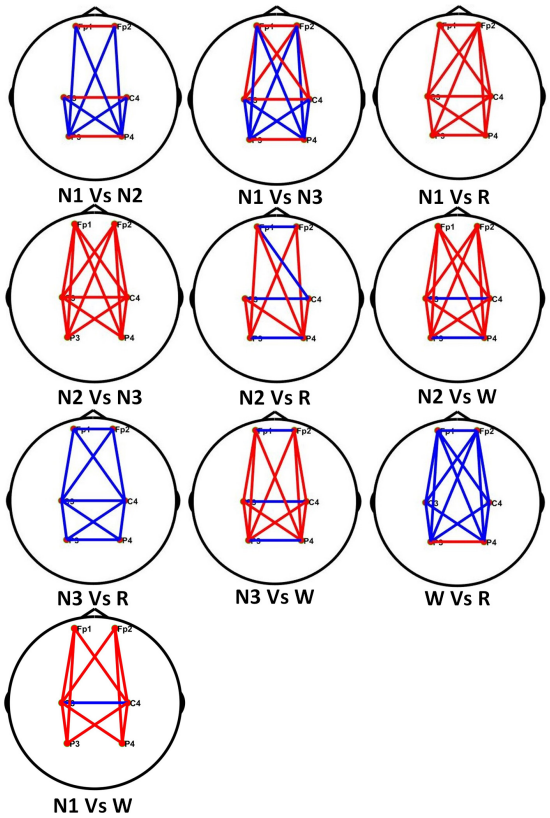


FIGURE 9
Comparison of brain network connectivity in different sleep stages in theα frequency band.

of connectivity in different sleep stages, where the red line represents the former with high connectivity, the blue line is the opposite.

In Figure 9, each line represents the connectivity coefficient between each pair of channels. It can be observed that the overall connectivity of the N1 stage is greater than that of the REM stage, that of the N2 stage is greater than that of the N3 stage, the connectivity of the N3 and W stages is generally smaller than that of the REM stage. The connectivity between the left and right brains in the N1 stage is stronger than that in the N2 stage, but the connectivity between the occipital, parietal, and frontal lobes is weaker than that in the N2 stage. Compared with the N3 stage, the left and right brains of the N1 stage are weaker The connectivity between them is stronger than that in the N3 stage, but the direct connectivity between the frontal and parietal lobes in different brain regions is also greater than that in the N3 stage. For the comparison of N2, W, and REM stages, the connectivity between the left and right brains in the N2 stage is smaller than that in the REM stage, and in the W stage, the connectivity between different brain regions is stronger.

5. Discussion

The main purpose of this study is to study the performance of multi-channel EEG signals on sleep staging, obtain effective fusion features using multiple sub-periods, propose a classification optimization strategy, and use the brain function network to analyze the physiological phenomenon of sleep staging. Therefore, a feature learning method of multi-sub-period brain functional network is

TABLE 7 Comparison of automatic sleep stage classification algorithms.

References	Description	Performance (%)
Sharma et al., 2021	Database: CAP	Accuracy: (A):92.8 (B):85.3
	Signals: EEG	
	Channel: F4-C4, C4-A1	
	Classification: 6-class A: Balance data B: Imbalance data	
	Features: wavelet decomposition	
	Classifier: EBT	
Tripathy et al., 2020	Database: CAP	Accuracy: 71.68
	Signals: EEG	
	Channel: F4-C4, C4-P4, P4-O2, C4-A1	
	Classification: 6-class	
	Features: Dispersion entropy and Bubble entropy	
	Classifier: Hybrid classifier	
Zhao et al., 2022	Database: CAP	Accuracy: 78.8
	Signals: EEG Channel: C3-A2, C4-A1	
	Classification: 5-class [W vs. S1 vs. S2 vs. (S3 + S4) vs. REM]	
	Features: Convolution feature	
	Classifier: RNN	
Our method	Database: CAP	Accuracy: 92.59
	Signals: EEG	
	Channel: FP2-F4,C4-P4,P4-O2,FP1-F3,C3-P3,P3-O1	
	Classification: 5-class [W vs. S1 vs. S2 vs. (S3 + S4) vs. REM]	
	Features: Interval connectivity coefficient	
	Classifier: LibSVM	

proposed, which can analyze the features of functional brain network in time series. Comparing the multi-sub-segment and non-multi-sub-segment classification results, the classification accuracy of 30-Sub of 98.63% is significantly better than the original data, and at the same time, it is higher than other classification performances with different numbers of sub-segments. Setting a different number of sub-segments will have different classification effects. For example, in the case of 20-Sub, the classification of the REM stage is better than the division of other sub-segments, which shows that the features extracted in 20-Sub are more suitable for the classification of the REM stage. Different sub-divisions have different classification effects on feature learning, so the relationship between the optimal number of sub-segments and features will be studied in future work. This manuscript explores multi-Sub EEG feature learning for multi-channel EEG sleep staging, which has important potential to improve the application of sleep staging.

Through the analysis of different frequency bands, it is found that the α frequency band has a good performance in classifying sleep stages. For example, in the AASM standard for the classification of sleep stages, the interpretation of the W stage is to record a series of sinusoidal brain waves of 8–13 Hz in the occipital area. The amplitude can be decreased when eyes are open. In addition, α activity may be more pronounced in REM than in N1, the α frequency in REM is usually 1–2 Hz slower than in W. At the same time, the related work of others also have revealed the important role of the α frequency band in sleep staging. Dkhil et al. (2017) proposed the importance of the α band in the assessment of drowsiness. Knaut et al. (2019) found that changes in α oscillations reflect different brain states associated with different levels of wakefulness and thalamic activity. **Figure 9** shows the differences in the brain connectivity of different sleep stages in the α frequency band. For example, in stages N2 and N3, the overall connectivity of N2 is greater than that of stage N3. This phenomenon indicates that the connections between brain regions are relatively close. The overall connectivity difference between the N1 and N2 stages is not very obvious, but there is a clear gap in the connectivity between the left and right brain regions.

The proposed framework is compared with the state-of-the-art in sleep stage classification studies, as shown in **Table 7**. Sharma et al. (2021) recently used the CAP database to decompose EEG epochs into sub-bands using a new class of optimized wavelet filters, the norm features were computed from the six sub-bands coefficients of the optimal wavelet filter bank, which were processed by ensemble of bagged tree (EBT). The ensemble of classifiers obtained 85.3% of unbalanced classification results and 92.8% of balanced data. Tripathy et al. (2020) used 25 subjects for 6 sleep-like stage classification. The CAP database was employed for processing and obtained the classification of 71.68%. This database includes 6 healthy, 7 insomniacs, 1 brux, 1 breathing disorder, and 10 REM behavior disorder patients. Zhao et al. (2022) used CNN to learn the representative features of each sleep stage, feedback on these feature sequences to recurrent neural network (RNN), and learn the context information of sleep stages in chronological order.

6. Conclusion

The framework of dividing sleep stages by multi-sub-segment brain functional connectivity has been proposed. The original EEG signal is filtered into different frequency bands, and the PLV is calculated for the processed data, respectively. The PLV value represents the connectivity coefficient between different channels, the PLV matrix calculated in different sub-periods is used as a feature to find the optimal frequency band according to the performance of sleep stage classification. Then, the filtered data is divided into different numbers of sub-periods, the PLV matrices of the sub-periods are calculated, the features of different numbers of sub-periods are feature-fused, the optimal number of sub-period classifications is found by the classification performance after feature fusion. Finally, the classification performance optimization strategy is used for classification, the brain network is constructed by PLV to explore the mechanism of brain functional connectivity. Firstly, to test the proposed method, extensive experiments have been performed on the sleep dataset CAP. The classification results are tested and analyzed using two test methods, random data partitioning and inter-subject cross-validation. The final results are 96.42 and 92.59%. These results have demonstrated the effectiveness and robustness of the proposed multi-channel EEG sleep staging algorithm. Secondly, the connectivity of N1 stage is larger than that of REM stage, the N2 stage is larger than that of N3 stage, the connectivity of N3 stage and W stage is smaller than that of REM stage. For the comparison of N2, W, and REM stages, the connectivity between the left and right brains in the N2 stage is smaller than that in the REM stage, and in the W stage, the connectivity between different brain regions is stronger. Finally, different numbers of sub-periods have different performances for distinguishing sleep stages. The case of 30-Sub shows good performance, but using 30-Sub in the α band has a higher error rate between N1 and W stages. In the future, multi-channel EEG signals in the CAP database will be planned to classify different sleep disorders, such as insomnia and REM dyskinesia. Furthermore, graph convolutional networks will be employed for automatic sleep stage monitoring to develop an online brain-computer interface system.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.physionet.org/static/published-projects/capslpdb/cap-sleep-database-1.0.0.zip.

Author contributions

FX, JT, SL, and JL had contributed to the conception and design of the study. JZ, ML, and XY had conducted the experiments. CW and YTL had collected the data. WS and YBL have processed the data. LG and BZ had drafted the manuscript. JL had reviewed the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by Natural Science Foundation of Shandong Province of China under (Grant Nos. ZR2022MF289, ZR2020MH160, and ZR2019MA037), Natural Science Foundation of China under (Grant Nos. 62271293 and 81471345), Introduce Innovative Teams of 2021 "New High School 20 Items" Project under (Grant No. 2021GXRC071), Program for Youth Innovative Research Team in the University of Shandong Province in China under (Grant No. 2019KJN010), Research Leader Program of Jinan Science and Technology Bureau under (Grant No. 2019GXRC061), Graduate Education and Degree Site Construction and Development Projects of Qilu University of Technology in 2022, School-level Teaching and Research Projects of Qilu University of Technology in 2021 under (Grant No. 2021yb08), and Talent Training and Teaching Reform Project of Qilu University of Technology in 2022 under (Grant No. P202204).

Acknowledgments

The authors thank Qianfoshan Hospital for his helpful comments on the early version of the manuscript.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.1088116/full#supplementary-material

References

Miyata, S., Noda, A., Nakata, S., Yagi, H., Yanagi, E., Honda, K., et al. (2007). Daytime polysomnography for early diagnosis and treatment of patients with suspected sleep-disordered breathing. *Sleep Breath* 2, 109–115. doi: 10.1007/s11325-006-0091-9

Younes, M. (2017). The case for using digital eeg analysis in clinical sleep medicine. *Sleep Sci.* 1, 1–15. doi: 10.1186/s41606-016-0005-0

Yan, R., Zhang, C., Spruyt, K., Wei, L., Wang, Z., Tian, L., et al. (2019). Multi-modality of polysomnography signals' fusion for automatic sleep scoring. *Biomed. Signal Process* 11, 13–23. doi: 10.1016/j.bspc.2018.10.001

Ronzhina, M., Janoušek, O., Koláøová, J., Nováková, M., Honzík, P., and Provazník, I. (2012). Sleep scoring using artificial neural networks. *Sleep Med. Rev.* 16, 251–263. doi: 10.1016/j.smrv.2011.06.003

Chapotot, F., and Becq, G. (2010). Automated sleep-wake staging combining robust feature extraction. Artificial neural network classification, and flexible decision rules. *Int. J. Adapt. Control Signal Process* 24, 409–423. doi: 10.1002/acs.1147

Wolpert, E. A. (1969). A manual of standardized terminology. techniques and scoring system for sleep stages of human subjects. *Arch. Gen. Psychiatry* 20, 246–247. doi: 10.1001/archpsyc.1969.01740140118016

Berry, R. B., Brooks, R., Gamaldo, C. E., Harding, S., Marcus, M. C., Vaughn, B. V., et al. (2012a). The AASM manual for the scoring of sleep and associated events. Rules, terminology and technical specifications. *Am. Acad. Sleep Med. (Darien)*. Available online at: http://www.aasmnet.org/scoringmanual/

Diykh, M., and Li, Y. (2016). Complex networks approach for EEG signal sleep stages classification. *Expert. Syst. Appl.* 63, 241–248. doi: 10.1016/j.eswa.2016.07.004

Seo, H., Back, S., Lee, S., Park, D., Kim, T., and Lee, K. (2020). Intra-and inter-epoch temporal context network (IITNet) using sub-epoch features for automatic sleep scoring on raw single-channel EEG. *Biomed. Signal Process* 61:102037. doi: 10.1016/j.bspc.2020.102037

Weber, F. D., Supp, G. G., Klinzing, J. G., Mölle, M., Engel, A. K., and Born, J. (2021). Coupling of gamma band activity to sleep spindle oscillations–a combined EEG/MEG study. *NeuroImage* 224:117452. doi: 10.1016/j.neuroimage.2020.117452

An, P., Yuan, Z., and Zhao, J. (2021). Unsupervised multi-subepoch feature learning and hierarchical classification for EEG-based sleep staging. *Expert Syst. Appl.* 186:115759. doi: 10.1016/j.eswa.2021.115759

Gunnarsdottir, K. M., Gamaldo, C. E., Salas, R. M., Ewen, J. B., Allen, R. P., and Sarma, S. V. (2018). "A novel sleep stage scoring system: combining expert-based rules with a decision tree classifier," in *Proceedings of the 40th annual international conference of the IEEE engineering in medicine and* biology society (EMBC). Honolulu: 3240–3243. doi: 10.1109/EMBC.2018.8513039

da Silveira, T. L., Kozakevicius, A. J., and Rodrigues, C. R. (2016). Single-channel EEG sleep stage classification based on a streamlined set of statistical features in wavelet domain. *Med. Biol. Eng. Comput.* 55, 343–352. doi: 10.1007/s11517-016-1519-4

Zhu, G., Li, Y., and Wen, P. (2014). Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal. *IEEE J. Biomed. Health Inform.* 18, 1813–1821. doi: 10.1109/JBHI.2014.2303991

Tabar, Y. R., Mikkelsen, K. B., Rank, M. L., Hemmsen, M. C., and Kidmose, P. (2021). Investigation of low dimensional feature spaces for automatic sleep staging. *Comput. Meth. Programs Biomed.* 205:106091. doi: 10.1016/j.cmpb.2021.106091

Mousavi, Z., Rezaii, T. Y., Sheykhivand, S., Farzamnia, A., and Razavi, S. (2019). Deep convolutional neural network for classification of sleep stages from single-channel EEG signals. *J. Neurosci. Methods.* 324:108312. doi: 10.1016/j.jneumeth.2019.108312

Khalili, E., and Asl, B. M. (2021). Automatic sleep stage classification using temporal convolutional neural network and new data augmentation technique from raw single-channel EEG. *Comput. Meth. Programs Biomed.* 204:106063. doi: 10.1016/j.cmpb.2021.106063

Stevner, A. B. A., Vidaurre, D., Cabral, J., Rapuano, K., Nielsen, S. F. V., Tagliazucchi, E., et al. (2019). Discovery of key whole-brain transitions and dynamics during human wakefulness and non-REM sleep. *Nat. Commun.* 10, 1–14. doi: 10.1038/s41467-019-08934-3

Tagliazucchi, E., Wegner, F., Morzelewski, A., Brodbeck, V., Jahnke, K., and Laufs, H. (2013). Breakdown of long-range temporal dependence in default mode and attention networks during deep sleep. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15419–15424. doi: 10.1073/pnas.1312848110

Tagliazucchi, E., and Laufs, H. (2014). Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep. *Neuron* 82, 695–708. doi: 10.1016/j.neuron.2014.03.020

Fu, M., Wang, Y., Chen, Z., Li, J., Xu, F., and Liu, X. (2021). Deep learning in automatic sleep staging with a single channel electroencephalography. *Front. Physiol.* 12:628502. doi: 10.3389/fphys.2021.628502

Lv, J., Liu, D., Ma, J., Wang, X., and Zhang, J. (2015). Graph theoretical analysis of BOLD functional connectivity during human sleep without EEG monitoring. *PLoS One* 10:106063. doi: 10.1371/journal.pone.0137297

Desjardins, M. È, Carrier, J., Lina, J. M., Fortin, M., Gosselin, N., Montplaisir, J., et al. (2017). EEG functional connectivity prior to sleepwalking: evidence of interplay between sleep and wakefulness. *Sleep* 40, 4. doi: 10.1093/sleep/zsx024

Terzano, M. G., Parrino, L., Smerieri, A., Chervin, R., Chokroverty, S., Guilleminault, C., et al. (2001). Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep. *Sleep Med.* 3, 187–199. doi: 10.1016/S1389-9457(02)00003-5

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

Berry, R. B., Budhiraja, R., Gottlieb, D. J., Gozal, D., Iber, C., Kapur, V. K., et al. (2012b). Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events. *Sleep Med. J. Clin. Sleep Med.* 8, 597–619. doi: 10.5664/jcsm.2172

Brignol, A., Al-Ani, T., and Drouot, X. (2012). "EEG-based automatic sleep-wake classification in humans using short and standard epoch lengths," in *Proceedings of the IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE)*. Larnaca: 276–281. doi: 10.1016/j.cmpb.2012.10.002

Siettos, C., and Starke, J. (2016). Multiscale modeling of brain dynamics: From single neurons and networks to mathematical tools. *Wiley Interdiscip. Rev. Syst.* 8, 438–458. doi: 10.1002/wsbm.1348

Rattenborg, N. C., Amlaner, C. J., and Lima, S. L. (2020). Behavioral, neurophysiological and evolutionary perspectives on unihemispheric sleep. *Neurosci. Biobehav. Rev.* 24, 817–842. doi: 10.1016/S0149-7634(00)00039-7

Baptista, M. S., Kakmeni, F. M., and Grebogi, C. (2010). Combined effect of chemical and electrical synapses in Hindmarsh-Rose neural networks on synchronization and the rate of information. *Phys. Rev. E.* 82:036203. doi: 10.1103/PhysRevE.82.036203

Guo, X., Simas, T., Lai, M. C., Lombardo, M. V., Chakrabarti, B., Ruigrok, A. N., et al. (2019). Enhancement of indirect functional connections with shortest path length in the adult autistic brain. *Hum. Brain Mapp.* 40, 5354–5369. doi: 10.1002/hbm.24777

Liu, Y., Yan, L., Zeng, B., and Wang, W. (2010). "Automatic sleep stage scoring using Hilbert-Huang transform with BP neural network," in *Proceedings of the 4th International Conference on Bioinformatics and Biomedical Engineering*. Chengdu: 1–4.

Gao, Q., Zhou, J., Ye, B., and Wu, X. (2015). Automatic sleep staging method based on energy features and least squares support vector machine classifier. *Int. J. Biomed. Eng. Technol.* 32, 531–536.

Diykh, M., Li, Y., Wen, P., and Li, T. (2018). Complex networks approach for depth of anesthesia assessment. *Measurement* 119, 178–189. doi: 10.1016/j.measurement.2018.01.024

Brázdil, M., Babiloni, C., Roman, R., Daniel, P., Bares, M., Rektor, I., et al. (2009). Directional functional coupling of cerebral rhythms between anterior cingulate and dorsolateral prefrontal areas during rare stimuli: A directed transfer function analysis of human depth EEG signal. *Hum. Brain Mapp.* 30, 138–146. doi: 10.1002/hbm.20491

Bortel, R., and Sovka, P. (2006). EEG–EMG coherence enhancement. *Signal Process.* 86, 1737–1751. doi: 10.1016/j.sigpro.2005.09.011

Chen, B., Ma, R., Yu, S., Du, S., and Qin, J. (2019). Granger causality analysis based on quantized minimum error entropy criterion. *IEEE Signal Process. Lett.* 26, 347–351. doi: 10.1109/LSP.2019.2890973

Caballero-Gaudes, C., Ville, D. V., Grouiller, F., Thornton, R., Lemieux, L., Seeck, M., et al. (2013). Mapping interictal epileptic discharges using mutual information between concurrent EEG and fMRI. *Neuroimage* 68, 248–262. doi: 10.1016/j.neuroimage.2012.12.011

Yüce, A. B., and Yaslan, Y. (2016). "A disagreement based co-active learning method for sleep stage classification," in *Proceedings of the 2016 international conference on systems, signals and image processing (IWSSIP)*. Bratislava, ACT: IEEE, 1–4.

Diykh, M., Li, Y., and Wen, P. (2016). EEG sleep stages classification based on time domain features and structural graph similarity. *IEEE Trans. Neural Syst. Rehabil. Eng* 24, 1159–1168. doi: 10.1109/TNSRE.2016.2552539

Phan, H., Andreotti, F., Cooray, N., Chén, O. Y., and De Vos, M. (2018). "DNN filter bank improves 1-max pooling CNN for single-channel EEG automatic sleep stage classification," in *Proceedings of the 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. Honolulu, ACT: IEEE, 453–456. doi: 10.1109/EMBC.2018.8512286

Zhou, J., Wang, G., Liu, J., Wu, D., Xu, W., Wang, Z., et al. (2020). Automatic sleep stage classification with single channel EEG signal based on two-layer stacked ensemble model. *IEEE Access* 8, 57283–57297. doi: 10.1109/ACCESS.2020.2982434

Lachaux, J. P., Rodriguez, E., Martinerie, J., and Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* 8, 194–208. doi: 10.1002/(SICI)1097-019319998:4

Quiroga, R. Q., Kraskov, A., Kreuz, T., and Grassberger, P. (2010). Performance of different synchronization measures in real data: a case study on electroencephalographic signals. *Phys. Rev. E* 82:036203. doi: 10.1103/PhysRevE.65.041903

Chang, C. C., and Lin, C. J. (2000). LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 1–27. doi: 10.1145/1961189.1961199

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297.

Dkhil, M. B., Chawech, N., Wali, A., and Alimi, A. M. (2017). "Towards an automatic drowsiness detection system by evaluating the α band of EEG signals," in *Proceedings of the IEEE 15th international symposium on applied machine intelligence and informatics (SAMI)*. Herl'any, ACT: IEEE, 000371–000376.

Knaut, P., von Wegner, F., Morzelewski, A., and Laufs, H. (2019). EEG-correlated fMRI of human α (de-) synchronization. *Clin. Neurophysiol.* 130, 1375–1386. doi: 10.1016/j.clinph.2019.04.715

Sharma, M., Tiwari, J., and Acharya, U. R. (2021). Automatic sleep-stage scoring in healthy and sleep disorder patients using optimal wavelet filter bank technique with EEG signals. *Int. J. Environ. Res. Public Health* 18:3087. doi: 10.3390/ijerph18063087

Tripathy, R. K., Ghosh, S. K., Gajbhiye, P., and Acharya, U. R. (2020). Development of automated sleep stage classification system using multivariate projection-based fixed boundary empirical wavelet transform and entropy features extracted from multichannel EEG signals. *Entropy* 22:1141. doi: 10.3390/e22101141

Zhao, C., Li, J., and Guo, Y. (2022). SleepContextNet: A temporal context network for automatic sleep staging based single-channel EEG. *Comput. Meth. Prog. Biomed.* 220:106806. doi: 10.1016/j.cmpb.2022.106806

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 03 February 2023
DOI 10.3389/fnins.2023.1113593

OPEN ACCESS

EDITED BY
Minpeng Xu,
Tianjin University, China

REVIEWED BY
Qingshan She,
Hangzhou Dianzi University, China
Chao Chen,
Tianjin University of Technology, China

*CORRESPONDENCE
Yongqing Zhang
✉ zhangyq@cuit.edu.cn

SPECIALTY SECTION
This article was submitted to Neuroprosthetics, a section of the journal Frontiers in Neuroscience

RECEIVED 01 December 2022
ACCEPTED 17 January 2023
PUBLISHED 03 February 2023

CITATION
Wang M, Zhou H, Li X, Chen S, Gao D and Zhang Y (2023) Motor imagery classification method based on relative wavelet packet entropy brain network and improved lasso. *Front. Neurosci.* 17:1113593. doi: 10.3389/fnins.2023.1113593

Motor imagery classification method based on relative wavelet packet entropy brain network and improved lasso

Manqing Wang[1,2], Hui Zhou[2], Xin Li[2], Siyu Chen[2], Dongrui Gao[1,2] and Yongqing Zhang[2]*

[1]School of Life Sciences and Technology, University of Electronic Science and Technology of China, Chengdu, China, [2]School of Computer Science, Chengdu University of Information Technology, Chengdu, China

Motor imagery (MI) electroencephalogram (EEG) signals have a low signal-to-noise ratio, which brings challenges in feature extraction and feature selection with high classification accuracy. In this study, we proposed an approach that combined an improved lasso with relief-f to extract the wavelet packet entropy features and the topological features of the brain function network. For signal denoising and channel filtering, raw MI EEG was filtered based on an R^2 map, and then the wavelet soft threshold and one-to-one multi-class score common spatial pattern algorithms were used. Subsequently, the relative wavelet packet entropy and corresponding topological features of the brain network were extracted. After feature fusion, mutcorLasso and the relief-f method were applied for feature selection, followed by three classifiers and an ensemble classifier, respectively. The experiments were conducted on two public EEG datasets (BCI Competition III dataset IIIa and BCI Competition IV dataset IIa) to verify this proposed method. The results showed that the brain network topology features and feature selection methods can retain the information of EEG more effectively and reduce the computational complexity, and the average classification accuracy for both public datasets was above 90%; hence, this algorithms is suitable in MI-BCI and has potential applications in rehabilitation and other fields.

KEYWORDS

motor imagery, brain function network, lasso, relief-f, brain-computer interface

1. Introduction

As a new interactive technology, brain-computer interface (BCI) combines biomedical and computer fields to establish a connection between human brain and computer, and continuously expand its application in recent years (Zhao et al., 2020). Among various BCI systems, motor imagery (MI) BCI collects the brain electrical signals during imaginary limb movements of subjects, which is proposed as a candidate approach in motor skill learning and medical rehabilitation (Bigirimana et al., 2020). However, compared with other BCI systems such as P300 and steady-state visual-evoked potential BCI, MI BCI presents a poor performance (Park et al., 2021).

Previous classification tasks of motor imagery primarily focused on improving the feature extraction algorithm. Owing to the characteristics of electroencephalogram (EEG) signals, the common spatial pattern (CSP) algorithm is often used to extract features in the spatial domain (Sharma et al., 2018). In 2018, David et al. proposed a regularized CSP method based on frequency bands and sorted the mutual information between the frequency bands to extract the features. Then they calculated the distance between the feature and label using the second normal form, and performed classification with the nearest neighbor (Park et al., 2018). Zhang et al. proposed a CSP algorithm that optimized both the filter band and the time window to extract features, and an accuracy rate of 88.5% was achieved on the BCI public four-category dataset with a support vector machine (SVM) classification (Jiang et al., 2020). In 2018, Vasilisa proposed a feature weighting and regularization method to optimize the current CSP method to avoid loss of feature information. After the minimum Mahalanobis distance classification, the accuracy of the four-class dataset reached 88.6% (Mishuhina and Jiang, 2018). These mentioned improved CSP algorithm overcomes some of the problems of the traditional CSP algorithm, it still exhibits certain shortcomings, such as it is unsuitable for processing multiclass EEG data.

In addition to feature extraction, studies have been made to improve the performance of feature selection and classification algorithm. In Udhaya Kumar and Hannah Inbarani (2017), the particle swarm optimization (PSO) algorithm combined with a rough set was used to retain features which contribute to the classification accuracy. With the neighborhood rough set classifier, the final average classification accuracy rate in the IIa dataset in BCI competition IV reached 73.1%. In Selim et al. (2018) selected the most distinctive CSP features and optimized SVM parameters by applying a hybrid attractor metagene algorithm and a bat optimization algorithm, and obtained an average classification accuracy rate of 78.3% in the same dataset as that mentioned above (Chu et al., 2018). At this stage, owing to the rapid development of the Riemannian geometry, researchers have used the Riemann minimum distance for pattern classification of EEG signals. In 2019, Javier proposed an improved contraction covariance matrix to handle small sample data more effectively, and subsequently processed the IIa dataset through the Riemann minimum mean distance classifier, and the average classification accuracy rate reached 79.6% (Olias et al., 2019). However, some problems persisted in Riemannian approaches, for example, as the number of the dimension of the covariance matrix rises, the worst the accuracy become (Yger et al., 2017).

To improve the accuracy of feature classification, a new algorithm model based on improved lasso and relief-F was designed in this study. During feature extraction, the relative wavelet packet energy entropy feature of the EEG signal, as well as the variance and mean of the multiclass score common spatial pattern (mSCSP) were extracted. These three features can not only effectively extract the time-frequency-spatial domain information of the signal, but also are suitable for analyzing biological non-stationary signals. Subsequently, feature fusion was performed on the obtained features to overcome the problem of low classification accuracy caused by a single feature. To address the redundancy and high computational complexity of fusion features, a feature selection method based on mutcorLasso and the relief-F algorithm was proposed to retain important features and eliminate redundant ones. Finally, four different classifiers were used to verify the effect of classification, including the K nearest neighbor (KNN), contraction linear discriminant analysis (sLDA), random forest (RF), and an ensemble classifier (Ensemble).

2. Materials and methods

In order to improve the MI EEG classification accuracy, a recognition method based on brain network and improved lasso was proposed in this paper. A flowchart of the proposed model is shown in Figure 1, which includes data introduction, preprocessing, feature extraction, feature selection, and classification. The feature extraction algorithm mentioned is based on the brain network model framework. The edge weight is set according to the relative wavelet packet entropy, and the threshold selection is based on the global network sparsity when the brain network is constructed. In addition, a feature selection method based on lasso method and presents some improvements to the traditional lasso was proposed. The mutual information and correlation between features are considered for the construction of the objective function of lasso, and then the relief-f algorithm is added for further feature selection.

2.1. Data preprocessing

The first step of preprocessing is to remove bad channels with low signal-to-noise ratio by interpolation or average. The next step is band-pass filtering which significantly affect the classification performance of EEG. In this study, the R^2 map is calculated using the power spectral density (PSD) to obtain the frequency band that contains the largest amount of information for each dataset (Choi et al., 2020). In addition, because the signal-to-noise ratio of EEG is extremely low, the data must be denoised and the wavelet soft threshold method was used to perform denoising. The above three steps are serial processing to avoid confusion caused by the entanglement of Midway data.

2.1.1. Wavelet soft threshold denoising algorithm

When the EEG signal undergoes a wavelet decomposition, the amplitude of the wavelet coefficients of EEG is greater than the noise. The noisy signal is decomposed by the orthogonal wavelet base at various scales at a low resolution (Khoshnevis and Sankar, 2020). For the decomposition value at high resolution, the wavelet coefficients whose amplitude is below the threshold were set to zero, and the wavelet coefficients above the threshold are reduced correspondingly or directly retained. Finally, the wavelet coefficients obtained after processing are reconstructed using the inverse wavelet technique, and the denoised EEG is restored.

2.1.2. Multi-score common spatial pattern

The spatial filtering technique is suitable for processing the multidimensional signals, such as EEG (Park et al., 2014). This algorithm mainly improves the CSP algorithm to select EEG channels. By calculating the score of the projection matrix for all the channels, the channel with the highest score for each class is selected and combined to obtain the optimal filter channels. The algorithm not only maximizes the variance difference between classes but also reduces the cost of computing resources.

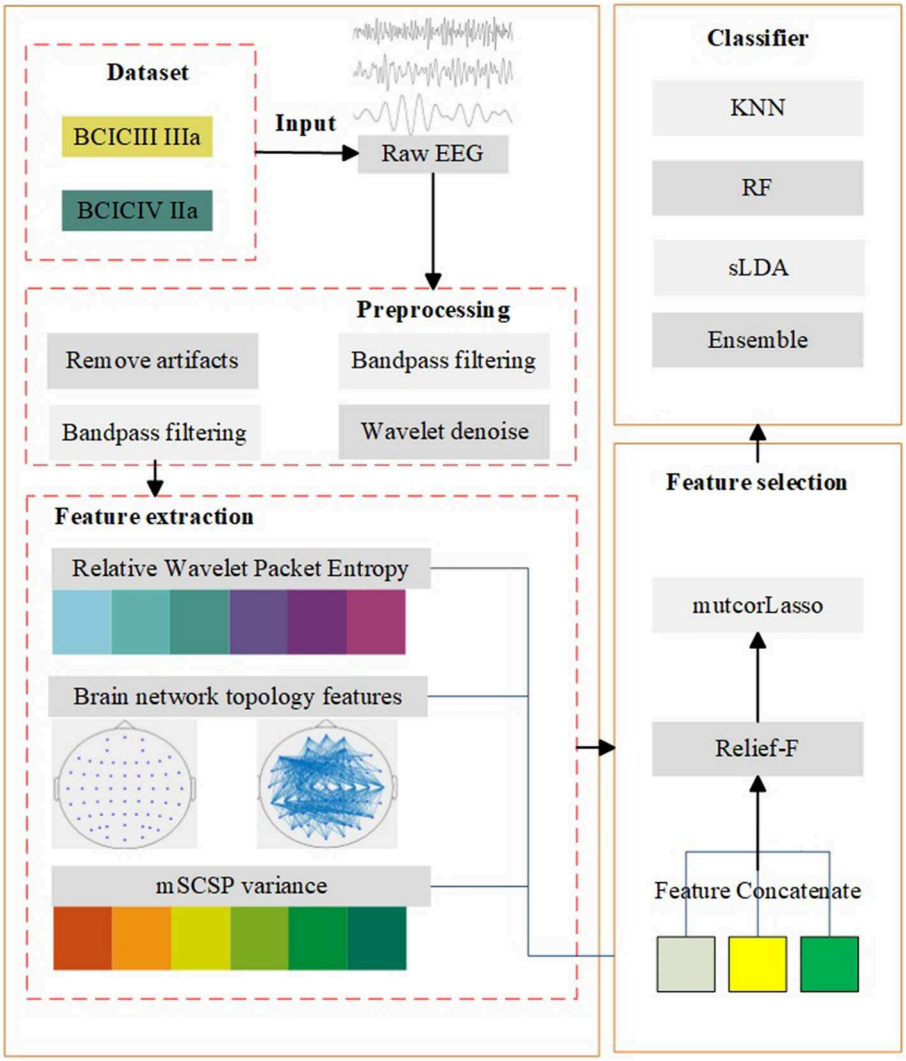


FIGURE 1
Algorithm model. The diagram consists of five main parts: data introduction, preprocessing, feature extraction, feature selection, and classification.

2.2. Feature extraction

In this paper, the wavelet packet method is used to extract the detail and approximate coefficients of EEG. The energy entropy values of these coefficients are calculated, and a brain function network based on these energy entropy values is constructed to extract the topological features. Because the mSCSP algorithm in the previous step amplifies the variance of different samples, the variance characteristics of each sample are also extracted. Finally, the three parts of features are fused to obtain a higher-dimensional matrix. However, the dimensions of the features extracted by the above three different feature extraction algorithms are different, resulting in the situation that some features with large dimensions may have a great impact on the screening results in the subsequent feature screening. Therefore, the feature matrix is standardized and the features with different dimensions are compressed to the range of [0,1] for subsequent processing. The two main feature extraction methods used in this study are as follows.

2.2.1. Relative wavelet packet entropy

Currently, relative wavelet packet entropy has been widely used in processing EEG data. It can efficiently extract the time–frequency domain information, and the low frequency of EEG can be reduced by wavelet packet decomposition technology. Meanwhile, the high-frequency information are extracted to reflect the time–frequency domain information of this part of the EEG signal more effectively. This wavelet packet decomposition method has no redundancy and omissions, therefore, it can perform an efficient time–frequency localization analysis on EEG that contain a large amount of medium and high-frequency information.

In this study, the EEG signal is decomposed into three layers. Therefore, the approximate and detail coefficients of the three layers are obtained, which are $A_j, j = 1, 2, 3$ and $D_j, j = 1, 2, 3$ where j represents the number of decomposition layers. The formula for calculating the energy coefficient of each layer was as follows:

$$E_j = (A_j(k) + \sum_k D_j(k)^2)/3 \quad (1)$$

where k represents the k-th channel, the approximation coefficients A_j are averaged, and the detail coefficients D_j are used in the second norm. Therefore, both the detail coefficients and approximation coefficients are considered as the energy value of each layer.

Furthermore, because the approximation coefficient is more important in the analysis of EEG signals, the original value of the approximation coefficient is directly used, whereas the detail coefficient is used as part of the energy coefficient. The formula to calculate the total energy is as follows:

$$E_t = \sum_j E_j^2, \tag{2}$$

where E_t represents the total wavelet packet energy value (Dimitrakopoulos et al., 2018). The relative wavelet energy value can be obtained from the two formulas above, and the specific formula is as follows:

$$P_j = \frac{E_j}{E_t} \tag{3}$$

where $\sum_j P_j = 1$, and the distribution of P_j can be used as an important feature of the EEG time–frequency domain. Next, based on the Shannon entropy theory, the wavelet packet energy entropy was calculated (Li and Zhou, 2016). The specific formula is as follows:

$$S_m = -\sum_j (P_j \ln(P_j)), \tag{4}$$

where S_m represents the relative wavelet packet energy entropy of channel m. Based on the formula, the value between channels can be calculated, which provide a foundation for building a brain function network for each dataset.

2.2.2. Brain network

The method to construct a brain function network can be primarily classified into following four steps:

Node definition: Each channel electrode after channel selection is used as a node to construct the brain network.

Weight calculation: The weight value of the edge in this experiment is the relative energy entropy of wavelet packet designed in the previous section.

Threshold definition: The threshold selection criterion used in this experiment is based on sparsity, which is determined as the 30% sparsity standard to ensure that each node is not an isolated node and that the network complexity is low. This is more suitable for subsequent processing.

Topological feature extraction: It is primarily aimed at several typical topological features of the constructed brain network, including the degree of the node, clustering coefficient of the node, global efficiency of the brain network, and characteristics of the first and spectral norms of the brain network. The specific formulas are as follows Lee et al. (2018):

The formula for node degree parameter is as follows:

$$k_i = \sum_{j \in N} (R_{ij}) + \sum_{j \in N} (R_{ji}), \tag{5}$$

where R_{ij} and R_{ji} indicate the edge from node i to node j and the edge from node j to node i exist, respectively. The N represents the total set of features extracted from the brain topology network, and k_i represents the degree of node i, which is calculated by the sum of the outgoing and incoming paths of the node. After calculating the degree of the node, it can be used to calculate the clustering coefficient of the brain network. The specific formula for the calculation is as follows Kakkos et al. (2019):

$$C = \frac{t}{k_i * (k_i - 1)} \tag{6}$$

where t represents the number of triangles around node i. The clustering coefficient can reflect the universality of cluster connections around a single node; therefore, it is often analyzed as a feature of the brain function network (Horn et al., 2014). Another feature is the global efficiency of the brain network, which can reflect the degree of connectivity of the entire brain network. The specific formula used for calculation is as follows:

$$E = \frac{1/\sum_{j \in N, j \neq i} d_{ij}}{N - 1}, \tag{7}$$

where d_{ij} represents the shortest distance from node i to node j. The shortest distance was calculated using Dijkstra's algorithm. The starting point is taken as the center and expand outward layer by layer (breadth first search idea) until it is extended to the end point. The order of increasing length produces the shortest path used in this algorithm. That is, after sorting the path lengths of all visible points each time, this algorithm select the shortest path from the corresponding vertex to the source point. Therefore, this algorithm is more suitable for EEG than prim algorithm or Freud algorithm.

The nodes of brain network are defined by reconstructing different node positions on the electrode cap and the corresponding path is composed of the relative wavelet packet entropy coefficient. Then the threshold is set to determine the sparsity of the brain network construction to avoid high computational complexity and feature redundancy. The topological characteristics of these three parts of the brain network can fit the information of entire brain network.

2.3. Feature selection

Owing to the higher dimension of the matrix after feature fusion, a significant amount of computing resources is consumed. Therefore, the lasso method based on mutual information and correlation combined with the relief-f method is used for feature selection. Finally, the feature matrix with smaller dimensions is selected, which could reduce the computational complexity and ensure a higher classification performance. The specific details of these two algorithms are as follows:

2.3.1. mutcorLasso

During data training, hundreds or even thousands of variables are involved. Therefore, there are possibilities of overfitting when the dependent variable of the objective function is measured using several variables. Lasso-based methods can be used to perform filtering more efficiently by eliminating some nonessential variables. Therefore, both discrete and continuous data can be processed. In this paper, we propose a lasso method based on mutual information and correlation,

which is an improvement on the traditional lasso algorithm. It considers the mutual information and correlation information of features and labels followed by optimization. We modified the objective function of the traditional lasso algorithm, and objective function proposed is as follows:

$$min(\|\mathbf{y}^\top - \mathbf{w}^\top X\|_2^2 + \alpha\|w\|_1 + \beta\mathbf{w}^\top Cw), \tag{8}$$

where y and X are formula elements in the traditional Lasso algorithm, y represents the label of the dataset, and X represents the characteristic matrix calculated according to the least squares method. C is the squared mutual information correlation matrix, and w is the weight coefficient of each feature vector. α and β are the learning rates that control the optimization speed of the entire objective function. If the setting is extremely small, the local optimal value can be obtained easily; if extremely large, the amplitude of result fluctuates significantly, and the global optimal value can be obtained easily. In this study, the initial value of α is set to 0.5 and β to 0.1, the values are updated to get a better accuracy based on these two parameter combinations. The formula to calculate matrix C is as follows:

$$C = R \odot R, \tag{9}$$

where R is the coefficient matrix of the mutual information and correlation, $\odot$ represents the Hadamard product, and the formula to calculate each element in the R matrix is as follows:

$$r_{kl} = \frac{\sum_{i=1}^{n} x_{ki}x_{li}}{\sqrt{\sum_{i=1}^{n} x_{ki}^2}\sqrt{\sum_{i=1}^{n} x_{li}^2}} + |mutInf\ (x_{ki}, x_{li})\ |, \tag{10}$$

where *mutInf* represents the mutual information between two feature vectors. Using this formula, each coefficient of matrix R is obtained and used as the basis for the optimization of subsequent objective function. After the general feature is selected using the algorithm above, the feature dimension is still large. Therefore, the experiment will be proceeded using the relief-f algorithm, which is typically used at this stage to perform further feature selection.

After feature filtering by the above method, the dimension of feature vector is reduced from 120 to 20, and the relevant redundant features are eliminated.

2.3.2. Relief-f

The basic principle of the algorithm is as follows: first, samples R are randomly selected from training set D and the k nearest neighbor samples H are obtained from the same type R. Subsequently, the k nearest neighbor samples M are selected from samples of different types from R. Finally, the feature weight is updated using this formula.

In view of the overall dimensions of the dataset and information from relevant studies, we set the k nearest neighbor samples to six. To ensure that each sample type is randomly selected, we control the random sampling rate required by the algorithm to be within 30–40%. The distance function is marginally modified, and the distance is set to the absolute value of the difference between elements in two feature vectors, thereby reducing calculation complexity and reflecting the difference between random and selected samples. Finally, the statistics on the w value after traversal are obtained, the w value of each feature vector is sorted, and feature matrix of lower dimensions is selected. In this algorithm, the update formula of feature weight w is as follows:

$$\begin{aligned} w = w &- \sum_{j=1}^{k} diff\ (A, R, H)\ /(mk) \\ &+ \sum_{c \in class(R)} \frac{p(C)}{1 - p(class(R))} \frac{\sum_{j=1}^{k} diff\ (A, R, H)}{(mk)^2} \end{aligned} \tag{11}$$

where *diff*() represents the difference between the R and H samples on feature A, and mk represents the number of total samples. According to the formula, the w coefficient can be continuously updated.

After feature filtering by the above method, the dimension of feature vector is reduced from 20 to 10, So it is better suitable for classification tasks with low time complexity.

2.4. Classifier

The last component pertains to classification. Four classifiers were used in the experiment, namely KNN, sLDA, RF, and the Ensemble obtained by integrating the three classifiers. These four classifiers can verify whether the proposed algorithm is universal.

KNN is determined by voting the unlabeled samples by the K nearest neighbors (Bablani et al., 2018). sLDA is an improved version of linear discriminant analysis, which is more applicable when the number of training samples is less than the number of features (Tjandrasa and Djanali, 2016). RF is an extension of the traditional decision tree classification algorithm that adds knowledge in the integrated learning field and performs decision classification based on multiple decision trees (Lanata et al., 2020). After verifying the classification accuracy for different number of decision tree on the datasets, we set the number to 10 in the RF. Ensemble is integrated according to the prediction labels finally obtained using the three classifiers mentioned, and it uses the voting method to predict the labels of final ensemble classifier.

The five components above are the specific description of the algorithm model. The following sections focus on the new algorithm proposed herein in feature extraction and feature selection. The pseudo code of the feature selection algorithm above is shown below Table 1.

2.5. Evaluating indicator

The evaluation indicators used in this experiment was accuracy. It is the most important index in the entire classification system and is obtained based on the confusion matrix. The specific formula to calculate it as follows:

$$Accuracy = \frac{True_{num}}{Total_{num}}, \tag{12}$$

where $True_{num}$ indicates the number of samples correctly classified, and the $Total_{num}$ indicates the total number of samples. All the data in the result tables are obtained through 10 fold cross validation.

TABLE 1 The pseudo code of the feature selection algorithm.

Algorithm: Procedure of mutcorLasso+Relief-f

Input: Feature vector matrix X ∈ R^{p*n}, corresponding label $y \in R^p$,

The maximum number of iterations $iter_{num}$

mutcorLasso method:

1: **For** i from 1 to n do:

2: **For** j from 1 to n do:

3: Calculate mutual information and correlation using Equation (10)

4: **end for**

5: Construct matrix C based on the above coefficients

6: Calculate the Hadamard product following Equation (9)

7: Obtain the fixed matrix B:

8: $B = X * \mathbf{X}^{\top} + \beta * C$

9: Initialize the w coefficient to a random decimal between 0 and 10

10: **While** i < 1,000

11: temp = w

12: Calculate diagonal matrix M:

13: $M^{(t)} = diag(\sqrt{w_1^{(t)}}, \sqrt{w_2^{(t)}}, ..., \sqrt{w_p^{(t)}})$

14: Update weight coefficient w:

15: $w^{t+1} = M^t[M^t B M^t + \frac{\alpha I_p}{2}]^{-1} M^t Xy$

16: **if** t > $iter_{num}$:

17: **break**

18: **end if**

19: According to the value of w, the features with w of 0 are eliminated

Relief-f method:

20: **For** i from 1 to 60:

21: Random select a sample

22: Find 6 neighbor samples from the same class as the sample

23: Find 6 neighbor samples that are different from the sample

24: Update w weight using Equation (11)

25: sort w

26: Select the first N-dimensional features according to the value of w

Output:feature matrix after selecting

TABLE 2 Data description.

Competition	Subjects	Train	Test
Dataset IIIa, BCI-III	K3b (45 Trials/task)	90	90
	K6b (30 Trials/task)	60	60
	L1b (30 Trials/task)	60	60
Dataset IIa, BCI-IV	A01~A09 (9 Subjects)	144	144

Both datasets include four motor imagery tasks of left hand, right hand, foot or tongue movement.

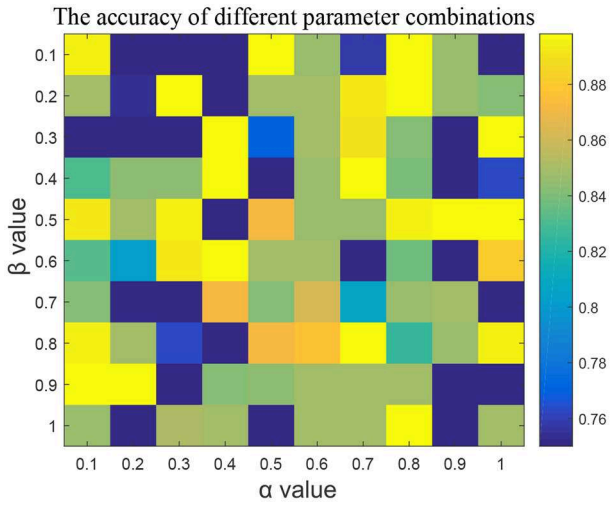


FIGURE 2
The effect of α and β parameters in mutcorLasso algorithm. As the color is closer to yellow, the higher the accuracy of the classification is.

3. Results

3.1. Data description

To demonstrate the effectiveness of the proposed method, we conducted the following experiments on the dataset IIIa in BCI competition III (Blankertz et al., 2006) and the dataset IIa in BCI competition IV (Tangermann et al., 2012), as detailed in Table 2.

3.2. Experimental parameter settings

3.2.1. Setting of α and β in mutcorLasso algorithm

In the feature selection part, we adjusted two parameters, α and β, used in the algorithm and used five-fold cross-validation to verify the results of RF classifier. Taken K3b dataset as example, Figure 2 represented the learning rates α and β of mutcorLasso algorithm, respectively. Although the parameters range between 0.1 and 1, which is relatively small, it affects the classification accuracy very significantly. It could be seen when α was 0.5 and β was 0.1, the accuracy of the K3b dataset was the best among all these values. Similarly, in the other dataset, the adjacency matrix graph was calculated to reflect a better accuracy based on different parameter combinations in mutcorLasso algorithm, and then the best combination of α and β was determined to improve accuracy.

3.2.2. Setting of bandpass filter parameter

We removed the artifacts from raw EEG data and then calculated the PSD of each sample to construct the R^2 chart that reflects the information of different frequency bands. The three graphs in Figure 3 show the bandpass filter for three datasets, in which the ordinate indicates the number of channels, the x-axis indicates the bandpass filter frequency band, and each square indicates the power of each channel in different filter frequency bands. Based on them, the filter band of the k3b dataset was set to 0.5–20, the k6b was set to 3–30, and the l1b was set to 4–40. Similarly, in the IV2a dataset, bandpass filtering was performed based on the relevant R^2 map to obtain more information.

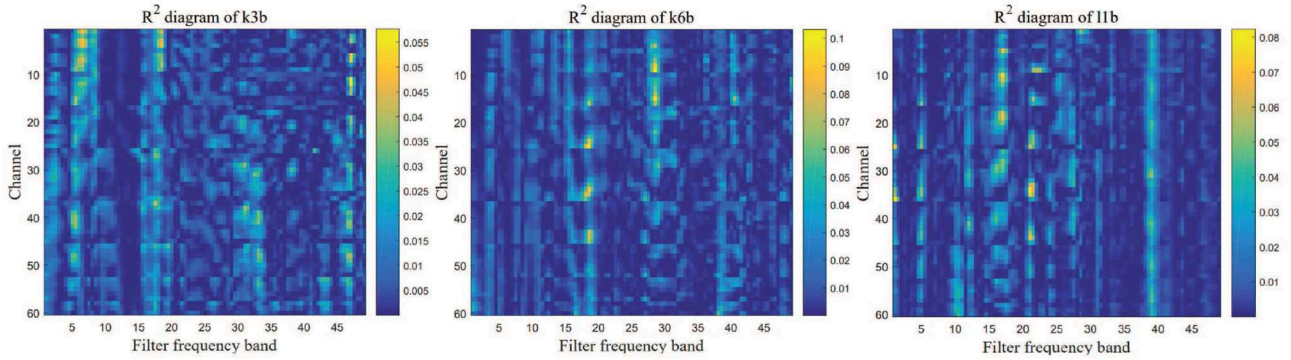


FIGURE 3
R^2 chart for three EEG datasets. The more information the frequency band contains, the closer is the color to yellow. Therefore, the bandpass filter parameters of each dataset can be determined.

3.2.3. Setting of wavelet base in wavelet soft threshold denoising

Because of the varied sampling numbers of the two competition datasets, the wavelet bases for these datasets were different. In this experiment, the wavelet soft threshold method was used to perform denoising. It can be inferred from the Figure 4 that the denoised signal can approximately retain the original value of the original signal, and some high-frequency noise signals are directly eliminated. For the 250 samples in the BCI3 IIIa four-class dataset, db10 was selected as the wavelet base, and for the 1,125 samples in the BCI4 IIa four-class dataset, db20 wavelet base was selected (Yang et al., 2018).

3.2.4. Setting of brain network sparsity parameter

The brain network parameter selection was verified based on the k3b dataset. After constructing the entire brain network model, we compared the effect of sparsity on the brain network model, as shown in Figure 5. The brain network with the sparsity of 10% contains isolated nodes, which could affect the subsequent extraction of the brain network features. In addition, the brain network with the sparsity of 50% shows an extremely dense overall connection of the brain network, which rises the calculation complexity. Thus, the brain network diagram with the sparsity of 30% will be used to build a brain network and the topological features.

3.3. Results of different classifiers on two datasets

3.3.1. Dataset IIIa (BCI-III)

The performance of the algorithm model was verified based on the 50% cross-validation method. This experiment was repeated 20 times and the average accuracy of the entire algorithm model was obtained. In Table 3, the classification accuracy of the existing corresponding algorithms used for Dataset IIIa (BCI-III) is mostly between 80 and 85%, and the accuracy of ensemble classifier exceeded 90% when the training set contained few samples.

3.3.2. Dataset IIa (BCI-IV)

Training set and test set were used at a ratio of 1:1 to verify the algorithm model on dataset IIa. The experiment was repeated 20 times to obtain the variance value of the entire model. In Table 4, the average accuracy of the nine datasets exceeded 80%, which is better than the optimal average value of 80.9% obtained in the previous paper. In particular, the average accuracy obtained by the RF classifier was approximately 90%, which is a significant improvement.The classification accuracy of the existing corresponding algorithms used for Dataset IIa (BCI-IV) was mostly above 85%, suggesting the algorithm model proposed in this paper can achieve good results on this data set.

3.4. Results of fused feature extraction algorithms

In the subsection, the classification accuracy of several feature extraction algorithms mentioned were verified through the five-fold cross-validation method. To avoid the influence of feature selection, the extraction features were directly classified by the RF classifier without feature selection. Table 5 shows that the classification effect of the combination of any two feature extraction methods is better than that of the single feature extraction algorithm alone, and the classification effect obtained by combining the three methods mentioned is the best, approximately 90%. What's more, it's discovered that the variance features obtained by the SCSP facilitated the classification to be the best, followed by the topological features of the brain network.

3.5. Results of feature extraction algorithms

We compared the proposed feature extraction method with other algorithms, including Sparse Filter Bank Common Spatial Pattern (SFBCSP) (Zhang et al., 2015), Temporally Constrained Sparse Group Spatial Pattern (TSGSP) (Yu et al., 2018) and Discrete Wavelet Decomposition (DWT) (Khatun et al., 2016). We randomly combined features extracted from these four methods with mSCSP variance features, then performed feature selection by the relief algorithm, finally obtained the average accuracy of the mentioned

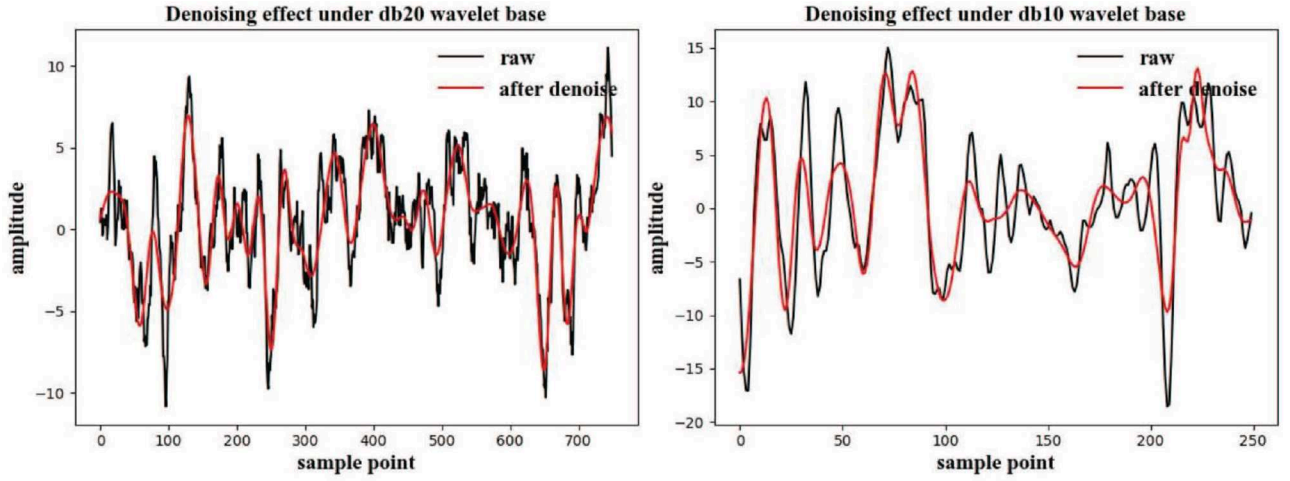


FIGURE 4
Wavelet soft threshold denoising results under different competition datasets. The black line in the figure represents the original signal, and the red line represents the result after the wavelet-based denoising.

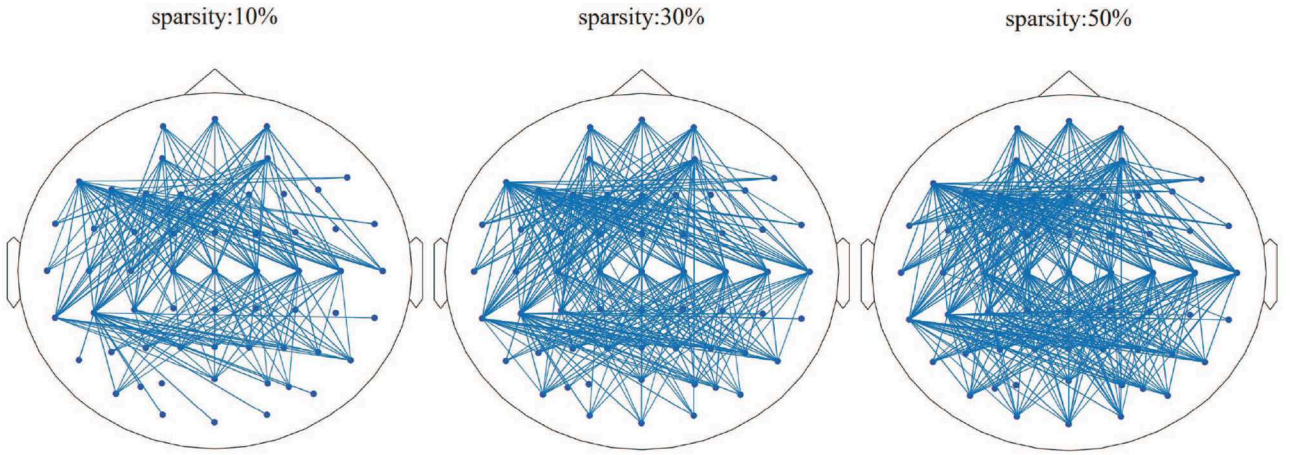


FIGURE 5
Brain network results in different sparsity situations, as 10, 30, and 50%, respectively.

TABLE 3 The classification accuracy of different classifiers on dataset IIIa.

Data	sLDA	RF	KNN	Ensemble
K3b	93.32	89.25	92.62	91.83
K6b	93.71	93.13	91.82	92.19
L1b	92.83	93.19	94.46	93.95

TABLE 4 The classification accuracy of different classifiers on dataset IIa.

Data	sLDA	RF	KNN	Ensemble
A01	86.42	89.03	91.06	90.55
A02	92.35	93.95	92.75	94.26
A03	72.26	88.89	85.74	86.41
A04	64.67	65.86	65.69	65.73
A05	95.81	99.11	97.94	97.94
A06	89.63	99.26	97.62	98.03
A07	90.36	98.33	98.36	97.88
A08	88.98	93.64	92.04	90.25
A09	66.01	77.18	66.35	73.14

classifiers after a five-fold cross-validation. As shown in the Table 6, in datasets IV2a, the average accuracy exceeds 90% by the proposed feature extraction method combining wavelet packet energy entropy and brain network features.The results suggest that the feature extraction method proposed is better than the other three feature extraction methods.

TABLE 5 The classification accuracy for different feature extraction algorithms.

Data	RWPEE	BNTC	mSCSP	RWPEE+BNTC	RWPEE+SCSP	BNTC+SCSP	Proposed
A01	56.16	70.16	79.93	76.96	80.51	85.54	**88.65**
A02	61.15	70.83	79.62	76.56	80.62	82.34	**82.64**
A03	63.84	81.94	84.61	82.24	88.91	84.64	**92.52**
A04	63.46	78.22	82.91	78.33	87.92	89.74	**91.96**
A05	64.04	84.86	83.36	84.85	81.81	**85.84**	**85.84**
A06	67.94	80.49	82.95	80.46	82.53	83.26	**93.76**
A07	68.46	73.12	83.46	73.24	82.13	84.04	**93.59**
A08	60.12	78.52	69.69	78.64	78.81	80.25	**82.21**
A09	67.56	74.43	86.38	75.31	82.21	84.65	**94.53**
Mean	63.61	76.91	81.43	78.35	82.86	84.45	**89.55**

RWPEE refers to relative wavelet packet energy entropy, BNTC refers to the brain network topological characteristics of brain network, mSCSP refers to the one-to-one scoring CSP algorithm variance characteristics.
Bold values represent the highest value.

TABLE 6 The classification accuracy different feature extraction algorithms.

Data	TSGSP	SFBCSP	DWT	Proposed
A01	86.03	82.73	74.83	**88.63**
A02	74.51	74.22	79.61	**82.63**
A03	91.75	88.55	87.71	**92.51**
A04	75.16	69.56	87.95	**91.95**
A05	81.91	76.34	83.15	**83.66**
A06	69.27	64.64	90.55	**93.78**
A07	89.89	86.88	**94.51**	93.51
A08	**94.48**	90.99	90.31	**92.23**
A09	81.95	72.84	89.91	**94.50**
Mean	82.75	78.55	86.73	**90.33**

Bold values represent the highest value.

TABLE 7 The classification accuracy for different feature selection algorithms on dataset IIIa.

Data	Relief-f	Lasso	Lasso+Relief-f	Proposed
K3b	87.84	82.21	89.05	**91.97**
K6b	88.57	82.15	90.25	**92.63**
l1b	86.74	83.33	88.75	**91.55**
Mean	87.75	82.53	89.35	**92.05**

Bold values represent the highest value.

TABLE 8 The classification accuracy for feature selection and comparison algorithms on dataset IV2a.

Data	GBDT	Pearson	Pso+svm	Proposed
A01	91.02	87.71	86.18	**91.91**
A02	84.30	83.42	76.19	**84.83**
A03	90.81	85.61	77.98	**91.52**
A04	**95.22**	88.74	87.41	90.36
A05	88.81	85.91	92.36	**92.75**
A06	**93.92**	90.15	91.68	93.74
A07	92.93	85.76	88.27	**93.49**
A08	85.51	82.88	81.02	**86.48**
A09	69.96	70.44	**74.85**	**74.87**
Mean	88.02	84.54	83.96	**88.85**

Bold values represent the highest value.

3.6. Results of feature selection algorithms

Table 7 compared the proposed mutcorLasso method with relief-f, lasso and the combination of these methods.The feature matrix obtained before feature selection is guaranteed to be exactly the same, but different features are adopted in the feature selection part. The selection algorithm controls the features in 20 dimensions to ensure that the feature dimensions selected using different feature selection algorithms are the same. After 50% cross-validation, the proposed algorithm achieved better accuracy of above 90% than other algorithms.

In Table 8, three feature selection algorithms were compared, including Gradient Boosting Decision Tree (GBDT) (Wang et al., 2019), Pearson correlation coefficient (Pearson) (Xu and Deng, 2018) and Particle swarm optimization (PSO) (Wang et al., 2020). These three algorithms are widely used and representative feature selection algorithms of different kinds. According to the results of these datasets, the feature selection algorithm proposed herein yielded better results on seven datasets, with an average accuracy rate of 88.8%, which is an improvement compared with other three feature selection algorithms, 0.8, 4.3, and 4.9%.

4. Conclusion

The proposed model effectively integrates seven components: bandpass filtering, wavelet denoising, channel filtering, feature extraction, feature fusion, feature selection, and pattern classification.

The main contributions of this study are as follows. Firstly, a complex brain network feature extraction method based on wavelet packet energy entropy was proposed, which not only extracts space–time domain features but also extracts the topological features of the brain network simultaneously, thereby retaining more EEG feature information. Then, a lasso method based on mutual information and correlation was proposed, and the subsequent relief-f algorithm was combined with feature filtering to improve the selected features. The proposed algorithm model can effectively mitigate the problem of low accuracy caused by the scarcity of the training set and achieve precise motion imaging classification. In the future, reducing the computational complexity of the algorithm model and realizing online analysis for a better application in medical rehabilitation will be another research direction of our work.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://www.bbci.de/competition/iii/results/.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. 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

MW and HZ: conceptualization, methodology, and writing—original draft preparation. HZ, XL, and SC: validation. MW: writing—review and editing. DG: visualization and funding acquisition. YZ: supervision. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the Foundation of Science and Technology Innovation 2030 major Project (No. 2022ZD0208500) and the Scientific Research Foundation of CUIT (No. KYQN202241).

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

Bablani, A., Edla, D. R., and Dodia, S. (2018). Classification of eeg data using k-nearest neighbor approach for concealed information test. *Procedia Comput. Sci.* 143, 242–249. doi: 10.1016/j.procs.2018.10.392

Bigirimana, A. D., Siddique, N., and Coyle, D. (2020). Emotion-inducing imagery versus motor imagery for a brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 850–859. doi: 10.1109/TNSRE.2020.2978951

Blankertz, B., Muller, K.-R., Krusienski, D., Schalk, G., Wolpaw, J., Schlogl, A., et al. (2006). The bci competition iii: validating alternative approaches to actual bci problems. *IEEE Trans. Neural Syst. Rehabil. Eng.* 14, 153–159. doi: 10.1109/TNSRE.2006.875642

Choi, H., Noh, G., and Shin, H. (2020). Measuring the depth of anesthesia using ordinal power spectral density of electroencephalogram. *IEEE Access* 8, 50431–50438. doi: 10.1109/ACCESS.2020.2980370

Chu, Y., Zhao, X., Zou, Y., Xu, W., Han, J., and Zhao, Y. (2018). A decoding scheme for incomplete motor imagery EEG with deep belief network. *Front. Neurosci.* 12, 680–688. doi: 10.3389/fnins.2018.00680

Dimitrakopoulos, G. N., Kakkos, I., Dai, Z., Wang, H., and Sun, Y. (2018). Functional connectivity analysis of mental fatigue reveals different network topological alterations between driving and vigilance tasks. *IEEE Trans. Neural Syst. Rehabil. Eng.* 55, 1–15. doi: 10.1109/TNSRE.2018.2791936

Horn, A., Ostwald, D., Reisert, M., and Blankenburg, F. (2014). The structural functional connectome and the default mode network of the human brain. *Neuroimage* 102, 142–151. doi: 10.1016/j.neuroimage.2013.09.069

Jiang, A., Shang, J., Liu, X., Tang, Y., Kwan, H. K., and Zhu, Y. (2020). Efficient csp algorithm with spatio-temporal filtering for motor imagery classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 1006–1016. doi: 10.1109/TNSRE.2020.2979464

Kakkos, I., Dimitrakopoulos, G. N., Gao, L., Zhang, Y., Qi, P., Matsopoulos, G. K., et al. (2019). Mental workload drives different reorganizations of functional cortical connectivity between 2d and 3d simulated flight experiments. *IEEE Trans. Neural Syst. Rehabil. Eng.* 15, 1–13. doi: 10.1109/TNSRE.2019.2930082

Khatun, S., Mahajan, R., and Morshed, B. I. (2016). Comparative study of wavelet-based unsupervised ocular artifact removal techniques for single-channel eeg data. *IEEE J. Transl. Eng. Health Med.* 4, 1–8. doi: 10.1109/JTEHM.2016.2544298

Khoshnevis, S. A., and Sankar, R. (2020). Applications of higher order statistics in electroencephalography signal processing: a comprehensive survey. *IEEE Rev. Biomed. Eng.* 13, 169–183. doi: 10.1109/RBME.2019.2951328

Lanata, A., Sebastiani, L., Di Gruttola, F., Di Modica, S., Scilingo, E. P., and Greco, A. (2020). Nonlinear analysis of eye-tracking information for motor imagery assessments. *Front. Neurosci.* 13, 1431. doi: 10.3389/fnins.2019.01431

Lee, M., Kim, D. Y., Chung, M. K., Alexander, A. L., and Davidson, R. J. (2018). Topological properties of the structural brain network in autism via ϵneighbor method. *IEEE Trans. Biomed. Eng.* 65, 2323–2333. doi: 10.1109/TBME.2018.2794259

Li, T., and Zhou, M. (2016). ECG classification using wavelet packet entropy and random forests. *Entropy* 18, 285. doi: 10.3390/e18080285

Mishuhina, V., and Jiang, X. (2018). Feature weighting and regularization of common spatial patterns in EEG-based motor imagery BCI. *IEEE Signal Process. Lett.* 25, 783–787. doi: 10.1109/LSP.2018.2823683

Olias, J., Martin-Clemente, R., Sarmiento-Vega, M. A., and Cruces, S. (2019). EEG signal processing in mi-bci applications with improved covariance matrix estimators. *IEEE Trans. Neural Syst. Rehabil. Eng.* 15, 1–10. doi: 10.1109/TNSRE.2019.2905894

Park, C., Took, C. C., and Mandic, D. P. (2014). Augmented complex common spatial patterns for classification of noncircular eeg from motor imagery tasks. *IEEE Trans. Neural Syst. Rehabil. Eng.* 22, 1–10. doi: 10.1109/TNSRE.2013.2294903

Park, S., Ha, J., Kim, D.-H., and Kim, L. (2021). Improving motor imagery-based brain-computer interface performance based on sensory stimulation training: an approach focused on poorly performing users. *Front. Neurosci.* 15, 732545. doi: 10.3389/fnins.2021.732545

Park, S., Lee, D., and Lee, S. (2018). Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 498–505. doi: 10.1109/TNSRE.2017.2757519

Selim, S., Tantawi, M. M., Shedeed, H. A., and Badr, A. (2018). A csp-ba-svm approach for motor imagery bci system. *IEEE Access.* 6, 49192–49208. doi: 10.1109/ACCESS.2018.2868178

Sharma, R. R., Varshney, P., Pachori, R. B., and Vishvakarma, S. K. (2018). Automated system for epileptic eeg detection using iterative filtering. *IEEE Sens. Lett.* 2, 1–14. doi: 10.1109/LSENS.2018.2882622

Tangermann, M., Müller, K.-R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., et al. (2012). Review of the BCI competition IV. *Front. Neurosci.* 6, 55. doi: 10.3389/fnins.2012.00055

Tjandrasa, H., and Djanali, S. (2016). "Classification of EEG signals using single channel independent component analysis, power spectrum, and linear discriminant analysis," in *Advances in Machine Learning and Signal Processing. Lecture Notes in Electrical Engineering, Vol. 387*, eds P. Soh, W. Woo, H. Sulaiman, M. Othman, M. Saat (Cham: Springer), 259–268.

Udhaya Kumar, S., and Hannah Inbarani, H. (2017). PSO-based feature selection and neighborhood rough set-based classification for bci multiclass motor imagery task. *Neural Comput. Appl.* 10, 590–598. doi: 10.1007/s00521-016-2236-5

Wang, W., Li, T., Wang, W., and Tu, Z. (2019). Multiple fingerprints-based indoor localization via gbdt: subspace and rssi. *IEEE Access* 7, 80519–80529. doi: 10.1109/ACCESS.2019.2922995

Wang, Y., Wang, D., and Tang, Y. (2020). Clustered hybrid wind power prediction model based on arma, pso-svm, and clustering methods. *IEEE Access* 8, 17071–17079. doi: 10.1109/ACCESS.2020.2968390

Xu, H., and Deng, Y. (2018). Dependent evidence combination based on shearman coefficient and pearson coefficient. *IEEE Access* 6, 11634–11640. doi: 10.1109/ACCESS.2017.2783320

Yang, W., Gao, X., Cao, C., Xiao, F., Hu, K., Zhang, X., et al. (2018). An effective approach for prediction of exposure to base stations using wavelets to fit the antenna pattern. *IEEE Trans. Antennas Propag* 66, 7519–7524. doi: 10.1109/TAP.2018.2874672

Yger, F., Berar, M., and Lotte, F. (2017). Riemannian approaches in brain-computer interfaces: a review. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1753–1762. doi: 10.1109/TNSRE.2016.2627016

Yu, Z., Nam, C. S., Zhou, G., Jing, J., Wang, X., and Andrzej, C. (2018). Temporally constrained sparse group spatial patterns for motor imagery BCI. *IEEE Trans Cybern*. 10, 1–11. doi: 10.1109/TCYB.2018.2841847

Zhang, Y., Zhou, G., Jin, J., Wang, X., and Cichocki, A. (2015). Optimizing spatial patterns with sparse filter bands for motor-imagery based brain computer interface. *J. Neurosci. Methods* 18, 85–91. doi: 10.1016/j.jneumeth.2015.08.004

Zhao, X., Liu, C., Xu, Z., Zhang, L., and Zhang, R. (2020). Ssvep stimulus layout effect on accuracy of brain-computer interfaces in augmented reality glasses. *IEEE Trans. Neural Syst. Rehabil. Eng.* 8, 5990–5998. doi: 10.1109/ACCESS.2019.2963442

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 22 February 2023
DOI 10.3389/fnins.2023.1129049

OPEN ACCESS

EDITED BY
Fangzhou Xu,
Qilu University of Technology, China

REVIEWED BY
Chao Chen,
Tianjin University of Technology, China
Yin Tian,
Chongqing University of Posts
and Telecommunications, China

*CORRESPONDENCE
Mingming Chen
✉ mmchen@zzu.edu.cn

SPECIALTY SECTION
This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neuroscience

RECEIVED 21 December 2022
ACCEPTED 03 February 2023
PUBLISHED 22 February 2023

CITATION
Zhang R, Chen Y, Xu Z, Zhang L, Hu Y and Chen M (2023) Recognition of single upper limb motor imagery tasks from EEG using multi-branch fusion convolutional neural network.
Front. Neurosci. 17:1129049.
doi: 10.3389/fnins.2023.1129049

Recognition of single upper limb motor imagery tasks from EEG using multi-branch fusion convolutional neural network

Rui Zhang, Yadi Chen, Zongxin Xu, Lipeng Zhang, Yuxia Hu and Mingming Chen*

Henan Key Laboratory of Brain Science and Brain-Computer Interface Technology, School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou, China

Motor imagery-based brain-computer interfaces (MI-BCI) have important application values in the field of neurorehabilitation and robot control. At present, MI-BCI mostly use bilateral upper limb motor tasks, but there are relatively few studies on single upper limb MI tasks. In this work, we conducted studies on the recognition of motor imagery EEG signals of the right upper limb and proposed a multi-branch fusion convolutional neural network (MF-CNN) for learning the features of the raw EEG signals as well as the two-dimensional time-frequency maps at the same time. The dataset used in this study contained three types of motor imagery tasks: extending the arm, rotating the wrist, and grasping the object, 25 subjects were included. In the binary classification experiment between the grasping object and the arm-extending tasks, MF-CNN achieved an average classification accuracy of 78.52% and kappa value of 0.57. When all three tasks were used for classification, the accuracy and kappa value were 57.06% and 0.36, respectively. The comparison results showed that the classification performance of MF-CNN is higher than that of single CNN branch algorithms in both binary-class and three-class classification. In conclusion, MF-CNN makes full use of the time-domain and frequency-domain features of EEG, can improve the decoding accuracy of single limb motor imagery tasks, and it contributes to the application of MI-BCI in motor function rehabilitation training after stroke.

KEYWORDS

single upper limb motor imagery, deep learning, brain-computer interface (BCI), convolutional neural network (CNN), feature fusion

1. Introduction

The brain-computer interface (BCI) establishes a channel for information exchange between the human brain and the outside world. It decodes the user's intent through reading and analyzing brain signals (Wolpaw et al., 2002), and has been linked to a wide range of devices, including the use of spellers, wheelchairs, robotic arms and robotic exoskeletons (Kaufmann and Kubler, 2014; Kwak et al., 2015; He et al., 2018; Kim et al., 2018; Penaloza and Nishio, 2018; Yu et al., 2018; Jeong et al., 2019; Yao et al., 2022). Among the various types of BCI paradigms, MI-BCI is one of the most important one because it has potential clinical application value. MI is a mental process that mimics motor intention without

actually eliciting motor behavior. It is an actively evoked EEG signal that has high application values in the field of neurorehabilitation because it can independently elicit potential activity in motor-related brain regions without external stimulation (Pfurtscheller and Neuper, 2001). The MI-BCI detects the user's motor intentions by capturing the potential changes, and the output command could be used to control functional electrical stimulation (FES), exoskeletons, or other rehabilitation assistive equipment (Biasiucci et al., 2018; Zhao et al., 2022). Thus MI-BCI is valuable in the medical rehabilitation pathway for patients with motor dysfunction through the provision of active rehabilitation training (Jeong et al., 2019; Romero-Laiseca et al., 2020). A large number of studies have shown that the addition of MI-BCI helps to promote the recovery of motor function and improve the quality of life of patients (Cervera et al., 2018; Yuan et al., 2021).

The majority of current researches on motor imagery EEG signal recognition focuses on movements of different body parts, such as the tongue, hands, and feet. These studies have produced excellent results, but it is uncommon to find studies on motor imagery EEG signal recognition of tasks that involve the same side of the limb. It is well known that limb motor dysfunction caused by stroke is often unilateral. In BCI-based rehabilitation training, motor imagery tasks using unilateral limbs are more natural and intuitive than motor imagery tasks between different body parts (Tavakolan et al., 2017; Ubeda et al., 2017). However, the classification of single limb motor imagery is more difficult and complex than that of different parts of the body, because similar brain regions are activated when performing different motor tasks for unilateral limbs (Bigdely-Shamlo et al., 2015; Jas et al., 2016; Taulu and Larson, 2021). Considering the low spatial resolution of EEG, it is not feasible to use the algorithms for multi-limb motor imagery EEG recognition to identify unilateral limb motor imagery EEG.

The issue of unilateral limb movement task recognition has begun to be focused on by some researchers in recent years. Edelman et al. (2016) reported that source space analysis can improve the classification accuracy of wrist movements, four different movements of the right hand (i.e., flexion and extension of the arm; left and right rotation of the wrist) were recognized with a classification accuracy of 81.4%. Ofner et al. (2017) encoded motor imagery tasks for the right hand into the time domain of low-frequency EEG signals to classify six different movements, including elbow flexion/extension, forearm left/right rotation, and hand opening/closing, and achieved an accuracy of 27%. A novel classification strategy using the combination of EMG and EEG signals was proposed by Li et al. (2017). They recognized a variety of upper limb movements such as hand open/close and wrist pronation/supination, and results showed that the classification performance achieved by the fusion features of EMG and EEG signals is significantly higher than that obtained by a single signal source of either EMG or EEG across all subjects. Loopez-Larraz et al. (2018) further used EMG activity as a complementary information to EEG to detect the motor intention, and also found that the fusion features achieved higher classification accuracy than EEG or EMG-based methods.

The end-to-end deep learning techniques provide a new development path for the recognition of motor imagery EEG. Inspired by the filter bank common spatial pattern (FBCSP), Schirrmeister et al. (2017) proposed three types of CNN-based models for motor imagery classification based on the number of layers. Jeong et al. (2020b) proposed a hierarchical flow convolutional neural network model consisting of a two-stage CNN for extracting relevant features for multi-class tasks and decoding arm rotation tasks. Zhang X. et al. (2019) proposed a network model CNN-LSTM, the motor imagery EEG data were spatially filtered by the FBCSP algorithm to extract the spatial domain feature information from the original data at first, then the extracted feature were fed into the CNN, and the final classification was performed by the LSTM. Cho et al. (2021) proposed a two-stage network structure called NeuroGrasp, which used six different CNN-BLSTM networks to implicitly map EEG signals to six muscle synergy features based on EMG and generated kinematic images corresponding to the EMG signals based on the extracted features. In the second stage, the generated images and real EMG features are used together as SiamNet network input to train the model, so as to realize the classification of single upper limb motor imagery tasks.

Most of the motor imagery EEG decoding methods based on deep learning used a single type of feature, including raw EEG signals, time-frequency maps, and power spectral density features. However, a single feature input often cannot fully and effectively mine the information related to motor imagery in EEG. Inspired by multimodal classification models, we proposed a multi-branch fusion convolutional network model (MF-CNN) for solving the classification problem of a single upper limb movement imagery task, which takes the EEG signals and the corresponding time-frequency maps as inputs simultaneously to make full use of the time-domain, frequency-domain and time-frequency-domain features of the EEG signal. The original EEG signal has high-resolution temporal information, and the discriminative features can be extracted by spatio-temporal convolution, while the two-dimensional time-frequency map contains rich time-frequency domain and spatial information. In this work, we first extracted the features of the above two inputs independently using two CNNs and then performed fusion classification, and the test results on the single upper limb motor imagery dataset showed that the proposed model achieved higher classification accuracy than single-input CNN.

2. Materials and methods

2.1. Datasets

The EEG data used in this work is the "Multimodal signal dataset for 11 intuitive movement tasks from single upper extremity during multiple recording sessions" from the Giga DB dataset completed by Jeong et al. (2020a). The dataset included intuitive upper limb movement data from 25 subjects, who were required to perform three types of motor tasks in a total of 11 categories, including 6 directions of arm extension movement (up, down, left, right, front, back), 3 kinds of object grasping action (cup, card, ball) and 2 kinds of wrist-twisting action (left rotation, right rotation), each type of movement was randomly executed 50 times, corresponding to 11 movements designed to be associated with each segmental movement of the arm, hand, and wrist, rather than continuous limb movements. The dataset included

not only EEG data but also magnetoencephalography (EMG) and electrooculogram (EOG) data, which are collected simultaneously in the same experimental setting while ensuring no interference between them. The data were acquired using a 60-channel EEG, 7-channel EMG, and 4-channel EOG. In the current work, only motor imagery EEG data were used, the EEG sensors were placed according to the international 10–20 system, and the sampling rate was set as 2,500 Hz. Our goal is to classify the motor imagery EEG of the three types of actions, so we selected forward extension of the arm, grasping the cup, and rotation of the wrist to the left from the three types of actions for the following study.

2.2. Algorithm framework

The workflow of the algorithm was shown in **Figure 1**. The time-frequency maps were firstly obtained by continuous wavelet transform (CWT) method, then both the EEG signals and the corresponding time-frequency maps were sent to the MF-CNN model, which consisted of two CNN network branches. After the process of convolution and pooling, the output features from the two branches were fused and combined into a one-dimensional vector. Finally, the one-dimensional feature vector was sent into a classifier to obtain the prediction results.

2.3. EEG signal pre-processing

When subjects were preparing or performing motor tasks, event-related desynchronization (ERD) and event-related synchronization (ERS) can be observed in the sensorimotor cortex of the brain (Pfurtscheller and da Silva, 1999; McFarland et al., 2000). Therefore, we selected 20 EEG channels on the sensorimotor cortex region to analyze (including FC1-6, C1-6, CP1-6, CZ, and CPZ). The selected EEG data were band-pass filtered within 8–30 Hz (Sreeja et al., 2017) and downsampled to 250 Hz. All the 4 s of EEG data during the motor task of a single trial were intercepted for subsequent processing, thus the EEG segment of each trial could be defined as a 20 × 1,000 matrix, where 20 was the number of channels and 1,000 was the length of the data. The preprocessed EEG signals were used as input for the EEG-CNN branch and the time-frequency map conversion.

In terms of time-frequency map transformation, Tabar and Halici (2017) proposed a method based on the short-time Fourier transform (STFT) to extract time-frequency features and constructed a three-channel stacked time-frequency map for subsequent classification. However, the time window of the STFT algorithm is fixed, so the time-frequency resolution is also fixed, which causes the problem of incompatibility between the time resolution and the spectral resolution. To solve this problem, wavelet transform based time-frequency analysis methods have been widely introduced to EEG signal feature extraction (Zhang et al., 2021). The wavelet transform replaced the infinite-length triangular function with a finite-length wavelet basis with attenuation, which made the window width inconsistent and thus enabled better local feature extraction. We chose Morlet wavelet as the basis function for the wavelet transform. As a single-frequency complex sinusoidal function under Gaussian envelope Morlet wavelet is the most commonly used complex-valued wavelet. Because it has a better local resolution in the time and frequency domain, it is often used in the decomposition of complex signals and time-frequency analysis (Lee and Choi, 2019). The features extracted from EEG signals through CWT include time and frequency information and are finally converted into two-dimensional time-frequency maps. **Figure 2** showed the example time-frequency maps of the three channels C3, C4, and CZ.

Since the conversion of the time-frequency map is generated for each channel individually, the 20 EEG channels we used could not all be combined into one image. And if only a few channels were selected, a lot of helpful information would be lost. To effectively utilize the information of each channel, we preprocessed the data to extract features and used CSP to filter the 20 channels of EEG signals in the spatial domain to obtain "virtual channels," and then generated the time-frequency maps. The basic principle of CSP is to find a set of optimal spatial filters for projection by diagonalizing matrices so that the difference in variance values between the two types of signals is maximized (Ramoser et al., 2000). For the three classification tasks we used the "One vs. Rest" strategy to extend the CSP to achieve multi-class CSP feature extraction (Dornhege et al., 2004). The spatially filtered EEG can be

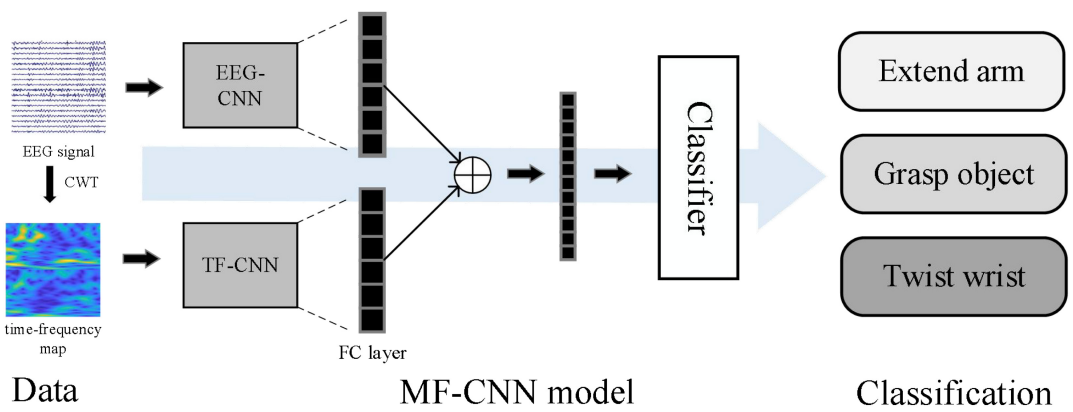


FIGURE 1
Workflow of the proposed algorithm.

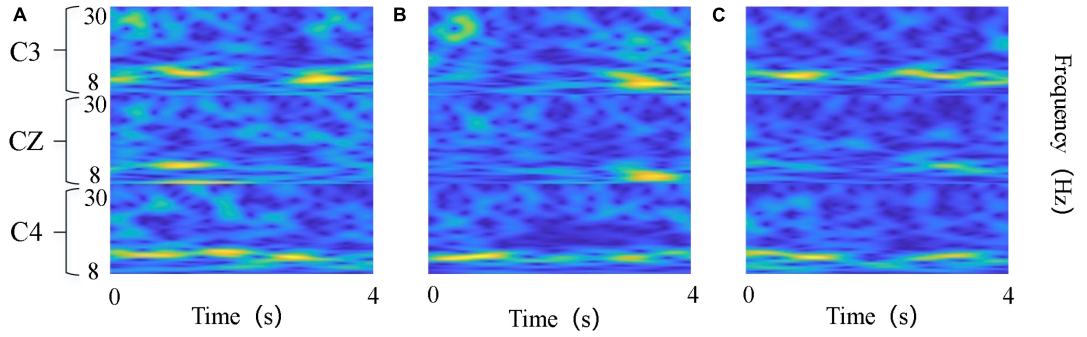


FIGURE 2
Time-frequency maps of the three kinds of tasks. **(A)** Left wrist rotation; **(B)** cup grasping; **(C)** forward arm extension. The abscissa denotes time points, and the ordinate denotes frequency bands.

calculated as:

$$Z_{M \times N} = W_{M \times M} E_{M \times N} \quad (1)$$

where W is the projection matrix of CSP, M is the number of EEG data channels; N is the data length; E is the EEG data matrix; Z is the obtained EEG on "virtual channels."

The information of the feature matrix generated by the CSP algorithm is not equivalent, and the feature information is mainly concentrated in the head and tail of the feature matrix, while the middle feature information is not obvious and can be ignored. Therefore, the first m rows and the last m rows (2 m $<$ M) of $Z_{M \times N}$ were usually selected. In this work, we chose m = 1, that is, the first and the last row of $Z_{M \times N}$ were selected to calculate the time-frequency map. The CWT was applied to the spatially filtered EEG data during the 4 s motor imagery to obtain time-frequency maps, and the maps were then saved as images with a resolution of 64 × 64. Such procedures were applied to all trials, and finally the motor imagery time-frequency map dataset was obtained. An example was shown in Figure 3.

2.4. Structure of MF-CNN

The classification of motor imagery EEG signals using deep learning networks based on CNN has proven successful and has good feature extraction capabilities (Lee and Kwon, 2016; Zhang P. et al., 2019). The common CNN models include convolutional layer, pooling layer, activation function, and fully connected layer. The convolution in the network is a local operation that can extract the deep features of the input signal through the kernel function, then the feature information can be obtained by the operation of each layer of the CNN model. In the convolution phase, the network input is convolved with the convolution kernel, and then the activation function *f(a)* is used to output the feature maps, which can be expressed for each convolution layer as:

$$h_{ij}^{k} = f(a) = f((W^{k} * x)_{ij} + b_{k}) \quad (2)$$

where x represents the input data, $W_k$is the weight matrix of the k_{th} convolution kernel, b_k corresponding to the deviation of the convolution kernel k, i and j denote the number of adjacent convolutional layers.

In the current work, the ReLU function was chosen as the activation function (Clevert et al., 2015), and it was defined as follows:

$$f(a) = ReLU(a) = ln(1 + e^{a}) \quad (3)$$

The main purpose of the fully connected layer in a CNN is classification. To merge the features acquired from the previous side, each node in the fully connected layer is connected to full nodes in the preceding layer. After a number of prior convolutions, it can combine the local information with category differentiation, and the output of the final fully-connected layer is then sent to the classifier to output the prediction result.

Figure 4 showed the network structures of MF-CNN proposed in this study, it extracted the features of the raw EEG data and the time-frequency map simultaneously by using two CNN branches, and could obtain more comprehensive information hidden in the motor imagery EEG.

The EEG-CNN branch extracted spatial and temporal features from the raw EEG data, the dimensionality of the input EEG signal was 20 × 1,000 (channels × points). The input was successively fed through a feature extraction module made up of two convolutional layers and a maximum pooling layer in this branch. A one-dimensional convolutional kernel along the horizontal axis was used to extract the features of each channel to obtain the feature map as the output of this layer. The size of the convolution kernel was set to 3 × 1, and the step size was 1. After convolution, a feature map of the form $N_w \times N_f$ could be obtained, where N_w is the vector and N_f is the number of convolution kernels. Then, the data from the convolutional layer was downsampled using the pooling layer, which set a kernel size of 2 × 1 and a step size of 2. Subsequently, the fully connected layer flattens the features extracted through the convolutional layer.

The TF-CNN branch performed feature extraction on the input time-frequency map, and the size of the time-frequency map was 64 × 64 × 3, which represented an RGB image of size 64 × 64. VGG16 was used as the basic network framework in this branch (Zhao-Hong et al., 2019), the main feature of which was the inclusion of convolutional kernel computation and feedforward structure. It contained 16 hidden layers (13 convolutional layers and 3 fully connected layers), the convolutional part used a convolutional kernel of size 3 × 3 with a step size of 1, and a max pooling layer of size 2 × 2 with a step size of 2.

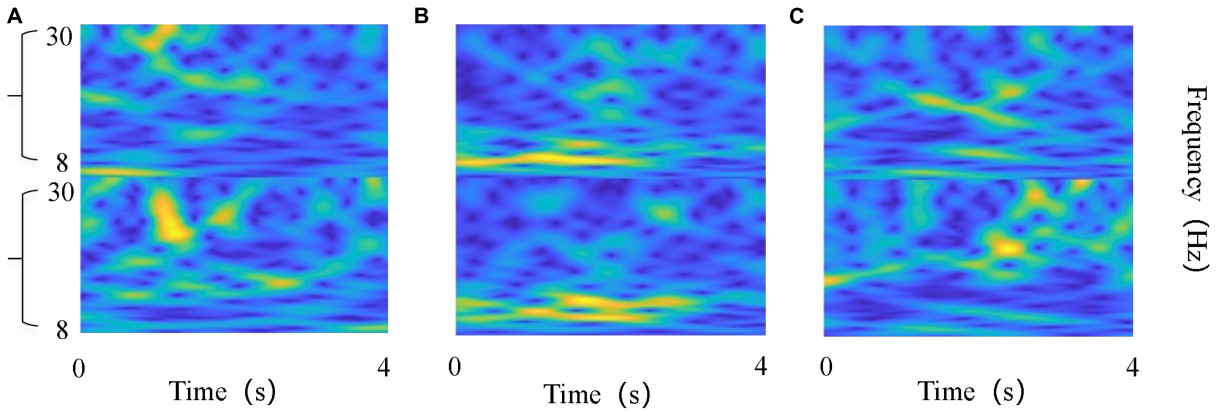


FIGURE 3
Time-frequency maps of the EEG on "virtual channel". **(A)** Left wrist rotation; **(B)** cup grasping; **(C)** forward arm extension. The abscissa denotes time points, and the ordinate denotes frequency bands.

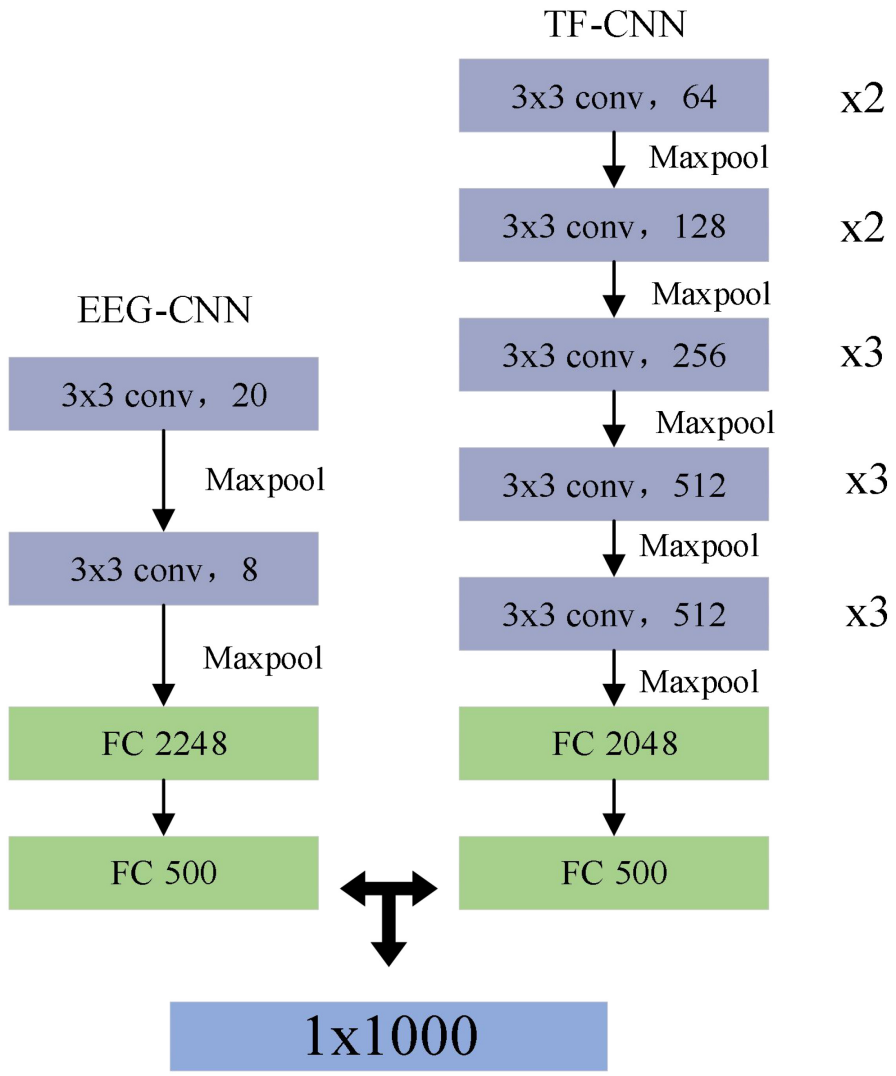


FIGURE 4
Structure of multi-branch fusion convolutional neural network (MF-CNN) model.

In the model training phase of the above two branches, the parameters of each network layer were updated using the Adam optimizer with β_1 0.9 and β_2 0.999, with an initial learning rate of 0.01.

2.5. Feature fusion method

Generally, fusion methods can be applied in two different ways: decision-level fusion and feature-level fusion. Decision-level fusion first trains different modalities with different models and then fuses the results of multiple model outputs. Feature-level fusion combines two or more feature vectors to construct a single feature vector to include more information (Zhang P. et al., 2019; Hatipoglu Yilmaz and Kose, 2021). In this study, feature-level fusion was selected. Before the feature fusion, the individual feature vector must have sufficient relevant features in order to provide a good classification model and achieve high classification performance. In CNN, the fully connected layer can integrate local information into global features for classification, which contains enough information. In addition, the output dimension of the last fully connected layer is consistent with the category of the sample, and the obtained information has been compressed, so it is not appropriate to serve as the final feature vector. Therefore, we chose to use the penultimate fully connected layer of these two branch networks as the fusion layer, and fused their outputs as the extracted features. Suppose the output feature vector of the EEG-CNN branch was $A = \{a_1, \cdots, a_m\}$, where m is the length of A, the feature vector output from the TF-CNN branch was $B = \{b_1, \cdots, b_n\}$, where n is the length of B. Then the fusion feature vector could be defined as $C = \{a_1, \cdots, a_m, b_1, \cdots, b_n\}$, and it is fed into the support vector machine (SVM) to complete the classification finally.

2.6. Performance evaluations

The classification accuracy was used as an evaluation criterion to compare the model's performance, which was calculated as follows.

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (4)$$

where T_P was the true-positives field in the confusion matrix, $T_N$was the true-negatives field, F_P was the false-positives field in the confusion matrix, $F_N$was the false-negatives field. It indicates the probability of correct prediction in all samples. In this paper, we compared the accuracy of six algorithms, including our proposed MF-CNN, the two single-branch CNNs (EEG-CNN and TF-CNN), EEGNET (Lawhern et al., 2018), ALEXNET (Iandola et al., 2016), and the classical CSP. EEG-CNN, EEGNET, and CSP used EEG signals as inputs, which are pre-processed in the same procedures as described in (section "2.3. EEG Signal pre-processing"). TF-CNN and ALEXNET used time-frequency maps as input for image classification.

In addition, we calculated kappa values (Tabar and Halici, 2017).

$$kappa = \frac{P_0 - P_e}{1 - P_e} \quad (5)$$

where p_0 represents the average classification accuracy and $p_e$represents the random classification accuracy for the n-class classification task.

3. Results

In this work, three-class and binary-class classification (grasp object vs. extend arm) test tasks were carried out separately to verify the performance of the proposed algorithm. The classification accuracies were calculated by using the five-fold cross-validation strategy, each subject's EEG data was divided into five equal subsets, one of which was randomly chosen as the testing dataset and the other subsets served as the training dataset. Such procedures were repeated five times, and the average accuracy was determined as the final classification accuracy. The three sessions for each subject were tested separately.

In order to verify the advantages brought by the dual-branch CNN, we compared the classification performance of MF-CNN

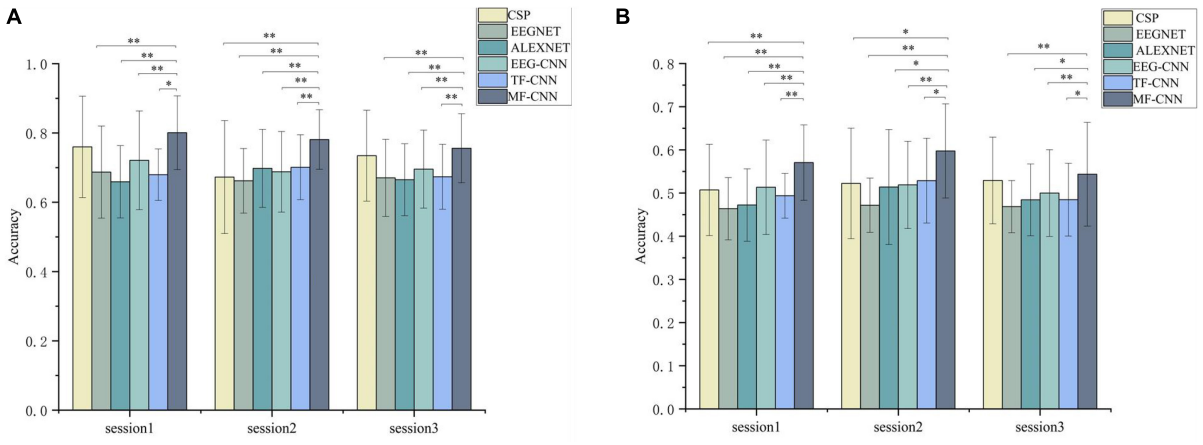


FIGURE 5
Comparison of the average classification results on the three sessions. **(A)** Binary-class classification experiment, **(B)** three-class classification experiments. *Denotes $p < 0.01$ and **denotes $p < 0.001$ (paired t-test).

model and single-branch CNN model. The single-branch CNN model was set up as an EEG-CNN branch for processing raw EEG signals and a TF-CNN branch for processing the time-frequency maps. The network architectures of these two single-branch CNN models were same as the EEG-CNN and TF-CNN branches in MF-CNN.

Figure 5 showed the classification results of the 25 subjects, the average classification accuracies of the single EEG-CNN branch were 70.8 and 51.08% separately for the binary-class and three-class classification experiments, while the single TF-CNN branch achieved 68.4 and 50.24%, respectively. It indicated that discriminative feature information can be extracted by the two kinds of single CNN branches. The accuracy obtained was higher than EEGNET and ALEXNET, but lower than CSP. After merging the features obtained from the two branches, MF-CNN achieved average accuracies of 78.52 and 57.06% for the two classification experiments, both of which were higher than that of the single CNN branch model, and also higher than CSP, EEGNET and ALEXNET.

The statistical analysis was further performed between the four algorithms using paired *t*-test. The results demonstrated that the accuracies achieved by MF-CNN were significantly higher than that of EEG-CNN and TF-CNN in all sessions. In addition, the accuracy of MF-CNN is higher than that of the deep learning algorithms EEGNET and ALEXNET used as comparisons.

The confusion matrix of the three deep learning network models were obtained. As shown in Figures 6, 7, the column represented the true label, and the row represented the predicted label. It can be seen that the probability of correct recognition of each motor imagery task by MF-CNN is higher than that of EEG-CNN and TF-CNN, and all the true positive values are greater than the true negative and false negative values for the three deep learning network models.

Finally, we calculated the kappa coefficient for each subject, and the mean results were shown in Table 1. The binary-class classification experiments obtained higher kappa values than the three-class classification experiment for all three deep learning models, and MF-CNN outperformed EEG-CNN and TF-CNN in the two experiments.

4. Discussion

In this study, we performed feature fusion at the feature level to recognize the single upper limb motor imagery tasks by using deep learning approach. The dataset we used include three different types of movements, including forward extension of the arm, grasping the cup, and rotation of the wrist to the left. These are complex movements of the upper limb of the body and are commonly used in daily life. The accurate classification on the motor imagery of these three movements is of great significance in the application of BCI-based upper limb motor rehabilitation training. In this paper, the MF-CNN model was proposed to extract fusion features from the original EEG signal and corresponding time-frequency map. In the comparative experiment conducted on the single upper limb motor imagery dataset, MF-CNN model achieved

FIGURE 6
Confusion matrix of the three-class classification experiment. **(A)** EEG-CNN, **(B)** TF-CNN, **(C)** MF-CNN.

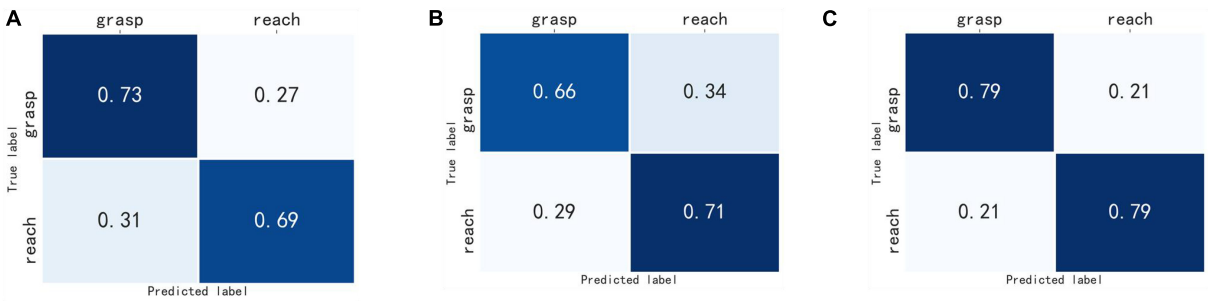


FIGURE 7
Confusion matrix of the binary-class classification experiment. **(A)** EEG-CNN, **(B)** TF-CNN, **(C)** MF-CNN.

TABLE 1 Kappa values of the three deep learning models.

	EEG-CNN	TF-CNN	MF-CNN
Three-class classification	0.2662 ± 0.01	0.2536 ± 0.03	0.3559 ± 0.04
Binary-class classification	0.415 ± 0.04	0.3696 ± 0.02	0.5704 ± 0.02

better classification performance than two single CNN branches, EEGNET, ALEXNET, and CSP.

The EEG signal is non-stationary and non-linear (Yang et al., 2022). One of the most valuable methods for analyzing EEG signals is to transform them from one-dimensional time-domain signal to two-dimensional time-frequency map, which can concurrently combine the frequency feature in the time-domain and frequency-domain. The STFT and WT are the typical approaches for time-frequency analysis (Tabar and Halici, 2017; Yang et al., 2022). The STFT is obtained by adding a window on the basis of the Fourier transform. It has the ability of time-frequency analysis by using a fixed window function to analyze the signal segment. However, there are some shortcomings in the determination of the window function. If the window function is too narrow, the frequency domain analysis will be inaccurate; if it is too wide, the signal features in the time domain will be imprecise, affecting the time resolution. The WT is based on the Fourier transform but replacing the infinitely long triangular function base with a finite length and decaying wavelet base, and introduces scale and translation factors so that the resolution of the window function can change with the frequency characteristics. Compared with STFT, WT has the ability to obtain the local characteristics of the signal in both the time domain and the frequency domain (Khorrami and Moavenian, 2010). CWT offers a greater time-frequency resolution and can express the 3–5 s MI-EEG signal more precisely. Therefore, the EEG signal is transformed into a two-dimensional time-frequency map using the CWT method in the current study.

Previous studies based on deep learning usually used multi-channel stacked time-frequency maps as input to recognize motor imagery EEG (Dai et al., 2019). We have also tried this method, but could not obtain higher accuracy, only about 50% accuracy was achieved when using the time-frequency maps of C3, CZ, and C4. The reason for this may be that the aim of this study is to discriminative the motor imagery EEG of unilateral upper limbs, rather than the recognition of bilateral upper limb motor imagery in most studies. The difference between different actions in the unilateral upper limb motor imagery EEG is more minor (Ofner et al., 2017; Cho et al., 2021), thus it is challenging to obtain discriminative features with fewer channels. In order to make full use of the hidden information in the unilateral limb motor imagery EEG, we selected the EEG signals of 20 channels covering the sensorimotor cortex of the brain for analysis. However, it is not suitable to directly stack the 20-channel time-frequency maps as the input of TF-CNN. To solve this problem, we proposed to convert the time-frequency map based on the virtual channel after CSP spatial filtering. CSP could extract the spatial distribution components of each class from the multi-channel EEG data (Ramoser et al., 2000), and the virtual channel signal generated after spatial filtering contained the discriminative information between classes. The results shown in **Figure 5** validated the effectiveness of this approach.

There are many successful applications for EEG signal classification using feature fusion methods of multi-modal signals. For instance, the feature fusions of facial pictures or sound signals with EEG signals have been proven to improve the classification accuracy of emotion recognition (Wagner et al., 2011; Xing et al., 2019). In the current study, the two-dimensional time-frequency maps converted by raw EEG signals were used as a supplement to the time-domain EEG signal. Since the time-frequency maps were calculated from the original EEG signals, this did not increase the complexity of the data acquisition and was suitable for rehabilitation training scenarios. In the processing of time-frequency images, TF-CNN was carried out from the perspective of image processing, which is quite different from the time-domain EEG signals processing of EEG-CNN. The information extracted from the two CNN branches were complementary, MF-CNN fused these information to make them complement each other. The results shown in **Figure 6** validated that the classification accuracy of single upper limb motor imagery EEG could be improved by such fusion strategy.

5. Conclusion

In this study, we proposed a deep learning framework named MF-CNN for classifying EEG signals associated with single upper limb motor imagery. There are two branches in MF-CNN, which can simultaneously extract features from the original EEG signal and the two-dimensional time-frequency map, and fully learn the time domain and time-frequency domain features of the EEG signal. The binary-class and three-class classification test results on the unilateral upper limb motor imagery dataset demonstrated that the proposed MF-CNN can improve the classification performance of unilateral upper limb motor imagery EEG effectively.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: http://gigadb.org/dataset/100788.

Ethics statement

This study was reviewed and approved by the Institutional Review Board at Korea University (1040548-KU-IRB-17-181-A-2). The patients/participants provided their written informed consent to participate in this study. 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

RZ and YC conceptualized the study, performed the majority of the experiments and analyses, made the figures, and wrote the first draft of the manuscript. ZX, LZ, YH, and MC performed some experiments, updated the figures, performed the statistics, and edited the manuscript. All authors approved the submitted version.

Funding

This work was supported by the MOST 2030 Brain Project (2022ZD0208500), the Technology Project of Henan Province (222102310031), and National Natural Science Foundation of China (62173310).

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

Biasiucci, A., Leeb, R., Iturrate, I., Perdikis, S., Al-Khodairy, A., Corbet, T., et al. (2018). Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke. *Nat. Commun.* 9:2421. doi: 10.1038/s41467-018-04673-z

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K. M., and Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Front. Neuroinform.* 9:16. doi: 10.3389/fninf.2015.00016

Cervera, M. A., Soekadar, S. R., Ushiba, J., Millán, J. D. R., Liu, M., Birbaumer, N., et al. (2018). Brain-computer interfaces for post-stroke motor rehabilitation: A meta-analysis. *Ann. Clin. Transl. Neurol.* 5, 651–663.

Cho, J. H., Jeong, J. H., and Lee, S. W. (2021). NeuroGrasp: Real-Time EEG classification of high-level motor imagery tasks using a dual-stage deep learning framework. *IEEE Trans. Cybern.* 52, 13279–13292. doi: 10.1109/tcyb.2021.3122969

Clevert, D.-A., Unterthiner, T., and Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (ELUs). *arXiv* [Preprint]. doi: 10.48550/arXiv.1511.07289

Dai, M., Zheng, D., Na, R., Wang, S., and Zhang, S. (2019). EEG classification of motor imagery using a novel deep learning framework. *Sensors* 19:551.

Dornhege, G., Blankertz, B., Curio, G., and Muller, K. R. (2004). Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. *IEEE Trans. Biomed. Eng.* 51, 993–1002. doi: 10.1109/tbme.2004.827088

Edelman, B. J., Baxter, B., and He, B. (2016). EEG source imaging enhances the decoding of complex right-hand motor imagery tasks. *IEEE Trans. Biomed. Eng.* 63, 4–14. doi: 10.1109/TBME.2015.2467312

Hatipoglu Yilmaz, B., and Kose, C. (2021). A novel signal to image transformation and feature level fusion for multimodal emotion recognition. *Biomed. Tech.* 66, 353–362. doi: 10.1515/bmt-2020-0229

He, Y., Eguren, D., Azorin, J. M., Grossman, R. G., Luu, T. P., and Contreras-Vidal, J. L. (2018). Brain-machine interfaces for controlling lower-limb powered robotic systems. *J. Neural Eng.* 15:021004. doi: 10.1088/1741-2552/aaa8c0

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. *arXiv* [Preprint]. doi: 10.48550/arXiv.1602.07360

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., and Gramfort, A. (2016). Autoreject: Automated artifact rejection for MEG and EEG data. *Neuroimage* 159, 417–429. doi: 10.1016/j.neuroimage.2017.06.030

Jeong, J.-H., Lee, B.-H., Lee, D.-H., Yun, Y.-D., and Lee, S.-W. (2020b). EEG classification of forearm movement imagery using a hierarchical flow convolutional neural network. *IEEE Access.* 8, 66941–66950.

Jeong, J. H., Cho, J. H., Shim, K. H., Kwon, B. H., Lee, B. H., Lee, D. Y., et al. (2020a). Multimodal signal dataset for 11 intuitive movement tasks from single upper extremity during multiple recording sessions. *Gigascience* 9:giaa098. doi: 10.1093/gigascience/giaa098

Jeong, J.-H., Shim, K.-H., Kim, D.-J., and Lee, S.-W. (2019). "Trajectory decoding of arm reaching movement imageries for brain-controlled robot arm system," in *Paper Presented at the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (Honolulu: EMBC), 2019. doi: 10.1109/EMBC.2019.8856312

Kaufmann, T., and Kubler, A. (2014). Beyond maximum speed-a novel two-stimulus paradigm for brain-computer interfaces based on event-related potentials (P300-BCI). *J. Neural Eng.* 11:056004. doi: 10.1088/1741-2560/11/5/056004

Khorrami, H., and Moavenian, M. (2010). A comparative study of DWT, CWT and DCT transformations in ECG arrhythmias classification. *Expert Syst. Appl.* 37, 5751–5757.

Kim, K.-T., Suk, H.-I., and Lee, S.-W. (2018). Commanding a brain-controlled wheelchair using steady-state somatosensory evoked potentials. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 654–665. doi: 10.1109/TNSRE.2016.2597854

Kwak, N.-S., Muller, K.-R., and Lee, S.-W. (2015). A lower limb exoskeleton control system based on steady state visual evoked potentials. *J. Neural Eng.* 12:056009. doi: 10.1088/1741-2560/12/5/056009

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

Lee, H., and Choi, Y. S. (2019). Application of continuous wavelet transform and convolutional neural network in decoding motor imagery brain-computer interface. *Entropy* 21:1199. doi: 10.3390/e21121199

Lee, H., and Kwon, H. (2016). Going deeper with contextual CNN for hyperspectral image classification. *IEEE Trans. Image Process.* 26, 4843–4855. doi: 10.1109/TIP.2017.2725580

Li, X. X., Samuel, O. W., Zhang, X., Wang, H., Fang, P., and Li, G. L. (2017). A motion-classification strategy based on sEMG-EEG signal combination for upper-limb amputees. *J. Neuroeng. Rehabil.* 14, 1–13. doi: 10.1186/s12984-016-0212-z

Loopez-Larraz, E., Birbaumer, N., and Ramos-Murguialday, A. (2018). "A hybrid EEG-EMG BMI improves the detection of movement intention in cortical stroke patients with complete hand paralysis," in *Paper Presented at the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (Honolulu: EMBC). doi: 10.1109/EMBC.2018.8512711

McFarland, D. J., Miner, L. A., Vaughan, T. M., and Wolpaw, J. R. (2000). Mu and beta rhythm topographies during motor imagery and actual movements. *Brain Topogr.* 12, 177–186. doi: 10.1023/a:1023437823106

Ofner, P., Schwarz, A., Pereira, J., and Muller-Putz, G. R. (2017). Upper limb movements can be decoded from the time-domain of low-frequency EEG. *PLoS One* 12:e0182578. doi: 10.1371/journal.pone.0182578

Penaloza, C. I., and Nishio, S. (2018). BMI control of a third arm for multitasking. *Sci. Robot.* 3:eaat1228. doi: 10.1126/scirobotics.aat1228

Pfurtscheller, G., and da Silva, F. H. L. (1999). Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/s1388-2457(99)00141-8

Pfurtscheller, G., and Neuper, C. (2001). Motor imagery and direct brain-computer communication. *Proc. IEEE* 89, 1123–1134. doi: 10.1109/5.939829

Ramoser, H., Muller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Rehabil. Eng.* 8, 441–446.

Romero-Laiseca, M. A., Delisle-Rodriguez, D., Cardoso, V., Gurve, D., Loterio, F., Posses Nascimento, J. H., et al. (2020). A low-cost lower-limb brain-machine interface triggered by pedaling motor imagery for post-stroke patients rehabilitation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 988–996. doi: 10.1109/TNSRE.2020.2974056

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

Sreeja, S. R., Rabha, J., Samanta, D., Mitra, P., and Sarma, M. (2017). "Classification of motor imagery based EEG signals using sparsity approach," in *Paper Presented at the 9th International Conference on Intelligent Human Computer Interaction (IHCI)*, (Evry: IHCI).

Tabar, Y. R., and Halici, U. (2017). A novel deep learning approach for classification of EEG motor imagery signals. *J. Neural Eng.* 14:016003. doi: 10.1088/1741-2560/14/1/016003

Taulu, S., and Larson, E. (2021). Unified expression of the quasi-static electromagnetic field: Demonstration with MEG and EEG signals. *IEEE Trans. Biomed. Eng.* 68, 992–1004. doi: 10.1109/TBME.2020.3009053

Tavakolan, M., Frehlick, Z., Yong, X., and Menon, C. (2017). Classifying three imaginary states of the same upper extremity using time-domain features. *PLoS One* 12:e0174161. doi: 10.1371/journal.pone.0174161

Ubeda, A., Azorin, J. M., Chavarriaga, R., and Millan, J. D. (2017). Classification of upper limb center-out reaching tasks by means of EEG-based continuous decoding techniques. *J. Neuroeng. Rehabil.* 14, 1–14. doi: 10.1186/s12984-017-0219-0

Wagner, J., Andre, E., Lingenfelser, F., and Kim, J. (2011). Exploring fusion methods for multimodal emotion recognition with missing data. *IEEE Trans. Affect. Comput.* 2, 206–218.

Wolpaw, R. J., Birbaumer, N., McFarland, J. D., Pfurtscheller, G., and Vaughan, T. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791.

Xing, B., Zhang, H., Zhang, K., Zhang, L., Wu, X., Shi, X., et al. (2019). Exploiting EEG signals and audiovisual feature fusion for video emotion recognition. *IEEE Access.* 7, 59844–59861.

Yang, J., Gao, S., and Shen, T. (2022). A two-branch CNN fusing temporal and frequency features for motor imagery EEG decoding. *Entropy* 24:376. doi: 10.3390/e24030376

Yao, D., Qin, Y., and Zhang, Y. (2022). From psychosomatic medicine, brain–computer interface to brain–apparatus communication. *Brain Apparatus Commun.* 1, 66–88.

Yu, Y., Liu, Y., Jiang, J., Yin, E., Zhou, Z., and Hu, D. (2018). An asynchronous control paradigm based on sequential motor imagery and its application in wheelchair navigation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 2367–2375. doi: 10.1109/TNSRE.2018.2881215

Yuan, K., Chen, C., Wang, X., Chu, W. C.-W., and Tong, R. K.-Y. (2021). BCI training effects on chronic stroke correlate with functional reorganization in motor-related regions: A concurrent EEG and fMRI study. *Brain Sci.* 11:56. doi: 10.3390/brainsci11010056

Zhang, H., Zhao, X., Wu, Z., Sun, B., and Li, T. (2021). Motor imagery recognition with automatic EEG channel selection and deep learning. *J. Neural Eng.* 18:016004. doi: 10.1088/1741-2552/abca16

Zhang, P., Wang, X., Zhang, W. H., and Chen, J. F. (2019). Learning spatial-spectral-temporal EEG features with recurrent 3D convolutional neural networks for cross-task mental workload assessment. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 31–42. doi: 10.1109/tnsre.2018.2884641

Zhang, R., Zong, Q., Dou, L., and Zhao, X. (2019). A novel hybrid deep learning scheme for four-class motor imagery classification. *J. Neural Eng.* 16:066004. doi: 10.1088/1741-2552/ab3471

Zhang, X., Shen, J., Din, Z. U., Liu, J., Wang, G., and Hu, B. (2019). Multimodal depression detection: Fusion of electroencephalography and paralinguistic behaviors using a novel strategy for classifier ensemble. *IEEE J. Biomed. Health Inform.* 23, 2265–2275. doi: 10.1109/JBHI.2019.2938247

Zhao, C.-G., Ju, F., Sun, W., Jiang, S., Xi, X., Wang, H., et al. (2022). Effects of training with a brain–computer interface-controlled robot on rehabilitation outcome in patients with subacute stroke: A randomized controlled trial. *Neurol. Ther.* 11, 679–695. doi: 10.1007/s40120-022-00333-z

Zhao-Hong, X. U., Liu, Y., Quan, J. C., and Chen, W. U. (2019). Buildings segmentation of remote sensing images based on VGG16 pre-encoding. *Sci. Technol. Eng.* 19, 250–255.

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 22 February 2023
DOI 10.3389/fnins.2023.1132290

OPEN ACCESS

EDITED BY
Minpeng Xu,
Tianjin University, China

REVIEWED BY
Dong Wen,
University of Science and Technology Beijing, China
Xiaolin Xiao,
Tianjin University, China

*CORRESPONDENCE
Penghai Li
✉ lph1973@tju.edu.cn
Longlong Cheng
✉ chenglonglong@cecdat.com

SPECIALTY SECTION
This article was submitted to Neuroprosthetics, a section of the journal Frontiers in Neuroscience

RECEIVED 27 December 2022
ACCEPTED 01 February 2023
PUBLISHED 22 February 2023

CITATION
Du P, Li P, Cheng L, Li X and Su J (2023) Single-trial P300 classification algorithm based on centralized multi-person data fusion CNN.
Front. Neurosci. 17:1132290.
doi: 10.3389/fnins.2023.1132290

Single-trial P300 classification algorithm based on centralized multi-person data fusion CNN

Pu Du[1], Penghai Li[1]*, Longlong Cheng[1,2]*, Xueqing Li[1] and Jianxian Su[1]

[1]School of Integrated Circuit Science and Engineering, Tianjin University of Technology, Tianjin, China, [2]China Electronics Cloud Brain Technology Co., Ltd., Tianjin, China

Introduction: Currently, it is still a challenge to detect single-trial P300 from electroencephalography (EEG) signals. In this paper, to address the typical problems faced by existing single-trial P300 classification, such as complex, time-consuming and low accuracy processes, a single-trial P300 classification algorithm based on multiplayer data fusion convolutional neural network (CNN) is proposed to construct a centralized collaborative brain-computer interfaces (cBCI) for fast and highly accurate classification of P300 EEG signals.

Methods: In this paper, two multi-person data fusion methods (parallel data fusion and serial data fusion) are used in the data pre-processing stage to fuse multi-person EEG information stimulated by the same task instructions, and then the fused data is fed as input to the CNN for classification. In building the CNN network for single-trial P300 classification, the Conv layer was first used to extract the features of single-trial P300, and then the Maxpooling layer was used to connect the Flatten layer for secondary feature extraction and dimensionality reduction, thereby simplifying the computation. Finally batch normalisation is used to train small batches of data in order to better generalize the network and speed up single-trial P300 signal classification.

Results: In this paper, the above new algorithms were tested on the Kaggle dataset and the Brain-Computer Interface (BCI) Competition III dataset, and by analyzing the P300 waveform features and EEG topography and the four standard evaluation metrics, namely Accuracy, Precision, Recall and F1-score,it was demonstrated that the single-trial P300 classification algorithm after two multi-person data fusion CNNs significantly outperformed other classification algorithms.

Discussion: The results show that the single-trial P300 classification algorithm after two multi-person data fusion CNNs significantly outperformed the single-person model, and that the single-trial P300 classification algorithm with two multi-person data fusion CNNs involves smaller models, fewer training parameters, higher classification accuracy and improves the overall P300-cBCI classification rate and actual performance more effectively with a small amount of sample information compared to other algorithms.

KEYWORDS

convolutional neural networks, centralized collaborative BCI, multi-person data fusion, single-trial, P300 classification

Introduction

Brain-Computer Interface is a new way of human-computer interaction, which provides a direct communication link between the brain and a computer or other external devices (McFarland and Wolpaw, 2011). The Event-Related potential (ERP) is a time-locked measure of electrical activity of the cerebral surface representing a distinct phase of cortical processing (Patel and Azzam, 2005), and it is an endogenous potential linked to a person's reaction to some stimuli or specific events. Typical examples of ERP are N200 and P300. P300 (Sutton et al., 1967), which is a positive peak waveform displayed at about 300 ms after being evoked by a small probability event, is one of the most studied, widely used and most prominent components of ERP (David et al., 2020; Kirasirova et al., 2020).

P300 classification detection is the focus of P300-BCI research, and fast and accurate recognition is crucial to improving the performance of P300-BCI (Huang et al., 2022). The P300 usually exhibits a low signal-to-noise ratio (SNR) (Zhang et al., 2022). In order to highlight its time-locked component and minimize the background noise, P300-BCI demands collecting, aggregating and averaging data from multiple trials to obtain a reliable output (Liu et al., 2018), which is time consuming and inefficient. Therefore it is a great challenge to correctly classify P300 in a single-trial. Up to now, the accuracy records of the single-trial P300 classification algorithms are as follows: Krusienski's average classification accuracy using stepwise linear discriminant analysis (SWLDA) is about 35%. Hoffmann's average classification accuracy using Bayesian Linear Discriminant Analysis (BLDA) is about 60%. Blankertz applied Shrinking Linear Discriminant Analysis (SKLDA) and achieved an average classification accuracy of about 70%. Zhang adopted spatiotemporal discriminant analysis (STDA) and attained an average classification accuracy of about 61%. The average classification accuracy of the support vector machine (SVM) algorithm developed by Kaper reaches 64.56%. And that value of discriminative canonical pattern matching (DCPM) proposed by Xiao comes to 71.23%, demonstrating that DCPM significantly outperformed other traditional methods in single-trial P300 classification with smaller training sample (Xu et al., 2018, 2021; Xiao et al., 2019a,b, 2021; Wang et al., 2020). Ma et al. (2021) proposed a capsule network-based model that improved the detection accuracy of single-trial P300, however, the calculation became complicated due to the increase in size. Zhang et al. (2022) filtered the data with xDAWN to improve the signal-to-noise ratio of EEG signals, but the spatial filtering method required manual selection of significant features after feature extraction, and then classifying them. It is highly specific to particular factors; however, the algorithm is often complex and its accuracy is influenced by feature selection (Zhang et al., 2022).

Deep learning is end-to-end learning with a simple structure that can be ported to a variety of tasks with high classification accuracy but high requirements for sample data. Nowadays, deep learning methods have made great progress in EEG-based target detection technology (Li et al., 2021), and based on this, some scholars have proposed other approaches for P300 classification, such as transfer learning (Wei et al., 2020), EEG Data Fusion (Panwar et al., 2020), Incep A-EEGNet (Xu et al., 2022), Combined Classifier (Yu et al., 2021), Principal Component Analysis (PCA) (Li et al., 2020) etc. At present, Daniela used CNN (Cecotti and Graser, 2010) with a large number of training samples to obtain an average accuracy of 78.19% for single-trial P300 (De Venuto and Mezzina, 2021) classification; For multiple trial P300 classification, Gao et al. (2021), proposed learning invariant patterns based on a CNN and big EEG data with an average accuracy of 80%. Liu et al. (2021) proposed a machine learning model based on one-dimensional convolutional capsule network (1D CapsNet), which attained a classification accuracy around 80%.

Currently, single-person BCI systems often fail to achieve the desired results because of significant individual differences and erratic execution due to the physical condition of the subjects. P300 usually has different temporal and spatial feature information, and to solve the single-trial P300 detection problem, suitable signal processing and classification algorithms are required to extract discriminative information from single-trial data (Zheng et al., 2020). Existing P300-BCI classification algorithms do not extract sufficient spatial and temporal information at the data level in feature extraction, and data must be collected from multiple trials to obtain summary and average values. With the development of complex BCI systems, the concept of multi-person cBCI has been proposed to improve overall BCI performance by fusing brain activity obtained from multiple subjects. Wang and Jung (2011) demonstrated that cBCI can improve the performance of single-trial P300 measurements by fusing brain activity from multiple subjects. Zheng et al. (2020) introduced an cross-session EEG dataset to improve the performance and utility of a collaborative RSVP-based BCI system. Song et al. (2022) proposed a Mutual Learning Domain Adaptation Network (MLDANet) cBCI framework with information interaction, dynamic learning, and individual transfer capabilities that exhibited superior population detection performance. Li P. et al. (2022) applied migration learning-based CNNs to steady-state visual evoked potentials (SSVEP). Li C. et al. (2022) proposed a fourth-order cumulative volume feature extraction method (CUM4-CSP) based on the common spatial pattern (CSP) algorithm.

In terms of BCI systems, Tian and Wang (2019) developed a multi-brain collaboration-based BCI music therapy system to help people with disabilities enjoy music and receive rehabilitation training services in the arts. Zhang et al. (2021) compared different group sizes, variations in integration strategies and their effects on group performance. Liu (2022) proposed a concrete mapping model based on human perception of sound and aesthetic transformation from sound to visual expression, forming a design representation method for interactive sound visualization practice. Currently, multi-person cBCI systems are not widely used in interactive control (Miao et al., 2020). Therefore, the research in this field can promote the development of BCI technology (Gu et al., 2021; Zhang et al., 2021).

Current research divides cBCI into two paradigms, namely distributed cBCI and centralized cBCI systems (Wang and Jung, 2011; Li P. et al., 2022). In distributed cBCI, the EEG information of the subjects is collected separately through the corresponding BCI subsystems for subsequent data pre-processing, feature extraction and pattern recognition. The results corresponding to each subject are then transmitted to the integrated classifier and the final decision is generated through a voting mechanism at the decision level, while in the centralized cBCI (Li P. et al., 2022), as shown in **Figure 1**, subjects' EEG information was collected individually for data pre-processing. The pre-processed EEG data from all subjects

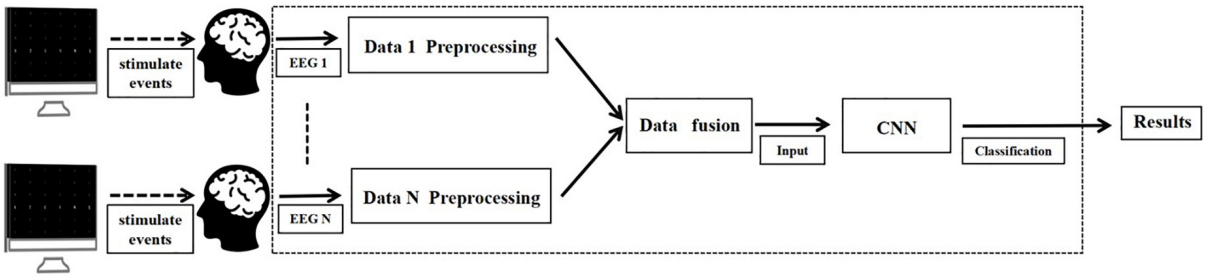


FIGURE 1
Structure of centralized collaborative brain-computer interfaces (cBCI) system.

were fused together for CNN classification identification to make the final decision for the group. The model used in this study is a centralized cBCI system, which does not rely on the voting mechanism of a distributed system, and classification is performed by a CNN-based algorithm model.

A series of experiments (Wang and Jung, 2011; Li et al., 2020, Li P. et al., 2022; Song et al., 2022) demonstrate that centralized cBCI improves overall BCI performance by fusing data from multiple subjects. To further improve the accuracy of single-trial classification, this paper combines a combination of centralized cBCI data fusion and CNN classification algorithm to identify single-trial P300. The two centralized cBCI data fusions, namely parallel data fusion and serial data fusion, can increase the effective information on the temporal and spatial domains of single-trial P300, and the CNN classification algorithm can effectively extract features on P300, hence improving the total classification accuracy and stability of P300-cBCI in the small sample case.

Materials and methods

Introduction to source datasets

Dataset I is derived from the Kaggle dataset, which includes raw data collected by electrodes, row/column numbers flickering as stimuli, and start and end time of flickers. The experimental subjects were eight healthy participants of different ages and genders, left-handed or right-handed. The experimental data acquisition process used a standard 6 × 6 Donchin and Farewell's P300 speller matrix stimulation interface with an interstimulus interval (ISI) of 0.125 ms. In the experiment, the acquisition channel selected eight lead channels Fz, Cz, P3, Pz, P4, PO7, PO8, Oz according to the international standard 10–20 system electrode location, and 35 alphanumeric characters were used for data acquisition. the stimulation went as follows: each row and column flickered once in a random order in one round of stimulation, so each stimulus includes 12 flickering rows/columns, and a subject was required to choose the correct row number and column number corresponding to a designated character, so as to produce 2 P300 signals. The stimulation repeated 10 times for each character, so the experimenter collected 4,200 (12*10*35) samples in total, among which 700 (2*10*35) were target stimuli. All subjects performed the same P300 stimulation evoked experiments. The stimulation interface and numbered row/column of the dataset are shown in Figure 2.

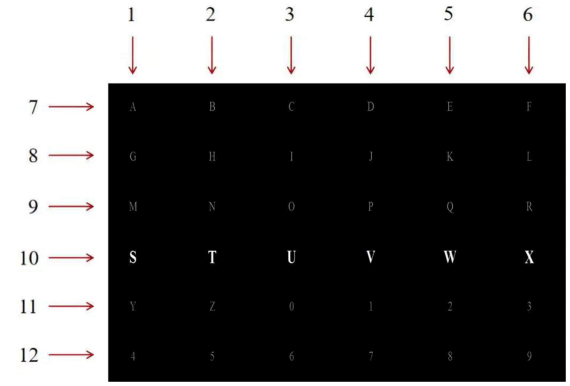


FIGURE 2
P300 speller matrix and corresponding row/column labels.

Dataset II is derived from BCI Competition III, including 50-min EEG recordings and speller matrix information of two subjects (subject A and B). One round of flickering of all the rows and columns is referred to as one trial, so each trial includes 12 row or column flickerings. Within each trial, the row or column flickers for 100 ms, with 75 ms interval between two flickering stimuli. The experiment repeats 15 times, producing 180 (12*15) row/column flickerings. The stimuli interface adopts the P300 speller matrix illustrated by Figure 2 and the corresponding row/column labels.

Data preprocessing and fusion

The P300 EEG signal has a very low signal-to-noise ratio and mainly lies within a specific frequency range of 0.5–7.5 Hz. Collected EEG signals often include fundamental noises in various frequencies, such as industrial frequency noise, or random noise. To remove the impact of these invalid noises and improve the signal-to-noise ratio of the P300 EEG signal, an individual trial's data extracted from a dataset are usually filtered and preprocessed with a 50 Hz trap filter and a (0.1–30 Hz) Butterworth bandpass filter. Besides the main 300 ms peak after stimulation, other peaks around it are also important, so the EEG signal in Dataset I is divided

into 1 s windows using 352 timestamps to better capture key information.

Downsampling is applied on data to reduce the data transmission rate and data size. Each element value is $X_{i,j}$, where $0 \leq i \leq N_{elec}$, $0 \leq j \leq N_t$. N_{elec} denotes the number of lead channels, N_t denotes the sampling frequency, and the sampling frequency used in the experiment is 240 Hz. The downsampling begins with data dimensionality reduction, specifically the data time domain sampling frequency is reduced from 240 Hz to 120 Hz. Then the data are normalized to prevent overfitting and avoid different data performing nearly identically in the same neural network. The calculation method is shown in formula (1).

$$X_{i,j} = \frac{X_{i,j} - \overline{X}}{\sigma_i} \quad (1)$$

In Formula (1), $\overline{X}$ represents the mean value of EEG signal recorded by electrode i and σ_i represents the standard deviation recorded by electrode i (Cecotti and Graser, 2010).

Two brain data fusion methods are proposed in this paper to merge the preprocessed data information in spatial and temporal domains. Specifically, both parallel data fusion and serial data fusion are performed on the data evoked by repeated identical experimental stimuli. As shown in Figure 3, n subjects labeled Single 1 to Single n were fused in two ways, and (n-2) sets of data were omitted from a total of n groups of data. Parallel data fusion increases the spatial domain feature information by fusing multi-person data stimulated by the same task, thus improving the overall performance of BCI. Serial data fusion can achieve the same goal by fusing multi-person data stimulated by the same task and adding feature information in the time domain without changing the number of leads.

Characteristic analysis

Two individual subjects' data was randomly selected from data set I, which was evoked by the same stimulus experiment. Then starting with the small probability stimulus moment, the wave form during 0–500 ms after the filtered small probability stimulus evoked response was drawn, and the single-trial P300 amplitude features and EEG topographic map in single-person mode and two-person centralized data fusion mode were analyzed and compared. As shown in Figures 4, 5. In Figure 4, different colored curves in each graph correspond to different lead signals. The position of the leads is shown in the upper left corner of the diagram, the upper right corner is the amplitude color scale measured in μv, the horizontal axis represents the time and the vertical axis represents the signal amplitude of each lead. Figures 4A,B represent the EEG signals of each lead for both subjects in single-person mode. As can be seen in Figure 4, the P300 EEG signal treated with the two centralized data fusion has a more pronounced wave at around 300 ms. In this case, Figure 4C shows the centralized parallel data fusion, as the international standard 10–20 lead system was used, so by assigning the eight leads Fz, Cz, P3, Pz, P4, PO7, PO8, Oz to the eight leads FCz, CPz, CP1, CP2, P5, P6, PO3, PO4, it was possible to draw 16 lead waveforms. The increase in lead (spatial domain) information by centralized parallel data fusion is evident in Figure 4C. Figure 4D shows the centralized serial data fusion. As the centralized serial data fusion is the information added in the time domain, in terms of the lead wave crest characteristics, it is first shown as the first one of the two fusion individuals.

Figure 5 illustrates the change in amplitude corresponding to each lead position in the EEG topography in single-person mode and two-person centralized data fusion mode, with Figures 5A,B representing single-person mode, Figure 5C representing two-person centralized parallel data fusion and Figure 5D representing

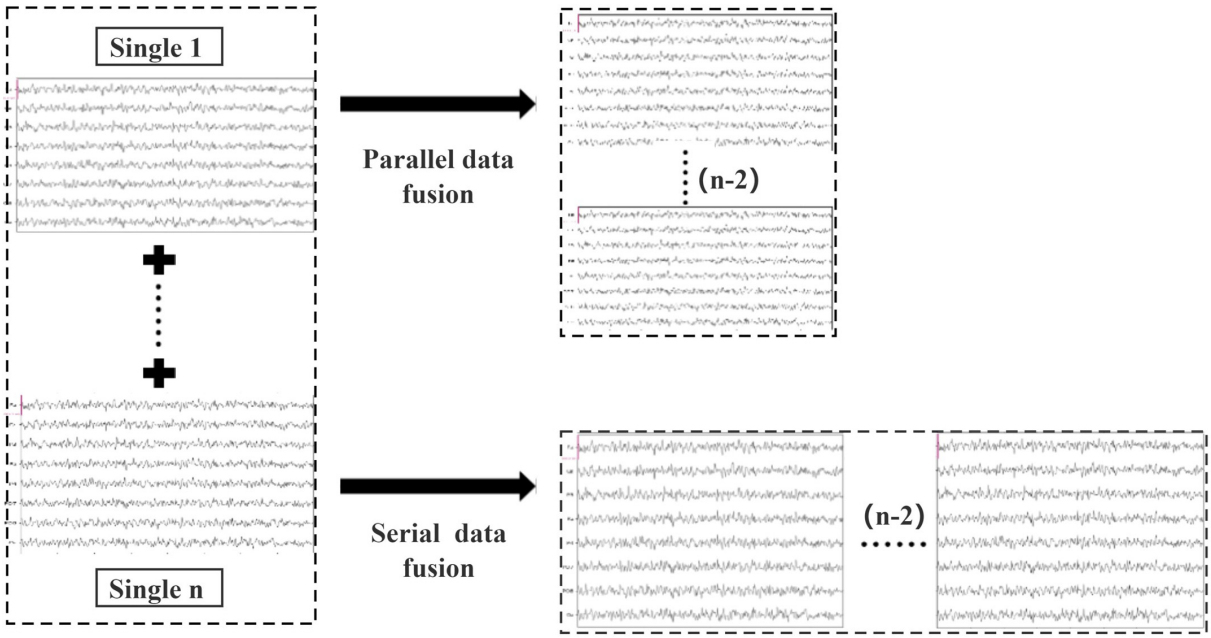


FIGURE 3
Schematic diagram of parallel data fusion and serial data fusion of multi-person data.

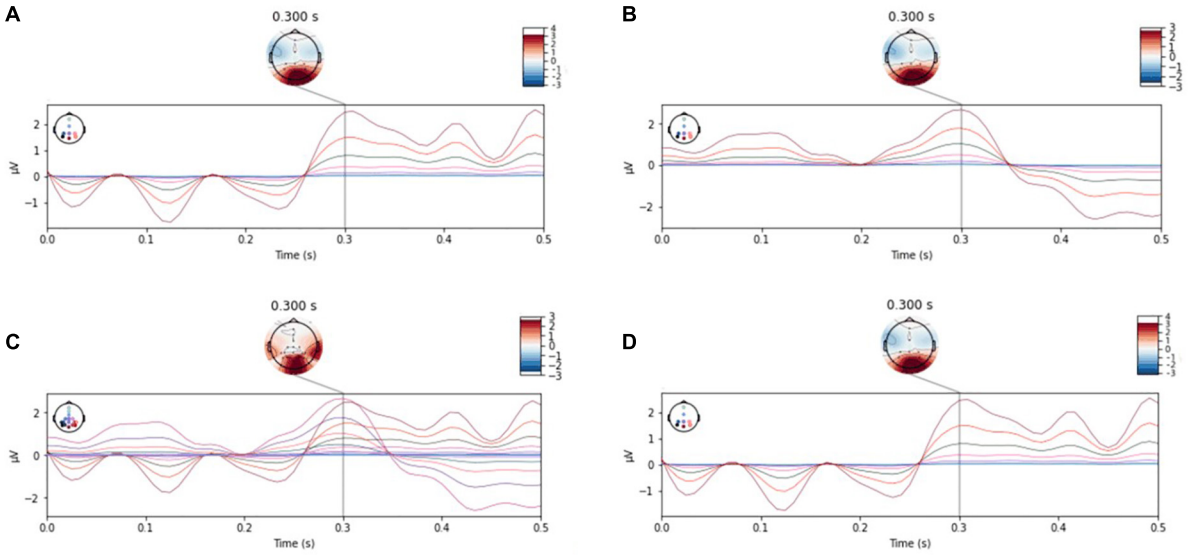


FIGURE 4
P300 characteristic distribution. **(A,B)** Single-person model. **(C)** Centralized parallel data fusion. **(D)** Centralized serial data fusion.

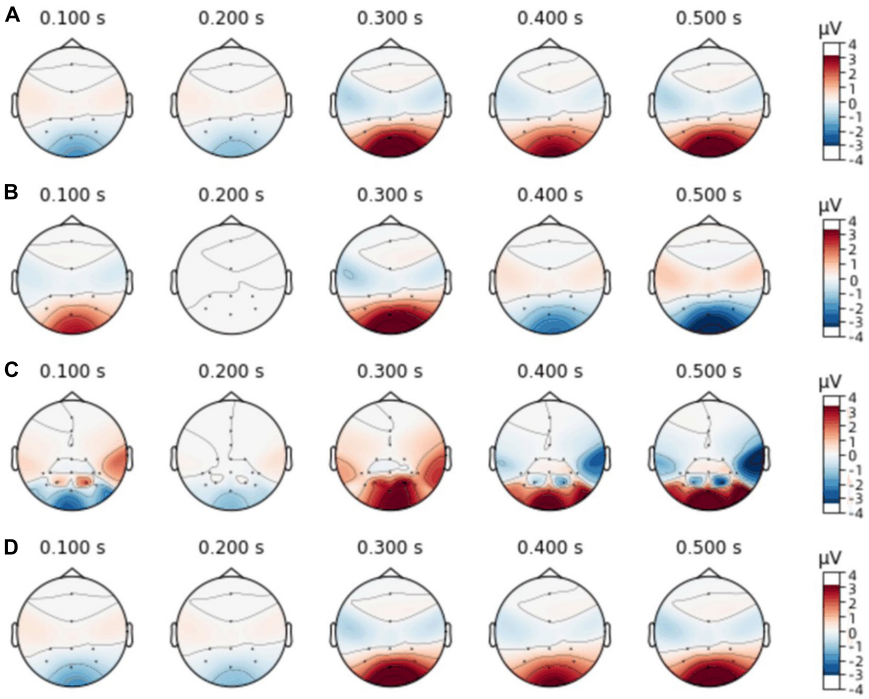


FIGURE 5
Electroencephalographic topography. **(A,B)** Single-person model. **(C)** Centralized parallel data fusion. **(D)** Centralized serial data fusion.

two-person centralized serial data fusion. It can be seen from Figures 4, 5 that this method is feasible.

CNN classification

In this paper, Dataset I was first used, referring to the CNN structure proposed by Cecotti and Graser (2010), and the network structure parameters were adjusted based on the data characteristics of Dataset I. Taking two-person parallel data fusion as an example, the 8-Lead data set is fused into 16 leads, which increases the characteristics of lead information and spatial domain. The CNN structure is composed of Input layer, Convolution layer, Dropout layer, Maxpooling layer, Flatten layer, and Dense layer. In the CNN structure, the first and third layers are the convolutional layers, and the convolutional operation can

TABLE 1 Convolutional neural network (CNN) structure of two-person parallel data fusion.

Layer	Input	Type	Output	#Parameters
L1	(None,351,16,1)	Conv2D+ReLU	(None,351,16,16)	1,040
L2	(None,351,16,16)	Dropout	(None,351,16,16)	0
L3	(None,351,16,16)	Conv2D+ReLU	(None,351,16,32)	18,464
L4	(None,351,16,32)	MaxPooling2D	(None,175,8,32)	0
L5	(None,175,8,32)	Flatten	(None,44800)	0
L6	(None,44800)	Dense+ReLU	(None,64)	2,867,264
L7	(None,64)	Dense+ReLU	(None,8)	520
L8	(None,8)	Dense+softmax	(None,2)	18

TABLE 2 Convolutional neural network (CNN) structure of two-person serial data fusion.

Layer	Input	Type	Output	#Parameters
L1	(None,702,8,1)	Conv2D+ReLU	(None,702,8,16)	1,040
L2	(None,702,8,16)	Dropout	(None,702,8,16)	0
L3	(None,702,8,16)	Conv2D+ReLU	(None,702,8,32)	18,464
L4	(None,702,8,32)	MaxPooling2D	(None,351,4,32)	0
L5	(None,351,4,32)	Flatten	(None,44928)	0
L6	(None,44928)	Dense+ReLU	(None,64)	2,875,456
L7	(None,64)	Dense+ReLU	(None,8)	520
L8	(None,8)	Dense+softmax	(None,2)	18

be regarded as the inner product of the input samples and the convolutional kernel, as shown in the formula (2).

$$Y_j^l = f(\sum_{i \in M_j} Y_i^{l-1} * \omega_{ij}^l + B_j^l) \quad (2)$$

In Formula (2), Y_j^l is the j th characteristic map of the l th convolution layer, $f()$ represents the activation function, ReLU activation function is used in this network, M_j represents all input characteristic maps, ω_{ij}^l represents the convolution kernel matrix between i and j, B_j^l represents offset, and * represents convolution operation.

The Dropout layer is used after the first convolution layer to prevent a decrease in sensitivity of the network model due to overfitting. The Maxpooling layer is added after the second Conv layer, which compresses the features obtained from the preceding layer with a pooling function, and selects the maximum value of all elements in each specific region of the feature map as the feature value of that region. This procedure resembles a secondary feature extraction process, retaining the main features of the data while of the data while lowering the dimensionality the dimensionality of the data, thus reducing the computational effort (He et al., 2020). So Maxpooling can effectively reduce the training parameters and over-fitting problems to form the final features. The Flatten layer is then connected with the Maxpooling layer to map the feature space calculated by the previous layer (convolution, pooling, etc.) to the sample marker space to produce the final classification result, and improve the generalization ability of the model. The specific parameters are shown in Table 1.

Also taking two-person serial data fusion as an example, the preprocessed single-person data is fused without changing the specific data of two person. Serial data fusion is mainly carried out in the time domain. That is, the time domain information can be greatly expanded without changing the number of leads. When constructing the CNN structure of two-person serial data fusion, in order to avoid errors caused by other reasons, only the corresponding time domain parameters are changed. The CNN structure and specific parameters of two-person serial data fusion are shown in Table 2.

The ReLU function is used as the activation function of the neurons in the CNN. This method can solve the gradient vanishing problem with fast calculation speed and fast convergence speed. As shown in formula (3), when the input x takes a negative value, the output is 0, and when it takes a positive value, the output remains that value of x.

$$ReLU(x) = max(x, 0) \quad (3)$$

The last layer of neurons uses the softmax function for binary classification. The function is given in formula (4) as follows, where x_i is the input.

$$Softmax(x) = \frac{e^{x_i}}{\sum_i e^{x_i}} \quad (4)$$

In this paper, the CNN adopts the most robust network optimizer for the neural network. Adam, and the cross-entropy function as the loss function. The learning rate is set at 0.001, the number of trainings is set as 75, and the random mini-batch size gradient descent is set to 32, which can enable the network to be well generalized and achieve faster classification.

Results

In order to evaluate the performance of the P300 classification algorithm, relevant evaluation criteria are considered. The standard metric for evaluating the P300 classification algorithm usually is the accuracy rate, and the formula for P300 recognition accuracy rate is given in Equation (5), which includes True Positive (*TP*), True Negative (*TN*), False Positive (*FP*), and False Negative (*FN*). TP indicates the number of samples correctly identified as positive in positive samples, *TN* indicates the number of samples correctly identified as negative in negative samples, *FP* indicates the number of samples misidentified as positive in negative samples, and *FN* indicates the number of samples misidentified as negative in positive samples (Cecotti and Graser, 2010; De Venuto and Mezzina, 2021; Liu et al., 2021).

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

The eight subjects contained in Dataset I were labeled in turn as Subjects 1–8, and their data was divided into four sets marked as C1, C2, C3, and C4, respectively, each including the data of two subjects. Then the four sets of data were used for parallel data fusion or serial data fusion. Table 3 lists the results of CNN's single-trial P300 classification of two centralized multi-person data fusion methods and single-person mode, respectively.

As shown in Table 3, the classification accuracy of single-trial P300 based on the fusion of two centralized multi-person data is

TABLE 3 Results of convolutional neural network (CNN) single-trial P300 classification for centralized multi-person data fusion and single-person mode.

CNN	Subject	1	2	3	4	5	6	7	8	Average
	Accuracy (%)	60.72	75.00	71.43	71.43	75.00	71.43	75.00	78.57	72.32
CNN+Parallel data fusion	Subject	C1		C2		C3		C4		Average
	Accuracy (%)	78.57		82.14		85.71		85.67		83.03
CNN+Serial data fusion	Subject	C1		C2		C3		C4		Average
	Accuracy (%)	71.43		78.57		78.57		85.71		78.57

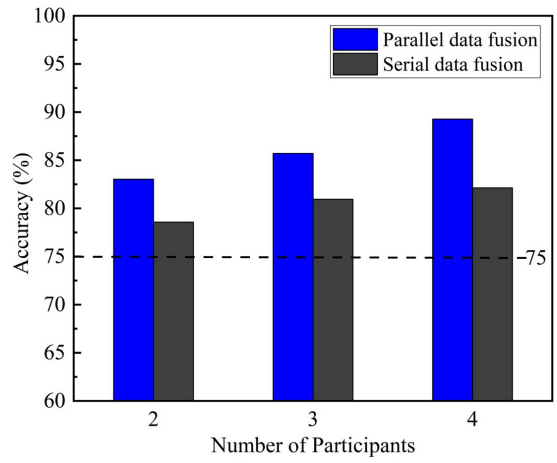


FIGURE 6
Single-trial P300 classification results of a centralized multi-person data fusion convolutional neural network (CNN) for groups with different numbers of participants.

higher than that of single-trial P300 based on the single-person mode. Specifically, the average accuracy of CNN for single-trial P300 of single-person is 71.88%, while the average classification accuracy of CNN with parallel data fusion reaches 83.03%, and that value of CNN with serial data fusion reaches 78.57%.

Figure 6 compares the single-trial P300 classification results of the two data fusion methods for two-person, three-person and four-person groups and the counterpart results of the single-person mode CNN. The dotted line 75% is the highest classification accuracy of the single-trial P300 given by the single-person mode CNN. It can be seen from Figure 6, as the number of participants in the experiment increases, the average classification accuracy of the centralized multi-person data fusion method for single-trial P300 keeps improving, and both of them exceed the dotted line 75%. When the number of participants was four-person, the average classification accuracy reached 89.13% for parallel data fusion and 82.14% for serial data fusion.

In addition to accuracy, some mainstream performance metrics for binary classification problems, such as *Recall*, *Precision*, and their summed average *F1- score,* are also considered relevant for further feature recognition (Cecotti and Graser, 2010; De Venuto and Mezzina, 2021; Liu et al., 2021). The calculation formula is shown in (6), (7), and (8):

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

TABLE 4 Single-trial P300 classification results of a centralized multi-person data fusion convolutional neural network (CNN) for different numbers of participants.

Method	N-participants	Recall (%)	Precision (%)	F1- score (%)
CNN	1	**66.7**	**72.7**	**69.6**
CNN+Parallel data fusion	2	81.2	72.2	76.5
		81.2	78.5	86.7
		68.7	70.5	81.4
		75.1	75.2	85.7
	Average	**76.5**	**74.1**	**82.5**
	3	75.3	85.7	83.3
		62.5	71.4	66.7
		66.7	75.6	75.0
	Average	**68.1**	**77.5**	**75.0**
	4	51.2	66.7	70.6
		51.4	66.9	70.8
	Average	**51.3**	**66.8**	**70.7**
CNN+Serial data fusion	2	75.1	71.4	76.9
		74.5	71.2	75.6
		68.7	66.7	75.8
		62.5	66.7	76.9
	Average	**70.2**	**69.0**	**76.3**
	3	75.2	85.7	76.9
		62.5	83.3	71.4
		66.7	75.6	75.0
	Average	**68.1**	**81.5**	**74.4**
	4	51.2	66.7	70.6
		51.4	66.9	70.8
	Average	**51.3**	**66.8**	**70.7**

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$F1 - score = 2\frac{Recall * Precision}{Recall + Precision} \tag{8}$$

Precision is the proportion of genuinely positive samples in all (P300) samples that are identified as positive, and *Recall* is the

TABLE 5 Results of centralized multi-person electroencephalography (EEG) data fusion convolutional neural network (CNN) and other classification algorithms.

Method	Subject	Accuracy (%)	Recall (%)	Precision (%)	F1- score (%)
ConvLSTM (Joshi et al., 2018)	A	75.1	64.3	36.1	46.2
	B	80.2	63.4	44.0	54.0
BN3 (Liu et al., 2018)	A	73.0	63.7	33.6	44.0
	B	79.7	67.0	42.9	52.3
CNN-1 (Cecotti and Graser, 2010)	A	70.4	67.4	31.7	43.1
	B	78.2	67.8	40.7	50.9
BN3(ns)+ConvLSTM (De Venuto and Mezzina, 2021)	A	75.7	63.4	36.8	46.7
	B	82.3	65.2	47.8	**55.2**
Autoencoded CNN (De Venuto and Mezzina, 2021)	A	75.1	67.2	36.6	47.4
	B	**82.7**	66.5	**48.6**	56.2
Autoencoded CNN no LBP (De Venuto and Mezzina, 2021)	A	75.2	65.0	36.3	46.6
	B	81.7	66.3	46.6	54.7
CNN+Parallel data fusion	$P_{(A+B)}$	**82.4**	**71.1**	**56.6**	**66.7**
CNN+Serial data fusion	$S_{(A+B)}$	81.4	**75.1**	44.4	53.3

proportion of positive samples that are detected from genuinely positive samples, where *Precision* and *Recall* influence each other, with both metrics being high if the detection algorithm is ideal. However, usually it is difficult to optimize both of them, when one is high the other will be low, so *F1- score* can be chosen as their combined metric. In Table 4, the bold values represent the average values of the three indicators in the single-person mode and the average values of the three indicators in the two centralized multi-person data fusion CNN for different participants. Since there are eight individual data in the data set, data fusion was carried out for four two-person groups, three three-person groups, and two four-person groups. In three-person group case, only two people were left for the last group, so one person was randomly selected from the other two groups that were already fused so as to make up three members. Then the single-trial P300 classification evaluation indicators Precision, Recall and F1- score for the two data fusion methods with two-person, three-person, and four-person groups were calculated, and the average value after centralized data fusion is taken in each case. The results are shown in Table 4. In Table 4, N-participants represent the number of participants in a group for centralized data fusion. In Table 4, N-participants represent the number of people who have undergone centralized data fusion. Since there are eight single persons in the data set, they are divided into four groups when the number of people fused is two, three groups when there are three, and two groups when there are four, and the average value after centralized data fusion is taken.

N-participants in Table 4 is 1, which represents the mean classification of single-trial P300 by the CNN in single-person mode. Although it can be seen from Figure 6 that the average classification accuracy of the two centralized data fusions increases as the number of participants increases, the three metrics Precision, Recall and F1-score all decrease to varying degrees as the number of participants increases. The reason behind this fact is that P300 and non-P300 data in the EEG data is unevenly distributed, even if all the recognition is made for non-P300 signals, the model can still achieve high accuracy, so the accuracy alone is not enough to achieve a scientific and persuasive evaluation, and all the four indicators should be considered comprehensively.

It can be seen from Table 4 that when the number of group member goes from 2 to 3 and 4, the recall of both centralized data fusions The highest recall rates are achieved in two-person group case with an average of 76.5% for parallel data fusion and 70.2% for serial data fusion. The F1- score averages for both centralized data fusion CNNs also reach the highest value in two-person group case, with parallel data fusion averaging 82.5% and serial data fusion averaging 76.3%. In two-person or three-person group cases, all the three metrics improved compared with those for the single-trial P300 classification in single-person mode. However, in four-person group case, the recall and precision of both centralized data fusions are slightly lower than the mean in the single-person mode, and the mean of F1- score is higher than in the single-person mode. In summary, the centralized multi-person data fusion classification algorithm has obvious advantages over the single-person mode classification algorithm. When the data of individual participants in the centralized data fusion is divided into four two-person groups, the F1–score reaches the highest when compared with the single-person mode and the number of group members is three and four, Combining the two indicators of Accuracy and F1-score, when the group members of individual participants in centralized data fusion are two, the classification single-trial P300 has the best effect. To explain the better experimental results using a data fusion group size of two compared with three and four, one possible reason could be the over-fitting of multi-dimensional data; another reason could be that the noisy nature of the EEG signal leads to saturation of the classification performance, resulting in reduced accuracy. EEG artifacts include electrode contact loosening, head movements, eye movements and muscle activity. It is known that noise levels may affect linear classification performance (Yun and Stoica, 2016).

Validation of the model on dataset I indicates that the best number of group members is two for the two centralized data fusion CNNs in single-trial P300 classification. In order to test the reproducibility of this method, this paper then applies the algorithm to the data of Dataset II. Each subject's single-trial P300 information of 15 repeated experiments was extracted, and two subjects' single-trial data is fused with the above-mentioned method to calculate the average classification accuracy after fusion. For consistency, the results of other advanced single-trial P300 classification algorithms analyzing the same dataset and using the same CNN structure were compared in terms of accuracy, recall, precision and F1-score. As shown in Table 5. In Table 5, bold values represent the highest two values of each column.

The comparison results in Table 5 show that the average accuracy of the CNN based on centralized parallel data fusion and serial data fusion in two-person group case reached 82.4 and 81.4%, respectively, both of which are slightly higher than that of other advanced single-trial P300 recognition algorithms. And the parallel data fusion always maintained the highest classification accuracy. In terms of recall, both algorithms based on the two centralized multi-person data fusion CNNs maintained a high level, with serial data fusion reaching around 75%. Compared with other methods, the proposed algorithms reached the higher level of accuracy, although the data was imbalanced. The two centralized brain data fusion CNNs also surpass the other algorithms in terms of F1- score, with the F1- score for parallel data fusion also maintaining the higher level. In summary, the method was shown to be reproducible. Compared with other classification algorithms, the spatial and temporal domain feature information of the single-trial P300 data layer can be increased after fusion of multi-person data, and the CNN constructed by connecting the Flatten layer with the Conv layer and Maxpooling layer can better extract and classify the features of the single-trial P300, which solves the problem of complex and time consuming operation as well as low accuracy in the process of recognizing the single-trial P300, thus achieving better recognition results.

Discussion

The single-trial P300 classification algorithm based on centralized multi-person data fusion CNN proposed in this paper uses CNN to classify the single-trial P300 signal after centralized parallel or serial fusion of multi-person EEG data. In the CNN network structure, a Dropout layer is added after the first Conv layer to prevent overfitting, and a Maxpooling layer is used after the second Conv layer to connect the Flatten layer, extracting the maximum of all elements in each region of the convolutional layer feature map as the feature value of this region, preserving the main features of the data while reducing the dimensionality of the data. Batch Normalization is adopted to train the data in small Batch, which makes it easier to generalize the network and classify P300 signals faster. The purpose is to improve the existing multi-trial P300 classification algorithm, which is time-consuming and complex in calculation, and the single-trial P300 classification algorithm which has low accuracy. This paper uses two centralized multi-person data fusion CNN approaches to fuse the EEG data of different number of participants ranging from 2 to 4 for P300 classification. The results are evaluated with four metrics, Accuracy, Recall, Precision and F1- score, respectively, and compared with those of single-person CNN model and other advanced single-trial P300 classification algorithms, which are validated on the available public dataset Kaggle dataset and BCI Competition III. The experimental results demonstrate that the classification results of both centralized multi-person data fusion CNNs outperform the CNN classification results in single-person mode, and the four metrics of Accuracy, Recall, Precision and F1-score for detecting single-trial P300 are improved by different margins compared with other classification algorithms, so the proposed approach can achieve high accuracy in identifying single-trial P300. Comparison among the results of fusing 2, 3, and 4 people's data as a group indicates that the best results are obtained for two-person groups.

Among the two data fusion methods used in this paper, the centralized P300-cBCI with parallel data fusion is the better choice in terms of applicability compared to the centralized P300-cBCI with serial data fusion, as it involves a smaller model and fewer training parameters. In summary, CNNs that undergo centralized two-person parallel data fusion can be more effective in improving the overall P300-cBCI classification accuracy and practical performance at small amounts of sample information. The single-trial P300 classification algorithm based on a centralized multi-person data fusion CNN proposed in this paper can be applied to online P300-cBCI systems, providing a new idea for building a more efficient P300-cBCI system, but this requires participating subjects to receive the same experimental stimuli under the same experimental conditions, and the same pre-processing of the data to be prepared for fusion. In the future, online P300-cBCI systems are to be built to enable efficient, fast and accurate classification of P300 for a number of applications, such as helping patients with text communication. This will improve the actual performance of the P300-cBCI system.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: http://www.bbci.de/competition/iii/ and https://www.kaggle.com/rramele/p300samplingdataset.

Ethics statement

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

PD: data curation, research design, data analysis, and manuscript writing. PL: conceptualization, supervision, and writing—review and editing. LC: methodology and validation. PL and PD: resources writing—original draft preparation. XL: production of figures and document retrieval. JS: data analysis and classification algorithm. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Key Research and Development Program of China (Grant No. 2021YFF1200600), the National Natural Science Foundation of China (No. 61806146), and the Natural Science Foundation of Tianjin City (Nos. 18JCYBJC95400 and 19JCTPJC56000).

Conflict of interest

LC was employed by the China Electronics Cloud Brain (Tianjin) Technology Co., Ltd.

The remaining 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

Cecotti, H., and Graser, A. (2010). Convolutional neural networks for P300 detection with application to brain-computer interfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 433–445. doi: 10.1109/TPAMI.2010.125

David, C., Omar, M., and Antelis, M. (2020). Single-option P300-BCI performance is affected by visual stimulation conditions. *Sensors* 20:7198. doi: 10.3390/S20247198

De Venuto, D., and Mezzina, G. (2021). A single-trial P300 detector based on symbolized EEG and autoencoded-(1D) CNN to improve ITR performance in BCIs. *Sensors* 21:3961. doi: 10.3390/S21123961

Gao, W., Yu, T., Yu, J., Gu, Z., Li, K., Huang, Y., et al. (2021). Learning invariant patterns based on a convolutional neural network and big electroencephalography data for subject-independent P300 brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 1047–1057. doi: 10.1109/TNSRE.2021.3083548

Gu, B., Xu, M., Xu, L., Chen, L., Ke, Y., Wang, K., et al. (2021). Optimization of task allocation for collaborative brain–computer interface based on motor imagery. *Front. Neurosci.* 15:683784. doi: 10.3389/FNINS.2021.683784

He, Q., Shao, D., Wang, Y., Zhang, Y., and Xie, P. (2020). Analysis of motor imagery EEG signals and intention recognition based on multi-feature convolutional neural nets. *J. Instrum.* 41, 138–146. doi: 10.19650/j.cnki.cjsi.J1905522

Huang, Z., Guo, J., Zheng, W., Wu, Y., Lin, Z., and Zheng, H. (2022). A calibration-free approach to implementing P300-based brain–computer interface. *Cogn. Comput.* 14, 887–899. doi: 10.1007/S12559-021-09971-1

Joshi, R., Goel, P., Sur, M., and Murthy, H. (2018). "Single trial P300 classification using convolutional LSTM and deep learning ensembles method," in *Intelligent human computer interaction. IHCI 2018*, ed. U. Tiwary (Cham: Springer), doi: 10.1007/978-3-030-04021-5_1

Kirasirova, L., Bulanov, V., Ossadtchi, A., Kolsanov, A., Pyatin, V., and Lebedev, M. (2020). A P300 brain-computer interface with a reduced visual field. *Front. Neurosci.* 14:604629. doi: 10.3389/FNINS.2020.604629

Li, C., Li, N., Qiu, Y., Peng, Y., Wang, Y., Deng, L., et al. (2022). Multimodal collaborative BCI system based on the improved CSP feature extraction algorithm. *Virtual Real. Intell. Hardw.* 4, 22–37. doi: 10.1016/J.VRIH.2022.01.002

Li, F., Li, X., Wang, F., Zhang, D., Xia, Y., and He, F. (2020). A novel P300 classification algorithm based on a principal component analysis-convolutional neural network. *Appl. Sci.* 10:1546. doi: 10.3390/app10041546

Li, P., Su, J., Belkacem, A., Cheng, L., and Chen, C. (2022). Steady-state visually evoked potential collaborative BCI system deep learning classification algorithm based on multi-person feature fusion transfer learning-based convolutional neural network. *Front. Neurosci.* 16:971039. doi: 10.3389/FNINS.2022.971039

Li, Z., Lan, Z., Tang, D., Yan, C., Xiang, X., and Zhou, H. (2021). A review of target detection based on ERP signal. *Comput. Eng. Appl.* 23, 37–49.

Liu, M. (2022). An EEG neurofeedback interactive model for emotional classification of electronic music compositions considering multi-brain synergistic brain-computer interfaces. *Front. Psychol.* 12:799132. doi: 10.3389/FPSYG.2021.799132

Liu, M., Wu, W., Gu, Z., Yu, Z., Qi, F., and Li, Y. (2018). Deep learning based on batch normalization for P300 signal detection. *Neurocomputing* 275, 288–297. doi: 10.1016/j.neucom.2017.08.039

Liu, X., Xie, Q., Lv, J., Huang, H., and Wang, W. (2021). P300 event-related potential detection using one-dimensional convolutional capsule networks. *Expert Syst. Appl.* 174:114701. doi: 10.1016/J.ESWA.2021.114701

Ma, R., Yu, T., Zhong, X., Yu, Z., Li, Y., and Gu, Z. (2021). Capsule network for ERP detection in brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 718–730. doi: 10.1109/TNSRE.2021.3070327

McFarland, D., and Wolpaw, J. (2011). Brain-computer interfaces for communication and control. *Commun. ACM* 54, 60–66. doi: 10.1145/1941487.1941506

Miao, Y., Yin, E., Allison, B., Zhang, Y., Chen, Y., Dong, Y., et al. (2020). An ERP-based BCI with peripheral stimuli: Validation with ALS patients. *Cogn. Neurodyn.* 14, 21–33. doi: 10.1007/s11571-019-09541-0

Panwar, S., Rad, P., Jung, T., and Huang, Y. (2020). Modeling EEG data distribution with a wasserstein generative adversarial network to predict RSVP events. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 1720–1730. doi: 10.1109/TNSRE.2020.3006180

Patel, S., and Azzam, P. (2005). Characterization of N200 and P300: Selected studies of the event-related potential. *Int. J. Med. Sci.* 2, 147–154.

Song, X., Zeng, Y., Tong Li, Shu, J., Yang, Q., Kou, J., et al. (2022). A collaborative brain-computer interface framework for enhancing group detection performance of dynamic visual targets. *Comput. Intell. Neurosci.* 2022:4752450. doi: 10.1155/2022/4752450

Sutton, S., Tueting, P., Zubin, J., and John, E. (1967). Information delivery and the sensory evoked potential. *Science* 155, 1436–1439. doi: 10.1126/science.155.3768.1436

Tian, K., and Wang, Z. (2019). The application of brain computer interface with multi brain coordination in music therapy. *Technol. Mark.* 26, 80–81. doi: 10.3969/j.issn.1006-8554.2019.08.030

Wang, K., Xu, M., Wang, Y., Zhang, S., Chen, L., and Ming, D. (2020). Enhance decoding of pre-movement EEG patterns for brain–computer interfaces. *J. Neural Eng.* 17:016033. doi: 10.1088/1741-2552/ab598f

Wang, Y., and Jung, T. (2011). A collaborative brain-computer interface for improving human performance. *PLoS One* 6:e20422. doi: 10.1371/journal.pone.0020422

Wei, W., Qiu, S., Ma, X., Li, D., Wang, B., and He, H. (2020). Reducing calibration efforts in RSVP tasks with multi-source adversarial domain adaptation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 2344–2355. doi: 10.1109/TNSRE.2020.3023761

Xiao, X., Xu, M., Han, J., Yin, E., Liu, S., Zhang, X., et al. (2021). Enhancement for P300-speller classification using multi-window discriminative canonical pattern matching. *J. Neural Eng.* 18:046079. doi: 10.1088/1741-2552/AC028B

Xiao, X., Xu, M., Jin, J., Wang, Y., Jung, T., and Ming, D. (2019a). Discriminative canonical pattern matching for single-trial classification of ERP components. *IEEE Trans. Biomed. Eng.* 67, 2266–2275. doi: 10.1109/TBME.2019.2958641

Xiao, X., Xu, M., Wang, Y., Jung, T., and Ming, D. (2019b). "A comparison of classification methods for recognizing single-trial P300 in brain-computer interfaces," in *2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, (Berlin: IEEE), 3032–3035. doi: 10.1109/EMBC.2019.8857521

Xu, M., He, F., Jung, T., Gu, X., and Ming, D. (2021). Current challenges for the practical application of electroencephalography-based brain–computer interfaces. *Engineering* 7, 1710–1712. doi: 10.1016/j.eng.2021.09.011

Xu, M., Wang, D., Li, Z., and Chen, Y. (2022). IncepA-EEGNet: P300 signal detection method based on fusion of Inception network and attention mechanism. *J. ZheJiang Univ.* 56, 745–753.

Xu, M., Xiao, X., Wang, Y., Qi, H., Jung, T., and Ming, D. (2018). A brain–computer interface based on miniature-event-related potentials induced by very small lateral

visual stimuli. *IEEE Trans. Biomed. Eng.* 65, 1166–1175. doi: 10.1109/TBME.2018.2799661

Yu, H., Xie, J., He, L., Yang, Y., Zhang, H., and Xu, G. (2021). Event related potential recognition method based on convolutional neural network and support vector machine. *J. Xi'an Jiaotong Univ.* 55, 47–54.

Yun, K., and Stoica, A. (2016). "Improved target recognition response using collaborative brain-computer interfaces," in *2016 IEEE international conference on systems, man, and cybernetics (SMC)* (Budapest: IEEE), 002220–002223. doi: 10.1109/SMC.2016.7844568

Zhang, H., Wang, Z., Yu, Y., Yin, H., Chen, C., and Wang, H. (2022). An improved EEGNet for single-trial EEG classification in rapid serial visual presentation task. *Brain Sci. Adv.* 8, 111–126. doi: 10.26599/BSA.2022.9050007

Zhang, L., Chen, X., Chen, L., Gu, B., Wang, Z., and Ming, D. (2021). Collaborative strategy research based on brain-computer interface of motor imagination. *J. Biomed. Eng.* 38, 409–416. doi: 10.7507/1001-5515.202007059

Zheng, L., Sun, S., Zhao, H., Pei, W., Chen, H., Gao, X., et al. (2020). A cross-session dataset for collaborative brain-computer interfaces based on rapid serial visual presentation. *Front. Neurosci.* 14:579469. doi: 10.3389/fnins.2020.579469

frontiers | Frontiers in Neuroscience

TYPE Opinion
PUBLISHED 09 March 2023
DOI 10.3389/fnins.2023.1153060

OPEN ACCESS

EDITED BY
Minpeng Xu,
Tianjin University, China

REVIEWED BY
Xulong Liu,
Northeastern University, China

*CORRESPONDENCE
Yanhong Zhou
✉ yhzhou168@163.com

SPECIALTY SECTION
This article was submitted to Neuroprosthetics, a section of the journal Frontiers in Neuroscience

RECEIVED 28 January 2023
ACCEPTED 15 February 2023
PUBLISHED 09 March 2023

CITATION
Wen D, Pang Z, Wan X, Li J, Dong X and Zhou Y (2023) Cross-task-oriented EEG signal analysis methods: Our opinion.
Front. Neurosci. 17:1153060.
doi: 10.3389/fnins.2023.1153060

Cross-task-oriented EEG signal analysis methods: Our opinion

Dong Wen[1,2], Zhenhua Pang[3], Xianglong Wan[1,2], Jingjing Li[3], Xianling Dong[4] and Yanhong Zhou[5]*

[1]School of Intelligence Science and Technology, University of Science and Technology Beijing, Beijing, China, [2]Key Laboratory of Perception and Control of Intelligent Bionic Unmanned Systems, Ministry of Education, Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing, China, [3]School of Information Science and Engineering, Yanshan University, Qinhuangdao, China, [4]Department of Biomedical Engineering, Chengde Medical University, Chengde, China, [5]School of Mathematics and Information Science and Technology, Hebei Normal University of Science and Technology, Qinhuangdao, China

KEYWORDS

EEG signal, cross-task, cross-subject, feature extraction, feature classification

1. Introduction

Research on cross-task EEG signals analysis methods has become a fast-growing research hotspot. In recent years, more and more researchers applied the features, which were widely used in EEG signal analysis to cross-task EEG signal analysis studies, including power spectral density (PSD) features (Touryan et al., 2016; Adewale and Panoutsos, 2019), fusion features (Kakkos et al., 2021), etc. The objective aimed to find ways to effectively deal with the differences between tasks. At the same time, some researchers have explored the classifiers which are more friendly to the differences between different tasks by comparing with the traditional feature classification methods, including multi-layer perceptron neural network (MLPNN) (Kamrud et al., 2021), domain adaptive methods (Zhou et al., 2022), sliding-window support vector machine (SVM) (Boring et al., 2020), etc. On the other hand, some new cross-task models based on deep learning models were proposed to narrow the differences between tasks, such as convolutional neural networks (CNNs) (Mota et al., 2021), recurrent neural networks (RNNs) (Gupta et al., 2021), metric-based methods (Jia et al., 2023), combinations of CNNs and RNNs (Zhang et al., 2019; Zhou et al., 2019; Taori et al., 2022), etc. However, there are still many unexplored areas in the field of cross-task EEG signal analysis methods, such as: task segmentation and complexity design (Kamrud et al., 2021), multi-source domain adaptive application (Zhou et al., 2022), multi-scale and multi-directional filter research (Taori et al., 2022), considering both feature extraction and feature classification, and increasing the amount of data. Furthermore, there are also some interconnections between cross-task analysis and relatively common cross-subject studies.

This study will review the literature related to cross-task EEG signal analysis from the perspective of feature extraction and feature classification, and discuss the relationship between cross-task research and cross-subject research for EEG signal analysis, and finally present the point of our original opinion in the purpose of providing useful suggestion for the research field of cross-task EEG signal analysis.

2. Cross-task EEG signal analysis based on feature extraction

With the development of EEG signal analysis methods, a series of studies on EEG signal analysis has found that many EEG signal feature extraction methods ignored the interference of different tasks on EEG signal analysis (Xing et al., 2022). Therefore, in order to improve cross-tasks results, more and more researchers are working to find features that perform better on cross-task.

2.1. Cross-task EEG signal analysis based on classical features

In EEG signal analysis research, as PSD is one of the most widely used features in EEG signal analysis, some cross-task studies started from PSD features for in-depth exploration. Touryan et al. used ICA to describe feature space, calculated PSD, and identified independent component (IC) sets in spectral properties using sequential forward floating selection (SFFS) (Touryan et al., 2016). The results showed that common components of cross-task EEG signals could be identified through this method. Furthermore, Adewale et al. designed a signal processing and feature extraction framework based on PSD (Adewale and Panoutsos, 2019), and found that PSD could be used as an excellent feature for mental workload estimation. Therefore, PSD features show excellent performance in cross-task EEG signal analysis.

2.2. Cross-task EEG signal analysis based on other features

In recent years, it has been found that cross-task EEG signal analysis using PSD features alone does not achieve the best results (Kakkos et al., 2021; Ke et al., 2021; Guan et al., 2022; Xing et al., 2022). Therefore, some studies have begun to use some features combined with PSD or propose new features.

(1) Research based on feature fusion. Kakkos et al. improved the performance of cross-task classification by combining PSD with functional connectivity (FC) features (Kakkos et al., 2021) and demonstrated that the use of brain feature fusion is more effective in cross-task. Ke et al. verified that task-independent auditory event-related potentials (tir-aERPs) have better adaptability than PSD (Ke et al., 2021), and will work on tir-aERPs and PSD feature fusion in their future studies.

(2) Research based on brain network features. Guan et al. proposed a dynamic brain network analysis method based on EEG microstates (Guan et al., 2022) and found that the use of dynamic functional connectivity metrics was more suitable for cross-task.

(3) Research based on fuzzy entropy features. Xing et al. used fuzzy entropy features for cross-task EEG signal analysis (Xing et al., 2022) and found that fuzzy entropy features are more adaptable to cross-tasks than other features.

In summary, the studies of cross-task EEG signals from the perspectives of PSD-based feature fusion and exploration of new features has garnered the attention of numerous researchers, yet further investigation is still required in terms of effective feature fusion and novel cross-task sharing features.

3. Cross-task EEG signal analysis based on feature classification

From the perspective of feature classification, although the differences between tasks corresponding to different EEG signals limit the cross-task versatility of existing classification models (Zhou et al., 2022), researchers are still committed to finding or constructing some relatively general cross-task feature classification models (Kamrud et al., 2021; Mota et al., 2021; Taori et al., 2022).

3.1. Cross-task EEG signal analysis based on classical classification methods

Classical classification models in the field of EEG signal analysis are emerging, but only a few methods exist in cross-task research field.

(1) Neural network method. Kamrud et al. studied the commonality of three different models in terms of cross-task: MLPNN, temporal convolutional network (TCN), TCN auto encoder (TCN-AE) (Kamrud et al., 2021), and the results showed that the best model for cross-task classification was the MLPNN frequency domain model.

(2) Domain adaptive method. Zhou et al. explored four domain adaptation methods to bridge differences between tasks (Zhou et al., 2022), and the results showed that the transfer joint matching method not only performed best, but always achieved the best performance compared to other methods.

(3) Support vector machine. Boring et al. compared the cross-task classification performance of SVM, linear discriminant analysis (LDA), and k-nearest neighbors (KNN) under sliding window (Boring et al., 2020), and the results showed that the performance of SVM was significantly better than that of other models.

The above studies analyzed the performance of some classical methods in the application of cross-task EEG signal analysis, and there may be some classical methods with better performance in the future that can be used for cross-task EEG signal analysis.

3.2. Cross-task EEG signal analysis based on other classification methods

In addition to the above classical methods, the following methods have also achieved excellent performance in cross-task EEG signal analysis.

(1) CNN. Mota et al. proposed a cross-task classification method based on CNN and compressed excitation blocks (Mota et al., 2021), and the results showed that compressed excitation blocks could be used to explore the dependence on EEG signal pathways.

(2) RNN. Gupta et al. proposed a deep RNN model (Gupta et al., 2021), which showed that the model could learn forward and reverse temporal dynamics and had long-term memory ability.

(3) Multi-classifier combination. Zhou et al. proposed a cross-task method for classification using raw data (Zhou et al., 2019), and the results showed that the method had good adaptability. In the same year, Wang et al. proposed a cascade structure (R3DCNN) of a deep recurrent and three-dimensional convolutional neural network (3DCNN) (Zhang et al., 2019), and the results showed that 3DCNN could be used to learn the spatial and spectral features of EEG signals, and the use of RNN layers that can obtain temporal representations improved the performance. Taori et al. proposed a structural model built on RNN and attention mechanisms (Taori

et al., 2022), and the results showed that the model could extract effective cross-task features from the space-time domain.

(4) Metric-based method. Jia et al. proposed a metric-based Spatial Filtering Transformer (MSFT) model, which used the angle margin loss function (Jia et al., 2023), and the results showed that the method had good application prospects in the field of cross-task EEG signal analysis.

The above methods showed excellent performance in cross-task research, and future research could tend to build new cross-task classification models.

4. Study on cross-task and cross-subject relationship for EEG signal analysis

The significant variability of EEG signals between individuals reduces the generalization ability of EEG analysis algorithms (Xu et al., 2021). Since 2005 or even earlier, cross-subject EEG signals analysis has flourished, a variety of cross-subject methods has been designed, Tangermann et al. proved that recursive channel elimination (RCE) can be used for cross-subject combinatorial data analysis (Tangermann et al., 2005), and Dyson et al. conducted cross-subject studies by sequential forward-floating search algorithms (Dyson et al., 2010). On the other hand, cross-task research has only been in its infancy in recent years, and the current cross-task approach has a mutually reinforcing relationship with the existing cross-scenario approach.

(1) Cross-subject methods are innovatively applied to cross-task research. In 2012, Khalighi et al. proposed a cross-subject method for unsupervised domain adaptation (Khalighi et al., 2012). In 2021, Zhao et al. proposed an aligned multi-source domain adaptation method for cross-subject (Zhao et al., 2021). Zhou et al. proposed a cross-task domain adaptive method based on the above (Zhou et al., 2022). In 2016, Hajinoroozi et al. proposed a cross-subject method for channel convolutional neural networks (Hajinoroozi et al., 2016), and 5 years later, Mota et al. proposed a CNN-based cross-task method (Mota et al., 2021). Similarly, in 2019, Hang et al. achieved feature fusion of cross-subject EEG signals (Hang et al., 2019), and 2 years later, Kakkos et al. also explored feature fusion of cross-task EEG signals (Kakkos et al., 2021).

(2) Cross-task methods are innovatively applied to cross-subject research. In 2016, Touryan et al. studied the PSD features of cross-task EEG signals (Touryan et al., 2016), and 2 years later, Booth et al. carried out in-depth research on the PSD features of cross-subject EEG signals (Booth et al., 2018).

In summary, there is a certain correlation between cross-task and cross-individual research methods, and combining cross-task and cross-subject research will make EEG analysis methods more versatile.

5. Discussion

In this study, the cross-task EEG signal analysis method was analyzed from three aspects: feature extraction, feature classification, and the relationship between cross-task and cross-subject methods. While these studies have yielded promising results, more exploration is needed before confident conclusions can be drawn. Therefore, this paper raises our own opinion on the future research of cross-task EEG signal analysis.

(1) Increasing the sample size. In the future, more EEG data can be collected or data augmentation techniques can be used to increase the sample size in order to improve the generalization performance of cross-task methods.

(2) Research from the perspectives of both feature extraction and feature classification. At present, most cross-task researches were carried out independently from the perspective of feature extraction or feature classification, and it is a valuable practice to find common ground from these two perspectives simultaneously in the future, such as multi-source domain adaptation (Zhou et al., 2022) and multi-scale and multi-directional filter (Taori et al., 2022) used for the study of single cross-task EEG signals.

(3) Task subdivision. In the future, it is necessary to subdivide different tasks to improve the practicability of cross-task research, such as the corresponding EEG signal datasets can be analyzed on tasks with the same cognitive domain but different cognitive training content, or tasks with different cognitive domains and different cognitive training content.

(4) The research of cross-task regression models can be explored in depth. The study results of Ke et al. suggested that regression models rather than classifiers should be used for obtaining optimal results in some cross-task studies (Ke et al., 2014). In the future, we can try to explore cross-task research on other types of regression models.

In conclusion, this paper introduced some research trends in the future. If cross-task research can continue to advance in these areas, it will take this type of research to a higher level.

Author contributions

DW, ZP, and YZ contributed to conception and design of the study. ZP searched the database. DW, ZP, and JL performed the analysis of literatures. DW and ZP wrote the first draft of the manuscript. XW and JL wrote sections of the manuscript. YZ and XD revised this paper. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work was supported in part by National Key Research and Development Program of China (2021YFF1200603) and National Natural Science Foundation of China (61876165, 62276022, 62206014, and 61503326).

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

References

Adewale, Q., and Panoutsos, G. (2019). Mental workload estimation using wireless EEG Signals. *bioRxiv*. 755033. doi: 10.1101/755033

Booth, B. M., Seamans, T. J., and Narayanan, S. S. (2018). An evaluation of EEG-based metrics for engagement assessment of distance learners. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 307–310.

Boring, M. J., Ridgeway, K., Shvartsman, M., and Jonker, T. R. (2020). Continuous decoding of cognitive load from electroencephalography reveals task-general and task-specific correlates. *J. Neural Eng.* 17, 056016. doi: 10.1088/1741-2552/abb9bc

Dyson, M., Sepulveda, F., and Gan, J. Q. (2010). Localisation of cognitive tasks used in EEG-based BCIs. *Clinical Neurophysiol.* 121, 1481–1493. doi: 10.1016/j.clinph.2010.03.011

Guan, K., Zhang, Z., Chai, X., Tian, Z., Liu, T., and Niu, H. (2022). EEG based dynamic functional connectivity analysis in mental workload tasks with different types of information. *IEEE Trans. Neural Syst. Rehabil. Eng.* 30, 632–642. doi: 10.1109/TNSRE.2022.3156546

Gupta, S. S., Taori, T. J., Ladekar, M. Y., Manthalkar, R., Gajre, S. S., and Joshi, Y. V. (2021). Classification of cross task cognitive workload using deep recurrent network with modelling of temporal dynamics. *Biomed. Signal Process. Control.* 70, 103070. doi: 10.1016/j.bspc.2021.103070

Hajinoroozi, M., Mao, Z., Jung, T., Lin, C., and Huang, Y. (2016). EEG-based prediction of driver's cognitive performance by deep convolutional neural network. *Signal Process. Image Commun.* 47, 549–555. doi: 10.1016/j.image.2016.05.018

Hang, W., Feng, W., Du, R., Liang, S., Chen, Y., Wang, Q., et al. (2019). Cross-subject EEG signal recognition using deep domain adaptation network. *IEEE Access* 7, 128273–128282. doi: 10.1109/ACCESS.2019.2939288

Jia, X., Song, Y., and Xie, L. (2023). Excellent fine-tuning: from specific-subject classification to cross-task classification for motor imagery. *Biomed. Signal Process. Control.* 79, 104051. doi: 10.1016/j.bspc.2022.104051

Kakkos, I., Dimitrakopoulos, G. N., Sun, Y., Yuan, J., Matsopoulos, G. K., Bezerianos, A., et al. (2021). EEG fingerprints of task-independent mental workload discrimination. *IEEE J. Biomed. Health Inform.* 25, 3824–3833. doi: 10.1109/JBHI.2021.3085131

Kamrud, A., Borghetti, B. J., Schubert-Kabban, C. M., and Miller, M. (2021). Generalized deep learning EEG models for cross-participant and cross-task detection of the vigilance decrement in sustained attention tasks. *Sensors* 21, 5617. doi: 10.3390/s21165617

Ke, Y., Jiang, T., Liu, S., Cao, Y., Jiao, X., Jiang, J., et al. (2021). Cross-task consistency of electroencephalography-based mental workload indicators: comparisons between power spectral density and task-irrelevant auditory event-related potentials. *Front. Neurosci.* 15, 703139. doi: 10.3389/fnins.2021.703139

Ke, Y., Qi, H., He, F., Liu, S., Zhao, X., Zhou, P., et al. (2014). An EEG-based mental workload estimator trained on working memory task can work well under simulated multi-attribute task. *Front. Hum. Neurosci.* 8.

Khalighi, S., Sousa, T., and Nunes, U. J. (2012). "Adaptive automatic sleep stage classification under covariate shift," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2259–2262.

Mota, M., Silva, P. H., Luz, E. J., Moreira, G. J., Schons, T., Moraes, L., et al. (2021). A deep descriptor for cross-tasking EEG-based recognition. *PeerJ. Comput. Sci.* 7, e549. doi: 10.7717/peerj-cs.549

Tangermann, M., Lal, T. N., Hinterberger, T., Bogdan, M., Hill, N. J., Birbaumer, N., et al. (2005). Robust EEG channel selection across subjects for brain-computer interfaces. *EURASIP J. Adv. Signal Process.* 2005, 1–10. doi: 10.1155/ASP.2005.3103

Taori, T. J., Gupta, S. S., Bhagat, S., Gajre, S. S., and Manthalkar, R. (2022). Cross-task cognitive load classification with identity mapping-based distributed CNN and attention-based RNN using gabor decomposed data images. *IETE J. Res.* doi: 10.1080/03772063.2022.2098191. [Epub ahead of print].

Touryan, J., Lance, B., Kerick, S. E., Ries, A. J., and McDowell, K. (2016). Common EEG features for behavioral estimation in disparate, real-world tasks. *Biol. Psychol.* 114, 93–107. doi: 10.1016/j.biopsycho.2015.12.009

Xing, Z., Dong, E., Tong, J., Sun, Z., and Duan, F. (2022). "Application of mental fatigue classification in cross task paradigm," in *2022 IEEE International Conference on Mechatronics and Automation (ICMA)*, 1750–1754.

Xu, M. P., He, F., Jung, T. P., Gu, X. S., Ming, D. (2021). Current challenges for the practical application of electroencephalography-based brain-computer interfaces. *Engineering* 7, 1710–1712. doi: 10.1016/j.eng.2021.09.011

Zhang, P., Wang, X., Zhang, W., and Chen, J. (2019). Learning spatial–spectral–temporal EEG features with recurrent 3D convolutional neural networks for cross-task mental workload assessment. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 31–42. doi: 10.1109/TNSRE.2018.2884641

Zhao, Y., Dai, G., Borghini, G., Zhang, J., Li, X., Zhang, Z., et al. (2021). Label-based alignment multi-source domain adaptation for cross-subject EEG fatigue mental state evaluation. *Front. Hum. Neurosci.* 15, 706270. doi: 10.3389/fnhum.2021.706270

Zhou, Y., Xu, T., Li, S., and Shi, R. P. (2019). Beyond engagement: an EEG-based methodology for assessing user's confusion in an educational game. *Univ. Access Inform. Society* 18, 551–563. doi: 10.1007/s10209-019-00678-7

Zhou, Y., Xu, Z., Niu, Y., Wang, P., Wen, X., Wu, X., et al. (2022). Cross-task cognitive workload recognition based on EEG and domain adaptation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 30, 50–60. doi: 10.1109/TNSRE.2022.3140456

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 09 March 2023
DOI 10.3389/fnins.2023.1105696

OPEN ACCESS

EDITED BY
Ernest N. Kamavuako,
King's College London, United Kingdom

REVIEWED BY
Sagarika Mukesh,
IBM Research, United States
Axel Steiger,
Ludwig Maximilian University of Munich, Germany
Vittorio Cuculo,
University of Milan, Italy

*CORRESPONDENCE
Zhaoyang Huang
✉ drhuangzy@gmail.com
Dong Ming
✉ richardming@tju.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Neuroprosthetics, a section of the journal Frontiers in Neuroscience

RECEIVED 23 November 2022
ACCEPTED 07 February 2023
PUBLISHED 09 March 2023

CITATION
Chen C, Meng J, Belkacem AN, Lu L, Liu F, Yi W, Li P, Liang J, Huang Z and Ming D (2023) Hierarchical fusion detection algorithm for sleep spindle detection. *Front. Neurosci.* 17:1105696. doi: 10.3389/fnins.2023.1105696

Hierarchical fusion detection algorithm for sleep spindle detection

Chao Chen[1,2†], Jiayuan Meng[1†], Abdelkader Nasreddine Belkacem[3], Lin Lu[4], Fengyue Liu[2], Weibo Yi[5], Penghai Li[2], Jun Liang[6], Zhaoyang Huang[7,8*] and Dong Ming[1*]

[1]Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin, China, [2]Key Laboratory of Complex System Control Theory and Application, Tianjin University of Technology, Tianjin, China, [3]Department of Computer and Network Engineering, College of Information Technology, United Arab Emirates University, Al Ain, United Arab Emirates, [4]Zhonghuan Information College Tianjin University of Technology, Tianjin, China, [5]Beijing Machine and Equipment Institute, Beijing, China, [6]Department of Rehabilitation, Tianjin Medical University General Hospital, Tianjin, China, [7]Department of Neurology, Xuanwu Hospital, Capital Medical University, Beijing, China, [8]Beijing Key Laboratory of Neuromodulation, Beijing, China

Background: Sleep spindles are a vital sign implying that human beings have entered the second stage of sleep. In addition, they can effectively reflect a person's learning and memory ability, and clinical research has shown that their quantity and density are crucial markers of brain function. The "gold standard" of spindle detection is based on expert experience; however, the detection cost is high, and the detection time is long. Additionally, the accuracy of detection is influenced by subjectivity.

Methods: To improve detection accuracy and speed, reduce the cost, and improve efficiency, this paper proposes a layered spindle detection algorithm. The first layer used the Morlet wavelet and RMS method to detect spindles, and the second layer employed an improved k-means algorithm to improve spindle detection efficiency. The fusion algorithm was compared with other spindle detection algorithms to prove its effectiveness.

Results: The hierarchical fusion spindle detection algorithm showed good performance stability, and the fluctuation range of detection accuracy was minimal. The average value of precision was 91.6%, at least five percentage points higher than other methods. The average value of recall could reach 89.1%, and the average value of specificity was close to 95%. The mean values of accuracy and F1-score in the subject sample data were 90.4 and 90.3%, respectively. Compared with other methods, the method proposed in this paper achieved significant improvement in terms of precision, recall, specificity, accuracy, and F1-score.

Conclusion: A spindle detection method with high steady-state accuracy and fast detection speed is proposed, which combines the Morlet wavelet with window

RMS and an improved k-means algorithm. This method provides a powerful tool for the automatic detection of spindles and improves the efficiency of spindle detection. Through simulation experiments, the sampled data were analyzed and verified to prove the feasibility and effectiveness of this method.

KEYWORDS

sleep spindle detection, hierarchical fusion detection algorithm, EEG, Morlet wavelet, SVM

1. Introduction

Sleep spindles refer to the recognizable 11–16 Hz sinusoidal periodic pulse sequence on an electroencephalogram (EEG) during sleep. According to the American Academy of Sleep Medicine (AASM) (Iber et al., 2007), the spindles mainly occur in stage two of Non-Rapid-Eye-Movement (NREM) sleep (Schilling et al., 2018; Chriskos et al., 2019). Spindle detection plays a crucial role in sleep staging research and clinical disease diagnosis (Dehnavi et al., 2019; Zhang et al., 2020). Studies have found that the number and density of spindles are associated with many diseases (Fogel and Smith, 2011), such as Parkinson's disease (Latreille et al., 2015), Alzheimer's disease, major depression, autism, insomnia, and schizophrenia (Limoges et al., 2005; Keshavan et al., 2011; Astori and Luthi, 2013; Davies et al., 2016; Spironelli et al., 2020). In addition, the function of sleep spindles is associated with human intelligence and sleep-dependent memory consolidation (Fogel et al., 2007; Ujma et al., 2014). Notably, spindles are a reflecting functional brain-state biomarker and have solid supplementary diagnostic value (Zhao et al., 2017; Mensen et al., 2018).

Currently, in clinical diagnosis, spindle detection mainly depends on the subjective experience of doctors, the so-called gold standard of spindle detection (Dakun et al., 2015). Generally, the manual detection of spindles allows many experts to select the spindles simultaneously. It is challenging to detect sleep spindles with this method due to the high detection cost and longer detection time (Lacourse et al., 2019). Thus, Wamsley et al. (2012) explored a method to detect spindles using the wavelet transform automatically and found an overall number and density of 62.5%, a result that was far from ideal. Athanasios and Clifford (2015) and others proposed a probabilistic wavelet estimation algorithm based on the wavelet algorithm for the automatic detection of spindles. However, due to the spindles' irregularities, the performance in recall rate was not ideal, with a minimum of 14.4% and a maximum of 83.2%. Hence, the stability of the recall rate was poor. Martin (Martin et al., 2012) et al. used the window root mean square (RMS) method to detect the spindle density of young and older people by calculating the RMS value. The accuracy was only 72%, but the recall rate was 83%. Furthermore, concerning age, it was found that the density, duration, and amplitude of spindles in young subjects were greater than those in older subjects, and age factors affected the detection of spindles. Mporas et al. (2013) proposed the hidden Markov model (HMM) and support vector machine (SVM) to process EEG signals. Through the fusion of HMM and SVM (HMM&SVM), the output recognition spindle results were combined to extract the final sleep spindle detection results. The average performance regarding precision was 88%, and recall was 76%, but the process was complex, and the operation was cumbersome, bringing great inconvenience to the experiment. Lacourse et al. (2019) proposed a detection method for spindles, also known as A7. The recall rate of this method reached 68%, with no crucial change compared with the average value of 72.7% within and between experts in sleep spindle detection (Devuyst et al., 2006).

The performance index of the above algorithms is in the range of 60–70%, and thus their performance is not ideal. Currently, there is a shortage of public databases for spindle detection, and the lack of databases has led to difficulties in validating the stability of different detection algorithms (Wendt et al., 2015). The performance of spindle detection can be further improved on the original basis.

Herein, a new notion of hierarchical fusion algorithm is proposed to improve the defects of expert manual detection and the automatic detection of spindles. It addresses the advantages and disadvantages of the automatic detection methods of sleep spindles by combining the advantages of the spindle detection method and overcoming the shortcomings of previous methods to detect spindles. The Morlet wavelet and RMS algorithms are used as the first layer of the basic algorithm. After fusing the results of the two automatic detection methods, the k-means algorithm in the second layer is used for clustering to get the final result. The Morlet wavelet detection method, window RMS detection method, HMM&SVM algorithm, and the newly proposed hierarchical fusion algorithm are compared in the detection results. The data results optimized by the hierarchical fusion algorithm substantially improve the performance of spindle automatic detection. In order to improve the efficiency and accuracy of spindle detection, we propose a new spindle detection algorithm, which combines the Morlet wavelet detection algorithm, RMS algorithms, and k-means algorithm. The average accuracy of this method is 91.6%, at least five percentage points higher than other methods.

2. Materials and methods

2.1. Data sources and methods

The experimental data came from the sleep monitoring room of Beijing Xuanwu Hospital. This experiment was designed to collect sleep data from 20 subjects with sleep disorders, all of whom were between 20 and 40 years old in order to avoid the effect

of age differences on the number of spindles. Their average age was 31 years old, 11 participants were female, and they were all recruited from the community. Twenty subjects were scored on the Pittsburgh Sleep Quality Index before EEG acquisition. Those with a score greater than or equal to 11 were patients with sleep disorders. Table 1 shows the Pittsburgh Sleep Quality Index scores of the 20 subjects in this paper. The higher the score, the worse the sleep quality.

The data collection equipment adopted polysomnography (PSG). The equipment could record many channels simultaneously during the subjects' sleep, such as EEG, ECG, EOG, EMG, airflow, and oxygen saturation. As shown in Figure 1, the international 10 / 20 standard electrode placement system shows the electrode positions of the relevant EEG signals collected in this paper. Based on the relevant research on the brain regions with frequent spindles, this study collected and analyzed the EEG signals collected from only six channels. Thus, the EEG data of the left and right channels collected in this experiment included F3 \F4 \C3 \C4 \O1 \O2. Here, F3 and F4 are frontal brain regions, C3 and C4 are central brain regions, and O1 and O2 represent the occipital regions. Spindles appear most frequently at these positions, and C3/C4 have the most spindle appearances. The sampling frequency of the EEG signal acquisition equipment used in this experiment was 1,024 Hz.

The data collection for each subject lasted from 9:30 p.m. to 6:30 a.m. The start and end times of the subjects entering sleep were not wholly consistent. Thus, sleep EEG data of up to 8 h from the first entering sleep stage were intercepted based on the stage of sleep for research to facilitate and accurately analyze the characteristic differences in individual and overall sleep spindles in the later stages. Professional sleep researchers operated and guided the monitor used in the sleep monitoring room. To ensure the accuracy of the physiological signal collection, the subjects were required to prohibit the intake of alcohol and the consumption of caffeine, sedatives, hypnotics, and other relevant drugs that could affect data collection 1 week before the sleep monitoring. Before data acquisition, the subjects were required to have taken a bath and cleaned their heads to ensure good contact between the electrodes and the skin to assist the collecting of data as much as possible. Furthermore, the subjects were required to urinate in advance or place a disposable night pot next to the hospital bed to avoid the disproportionate impact of large-scale activities at night on data acquisition. Next, the experimenter recorded the subject's name, weight, gender, age, and other essential information. The relevant electrode connection points were wiped with a cleaning paste specifically used to clean the electrodes before the electrodes were connected. When the electrode was placed, the accuracy and firmness of the electrode's position could be secured, and the electrodes were placed in a specified order. After the work preparation, before collection was completed, the monitor was opened to record the data collected by the software. During the acquisition process, the cell phones of the subject and experimenter were turned off to keep the environment quiet to avoid the interference of external environmental sound on the experimental data. After data collection in the morning, the experimenter turned off the equipment and woke up the subject. At this point, the experimental data collection was finished.

The sampling frequency of the experimental data was 1,024 Hz, which was downsampled to 512 Hz during the pre-processing for this paper. In this experiment, the bilateral mastoid was used as the mean reference for re-referencing. The Morlet wavelet-based and RMS automatic detection methods used in this experiment both require band-pass filtering of the data prior to spindle wave detection. The raw EEG signal was pre-processed with band-pass filtering from 5 to 35 Hz prior to automatic spindle wave detection using Morlet wavelets. Pre-processing of the raw data with 11–16 Hz bandpass filtering was conducted before using the RMS algorithm, where the frequency band was chosen based on the standard definition of the spindle frequency distribution.

2.2. Proposed algorithm

Figure 2 is the flowchart of the newly proposed hierarchical fusion spindle automatic detection algorithm. Two single detection algorithms, the Morlet wavelet and window RMS were used in the fusion algorithm, merged with the improved k-means algorithm.

The specific steps of the hierarchical fusion algorithm were as follows:

Firstly, the collected sleep EEG signal was pre-processed. The resultant sleep signal was transmitted to the Morlet wavelet and window RMS spindle automatic detector to judge the true and false spindle of the two detectors' output results.

The Morlet wavelet function was closer to the spindles and more conducive to spindle detection. Spindle detection function based on the Morlet wavelet was defined as:

$$f(x) = (\pi F_B)^{-0.5} \exp(2\pi i F_C x) \exp(-x^2/F_B) \quad (1)$$

Where F_B is the bandwidth of wavelet transform, $F_B = 2s^2$, and $s = n/2\pi F_C$. The value of F_B depends on the magnitude of the values of n and F_C. The n represents the number of cycles of the Morlet wavelet. F_C is the center frequency. Here, set $n = 7$ is a typical default value when balancing the time-frequency domain.

The Morlet wavelet function performed time-frequency conversion on the pre-processed EEG signal, and a threshold function was used to detect the spindle. After all pre-processing, the threshold was defined as 4.5 times the average signal amplitude. The average moving value was calculated using a 0.1 s sliding window to extract the spindle in the frequency band. When the wavelet signal exceeded the threshold and the duration was in the range of 0.5–3 s, it was deemed a spindle. If the distance between the two spindles was less than 1 s and the duration was less than 3 s, the spindle was combined. This detection result was reserved for the fusion of later experiments.

The window RMS algorithm used a linear phase finite impulse response filter for 11–16 HZ band-pass filtering of the EEG original signal in the NREM period of the C3 channel, doubling the order of the filter. The filtered EEG signal was determined using a time window of 0.25 s, and the threshold value was 0.95 times the mean value. The spindle was identified as two consecutive root mean squares calculated time points, exceeding the threshold and lasting between 0.5 and 3 s. The RMS value was calculated every 5 s by employing the following formula:

$$RMS - A = \sqrt{\frac{\sum_{i=1}^{N} X_i^2}{N}} \quad (2)$$

TABLE 1 Pittsburgh sleep quality index score.

Subject	Age	Sex	Score	Subject	Age	Sex	Score
DS1	21	Male	14	DS11	24	Male	12
DS2	35	Male	17	DS12	40	Female	18
DS3	40	Female	19	DS13	31	Female	20
DS4	25	Male	16	DS14	22	Male	16
DS5	34	Female	16	DS15	30	Female	15
DS6	20	Female	18	DS16	37	Female	20
DS7	36	Female	11	DS17	25	Male	18
DS8	29	Male	13	DS18	33	Female	17
DS9	38	Female	16	DS19	34	Male	14
DS10	32	Male	14	DS20	29	Female	16

DS represents subjects with sleep disorders. Subjects with a score greater than or equal to 11 are subjects with sleep disorders.

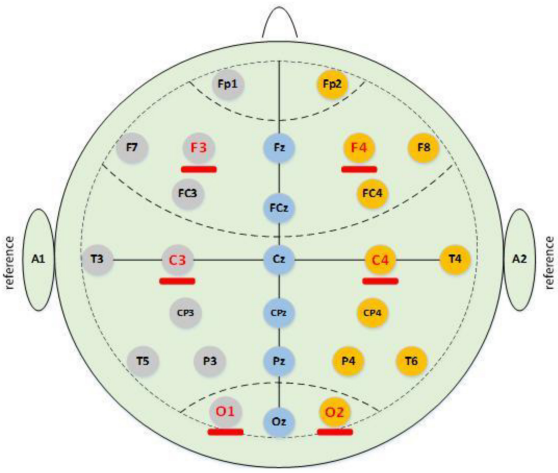


FIGURE 1
International 10 / 20 standard electrode placement system.

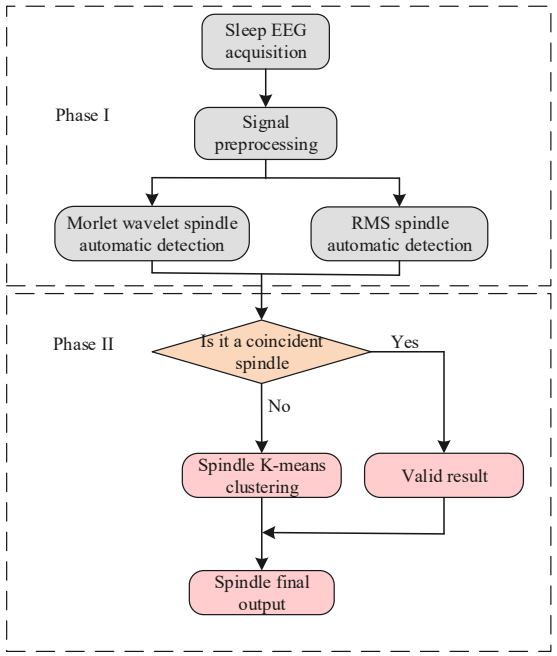


FIGURE 2
Flowchart of the fusion algorithm.

RMS-A refers to the root mean square of the spindle wave frequency band. Where X_i^2 is the square of the amplitude of the sampling point i, and N is the number of sampling points within 5 s.

Let us repeat the spindle detection link. The spindle detected by the Morlet wavelet and RMS method was divided into coincident and non-coincident spindles. When the spindle detected using the two methods overlapped in the time series, they were considered coincident spindles. As can be seen in Figure 3A, the coincident spindle was regarded as the effective result of the fusion algorithm detection. When the spindle time did not repeat, it was considered a non-coincident spindle, as can be seen in Figure 3B.

This study identified the coincident spindle as the same spindle and directly classified it into the final automatic detection result set. The non-coincident spindle set was treated as the sample of clustering input. After cluster analysis, many non-spindle clusters were removed, and the samples of the remaining clusters and the coincident spindle set were taken as the final output result.

The non-coincident spindle needed to be further analysed and processed by the K-means clustering method. The K-means algorithm was relatively stable, and had a very smooth clustering effect. The K-means clustering algorithm needed to calculate the distance between the points of each cluster. The distance metric commonly used in the k-means algorithm is Euclidean distance.

Euclidean distance (L2):

$$d_{12} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (3)$$

The density was the distance between points to judge the abnormal points. After removing the outliers, the data were put into K-means clustering, improving the accuracy of clustering, reducing the amount of clustering data, and enhancing data processing speed.

The sum of squared errors (SSE) was used as the objective function to evaluate the clustering effect so that the clustering result can obtain the minimum SSE value (Klampanos et al., 2009).

The local outlier factor detection method optimized K-means clustering, addressing the detection problem of amplitude affecting the spindle.

The amplitude of the spindles was in the range of 10–60 μV. When clustering the non-coincident spindles, the amplitude was used as the input. Five categories were clustered based on the amplitude, and the maximum and minimum were removed. The remaining three categories were used as data for further analysis. After clustering, clusters with many non-spindles were discarded, leaving the actual spindle clusters.

Finally, the final fusion result was the clustering and the coincident spindle.

In this paper, when using the k-means algorithm for clustering, the method of calculating the difference statistics was adopted to select the k value.

2.3. Evaluation method

Twenty subjects were tested manually using the gold standard of spindle detection to prove the reliability of the hierarchical fusion spindle automatic detection method. The gold standard adopted was three experts conducting the sample's artificial spindle detection. The intersection of the three experts' detection results were determined as spindles here, and the detection results not within the intersection were regarded as non-spindle sets.

Then according to the confusion matrix and the gold standard of spindle detection, the evaluation indexes, such as TPR (recall), specificity, accuracy, precision, and F1-score, were calculated to evaluate the detector's performance.

The following was the meaning and relevant calculation formula of these evaluation indicators:

$$\text{recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{4}$$

$$\text{precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{5}$$

$$\text{specificity} = TN/(TN + FP) \tag{6}$$

$$\text{accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{TN} + \text{FN}) \tag{7}$$

$$\text{F1} = 2 * \frac{\text{pression} * \text{recall}}{\text{pression} + \text{recall}} \tag{8}$$

TP plus FN was the sample set of the actual spindles detected by experts, and TP plus FP was the sample set predicted as actual spindles by the detection algorithm. TP represents samples that were actually true spindle waves and detected by the automatic detection algorithm as true spindle waves; FP represents samples that were non-spindle waves but detected as true spindle waves; FN represents samples that were actually true spindle waves but predicted by the detection algorithm as non-spindle waves; and TN represents samples that are actually non-spindle waves and predicted by the detection algorithm as non-spindle waves.

3. Analysis of experimental results

The spindle detector automatically detected the spindle and contrasted it with the spindle detected by experts. R-spindles represented the actual spindle in the automatically detected spindles and A-spindles denoted all spindles automatically detected by the algorithm. The E-spindles symbolized the actual spindle detected by experts, and the intersections of the three experts' two or three detection results were considered E-spindles.

Table 2 shows the number of spindle waves detected by the Morlet wavelet, windowed RMS, HMM&SVM, and hierarchical fusion algorithms compared to the real spindle waves labeled by experts. In Table 2, A-spindle represents the number of all spindle waves automatically detected by the algorithm, R-spindle represents the number of true spindle waves among the automatically detected spindle waves, and E-spindle represents the true spindle waves detected by the expert.

As shown in Table 2, the total number of spindles automatically detected by the four methods differed for the same data. The four methods detected the most spindles in sample DS15, while the total spindles detected in sample DS6 were the least. The number of samples detected by the Morlet wavelet algorithm was small. Compared with the wavelet algorithm, the RMS-based algorithm could detect more spindles. The average number of R-spindles detected by HMM&SVM was 712, higher than the Morlet wavelet and RMS

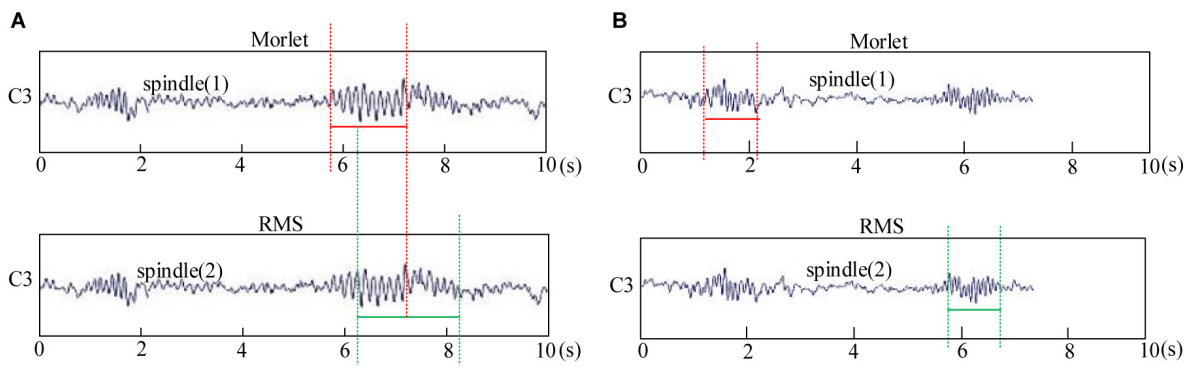


FIGURE 3
Spindle automatic detection results of the Morlet wavelet and root mean square (RMS). **(A)** Morlet wavelet and RMS detect the coincident spindle. **(B)** Morlet wavelet and RMS detect non-coincident spindle.

TABLE 2 Comparison of the results of different detection algorithms.

Subject	Morlet wavelet algorithm		RMS algorithm		HMM-SVM algorithm		Fusion algorithm		E-spindle
	A-spindle	R-spindle	A-spindle	R-spindle	A-spindle	R-spindle	A-spindle	R-spindle	
DS1	732	647	1409	998	1258	994	1124	1005	1136
DS2	467	424	831	602	802	685	725	668	730
DS3	263	229	518	383	617	527	511	472	551
DS4	231	204	589	406	523	436	456	429	493
DS5	439	376	977	747	821	765	792	763	848
DS6	184	160	632	491	635	532	530	487	590
DS7	702	605	1346	984	866	603	1063	994	1073
DS8	479	427	746	590	733	616	630	581	657
DS9	212	187	692	526	805	668	621	578	706
DS10	643	574	846	669	762	694	792	706	779
DS11	752	629	1415	1070	1238	1035	1248	1125	1265
DS12	525	481	968	708	774	688	834	761	819
DS13	596	528	724	571	729	605	634	583	676
DS14	381	334	898	639	880	763	759	702	798
DS15	1066	938	1635	1323	1458	1237	1507	1359	1491
DS16	539	456	736	608	664	503	687	643	736
DS17	551	492	939	712	875	756	835	765	839
DS18	572	489	929	674	802	701	790	688	763
DS19	360	324	727	598	768	654	689	640	702
DS20	736	635	965	763	1033	789	856	784	891
DS-average	521	457	926	703	852	712	804	737	827

algorithm, demonstrating that this algorithm could recall more spindles.

Compared with the Morlet wavelet method, the R-spindle/A-spindle ratio of the hierarchical fusion algorithm was higher, reaching 91.67%, indicating that the hierarchical fusion algorithm improved the recall rate. Compared with the RMS algorithm and the HMM&SVM algorithm, the number of A-spindles of the hierarchical fusion algorithm decreased by 122 and 48, respectively, without a substantial change. The number of R-spindles did not decrease but instead increased by 34 and 25, respectively, showing that the accuracy of the newly proposed algorithm had been substantially enhanced.

The A-spindle of the hierarchical fusion algorithm had the recombined set of the two spindle detection results. After clustering, they had the Morlet, the RMS, and the non-recombined sets. The maximum number of spindles detected reached 1,507 in DS15 samples and 530 in DS6 samples, and the average value was 804. However, the maximum value of R-spindles reached 1,359 in DS15 samples, and the average value was 737. The fusion algorithm was closest to the actual spindle value detected by experts.

Figure 4 contrasts the hierarchical fusion algorithm with the other three automatic detection algorithms by calculating five performance evaluation indexes: Precision, Recall, Specificity, Accuracy, and F1-score.

The fusion algorithm represented the newly proposed hierarchical fusion algorithm, and Morlet wavelet, RMS, and HMM&SVM represented the Morlet wavelet, window RMS, and HMM and SVM, respectively. Figure 4 shows that the Precision of the spindle detection based on the Wavelet algorithm could meet the accuracy requirements under the current standard, scoring more than 85%. However, a critical gap existed between the maximum and minimum values of Recall, which had large fluctuations, and its stability needs to be improved. Only the evaluation index of Specificity of spindle automatic detection based on the RMS algorithm reached more than 90%. Likewise, the Specificity evaluation index of the HMM&SVM algorithm was more than 90% but lower than the other three detection algorithms. In the subject samples, the Recall index of the hierarchical fusion algorithm reached 92.9% at its highest, 81.9% at its lowest, and with an average of 89.1 $\pm$ 8.84%. In the performance of Precision, the maximum value reached was 96.3%, the minimum value was 87.1%, and the average value was 91.6 $\pm$ 3.06%. This Precision average value was better than those with the RMS and HMM&SVM algorithms and slightly better than those with the Morlet wavelet algorithm. The performances of Specificity, Accuracy, and F1-score were also ideal, and the average values were higher than those of the other three methods. The average values in the subject samples were 96 $\pm$ 2.89, 90.4 $\pm$ 6.50, and 90.3 $\pm$ 2.49%, respectively. The maximum, minimum, and average of each evaluation index of

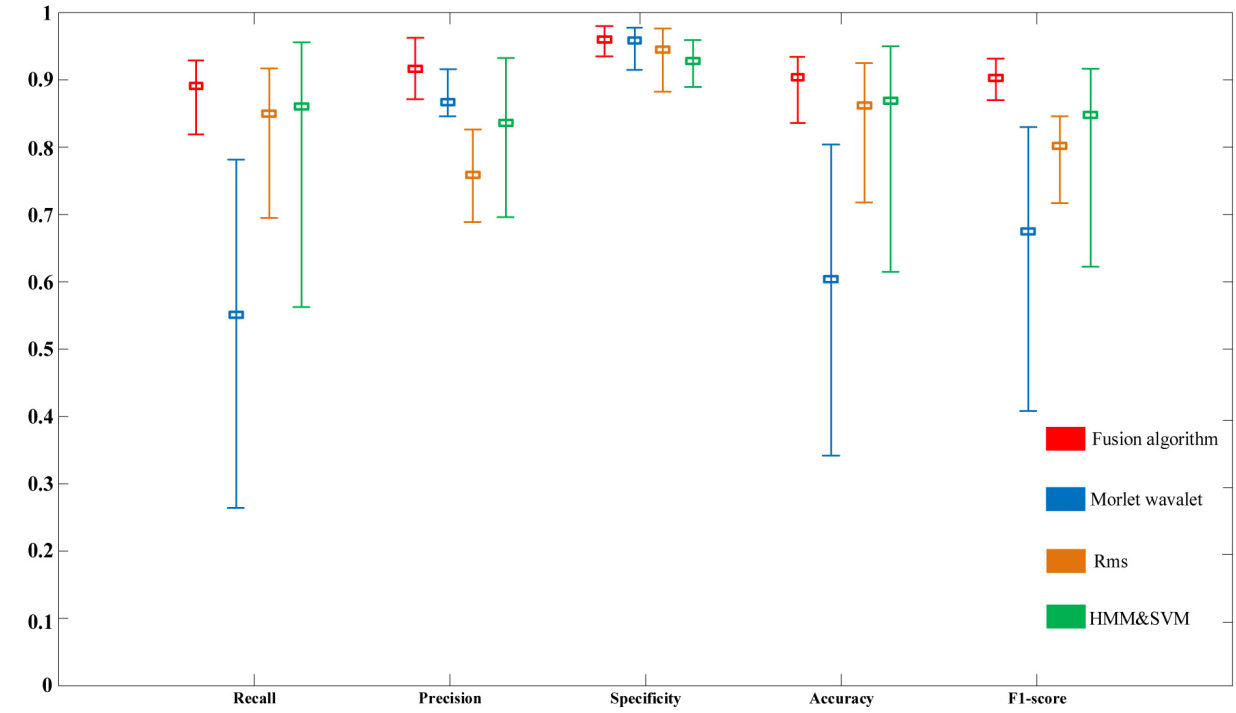


FIGURE 4
Performance comparison of detection algorithms based on subject data.

the hierarchical fusion algorithm were substantially improved compared with the other three algorithms.

4. Discussion

As sleep spindles are a characteristic index to evaluate sleep quality, the pursuit of the accurate detection of sleep spindles is imperative. The fusion method proposed in this paper was based on the Morlet wavelet and the Window RMS algorithm, combined with the improved k-means algorithm, and then the data results of algorithmic optimization. The fusion detection spindle algorithm combined the advantages of the three algorithms. Compared with previous research on spindle automatic detection, it improved calculation and effectiveness. The spindle detection results of the hierarchical fusion algorithm are shown in **Figure 4**, demonstrating better consistency with the expert detection results than the other methods. The detection rate of actual spindles was critically improved compared with the previous detection methods, effectively improving the accuracy and speed of the automatic spindle detection. This proposed method improved the shortcomings of the existing spindle detection methods and effectively enhances the detection efficiency of doctors, and reduces the visual inspection workload of sleep clinicians and the cost of detection (Jiang et al., 2021).

In the study of Warby et al. (2014), the amplitude and duration of the spindle decreased with age, probably damaging the spindle recognition performance. Therefore, in this experiment, we also paid attention to the interference of other unnecessary sample factors. PSG equipment was used to experiment on 20 subjects, and sleep data of 8 h were intercepted for research and analysis. The age of the subjects was controlled between 20 and 40 years old. The experimental data were truncated (Herrmann et al., 2016) to ensure a specific length of sample time. The wavelet automatic spindle detection method proposed by Wamsley et al. (2012) is more hierarchical than the fusion algorithm. The wavelet automatic detection algorithm needs much calculation and cumbersome experiments. The hierarchical fusion algorithm simplified the calculation, solving this problem. After completing the automatic spindle detection, the outliers and misjudged spindles were first eliminated. Then clustering processing was conducted to improve the effectiveness of spindle detection and simplify the process.

The number of actual spindles detected by the Wavelet algorithm proposed by Athanasios and Clifford (2015) was less than that detected by experts, so it cannot replace expert detection methods. This study used a sliding window to calculate the corrected moving average of the signal for the threshold setting. This method shortened the detection time and improved the accuracy. The spindles detected by the Martin et al. (2012) window root mean square method contained more false spindles. The hierarchical fusion algorithm clustered the non-coincident spindles automatically detected, improving the stability of spindle detection. The hierarchical fusion algorithm combined the Morlet wavelet and window RMS. It adopted the ideal accuracy of the wavelet method and the ideal recall rate of the window RMS detection method. Thus, the fusion algorithm realized both high levels of precision and recall and could achieve high evaluation indexes.

The HMM&SVM algorithm (Löfhede et al., 2008) also inspired the improvement of the algorithm in this paper. The influence of spindle amplitude on spindle detection was avoided, as the fluctuation range is too large to be ideal for the stability of the

accuracy of the HMM&SVM algorithm. The hierarchical fusion algorithm advanced the experimental data to sample clustering, improving the detection speed through iterative clustering (Ding et al., 2018), and optimized the algorithm before clustering.

The figures reveal that the average Recall rate of the wavelet fusion method was 91.4%. The average Recall rate of the Precision method was 91.4%, which can be improved by 91.4% compared with that of the previous method, and the average Recall rate of the Precision method is improved by 90.4% compared with that of the Precision method. It met the requirements of improving the performance index and stability of the spindles.

Accurate and effective detection of sleep spindles is a methodological challenge. The spindles have a necessary judgment basis for diagnosing human diseases (Manoach et al., 2020). The hierarchical fusion algorithm is a favorable and feasible method for liberating the "gold standard" detection of experts, and reducing the shortcomings of the cumbersome, expensive, and strongly subjective spindle detection methods of the past (Parekh et al., 2017). This method could be popularized for clinical disease diagnosis instead of artificial spindle detection as it improves the speed of disease diagnosis and enables patients to receive rapid treatment (Imtiaz and Rodriguez-Villegas, 2014). At the same time, according to this test, the study of spindles on human intelligence and memory can save substantial experimental time (Wei et al., 2020). Therefore, effective and rapid spindle detection method is a common research direction.

The experiment mentioned in this paper only used the spindle samples of 20 subjects for analysis due to the limitation of the number and age of subjects and limited conditions; as a sample base, this is insufficient. In future research, increasing the sample base will improve the credibility of the results further. In the spindle detection algorithm, the spindle detection used in this paper was based on a single-channel C3, which uses too few channels. We could use dual-channel or multi-channels to detect the spindle automatically in future research. Concurrently, we could combine more deep learning models to classify the spindles and explore the prospects of deep learning in spindle classification.

Data availability statement

The original contributions presented in this study are included in this article/supplementary material, further inquiries can be directed to the corresponding authors upon reasonable request, and according the policies of Tianjin University, Tianjin university of Technology, Tianjin Medical University General Hospital, Xuanwu Hospital, Capital Medical University.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Xuanwu Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

CC and JM completed the ethics files of this experiment. CC and ZH recorded the original experiment data, analyzed the experiment data, and penned the manuscript. CC, LL, DM, and AB wrote parts of the manuscript. FL, WY, PL, and JL designed the experiment and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was financially supported in part by the National Key Research and Development Program of China (2022YFF1202500 and 2022YFF1202501), the National Natural Science Foundation of China (61806146, 62106173, 62006014, and 82101448), the United Arab Emirates University and ASPIRE (AYIA20-002 and 21T057), and the Graduate Research and Innovation Project of Tianjin City (2019YJSS052).

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

Astori, S., and Luthi, A. (2013). Synaptic plasticity at intrathalamic connections via CaV3.3 T-type Ca2+ channels and glun2b-containing nmda receptors. *J. Neurosci.* 33, 624–630. doi: 10.1523/JNEUROSCI.3185-12.2013

Athanasios, T., and Clifford, G. D. (2015). Stage-independent, single lead EEG sleep spindle detection using the continuous wavelet transform and local weighted smoothing. *Front. Hum. Neurosci.* 9:181. doi: 10.3389/fnhum.2015.00181

Chriskos, P., Frantzidis, C. A., Gkivogkli, P. T., Bamidis, P. D., and Kourtidou-Papadeli, C. (2019). "Automatic sleep staging employing convolutional neural networks and cortical connectivity images," in *IEEE transactions on neural networks & learning systems*, (New Yotk, NY: IEEE), 1–11. doi: 10.1109/TNNLS.2019.2899781

Dakun, T., Rui, Z., Jinbo, S., and Qin, W. (2015). Sleep spindle detection using deep learning: A validation study based on crowdsourcing. *Annu. Int.*

Conf. IEEE Eng. Med. Biol. Soc. 37, 2828–2831. doi: 10.1109/EMBC.2015.7318980

Davies, G., Haddock, G., Yung, A. R., Mulligan, L. D., and Kyle, S. D. (2016). A systematic review of the nature and correlates of sleep disturbance in early psychosis. *Sleep Med. Rev.* 31, 25–38. doi: 10.1016/j.smrv.2016.01.001

Dehnavi, F., Moghimi, S., Haghighi, S. S., Safaie, M., and Ghorbani, M. (2019). Opposite effect of motivated forgetting on sleep spindles during stage 2 and slow wave sleep. *Sleep* 42:zsz085. doi: 10.1093/sleep/zsz085

Devuyst, S., Dutoit, T., Didier, J. F., Meers, F., Stanus, E., Stenuit, P., et al. (2006). Automatic sleep spindle detection in patients with sleep disorders. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 1, 3883–3886.

Ding, H., Wang, W., and Califano, A. (2018). IterClust: A statistical framework for iterative clustering analysis. *Bioinformatics* 34, 2865–2866. doi: 10.1093/bioinformatics/bty176

Fogel, S. M., Nader, R., Cote, K. A., and Smith, C. T. (2007). Sleep spindles and learning potential. *Behav. Neurosci.* 121, 1–10.

Fogel, S. M., and Smith, C. T. (2011). The function of the sleep spindle: A physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neurosci. Biobehav. Rev.* 35, 1154–1165. doi: 10.1016/j.neubiorev.2010.12.003

Herrmann, C. S., Strüber, D., Helfrich, R. F., and Engel, A. K. (2016). EEG oscillations: From correlation to causality. *Int. J. Psychophysiol.* 103, 12–21.

Iber, C., Ancoli-Israel, S., Chesson, A., and Quan, S. (2007). *The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications*. Westchester, IL: American Academy of Sleep Medicine.

Imtiaz, S. A., and Rodriguez-Villegas, E. (2014). Evaluating the use of line length for automatic sleep spindle detection. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2014, 5024–5027. doi: 10.1109/EMBC.2014.6944753

Jiang, D., Ma, Y., and Wang, Y. (2021). A robust two-stage sleep spindle detection approach using single-channel EEG. *J. Neural Eng.* 18:026026. doi: 10.1088/1741-2552/abd463

Keshavan, M. S., Montrose, D. M., Miewald, J. M., and Jindal, R. D. (2011). Sleep Correlates of cognition in early course psychotic disorders. *Schizophr. Res.* 131, 231–234.

Klampanos, I. A., Christopher, M., and Prabhakar, R. (2009). Hinrich schütze: Introduction to information retrieval. *Inform. Retriev.* 12, 609–612.

Lacourse, K., Delfrate, J., Beaudry, J., Peppard, P., and Warby, S. C. (2019). A sleep spindle detection algorithm that emulates human expert spindle scoring. *J. Neurosci. Methods* 15, 3–11. doi: 10.1016/j.jneumeth.2018.08.014

Latreille, V., Carrier, J., Lafortune, M., Postuma, R., Bertrand, J., Panisset, M., et al. (2015). Sleep spindles in Parkinson's disease may predict the development of dementia. *Neurobiol. Aging* 36, 1083–1090.

Limoges, E., Mottron, L., Bolduc, C., Berthiaume, C., and Godbout, R. (2005). Atypical sleep architecture and the autism phenotype. *Brain* 128, 1049–1061. doi: 10.1093/brain/awh425

Löfhede, J., Degerman, J., Löfgren, N., Thordstein, M., Flisberg, A., Kjellmer, I., et al. (2008). Comparing a supervised and an unsupervised classification method for burst detection in neonatal EEG. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2008, 3836–3839. doi: 10.1109/IEMBS.2008.4650046

Manoach, D. S., Mylonas, D., and Baxter, B. (2020). Targeting sleep oscillations to improve memory in schizophrenia-sciencedirect. *Biol. Psychiatry* 221, 63–70.

Martin, N., Lafortune, M., Godbout, J., Barakat, M., Robillard, R., Poirier, G., et al. (2012). Topography of age-related changes in sleep spindle. *Neurobiol. Aging* 34, 468–476.

Mensen, A., Poryazova, R., Huber, R., and Bassetti, C. L. (2018). Individual spindle detection and analysis in high-density recordings across the night and in thalamic stroke. *Sci. Rep.* 8:17885. doi: 10.1038/s41598-018-36327-x

Mporas, I., Korvesis, P., Zacharaki, E. I., et al. (2013). "Sleep Spindle Detection in EEG Signals Combining HMMs and SVMs," in *Engineering applications of neural networks. EANN 2013. Communications in computer and information science*, eds L. Iliadis, H. Papadopoulos, and C. Jayne (Berlin: Springer), 40–49.

Parekh, A., Selesnick, I. W., Osorio, R. S., Varga, A. W., Rapoport, D. M., and Ayappa, I. (2017). Multichannel sleep spindle detection using sparse low-rank optimization. *J. Neurosci. Methods* 1, 1–16. doi: 10.1016/j.jneumeth.2017.06.004

Schilling, C., Gappa, L., Schredl, M., Streit, F., Treutlein, J., Frank, J., et al. (2018). Fast sleep spindle density is associated with rs4680 (Val108/158Met) genotype of catechol-O-methyltransferase (COMT). *Sleep* 41. doi: 10.1093/sleep/zsy007

Spironelli, C., Maffei, A., Romeo, Z., Piazzon, G., Padovan, G., Magnolfi, G., et al. (2020). Evidence of language-related left hypofrontality in major depression: An EEG beta band study. *Sci. Rep.* 10:8166. doi: 10.1038/s41598-020-65168-w

Ujma, P. P., Konrad, B. N., Genzel, L., Bleifuss, A., Simor, P., Pótári, A., et al. (2014). Sleep spindles and intelligence: Evidence for a sexual dimorphism. *J. Neurosci.* 34, 16358–16368. doi: 10.1523/JNEUROSCI.1857-14.2014

Wamsley, E. J., Tucker, M. A., Shinn, A. K., Ono, K. E., McKinley, S. K., Ely, A. V., et al. (2012). Reduced sleep spindles and spindle coherence in schizophrenia: Mechanisms of impaired memory consolidation. *Biol. Psychiatry* 71, 154–161. doi: 10.1016/j.biopsych.2011.08.008

Warby, S. C., Wendt, S. L., Welinder, P., Munk, E. G., Carrillo, O., Sorensen, H. B., et al. (2014). Sleep-spindle detection: Crowdsourcing and evaluating performance of experts; non-experts and automated methods. *Nat. Methods* 11, 385–392.

Wei, L., Ventura, S., Lowery, M., Ryan, M. A., Mathieson, S., Boylan, G. B., et al. (2020). Random forest-based algorithm for sleep spindle detection in infant EEG. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2020, 58–61.

Wendt, S. L., Welinder, P., Sorensen, H. B., Peppard, P. E., Jennum, P., Perona, P., et al. (2015). Inter-expert and intra-expert reliability in sleep spindle scoring. *Clin. Neurophysiol.* 126, 1548–1556. doi: 10.1016/j.clinph.2014.10.158

Zhang, Y., Quiñones, G. M., and Ferrarelli, F. (2020). Sleep spindle and slow wave abnormalities in schizophrenia and other psychotic disorders: Recent findings and future directions. *Schizophr. Res.* 221, 29–36. doi: 10.1016/j.schres.2019.11.002

Zhao, R., Sun, J., Zhang, X., Wu, H., Liu, P., Yang, X., et al. (2017). Sleep spindle detection based on non-experts: A validation study. *PLoS One* 12:e0177437. doi: 10.1371/journal.pone.0177437

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 09 March 2023
DOI 10.3389/fnins.2023.1136609

OPEN ACCESS

EDITED BY
Peng Xu,
University of Electronic Science and Technology of China, China

REVIEWED BY
Mingyi Chen,
Shanghai Jiao Tong University, China
Xiaohua Wan,
Beijing Institute of Technology, China

*CORRESPONDENCE
Yongqing Zhang
✉ zhangyq@cuit.edu.cn

SPECIALTY SECTION
This article was submitted to Neuroprosthetics, a section of the journal Frontiers in Neuroscience

RECEIVED 03 January 2023
ACCEPTED 07 February 2023
PUBLISHED 09 March 2023

CITATION
Gao D, Tang X, Wan M, Huang G and Zhang Y (2023) EEG driving fatigue detection based on log-Mel spectrogram and convolutional recurrent neural networks.
Front. Neurosci. 17:1136609.
doi: 10.3389/fnins.2023.1136609

EEG driving fatigue detection based on log-Mel spectrogram and convolutional recurrent neural networks

Dongrui Gao[1], Xue Tang[1], Manqing Wan[1], Guo Huang[2] and Yongqing Zhang[1]*

[1]School of Computer Science, Chengdu University of Information Technology, Chengdu, China, [2]School of Electronic Information and Artificial Intelligence, Leshan Normal University, Leshan, China

Driver fatigue detection is one of the essential tools to reduce accidents and improve traffic safety. Its main challenge lies in the problem of how to identify the driver's fatigue state accurately. Existing detection methods include yawning and blinking based on facial expressions and physiological signals. Still, lighting and the environment affect the detection results based on facial expressions. In contrast, the electroencephalographic (EEG) signal is a physiological signal that directly responds to the human mental state, thus reducing the impact on the detection results. This paper proposes a log-Mel spectrogram and Convolution Recurrent Neural Network (CRNN) model based on EEG to implement driver fatigue detection. This structure allows the advantages of the different networks to be exploited to overcome the disadvantages of using them individually. The process is as follows: first, the original EEG signal is subjected to a one-dimensional convolution method to achieve a Short Time Fourier Transform (STFT) and passed through a Mel filter bank to obtain a logarithmic Mel spectrogram, and then the resulting logarithmic Mel spectrogram is fed into a fatigue detection model to complete the fatigue detection task for the EEG signals. The fatigue detection model consists of a 6-layer convolutional neural network (CNN), bi-directional recurrent neural networks (Bi-RNNs), and a classifier. In the modeling phase, spectrogram features are transported to the 6-layer CNN to automatically learn high-level features, thereby extracting temporal features in the bi-directional RNN to obtain spectrogram-temporal information. Finally, the alert or fatigue state is obtained by a classifier consisting of a fully connected layer, a ReLU activation function, and a softmax function. Experiments were conducted on publicly available datasets in this study. The results show that the method can accurately distinguish between alert and fatigue states with high stability. In addition, the performance of four existing methods was compared with the results of the proposed method, all of which showed that the proposed method could achieve the best results so far.

KEYWORDS

driving fatigue detection, EEG, convolutional neural network, recurrent neural network, log-Mel spectrogram

1. Introduction

Mental fatigue is a psychobiological state caused by prolonged and demanding cognitive activity (Van Cutsem et al., 2017). This mental fatigue reduces a driver's ability to concentrate and make decisions, making it impossible to drive effectively. According to data released by China's National Bureau of Statistics, more than 60,000 people will die in traffic accidents nationwide in 2020 alone (Bureau, 2021). Traffic accidents cause great harm and loss to individuals, the country, and society. Therefore, fatigue detection has become an effective means of reducing accidents and improving transport safety.

To date, several widely used indicators have been proposed for detecting driving fatigue, such as monitoring fatigue-related facial expressions, blinking, yawning (Wang Z. et al., 2020), muscle (Zhang et al., 2021a), and measuring fatigue-related physiological variables such as electrooculography (Zheng and Lu, 2017), heart rate variability (Du et al., 2020a), and electroencephalography (Zhang et al., 2020). Of the many fatigue detection indicators, EEG signals are good mental indicators (Zhang Y. et al., 2022). Because the EEG signal is closely related to brain activity (Zhang X. et al., 2022). Therefore, this paper proposes a driving fatigue detection task based on EEG signals. In fatigue detection tasks, the mental state is usually classified into two categories of alertness and fatigue, or three categories of alertness, fatigue, and drowsiness. Several excellent research results have been presented to achieve this goal (Tuncer et al., 2021). However, current research on EEG driving fatigue detection is still at the point where only complex preprocessed EEG signals can be used for fatigue detection with good recognition results. Due to the instability of the EEG signal, it is challenging to obtain good results if these models are applied to different acquisition devices and scenarios. The task of inputting raw signals for fatigue detection is more challenging than preprocessed EEG models. The models should be better trained than the raw signals with the noise and artifacts removed. Unlike existing methods, the proposed approach in this paper focuses on extracting features from the raw signal for the driving fatigue detection task.

Traditional methods to implement EEG driving fatigue detection are mostly shallow models. A shallow model provides reasonable predictive power with minimal complexity. It consists of a few layers and requires limited training data. However, it requires predefined features with discriminative power. Artificial neural networks (ANNs) with a hidden layer and support vector machines (SVM) are well-known shallow models. Of these, ANNs have been widely used in EEG fatigue detection systems (Chai et al., 2014). In the literature (Vuckovic et al., 2002), to predict fatigue status from EEG signal, time series of inter- and intra-hemispheric cross-spectrogram densities of EEG signal are fed as input to an ANN, which then classifies driver status as either fatigue or alertness. In an alternative approach (King et al., 2006), the time domain EEG data is converted into frequency bands, delta, theta, alpha, and beta bands, and the frequency domain data is then fed to the ANN for fatigue detection. The literature (He et al., 2016) integrated multiple indicators of fatigue to build an ANN-based driver fatigue assessment model, where the EEG indicators were labeled as awake or fatigued. The average power spectrogram ratio A(theta + alpha)/beta for the theta, alpha, and beta bands was derived by fast Fourier transform and fed into the ANN for classifying the driver status as awake or fatigued. Alternatively, a Support Vector Machine (SVM) is specifically designed for the two-class problem. SVM has been employed in many fatigue detection systems to classify driver states according to different fatigue levels. The literature (Mu et al., 2017) extracted the spectrogram entropy, approximate entropy, sample entropy, and fuzzy entropy of EEG signals and fused them as feature vectors to feed into the SVM to classify driver fatigue states, and fusing the four feature entropies obtained the best classification results. In the literature (Yeo et al., 2009), the researcher trained the SVM for binary classification of fatigue states, with significant beta wave activity representing the alert EEG signals and a drop in the alpha wave representing the fatigued EEG signals, and the method obtained better classification results. However, the training set for shallow models is usually small, and when the dataset is too large, the model needs help building better classifiers and often relies on hand-extracted unique features. Deep learning (DL) models are widely used in fatigue driving detection to address this problem.

Recent research has shown that deep learning methods have yielded better results (Ed-Doughmi et al., 2020; Zhang et al., 2021b) and are widely used in various fields. Deep learning models incorporate a learned representation of the data rather than a task-specific approach. In contrast to shallow models, deep models can extract features from training data. Convolutional Neural Networks (CNN) (LeCun and Bengio, 1995) was the first deep model used for driver fatigue detection. Among them, the literature (Wu et al., 2021). The proposed 3D EEG signals-based CNN for driving fatigue classification has achieved remarkable success by projecting the 3D brain topology and its corresponding fatigue features into 2D space to form a brain fatigue topography fed into a coupled CNN-LSTM structure for fatigue classification. Wang H. et al. (2020) proposed an attention-based multiscale convolutional neural network combined with a graph convolutional network for driving fatigue classification, using a time-domain convolutional block to learn the features of each channel and a graph convolutional block network to learn the spatial filters. Finally, the features were passed to a classifier consisting of a softmax layer and a fully connected layer for fatigue classification of the EEG signal. Zeng et al. (2018) proposed two convolutional neural network models, EEG-Conv and EEG-Conv-R. The first model is a conventional CNN. The second combines a CNN with a deep-learning residual network for driver fatigue classification. The results show that both network models outperform the classifier based on the support vector machine. The convolutional neural network combined with the residual network has better generalization capability. Du et al. (2020b) developed a TK-type Convolution Recurrent Fuzzy Network (TCRFN). This method uses convolutional neural networks to deal with noise and improve fatigue classification, in which the authors projected the three-dimensional coordinates of electrodes of EEG and fatigue-related frequency bands theta, alpha, pre-beta, post-beta, gamma onto a two-dimensional plane so that the EEG data is converted into a series of images as feature vectors, the literature concludes that convolutional neural networks are effective in reducing the effect of noise on the model TCRFN. However, due to the massive amount of EEG data, the

current methods for extracting EEG signal features may overlook various important information. To solve the above problems, this paper proposes a model that combines spectrogram feature and temporal feature information, which can effectively utilize time-frequency information to improve the accuracy and precision of fatigue detection. Due to the many hyperparameters and complex structures involved, many DL structures often suffer from time-consuming training processes. We adjust the hyperparameters and model structure to achieve better results in terms of both training time and classification results.

The main contributions of this paper are as follows. (1) In the feature extraction stage, the traditional Fourier transform calculation method is abandoned, and a one-dimensional (1D) convolutional neural network is used to implement the short-time Fourier transform, obtaining more discriminative features. (2) Application of the speech signal technique log-Mel spectrogram to EEG signals to extract the time-frequency characteristics of EEG signals. (3) Combining convolutional neural networks and bi-directional recurrent neural networks to learn the high-level features of the signal, the implementation results show that adding a bi-directional recurrent neural network to the model yields the current optimal results.

The paper is structured as follows. Section 2 describes the related work in this paper. In Section 3, introduce the driving fatigue dataset. Section 4 describes the proposed method and feature extraction. In Section 5, experimental results are given, and the four existing methods are compared with the method proposed in this paper. Finally, Section 6 concludes with a summary and an outlook of the new model in practical driving applications.

2. Related work

This paper's feature extraction of the proposed method mainly involves the Short Time Fourier Transform and the log-Mel spectrogram. Therefore, this section focuses on the fatigue detection research work related to the short-time Fourier transform and log-Mel spectrogram and the proposed method and contributions.

Since EEG signals contain rich brain function information, many research methods can analyze EEG signals with spectrum technology. The literature (Kıymık et al., 2005) applied short-time Fourier transform (STFT) and wavelet transform (WT) to the EEG signal of normal children and children with epileptic seizures. The results showed that STFT has a short processing time and is more suitable for the real-time processing of EEG signals. In the methods (Sparto et al., 2000), the authors used STFT and wavelet transform to process surface EMG signals from the medial, lateral, and latissimus dorsi sites of the erector spinae. The results showed that both methods could detect and quantify fatigue. In another paper (Hajinoroozi et al., 2016), the authors used the fast Fourier transform (FFT) with a Hamming window to obtain frequency features and associated eigenvectors. Numerous advances in brain-computer interface (BCI) technology have demonstrated the feasibility of classification methods based on the short-time Fourier transform. Therefore, in this paper, we will use the short-time Fourier transform to extract the spectrogram features of EEG signals. Still, unlike traditional methods, we use a novel one-dimensional convolutional neural network to implement the Fourier transform work.

Another part of the extraction of features is the extraction of the signal's log-Mel spectrogram. Mel Spectrogram (Li et al., 2001) was first applied to the research of speech recognition. Due to the nonlinearity of the signal and the relationship between the time-frequency domain, the Mel filter bank or Bark filter bank method is usually used, and the extracted spectrogram can be used in speech recognition. The Mel spectrogram can represent the frequency energy independent of the input signal source. Other research areas use log Mel spectrograms to extract features and for deep learning classification tasks. In the literature (Dehzangi and Taherisadr, 2018), the authors proposed a system for detecting distracted drivers based on galvanic skin response (GSR) detection, which converts one-dimensional EEG signals into a two-dimensional spectrogram feature map by extracting the Mel spectrogram of the original GSR as a feature, and obtains good classification results. The literature (Kumar et al., 2021) removes the Mel spectrogram, STFT, and Croma of the emotional speech signal as input to the model. The method experiments with each of the three features. The experimental results show that the proposed model gets the best results using the Mel spectrogram features. In the method (Woo et al., 2022), sleep stage classification based on single-channel EEG signals was studied using the frequency-domain feature extraction method Mel Spectrum. Experiments show that using the Mel spectrogram, the number of input samples will be significantly reduced, the neural network training will be accelerated, and more discriminative features will be obtained. In the literature (Meng et al., 2019), an algorithm based on a 3D log Mel spectrogram is proposed for speech emotion recognition. Experimental results show a 4.58% improvement in recognition accuracy and processing time, showing that log Mel spectrogram maps are practical features for classification tasks. Therefore, we applied the log-Mel spectrogram features to the EEG-based fatigue classification task. The addition of the log-Mel spectrogram improved the accuracy of fatigue detection. The relationship between the learned features and fatigue information could be established through the fatigue classification model once the feature extraction was completed.

3. The proposed model

3.1. Feature extraction

Feature extraction takes the most relevant information from the original data and assigns that information to a lower dimensional space. When the input data is too large and not informative, the data is considered redundant. The input data is then transformed into a simplified representation of the features, also known as a feature vector. This conversion process is called feature extraction. Classification is performed based on the selected features. The classifier's performance depends on the signal's quality and the soundness of the feature selection. As the EEG signal has significant features in the time-frequency domain, to obtain the feature vector set, this paper implements the STFT by using one-dimensional convolution to get the signal's spectrogram, then uses the Mel filter bank to obtain the Mel spectrogram. A spectrogram is a visual representation of a signal's frequency spectrogram as a time

function. The STFT can capture local features accurately, similar to a convolutional filter. STFT is the essential operation for calculating the Mel Spectrogram. To convert the STFT spectrogram to a Mel Spectrogram, the spectrogram is multiplied by the Mel filter bank kernel. We then explain using a 1D convolutional neural network to compute the STFT and extract the Mel spectrogram using a Mel filter bank.

3.1.1. STFT

We implement STFT using a 1D convolutional network, where the convolutional kernel is initialized as the product of a DFT matrix and a window function in a 1D convolutional operation. The DFT matrix is an n*n matrix that the following equation can express.

$$\begin{cases} (F)_{nm} = \omega^{-nm} \\ \omega = e^{2\pi i \frac{1}{N}} \end{cases} \tag{1}$$

Where nm is the matrix of the n columns. In this paper, we apply the Hanning window function, which is calculated in the following form:

$$\omega(n) = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi\ (n-1)}{N}\right)\right], 0 \le n \le N-1 \tag{2}$$

Where n denotes the total length of the window function, and N represents the effective length of the window function. The size of the convolution kernel is equal to the transform size of the STFT, and the step size is the hop length. The EEG signal is an actual signal, so the real and imaginary parts obtained can be seen as the result of the separate action of the real and imaginary parts of the convolution kernel, expressed as:

$$\begin{cases} Y^R = Conv1d\left(x, F^R_{DFT} \odot w, S\right) \\ Y^I = Conv1d\left(x, F^I_{DFT} \odot w, S\right) \end{cases} \tag{3}$$

where x is the EEG input signal, w is the window function, and S is the step size. A convolution operation is synthesized to obtain the signal after STFT:

$$STFT(x) = \left[Y^R; Y^I\right] = Conv1d\left(x, \left[F^R_{DFT}; F^I_{DFT}\right]^T \odot w, S\right) \tag{4}$$

At this point, the returned STFT is in the form of a full FFT. For general feature extraction, only the first $N/2+1$ part of the DFT matrix is taken when initializing the convolution kernel. To obtain the spectrogram, the above equation also needs to be squared.

$$Spectrogram(x) = \left[Y^R\right]^2 + \left[Y^I\right]^2 \tag{5}$$

There are two main advantages of using a short-time Fourier transform based on a one-dimensional convolutional neural network. First, it supports batch processing. Using a neural network-based framework, we can enable tensor operations to convert a tensor of EEG signal segments into a tensor of spectrograms. Secondly, it is trainable. We will discuss how the prediction accuracy of the model can be improved by adjusting the number of convolution kernels of the one-dimensional convolution.

3.1.2. Log-Mel spectrogram

In order to extract the unique perceptual features from the EEG signal, this work further extracts the signal's Mel spectrogram based on the spectrogram for use in a deep learning system for fatigue detection. The traditional frequency-to-Mel scale conversion is in D. O'shaughnessy's book (Douglas, 2000), as follows.

$$Mel(f) = 2595\log_{10}\left(1 + \frac{f}{700}\right) \tag{6}$$

Once the conversion from frequency to Mel scale has been achieved, we can create the Mel filter bank. The equation to implement the mth filter of the filter bank is expressed as:

$$\begin{cases} 0, k < f_{m-1} and k > f_{m+1} \\ \frac{k-f_{m-1}}{f_m - f_{m-1}}, f_{m-1} \le k \le f_m \\ \frac{f_{m+1}-k}{f_{m+1}-f_m}, f_m \le k \le f_{m+1} \end{cases} \tag{7}$$

Where f is the Mel-scale frequency, and m is the total number of filters in the band. These filter banks were multiplied with the spectrograms of the results obtained by STFT above to obtain the Mel-scale spectrograms, as follows.

$$MS(x) = Spectrogram(x) \odot B_m(k) \tag{8}$$

We also need to convert the power spectrogram to dB units by performing a logarithmic operation to obtain a log-Mel spectrogram. The purpose of taking the logarithm is to have the low-amplitude components pulled higher relative to the high-amplitude components to observe periodic signals masked by low-amplitude noise, as shown in the logarithm below:

$$LogMelSpec = 10 * \log_{10}(MS(x)) - 10 * \log_{10}(ref) \tag{9}$$

where ref is the reference value by which the amplitude $MS(x)$ is scaled relative to ref. The extracted log-Mel spectrogram features are fed into the model for the next step of training. The details of the model proposed in this paper are presented below.

3.2. Model description

CNN's have been widely used in EEG-based classification, such as driving fatigue detection (Wang H. et al., 2020; Wu et al., 2021). A convolutional neural network consists of several blocks, with one convolutional block composed of several layers of convolutional layers. High-level local features are obtained from the input feature vector. The log-Mel spectrogram features extracted above are used as input to the CNN.

3.2.1. CNNs

The six-layer CNN proposed in this paper consists of three convolutional blocks, derived from a VGG-like CNN (Simonyan and Zisserman, 2015) perception. Each convolutional block consists of two convolutional layers with a kernel size of 3 × 3, a step size of 1 × 1, and a padding of 1 × 1. Batch normalization is applied between each convolutional layer (Ioffe and Szegedy, 2015).

The mathematical definition of continuous convolution is given in Equation (12) and, in the discrete case, in Equation (13).

$$f * g(n) = \int_{-\infty}^{+\infty} f(\tau)\, g(n-\tau)\, d\tau \tag{10}$$

$$f * g(n) = \sum_{\tau=-\infty}^{+\infty} f(\tau)\, g(n-\tau) \tag{11}$$

where f and g do convolution operations. Suppose the batch normalization layer is not used. In that case, it will easily lead to the model's training to reach the activation function's gradient saturation zone. This is because when the network reaches a certain depth and has a certain complexity, the accumulated changes in the underlying network will affect the upper network, and the normalization operation can make the input data of the activation function fall in the gradient non-saturation zone and alleviate the problem of gradient disappearance. Finally, a $2 * 2$ averaging pooling layer is used for downsampling, and the literature (Kong et al., 2019) demonstrates that $2 * 2$ averaging pooling is a better choice than $2 * 2$ max pooling. A dropout is applied between each convolutional block (Hinton et al., 2016) regularization technique, which is mainly used to avoid complex mutual adaptation on the training data to combat overfitting in the network, with probability p set to 0.2, i.e., we drop neurons in the convolutional layer with probability 0.2.

The signal of 17 channels will be applied with a sliding window size of 1,600 to segment the data. After a small amount of trial and error, the patch size in training is finally set to 200. Each segment will get a sequence of eight small frames. Each frame will be input to the CNNs after extracting the above log-Mel spectrogram features. The next spliced output feature vector will be the extracted high-level features X. X will be used as the input to the recurrent neural network. The convolutional block model is shown in Figure 1.

3.2.2. LogMel-CRNN

In this work of extracting temporal features, we choose a recurrent neural network that captures temporal relationships, which is an end-to-end model that is good at processing time series, such as the EEG signal or speech signal we use, that is, a sample where the preceding and following inputs are correlated. RNNs can be used to extract high-level features in the temporal domain. Schuster and Paliwal (1997) proposed a bi-directional RNN, which uses information from both ends of the sequence to estimate the output since the current value of the sequence depends not only on the information of the previous sequence but also on the sequence at the future moment so that the RNN structure can capture more of the long-term dependence of the sequence and obtain more information about the sequence, which is beneficial to the classification results. In this study, we constructed a two-layer bi-directional RNN with 128 hidden cell counts to model features in the time domain, with the aim of exploring the intrinsic relationships between consecutive time sequences. The values of forward and backward propagation determine the output values of a bi-directional RNN. A deep bi-directional RNN is shown below, denoted by i denoting the ith layer.

$$\begin{cases} s_t^{(1)} = f\left(U^{(1)} x_t + W^1 s_{t-1}^{(1)} + b^{(1)}\right) \\ s_t^{'(1)} = f\left(U^{'(1)} x_t + W^{'(1)} s_{t-1}^{'(1)} + b^{'(1)}\right) \\ \cdots \\ s_t^{(i)} = f\left(U^{(i)} s_t^{(i-1)} + W^i s_{t-1}^{(i)} + b^{(i)}\right) \\ s_t^{'(i)} = f\left(U^{'(i)} s_t^{'(i-1)} + W^{'(i)} s_{t+1}^{'(i)} + b^{'(i)}\right) \\ o_t = g\left(V^{(i)} s_t^{(i)} + V^{'(i)} s_t^{'(i)} + c^{(i)}\right) \end{cases} \tag{12}$$

Of these, the U, V, W and U', V', W' are both weight matrices, and b, b', c is the weight vector. s_t indicates that a forward calculation is being performed, and $s_t^{'}$ prime is the inverse calculation, and o_t is the value of the output layer. For deep bi-directional RNNs, the more layers, the better the learning ability, but this requires more training data. In summary, the features extracted by the convolutional neural network are fed into the bi-directional RNN to obtain time-frequency domain features, which then enter the classifier, which consists of two fully connected layers and softmax functions, with the final predictions defined as follows.

$$Prob = Softmax\left(W_s \times F + b_s\right), Prob \in R^z \tag{13}$$

Where $W_s \in R^{z \times l}$ is the weight matrix, and $b_s \in R^k$ is the bias vector, l denotes the size of the fully connected layer, z is the number of classifications in the model, and k represents the dimension of the feature F.

Overall, in this paper, the extracted frequency domain features are fed into a six-layer CNN to extract high-level features, which are then provided into a two-layer bi-directional RNN. The resulting temporal features are finally fed into a classifier consisting of a fully connected layer, a ReLU activation function, and a softmax function to produce the final classification prediction results. The schematic diagram of the proposed model, LogMel-CRNN, is shown in Figure 2.

4. Experiments and analysis of experimental results

4.1. Datasets

Experimental paradigm: This study used the SEED-VIG (Zheng and Lu, 2017) dataset from the Human Brain Computing and Machine Intelligence Commons at Shanghai Jiao Tong University. The experiments were based on EEG signals collected by a virtual reality driving simulation system, in which each participant was tested for ~2 h. Collect EEG signals: 21 EEG data sets were collected from 23 volunteer participants. Twelve channels of EEG signals were recorded from posterior sites (CP1, CPZ, CP2, P1, PZ, P2, PO3, POZ, PO4, O1, OZ, and O2), and six channels of electrical signals were recorded from temporal sites (FT7, FT8, T7, T8, TP7, and TP8). One of the electrodes, CPZ, was the reference electrode. Raw EEG signals were acquired from 18 channels at a sampling rate of 1,000 Hz. Labeling: The SEED-VIG dataset includes both EEG and EEG signals. Where the labels were obtained using SensoMotoric Instruments (SMI) eye-tracking glasses, the

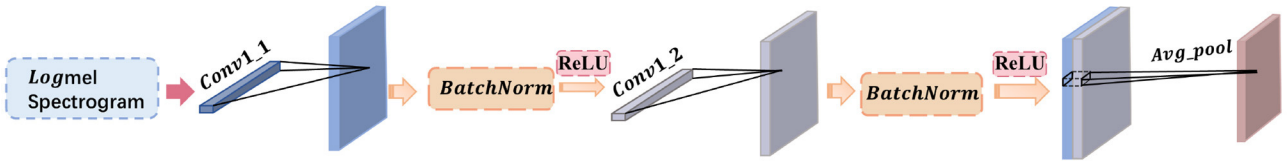


FIGURE 1
Schematic diagram of the convolutional block model.

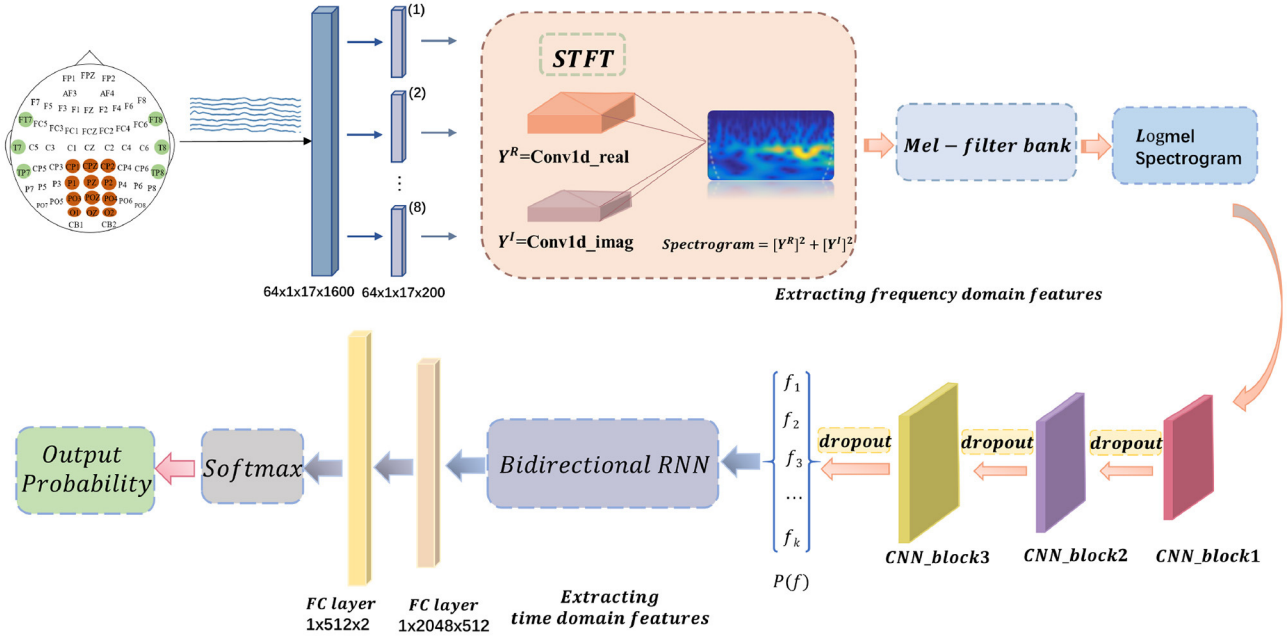


FIGURE 2
Schematic diagram of the whole model of the proposed LogMel-CRNN model.

data were labeled as the percentage of eyelids closed over the pupil over some time, PERCLOS (Gao et al., 2015), which was calculated as follows:

$$PERCLOS = \frac{blink + CLOS}{interval} \tag{14}$$

$$interval = blink + fixation + saccade + CLOS \tag{15}$$

Among them, *blink* is the blink time, fixation is the gaze time, a saccade is the saccade time, and *CLOS* is the closed eye time. PERCLOS values were classified into three categories: wakefulness, fatigue, and somnolence, through thresholds of 0.35 and 0.7. The smaller the PERCLOS value, the higher the driver's alertness. In this paper, two groups of experiments are done. One is that the binary classification is divided into alertness and fatigue state through the threshold of 0.35; the second is that the multi-class category is divided into the state of alertness, fatigue, and sleepiness through the thresholds of 0.35 and 0.7.

4.2. Experimental settings

During the initial training of the experiments, the Xavier normal distribution was used to initialize the network weights. LogMel-CRNN used the Adam optimizer and a learning rate set to 1.0e-4. To perform a comprehensive evaluation of the model, the leave-out validation method was used to obtain the model with the highest accuracy. After disrupting the dataset, the ratio of the training set to the test set was 8:2. Considering the training time of the network, the batch size was set to 64 when training the network, and 50 epochs were required. The cross-entropy function was the model's loss function and was calculated as follows.

$$CE = \sum_{k=1}^{K} -P_k \log\left(p_k\right) \tag{16}$$

Where P is the true distribution, which is a one-hot vector of length K a one-hot vector of length $P_k \in (0, 1)$, p is the predictive distribution.

We applied four standard metrics to measure the performance of our model from distinct perspectives: Accuracy indicates the precision of the prediction results; Precision suggests the

probability of correctly predicting a positive sample among the samples predicted to be positive; Recall suggests the probability of being correctly predicted as the last positive sample among the positive samples of the original sample; F1- Score is the summed average of precision and recall.

4.3. Model ablation

To study the utility of each component in our model, we decomposed our proposed LogMel-CRNN model with a set of variants, including the following:

1. Model 1: Fourier transforms and comparing Fourier transform types

   In order to explore the effect of frequency domain information on feature extraction, in the model LogMel-CRNN basis, the model with frequency domain information obtained by the Fourier transform is compared with the model without FT, and the results show that the performance indicators of the former results are all greater than those of the latter. And the comparison will be made with the extraction of frequency domain information using the conventional STFT, which performs less well than the STFT using one-dimensional convolution, which is consistent with the previous (Kumar et al., 2021) experimental results of extracting frequency domain information to obtain better classification results that are consistent. Figure 3 shows the results of the experimental comparison of the average evaluation metric with and without the Fourier transform and the experimental classification evaluation results using the conventional short-time Fourier transform. In Figures 3–6, the abscissa is the four classification indicators, the ordinate is the evaluation result value, and the results are averaged. The graph shows that the model using the Fourier transform implemented using one-dimensional convolution achieves the highest accuracy, with the remaining three indicators being higher than the other two experiments.
2. Model 2: A Mel filter bank

   The Mel filter bank converts frequencies to a Mel scale, which results in a Mel spectrogram. The Mel scale relates the perceived frequency of the original signal to the actual measured frequency. Humans are much better at recognizing small pitch changes at low frequencies than at high frequencies (Wu and Cao, 2005). So the Mel scale makes the features much closer to what humans perceive. In this model, we verified that extracting the log Mel spectrogram plays a positive role in the model. The model has better robustness when it is used compared to the model without the Mel filter bank. Figure 4 shows the results of the experimental classification evaluation of the model with and without the Mel filter bank. The classification results are much higher for the experiments with the model with the addition of the Mel filter set than for the experiments without it. In particular, the F1 score is even higher than 34.13%. This shows that the Mel filter bank is able to extract significant fatigue features.
3. Model 3: Bi-directional RNN

   This experiment aims to determine whether the time domain contains valuable information. In this study, a bi-directional RNN was used to extract temporal features. The RNN has two hidden layers, each with a hidden cell size of 128. We find a relationship between temporal order and fatigue classification as time changes. The experimental results show that using RNNs to extract features and then feeding them into the classifier achieves better classification results. Figure 5 shows a plot of the evaluation results of the experimental results of the model with and without the RNN. It can be seen that adding the RNN module resulted in a 4.42% increase in fatigue classification accuracy precisely because the RNN can provide learning training using temporal information to improve model performance.
4. Model 4: Increased attention mechanism

   In this study, we also experimented with the use of attention mechanisms after extracting temporal features. The attention mechanism in deep learning borrows from the four-dimensional approach to human attention and has been used in a variety of different types of deep learning tasks, such as natural language, image classification, and speech recognition, with significant results. However, the experimental results did not show satisfactory results, so the attention mechanism module was not used in the model. Figure 6 shows the experimental evaluation results of adding the attention mechanism to the model versus not employing it. As seen in the table, the classification results with the attention mechanism will be slightly worse than the model without it. Therefore, the final LogMel-CRNN model did not adopt the attention mechanism.

4.4. Influence of important parameters

In this section, we studied the influence of five parts: (1) The number of STFT-1D convolutional kernel size. (2) The size of the STFT-Hop length of our model. (3) The number of Mel filter banks. (4) The sampling rate of Mel filter banks. (5) The patch size of CNN model.

1. STFT-1D convolutional kernel size

   In the STFT calculation, this is achieved by one-dimensional convolution, in which the size of the convolution kernel is equal to the size of the FT (n_fft). The experiments are performed by adjusting the n_fft parameter to change the perceptual field during the convolution. Therefore, this parameter directly changes the output of the frequency spectrogram and is an essential parameter for this experiment. Table 1 shows the classification evaluation indicators after four parameter adjustments. The first column of the table shows the size of the set convolution kernel parameters. The first row shows the four average evaluation indicators. Each row indicates the average evaluation of the classification results at that convolution kernel size. Obviously, in the four adjustment parameters, when the size of the convolution kernel, that is, the size of the FT,

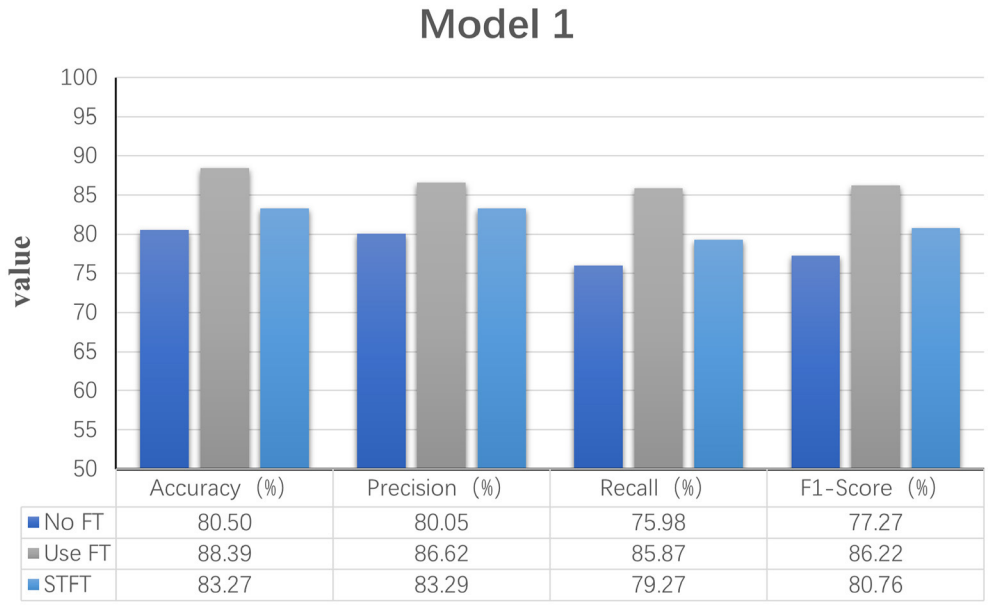


FIGURE 3
Statistical graphs of the evaluation results of the model with the addition of FT vs. without FT.

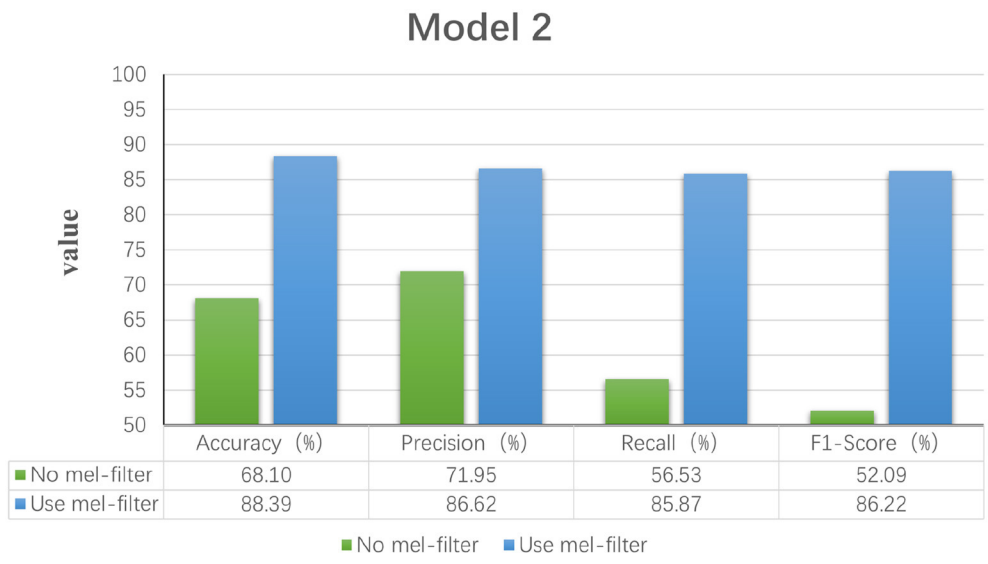


FIGURE 4
Statistical graphs of the evaluation results of the model with the addition of Mel filter vs. without.

is 50, the model achieves better results, with an accuracy rate of 88.39%. When the convolution kernel size is 100, the accuracy rate also reaches 87.34%. The table shows that the method of implementing the short-time Fourier transform using one-dimensional convolution makes the short-time Fourier transform trainable.

2. STFT-Hop length

Hop length is the second important parameter in the STFT calculation. It is the distance that the window moves, also known as the frameshift. The window size for this experiment was set to 10 samples, and the window was changed to overlap or not by adjusting the window shift distance. When the Hop length is five samples, which is equal to one-half of the window size, there is an overlap of five frames. Table 2 shows the average evaluation indicators obtained after adjusting the hop length four times. The first column of the table shows the set hop length, the first row shows the four average evaluation indicators, and each of the remaining rows shows the results of the evaluation indicators for that hop length size. The table shows that the best

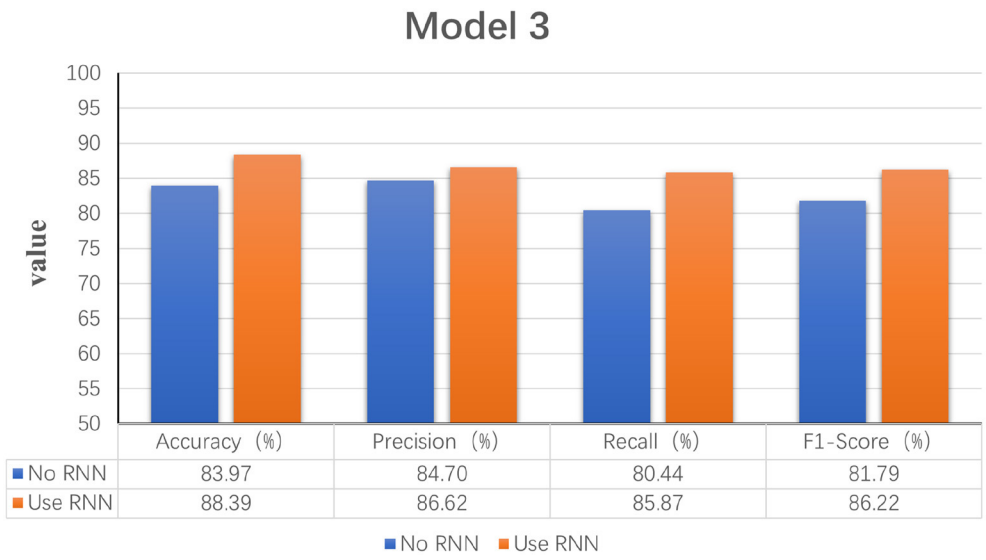


FIGURE 5
Statistical graphs of the evaluation results of the model with the addition of RNN vs. without RNN.

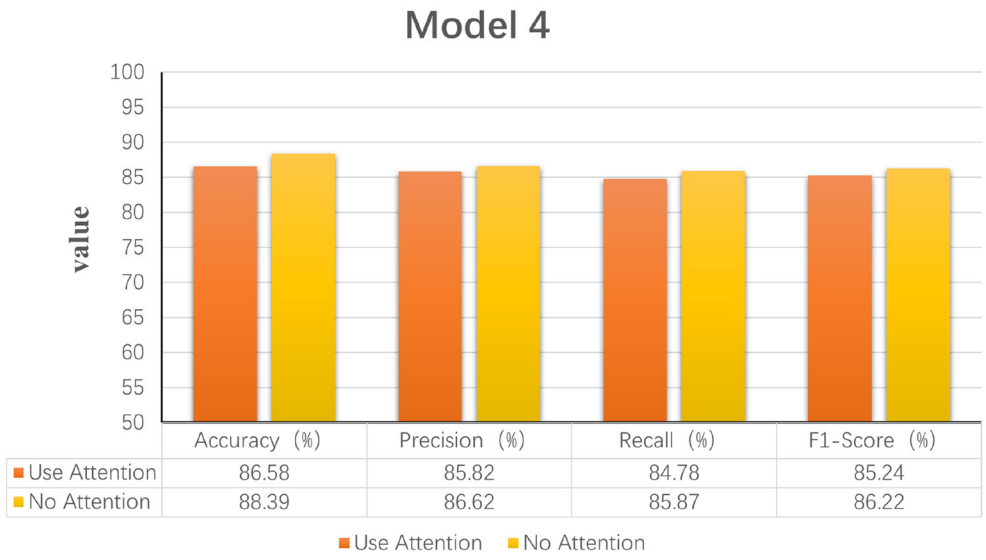


FIGURE 6
Statistical graphs of the evaluation results of the model with the addition of attention mechanism vs. without.

classification accuracy was obtained when the Hop length size was 1, while the F1 score got the best result of 86.37% when the Hop length was 5.

3. Mel filter banks - number of filters

The final step in calculating the filter banks is to apply the triangular filter on the Mel scale to the power spectrogram to extract the frequency bands, mapping n_fft into the mel bin. According to Equation (9), after the filter bank has been created, the first filter bank will start at the first point, peak at the second point, and then return to 0 at the third point. The second filter bank will begin with the second point, reach a maximum at the third point, then be 0 at the fourth point, and so on. To obtain a Mel spectrogram plot with 12 mel bins, we would need 12 Mel filter banks. Table 3 shows the average evaluation indicators for the three different numbers of Mel filter banks set up for this experiment. The first column of the table shows the number of Mel filters set up, and the first row shows the four different evaluation indicators. It is clear from the table that the four evaluation indicators for fatigue classification are much higher than the other two groups when the number of Mel filters is 12. This shows that the number of Mel filters is an essential parameter for the model's training.

TABLE 1 Average performance results based on differences in the size of the convolution kernel for 1D convolution in the Fourier transform.

Kernel size	Accuracy %	Precision %	Recall %	F1 score %
10	85.91	84.66	84.86	84.73
20	86.99	86.73	84.29	85.28
50	**88.39**	**86.62**	**85.87**	**86.22**
100	87.34	86.13	86.05	86.07

In parameter experiments, bold values indicate the best classification results are obtained for the parameters at their current values.

TABLE 2 Table of average performance results generated by adjusting the parameter Hop Length model in the Fourier transform.

Hop-length	Accuracy %	Precision %	Recall %	F1 score %
1	**88.39**	**86.62**	**85.87**	**86.22**
5	88.47	86.49	86.26	86.37
10	86.66	85.69	85.38	85.53
20	86.56	85.67	84.89	85.25

In parameter experiments, bold values indicate the best classification results are obtained for the parameters at their current values.

TABLE 3 Table of average performance results generated by adjusting the parametric Mel filter number model.

N-filter	Accuracy %	Precision %	Recall %	F1 score %
6	82.44	82.47	80.10	80.49
12	**88.39**	**86.62**	**85.87**	**86.22**
24	82.98	82.55	79.66	80.71

In parameter experiments, bold values indicate the best classification results are obtained for the parameters at their current values.

TABLE 4 Table of average performance results generated by the sample rate model for adjusting the parametric Mel filter.

Sampling rate	Accuracy %	Precision %	Recall %	F1 score %
100	85.45	84.50	83.87	84.17
200	85.91	84.66	84.86	84.73
300	85.91	84.82	84.62	84.72
400	**88.39**	**86.62**	**85.87**	**86.22**
800	85.80	84.82	84.55	84.68

In parameter experiments, bold values indicate the best classification results are obtained for the parameters at their current values.

TABLE 5 Average performance results generated by patch size model in tuned parameter CNN.

Patch size	Accuracy %	Precision %	Recall %	F1 score %
100	84.02	83.72	83.10	83.39
200	**88.39**	**86.62**	**85.87**	**86.22**
400	86.67	84.70	84.33	84.54

In parameter experiments, bold values indicate the best classification results are obtained for the parameters at their current values.

4. Mel filter banks - sampling rate

The sampling rate parameter of the Mel filter bank, i.e., the sampling rate of the input signal. The sampling rate of the original EEG signal was 1,000 Hz, and the best input sampling rate parameter was obtained by five experiments of adjusting the input signal's sampling rate. Table 4 shows the results of the average evaluation metric for the five experiments. The first column indicates the size of the set sampling rate, and the first row shows the four different evaluation indicators. As can be seen from the table, the four indicators for fatigue classification are higher than the other four experiments when the input signal has a sampling rate of 400 Hz.

5. CNN-Patch size

Patch size: Previous research has shown that using a larger patch size in a CNN can result in better classification accuracy because the CNN can capture more contextual information to make decisions (Farabet et al., 2012; Li et al., 2014). This is because CNNs can capture more contextual information to make decisions. Therefore, testing a larger network and increasing the patch size also requires changes to the network, as some layers (e.g., fully connected layers) are, by definition, required to have a fixed-size input. So, a patch size change also implies a layered architecture change. Table 5 shows the results of the average evaluation metric for three experiments with the

TABLE 6 Table of optimal experimental parameters for the LogMel-CRNN model proposed in this paper.

Experiment parameter	
FT_kenel size	50
FT_Hop-Length	1
Mel-N-filter	12
Mel Sampling rate	400
CNN Patch size	200

TABLE 7 Four assessment indicators for classifying EEG signals into alert and fatigue states.

Accuracy %	Precision %	Recall %	F1 Score %
Two-class: Vigilance or fatigue			
88.39	86.62	85.87	86.22

TABLE 8 Four assessment indicators for classifying EEG signals into three states of alertness, fatigue, and drowsiness.

Accuracy %	Precision %	Recall %	F1 Score %
Three-class: Vigilance, fatigue, or drowsiness			
81.30	81.67	81.97	81.80

Patch size parameter adjusted. The first column of the table sets the size of the patch size and the first row shows the four average evaluation indicators for fatigue classification. From this table, the best precision value of 86.62% was obtained for a patch size of 200 in the convolutional neural network in this experiment. The performance of the model was also improved and the evaluation indicators of the model were optimized.

After five parameter tuning experiments, the best parameter combinations for this study are summarized in Table 6, where the left-hand side of the table indicates the names of the different parameters, and the right-hand side shows the optimal values for each parameter. Table 7 provides the results of the average fatigue assessment metric for the two-classification hybrid experiment of LogMel-CRNN, with an accuracy of 88.39%, a precision of 86.62%, recall of 85.87%, and an f1 score of 86.22% for the two-classification. Table 8 shows the results of the average assessment indicators for fatigue for the three-classification mixed experiment of LogMel-CRNN. As seen in this table, accuracy is 81.30%, precision is 81.67%, recall is 81.97%, and f1 score is 81.80%. The assessment indicators for the three-classification experiment were much lower than those for the two-classification because the EEG signals for fatigue and drowsiness in the three-classification were very similar. The model needed help to distinguish the significant difference between the two.

4.5. Comparison with existing methods

This section mainly discusses the differences between the proposed and related fatigue detection methods. Here, we have selected four existing methods for comparison: (1) LSTM (Hefron et al., 2017); (2) ESTCNN (Gao et al., 2019); (3) EEG_Conv (Zeng et al., 2018); (4) EEG_Conv_R (Zeng et al., 2018).

LSTM: In the literature (Hefron et al., 2017), a long short-term memory network, a particular type of RNN model, is proposed. The network considers the temporal correlation between improving features' smoothness and solving the problem of gradient disappearance and gradient explosion during backpropagation in simple recurrent networks. LSTM provides algorithms with fine-grained control over what is put into and removed from memory in hidden layers called memory cells. It does this through a combination of three gates: an input gate, a forgotten gate, and an output gate. The forgot gate determines when inputs are remembered or ignored in the hidden state through a dedicated mechanism. The network is adequate for the task of "long-term memory." Driving fatigue detection is a type of long-term memory task. We use this method to compare with the one proposed in this paper, but also because it introduces a lot of content, resulting in more parameters, which makes training much more difficult.

ESTCNN: In this research (Gao et al., 2019) proposed a Spatio-temporal convolutional neural network (ESTCNN) based on EEG signals. The network introduces a core block to extract temporal correlations from the EEG signal, which consists of three convolutional blocks and a pooling layer. The combination of the core block and the dense layer is then used to learn valid information related to fatigue. The model is a 14-layer network consisting of three core blocks, two dense layers, and a softmax layer, respectively. The results show that the Spatio-temporal structure of the framework offers advantages in terms of computational efficiency and reference time.

EEG_Conv: In the literature (Zeng et al., 2018) the authors proposed two models for predicting the mental state of drivers in EEG signals. EEG data were collected on a driving simulation platform constructed themselves and applied to models EEG-Conv and EEG-Conv-R, respectively. Where EEG-Conv has a total of eight layers, consisting of an input layer, three convolutional layers, a pooling layer, a local response normalization layer, a fully connected layer, and an output layer. In the paper, the prediction performance of the proposed classifier is investigated for both within-subject and between-subject EEG data. Inter-subject prediction refers to training and test data from the same subject, while inter-subject prediction refers to training and test data from different subjects. The results show that the proposed method has better generality for detecting mental states from various subjects.

EEG_Conv_R: In the literature (Zeng et al., 2018), in order to further improve the classification accuracy based on the model EEG-Conv, researchers have combined EEG-Conv with residual learning and proposed the EEG-Conv-R model. Residual learning explicitly re-represents layers as learning residual functions concerning the layer inputs rather than learning unreferenced functions. In other words, the residual layer knows the changes in the perturbations. From the experimental results, EEG-Conv-R converges faster than EEG-Conv and takes less time to extract features in the training phase.

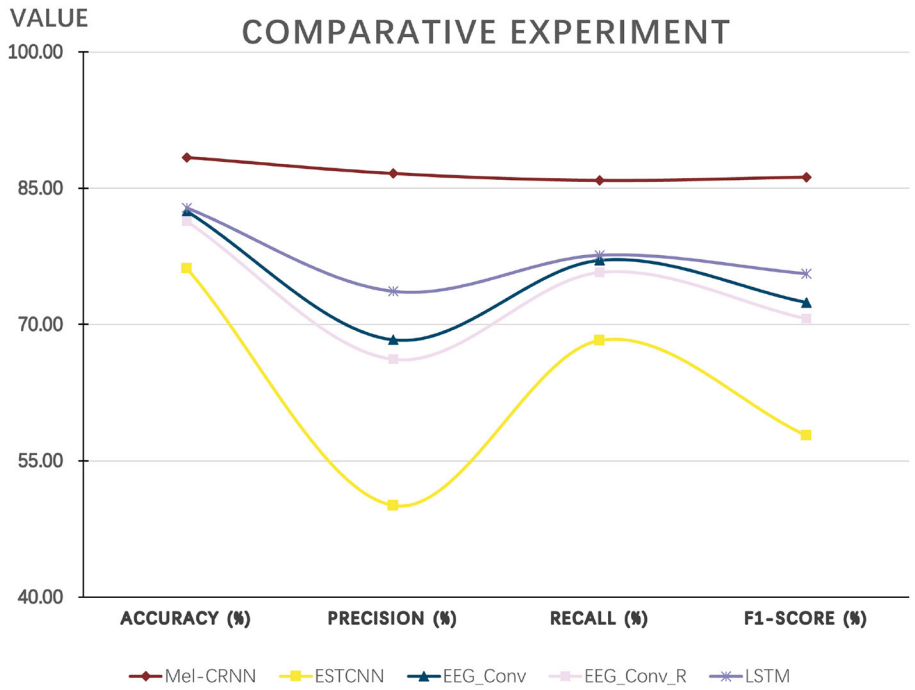


FIGURE 7
Experimental results compared with existing methods.

In Figure 7, the abscissa represents the four indicators of fatigue predicted by the model, and the ordinate represents the average value of each indicator. As seen in Figure 7, the method proposed in this paper outperforms the other four existing methods in all four indicators compared with the current four methods.

5. Conclusion

This paper proposes using Convolutional Recurrent Neural Network based on Log Mel Spectrogram (LogMel-CRNN) for driver EEG fatigue detection. In particular, we use one-dimensional convolution to compute STFT, which improves the accuracy of extracting features by 5.12% compared to the traditional STFT method. To build LogMel-CRNN, we investigated four model structures and obtained different recognition results for each of them, where the STFT of extracted EEG signals and log Mel spectrograms as features input to the model got the highest accuracy of 88.39%, the accuracy of 86.62%, recall of 85.87%, and F1 score of 86.22%, thereby the EEG-based fatigue detection model was thus finalized. The use of log-Mel spectrogram features was shown to improve classification accuracy and model performance in the tuning structure experiments. We demonstrate that the model outperforms several previous state-of-the-art methods through comparative experiments. In practical driving applications, lightweight EEG signal acquisition methods are an essential issue, as is noise handling for real-time fatigue detection. In the future, we will further optimize LogMel-CRNN to obtain better detection results for fatigue triple classification, to be able to implement the algorithm in practical applications as well as to extend it to more recognition tasks.

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/s.

Author contributions

DG: conceptualization, methodology, and writing. XT: methodology and writing. MW: methodology and data curation. GH: validation. YZ: supervision, conceptualization, and writing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China under Grant No. 62272067, the Sichuan Science and Technology Program under Grant Nos. 2023NSFSC0499 and 2023YFG0018, the LOST 2030 Brain Project No. 2022ZD0208500, the Scientific Research Foundation of Chengdu University of Information Technology under Grant Nos. KYQN202208 and KYQN202206, and the 2011 Collaborative Innovation Center for Image and Geospatial Information of Sichuan Province.

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

References

Bureau, C. S. (2021). *China Statistical Yearbook 2021. National Bureau of Statistics of China*. Available online at: http://www.stats.gov.cn/tjsj/ndsj/2021/indexch.htm

Chai, R., Tran, Y., Craig, A., Ling, S. H., and Nguyen, H. T. (2014). "Enhancing accuracy of mental fatigue classification using advanced computational intelligence in an electroencephalography system," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Chicago, IL), 1318–1341.

Dehzangi, O., and Taherisadr, M. (2018). "Driver distraction detection using Mel cepstrum representation of galvanic skin responses and convolutional neural networks," in *2018 24th International Conference on Pattern Recognition (ICPR)* (Beijing), 1481–1486. doi: 10.1109/ICPR.2018.8545082

Douglas, O. (2000). *Speech Communications: Human and Machine*. Available online at: http://ieeexplore.ieee.org/document/5312112

Du, G., Li, T., Li, C., Liu, P. X., and Li, D. (2020a). Vision-based fatigue driving recognition method integrating heart rate and facial features. *IEEE Trans. Intell. Trans. Syst.* 22, 3089–31007. doi: 10.1109/TITS.2020.2979527

Du, G., Wang, Z., Li, C., and Liu, P. X. (2020b). A tsk-type convolutional recurrent fuzzy network for predicting driving fatigue. *IEEE Trans. Fuzzy Syst.* 29, 2100–2111. doi: 10.1109/TFUZZ.2020.2992856

Ed-Doughmi, Y., Idrissi, N., and Hbali, Y. (2020). Real-time system for driver fatigue detection based on a recurrent neuronal network. *J. Imaging* 6:8. doi: 10.3390/jimaging6030008

Farabet, C., Couprie, C., Najman, L., and LeCun, Y. (2012). Learning hierarchical features for scene labeling. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 1915–1929. doi: 10.1109/TPAMI.2012.231

Gao, X.-Y., Zhang, Y.-F., Zheng, W.-L., and Lu, B.-L. (2015). "Evaluating driving fatigue detection algorithms using eye tracking glasses," in *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)* (Montpellier), 767–770. doi: 10.1109/NER.2015.7146736

Gao, Z., Wang, X., Yang, Y., Mu, C., Cai, Q., Dang, W., et al. (2019). EEG-based spatio-temporal convolutional neural network for driver fatigue evaluation. *IEEE Trans. Neural Netw. Learn. Syst.* 30, 2755–2763. doi: 10.1109/TNNLS.2018.2886414

Hajinoroozi, M., Mao, Z., Jung, T.-P., Lin, C.-T., and Huang, Y. (2016). Eeg-based prediction of driver's cognitive performance by deep convolutional neural network. *Signal Process.* 47, 549–555. doi: 10.1016/j.image.2016.05.018

He, Q., Li, W., Fan, X., and Fei, Z. (2016). Evaluation of driver fatigue with multi-indicators based on artificial neural network. *IET Intell. Transp. Syst.* 10, 555–561. doi: 10.1049/iet-its.2015.0021

Hefron, R. G., Borghetti, B. J., Christensen, J. C., and Kabban, C. M. S. (2017). Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation. *Pattern Recogn. Lett.* 94, 96–104. doi: 10.1016/j.patrec.2017.05.020

Hinton, G. E., Krizhevsky, A., Sutskever, I., and Srivastva, N. (2016). *System and Method for Addressing Overfitting in a Neural Network. U. S. Patent 9,406,017*.

Ioffe, S., and Szegedy, C. (2015). "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning* (Lille), 448–456.

King, L. M., Nguyen, H. T., and Lal, S. (2006). "Early driver fatigue detection from electroencephalography signals using artificial neural networks," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society* (New York, NY), 2187–2190. doi: 10.1109/IEMBS.2006.259231

Kıymık, M. K., Güler, İ., Dizibüyük, A., and Akín, M. (2005). Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application. *Comput. Biol. Med.* 35, 603–616. doi: 10.1016/j.compbiomed.2004.05.001

Kong, Q., Cao, Y., Iqbal, T., Xu, Y., Wang, W., and Plumbley, M. D. (2019). Cross-task learning for audio tagging, sound event detection and spatial localization: dcase 2019 baseline systems. *arXiv preprint arXiv:1904.03476*. doi: 10.48550/arXiv.1904.03476

Kumar, N., Kaushal, R., Agarwal, S., and Singh, Y. B. (2021). "CNN based approach for speech emotion recognition using MFCC, Croma and STFT hand-crafted features," in *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)* (Greater Noida), 981–985. doi: 10.1109/ICAC3N53548.2021.9725750

LeCun, Y., and Bengio, Y. (1995). "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks*, ed M. A. Arbib (MIT Press), 3361.

Li, D., Sethi, I. K., Dimitrova, N., and McGee, T. (2001). Classification of general audio data for content-based retrieval. *Pattern Recogn. Lett.* 22, 533–544. doi: 10.1016/S0167-8655(00)00119-7

Li, H., Zhao, R., and Wang, X. (2014). Highly efficient forward and backward propagation of convolutional neural networks for pixelwise classification. *arXiv preprint arXiv:1412.4526*.

Meng, H., Yan, T., Yuan, F., and Wei, H. (2019). Speech emotion recognition from 3D log-mel spectrograms with deep learning network. *IEEE Access* 7, 125868–125881. doi: 10.1109/ACCESS.2019.2938007

Mu, Z., Hu, J., and Min, J. (2017). Driver fatigue detection system using electroencephalography signals based on combined entropy features. *Appl. Sci.* 7:150. doi: 10.3390/app7020150

Schuster, M., and Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE* 45, 2673–2681. doi: 10.1109/78.650093

Simonyan, K., and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* 7, 1–14.

Sparto, P. J., Parnianpour, M., Barria, E. A., and Jagadeesh, J. M. (2000). Wavelet and short-time Fourier transform analysis of electromyography for detection of back muscle fatigue. *IEEE Trans. Rehabil. Eng.* 8, 433–436. doi: 10.1109/86.867887

Tuncer, T., Dogan, S., and Subasi, A. (2021). Eeg-based driving fatigue detection using multilevel feature extraction and iterative hybrid feature selection. *Biomed. Signal Process. Control* 68:102591. doi: 10.1016/j.bspc.2021.102591

Van Cutsem, J., Marcora, S., De Pauw, K., Bailey, S., Meeusen, R., and Roelands, B. (2017). The effects of mental fatigue on physical performance: a systematic review. *Sports Med.* 47, 1569–1588. doi: 10.1007/s40279-016-0672-0

Vuckovic, A., Radivojevic, V., Chen, A. C., and Popovic, D. (2002). Automatic recognition of alertness and drowsiness from EEG by an artificial neural network. *Med. Eng. Phys.* 24, 349–360. doi: 10.1016/S1350-4533(02)00030-9

Wang, H., Xu, L., Bezerianos, A., Chen, C., and Zhang, Z. (2020). Linking attention-based multiscale CNN with dynamical GCN for driving fatigue detection. *IEEE Trans. Instrum. Measure.* 70, 1–11. doi: 10.1109/TIM.2020.3047502

Wang, Z., Hong, Q., and Wang, X. (2020). A memristive circuit implementation of eyes state detection in fatigue driving based on biological long short-term memory rule. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 2218–2229. doi: 10.1109/TCBB.2020.2974944

Woo, S.-W., Kang, M.-K., Park, B.-J., and Hong, K.-S. (2022). "Sleep stage classification using electroencephalography *via* Mel frequency cepstral coefficients," in *2022 13th Asian Control Conference (ASCC)* (Jeju: IEEE), 42–47. doi: 10.23919/ASCC56756.2022.9828340

Wu, E. Q., Xiong, P., Tang, Z.-R., Li, G.-J., Song, A., and Zhu, L.-M. (2021). Detecting dynamic behavior of brain fatigue through 3-D-CNN-LSTM. *IEEE Trans. Syst. Man Cybern.* 52, 90–100. doi: 10.1109/TSMC.2021.3062715

Wu, Z., and Cao, Z. (2005). Improved mfcc-based feature for robust speaker identification. *Tsinghua Sci. Technol.* 10, 158–161. doi: 10.1016/S1007-0214(05)70048-1

Yeo, M. V., Li, X., Shen, K., and Wilder-Smith, E. P. (2009). Can SVM be used for automatic EEG detection of drowsiness during car driving? *Saf. Sci.* 47, 115–124. doi: 10.1016/j.ssci.2008.01.007

Zeng, H., Yang, C., Dai, G., Qin, F., Zhang, J., and Kong, W. (2018). EEG classification of driver mental states by deep learning. *Cogn. Neurodyn.* 12, 597–606. doi: 10.1007/s11571-018-9496-y

Zhang, X., Lu, D., Pan, J., Shen, J., Wu, M., Hu, X., et al. (2020). Fatigue detection with covariance manifolds of electroencephalography in transportation industryg. *IEEE Trans. Indus. Inform.* 17, 3497–3507. doi: 10.1109/TII.2020.3020694

Zhang, X., Pan, J., Shen, J., Din, Z. U., Li, J., Lu, D., et al. (2022). Fusing of electroencephalogram and eye movement with group sparse canonical correlation analysis for anxiety detection. *IEEE Trans. Affect. Comput.* 13, 958–971. doi: 10.1109/TAFFC.2020.2981440

Zhang, Y., Cao, W., Feng, L., Wang, M., Geng, T., Zhou, J., et al. (2022). SHNN: a single-channel EEG sleep staging model based on

semi-supervised learning. *Expert Syst. Appl.* 183:119288. doi: 10.1016/j.eswa.2022.119288

Zhang, Y., Chen, S., Cao, W., Guo, P., Gao, D., Wang, M., et al. (2021a). MFFNET: multi-dimensional feature fusion network based on attention mechanism for SEMG analysis to detect muscle fatigue. *Expert Syst. Appl.* 185:115639. doi: 10.1016/j.eswa.2021.115639

Zhang, Y., Qiao, S., Zeng, Y., Gao, D., Han, N., and Zhou, J. (2021b). CAE-CNN: predicting transcription factor binding site with convolutional autoencoder and convolutional neural network. *Expert Syst. Appl.* 183:115404. doi: 10.1016/j.eswa.2021.115404

Zheng, W.-L., and Lu, B.-L. (2017). A multimodal approach to estimating vigilance using EEG and forehead EOG. *J. Neural Eng.* 14:026017. doi: 10.1088/1741-2552/aa5a98

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 15 March 2023
DOI 10.3389/fnins.2023.1133933

OPEN ACCESS

EDITED BY
Peng Xu,
University of Electronic Science and Technology of China, China

REVIEWED BY
Fangzhou Xu,
Qilu University of Technology, China
Qi Li,
Changchun University of Science and Technology, China
Erwei Yin,
Tianjin Artificial Intelligence Innovation Center (TAIIC), China
Yunfa Fu,
Kunming University of Science and Technology, China

*CORRESPONDENCE
Shouliang Qi
✉ qisl@bmie.neu.edu.cn

SPECIALTY SECTION
This article was submitted to Neuroprosthetics, a section of the journal Frontiers in Neuroscience

RECEIVED 29 December 2022
ACCEPTED 27 February 2023
PUBLISHED 15 March 2023

CITATION
Bai X, Li M, Qi S, Ng ACM, Ng T and Qian W (2023) A hybrid P300-SSVEP brain-computer interface speller with a frequency enhanced row and column paradigm.
Front. Neurosci. 17:1133933.
doi: 10.3389/fnins.2023.1133933

A hybrid P300-SSVEP brain-computer interface speller with a frequency enhanced row and column paradigm

Xin Bai[1,2], Minglun Li[3], Shouliang Qi[1,2]*, Anna Ching Mei Ng[4], Tit Ng[4] and Wei Qian[1]

[1]College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, China, [2]Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education, Northeastern University, Shenyang, China, [3]Department of Biomedical Engineering, College of Precision Instruments and Optoelectronics Engineering, Tianjin University, Tianjin, China, [4]Shenzhen Jingmei Health Technology Co., Ltd., Shenzhen, China

Objective: This study proposes a new hybrid brain-computer interface (BCI) system to improve spelling accuracy and speed by stimulating P300 and steady-state visually evoked potential (SSVEP) in electroencephalography (EEG) signals.

Methods: A frequency enhanced row and column (FERC) paradigm is proposed to incorporate the frequency coding into the row and column (RC) paradigm so that the P300 and SSVEP signals can be evoked simultaneously. A flicker (white-black) with a specific frequency from 6.0 to 11.5 Hz with an interval of 0.5 Hz is assigned to one row or column of a 6 × 6 layout, and the row/column flashes are carried out in a pseudorandom sequence. A wavelet and support vector machine (SVM) combination is adopted for P300 detection, an ensemble task-related component analysis (TRCA) method is used for SSVEP detection, and the two detection possibilities are fused using a weight control approach.

Results: The implemented BCI speller achieved an accuracy of 94.29% and an information transfer rate (ITR) of 28.64 bit/min averaged across 10 subjects during the online tests. An accuracy of 96.86% is obtained during the offline calibration tests, higher than that of only using P300 (75.29%) or SSVEP (89.13%). The SVM in P300 outperformed the previous linear discrimination classifier and its variants (61.90–72.22%), and the ensemble TRCA in SSVEP outperformed the canonical correlation analysis method (73.33%).

Conclusion: The proposed hybrid FERC stimulus paradigm can improve the performance of the speller compared with the classical single stimulus paradigm. The implemented speller can achieve comparable accuracy and ITR to its state-of-the-art counterparts with advanced detection algorithms.

KEYWORDS

brain-computer interface, speller, electroencephalography, machine learning, Neural decoding

1. Introduction

The brain-computer interface (BCI) is a communication system that does not rely on peripheral nerves and muscles to send information and commands to the outside world (Wolpaw et al., 2000). The BCI system can directly control and communicate with other entities by reading and transducing brain signals. By providing patients with conditions such as amyotrophic lateral sclerosis (ALS) and locked-in syndrome (LIS) with a way to restore their communication, the BCI system can help patients restore certain motor functions, improve their quality of life, and even enable them to enjoy life as healthy people do. Meanwhile, electroencephalography (EEG) signals are secure, non-invasive, easy to use, easy to collect, and time-resolving, making them ideal for BCI systems. At present, various BCI systems based on EEG have been developed, such as the speller systems (Farwell and Donchin, 1988; Cecotti, 2010), wheelchair control systems (Li J. et al., 2013; Li Y. et al., 2013; Li et al., 2014; Kaufmann et al., 2014), prostheses, and mechanical arm control systems (Pfurtscheller et al., 2010; Hochberg et al., 2012; Onose et al., 2012).

Electroencephalography-based BCI systems can be divided into two types: spontaneous or evoked. A representative of spontaneous BCI is Motor Imagery (MI), in which the user autonomously controls their thought activity to form certain identifiable potential information to control an external device (Jiang et al., 2020; Chen et al., 2021; Pei et al., 2021). However, the user often needs sufficient training to become proficient. The representatives of evoked BCI systems are based on the event-related potential (ERP) and steady-state visually evoked potential (SSVEP). These BCIs are characterized by the need to rely on external stimuli to evoke certain specific potential information in the human brain. However, the advantage is that the evoked signals are often stable, and the user generally only needs to know the basic operation process, so evoked BCIs have received more attention, and their technology is relatively mature.

Event-related potential is a transient characteristic potential evoked by a small probability event (Squires et al., 1975), and the most commonly detected potential is the P300 potential. This is excited as a positive voltage approximately 300 ms after the onset of the target stimulus (Picton, 1992; Luck and Kappenman, 2011). In the "Oddball" paradigm, this component is evoked when a rare stimulus (target stimulus) appears in between several relevant stimuli (non-target stimulus). The task of the subjects was to focus on the target stimulus and count its occurrences (Sur and Sinha, 2009). The P300-based speller systems have been proposed early on, and most of them are based on the "Oddball" paradigm. The earliest P300 speller system using the row and column (RC) paradigm was proposed by Farwell and Donchin (1988), whose paradigm is in the form of a 6×6 matrix. Rows and columns are blinked once in each stimulus trial in a random order to induce the P300 signal, and the system determines the user's target character (intersection of rows and columns) by determining which row and which column triggered the P300 signal. Most previous studies have increased their accuracy by superimposed averaging of EEG signals, but resulting from the inefficiency, single-trail recognition of P300 has become a hot topic.

Steady-state visually evoked potential is a periodic response induced by a stimulus at a specific frequency. When the subject looks at a target flashing at a specific frequency, the subject's EEG signal is significantly enhanced, and significant peaks can be observed at the harmonics of the frequency after the time-to-frequency conversion. Meanwhile, it was shown that the SSVEP frequency located in the center of the visual field has the most pronounced energy increase and gradually decays toward the periphery in an approximately Gaussian distribution (Sutter, 1992) so that the subject needs to gaze at the stimulus target constantly. In addition to frequency, the phase can encode the stimulus frequency (Chen et al., 2014) and the single-stage paradigm. This can present many targets simultaneously and is applied in the design of SSVEP-based spellers. However, in later studies, the Multi-stage paradigm (Multi-stage) proved more prevalent and efficient (Kick and Volosyak, 2014; Li et al., 2021). Cecotti (2010) proposed a multilevel paradigm-based SSVEP speller system, while Nakanishi et al. (2014) proposed a hybrid frequency and phase coding for a frequency and phase-based SSVEP speller system. Chen et al. (2014) proposed a stimulus paradigm with 40 target characters and compared two frequency-phase mixing patterns.

Many studies have been performed to improve the performance and create a hybrid BCI paradigm. Since both SSVEP and P300 are EEG signals and their detection regions are independent of each other, that is, they come from the time and the frequency domain, respectively, it is feasible to design hybrid triggered systems combining P300 and SSVEP without additional data acquisition equipment (Xu et al., 2016; Kundu and Ari, 2022). Panicker et al. (2011) first introduced a hybrid P300/SSVEP BCI, which uses a P300 signal for target detection and an SSVEP signal for asynchronous spelling control, and only starts P300 spelling when SSVEP reaches a certain threshold. Xu et al. (2013) developed a hybrid BCI speller that, with the same target stimulus, evokes P300 spelling in different ways to evoke both P300 and SSVEP blocking (SSVEP-B). Yin et al. (2013) have developed a hybrid BCI speller system that divides the conventional P300 speller into six groups, each flashing at a different frequency, that combines distinct features of P300 and SSVEP to reduce the number of errors in the same row or column relative to the target errors that occur. More recently, a few hybrid BCI speller systems were constructed (Kapgate et al., 2020; Xu et al., 2020; Katyal and Singla, 2021; Han et al., 2022). Han et al. (2023) even developed one high-speed system with over 200 targets, greatly expanding the instruction set.

It is noted that there is a significant competing effect when P300 and SSVEP visual stimuli are presented simultaneously; SSVEP stimuli will reduce the amplitude of the P300 signal, while P300 stimuli significantly reduce the band power of the SSVEP signal. However, as reported in the previous study, this competing effect will not result in a significant decrease in decoding accuracy because the extracted features from the reduced signals are still able to discriminate between target categories (Lee et al., 2018).

In this study, we designed and implemented a hybrid BCI speller system with high compatibility and scalability. The contributions are in four aspects. First, a frequency enhanced row column (FERC) paradigm is proposed as a new hybrid stimulus paradigm. Second, the frequency coding is incorporated into the RC paradigm so that P300 and SSVEP signals can be evoked simultaneously. Third, the new hybrid P300-SSVEP speller outperforms that only using P300 or SSVEP. Fourth, advanced detection algorithms of P300 and SSVEP and their further fusion are done. The remainder of this work is organized

as follows. Section "2. Experimental methods" describes our methodology, the experimental dataset used in this work, and the corresponding data processing methods. Section "3. Results" describes the results. Section "4. Discussion" gives our conclusion and provides a discussion.

2. Experimental methods

2.1. The P300/SSVEP hybrid stimulation paradigm

A new P300/SSVEP hybrid stimulation paradigm is proposed and named the FERC paradigm. **Figure 1** shows that the stimulation interface is divided into six rows and six columns of 36 characters or targets (A to Z, 0 to 9). Each symbol is displayed equally spaced. To induce the P300 signal, the FERC paradigm uses the same principle as the Oddball paradigm, where rows and columns flash alternately in random order within a trial. Each row or column will only flash once in the same trial.

Simultaneously, to induce the SSVEP signal, the rows and columns were encoded with different frequencies of stimuli with continuous flashing. The frequency ranged from 6.0 to 11.5 Hz with an interval of 0.5 Hz, and there were 12 frequency groups. Specifically, the frequency of columns is 6.0, 6.5, 7.0, 7.5, 8.0, and 8.5 Hz; the frequency of rows is 9.0, 9.5, 10.0, 10.5, 11.0, and 11.5 Hz. For any row or column, the flashing process lasts one second. In this way, once the classifier recognizes the column and row, the target can be determined. Although the information of two stimuli is used, these stimuli are applied simultaneously in one trial. Therefore, the FERC paradigm belongs to the one-stage rather than the multi-stage paradigm.

In the stimulus interface, all characters are gray, and the background is black when there is no target flashing. The font size of characters will be enlarged by 16.67% (from 60 to 70), and their background intensity will change according to the time when the stimulus flashes appear.

$$Intensity = \frac{\mathrm{Cos}\left(2\pi \times f \times t\right) + 1}{2} \quad (1)$$

where f is the frequency of the stimulus and t is the duration of the visual stimulus. The reason for taking the cosine function is to ensure that the background intensity of the target character constantly changes periodically at any moment of the screen refresh and is not subject to frame loss because of the phase difference between the screen refresh and the operation.

2.2. Subjects

Eleven healthy volunteers (male, 19–24 years old, mean 20.7 years old) participated in our experiment, and data from 10 subjects were analyzed. All subjects provided written notification permission. They had no history of eye problems or neurological disorders, and nine subjects had no experience with the BCI system. The Medical Ethics Committee of Northeastern University approved this study. The participants provided their written informed consent to participate in this study. Before the start of the experiment, subjects were asked to minimize eye movements and to sit comfortably in a chair facing the screen.

2.3. Data acquisition

In this study, an actiCHamp EEG signal amplifier from Brainproducts (actiCHamp 32ch, Gilching, Germany), was used, along with the accompanying EasyCap electrode cap, which has an electrode setup in the cap corresponding to the internationally accepted standard 64 leads and a signal sampling rate of 250 Hz. In some other studies, only single-channel EEG signals have been used for BCI detection (Gao et al., 2003), but we found that multichannel data would not only yield better and more stable performance. The multichannel method is convenient for BCI applications because all users can use the same electrode cap (Bin et al., 2009).

In practice, 15 electrodes were used, including one ground electrode and one reference electrode (**Figure 2**). The ground electrode is located in the FPz of the international standard electrode, and the reference electrode is located in TP10 of the International standard electrode. The other electrodes were used to collect EEG signals. Fz, Cz, P3, P4, PO7, PO8, Pz, and Oz were used to collect P300 signals located in the parietal and occipital regions. Nine electrodes, PO7, PO8, Pz, Oz, O1, O2, PO3, PO4, and POz, were used to collect SSVEP signals in the occipital region. It is known from previous studies that P300 signals are mostly collected from the parietal and occipital regions (Luck and Kappenman, 2011) and SSVEP signals are mostly collected from the occipital region (Allison et al., 2014). Therefore, we selected the most helpful electrodes for the experimental results from the above regions and finally determined the above 15 electrodes.

The stimulus is displayed on a 27-inch monitor with a refresh rate of 240 Hz. All operational analysis is carried out in Matlab 2018b. After the collection, the data will be sent to Matlab software in the data processing equipment through TCP/IP protocol by the Remote Data Access (RDA) module, and the marked data will be shunted.

2.4. Signal preprocessing

There are four main steps in the preprocessing stage of EEG signals. Firstly, the raw EEG signal was filtered with a 60 Hz notch filter in order to remove the effects of the industrial frequency interference. Secondly, the filtered data were corrected to the baseline by "Detrending." Thirdly, the data were cut into approximately 0–800 ms segments after stimulus onset. Finally, the data were divided into two channels, P300 and SSVEP, and entered into the next step of the experiment.

2.5. Feature extraction and classification

The data processing procedure for a single trial in our speller system is given in **Figure 3**. The triaged P300 and SSVEP data are passed into their respective data processing modules and analyzed separately with different processing methods to extract the feature vectors. The processed P300 data are

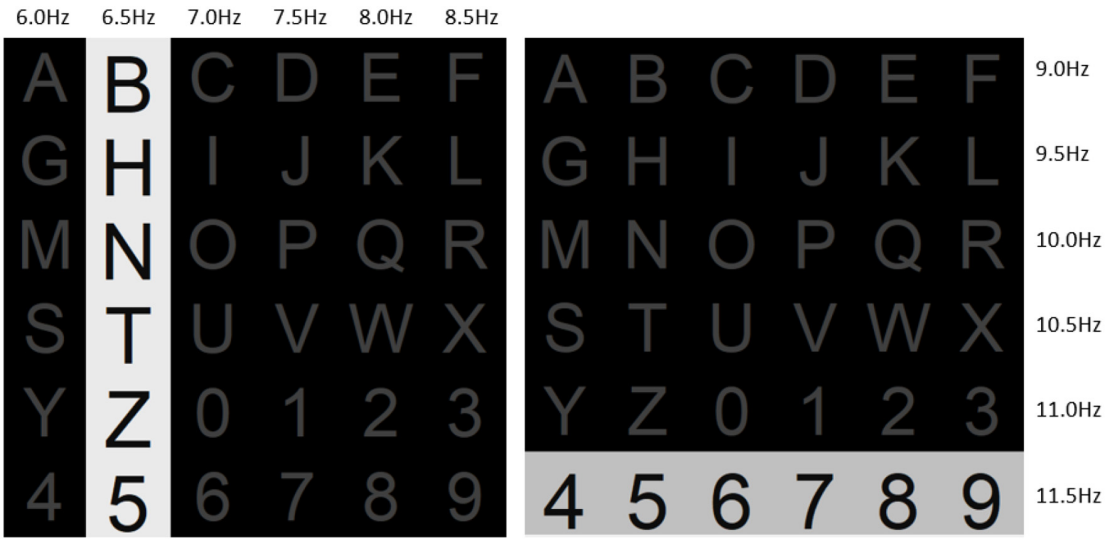


FIGURE 1
The proposed frequency enhanced row and column stimulus paradigm. A flicker (white-black) with a specific frequency is assigned to one row or column of a 6 × 6 layout, and the row/column flashes are carried out in a pseudorandom sequence. In this figure, Column 2 and Row 6 are activated, and the target output is the number "5".

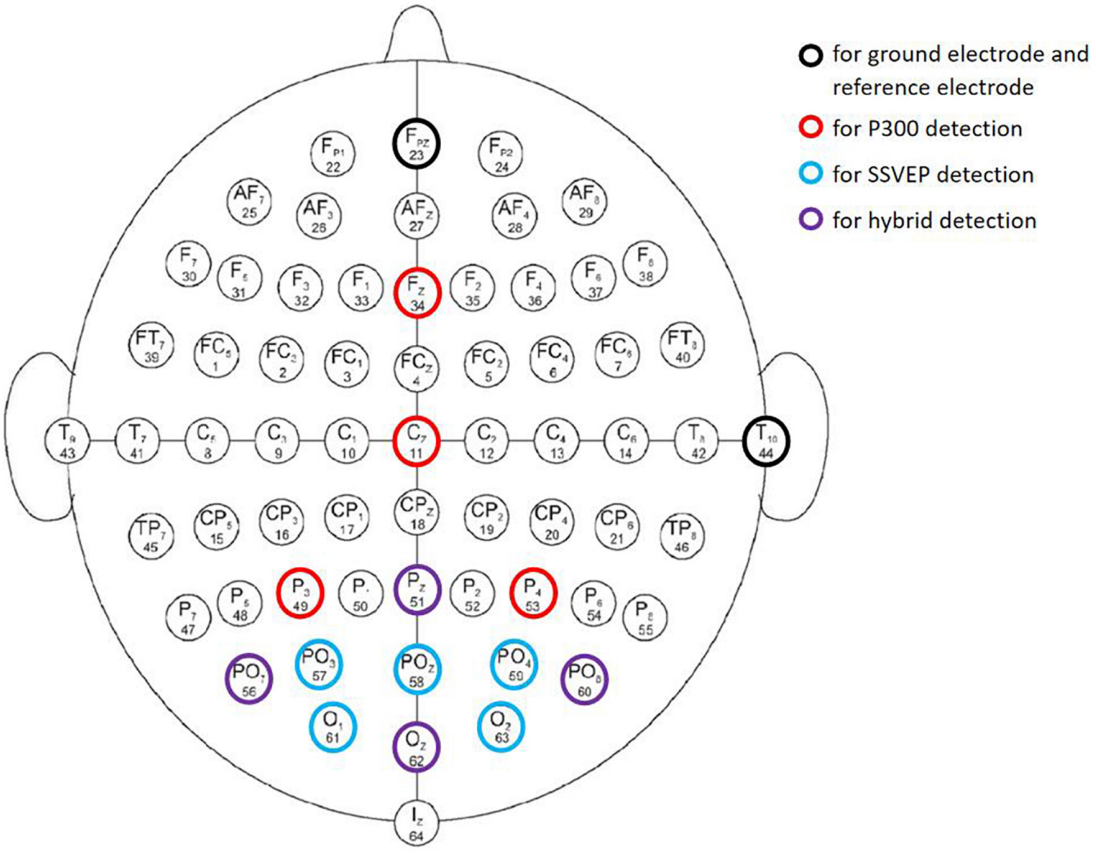


FIGURE 2
Schematic diagram of arrangements of the 15 electrodes used in the experiment.

passed into the five support vector machine (SVM) classifiers trained in advance, and then the classification results are fed into the modified weight controller to derive the P300 score. For SSVEP data, they are fed into the pre-trained task-related component analysis (TRCA) model at the end of the pre-processing procedure to derive the SSVEP score. The two scores

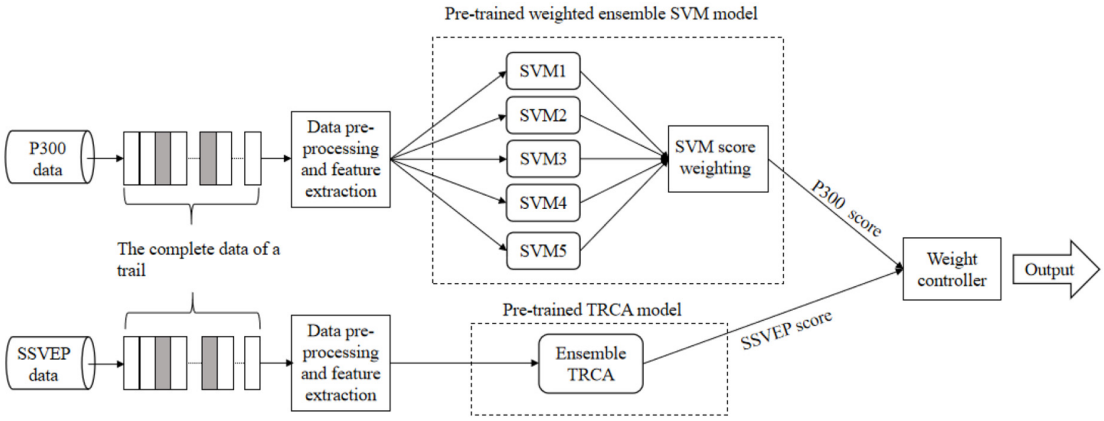


FIGURE 3
The data processing process for a single trial in the hybrid speller system. The gray blocks in the data section indicate target stimuli, and the white blocks indicate non-target stimuli. The two models and one weight controller are pre-trained by the calibration data.

are passed into the total weight controller to get the final output.

2.5.1. P300 feature extraction and classification

The superimposed averaging helps to improve the signal-to-noise ratio of P300 but often implies a more extended output character period, which is very detrimental to the efficiency of online systems. Therefore, a single detection of the P300 signal is chosen to improve the overall efficiency. It is well known that detecting single-trial P300 is a complex problem, so we used a follow-up operation to improve the accuracy.

The pre-processed data were used for feature extraction by wavelet transform. Since the frequency domain of the P300 signal ranges from 0.3 to 15 Hz, which is a low-frequency signal, the original signals are decomposed to a frequency scale of 0–12.5 Hz by wavelet transform. The m low-frequency wavelet coefficients of n channels are concatenated to form a feature vector of length $m \times n$ as the features of P300 and fed into the classifier for classification. Here, m and n are 13 and 8, respectively.

After the feature extraction, the SVM method is adopted to detect the occurrence of P300 because of its better generalization ability compared with other machine learning algorithms (Theodoridis and Koutroumbas, 2006). The principle of SVM is to find the hyperplane with the maximum distance that can separate sample categories from the high-dimensional space after mapping samples to the high-dimensional space. The Gaussian kernel is selected as the kernel function of SVM, which mainly considers the unique advantages in solving nonlinear problems when we are uncertain whether the P300 single detection problem is linearly separable.

Considering the single detection of the P300 signal, a weighted ensemble SVM method is further proposed to counteract the instability of the individual classifier to enhance the classification accuracy. Determined by the proposed paradigm in this study, the ratio of Target Stimulus Signal (TSS) samples containing the P300 signal to Non-target Stimulus Signal (NSS) samples in the collected EEG data is 1:5. Therefore, an undersampling method similar to the "EasyEnsemble" algorithm is adopted. Specifically, the NSS samples are randomly and evenly divided into five parts, the TSS samples are copied five times, and one part of NSS samples and one copy of TSS samples are combined into a training set to train one SVM classifier. In this way, the problem of sample imbalance is transformed into a problem of combining five classifiers (SVM1, SVM2, . . ., SVM5) with different training sets, which requires ensemble learning techniques.

The core idea of ensemble learning is to combine multiple individual learners to solve the same problem. A modified weighted ensemble method is used based on the theory of ensemble learning and the actual situation of the paradigm used in this study. It can dynamically adjust the weights of the classifiers according to the actual situation.

Suppose there are n trained classifiers (classifier_1, classifier_2, ..., classifier_n), where ($n > 0$) and their cross-validation accuracies are (acc_1, acc_2, ..., acc_n), let their weights be (w_1, w_2, ..., w_n) respectively, and the sum of the weights is 1. To make the weight of the classifiers with high accuracy greater, one can make their weights proportional to the accuracy distribution.

$$\begin{cases} \frac{w_1}{acc_1} = \frac{w_2}{acc_2} = \frac{w_3}{acc_3} = \ldots = \frac{w_n}{acc_n} \\ \sum_{i=1}^{n} w_i = 1 \end{cases} \quad (2)$$

Then for any m $\in$[1, n], the system of equations has the following solutions.

$$w_m = \frac{acc_m}{\sum_n^{i=1} acc_i} \quad (3)$$

However, suppose weights are assigned directly according to Equation 3. In that case, there is the problem that classifiers with better performance do not receive higher weights, which may make the result after integration inferior to using a single sub-classifier with better performance and render the integration useless. Therefore, this study improves on Equation 3 by introducing a theoretical accuracy to correct the weights to ensure the performance of the integration method.

Knowing the accuracy P of the classifier in the case of completely random target selection, we can assume that the classifier with classification accuracy $acc_i > P$ plays a positive role. The classifier with $acc_i < P$ is ineffective and cannot play a role, and the classifier with $acc_i = P$ plays a negative role. Based on this

conclusion, we improve the set of equations in Equation 2 and obtain the following equation.

$$\begin{cases} \frac{w_1}{acc_1-P} = \frac{w_2}{acc_2-P} = \frac{w_3}{acc_3-P} = \ldots = \frac{w_n}{acc_{n-P}} \\ \sum_{i=1}^{n} w_i = 1 \end{cases} \quad (4)$$

For any m∈[1, n], w_m can be calculated by

$$w_m = \frac{acc_m - P}{\sum_n^{i=1}(acc_i - P)} \quad (5)$$

According to the modified assignment method, the weights of those with higher classification accuracy are effectively scaled up, which is more helpful in improving the minimum level of the integrated classifier and saving it from being affected by the classification performance of classifiers with low accuracy.

2.5.2. SSVEP feature extraction and classification

For the SSVEP data, the filter bank technique is utilized first. The filtered data are input in the TRCA model pre-trained by the calibration data. Ten sub-bands (Sub1–Sub10) are set as Sub(k): [k × fa, 90] Hz, k = 1, 2,..., 10. To avoid distortion resulting from bandpass filtering, each sub-band filter has an extra width of 2 Hz on the low-frequency side, such as the actual sub-bands have a lower frequency limit of (k × fa − 2) Hz. Subsequently, a set of bandpass filters for these sub-bands are designed, and the Chebyshev I-type bandpass filter is used. The filter bank processing enhances the relative strength of the stimulus frequency harmonics, improving its detection accuracy and making the harmonics of the stimulus frequency much more usable for SSVEP detection. This data up-dimensioning method can improve the accuracy of recognition when the time required for calculation is not much or not limited.

The TRCA has previously been used for the BCI speller by Nakanishi et al. (2017). In our study, the calibration data of stimulus "n" used to perform SSVEP detection are defined as a four-dimension tensor $x = (\chi)_{njkh} \in R^{Nf \times Nc \times Ns \times Nt}$, and its corresponding test data or data to be detected are defined as a two-dimensional matrix $X \in R^{c \times Ns}$. Here n and N_f denote the identifier and number of target stimuli, j and N_c denote the identifier and number of channels, k and N_s denote the identifier and number of sampling points, and h and N_t denote the identifier and number of experimental trials. TRCA extracts task-relevant components by spatially filtering the training data. The spatial filter $\omega_f \in R^{Nc \times 1}$ at the stimulus frequency f can be calculated by the following equation.

$$arg_{\omega_f} max \frac{\omega_f^T A^T A \omega_f}{\omega_f^T B^T B \omega_f} \quad (6)$$

where $A \in R^{Ns \times Nc}$ denotes the result of averaging over N_t blocks at a frequency f in Z.

$$A = \frac{1}{N_t} \sum_{i=1}^{N_t} Z_{i,f} \quad (7)$$

$B=[Z^T{}_{1,f}\ Z^T{}_{2,f}\ Z^T{}_{3,f}\ \ldots\ Z^T{}_{Nt,f}]^T \in R^{Nt \times Ns \times Nc}$, and $Z_{i,f} \in R^{Ns \times Nc}$ denotes the multichannel EEG signal with stimulation frequency f in block i. After calculating the spatial filter ω_f at frequency f, TRCA uses the Pearson correlation coefficient between the averaged training data across trials for n-th visual stimulus and the test signal $X \in R^{Ns \times Nc}$ as the final discriminant.

$$p' = corr(X\omega_f, A\omega_f) \quad (8)$$

The pre-trained TRCA model is a classifier with 12 categories.

2.6. Weight controller combining P300 with SSVEP

The probability of each category is obtained and represented as the SSVEP score in the form of a 12 × 1 vector. Similarly, the P300 score is also represented by a 12 × 1 vector. The weight controller combines these two scores to yield the final prediction. The weights in the controller are determined by Equations 9, 10.

$$\begin{cases} \frac{w_P}{acc_P-E} = \frac{w_S}{acc_S-E} \\ w_P + w_S = 1 \end{cases} \quad (9)$$

$$\begin{cases} w_P = \frac{acc_P-E}{acc_P+acc_S-2*E} \\ w_S = \frac{acc_S-E}{acc_P+acc_S-2*E} \end{cases} \quad (10)$$

where acc_P and acc_S are the system's recognition accuracy for the P300 and SSVEP signal, respectively; w_P and w_S are the weights of the kinds of signals; E is the accuracy of the completely randomized system.

2.7. Experimental setup

The experimental procedure was divided into two phases for each subject: offline calibration and online testing. In the first phase, 11 epochs of data are collected, and the procedure is shown in **Figure 4**. During the experiment, the subjects only need to focus on their target and keep their eyes on the target location, ignoring the rest of the stimuli. Each epoch consists of 12 trials corresponding to a character input. Before each trial, two seconds are given to the subject for identifying, locating, and gazing at the target character. Each trial contains 12 flash stimuli (6 rows and 6 columns). The order and flashing frequency of the flashing stimulus were determined by the paradigm defined in section "2.1. The P300/SSVEP hybrid stimulation paradigm." The duration of each visual stimulus flashing is 1 s, and the stimulus interval is 100 ms.

No feedback is given to subjects throughout the calibration phase. For each subject, 1,584 trials (samples) were collected. Using these samples, the two models (Pre-trained weighted ensemble SVM model and Pre-trained TRCA model) and one weight controller are trained and tested offline in a cross-validation fashion. Specifically, 10 trials of data are used for training, and 1 trial of data is used for testing with 11 cross-validations.

The second phase of the test was divided into two parts: a copy speller test and a free spelling test. In the copy speller test, subjects are asked to type "BCISPELLER," "HELLOWORLD," and "NEUBMIE" in sequence (27 characters in total). For each subject, 972 flashes (27 characters × 12 flashes × 3 repetitions) are collected. The EEG signal of each flash collection is called a data sample and is sequentially input to the trained models and weight controller

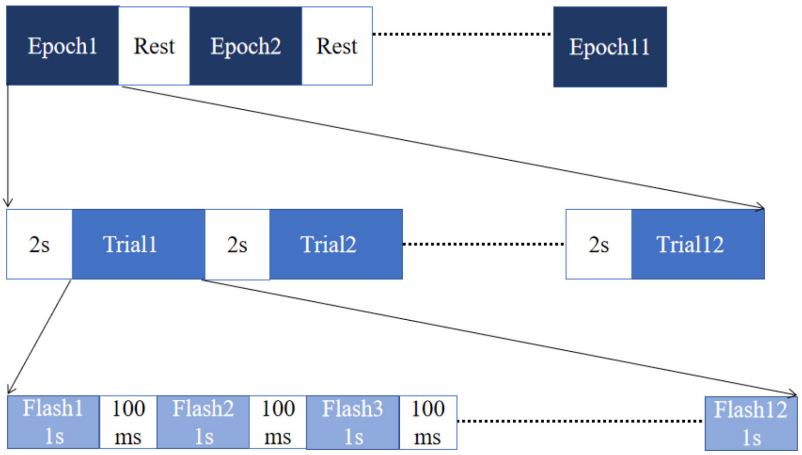


FIGURE 4
The time sequence diagram of the offline calibration.

for real-time online judgment. In the free input test, subjects are asked to spell any number of arbitrary characters and to inform the experimenter of the characters in mind at the end of the experiment, which the experimenter recorded.

During the test phase, at the end of each trial, as real-time feedback, the characters identified from the EEG signal of that trial will be displayed on the screen. In summary, the graphical user interface used by the hybrid speller system is given in **Figure 5**.

In both the offline calibration and online testing phases, subjects are given a 3–7 min break between each epoch of the test to relieve visual fatigue and relax their mind to avoid the impact of nervousness or excitement on the subsequent test, with the exact duration of the break determined by the subject.

2.8. Evaluation and comparison

Two important metrics are usually used to evaluate the performance of a BCI speller system: accuracy (ACC) and the information transfer rate (ITR). These two metrics represent the input accuracy and speed of the system, respectively. ACC is defined as the ratio of the number of successful target selections (X_1) to the total number of system inputs (X)

$$ACC = \frac{X_1}{X} \times 100\% \quad (11)$$

The input speed and the number of characters should also be considered, so Wolpaw introduced the metric ITR, which shows the amount of information that can be transmitted in 1 min (Wolpaw, 2007)

$$ITR = \{log_2 N + Plog_2 P + (1-P)log_2 \frac{1-P}{N-1}\}/T \quad (12)$$

here N is the number of categories of output commands available in the system, P is the probability of correctly selecting the target option, and T denotes the time of each trial of experiments.

The performance of the weighted ensemble SVM model for P300 and the ensemble TRCA model for SSVEP is compared with that of its three counterparts. They are FLDA (Fisher Linear Discriminant Analysis), BLDA (Bayesian Linear Discriminant Analysis), and CCA (Canonical Correlation Analysis). FLDA projects the data to lower dimensions, projects mean values of classes far apart, and the diffusion of projected data has been used for P300 detection (Panicker et al., 2011). BLDA uses regularization to prevent the overfitting of noisy data sets. With Bayesian analysis, the degree of regularization can be estimated automatically and quickly from the training data without time-consuming cross-validation (Hoffmann et al., 2008). CCA is a standard algorithm in SSVEP BCI (Lin et al., 2006; Bin et al., 2009), a multivariate statistical algorithm that attempts to reveal the correlation between two data sets. Moreover, our hybrid speller is compared with some state-of-the-art counterparts regarding the number of subjects, detection algorithms, stimulus paradigms, ACC, and ITR.

3. Results

3.1. Performance of the hybrid speller in the offline calibration tests

Table 1 presents the performance of the hybrid speller and its counterparts in the offline calibration tests, the standard deviation of ACC and ITR is also given for different classification methods. Among the 10 subjects, the hybrid speller accuracy ranges from 93.89 to 99.31%, and the mean reaches 96.86%, much higher than that only by P300 (75.29%) and only SSVEP (89.13%). Meanwhile, the hybrid speller yields a mean ITR of 30.08 bits/min, outperforming P300 only (19.36 bits/min) and SSVEP only (25.83 bits/min). The speller by SSVEP alone outperforms P300 alone.

3.2. Performance of the hybrid speller in the online tests

In the online tests, our hybrid speller achieves a mean accuracy of 94.29% and a mean ITR of 28.64 bits/min (**Table 2**). One out of

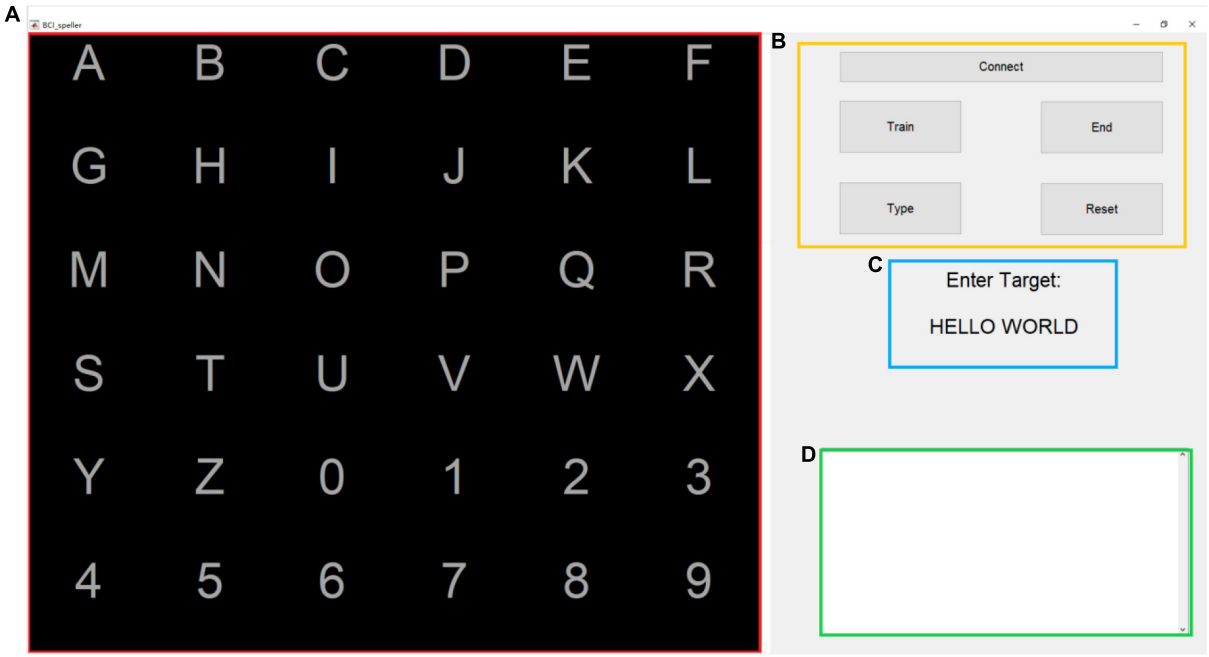


FIGURE 5
The graphical user interface used by the hybrid speller system. The area marked in red **(A)** is the visual stimulus interface; the area marked in yellow **(B)** is the control panel; the area marked in blue **(C)** is the input target prompt text box, and the area marked in green **(D)** is the input text box.

TABLE 1 Performance of the hybrid speller and its counterparts in the offline calibration tests.

Subject	ACC (%)			ITR (bits/min)		
	P300	SSVEP	Hybrid	P300	SSVEP	Hybrid
S1	76.51	87.12	95.33	19.86	24.72	29.11
S2	75.00	95.83	98.33	19.23	29.41	31.01
S3	75.00	80.42	93.89	19.23	21.58	28.28
S4	74.31	87.50	98.03	18.94	24.91	30.81
S5	72.92	96.67	99.31	18.36	29.93	31.72
S6	75.00	87.50	95.14	19.23	24.91	29.00
S7	74.31	88.75	97.50	18.94	25.53	30.46
S8	79.17	89.17	96.53	21.02	25.75	29.84
S9	76.37	86.25	96.80	19.81	24.29	30.01
S10	74.31	92.08	97.78	18.94	27.28	30.64
Mean (SD)	75.29 (1.63)	89.13 (4.51)	96.86 (1.58)	19.36 (0.69)	25.83 (2.35)	30.08 (0.99)

TABLE 2 Performance of the hybrid speller in the online tests.

Subject	ACC (%)	ITR (bits/min)
S1	88.89	25.61
S2	94.90	28.86
S3	90.74	26.56
S4	92.59	27.56
S5	100.00	32.31
S6	98.15	30.89
S7	93.75	28.20
S8	90.48	26.43
S9	95.31	29.10
S10	98.15	30.89
Mean (SD)	94.29 (3.51)	28.64 (2.08)

10 subjects has an accuracy of 100% at an ITR of 32.31 bits/min, and nine subjects have an accuracy higher than 90%. In addition, all subjects had an ITR greater than 25 bits/min.

Figure 6 shows the accuracy of the hybrid speller with the number of tests in the online testing phase. In the copy speller tests, the accuracy increases with the number of tests, and the accuracy in the copy speller test 3 (mean value, 96.48%) is significantly higher than that in the copy speller test 1 (mean value, 89.07%). This may indicate that most subjects become increasingly proficient in using the hybrid speller system as the number of tests increases and the learning and training have performance-enhancing effects for speller users.

Interestingly, in a few cases, spelling the same character too much caused a decrease in accuracy. For example, for S5, the average accuracy of three copy speller tests is 100, 98.14, and 96.30%. The possible reason might be visual fatigue.

The free spelling test yields higher accuracy than the copy speller test 1 (89.07 versus 95.69%, $p < 0.05$). There is no significant difference between the free spelling test and the second and third copy speller tests ($p > 0.05$). This situation may be related to the subjects' attention; specifically, too much input from the same character causes inattention, and free input may encourage the subjects to focus more actively on the target.

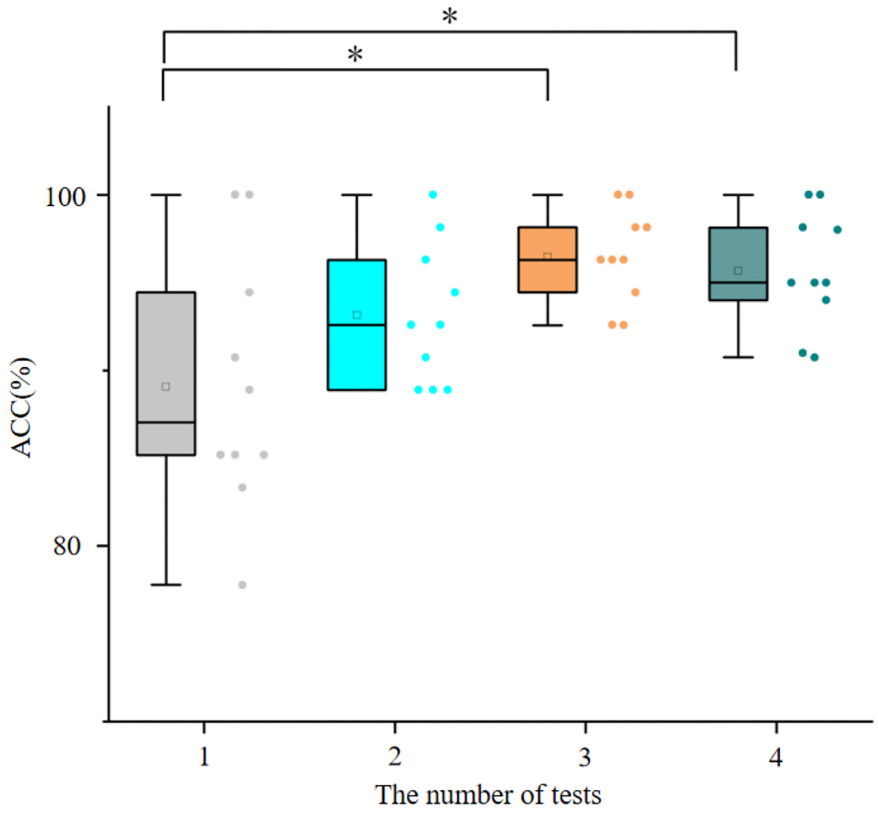


FIGURE 6
The accuracy of the hybrid speller with the number of tests in the online tests. Asterisk represents a significant difference ($p < 0.05$), 1–3 are the copy speller tests, and 4 is the free spelling test.

Many studies present the difference in the average accuracy and average ITR between the online and offline experiments (Chang et al., 2016; Katyal and Singla, 2021). In line with our results, the online performance is worse than the offline. The reason for this difference is unclear. We speculate that the performance might be affected by the subjects' emotional state at different times, the wearing of soft electrode caps, and other issues. It is noted that the differences are within acceptable limits.

Figure 7 shows the corresponding confusion matrix in the online experiments to compare the recognition accuracy of different targets. The diagonal line in this matrix shows the accuracy when the predicted value is true. Because the rows and columns are calculated separately, there are no elements in the upper right and lower left parts of the confusion matrix. In most cases (10/12), the wrong identification usually occurs at the adjacent targets. Here is an example: the highest 3% error appears in the adjacent column (the third column) while the real target is located in the second column with a frequency of 6.5 Hz. There are only two special cases (the first column and the third row) and the highest error does not appear within two adjacent targets.

3.3. Results of the comparative experiments and spellers

Using the same dataset collected in our calibration tests, three comparative algorithms are compared with our SVM and TRCA methods (Table 3). Our SVM method achieves higher ACC compared with FLDA and BLDA in P300 detection (75.29 versus 72.22 and 61.90%) and higher ITR (19.36 versus 18.08 and 14.11). Our TRCA outperforms CCA in SSVEP detection (ACC: 89.13 versus 73.33%; ITR: 25.83 versus 18.53).

Our hybrid speller is compared with eight counterparts regarding the number of subjects, detection algorithms, stimulus paradigms, ACC, and ITR (Table 4). The number of subjects in previous studies ranged between three and twenty. The ACC is between 79.17 and 9.90%, and ITR is between 19.8 and 164.00 bits/min. Our hybrid speller performs comparably to the-state-of-art hybrid spellers (P300 and SSVEP). It is worth mentioning that we use a single-trial P300 to increase ITR and our counterparts use the method of superimposed averaging. Our hybrid speller uses a one-stage paradigm similar to the other four studies (Panicker et al., 2011; Yin et al., 2013; Xu et al., 2014; Jalilpour et al., 2020), and the other four spellers use the multi-stage paradigm (Yin et al., 2015; Chang et al., 2016; Xu et al., 2020; Katyal and Singla, 2021). The one-stage paradigm has the outstanding feature of high speed, while the multi-stage paradigm can have a higher capacity of characters (Li et al., 2021).

The FERC paradigm is proposed and implemented in our speller. This paradigm differs from the four previous one-stage paradigms (Panicker et al., 2011; Yin et al., 2013; Xu et al., 2014; Jalilpour et al., 2020). To the best of our knowledge, this paradigm has not been used for the speller though a similar paradigm has

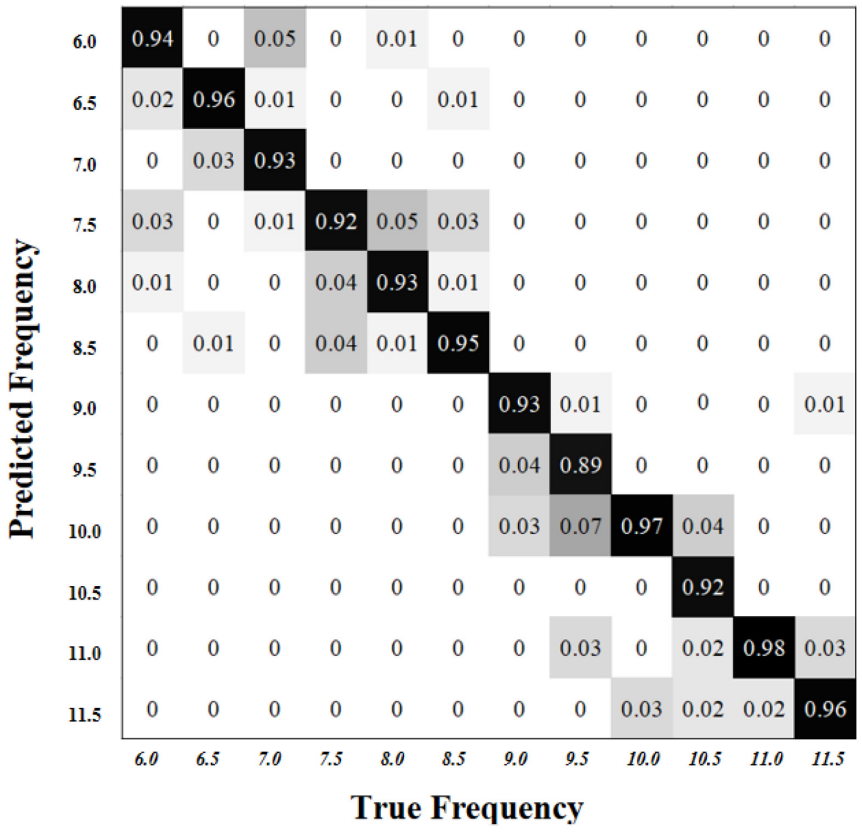


FIGURE 7
The confusion matrix in the online experiments to compare the recognition accuracy of different targets.

been used for the BCI for other purposes (Allison et al., 2014; Wang et al., 2015).

Our work is different from previous studies mainly in two aspects: (1) the stimulation paradigm; (2) the method of feature extraction and classification of EEG signals. First, our paradigm triggers both P300 and SSVEP signals simultaneously by the flickers (white-black) of row/column with different frequencies. It is different from the previous four one-stage paradigms (Panicker et al., 2011; Yin et al., 2013; Xu et al., 2014; Jalilpour et al., 2020). Despite belonging to the same category of paradigms, they have significantly distinct features. In the paradigm of Panicker et al. (2011), P300 and SSVEP signals have different roles and P300 spelling started only when SSVEP reached a certain threshold. Yin et al. (2013) directly combined SSVEP and P300 signals triggered by different stimuli (i.e., the flashing of row/column and orange crosses, respectively). Xu et al. (2014) used the paradigm suitable for capturing signals of P300 and SSVEP blocking, while Jalilpour et al. (2020) specified the paradigm using the spatial characteristics of the SSVEP signal response. Second, for the method of feature extraction and classification of EEG signals, our work combined SVM for P300 and TRCA for SSVEP with the more advanced fusion algorithm. In previous studies, different linear discriminant analysis (LDA) algorithms [e.g, FLDA, SWLDA (step-wise LDA), RLAD (regularized LDA), BLAD] have been commonly utilized for P300 signal (Panicker et al., 2011; Yin et al., 2013, 2015; Xu et al., 2014; Chang et al., 2016; Jalilpour et al., 2020; Katyal and Singla, 2021). For SSVEP signals, the CCA method is usually adopted (Panicker et al., 2011; Yin et al., 2013, 2015; Xu et al., 2014; Chang et al., 2016; Katyal and Singla, 2021) while SVM is also used (Jalilpour et al., 2020).

TABLE 3 Performance comparison of the proposed algorithm and its counterparts.

Performance	P300			SSVEP	
	FLDA	BLDA	SVM (our method)	CCA	TRCA (our method)
ACC [Mean (SD)]	72.22 (2.44)	61.90 (4.62)	75.29 (1.63)	73.33 (3.13)	89.13 (4.51)
ITR [Mean (SD)]	18.08 (1.14)	14.11 (2.89)	19.36 (0.69)	18.53 (1.87)	25.83 (2.35)

4. Discussion

This article proposes a FERC as a new hybrid stimulus paradigm, and a hybrid speller is implemented. The frequency coding is incorporated into the RC paradigm so that P300 and SSVEP signals can be evoked simultaneously. The new hybrid P300-SSVEP speller outperforms that using P300 or SSVEP alone. Advanced detection algorithms of P300 and SSVEP (such as the weighted ensemble SVM model and the ensemble TRCA model) and their further fusion by the weight controller are crucial. In the end, the implemented hybrid speller presents a comparable

TABLE 4 Comparison of our hybrid speller to the previous ones.

Study	Key aspects	Performance
Our method	Subjects: 10; P300: SVM; SSVEP: ensemble TRCA Our paradigm causes both P300 and SSVEP signals to be hybrid by weights.	ACC = 94.29% ITR = 28.64
Panicker et al., 2011	Subjects: 3; P300: FLDA; SSVEP: CCA It uses a P300 signal for target detection and an SSVEP signal for asynchronous spelling control and only starts P300 spelling when SSVEP reaches a certain threshold.	ACC = 97.00% ITR = 20.13
Yin et al., 2013	Subjects: 12; P300: SWLDA; SSVEP: CCA SSVEP is stimulated by flashing between white and black at different frequencies, and P300 is stimulated by highlighting rows and columns using orange crosses.	ACC = 93.85% ITR = 56.44
Xu et al., 2014	Subjects: 11; P300: SWLDA; SSVEP: CCA P300 and SSVEP blocking were simultaneously evoked in different ways under the same target stimulus.	ACC = 87.80% ITR = 54.00
Jalilpour et al., 2020	Subjects: 6; P300: RLDA; SSVEP: SVM The P300 signal determines the target symbol group, and the character position is determined by the spatial characteristics of the SSVEP signal response.	ACC = 93.06% ITR = 23.41
Yin et al., 2015	Subjects: 13; P300: SWLDA; SSVEP: CCA In the multi-step SSVEP paradigm, the frequencies of SSVEP stimuli switch once midway through each selection, such as the first half of the selection is for step one and the second half for step two.	ACC = 95.18% ITR = 50.41
Chang et al., 2016	Subjects: 10; P300: SWLDA; SSVEP: CCA In the multi-stage paradigm, the flickering stimulus and periodic change of the character evoke dual-frequency SSVEP, while the oddball stimulus of the target character evokes P300.	ACC = 93.00% ITR = 31.80
Xu et al., 2020	Subjects: 10; ensemble TRCA In the multi-stage paradigm, different sub-spellers use different frequencies and initial phases, and different characters blink at different times.	ACC = 79.17% ITR = 164.00
Katyal and Singla, 2021	Subjects: 20; P300: BLDA; SSVEP: CCA In the multi-stage paradigm, P300 estimates the tier of alpha-numeric symbol sets intended by the subject. The secondary selection phase used a traditional SSVEP paradigm to elicit SSVEP markers.	ACC = 96.42% ITR = 131.00

performance in terms of both ACC and ITR to the-state-of-art hybrid spellers (P300 and SSVEP).

Although P300 spellers have been extensively studied and can achieve good accuracy with multiple trials per symbol, using single-trial spellings is still a challenging problem. In our study, the hybrid system incorporating SSVEP achieved an online spelling performance of 94.29% accuracy, ITR of 28.64 bits/min using a single trial, and 96.8% accuracy after two training sessions. These results show that our BCI speller is expected to enable fast spelling in stimulus-driven BCI applications.

We used data acquisition in parallel with stimulus generation and labeling for the system's design to speed up the feedback. For the choice of paradigm, we designed the FERC to incorporate frequency coding into the RC paradigm to achieve the effect of evoking P300 and SSVEP signals simultaneously. For the selection of frequency, we chose a relatively low 6.0–11.5 Hz with a spacing of 0.5 Hz to get the best possible classification results while also considering the problem of visual fatigue of the subjects. For the page layout, we used the classic 6×6 layout and finally simplified all the operations into a few buttons, making it easy for people without professional training to operate.

Simultaneously, the system is not dependent on existing external platforms, so it is highly scalable and compatible, supporting many different application scenarios, and can be used within the system, in-text editing software, and social software. The application of a virtual keyboard in the feedback control module allows the system to support the input of text in various languages using the input methods already installed on the computer, meeting the user's needs while ensuring that it is as consistent as possible with the input habits of healthy people in their lives. In subsequent performance tests, the system performed well.

The most likely explanation for the performance improvement is three aspects. First, the clever design of the hybrid stimulus mechanism can detect both signals without causing performance degradation. Second, the modified weighted integrated SVM method can classify P300 signals more efficiently compared with the traditional SVM algorithm for processing EEG signal effects. Third, the addition of the SSVEP component provides additional information that helps predict targets versus non-targets. These analyses conclude that our hybrid BCI method yields better and more stable performance than the P300-only and SSVEP-only methods.

One of the 11 subjects was discarded because of the low accuracy (33.33%). By reviewing the literature, we identified a phenomenon called "BCI Illiteracy," which refers to the presence of a proportion of people who cannot trigger the two signal-evoked signals P300 or SSVEP, as well as the spontaneous signal MI (Lee et al., 2019). However, there is still no standardized definition of BCI blindness, so we believe this is a possible explanation.

We collected feedback from subjects after the experiment on their feelings about using the system and their experience with the experimental process and stimulation paradigm. The following two main feedback comments are obtained. First, all subjects reported significant visual fatigue during the overall length of data collection in the first phase of the experiment, and some subjects would feel annoyed or lose focus uncontrollably during the process. However, from an experimental design perspective, if the number of training sessions is reduced, it may lead to poor generalization of the model. During model calibration, we found that the accuracy of model

cross-validation improved with the increase in data. Second, several volunteers reported that their attention was affected to some extent by the flickering of adjacent characters while gazing at the target characters, both during calibration and testing. They indicated that they would unconsciously shift their eyes to gaze at or look at these targets with their afterglow when the adjacent characters were illuminated, and the low accuracy of some volunteers may have been related to this factor.

Combined with the above feedback, the study of reducing the collected training data is an important future research direction in which the generated spurious data will supplement the missing training data. In addition, to further improve the performance of our BCI speller, in the future, we will conduct more research with more advanced signal processing algorithms (Pei et al., 2022), reduce the current electrode set, and select the optimal stimulus onset asynchrony for the flash frequency.

5. Conclusion

A hybrid BCI speller system based on a single-trial P300 and SSVEP has been designed and implemented. The frequency coding is incorporated into the RC paradigm so that the P300 and SSVEP signals can be evoked simultaneously. Advanced detection algorithms of P300 (the weighted ensemble SVM) and SSVEP (the ensemble TRCA) and their further fusion lead to good performance (average 94.8%, maximum accuracy of 100%, and ITR of 28.64 bits/min). The new hybrid P300-SSVEP speller outperforms the P300 or SSVEP alone and shows comparable performance to its state-of-the-art counterparts. These results demonstrate that our speller system has specific application prospects and practical value.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of Northeastern University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

XB performed the experiments and analyzed the data along with ML and SQ. SQ, AN, TN, and WQ conceived the study, presented the results, and wrote the manuscript along with XB. XB collected and analyzed the data. SQ and WQ supervised the algorithm development and analyzed the data. All authors read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (82072008), the Fundamental Research Funds for the Central Universities (N2119010 and N2224001-10), and Open funding from Shenzhen Jingmei Health Technology Co., Ltd. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Conflict of interest

AN and TN were employed by Shenzhen Jingmei Health Technology Co., Ltd.

The remaining 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

References

Allison, B., Jin, J., Zhang, Y., and Wang, X. (2014). A four-choice hybrid P300/SSVEP BCI for improved accuracy. *Brain-Comp. Interfaces* 1, 17–26. doi: 10.1080/2326263X.2013.869003

Bin, G., Gao, X., Yan, Z., Hong, B., and Gao, S. (2009). An online multi-channel SSVEP-based brain–computer interface using a canonical correlation analysis method. *J. Neural Eng.* 6:046002. doi: 10.1088/1741-2560/6/4/046002

Cecotti, H. (2010). A self-paced and calibration-less SSVEP-based brain–computer interface speller. *IEEE Trans. Neural Systems Rehabil. Eng.* 18, 127–133. doi: 10.1109/TNSRE.2009.2039594

Chang, M. H., Lee, J. S., Heo, J., and Park, K. S. (2016). Eliciting dual-frequency SSVEP using a hybrid SSVEP-P300 BCI. *J. Neurosci. Methods* 258, 104–113. doi: 10.1016/j.jneumeth.2015.11.001

Chen, L., Chen, P., Zhao, S., Luo, Z., Chen, W., Pei, Y., et al. (2021). Adaptive asynchronous control system of robotic arm based on augmented reality-assisted brain–computer interface. *J. Neural Eng.* 18:066005. doi: 10.1088/1741-2552/ac3044

Chen, X., Wang, Y., Nakanishi, M., Jung, T. P., and Gao, X. (2014). Hybrid frequency and phase coding for a high-speed SSVEP-based BCI speller. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2014, 3993–3996. doi: 10.1109/EMBC.2014.6944499

Farwell, L. A., and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography Clin. Neurophysiol.* 70, 510–523. doi: 10.1016/0013-4694(88)90149-6

Gao, X., Xu, D., Cheng, M., and Gao, S. (2003). A BCI-based environmental controller for the motion-disabled. *IEEE Trans. Neural Systems Rehabil. Eng.* 11, 137–140. doi: 10.1109/TNSRE.2003.814449

Han, J., Liu, C., Chu, J., Xiao, X., Chen, L., Xu, M., et al. (2022). Effects of inter-stimulus intervals on concurrent P300 and SSVEP features for hybrid brain-computer interfaces. *J. Neurosci. Methods* 372:109535. doi: 10.1016/j.jneumeth.2022.109535

Han, J., Xu, M., Xiao, X., Yi, W., Jung, T. P., and Ming, D. (2023). A high-speed hybrid brain-computer interface with more than 200 targets. *J. Neural Eng.* 20:016025. doi: 10.1088/1741-2552/acb105

Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., et al. (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* 485, 372–375. doi: 10.1038/nature11076

Hoffmann, U., Vesin, J. M., Ebrahimi, T., and Diserens, K. (2008). An efficient P300-based brain–computer interface for disabled subjects. *J. Neurosci. Methods* 167, 115–125. doi: 10.1016/j.jneumeth.2007.03.005

Jalilpour, S., Sardouie, S. H., and Mijani, A. (2020). A novel hybrid BCI speller based on RSVP and SSVEP paradigm. *Comp. Methods Prog. Biomed.* 187:105326. doi: 10.1016/j.cmpb.2020.105326

Jiang, J., Wang, C., Wu, J., Qin, W., Xu, M., and Yin, E. (2020). Temporal combination pattern optimization based on feature selection method for motor imagery BCIs. *Front. Hum. Neurosci.* 14:231. doi: 10.3389/fnhum.2020.00231

Kapgate, D., Kalbande, D., and Shrawankar, U. (2020). An optimized facial stimuli paradigm for hybrid SSVEP+ P300 brain computer interface. *Cogn. Systems Res.* 59, 114–122. doi: 10.1016/j.cogsys.2019.09.014

Katyal, E., and Singla, R. (2021). EEG-based hybrid QWERTY mental speller with high information transfer rate. *Med. Biol. Eng. Comp.* 59, 633–661. doi: 10.1007/s11517-020-02310-w

Kaufmann, T., Herweg, A., and Kübler, A. (2014). Toward brain-computer interface based wheelchair control utilizing tactually-evoked event-related potentials. *J. Neuroeng. Rehabil.* 11:7. doi: 10.1186/1743-0003-11-7

Kick, C., and Volosyak, I. (2014). Evaluation of different spelling layouts for SSVEP based BCIs. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2014, 1634–1637. doi: 10.1109/EMBC.2014.6943918

Kundu, S., and Ari, S. (2022). Brain-Computer interface speller system for alternative communication: a review. *IRBM* 43, 317–324. doi: 10.1016/j.irbm.2021.07.001

Lee, M. H., Kwon, O. Y., Kim, Y. J., Kim, H. K., Lee, Y. E., Williamson, J., et al. (2019). EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy. *GigaScience* 8:giz002. doi: 10.1093/gigascience/giz002

Lee, M. H., Williamson, J., Lee, Y. E., and Lee, S. W. (2018). Mental fatigue in central-field and peripheral-field steady-state visually evoked potential and its effects on event-related potential responses. *NeuroReport* 29:1301. doi: 10.1097/WNR.0000000000001111

Li, J., Ji, H., Cao, L., Zang, D., Gu, R., Xia, B., et al. (2014). Evaluation and application of a hybrid brain computer interface for real wheelchair parallel control with multi-degree of freedom. *Int. J. Neural Systems* 24:1450014. doi: 10.1142/S0129065714500142

Li, J., Liang, J., Zhao, Q., Li, J., Hong, K., and Zhang, L. (2013). Design of assistive wheelchair system directly steered by human thoughts. *Int. J. Neural Systems* 23:1350013. doi: 10.1142/S0129065713500135

Li, M., He, D., Li, C., and Qi, S. (2021). Brain–computer interface speller based on steady-state visual evoked potential: a review focusing on the stimulus paradigm and performance. *Brain Sci.* 11:450. doi: 10.3390/brainsci11040450

Li, Y., Pan, J., Wang, F., and Yu, Z. (2013). A hybrid BCI system combining P300 and SSVEP and its application to wheelchair control. *IEEE Trans. Biomed. Eng.* 60, 3156–3166. doi: 10.1109/TBME.2013.2270283

Lin, Z., Zhang, C., Wu, W., and Gao, X. (2006). Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. *IEEE Trans. Biomed. Eng.* 53, 2610–2614. doi: 10.1109/TBME.2006.886577

Luck, S. J., and Kappenman, E. S. (eds) (2011). *The Oxford Handbook of Event-related Potential Components*. Oxford: Oxford university press.

Nakanishi, M., Wang, Y., Chen, X., Wang, Y. T., Gao, X., and Jung, T. P. (2017). Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis. *IEEE Trans. Biomed. Eng.* 65, 104–112. doi: 10.1109/TBME.2017.2694818

Nakanishi, M., Wang, Y., Te Wang, Y., Mitsukura, Y., and Jung, T. P. (2014). A high-speed brain speller using steady-state visual evoked potentials. *Int. J. Neural Systems* 24:1450019. doi: 10.1142/S0129065714500191

Onose, G., Grozea, C., Anghelescu, A., Daia, C., Sinescu, C. J., Ciurea, A. V., et al. (2012). On the feasibility of using motor imagery EEG-based brain–computer interface in chronic tetraplegics for assistive robotic arm control: a clinical test and long-term post-trial follow-up. *Spinal Cord* 50, 599–608. doi: 10.1038/sc.2012.14

Panicker, R. C., Puthusserypady, S., and Sun, Y. (2011). An asynchronous P300 BCI with SSVEP-based control state detection. *IEEE Trans. Biomed. Eng.* 58, 1781–1788. doi: 10.1109/TBME.2011.2116018

Pei, Y., Luo, Z., Yan, Y., Yan, H., Jiang, J., Li, W., et al. (2021). Data augmentation: using channel-level recombination to improve classification performance for motor imagery EEG. *Front. Hum. Neurosci.* 15:645952. doi: 10.3389/fnhum.2021.645952

Pei, Y., Sheng, T., Luo, Z., Xie, L., Li, W., Yan, Y., et al. (2022). "A tensor-based frequency features combination method for brain–computer interfaces," in *Proceedings of the Cognitive Systems and Information Processing: 6th International Conference, ICCSIP 2021, Suzhou, China, November 20–21, 2021, Revised Selected Papers 6*, (Singapore: Springer).

Pfurtscheller, G., Solis-Escalante, T., Ortner, R., Linortner, P., and Muller-Putz, G. R. (2010). Self-paced operation of an SSVEP-Based orthosis with and without an imagery-based "brain switch:" a feasibility study towards a hybrid BCI. *IEEE Trans. Neural Systems Rehabil. Eng.* 18, 409–414. doi: 10.1109/TNSRE.2010.2040837

Picton, T. W. (1992). The P300 wave of the human event-related potential. *J. Clin. Neurophysiol.* 9, 456–479. doi: 10.1097/00004691-199210000-00002

Squires, N. K., Squires, K. C., and Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephal. Clin. Neurophysiol.* 38, 387–401. doi: 10.1016/0013-4694(75)90263-1

Sur, S., and Sinha, V. K. (2009). Event-related potential: an overview. *Industrial Psychiatry J.* 18:70. doi: 10.4103/0972-6748.57865

Sutter, E. E. (1992). The brain response interface: communication through visually-induced electrical brain responses. *J. Microcomp. Appl.* 15, 31–45. doi: 10.1016/0745-7138(92)90045-7

Theodoridis, S., and Koutroumbas, K. (2006). *Pattern Recognition*. Amsterdam: Elsevier.

Wang, M., Daly, I., Allison, B. Z., Jin, J., Zhang, Y., Chen, L., et al. (2015). A new hybrid BCI paradigm based on P300 and SSVEP. *J. Neurosci. Methods* 244, 16–25. doi: 10.1016/j.jneumeth.2014.06.003

Wolpaw, J. R. (2007). "Brain-computer interfaces (BCIs) for communication and control," in *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, (New York, NY), doi: 10.1016/S1388-2457(02)00057-3

Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., et al. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE Trans. Rehabil. Eng.* 8, 164–173. doi: 10.1109/TRE.2000.847807

Xu, M., Chen, L., Zhang, L., Qi, H., Ma, L., Tang, J., et al. (2014). A visual parallel-BCI speller based on the time–frequency coding strategy. *J. Neural Eng.* 11:026014. doi: 10.1088/1741-2560/11/2/026014

Xu, M., Han, J., Wang, Y., Jung, T. P., and Ming, D. (2020). Implementing over 100 command codes for a high-speed hybrid brain-computer interface using concurrent P300 and SSVEP features. *IEEE Trans. Biomed. Eng.* 67, 3073–3082. doi: 10.1109/TBME.2020.2975614

Xu, M., Jia, Y., Qi, H., Hu, Y., He, F., Zhao, X., et al. (2016). Use of a steady-state baseline to address evoked vs. oscillation models of visual evoked potential origin. *Neuroimage* 134, 204–212. doi: 10.1016/j.neuroimage.2016.03.073

Xu, M., Qi, H., Wan, B., Yin, T., Liu, Z., and Ming, D. (2013). A hybrid BCI speller paradigm combining P300 potential and the SSVEP blocking feature. *J. Neural Eng.* 10:026001. doi: 10.1088/1741-2560/10/2/026001

Yin, E., Zeyl, T., Saab, R., Chau, T., Hu, D., and Zhou, Z. (2015). A hybrid brain–computer interface based on the fusion of P300 and SSVEP scores. *IEEE Trans. Neural Systems Rehabil. Eng.* 23, 693–701. doi: 10.1109/TNSRE.2015.2403270

Yin, E., Zhou, Z., Jiang, J., Chen, F., Liu, Y., and Hu, D. (2013). A novel hybrid BCI speller based on the incorporation of SSVEP into the P300 paradigm. *J. Neural Eng.* 10:026012. doi: 10.1088/1741-2560/10/2/026012

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 06 April 2023
DOI 10.3389/fnins.2023.1143495

OPEN ACCESS

EDITED BY
Peng Xu,
University of Electronic Science and Technology of China, China

REVIEWED BY
Yingwei Li,
Yanshan University, China
Li Yizhou,
Sichuan University, China

*CORRESPONDENCE
Yong-Qing Zhang
✉ zhangyq@cuit.edu.cn

SPECIALTY SECTION
This article was submitted to Neuroprosthetics, a section of the journal Frontiers in Neuroscience

RECEIVED 13 January 2023
ACCEPTED 13 March 2023
PUBLISHED 06 April 2023

CITATION
Gao D-R, Li J, Wang M-Q, Wang L-T and Zhang Y-Q (2023) Automatic sleep staging of single-channel EEG based on domain adversarial neural networks and domain self-attention. *Front. Neurosci.* 17:1143495. doi: 10.3389/fnins.2023.1143495

Automatic sleep staging of single-channel EEG based on domain adversarial neural networks and domain self-attention

Dong-Rui Gao[1,2], Jing Li[1], Man-Qing Wang[1,2], Lu-Tao Wang[1] and Yong-Qing Zhang[1]*

[1]School of Computer Science, Chengdu University of Information Technology, Chengdu, China, [2]School of Life Sciences and Technology, University of Electronic Science and Technology of China, Chengdu, China

The diagnosis and management of sleep problems depend heavily on sleep staging. For autonomous sleep staging, many data-driven deep learning models have been presented by trying to construct a large-labeled auxiliary sleep dataset and test it by electroencephalograms on different subjects. These approaches suffer a significant setback cause it assumes the training and test data come from the same or similar distribution. However, this is almost impossible in scenario cross-dataset due to inherent domain shift between domains. Unsupervised domain adaption was recently created to address the domain shift issue. However, only a few customized UDA solutions for sleep staging due to two limitations in previous UDA methods. First, the domain classifier does not consider boundaries between classes. Second, they depend on a shared model to align the domain that could miss the information of domains when extracting features. Given those restrictions, we present a novel UDA approach that combines category decision boundaries and domain discriminator to align the distributions of source and target domains. Also, to keep the domain-specific features, we create an unshared attention method. In addition, we investigated effective data augmentation in cross-dataset sleep scenarios. The experimental results on three datasets validate the efficacy of our approach and show that the proposed method is superior to state-of-the-art UDA methods on accuracy and MF1-Score.

KEYWORDS

unsupervised domain adaptation, automatic sleep staging, EEG data, adversarial training, attention mechanism, data augment frontiers

1. Introduction

Sleep appears indispensable in all mammals, and many studies try to unravel the regularity of sleep (Harding et al., 2019; Peng et al., 2020; Bowles et al., 2022). People have been working on this research in the past decade, among which sleep staging has significantly progressed. Sleep staging contributes to detecting sleep disorders, which are usually collected through noninvasive brain-computer interface devices. It keeps track of cerebral cortex activity using a polysomnogram (PSG), a collection of bio-signals including an electrocardiogram (EEG), electromyogram, and an electroencephalogram. According to the American Sleep Society (AASM) (Iber et al., 2007), sleep is divided into three stages: wake (W), rapid eye movement (REM), and non-rapid eye movement (NREM), and N1, N2,

and N3 are the three substages of NREM. In the clinical setting, expert clinicians mainly interpret manually for sleep records. It takes a lot of time and effort for professionals to check, segment, and classify each segment 8–24 h multichannel signals into continuous, fixed-length periods of 30 seconds' epoch.

In recent years, the sleep staging problem has benefited from improvements in machine learning methods (Supratak et al., 2017; Sors et al., 2018; Phan et al., 2019a; Sun et al., 2019), which processed EEG data using various network topologies and properly trained classification models to function effectively in testing. Those methods aim at automating the tedious process. However, many sleep laboratories still rely on the manual scoring of EEG data. There are two main reasons: First, automated sleep staging algorithms still require a large number of labeled data to train the models, which needs to be done manually by sleep technicians or expert clinicians. In this context, attempting to train a deep network on a large labeled source domain and transfer it to the target domain is a good compromise. Yet, this method gives lower performances than expected, which is the second question. The data produced in sleep labs and the publicly available training data differ significantly (Nasiri and Clifford, 2020; Phan et al., 2020a). It can happen for several reasons, including various measuring sites on the skull, different sampling rates of measuring instruments (Azab et al., 2018), or inherent variability between subjects. The training (source) and the test (target) distributions are different, referred to as a domain shift problem.

Unsupervised domain adaptation (UDA) (Ganin and Lempitsky, 2015) has lately shown great potential in enhancing deep learning models when labeled data is scarce. It can solve the two problems above simultaneously. Firstly, it does not need mass-labeled data cause it transfers knowledge from the domain source with rich labels to the target domain with imperfect labels. Secondly, it solves the problem of distribution differences between source and target domains by aligning the distribution of source features and targets. From Figure 1, UDA uses both the labeled source domain and the unlabeled target domain to train the model to perform well on both the source and target domains. UDA is widely used in machine vision (Wang and Deng, 2018) to reduce the discrepancy between the source and the target distributions without utilizing any labels from the target domain. Some research has examined UDA's role in classifying sleep stages thus far. For instance, Chambon et al. (2018) improved the feature transferability between source and target domains using the best transport domain adaptation. Besides, Nasiri and Clifford (2020) used adversarial-based domain adaptation to increase feature transferability.

In UDA, a domain classifier (also known as a discriminator) and a feature generator are used as two players to align allocations in an adversarial way. However, applying these methods to automatic sleep staging continues to have the following drawbacks. First, they disregard the connection between the decision border and the target samples when aligning distributions. In other words, the main task of the generator is to match the distribution between the source and the target. This theory assumes that the classifier can correctly classify these target features because they are consistent with the source samples. So, they do not consider the relationship between the target sample and the decision boundary for distribution alignment. As shown on the right in Figure 1, because the generator is only attempting to make the two distributions close rather than the categorization boundary, it can yield ambiguous features close to the boundary. Secondly, they rely entirely on a common frame to extract features from the source and target domains. This could result in the loss of source and target domain-specific features, which is detrimental to classification tasks in the target domain.

To overcome the challenges above, this article presents a novel framework named Task-Domain Specific Adversarial Network (TDSAN), composed of a feature generator, a domain discriminator, and a dual classifier. It aims to align feature distributions from the source and target domains by combining the classifier's output for the target data and the domain discriminator's output for the domain identifier. Second, we create a domain-specific attention module to maintain source and target-specific features.

Specifically, we train a domain discriminator to predict the input domain and a dual classifier to predict task-specific class labels. We use domain pseudo-labels (i.e., source domain as "1" and target domain as "0") as input. The domain discriminator is trained until it cannot distinguish between the distributions of its training and test domain examples. At the same time, the dual classifier is used to correctly classify the source samples while being trained to find target samples that are located far from the source of support. Because they are not grouped into any classes, samples far from the support do not have traits that can be used to differentiate between them. That is to say. While considering the classifier's output to the target samples, it is instructed to produce desired features close to the support points simultaneously. Therefore, our approach uses a domain discriminator to distinguish the features between samples drawn from the source domain and drawn from the target domain by predicting the domain label and a dual classifier to generate the discriminative features of the target sample because it considers how the decision boundary and the target data relate, and training is adversarial. Additionally, we use domain-specific attention to clean up the extracted features so that each domain keeps its essential characteristics.

The contributions of our paper are summarized as follows: a novel cross-dataset sleep classification framework is proposed that simultaneously changes the categorization boundaries between classes and the conditional distribution between domains. The algorithm adopts a non-shared attention module to keep critical features during adaptation, thereby improving adversarial performance on the target domain. Aiming at the data imbalance in sleep staging, we applied data augment to effectively improve the impact of sample skew on the classification network. Numerous tests show that our TDSAN delivers more excellent cross-domain sleep stage classification performance compared to cutting-edge UDA techniques.

The rest of this paper is organized as follows. Section 2 introduces related work of EEG sleep classification on domain adaptation and describes the proposed model. In Section 3,

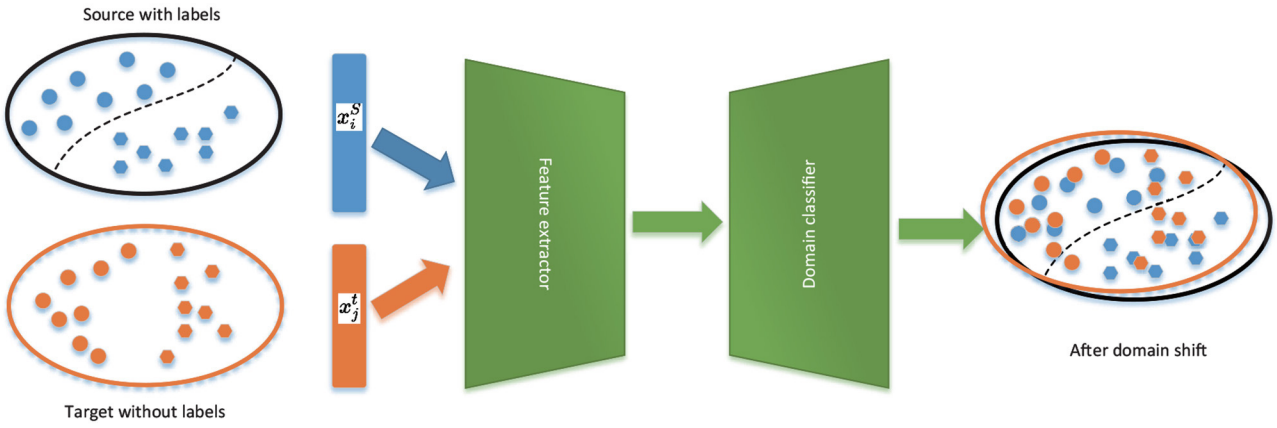


FIGURE 1
Domain adversarial neural networks.

the experimental results are presented and debated. Section 4 concludes this study.

2. Materials and methods

2.1. Related work

The use of single-channel EEG for automated sleep staging has received much attention in the literature. Specifically, deep learning-based techniques (Sors et al., 2018; Kuo and Chen, 2020; Fan et al., 2021; Lee et al., 2021) have made significant progress. These approaches create various network structures to extract characteristics from EEG data and capture temporal dependencies. However, these methods often require enough labeled data to train networks with thousands of parameters. Furthermore, all the above scenarios assume that the training and test data distribution is the same or very comparable, which frequently is not the case because different psychological states or complex equipment noise may cause changes in data distribution. Although these methods have been successful in dealing with complex EEG data, they have limited results in sleep stage classification across domains (Wu et al., 2020) (e.g., cross-datasets and cross-devices) because of domain shift. As a result, numerous studies were told to use transfer learning techniques to address this problem.

There have been a few studies investigating the problem of individual sleep staging using transfer learning (Mikkelsen and De Vos, 2018; Phan et al., 2020b) to increase the specific subject's classification accuracy within a similar dataset. They exclude two nights of test subjects for datasets with two nights of recordings per subject and pretrain the model. Then, the data from the other night is used for evaluation, while the data from the first night is used to fine-tune the model.

The cross-dataset scenario, which involves training a model on data from one dataset and evaluating it on another dataset, has yet to receive much attention. Using a sizable source dataset and another labeled but small target dataset, Phan et al. (2020a) investigated the problem of data variability. Abou Jaoude et al. (2020) also applied a similar transfer learning strategy for extended scalp EEG recordings. They used the larger source dataset to train their model and the smaller target dataset to refine it. Similar to the problem scenario, Phan et al. (2019b) proposed using deep transfer learning to overcome the channel mismatch between the two domains. Abdollahpour et al. (2020) also used this idea to predict sleep staging on fused features on pretrained models. Guillot and Thorey (2021) composed eight heterogeneous sleep staging datasets into a large corpus, which solved the problem of incompatible input data shapes on tasks across datasets and improved the classification accuracy in the target domain. Moreover, even though some studies were not focused on sleep staging, they classify EEG/EMG for fatigue or motor image research and provide some practical examples. For example, Soroushmojdehi et al. (2022) proposes a subject-transfer framework. It uses the information learned from other subjects to make up for the data from the target subject. This article is about a study of hand movement intention identification based on EMG signals. Perry Fordson et al. (2022) also propose a domain adaptation method. It tries to individually treat features from auditory and visual brain regions, which successfully tackles subject-to-subject variations.

Both a labeled target dataset and a sizable corpus of source datasets are necessary for these techniques to fine-tune their models. To solve these problems, UDA strategies that align the traits from several domains with a few annotation data were presented. These methods can be classified as discrepancy-based approaches and adversarial-based approaches. Discrepancy-based methods, such as Maximum Mean Difference (MMD) (Long et al., 2015, 2016) and Correlation Alignment (CORAL) (Sun et al., 2016), strive to reduce the distance measured between the source and target distributions. On the other hand, adversarial-based methods are like Generative Adversarial Networks (GAN) (Goodfellow et al., 2014). This approach trains a domain classifier to predict the input domain and a class classifier to predict task-specific class labels. Both classifiers share the feature extraction layer. The two layers are trained to predict the labels of the source samples correctly as well as to fool the domain classifier. This method is used in current sleep staging works. It was recommended by Zhao et al. (2021) to employ adversarial UDA with a domain discriminator and several classifiers fed from different feature

extractor layers. Nasiri and Clifford (2020) employed adversarial training and strategies for focusing local and global attention to extract transferable personal information. Yoo et al. (2021) used three discriminators for adversarial domain adaptation. One is the global discriminator, and two are local discriminators. The local discriminators will preserve the intrinsic structure of sleep data and reduce local misalignment. Eldele et al. (2022) used a dual classifier for the adversarial domain adaptation framework to improve the accuracy of the decision boundary.

The distribution alignment approaches based on GAN or MMD do not consider the connection between decision boundaries and target samples. Saito et al. (2018) presented an unsupervised adaptation method on the bias of Maximum Classifier Difference (MCD). The technique has a generator module to extract high-level features from the source and target domains. MCD matches distributions by producing representations within similar task-specific boundaries. Dual classifiers are proposed following the same structure. First, the annotated source data is trained to obtain two different classifiers. Second, target samples that are not supported by the source are found. Then, the L1 distance of the probability output is employed to measure the difference between the two classifiers. In this stage, the difference is maximal. It will result in the separation of the two classification boundaries. The generator will relocate the outliers within the target domain. Third, the exact differences are then minimized. Due to the two distributions widely overlapping in stage 2, the target domain is a component of the source domain in this stage. Training deals with the minmax issue by maximizing the target variance and generating a representation that minimizes features. Inspired by this, we set two discriminators. The domain discriminators for domain adaption and dual classifiers for classification boundaries adapting. The classification boundaries are between classes. A generator for generating the minimized differences features. Furthermore, we improve the adversarial training process by maintaining domain-specific features through domain-specific attention.

2.2. Proposed method

Section 2.2.1 briefly introduces the notations and definitions. In Section 2.2.2, we outline the structure and the details. Finally, the whole training process is described in Section 2.2.3.

The definition and notations of EEG-based sleep staging are first briefly discussed in this section. We denote $Xs = \left\{\left(\mathbf{x}_s^i, y_s^i\right)\right\}_{i=1}^{m_s}$ with m_s labeled source data and $X_t = \left\{\left(\mathbf{x}_t^i,\right)\right\}_{i=1}^{m_s}$ with m_t unlabeled samples. In the context of EEG data, x_s^i and $x_t^j \in R^{1*T}$, since EEG data is 2D time series data, symbol 1 means channels (electrodes) and T means how many numbers of timesteps are in each 30-second EEG epochs. Feature extractor F, receives X_s or X_t as input. Domain attention module A receives the output of the feature generator, dual classifier networks $C1$, $C2$, and domain classifier D extract features from A and classify them. The dual classifier networks classify the extracted features into K classes, i.e., output a 5-dimensional logarithmic vector. The domain classifier classifies them into two classes, which are set to be one if the data come from the source domain and set to 0 otherwise. All the class probabilities are obtained through the softmax function. Here, the softmax function's activation of the L1 distance between the probabilities of the two classifiers serves as the discrepancy loss. Following the experience of Saito et al. (2018), we denote the discrepancy loss as follows:

$$d\left(p_1, p_2\right) = \frac{1}{K}\sum_{k=1}^{K}\left|p_{1k} - p_{2k}\right| \tag{1}$$

where p_1k and p_2k represent the probability outputs of p_1 and p_2 of k classes, respectively. We also have an adversarial loss, which tries to deceive the domain classifier by confusing the two data domains.

We aim to acquire a feature generator that minimizes the variance of the target samples.

2.2.1. Network framework

We propose the TDSAN model, which consists of a feature extractor, a domain-specific attention module, a dual classifier, and a domain classifier. The overall model of TDSAN is shown in Figure 2. We first extract the shared feature to generate high-level features representation using both source data and target data. Then, domain-specific attention is put in charge of calculating the relevance of the time sequence. It plays a crucial role in keeping each domain's useful features by fine-tuning the extracted features. The feature extractor and the domain-specific attention together form a generator module. The dual classifier has the following two functions. First, complete the classification task. Second, it is iteratively trained with the generator as a discriminator. It tries to align the distribution on task-specific boundaries. The domain classifier recognizes the domain ID to align domain feature distribution.

2.2.1.1. Feature extractor

The feature extractor consists of three CNN-based convolution blocks, each in 1 dimension. To speed up training and keep CNN Network stable, we use a batch normalization layer. It can reduce internal covariate shift (Perry Fordson et al., 2022). We use leaky ReLU as an activation function and MaxPooling to reduce information redundancy. Given an input source sample $x \in R^{1\cdot T}$, it generates source features through a feature extractor, that is, $F(x) = (fl, \ldots, fl) \in R^{1\cdot l}$ where l is the length of the feature.

2.2.1.2. Domain-specific attention

After convolution, EEG data output is shorter 2D EEG data usually contains the temporal dimension. We do not simply treat the EEG series as a particular image. Based on such consideration, we attempt to extract recessive temporal features from EEG using temporal self-attention mechanisms. In addition, we get shared features after feature extraction. However, different datasets may have different temporal features. Therefore, we use a non-shared attention module to extract domain-specific information. Extracting different domain time temporal also plays a vital role in fine-tuning features. The attention module computes a weighted sum of the features at all locations with the bit of computational cost for each location in the feature space. As a result, each site's features contain intricate details that correspond to fine details in the feature's distant sections.

Inspired by Zhang et al. (2019), as seen in Figure 2, we use a convolutional attention method. The attention operation

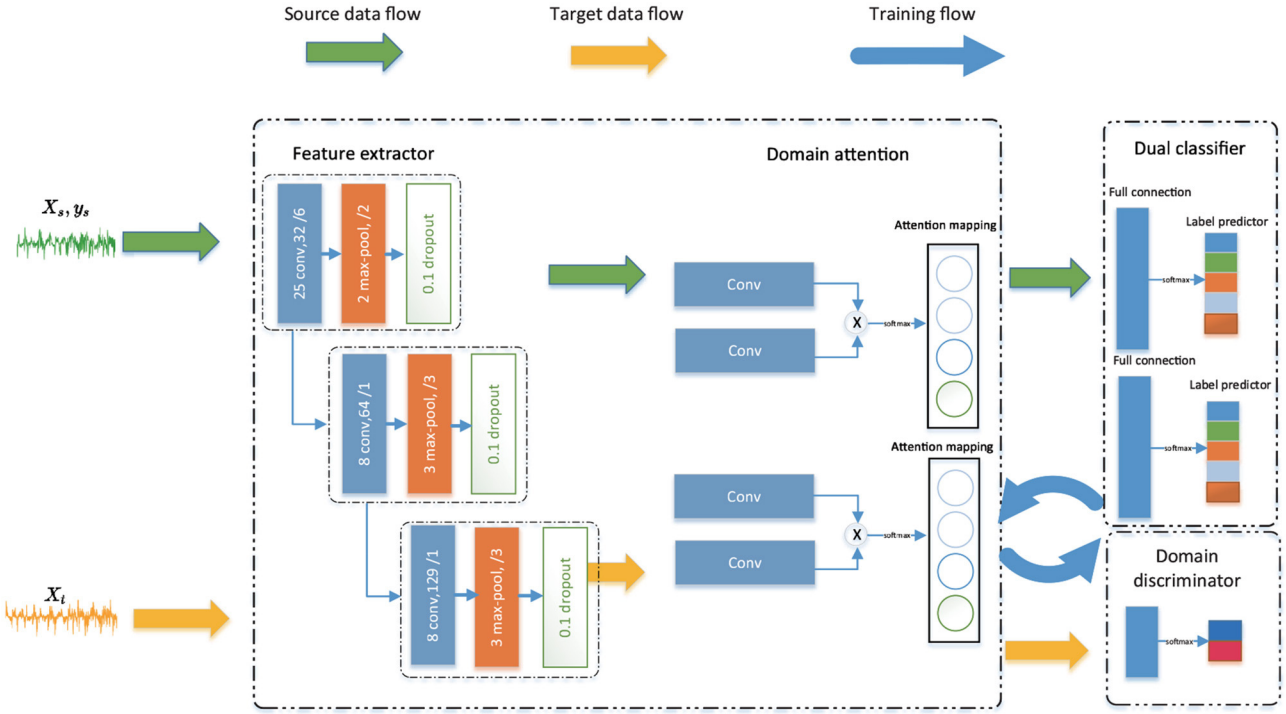


FIGURE 2
Overall architecture of the proposed TDSAN framework.

obtains a feature representation at each location based on two 1D convolutions (e.g., M_1 and M_2). Particularly, $\mathbf{p}_{sx}, \mathbf{p}_{sy} \in \mathbb{R}^d$ are the eigenvalues of positions x and y, which are transferred into $Q_{sx} = M_1\left(\mathbf{p}_{sx}\right)$ and $Q_{sy} = M_2\left(\mathbf{p}_{sy}\right)$. Note that the score is expressed as:

$$\mathcal{W}_{yx} = \frac{\exp\left(Q_{sx}^{\top} Q_{sy}\right)}{\sum_{k=1}^{l} \exp\left(Q_{sk}^{\top} Q_{sy}\right)} \tag{2}$$

where $\mathcal{W}_{yx}$ represents the degree to which the y-*th* location pays attention to the x-*th* position. The output is

$$\mathbf{o}_{sy} = \sum_{x=1}^{l} \mathcal{W}_{yx} \mathbf{p}_{sx} \tag{3}$$

Equations 1 and 2 refer to the attention process as $A(\cdot)$; therefore, $O_s = A_s(F(x_s))$. The same procedure applies to the target domain data stream to train A_t.

2.2.1.3. Dual classifier

The dual classifier consists of three fully connected layers. A dual classifier has two tasks: acting as a classifier for the source data and detecting the target samples outside the source support when we combine two category classifiers. The dual classifiers C1 and C2 first try to maximize the difference for a given target feature to find the target samples away from classification boundaries. Then it minimizes the discrepancy to make them close to the classification boundary. By iterating this process repeatedly, we train a generator that has rare differences with the classification Boundary of the source domain.

2.2.1.4. Domain discriminator

It is widely used in UDA (Tzeng et al., 2017). The main goal is to minimize and regularize the distance between the empirical mapping distributions by training iteratively with the generator. Specifically, We first introduce the domain discriminator by using source samples and target samples with their domain labels. Subsequently, we fix the domain discriminator and train the generator to deceive the discriminator confusing the data domain. We finally get the domain-invariant representations through iterative training.

The details of these four modules are shown in Table 1.

2.2.2. Training steps and loss function

We described the four modules in detail in Section 2.2.1. This section describes the training steps and loss functions to train the entire network. First, the framework of the overall network is to solve a maximize problem. It should be noted that the training is iterative between the generator and the dual classifier and between the generator and domain Discriminator. We solve this problem in four steps.

2.2.2.1. Train the dual classifier

First, the CNN and the attention mechanism network extract the high-level representation. Then we put them in the dual classifier and train it to get two different decision boundaries with source data. We use cross-entropy loss since the data is labeled. It should be noted that C1 and C2 are initialized differently to obtain

TABLE 1 Model parameters of the feature extractor and classifiers.

Modules	Layers	Kernel
Feature extractor	Conv1d	1x25,32
	BatchNorm1d	
	MaxPool1d	1*2,stride 2
	Conv1d	1x8,64
	BatchNorm1d	
	MaxPool1d	1*3,stride 3
	Conv1d	1x8,128
	BatchNorm1d	
	MaxPool1d	1*3,stride 3
Domain attention	/	/
Dual classifier	FC	1,024
	FC	512
	FC	5
Domain discriminator	FC	1,024
	FC	512
	FC	2

two different decision boundaries.

$$\min_{G,c_1,c_2} \mathcal{L}\left(X_s, Y_s\right) \tag{4}$$

$$\mathcal{L}\left(X_s, Y_s\right) = -\mathbb{E}_{(\mathbf{x}_s,y_s)\sim(X_s,Y_s)} \sum_{k=1}^{K} \mathbb{I}_{[k=y_s]} \log p\left(\boldsymbol{y} \mid \mathcal{O}\left(\boldsymbol{x}_s\right)\right) \tag{5}$$

2.2.2.2. Fixed generator, training domain discriminator

In this step, the source domain ID is one, and the target domain ID is 0. It aims to minimize the domain confusion loss for source and target samples on a fixed G. This is a supervised learning process.

$$\mathcal{L} = -\sum_{i=1}^{N} y^{(i)} \log \hat{y}^{(i)} + \left(1 - y^{(i)}\right) \log\left(1 - \hat{y}^{(i)}\right) \tag{6}$$

2.2.2.3. Fixed generator, dual training classifier

In this step, we train classifiers $C1$ and $C2$ to minimize discrepancy on target samples for a fixed generator G. They can detect target samples far from source support and force them to relocate to the corresponding category.

$$\mathcal{L}_{\text{adv}}\left(X_t\right) = \mathbb{E}_{\mathbf{x_t}\sim X_t}\left[d\left(p_1\left(\mathbf{y} \mid \mathcal{O}\left(\mathbf{x_t}\right)\right), p_2\left(\mathbf{y} \mid \mathcal{O}\left(\mathbf{x_t}\right)\right)\right)\right] \tag{7}$$

2.2.2.4. Fixed dual classifier and domain discriminator, training generator

In this step, we train generator G to maximize discrepancy on target samples. It identifies the target samples that the source's support has eliminated.

$$\max_{c1,c2} \mathcal{L}_{\text{adv}}\left(X_t\right) \tag{8}$$

In our system, these three phases are repeated. Our primary focus is on adversarially training classifiers and generators to identify source samples and confusing domain distribution. Algorithm 1 summarizes the complex algorithm of TDSAN.

Input: source data Xs, Ys, target data Xt
for *epoch in maxepoch* **do**
  for *each mini-batch* **do**
    pretrain $C1$, $C2$ with source Xs, Ys;
    reduce $\mathcal{O}(f_s)$ and $\mathcal{O}(f_t)$;
    for *p in F)* **do**
      train D for $\mathcal{O}(f_s)$ and $\mathcal{O}(f_t)$;
    end
    train $C1$, $C2$ with source Xs, Ys;
    maximize discrepancy using the class probability of target f_t;
    reduce task-specific variance based on the class probability of target f_t;
  end
end

Algorithm 1. Training Procedure for TDSAN

3. Results

3.1. Data

We evaluated the proposed framework on three datasets, including two public and one private dataset, namely sleep-EDF-SC (EDF), sleep-EDF-ST(ST), and self-collection datasets. Before downsampling, a summary of the three datasets above is displayed in Table 2.

3.1.1. Public datasets

The sleep-EDFx dataset (Goldberger et al., 2000) is made up of 42 subjects' 61 polysomnographic (PSG) data and corresponding hypnograms (annotations by sleep experts). European data format (EDF) is used to store PSG records, while EDF+ is used to store hypnograms. A horizontal EOG channel, a sub-mental chin EMG, two EEG channels, Fpz-Cz and Pz-Oz, an event marker, and an EOG channel are all included in each record. At 100 Hz, the EEG signal is captured. The annotations for each stage of sleep in the hypnogram file are W (wake), R (REM), 1 (N1), 2 (N2), 3 (N3), 4 (N4), M (movement time), and not scored (denoted as?). 42 participants from two separate groups—the Sleep Cassette (SC) group and the Sleep Telemetry (ST) group—were employed in the study. The ST group consists of the remaining 22 participants with modest sleep problems. In contrast, the SC group consists of the remaining 20 healthy subjects who are not taking any medicine. Every SC subject in the EDFx dataset has two nights' sleep records, except one, which only has one. 22 sick participants were recorded for one night of sleep by the ST group. Based on the experience of

TABLE 2 A brief description about the datasets.

Dataset	Subject	Recordings	Sample rate	Channel	Scoring
EDF	20	39	100	Fpz-Cz	R&K
ST	22	44	100	Fpz-Cz	R&K
self-collection	6	30	250	C4-A1	AASM

Supratak et al. (2017) and Eldele et al. (2021), we evaluate our model on the channel Fpz-Cz.

3.1.2. Self-collection datasets

This study conducted 30 nights of polysomnography trials following clinical recommendations and the AASM guideline. There are six subjects, each under observation for five nights. The subjects were all young, healthy adults between the ages of 20 and 24, with a male-to-female ratio of 1:0.67. Each participant voluntarily agreed to take part in this sleep study. They had to bathe and do their hair before the investigation to keep their heads tidy. The time ranged from 23:00 to 07:30, exceeding 8 h. All participants ensured they were in good health, had no medical history, and hadn't done any strenuous exercise the hour before the sleep experiment began. The 3-channel EEG, 2-channel EOG, 1-channel EMG, and 1-channel ECG signals were accurately acquired and stored. The system gain is 24, and the sampling rate is 250 Hz. In addition, conductive gel paste is applied to all the gold-plated disc electrodes used in EEG electrodes for signal acquisition. Patch electrodes are used in EOG, EMG, and ECG electrodes. The subject's dormitory bed served as the site for the entire sleep experiment, and the AASM suggested all electrode implantation settings. Three EEG channels, i.e., C4-A1, F4-A1, and O2-A1, were combined to form the electrode title, which is now only C4, F4, and O2. Then, EOG-R and EOG-L are combined to form REOG and LEOG from the two EOG channels. We evaluated our model on channel C4-A1.

3.2. Data preprocessing

In the 30 s/epoch time series, the filtered data were separated into non-overlapping pieces according to the AASM staging criteria.

- Epochs classified as being in motion, artifacts, or unknown were eliminated.
- To meet the AASM norm, sleep stages S3 and S4 were combined into a single N3 stage.
- Only the first and last 30 mins of wake time were included.
- Downsampling the data with a sampling frequency higher than 100 Hz, and the length of a single epoch is 30 s × 100 Hz (T = 3,000).
- Cutoff frequency design: based on high-pass and low-pass filters (0.3-35 Hz) to reduce the noise.

3.3. Experimental settings

We used the macro-averaged F1-score and the classification accuracy (ACC) to assess the proposed performance. The metrics are denoted as follows:

$$ACC = \frac{\sum_{i=1}^{K} TP_i}{M} \tag{9}$$

$$MF1 = \frac{1}{K}\sum_{i=1}^{K}\frac{2 \times \text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{10}$$

where $\text{Precision} = \frac{TP}{TP+FP}$, $\text{Recall} = \frac{TP}{TP+FN}$. TP, FP, TN, and FN denote True Positive, False Positive, True Negative, and False Negative, respectively. The whole sample number is M, and the total class number is K. The experiment was model initialized from various random seeds and repeated five times. The average final result (ACC and MF1) was then presented with the standard deviation.

We divide the experimental data into 80% and 20% for training and testing. We do not disrupt the order of epochs of subjects so that domain-attention-specific can capture the relationship between different sleep stages. We employ the Adam optimizer with a batch size of 128 and a learning rate of 1e−3. We did not fine-tune these hyperparameters for a fair comparison. Another hyperparameter is n, which represents the number of times this operation is repeated for the same mini-batch. This value represents the trade-off between the generator and the classifier. All experiments are done by pytorch1.12 on an NVIDIA GeForce RTX 2080 Ti GPU.

3.4. Baselines

We analyze our suggested TDSAN by contrasting it with different baselines. We started by including the Direct Transfer (DT) findings from DeepSleepNet's (Supratak et al., 2017) three sleep staging methods. In addition, we adopted four state-of-the-art baselines based on adversarial domain adaptation (DA). We briefly describe these baselines:

- DANN (Eldele et al., 2021): It simultaneously trains a feature extractor and a domain classifier using a gradient reversal layer (GRL) to remove the gradient of the domain classifier.
- ADDA (Ganin et al., 2016): It accomplishes a comparable task to DANN but reverses the labels.
- CDAN (Tzeng et al., 2017): Minimize the cross-covariance between feature representations and classifier predictions.

TABLE 3 Comparison of different baselines on ACC. Bold: the best results; Underlined: the second best results.

	Baselines	EDF→ST	EDF→SC	ST→EDF	ST→SC	SC→EDF	SC→ST	ACC
DT	DeepSleepNet	72.34	61.53	68.35	50.23	62.75	49.85	60.84
DA	DANN	73.70	65.98	64.23	58.94	65.93	67.53	66.05
	ADDA	73.72	77.40	64.14	65.93	58.43	68.65	68.71
	CDAN	**76.42**	67.36	66.68	70.36	62.89	**70.06**	68.96
	ADAST	72.85	70.41	71.23	70.06	**68.41**	69.88	70.47
	TDSAN(ours)	73.97	**79.32**	**73.90**	**70.89**	65.12	68.19	**71.89**

TABLE 4 Comparison of different baselines on MF1. Bold: the best results; Underlined: the second best results.

	Baselines	EDF→ST	EDF→SC	ST→EDF	ST→SC	SC→EDF	SC→ST	MF1
DT	DeepSleepNet	56.58	45.88	53.86	44.83	55.32	40.30	50.12
DA	DANN	61.35	55.79	53.49	50.63	55.19	56.63	55.51
	ADDA	59.69	55.5	55.69	54.79	48.89	**63.86**	56.40
	CDAN	**63.06**	54.73	53.91	64.01	52.61	59.71	57.95
	ADAST	59.86	**61.89**	**60.72**	**64.29**	**56.65**	60.33	**60.62**
	TDSAN (ours)	60.00	60.66	59.91	58.95	53.29	55.30	58.01

- ADAST (Yoo et al., 2021): It uses dual classification on top of domain obfuscation to consider class-conditional distributions.

3.5. Results

Tables 3, 4 show the comparison results among various methods. It suggested that the direct transfer results are usually the worst. This result indicated that the domain shift issue has a significant impact and needs to be handled independently. The findings of the other 4 DA baselines confirm the need for domain adaptation to overcome the domain shift issue.

It is important to emphasize that we use the proposed backbone feature extractor on four baseline methods except DeepSleepNet to ensure a fair comparison. In this setting, we note that methods that consider class-conditional distributions: such as CDAN and ADAST, outperform the globally aligned source and target domains, namely DANN and ADDA. This shows that taking class distribution into account is crucial to improving classification performance on the target domain, mainly when dealing with imbalanced sleep data. Our proposed TDSAN outperforms all baselines in accuracy in four of the six cross-domain scenarios and achieves the second-best average score among all baselines on *MF1-Score*.

We consider possible reasons: First, by iteratively training the generator and classifier, we obtain a feature extractor that can extract the smallest difference between the source and target domains. At the same time, the task-specific classifier fully considers the relationship between the target task and the decision boundary. Second, performance is enhanced by TDSAN because it uses a non-shared attention module to preserve domain-specific features.

In Tables 3, 4, we also find essential clues on different cross-dataset situations. In SC→EDF and SC→ST, they are generally lower than other scenarios in various cross-domain scenarios. This may be because the dataset is too small, and the classification performance of the classifier on the source domain data is poor, which is insufficient to correct the do-main offset results. In addition, we also observed that the transfer results of datasets between SC and ST are relatively poor. We observed from the characteristics of datasets that EDF and SC datasets are the sleep EEG data of healthy people, while ST data sets are collected from people with mild sleep disorders. In addition, the acquisition channels and frequencies of ST and SC are different, so they are a relatively remote domain from EDF and SC. These findings suggest that adapting to distant domains is still exceedingly difficult.

3.6. Data augmentation

Each stage's length varies for a sleep recording. Mainly, stage N2 makes up between 45% and 55% of the entire sleep time and contributes to the majority class. N1, on the other hand, only makes up roughly 2%–5% [36]. Every sleep dataset that is accessible has this problem, including public datasets we can retrieve and our self-collection dataset. As mentioned in many studies (Tsinalis et al., 2016a,b; Supratak et al., 2017; Sun et al., 2019), class imbalance may hinder the classifier's performance, limiting the improvement of automatic sleep staging algorithms. Following the suggestion of Ko et al. (2021), we use the sliding window method to augment the N1 stage data. The window size is 30s, and the step size is 25s. We are also experimenting with six cross-domain scenarios to see if augmentation affects the classification results. Table 5 shows the comparison before and after data augmentation on six cross-domain scenarios with three indexes, precision, recall, and

TABLE 5 Classification performance before and after data augment.

Cross-domain	Slide windows	Sleep score	W (%)	N1 (%)	N2 (%)	N3 (%)	REM (%)	ACC (%)	MF1 (%)
EDF→ST	Before	Precision	50.00	34.42	84.24	97.26	66.55	73.97	66.49
		Recall	71.95	43.44	81.83	27.95	88.94	73.97	62.82
		F1-score	59.00	38.41	83.02	43.43	76.13	73.97	60.00
		Support	82	122	1156	254	434	73.97	2048
	After	Precision	80.17	60.00	94.36	96.19	61.70	75.12	68.10
		Recall	96.04	28.68	95.81	55.04	73.32	75.12	53.03
		F1-score	87.39	38.81	95.08	70.02	67	75.12	55.87
		Support	82	205	1156	254	434	75.12	2131
EDF→SC	Before	Precision	37.25	15.79	87.32	92.53	62.46	79.32	59.07
		Recall	66.67	11.11	84.22	82.29	78.60	79.32	64.58
		F1-score	47.80	13.04	85.74	87.11	69.61	79.32	60.66
		Support	323	191	925	201	408	79.32	2048
	After	Precision	56.06	22.22	85.55	91.55	61.08	80.26	63.30
		Recall	64.91	7.41	86.98	78.47	83.39	80.26	64.23
		F1-score	60.16	11.11	86.26	84.51	70.51	80.26	62.51
		Support	323	215	925	201	408	80.26	2072
ST→EDF	Before	Precision	62.18	38.24	88.77	68.77	58.63	73.90	64.92
		Recall	30.03	6.81	86.27	97.51	85.29	73.90	61.18
		F1-score	40.50	11.56	87.50	80.66	67.33	73.90	59.91
		Support	101	77	1170	36	517	73.90	1901
	After	Precision	90.67	40.63	83.95	82.99	64.03	77.94	72.45
		Recall	54.40	12.68	91.71	78.74	78.34	77.94	63.17
		F1-score	68.00	19.33	87.66	80.81	70.47	77.94	65.25
		Support	101	136	1170	36	517	77.94	1960
ST→SC	Before	Precision	62.18	38.24	88.77	68.77	61.63	70.90	61.92
		Recall	30.03	6.81	86.27	97.51	85.29	70.90	61.18
		F1-score	40.50	11.56	87.50	80.66	64.33	70.90	58.91
		Support	323	191	925	201	408	70.90	2048
	After	Precision	55.93	47.83	83.86	98.85	47.87	75.42	64.25
		Recall	76.74	8.21	91.21	80.54	94.07	75.42	60.83
		F1-score	64.71	14.01	86.37	88.32	62.31	75.42	58.39
		Support	323	215	925	201	408	75.42	2072
SC→EDF	Before	Precision	100.00	0.00	77.22	96.15	37.27	65.12	62.13
		Recall	19.02	0.00	86.52	87.04	97.35	65.12	56.99
		F1-score	30.59	0.00	81.60	91.37	53.91	65.12	53.29
		Support	101	77	1170	36	517	65.12	1901
	After	Precision	97.59	4.35	80.87	96.15	49.28	72.28	65.65
		Recall	43.67	1.95	92.30	91.88	90.00	72.28	63.96
		F1-score	60.34	2.70	86.21	93.97	63.68	72.28	61.38
		Support	101	136	1170	36	517	72.28	1960
SC→ST	Before	Precision	98.70	0.00	79.69	98.37	45.80	68.19	63.51
		Recall	35.49	0.00	81.22	81.82	99.12	68.19	56.53

(Continued)

TABLE 5 (Continued)

Cross-domain	Slide windows	Sleep score	W (%)	N1 (%)	N2 (%)	N3 (%)	REM (%)	ACC (%)	MF1 (%)
		F1-score	37.93	0.00	80.45	89.33	58.80	68.19	55.30
		Support	82	122	1156	254	434	68.19	2048
	After	Precision	81.82	0.00	92.94	47.39	68.54	73.44	58.14
		Recall	21.95	0.00	72.84	100.00	89.86	73.44	56.93
		F1-score	34.62	0.00	81.67	64.30	77.77	73.44	51.67
		Support	82	205	1156	254	434	73.44	2131

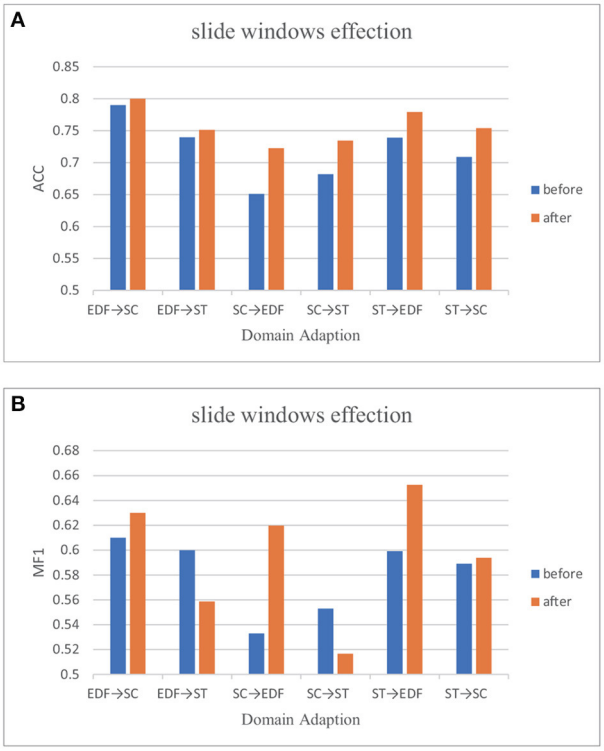


FIGURE 3
Classification performance before and after data augment. **(A)** Performances on ACC before and after sliding window. **(B)** Performances on MF1 before and after sliding window.

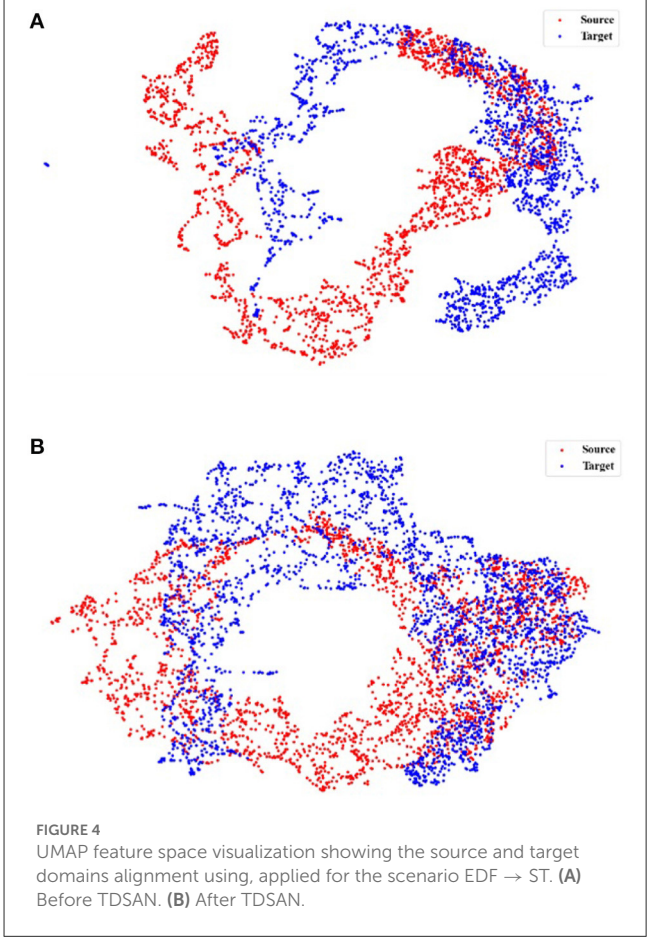


FIGURE 4
UMAP feature space visualization showing the source and target domains alignment using, applied for the scenario EDF → ST. **(A)** Before TDSAN. **(B)** After TDSAN.

MF1_score in detail. For easier reading, we use Figure 3 to present these results.

The result suggested sample skew affects the performance of the classification network. In domain adaption, data augmentation is still optional when we can not get enough EEG data.

3.7. Feature visualization

Further, we use UMAP to visualize the feature representations learned to make the Comparison more intuitive.

Initially, we investigate the alignment quality. Figure 4 shows the alignment between the source and target domains in the EDF→ST scenario, where Figure 4A shows the feature distribution of the source and target domains before training. Our TDSAN framework alignment is shown in Figure 4B. The blue dots in these pictures offer the target domain, whereas the red dots show the source domain. The source and target domain feature distributions have significant differences before alignment. After alignment, the feature overlap between the source and target domains increases.

Additionally, after the alignment in Figure 4, we investigate the target domain classification performance under the above scenarios. In particular, Figure 5A is the class distribution of the source and target domains before training, and Figure 5B is the distribution after our alignment. The symbol (·) represents the

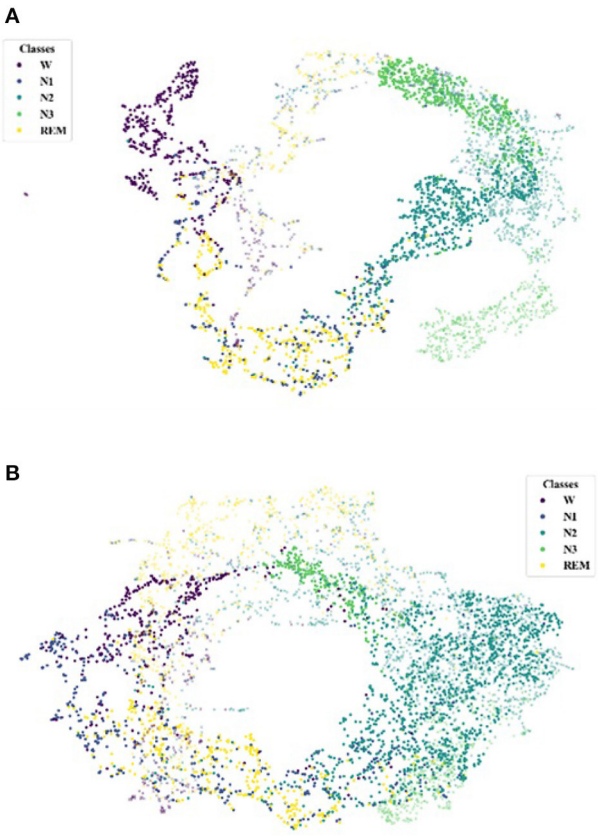


FIGURE 5
UMAP feature space visualization showing the target domains classification performance, applied for the scenario EDF → ST. **(A)** Before TDSAN. **(B)** After TDSAN.

classification of the source domain data, and the emblem (×) represents the classification of the target domain. We note that Figure 5A shows that the type of source and target domains differs considerably. And after training, the classification of the source and target domains is much higher. This is achieved through an iterative self-training strategy.

4. Conclusion

This paper proposes a novel UDA method, TDSAN, to address the sleep EEG staging scores on unlabeled data. TDSAN is an adversarial learning method that uses a specific classifier as a discriminator whose target samples are remote from the source support detected. To trick the classifier, the feature generator masters to produce target features close to supports. The generator will prevent creating target features close to class boundaries since it incorporates feedback from task-specific classifiers. Meanwhile, a non-shared attention mechanism preserves domain-specific features, which can capture the relationship between different sleep stages. Experiments show that we can achieve the same accuracy on unlabeled sleep data as on labeled data.

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

D-RG and JL: conceptualization and writing—review and editing. JL: methodology, investigation, data curation, and writing—original draft preparation. D-RG, JL, and M-QW: visualization and validation. L-TW: formal analysis and resources. Y-QZ: supervision and writing—review and editing. All authors have read and approved the final draft of the manuscript.

Funding

This work is supported by the National Natural Science Foundation of China under Grant No. 62272067, the Sichuan Science and Technology Program under Grant Nos. 2023NSFSC0499 and 2023YFG0018, the LOST 2030 Brain Project No. 2022ZD0208500, and the Scientific Research Foundation of Chengdu University of Information Technology under Grant Nos. KYQN202208 and KYQN202206.

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

Abdollahpour, M., Rezaii, T. Y., Farzamnia, A., and Saad, I. (2020). Transfer learning convolutional neural network for sleep stage classification using two-stage data fusion framework. *IEEE Access* 8, 180618–180632. doi: 10.1109/ACCESS.2020.3027289

Abou Jaoude, M., Sun, H., Pellerin, K. R., Pavlova, M., Sarkis, R. A., Cash, S. S., et al. (2020). Expert-level automated sleep staging of long-term scalp electroencephalography recordings using deep learning. *Sleep* 43, zsaa112. doi: 10.1093/sleep/zsaa112

Azab, A. M., Toth, J., Mihaylova, L. S., and Arvaneh, M. (2018). "A review on transfer learning approaches in brain-computer interface," in *Processing and Machine Learning for Brain-Machine Interfaces*. p. 81–101. doi: 10.1049/PBCE114E_ch5

Bowles, N. P., Thosar, S. S., Butler, M. P., Clemons, N. A., Robinson, L. D., Ordaz, O. H., et al. (2022). The circadian system modulates the cortisol awakening response in humans. *Front. Neurosci*. 16, 995452. doi: 10.3389/fnins.2022.995452

Chambon, S., Galtier, M. N., and Gramfort, A. (2018). "Domain adaptation with optimal transport improves eeg sleep stage classifiers," in *2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, 1–4. IEEE. doi: 10.1109/PRNI.2018.8423957

Eldele, E., Chen, Z., Liu, C., Wu, M., Kwoh, C. -K., Li, X., et al. (2021). "An attention-based deep learning approach for sleep stage classification with single-channel EEG," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 29* (IEEE), 809–818. doi: 10.1109/TNSRE.2021.3076234

Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C.-K., Li, X., et al. (2022). Adast: attentive cross-domain eeg-based sleep staging framework with iterative self-training. *IEEE Trans. Emerg. Topics Comput. Intell*. doi: 10.1109/TETCI.2022.3189695

Fan, J., Sun, C., Long, M., Chen, C., and Chen, W. (2021). Eognet: a novel deep learning model for sleep stage classification based on single-channel eog signal. *Front. Neurosci*. 15, 573194. doi: 10.3389/fnins.2021.573194

Ganin, Y., and Lempitsky, V. (2015). "Unsupervised domain adaptation by backpropagation," in *International Conference on Machine Learning*, 1180–1189. PMLR.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., et al. (2016). Domain-adversarial training of neural networks. *J. Mach. Learn. Res*. 17, 2096–2030. doi: 10.48550/arXiv.1409.7495

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. doi: 10.1161/01.CIR.101.23.e215

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. *Adv. Neural Inform. Process. Syst*. 63, 139–144. doi: 10.48550/arXiv.1406.2661

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

Harding, E. C., Franks, N. P., and Wisden, W. (2019). The temperature dependence of sleep. *Front. Neurosci*. 13, 336. doi: 10.3389/fnins.2019.00336

Iber, C., Ancoli-Israel, S., Chesson, A., and Quan, S. (2007). *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Westchester, IL: American Academy of Sleep Medicine.

Ko, W., Jeon, E., Jeong, S., Phyo, J., and Suk, H.-I. (2021). A survey on deep learning-based short/zero-calibration approaches for eeg-based brain-computer interfaces. *Front. Human Neurosci*. 15, 643386. doi: 10.3389/fnhum.2021.643386

Kuo, C.-E., and Chen, G.-T. (2020). Automatic sleep staging based on a hybrid stacked lstm neural network: verification using large-scale dataset. *IEEE Access* 8, 111837–111849. doi: 10.1109/ACCESS.2020.3002548

Lee, C.-H., Kim, H.-J., Heo, J.-W., Kim, H., and Kim, D.-J. (2021). "Improving sleep stage classification performance by single-channel eeg data augmentation via spectral band blending," in *2021 9th International Winter Conference on Brain-Computer Interface (BCI)*, 1–5. IEEE.

Long, M., Cao, Y., Wang, J., and Jordan, M. (2015). "Learning transferable features with deep adaptation networks," in *International Conference on Machine Learning*, 97–105. PMLR.

Long, M., Zhu, H., Wang, J., and Jordan, M. I. (2016). "Unsupervised domain adaptation with residual transfer networks," in *Advances in Neural Information Processing Systems 29 (NIPS 2016)*.

Mikkelsen, K., and De Vos, M. (2018). Personalizing deep learning models for automatic sleep staging. *arXiv [preprint] arXiv*:1801.02645.

Nasiri, S., and Clifford, G. D. (2020). "Attentive adversarial network for large-scale sleep staging," in *Machine Learning for Healthcare Conference*, 457–478. PMLR.

Peng, Z., Dai, C., Ba, Y., Zhang, L., Shao, Y., and Tian, J. (2020). Effect of sleep deprivation on the working memory-related n2-p3 components of the event-related potential waveform. *Front. Neurosci*. 14, 469. doi: 10.3389/fnins.2020.00469

Perry Fordson, H., Xing, X., Guo, K., and Xu, X. (2022). Not all electrode channels are needed: knowledge transfer from only stimulated brain regions for eeg emotion recognition. *Front. Neurosci*. 16, 865201. doi: 10.3389/fnins.2022.865201

Phan, H., Andreotti, F., Cooray, N., Chén, O. Y., and De Vos, M. (2019a). 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., Chén, O. Y., Koch, P., Lu, Z., McLoughlin, I., Mertins, A., et al. (2020a). Towards more accurate automatic sleep staging via deep transfer learning. *IEEE Trans. Biomed. Eng*. 68, 1787–1798. doi: 10.1109/TBME.2020.3020381

Phan, H., Chén, O. Y., Koch, P., Mertins, A., and De Vos, M. (2019b). "Deep transfer learning for single-channel automatic sleep staging with channel mismatch," in *2019 27th European Signal Processing Conference (EUSIPCO)*, 1–5. IEEE.

Phan, H., Mikkelsen, K., Chén, O. Y., Koch, P., Mertins, A., Kidmose, P., et al. (2020b). Personalized automatic sleep staging with single-night data: a pilot study with kullback-leibler divergence regularization. *Physiol. Measurement* 41, 064004. doi: 10.1088/1361-6579/ab921e

Saito, K., Watanabe, K., Ushiku, Y., and Harada, T. (2018). "Maximum classifier discrepancy for unsupervised domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3723–3732.

Soroushmojdehi, R., Javadzadeh, S., Pedrocchi, A., and Gandolla, M. (2022). Transfer learning in hand movement intention detection based on surface electromyography signals. *Front. Neurosci*. 16, 977328. doi: 10.3389/fnins.2022.977328

Sors, A., Bonnet, S., Mirek, S., Vercueil, L., and Payen, J.-F. (2018). A convolutional neural network for sleep stage scoring from raw single-channel eeg. *Biomed. Signal Process. Control* 42, 107–114. doi: 10.1016/j.bspc.2017.12.001

Sun, B., Feng, J., and Saenko, K. (2016). "Return of frustratingly easy domain adaptation," in *Proceedings of the AAAI Conference on Artificial Intelligence*. doi: 10.1609/aaai.v30i1.10306

Sun, C., Fan, J., Chen, C., Li, W., and Chen, W. (2019). A two-stage neural network for sleep stage classification based on feature learning, sequence learning, and data augmentation. *IEEE Access* 7, 109386–109397. doi: 10.1109/ACCESS.2019.2933814

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

Tsinalis, O., Matthews, P. M., and Guo, Y. (2016a). Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders. *Ann. Biomed. Eng*. 44, 1587–1597. doi: 10.1007/s10439-015-1444-y

Tsinalis, O., Matthews, P. M., Guo, Y., and Zafeiriou, S. (2016b). Automatic sleep stage scoring with single-channel eeg using convolutional neural networks. *arXiv* [preprint] *arXiv*:1610.01683.

Tzeng, E., Hoffman, J., Saenko, K., and Darrell, T. (2017). "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7167–7176.

Wang, M., and Deng, W. (2018). Deep visual domain adaptation: a survey. *Neurocomputing* 312, 135–153. doi: 10.1016/j.neucom.2018.05.083

Wu, D., Xu, Y., and Lu, B.-L. (2020). Transfer learning for EEG-based brain-computer interfaces: a review of progress made since 2016. *IEEE Trans. Cogn. Dev. Syst*. 14, 4–19. doi: 10.1109/TCDS.2020.3007453

Yoo, C., Lee, H. W., and Kang, J.-W. (2021). Transferring structured knowledge in unsupervised domain adaptation of a sleep staging network. *IEEE J. Biomed. Health Inform*. 26, 1273–1284. doi: 10.1109/JBHI.2021.3103614

Zhang, H., Goodfellow, I., Metaxas, D., and Odena, A. (2019). "Self-attention generative adversarial networks," in *International Conference on Machine Learning*, 7354–7363. PMLR.

Zhao, R., Xia, Y., and Zhang, Y. (2021). Unsupervised sleep staging system based on domain adaptation. *Biomed. Signal Process. Control* 69, 102937. doi: 10.1016/j.bspc.2021.102937

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 17 April 2023
DOI 10.3389/fnins.2023.1135995

OPEN ACCESS

EDITED BY
Peng Xu,
University of Electronic Science and Technology of China,
China

REVIEWED BY
Liang Gong,
Chengdu Second People's Hospital,
China
Xiufeng Xu,
The First Affiliated Hospital of Kunming Medical University,
China

*CORRESPONDENCE
Yi Guo
✉ xuanyi_guo@163.com

SPECIALTY SECTION
This article was submitted to Neuroprosthetics, a section of the journal Frontiers in Neuroscience

RECEIVED 02 January 2023
ACCEPTED 28 March 2023
PUBLISHED 17 April 2023

CITATION
Zhu L, Dang G, Wu W, Zhou J, Shi X, Su X, Ren H, Pei Z, Lan X, Lian C, Xie P and Guo Y (2023) Functional connectivity changes are correlated with sleep improvement in chronic insomnia patients after rTMS treatment. *Front. Neurosci.* 17:1135995. doi: 10.3389/fnins.2023.1135995

Functional connectivity changes are correlated with sleep improvement in chronic insomnia patients after rTMS treatment

Lin Zhu[1], Ge Dang[1], Wei Wu[2], Junhong Zhou[3], Xue Shi[1], Xiaolin Su[1], Huixia Ren[4], Zian Pei[5], Xiaoyong Lan[5], Chongyuan Lian[5], Peng Xie[6] and Yi Guo[1,5]*

[1]Department of Neurology, Shenzhen People's Hospital, The Second Clinical Medical College, Jinan University, The First Affiliated Hospital, Southern University of Science and Technology, Shenzhen, Guangdong, China, [2]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, United States, [3]Hebrew Seniorlife, Hinda and Arthur Marcus Institute for Aging Research, Harvard Medical School, Boston, MA, United States, [4]Department of Geriatrics, Shenzhen People's Hospital, The Second Clinical Medical College, Jinan University, The First Affiliated Hospital, Southern University of Science and Technology, Shenzhen, Guangdong, China, [5]Shenzhen Bay Laboratory, Shenzhen, Guangdong, China, [6]NHC Key Laboratory of Diagnosis and Treatment on Brain Functional Diseases, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China

Background: Repetitive transcranial magnetic stimulation (rTMS) has been increasingly used as a treatment modality for chronic insomnia disorder (CID). However, our understanding of the mechanisms underlying the efficacy of rTMS is limited.

Objective: This study aimed to investigate rTMS-induced alterations in resting-state functional connectivity and to find potential connectivity biomarkers for predicting and tracking clinical outcomes after rTMS.

Methods: Thirty-seven patients with CID received a 10-session low frequency rTMS treatment applied to the right dorsolateral prefrontal cortex. Before and after treatment, the patients underwent resting-state electroencephalography recordings and a sleep quality assessment using the Pittsburgh Sleep Quality Index (PSQI).

Results: After treatment, rTMS significantly increased the connectivity of 34 connectomes in the lower alpha frequency band (8–10Hz). Additionally, alterations in functional connectivity between the left insula and the left inferior eye junction, as well as between the left insula and medial prefrontal cortex, were associated with a decrease in PSQI score. Further, the correlation between the functional connectivity and PSQI persisted 1month after the completion of rTMS as evidenced by subsequent electroencephalography (EEG) recordings and the PSQI assessment.

Conclusion: Based on these results, we established a link between alterations in functional connectivity and clinical outcomes of rTMS, which suggested that EEG-derived functional connectivity changes were associated with clinical improvement of rTMS in treating CID. These findings provide preliminary evidence that rTMS may improve insomnia symptoms by modifying functional connectivity, which can be used to inform prospective clinical trials and potentially for treatment optimization.

KEYWORDS

chronic insomnia disorder, rTMS, EEG, functional connectivity, DLPFC

Introduction

Insomnia is the most prevalent sleep disorder and is associated with difficulties in initiating or maintaining sleep, as well as a decline in daytime performance and cognitive impairment, leading to many health issues (Spiegelhalder et al., 2015). Hypnotics are often prescribed as a first-line treatment for acute insomnia (Glass et al., 2005); however, the long-term use of hypnotics has adverse effects and may confer a risk of dependence. Cognitive behavioral therapy (CBT) is an alternative to chronic pharmacological insomnia treatment. Although there is evidence for long-lasting improvement in sleep after CBT, the strict time commitment and insufficient number of qualified clinicians to employ it, limit its use in clinical practice (Morin and Benca, 2012). Therefore, better treatment options for chronic insomnia, a long-term pattern of difficulty sleep, are urgently needed.

Insomnia is rarely an isolated psychobiological disorder and is associated with measurable aberration in functional brain mechanisms (Riemann et al., 2010). Neuroimaging studies, such as electroencephalography (EEG) or functional magnetic resonance imaging (fMRI), showed that patients with insomnia have abnormal connectivity patterns in emotional circuits (Huang et al., 2012), salience networks (Chen et al., 2014), and brain network topology in general (Li et al., 2018). Moreover, previous studies have suggested that increased cortical excitability was associated with chronic insomnia (Lanza et al., 2015; Ly et al., 2016).

Repetitive transcranial magnetic stimulation (rTMS) can noninvasively modulate cortical activity by delivering a sequence of magnetic pulses. Normally, low frequency (<1 Hz) is thought to inhibit, and high frequency (≥5 Hz) to facilitate motor cortical excitability (Fitzgerald et al., 2006). The inhibitory effect of low frequency (i.e., 1 Hz) rTMS on cortical excitability has therefore led to it being increasingly considered for the treatment of insomnia disorders.

Stimulation site is another essential factor in the field of clinical application of rTMS. Dorsolateral prefrontal cortex (DLPFC) is the most widely used rTMS target for the treatment of neuropsychiatric disorders, including depression (Eshel et al., 2020). As a node of the frontoparietal network, it plays a critical role in integrating cognition and emotion (Mars and Grol, 2007; Gong et al., 2020). According to a previous neuroimaging study, the DLPFC of insomnia patients shows hyperexcitability compared with those who were well-slept (Spiegelhalder et al., 2013). As a result, low frequency DLPFC rTMS appears to be a reasonable strategy for the treatment of insomnia (Gong et al., 2020).

The efficacy of rTMS in treating insomnia has been investigated in several clinical studies. Most of them chose the DLPFC as the target. These studies found that 10 daily sessions of low frequency rTMS stimulation applied to the right or bilateral DLPFC resulted in a significant decrease in the Pittsburgh Sleep Quality Index (PSQI; lower index indicates better sleep) (Jiang et al., 2013; Feng et al., 2019; Shi et al., 2021). Only one study selected the right posterior parietal cortex as the stimulation site and reported similar results: 14 consecutive low frequency rTMS sessions could lower down PSQI and Insomnia Severity Index significantly (Song et al., 2019).

However, despite the aforementioned clinical evidences, there are limited neurophysiological studies investigating the underlying mechanisms of low frequency DLPFC rTMS in patients with chronic insomnia. In this context, we conducted this single-arm, open-label, interventional study to determine the neural mechanisms of rTMS other than to demonstrate its clinical efficacy. We used the resting-state electroencephalography (rsEEG) to assess the functional connectivity characteristics in patients with chronic insomnia disorder (CID) before, after and 1-month after 10 daily rTMS sessions. rsEEG is a promising paradigm for studying abnormal functional architectures in various disorders due to its task-independent properties (Zhang et al., 2021). The overarching goal was to investigate the alterations of functional connectivity induced by rTMS in patients with CID, which is critical for linking translationally relevant discoveries that can be applied in a clinical setting. The secondary aim of this study was to examine whether the clinical improvement observed 1 month after rTMS was related to any previously identified connectivity characteristics. We hypothesized that rTMS would induce connectivity changes and that such changes would correlate significantly with symptom improvements.

Materials and methods

Participants

Participants with CID were recruited at the Neurology Department of Shenzhen People's Hospital. A total of 47 patients were screened for eligibility, of whom 37 gave informed consent to participate in this study. All experimental details were approved by the Ethics Committee of Shenzhen People's Hospital (see chictr.org.cn registration: ChiCTR1900026904). Nine patients dropped out during the 1-month follow-up period; therefore, only 28 patients were included in the statistical analysis of follow-up.

All patients were required to meet the diagnostic criteria for CID according to the International Classification of Sleep Disorders, Third Edition (ICSD-3). The inclusion criteria were as follows: (1) aged 18–70 years, right-handed; (2) the sleep disturbances occur at least three times per week and present for the last 3 months; (3) PSQI ≥7 (Buysse et al., 1989); (4) scored <25 on 24-item Hamilton Depression Rating Scale (HAMD); (5) no other sleep disorders like sleep apnea, etc.; and (6) failure of at least one adequate sleep medication trial. The exclusion criteria were: (1) any contraindication to TMS (history of seizures, metallic implants, etc.); and (2) prior history of neurological or psychiatric disorders.

In addition, 40 healthy controls (HC) without sleep problems participated in the baseline assessment, serving as a reference for changes in functional connectivity. HCs needed to meet the following criteria: (1) no history of sleep disorders; (2) PSQI <7; (3) HAMD ≤7 and Hamilton Anxiety Rating Scale (HAMA) ≤7 (Matza et al., 2010; Zimmerman et al., 2013); and (4) no neurological or psychiatric disorders.

During the study, patients were allowed to take concomitant medications, and were asked to remain constant throughout the clinical trial (see Supplementary Methods for details).

rTMS treatment

Stimulation was performed using a figure 8-shaped focal coil attached to a MagPro 100 magnetic stimulator (MagVenture, Copenhagen, Denmark). All patients received 1 Hz (10 s trains, 1 s inter-train interval, 1,360 pulses per session) rTMS treatment once

daily on weekdays for 2 consecutive weeks. rTMS was delivered over the right DLPFC (F4 electrode site according to the International 10–20 EEG system) at 100% of the resting motor threshold (RMT) (Mir-Moghtadaei et al., 2015). To determine the RMT, stimulus intensity was gradually increased until 5 out of 10 trials elicited motor evoked potentials with peak-to-peak amplitudes over 50 μV in the contralateral abductor pollicis brevis muscle (Rossini et al., 2015). Adverse events attributed to rTMS were documented and reported.

Clinical assessment

All participants received a pre-treatment assessment with the PSQI, HAMA, and HAMD. We used the PSQI to measure sleep quality, as well as HAMA and HAMD to assess participant's anxious and depressive states. Likewise, the PSQI was measured post-treatment (upon completion of the final treatment session) and at 1-month follow-up. In order to determine the magnitude of the rTMS response, the percent reduction in PSQI from pre- to post-treatment, and from pre-treatment to follow-up, were calculated (see Figure 1 for the experimental design).

Electroencephalography

For the HC group, EEG data were recorded once at baseline. For the CID patients, EEG data were acquired at three time points concurrently with PSQI. Equipment setting and environment were uniform across all three acquisitions.

EEG recordings were acquired from 64 channels under closed-eye conditions for 8 min using a BrainAmp DC amplifier (Brain Products GmbH, Germany). During the recording, the participants were asked to relax on a comfortable chair in a metal-shielded room. The data were referenced online to the FCz channel with the ground at AFz. Data were initially sampled at 5,000 Hz with impedances kept below 5KΩ throughout the data collection period. Participants were instructed to refrain from consuming any caffeinated (or energy) drinks within 24 h of the EEG recording sessions.

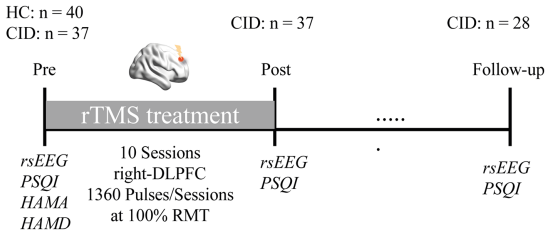


FIGURE 1
Experimental protocol. Healthy controls (HC) and patients with chronic insomnia disorder (CID) participated in an EEG session before rTMS treatment, as well as clinical assessments. Afterwards, patients received a course of low frequency rTMS treatment targeting the right DLPFC. Subsequently, EEG data and PSQI were collected again after treatment, and at 1-month follow-up, respectively.

The details of the pre-processing and source localization are described in the Supplementary Methods. Based on previous findings, the resulting EEG data were filtered into five frequency bands: theta (4–8 Hz), low-alpha (8–10 Hz), high-alpha (10–13 Hz), beta (13–30 Hz), and gamma (30–45 Hz).

Estimating functional connectivity

All connectivity analyses were computed at the source level using 3,003 vertices and then projected into 31 cortical regions of interest (ROI) using the Montreal Neurological Institute template (Chen et al., 2013); (see Supplementary Table 1). Here, we chose the debiased weighted phase-lag index (dwPLI) to represent the non-zero phase-lag statistical interdependencies between each pair of ROIs (Vinck et al., 2011). dwPLI is an optimized phase lag index that minimizes the influence of volume conduction and field spread, which could affect the estimation even at the source level. The connectivity between each pair of regions was calculated by averaging the dwPLI values over all possible vertex pairs. Accordingly, we identified 465 edges representing each participant's regional pairwise connectivity.

Statistical analysis

To compare the difference in clinical outcomes, we analyzed the PSQI of pre-, post-treatment and follow-up using a linear mixed model with a fixed effect of time and a random intercept.

The differences between pre- and post-functional connectivity matrices were analyzed using the Network-based Statistics (NBS) (Zalesky et al., 2010), a nonparametric statistical test to control for the family-wise error rate resulting from multiple comparisons. For each comparison, 5,000 random permutations were used.

To further investigate the association of functional connectivity changes with clinical outcomes, a multiple linear regression model was constructed between the pre- to post-network differences and the percentage PSQI change with age, sex, HAMA, and HAMD as covariates. Here, all edges that varied significantly (NBS-corrected, $p < 0.01$) between the pre- and post- functional connectivity matrices were treated as independent inputs. When implementing feature selection, we used the least absolute shrinkage and selection operator (LASSO) to define a low-dimensional representation of the selected connectivity features (Tibshirani, 1996). Correlation analysis between the estimated and the actual PSQI value of post-treatment and follow-up were performed using Pearson's correlation.

Results

Clinical results

The demographic and clinical characteristics of the two groups are summarized in Table 1. There were significant differences in the PSQI, HAMA, and HAMD scores ($p < 0.05$) between the HC and CID groups. No significant differences were found in age or sex. Among these characteristics, the PSQI global score correlated appreciably with HAMD score ($p < 0.01$, $r = 0.447$) but was not associated with the other measures.

TABLE 1 Demographic and clinical characteristics of the participants.

Variables	HC (n=40)	CID (n=37)	p-value
Age, year	23–72, 46.1 ± 9.4	22–69, 48.9 ± 11.1	0.23
Sex (M/F)	18/22	15/22	0.69
PSQI	3.2 ± 1.6	15.1 ± 3.4	<0.001
HAMA	2.6 ± 1.8	13.6 ± 6.2	<0.001
HAMD	2.7 ± 1.8	13.3 ± 5.0	<0.001
RMT	/	39.0 ± 12.5	

Data are presented as mean ± SD. The p values were obtained by independent sample t-test and chi-square test (for sex only). HC, healthy control; CID, chronic insomnia disorder; PSQI, Pittsburgh Sleep Quality Index; HAMA, Hamilton Anxiety Rating Scale; HAMD, Hamilton Depression Rating Scale; RMT, resting motor threshold.

All patients tolerated the rTMS well, and no adverse effects were reported. A mixed-effects model assessing PSQI differences revealed a significant main effect of rTMS treatment [$F(1.84, 57.88) = 12.38$, $p < 0.0001$]. *Post hoc* comparisons indicated that rTMS facilitated a significant reduction in insomnia symptoms on the PSQI between pre- and post-treatment (delta-PSQI = 3.243, $p = 0.0002$, Cohen's $d = 0.857$), and between scores pre- and 1-month later (delta-PSQI = 3.911, $p = 0.0003$, Cohen's $d = 1.132$) (see the Supplementary Figure 1).

rTMS-induced functional connectivity changes and association to clinical outcomes

There were no significant differences in the functional connectivity matrices between the pre- and post-treatment measures that survived multiple corrections at four frequency bands: theta (4–8 Hz), high-alpha (10–13 Hz), beta (13–30 Hz), and gamma (30–45 Hz). Instead, a comparison of the functional connectivity of low-alpha frequency band (8–10 Hz) identified 34 edges that had been significantly changed at $p < 0.01$ level after NBS correction (Figure 2; also see Supplementary Table 2). Specifically, CID patients had a lower mean functional connectivity pre-treatment. On average, these rTMS-induced connectivity changes brought patients' patterns aligned more closely with healthy controls. At follow-up, the strength of the connectivity matrix remained relatively stable (see Figure 2C). The most frequently occurring connectomes were located in the frontal lobe, preferentially involving the frontoparietal network (FPN), dorsal attention network (DAN), and ventral attention network (VAN) (see Figure 2D). Moreover, a dominant interhemispheric functional connectivity change was also notable.

Regression analyses were performed to identify connections that were significantly correlated with improvement in PSQI scores at the end of rTMS treatment. That is, the functional connectivity changes between pre- and post-treatment of these 34 identified edges were taken as independent variables (see Supplementary Table 3 for coefficients of regression model). Notably, two edges appeared to be significantly correlated with PSQI change ($r = 0.62$, $p = 0.02$) (see Figure 3B): the left inferior frontal junction to left insula (LIFJ-LINS) and the medial prefrontal cortex to left insula (MPFC-LINS) (see Figure 3A).

Analysis of connectivity changes and clinical outcomes at follow-up

EEG data acquired 1 month after treatment were used to validate the association between the connectomes obtained in the previous regression model and clinical measures. The connectivity changes of these two edges between the pre- and follow-up assessments were taken as independent inputs, including age, sex, HAMA, and HAMD at baseline, as covariates. The estimated PSQI score was significantly correlated with the actual PSQI score collected 1 month after rTMS completion ($r = 0.41$, $p = 0.032$) (Figure 3B).

Longitudinal connectivity analysis

To gain further insight into the rTMS-induced changes in LIFJ-LINS and MPFC-LINS connectivity, we examined mean values at each time point for both HC and CID groups. Significant difference was observed between the HC and CID pre-treatment for both LIFJ-LINS (average 0.46 vs. 0.27) and MPFC-LINS (average 0.41 vs. 0.30) connections (see Figure 4). After rTMS, the connectivity of the LIFJ to the LINS increased significantly at post-treatment and follow-up. The trend of MPFC-LINS connectivity resembled a similar pattern, which increased significantly at follow-up compared to that at baseline. These active rTMS-induced changes shifted the CID patient profile closer to that of HCs (see pairwise statistics in the Figure 4 legend).

Discussion

To better understand the regulatory role of rTMS in patients with CID, we examined longitudinal changes in functional connectivity induced by low frequency rTMS of the right DLPFC. In agreement with previous studies, our results confirmed the efficacy of this low frequency rTMS protocol for treating clinical CID. We found that rTMS to the right DLPFC was associated with widespread alterations in functional connectivity, which correlated with clinical outcomes, and this association persisted 1 month after the cessation of rTMS stimulation. These results offer promising preliminary evidence linking the altered brain function patterns observed in CID with clinical symptoms and expands on how changes in both aspects are associated.

Unlike other studies that focused only on *a priori* brain regions, we discovered connectivity biomarkers based directly on EEG data without making any assumptions. This data-driven approach allows us to account for uncertainties in the spatial distribution of connectivity changes correlated with clinical outcomes. Our findings suggest that the significant increase in connectivity is not confined to the ones within the stimulated network, but instead spreads across networks, involving the FPN, DAN, and VAN. Remote functional connections are impacted more than local ones, indicating that rTMS ultimately modulates insomnia-related connectivity (Castrillon et al., 2020). Additionally, we observed a dominant interhemispheric functional connectivity change, which is consistent with that reported in previous studies (Watanabe et al., 2014). Of note, since all the significant edges show higher connectivity after rTMS than before it, this observation corroborates that rTMS treatment may rewire impaired intracortical connections.

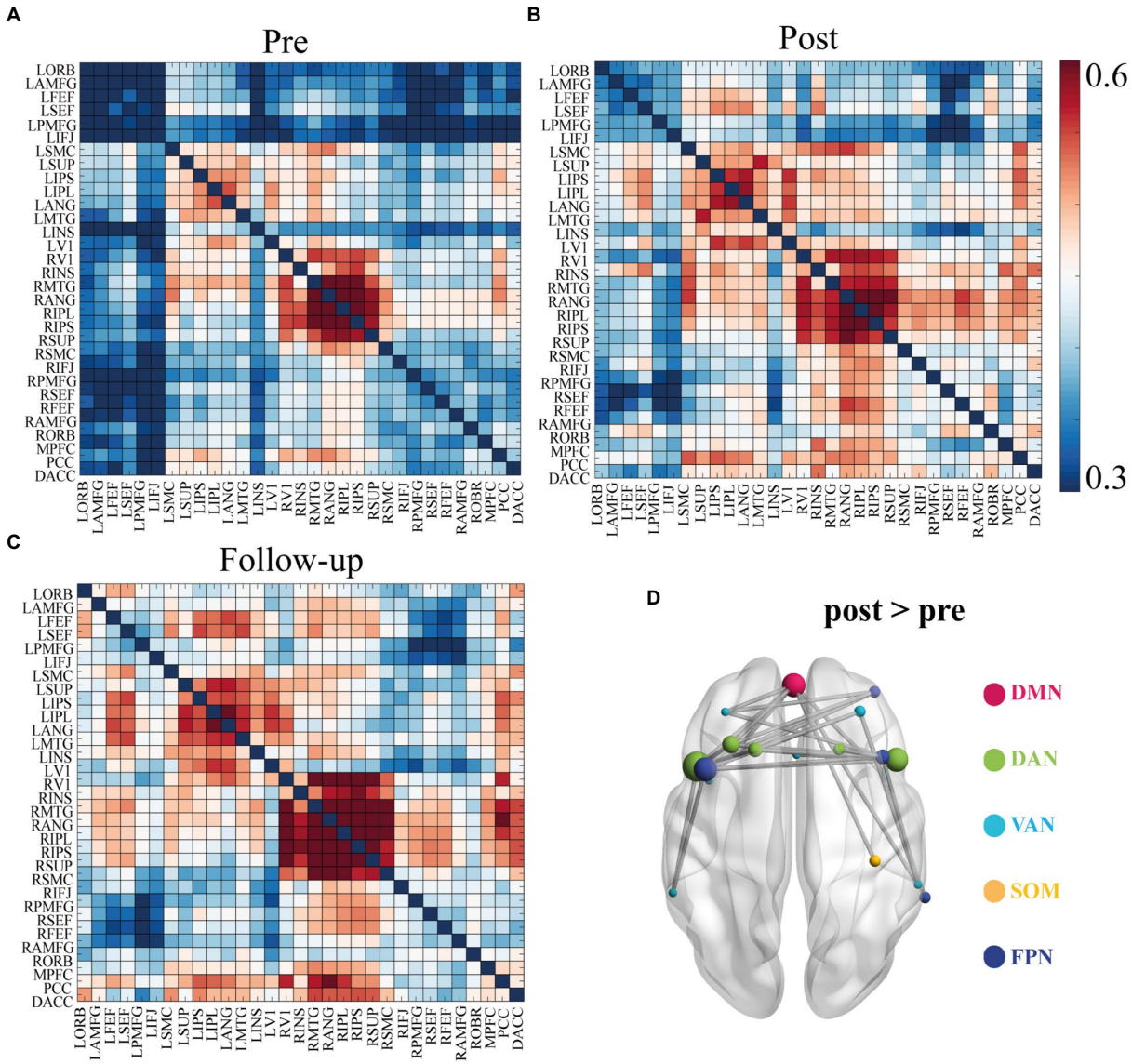


FIGURE 2
Functional connectivity matrices in low alpha band (8–10Hz) and 34 identified edges. **(A)** Connectivity matrices of pre-treatment EEG scan. **(B)** Connectivity matrices of post-treatment EEG scan. **(C)** Connectivity matrices of the follow-up EEG scan. **(D)** Significant functional connectivity changes between the pre- and post- EEG scans included 34 edges. Significance was determined by network-based statistics (NBS) correction at $p<0.01$. List of the brain regions are presented in the Supplementary Table 1. List of the 34 pairs showed significant differences are presented in the Supplementary Table 3. DMN, default mode network; DAN, Dorsal attention network; VAN, Ventral attention network; SOM, somatosensory network; FPN, frontoparietal network.

Neuroplasticity describes the ability of the brain to change and adapt its organization as a result of external influences. The results of the longitudinal analysis suggest that rTMS has a long-lasting plastic effect; hence, decrease in clinical symptoms and increase in functional connectivity characteristics could persist even 1-month after the completion of rTMS (Ge et al., 2020). It also proves the robustness of these two indicators, LINS-MPFC, LINS-LIFJ, over a long period of time.

The insula is a key node in the salience network, and its abnormalities have been pointed out in many studies of insomnia (Motomura et al., 2021). An fMRI research demonstrated decreased connectivity between the amygdala and insula during resting states, indicating a possible neural mechanism responsible for the sleep-related affective disorders and dysregulation of emotional control (Li et al., 2014). We found that patients with CID suffered significantly higher HAMA and HAMD than those with HCs, which reflects the true situation in real life. Our study reinforces earlier findings about the insula in CID and adds to the evidence that it plays a role in the disease's pathophysiology. The MPFC and its ample connections with other regions, play critical roles in long-range, recent, and short-term memory over a wide range of activities (Euston et al., 2012). It has also been previously discussed regarding insomnia with the conclusion that circuit dysfunction involving the MPFC was associated with poor sleep quality as measured by the PSQI (Shao et al., 2020). The IFJ is part of the cognitive control network that co-activates with the DLPFC, ventrolateral prefrontal cortex, anterior insula, and posterior parietal cortex (Sundermann and Pfleiderer, 2012).

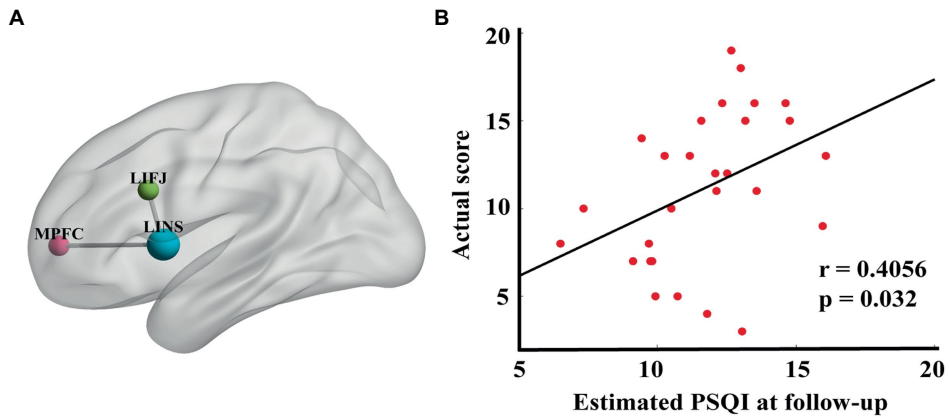


FIGURE 3
(A) Edges that were significantly correlated with percentage PSQI change before and after treatment (LIFJ-LINS, $p<0.01$, MPFC-LINS, $p<0.01$). **(B)** The estimated PSQI at follow-up were significantly correlated with the actual score ($r=0.4056$, $p=0.032$), with age, gender, HAMA and HAMD at baseline as the nuisance covariates.

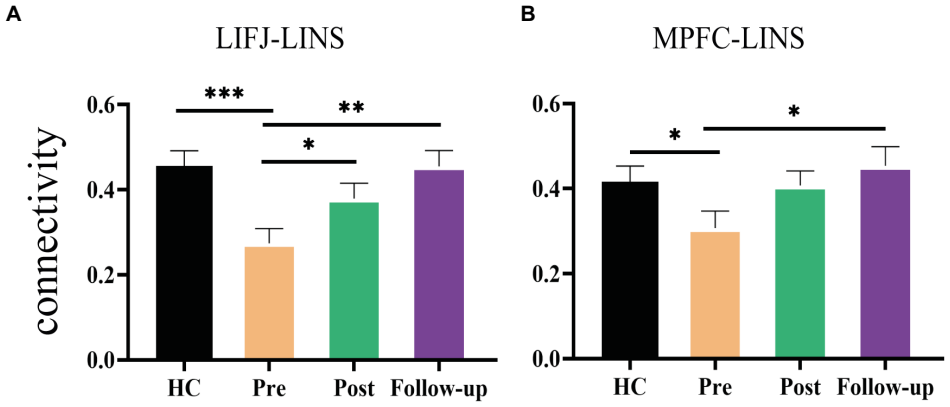


FIGURE 4
Significant rTMS effect of group and time. **(A)** Changes of the mean functional connectivity between left inferior eye junction (LIFJ) and left insula (LINS). Mixed effect model revealed significant rTMS treatment effect in the functional connection [$F(1.8, 55.83)=7.256$, $p=0.0022$]. *Post hoc* comparison showed significant differences between pre- and post-treatment ($p=0.0236$), and pre-treatment to follow-up ($p=0.0076$). The functional connectivity of HC was significantly higher than CID patients at baseline ($t=3.641$, $p=0.0005$). **(B)** Changes of the mean functional connectivity between medial prefrontal cortex (MPFC) and left insula (LINS). Mixed effect model revealed significant rTMS treatment effect in the functional connection [$F(1.92, 59.53)=3.862$, $p=0.028$]. *Post hoc* comparison showed a significant difference between pre-treatment and follow-up ($p=0.0497$). Independent *t*-test showed that CID patients had significantly lower functional connectivity at baseline than HCs ($t=2.016$, $p=0.0474$). Error-bars represents SEM. *$p<0.05$, **$p<0.01$, ***$p<0.001$.

In light of our understanding of how alpha oscillations correlate with resting wakefulness, the therapeutic benefits of rTMS typically focus on alpha rhythms. The alpha rhythms promote coordination between cortical areas and between the cortex and subcortical structures, such as the thalamus (Buzsáki and Draguhn, 2004). A highly synchronized alpha rhythms can facilitate the coordination of brain regions in an event-related, synchronized readiness manner, prior to engaging in different type of tasks (Klimesch et al., 2007). We found that DLPFC-rTMS significantly enhanced the connectivity of low-frequency alpha rhythms (8–10 Hz), but not of the entire alpha band (8–13 Hz) (see Supplementary Figure 2). In the thalamo-cortical and cortico-cortical loops, alpha rhythms carry out different functions; for example, low-frequency alpha rhythms execute inhibitory functions (Pfurtscheller and Lopes da Silva, 1999). Researchers have proposed that the lower alpha band diffusely regulates alertness and arousal in the brain (Klimesch et al., 1998). Our results show that inhibitory 1 Hz rTMS stimulation of the DLPFC increases lower alpha synchronization in the resting state. This inhibitory synchronization state could facilitates recruitment of specific regions of the cortex to transmit and retrieve task information before preparing for subsequent tasks (Klimesch, 2012). A previous study reported a similar phenomenon. rTMS targeting in the angular gyrus, core regions of the DAN, enhanced intrahemispheric alpha coherence of 8–10 Hz, suggesting the causal role of the angular gyrus in modulating of dominant low-frequency alpha rhythms under the resting conditions (Capotosto et al., 2014).

CID appears to be associated with altered cortico-thalamic connectivity, which is partly responsible for cognitive regulation and circadian processes (Zou et al., 2021). Indeed, the thalamus is the pacemaker of cortical alpha rhythms. Taken together, we infer that DLPFC-rTMS regulates altered alpha connectivity by top-down control of DLPFC to the thalamus. While path analysis indicates a causal relationship, our study cannot infer causality since no connectivity with subcortical regions was observed. These results should therefore, only be considered as supportive. Further neuroimaging study is required to elucidate the potential involvement of the thalamus in the improvement of cortical alpha band connections following treatment. A previous study supports our hypothesis reporting that DLPFC-rTMS modulates the functional connectivity of the insula and thalamus in smokers (Li et al., 2017).

rTMS is a promising strategy for treating various neuropsychiatric disorders. The effects of rTMS depend on its intensity, frequency and stimulation site, in particular the stimulation site. However, there is no consensus on the optimal stimulation site for insomnia due to the prior lack of therapeutic mechanisms of rTMS. These findings offer implications for optimizing brain stimulation therapy, which should be examined in future clinical trials. Besides, as the optimal treatment parameters for insomnia have not yet been determined, the schedule implemented in this study is a commonly used protocol to ensure treatment adherence. It is essential to identify efficacious dosing parameters that are feasible for further studies.

This study has some limitations. First, the lack of a sham-control group prevents us from determining whether the results can be partly explained by placebo effects. In effect, the primary goal of this study was to determine the neural mechanism of rTMS rather than to demonstrate differential clinical outcome with active versus sham conditions. Second, chronic insomnia is quite a heterogenic disorder, often comorbid with other symptoms. To ensure the homogeneity of patients enrolled in the study, we strictly enrolled patients who had difficulties with sleep for over 3 months and had failed at least one medication trial. Another potential weakness of this study was that all patients were medicated. The drug itself affects EEG parameters. Ideally, the inclusion of medication-free patients would eliminate any potential confounding effect of medication; this is not an ethically viable option for patients with medication-resistant insomnia. It remains uncertain whether the functional connectomes found in the present study are general biomarkers of response across all treatments. Hence, these findings need to be replicated in equally large samples obtained from multiple sites.

Conclusion

Despite its high prevalence, there has been little progress in the treatment of CID. We found that low frequency rTMS treatment over the right DLPFC significantly increased the EEG-derived functional connectivity in patients with CID. Furthermore, changes in connections, particularly the LIFJ-LINS and MPFC-LINS, were significantly associated with clinical measures and could predict the PSQI in the subsequent assessment at 1-month follow-up. These results could be very valuable in clinical treatment of insomnia, especially in patients that are resistant to medications or have a history of drug failure. These findings expand our understanding of neural response in patients with chronic insomnia treated with rTMS and lay the foundation for future studies. Further research using a rigorous design is required to address the aforementioned issues.

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 human participants were reviewed and approved by the Shenzhen People's Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

LZ contributed to the data analysis and original draft. GD, XSh, and XSu were responsible for acquisition and analysis. WW and JZ assisted with analysis and manuscript revision. XL and CL contributed to the data collections. ZP and HR contributed to the data curation. PX provided clinical study consultation. YG was responsible for the conceptualization and funding acquisition. All authors critically reviewed the final content and approved for publication.

Funding

This study was supported by the Natural Science Foundation of Guangdong Province (grant number 2021A1515010983), the Sanming Project of Medicine in Shenzhen (grant number SZSM202111009), the Shenzhen Science and Technology Innovation Program (grant number KCXFZ20201221173400001), and Shenzhen Key Medical Discipline Construction Fund (grant number SZXK005).

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2023.1135995/full#supplementary-material

References

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., and Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Res.* 28, 193–213. doi: 10.1016/0165-1781(89)90047-4

Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi: 10.1126/science.1099745

Capotosto, P., Babiloni, C., Romani, G. L., and Corbetta, M. (2014). Resting-state modulation of α rhythms by interference with angular gyrus activity. *J. Cogn. Neurosci.* 26, 107–119. doi: 10.1162/jocn_a_00460

Castrillon, G., Sollmann, N., Kurcyus, K., Razi, A., Krieg, S. M., and Riedl, V. (2020). The physiological effects of noninvasive brain stimulation fundamentally differ across the human cortex. *Sci. Adv.* 6:eaay2739. doi: 10.1126/sciadv.aay2739

Chen, M. C., Chang, C., Glover, G. H., and Gotlib, I. H. (2014). Increased insula coactivation with salience networks in insomnia. *Biol. Psychol.* 97, 1–8. doi: 10.1016/j.biopsycho.2013.12.016

Chen, A. C., Oathes, D. J., Chang, C., Bradley, T., Zhou, Z.-W., Williams, L. M., et al. (2013). Causal interactions between fronto-parietal central executive and default-mode networks in humans. *Proc. Natl. Acad. Sci. U. S. A.* 110, 19944–19949. doi: 10.1073/pnas.1311772110

Eshel, N., Keller, C. J., Wu, W., Jiang, J., Mills-Finnerty, C., Huemer, J., et al. (2020). Global connectivity and local excitability changes underlie antidepressant effects of repetitive transcranial magnetic stimulation. *Neuropsychopharmacology* 45, 1018–1025. doi: 10.1038/s41386-020-0633-z

Euston, D. R., Gruber, A. J., and McNaughton, B. L. (2012). The role of medial prefrontal cortex in memory and decision making. *Neuron* 76, 1057–1070. doi: 10.1016/j.neuron.2012.12.002

Feng, J., Zhang, Q., Zhang, C., Wen, Z., and Zhou, X. (2019). The effect of sequential bilateral low-frequency rTMS over dorsolateral prefrontal cortex on serum level of BDNF and GABA in patients with primary insomnia. *Brain Behav* 9:e01206. doi: 10.1002/brb3.1206

Fitzgerald, P., Fountain, S., and Daskalakis, Z. (2006). A comprehensive review of the effects of rTMS on motor cortical excitability and inhibition. *Clin. Neurophysiol.* 117, 2584–2596. doi: 10.1016/j.clinph.2006.06.712

Ge, R., Downar, J., Blumberger, D. M., Daskalakis, Z. J., and Vila-Rodriguez, F. (2020). Functional connectivity of the anterior cingulate cortex predicts treatment outcome for rTMS in treatment-resistant depression at 3-month follow-up. *Brain Stimul.* 13, 206–214. doi: 10.1016/j.brs.2019.10.012

Glass, J., Lanctôt, K. L., Herrmann, N., Sproule, B. A., and Busto, U. E. (2005). Sedative hypnotics in older people with insomnia: Meta-analysis of risks and benefits. *BMJ* 331:1169. doi: 10.1136/bmj.38623.768588.47

Gong, L., Xu, R., Qin, M., Liu, D., Zhang, B., Bi, Y., et al. (2020). New potential stimulation targets for noninvasive brain stimulation treatment of chronic insomnia. *Sleep Med.* 75, 380–387. doi: 10.1016/j.sleep.2020.08.021

Huang, Z., Liang, P., Jia, X., Zhan, S., Li, N., Ding, Y., et al. (2012). Abnormal amygdala connectivity in patients with primary insomnia: Evidence from resting state fMRI. *Eur. J. Radiol.* 81, 1288–1295. doi: 10.1016/j.ejrad.2011.03.029

Jiang, C., Zhang, T., Yue, F., Yi, M., and Gao, D. (2013). Efficacy of repetitive transcranial magnetic stimulation in the treatment of patients with chronic primary insomnia. *Cell Biochem. Biophys.* 67, 169–173. doi: 10.1007/s12013-013-9529-4

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* 16, 606–617. doi: 10.1016/j.tics.2012.10.007

Klimesch, W., Doppelmayr, M., Russegger, H., Pachinger, T., and Schwaiger, J. (1998). Induced alpha band power changes in the human EEG and attention. *Neurosci. Lett.* 244, 73–76. doi: 10.1016/s0304-3940(98)00122-0

Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition-timing hypothesis. *Brain Res. Rev.* 53, 63–88. doi: 10.1016/j.brainresrev.2006.06.003

Lanza, G., Cantone, M., Lanuzza, B., Pennisi, M., Bella, R., Pennisi, G., et al. (2015). Distinctive patterns of cortical excitability to transcranial magnetic stimulation in obstructive sleep apnea syndrome, restless legs syndrome, insomnia, and sleep deprivation. *Sleep Med. Rev.* 19, 39–50. doi: 10.1016/j.smrv.2014.04.001

Li, Z., Chen, R., Guan, M., Wang, E., Qian, T., Zhao, C., et al. (2018). Disrupted brain network topology in chronic insomnia disorder: A resting-state fMRI study. *Neuroimage Clin.* 18, 178–185. doi: 10.1016/j.nicl.2018.01.012

Li, X., Du, L., Sahlem, G. L., Badran, B. W., Henderson, S., and George, M. S. (2017). Repetitive transcranial magnetic stimulation (rTMS) of the dorsolateral prefrontal cortex reduces resting-state insula activity and modulates functional connectivity of the orbitofrontal cortex in cigarette smokers. *Drug Alcohol Depend.* 174, 98–105. doi: 10.1016/j.drugalcdep.2017.02.002

Li, Y., Wang, E., Zhang, H., Dou, S., Liu, L., Tong, L., et al. (2014). Functional connectivity changes between parietal and prefrontal cortices in primary insomnia patients: Evidence from resting-state fMRI. *Eur. J. Med. Res.* 19:32. doi: 10.1186/2047-783X-19-32

Ly, J. Q. M., Gaggioni, G., Chellappa, S. L., Papachilleos, S., Brzozowski, A., Borsu, C., et al. (2016). Circadian regulation of human cortical excitability. *Nat. Commun.* 7, 1–10. doi: 10.1038/ncomms11828

Mars, R. B., and Grol, M. J. (2007). Dorsolateral prefrontal cortex, working memory, and prospective coding for action. *J. Neurosci.* 27, 1801–1802. doi: 10.1523/JNEUROSCI.5344-06.2007

Matza, L. S., Morlock, R., Sexton, C., Malley, K., and Feltner, D. (2010). Identifying HAM-A cutoffs for mild, moderate, and severe generalized anxiety disorder. *Int. J. Methods Psychiatr. Res.* 19, 223–232. doi: 10.1002/mpr.323

Mir-Moghtadaei, A., Caballero, R., Fried, P., Fox, M. D., Lee, K., Giacobbe, P., et al. (2015). Concordance between BeamF3 and MRI-neuronavigated target sites for repetitive transcranial magnetic stimulation of the left dorsolateral prefrontal cortex. *Brain Stimul.* 8, 965–973. doi: 10.1016/j.brs.2015.05.008

Morin, C. M., and Benca, R. (2012). Chronic insomnia. *Lancet* 379, 1129–1141. doi: 10.1016/S0140-6736(11)60750-2

Motomura, Y., Katsunuma, R., Ayabe, N., Oba, K., Terasawa, Y., Kitamura, S., et al. (2021). Decreased activity in the reward network of chronic insomnia patients. *Sci. Rep.* 11:3600. doi: 10.1038/s41598-020-79989-2

Pfurtscheller, G., and Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/s1388-2457(99)00141-8

Riemann, D., Spiegelhalder, K., Feige, B., Voderholzer, U., Berger, M., Perlis, M., et al. (2010). The hyperarousal model of insomnia: A review of the concept and its evidence. *Sleep Med. Rev.* 14, 19–31. doi: 10.1016/j.smrv.2009.04.002

Rossini, P. M., Burke, D., Chen, R., Cohen, L. G., Daskalakis, Z., di Iorio, R., et al. (2015). Non-invasive electrical and magnetic stimulation of the brain, spinal cord, roots and peripheral nerves: Basic principles and procedures for routine clinical and research application. An updated report from an I.F.C.N. Committee. *Clin. Neurophysiol.* 126, 1071–1107. doi: 10.1016/j.clinph.2015.02.001

Shao, Z., Xu, Y., Chen, L., Wang, S., Zhang, M., Liu, S., et al. (2020). Dysfunction of the NAc-mPFC circuit in insomnia disorder. *Neuroimage Clin.* 28:102474. doi: 10.1016/j.nicl.2020.102474

Shi, X., Guo, Y., Zhu, L., Wu, W., Hordacre, B., Su, X., et al. (2021). Electroencephalographic connectivity predicts clinical response to repetitive transcranial magnetic stimulation in patients with insomnia disorder. *Sleep Med.* 88, 171–179. doi: 10.1016/j.sleep.2021.10.017

Song, P., Lin, H., Li, S., Wang, L., Liu, J., Li, N., et al. (2019). Repetitive transcranial magnetic stimulation (rTMS) modulates time-varying electroencephalography (EEG) network in primary insomnia patients: A TMS-EEG study. *Sleep Med.* 56, 157–163. doi: 10.1016/j.sleep.2019.01.007

Spiegelhalder, K., Regen, W., Baglioni, C., Nissen, C., Riemann, D., and Kyle, S. D. (2015). Neuroimaging insights into insomnia. *Curr. Neurol. Neurosci. Rep.* 15:9. doi: 10.1007/s11910-015-0527-3

Spiegelhalder, K., Regen, W., Baglioni, C., Riemann, D., and Winkelman, J. W. (2013). Neuroimaging studies in insomnia. *Curr. Psychiatry Rep.* 15:405. doi: 10.1007/s11920-013-0405-0

Sundermann, B., and Pfleiderer, B. (2012). Functional connectivity profile of the human inferior frontal junction: Involvement in a cognitive control network. *BMC Neurosci.* 13:119. doi: 10.1186/1471-2202-13-119

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* 58, 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x

Vinck, M., Oostenveld, R., van Wingerden, M., Battaglia, F., and Pennartz, C. M. A. (2011). An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *Neuroimage* 55, 1548–1565. doi: 10.1016/j.neuroimage.2011.01.055

Watanabe, T., Hanajima, R., Shirota, Y., Ohminami, S., Tsutsumi, R., Terao, Y., et al. (2014). Bidirectional effects on interhemispheric resting-state functional connectivity induced by excitatory and inhibitory repetitive transcranial magnetic stimulation. *Hum. Brain Mapp.* 35, 1896–1905. doi: 10.1002/hbm.22300

Zalesky, A., Fornito, A., and Bullmore, E. T. (2010). Network-based statistic: Identifying differences in brain networks. *Neuroimage* 53, 1197–1207. doi: 10.1016/j.neuroimage.2010.06.041

Zhang, Y., Wu, W., Toll, R. T., Naparstek, S., Maron-Katz, A., Watts, M., et al. (2021). Identification of psychiatric disorder subtypes from functional connectivity patterns in resting-state electroencephalography. *Nat. Biomed. Eng.* 5, 309–323. doi: 10.1038/s41551-020-00614-8

Zimmerman, M., Martinez, J. H., Young, D., Chelminski, I., and Dalrymple, K. (2013). Severity classification on the Hamilton depression rating scale. *J. Affect. Disord.* 150, 384–388. doi: 10.1016/j.jad.2013.04.028

Zou, G., Li, Y., Liu, J., Zhou, S., Xu, J., Qin, L., et al. (2021). Altered thalamic connectivity in insomnia disorder during wakefulness and sleep. *Hum. Brain Mapp.* 42, 259–270. doi: 10.1002/hbm.25221

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 05 June 2023
DOI 10.3389/fnins.2023.1178283

OPEN ACCESS

EDITED BY
Jing Jin,
East China University of Science and Technology, China

REVIEWED BY
Mengfan Li,
Hebei University of Technology, China
Dong Wen,
University of Science and Technology Beijing, China

*CORRESPONDENCE
Minpeng Xu
✉ xmp52637@tju.edu.cn

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

RECEIVED 02 March 2023
ACCEPTED 20 March 2023
PUBLISHED 05 June 2023

CITATION
Xiao X, Gao R, Zhou X, Yi W, Xu F, Wang K, Xu M and Ming D (2023) A novel visual brain-computer interfaces paradigm based on evoked related potentials evoked by weak and small number of stimuli.
Front. Neurosci. 17:1178283.
doi: 10.3389/fnins.2023.1178283

A novel visual brain-computer interfaces paradigm based on evoked related potentials evoked by weak and small number of stimuli

Xiaolin Xiao[1,2†], Runyuan Gao[1†], Xiaoyu Zhou[1], Weibo Yi[3], Fangzhou Xu[4], Kun Wang[2], Minpeng Xu[1,2,4*] and Dong Ming[1,2]

[1]School of Precision Instruments and Optoelectronics Engineering, Tianjin University, Tianjin, China, [2]Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin, China, [3]Beijing Institute of Mechanical Equipment, Beijing, China, [4]International School for Optoelectronic Engineering, Qilu University of Technology, Shandong Academy of Sciences, Jinan, China

Introduction: Traditional visual Brain-Computer Interfaces (v-BCIs) usually use large-size stimuli to attract more attention from users and then elicit more distinct and robust EEG responses, which would cause visual fatigue and limit the length of use of the system. On the contrary, small-size stimuli always need multiple and repeated stimulus to code more instructions and increase separability among each code. These common v-BCIs paradigms can cause problems such as redundant coding, long calibration time, and visual fatigue.

Methods: To address these problems, this study presented a novel v-BCI paradigm using weak and small number of stimuli, and realized a nine-instruction v-BCI system that controlled by only three tiny stimuli. Each of these stimuli were located between instructions, occupied area with eccentricities subtended 0.4°, and flashed in the row-column paradigm. The weak stimuli around each instruction would evoke specific evoked related potentials (ERPs), and a template-matching method based on discriminative spatial pattern (DSP) was employed to recognize these ERPs containing the intention of users. Nine subjects participated in the offline and online experiments using this novel paradigm.

Results: The average accuracy of the offline experiment was 93.46% and the online average information transfer rate (ITR) was 120.95 bits/min. Notably, the highest online ITR achieved 177.5 bits/min.

Discussion: These results demonstrate the feasibility of using a weak and small number of stimuli to implement a friendly v-BCI. Furthermore, the proposed novel paradigm achieved higher ITR than traditional ones using ERPs as the controlled signal, which showed its superior performance and may have great potential of being widely used in various fields.

KEYWORDS

v-BCI, weak stimuli, evoked related potential, discriminative spatial patterns, information transfer rate

1. Introduction

Brain-Computer Interfaces (BCIs), which provide a new communication pathway between humans and the outside world, and received great attention in recent years (Lebedev and Nicolelis, 2006; Wolpaw, 2014). Among the various BCIs, the electroencephalogram (EEG) based BCI has become a popular solution in BCI research, which is considered non-invasiveness, low-cost, and more convenient to set up (Xu et al., 2013; Chen et al., 2015). Notably, visual BCIs (v-BCIs) can achieve higher information transfer rate (ITR) than other EEG-BCIs, due to the high signal-noise rate (SNR) of visual evoked potentials (VEPs) (Nakanishi et al., 2017). However, although the current v-BCI has achieved superior performance (Han et al., 2023), it is always applied in a laboratory environment. The weak friendliness of human-computer interaction is an important reason, which limits BCIs' daily use in real life (Miao et al., 2020). For most traditional v-BCIs, the stimulus blocks and instructions are spatially overlapping. Therefore, they commonly need users to stare at the irritating flash stimulus to output commands, which inevitably brings out visual fatigue for users (Chen et al., 2015). In addition, most traditional v-BCIs commonly used large-size or a relatively large numbers of visual stimuli to attract users' attention and elicit distinct EEG features (Xu et al., 2018), which further brings a strong burden on users.

Recently, lateralized visual stimuli away from the central field of view has gained considerable interest in BCI studies because of the theory of retinotopic mapping (Yoshimura et al., 2011; Chen et al., 2017). According to retina-cortical mapping, the spatial pattern of VEPs is closely related to the position of visual stimuli in the visual field (Wurtz and Kandel, 2000). That is, there is a spatial mapping relation between the position of visual stimuli and the gaze position. Therefore, retina-cortical-based BCIs always arrange lateralized visual stimuli spatially separated from instructions, which no longer request subjects to stare at the irritating flash stimulus and therefore reduce the visual burden on the users. However, stimuli away from instructions may evoke EEG responses with lower SNR and discriminability. To address this problem, current retina-cortical-based BCIs commonly increase the size or number of visual stimuli, which brings out visual burden on users. For instance, in 2018, Chen accomplished a four-instructions speller that identified by a single motion stimulus (Chen et al., 2018). The instructions were fixed around a vertical bar ($6° \times 0.32°$), which appeared from left border of the central square and moved rightward. However, the size of the sliding stimuli had eccentricity with 6Ɣ, even larger than the $4° \times 4°$ of the stimulation block in traditional v-BCIs (Chen et al., 2015). In the same year, we developed a 32-command BCI system using very small lateral visual stimuli, and each stimuli only subtended 0.4Ɣ of the visual angle (Xu et al., 2018). The results of this study firstly demonstrated that the weak spatial evoked related potentials (ERPs) induced by small visual stimuli were also detectable for BCIs. For each command in this study, there was a set of left and right lateral visual stimuli flicking in different time sequences. That means two stimuli were used to code one command, which increased coding time and experimentation time. In 2021, we further designed a 16-command ERP-BCI with only one small visual stimulus. In this study, the spatial ERPs induced by a small visual stimulus was highly related to the distances and directions between the stimulus and the 16 gazed positions (Zhou et al., 2021). Inspired by these results, we believe that there is huge potential to design user-friendly and multi-instructions ERP-BCIs using tiny lateralized ERPs.

In this study, we designed a novel 3×3 ERP-BCI paradigm using small and lateralized visual stimuli, and the weak stimuli located between instruction rows/columns flashed in a row-column paradigm. Each of the stimuli had a small size with 0.4° of visual angle. Meanwhile, this paradigm used only three visual stimuli to code nine instructions, which reduced intensity of single-round stimulation. Furthermore, we conducted offline and online experiments with this novel paradigm, and used a template-matching method to recognize weak ERPs.

2. Method and materials

2.1. Participants

Nine healthy volunteers from Tianjin University (five males and four females, aged from 20 to 25 years) with normal or corrected-to-normal vision participated in the offline and online experiments. They had read and signed informed consents which were approved by the Research Ethics Committee of Tianjin University before the experiment.

2.2. Experimental procedure

Before the experiment, subjects were seated 70 centimeters in front of the screen and they were told to try their best to avoid behavior such as blinking and body shaking during the experiment. The paradigm was presented on a 24-inch LCD computer monitor with a 1920×1080-pixel resolution and a 120 Hz refresh rate. During the experiment, the subjects were told only need to stare at the instruction rather than flickering stimuli, and their eyes were always at the same height as the center of the paradigm.

Four stimulus patterns of the offline experimental procedure are shown in the top panel of Figure 1. A 3×3 (each with a visual field of $1.64° \times 1.64°$) character matrix was presented in the center of the screen. Three small visual stimuli which only subtended 0.4Ɣ of visual angle were located between instruction rows/columns and flashed in the row-column paradigm in the order of four stimulus patterns. Each of four kinds of stimulus patterns were composed of three stimuli, and were tagged with S1, S2, S3, and S4, respectively. These four patterns divided the character matrix into four areas (upper, lower, left and right) in the paradigm. The flowchart of the offline experiment is shown in the bottom panel of Figure 1. In this experiment, one session was divided into the stage of cue and the stage of flicker. Cueing duration lasted 500 ms for subjects to shift their point of view, subjects were told to stare at the characters cued with a specific symbol '∇'. Flashing process consisted of 5 trials in offline experiment. Each trial consisted of 4 specific stimulus patterns which corresponded with the stimulus shown in the top panel of Figure 1. The stimulus onset asynchrony (SOA) between two consecutive stimulation was fixed at 100 ms,

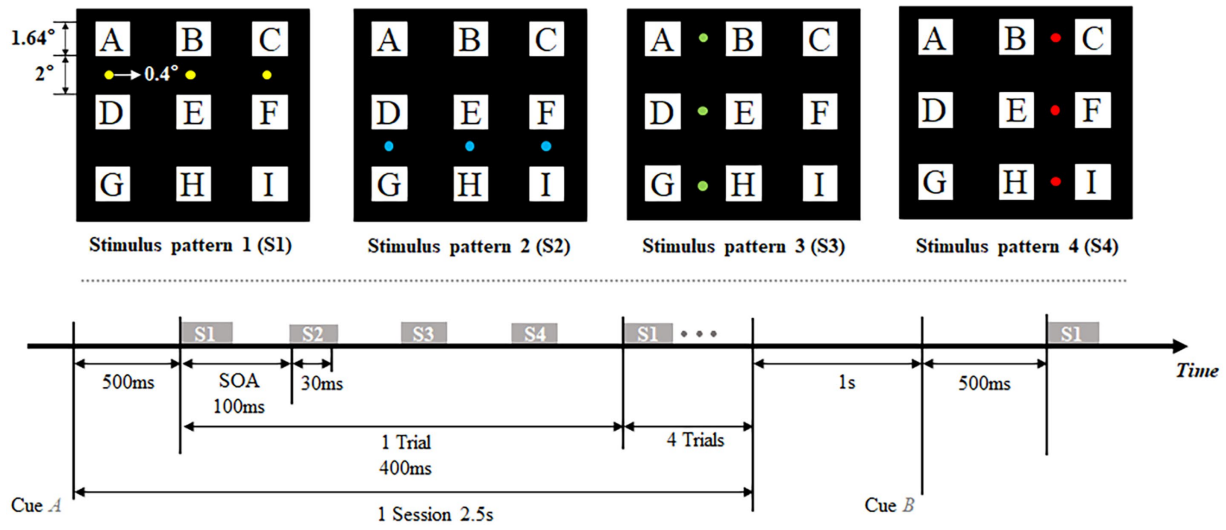


FIGURE 1
Stimulus patterns (top) and flowchart of experiment (bottom).

including 30 ms presentation of stimulation and 70 ms blank without stimulus. Thus, one trial lasted 400 ms that consisted of four stimuli, and the process of flashing lasted 2 s that consisted of 5 trials. In addition, the inter-session interval was fixed at 1 s, which allowed subjects' activities such as blinking and avoided interference between each session.

In the offline experiment, data of 30 sessions were collected for each instruction which led to an experimental duration of 30 min for each subject. Subjects were told to take a short break to reduce visual fatigue between consecutive sessions. After the offline experiment, an online spelling task of randomly spelling nine characters twice was conducted to further evaluate the performance of the proposed ERP-BCI. All subjects were asked to make opinions on the comfort level of the proposed paradigm after completing character spelling.

2.3. Signal recording and pre-processing

In this study, the EEG data was recorded by the Neuroscan Synamps2 system with 64 electrodes. One referenced electrode was in the central area near Cz and one ground electrode was on the frontal lobe. The sampling rate of the system was set to 1,000 Hz and a 50 Hz notch filter was applied during the EEG acquisition. In the pre-processing, the EEG was firstly re-referenced to the average value of left and right mastoids M1, M2, and then filtered at 1–15 Hz by Chebyshev II filters. To make the data easy to calculate and store, the filtered EEG was down sampled at 500 Hz. This study used MetaBCI for realizing real-time high-speed data retrieving, data offline/online processing and output feedback. MetaBCI is a one-stop software for the construction of BCIs which was proposed as the first open-source software for BCIs in China.[1]

1 https://github.com/TBC-TJU/MetaBCI

2.4. Classify algorithm of character-by-character recognition

It has been demonstrated that Discriminative Spatial Pattern (DSP) could effectively identify the spatial feature of EEG evoked by very small stimuli (Xu et al., 2018; Zhou et al., 2021). Thus, this study used DSP for extracting and recognizing EEG features, which consisted of two major parts: (1) the construction of DSPs and (2) pattern matching. It is well known that DSP finds a projection matrix W, which extracts the spatial features and removes the common noise from EEG (Duda et al., 2001). The projection matrix W can be obtained by maximizing the Fisher's linear discriminant criterion as follows (Duda and Hart, 2006):

$$J(\mathrm{W}) = \frac{|S_B|}{|S_W|} = \frac{|W^T S_B W|}{|W^T S_W W|} \tag{1}$$

where S_B indicates between-class scatter matrix and S_W indicates within-class scatter matrix (Liao et al., 2007). In the proposed paradigm, distinguishable neural response patterns could be elicited when gazing at each of nine instructions in each stimulus pattern. In other words, nine instructions could be recognized in each of four stimulus patterns. Therefore, we calculated projection matrix $W_K^i \in R^{N_C \times N_C}$ $(i = 1 \ldots 4, K = 1 \ldots 9)$ of each stimulus patterns and then obtained training template $\hat{X}_K^i \in R^{N_C \times N_T}$ after filtering of noise removal, where W_K^i indicates the projection matrix of instruction K from stimulus pattern i and $\hat{X}_K^i$ is the average of training samples from the stimulus pattern i and instruction K, N_C is the number of channels and N_T is the number of time points.

In pattern matching, test sample $Y \in R^{N_C \times N_T}$ can be divided into four segments Y_i according to the time windows of four stimulus patterns. Each segment data can be template-matched according to its stimulus pattern, then the decision value of specific characters $\tilde{\rho}_K$ can be calculated by the $W_K^{iT} X_K$ and $Y^i W_K^i$ as follows:

$$\tilde{\rho}_K = \sum_{i=1}^{4} \rho_i = \sum_{i=1}^{4} corr\left(W_K^{i\,T} \hat{X}_K^i, Y^i W_K^i \right) \tag{2}$$

And the predicted instruction of Y is:

$$\tilde{k} = \max \left\{ \tilde{\rho}_K, K = 1,2,..,9 \right\} \tag{3}$$

2.5. Information transfer rate

To further evaluate the performance of the proposed system, this study investigated the Information Transfer Rate (Wolpaw et al., 2000), which is defined as

$$ITR = \left(\log_2 N + P \log_2 P + (1-P) \log_2 \frac{1-P}{N-1} \right) \times \left(\frac{60}{T} \right) \tag{4}$$

where N is the total number of system commands, P is the accuracy of target identification and T is the consuming time for each command outputting, which includes 500 ms for the subjects to shift their focus additionally.

3. Results and analysis

3.1. Analysis of ERPs with spatial specificity in the novel paradigm

Previous studies indicated that spatial features of ERPs commonly appeared at 100–200 ms after the presentation of stimuli (Chen et al., 2018; Xu et al., 2018; Zhou et al., 2021). For one trial arranged in this paradigm, various stimuli presented at 0 ms, 100 ms, 200 ms, and 300 ms corresponding to four stimulus patterns S1, S2, S3, and S4 individually (Figure 1). Thus, timing windows of ERP responses, respectively, concentrated in 100–200 ms, 200–300 ms, 300–400 ms, and 400–500 ms corresponding to each of four stimulus patterns in one trial. In Figure 2, code 'E' among all characters was firstly chosen for analysis as its position which has an equal visual angle from all stimulus patterns. We first investigated waveforms and typical topographies for code 'E' during 0–500 ms after the presentation of visual stimuli across nine subjects. ERP features at electrodes of PO7 and PO8 were analyzed because of their stronger and more typical spatial responses than other electrodes (Chen et al., 2018). As shown in Figure 2, ERP responses reached the first wave crest around 120 ms that evoked by S1. Meanwhile, most signals in the central occipital region presented a positive amplitude in the topography at that moment, which is generally acknowledged as the responses evoked by stimulus located above the line of sight. Subsequently, during 200–300 ms, potentials showed a negative variation, and obvious negative waves could be observed near 230 ms. Most signals in the central occipital region presented a negative amplitude in the topography at that moment. The polarities of most electrodes were observed inverted from 100–200 ms to 200–300 ms, which corresponded to ERP responses for S1 and S2, respectively. Then during 300–400 ms, ERP responses were concentrated in the right hemisphere of the brain in the view of the topography and the waveform at PO8. A laterally symmetric response could be found during 400–500 ms after S4, which were concentrated in the left hemisphere of the brain. The ERP responses with spatial specificity appeared after S3 and S4 showed contralateral maximal responses when the stimuli were located on the left of the visual field (stimulus pattern 3) and right of the visual field (stimulus pattern 4).

Furthermore, Figure 3 showed average waveforms across all subjects when gazing at each of nine instructions. For a clear presentation of results, the location of subgraphs in Figure 3 referenced the real position of the instruction in the character matrix in paradigm. It could be observed that the spatial features of ERPs at PO7 and PO8 were similar when the subjects noted the specific characters in the same time window of S1 and S2. For characters A, B and C in the top row, stimuli were presented below the instruction in pattern of S1, and spatial responses concentrated during 100–200 ms after the

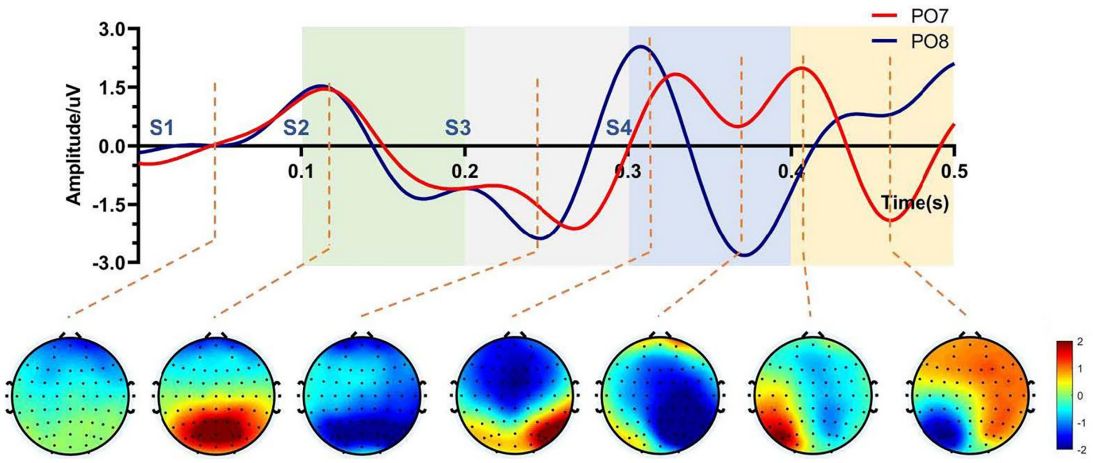


FIGURE 2
Spatial response of character E in one trial, which contains four stimulus patterns S1, S2, S3, and S4. Four colors represent timing windows of spatial responses for four stimulus patterns.

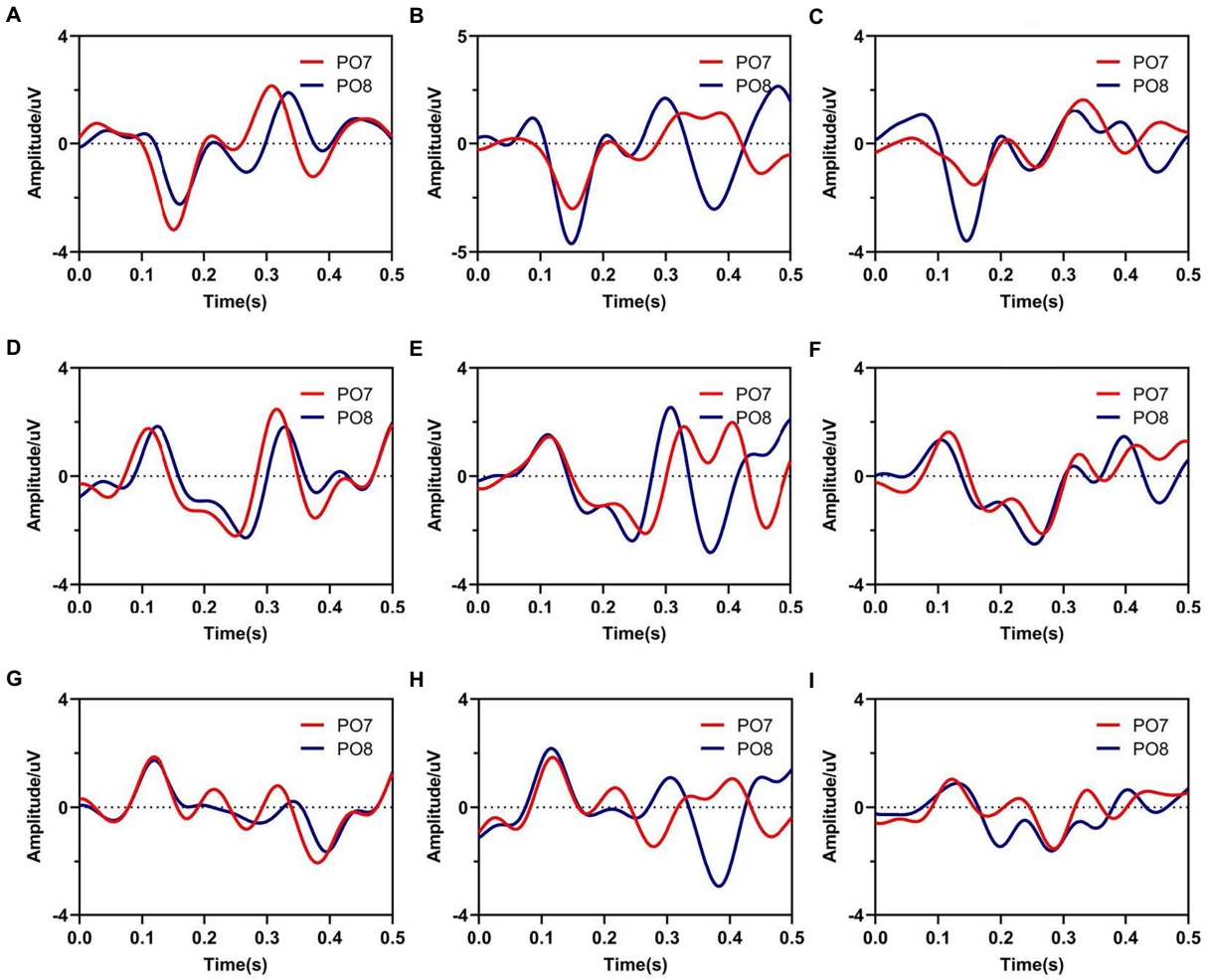


FIGURE 3
Average waveforms from electrodes of PO7 and PO8 when subjects gazed at nine instructions. Subfigures **(A–I)** correspond to real characters **(A–I)** respectively.

presentation of visual stimuli, meanwhile, electrodes of PO7 and PO8 showed negative potentials and peaked around 170 ms. For characters D, E, F, G, H, and I in the middle and bottom row, stimuli were presented above the instruction in pattern of S1, the polarities of PO7 and PO8 were observed inverted in the same time window. These results were highly related to the position between instructions and visual stimuli. For stimulus pattern 1, visual stimuli divided the character matrix into upper and lower parts. Thus, visual stimuli were commonly located in the upper or lower visual field, which resulted in polarity-inverted when subjects gazed at some characters. Similarly, visual stimuli of stimulus pattern 2 also divided the character matrix into upper and lower parts. For responses of S2 during 200–300 ms, characters A, B, C, D, E and F in the top and middle row had similar waveform variation, which was different from characters G, H and I. For stimulus patterns 3 and 4 (EEG responses during 300–400 ms and 400–500 ms), visual stimuli divided the character matrix into left and right parts. Thus, visual stimuli were commonly located in the left or right visual field, which resulted in the lateralization spatial responses of the brain.

In order to indicate the characteristic of spatial responses for all instructions under four stimulus patterns, we further explored the brain topographies elicited by four stimulus patterns. Figure 4 showed the spatial responses at 120 ms and 170 ms after the stimulation of each stimulus pattern, because Figure 3 showed that the responses of those moments had stronger responses after one stimulation. For stimulus patterns 1 and 2, visual stimuli divided the character matrix into upper and lower parts, and the amplitude of most electrodes in the occipital region was reversed in polarity when subjects gazed at corresponding parts. The phenomenon was more significant at 120 ms but weak at 170 ms after the presentation of stimulation. For stimulus patterns 3 and 4, the visual stimuli divided instructions into left and right parts. When visual stimuli were presented to the left or right sides of instructions, maximum responses were elicited in the contralateral hemisphere, which could be found at both 120 ms and 170 ms. At 120 ms, most electrodes of the occipital region performed lateralization of positive potentials, while at 170 ms, they performed negative potentials.

In addition, characters with similar positions around stimuli showed similar spatial responses. This was consistent with the performance of average waveforms shown in Figure 3. For example, in stimulus pattern 1 and stimulus pattern 2, which divided instructions into upper and lower parts, instructions D, E, and F performed similar spatial responses, as were located on the upper

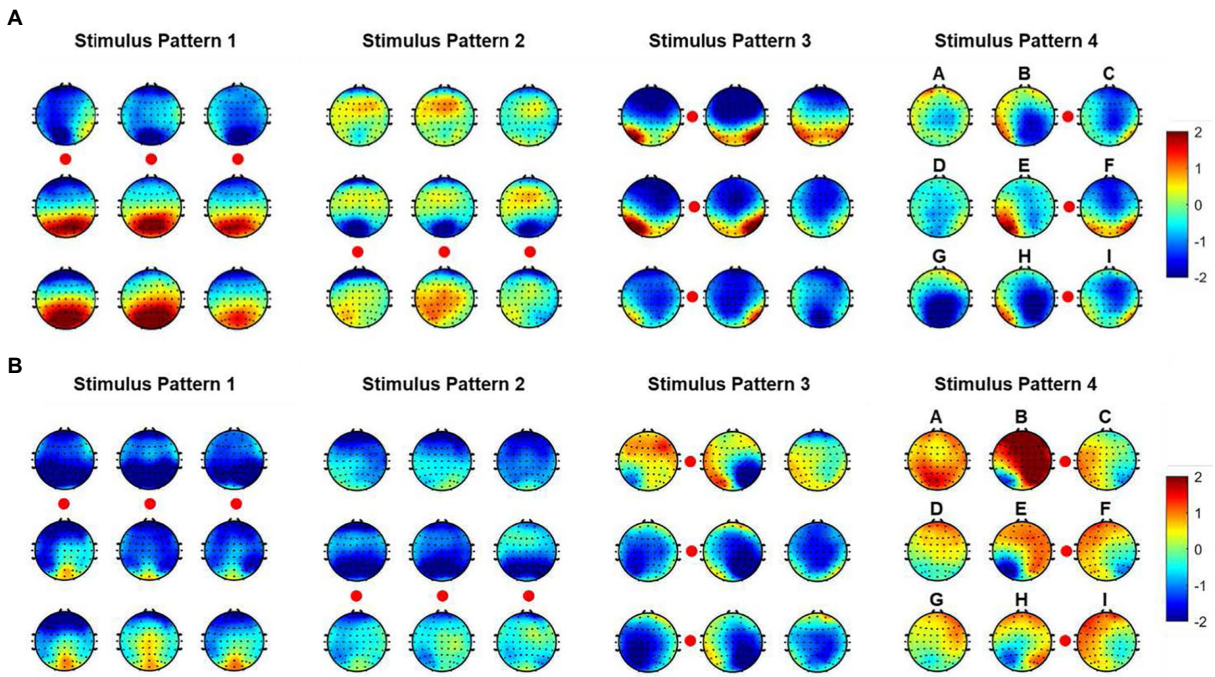


FIGURE 4
Spatial responses at 120 ms and 170 ms after stimulation of each stimulus patterns. Nine topographies correspond to nine instructions, and the red circles correspond to the stimulation locations. **(A)** Spatial responses at 120 ms. **(B)** Spatial responses at 170 ms.

(lower) side of the visual stimulus at the same time. While in stimulus pattern 3 and stimulus pattern 4, which divided instructions into left and right parts, characters B, E, and H showed similar spatial responses. In addition, it could be observed that the strength of spatial responses decreased with increasing distance between instructions and visual stimuli as shown in Figure 4.

Our analysis indicated that when gazing at different characters, ERPs with different spatial specificity could be evoked by four stimulus patterns which were composed of three tiny stimuli. Meanwhile, the characteristic of spatial response was mainly concentrated in the parietal and occipital region, which provided basic reference to electrode selection.

3.2. Classification results for different stimulus patterns

Based on the results of feature analysis in section 3.1, 21 electrodes (P1, P3, P5, P7, Pz, P2, P4, P6, P8, PO3, PO5, PO7, POz, PO4, PO6, PO8, O1, Oz, O2, CB1, CB2) in parietal and occipital region were selected for target identification. Figure 5 indicated the classification results for each of four stimulus patterns based on DSPs. The horizontal coordinate represents the repetition times of trials utilized in pattern matching. The nine-character classification accuracy increased with the repetition times for all stimulus patterns. Specifically, as shown in Figure 5A, the average accuracy of stimulus pattern 1 was 29.63%, 61.07%, 72.39%, 78.60%, 82.18% for 1 to 5 repetitions, respectively. The highest individual accuracy achieved 96.67% for Sub5 at 5 repetitions. In Figure 5B, the average accuracy of stimulus pattern 2, respectively, achieved 51.56%, 70.53%, 76.54%, 80.49%, 83.25% for 1 to 5 repetitions with the highest individual accuracy of 95.19% for Sub7 at 5 repetitions. For Figures 5C,D, the average accuracy performed higher than the accuracy in Figures 5A,B, which achieved 89.42 and 84.49% after 5 repetitions, respectively. Overall, grand-accuracy of all stimulus patterns for 1 to 5 repetitions were higher than 11.11% (random level), which further demonstrated the divisibility of 9 characters in each of stimulus patterns.

3.3. Classification results for target identification in the novel paradigm

In pattern matching for target identification of nine-character matrix, classification results of four stimulus patterns in section 3.2 were employed to joint determine the specific instruction. During the process of target character identification, we constructed templates and spatial filters from training set for each of nine characters under each condition of four stimulus patterns, then used pattern matching to calculate the correlation between the template and test sample after spatial filtering, and lastly confirm the target character based on formula (2) and (3). Figure 6A showed the classification accuracies of all subjects in the offline experiment. The horizontal coordinate represents the repetition times of trials employed in classification, which corresponds to stimulus durations of 0.4 s, 0.8 s, 1.2 s, 1.6 s, and 2 s, respectively. It could be observed that the character accuracies increased with the repetition times, the average classification accuracies across nine participants were 72.06%, 84.98%, 88.77%, 91.97%, and 93.46% for 1 (0.4 s stimulation duration) to 5 (2 s stimulation duration) repetition times. Most subjects performed well with accuracies between 60% and 100%,

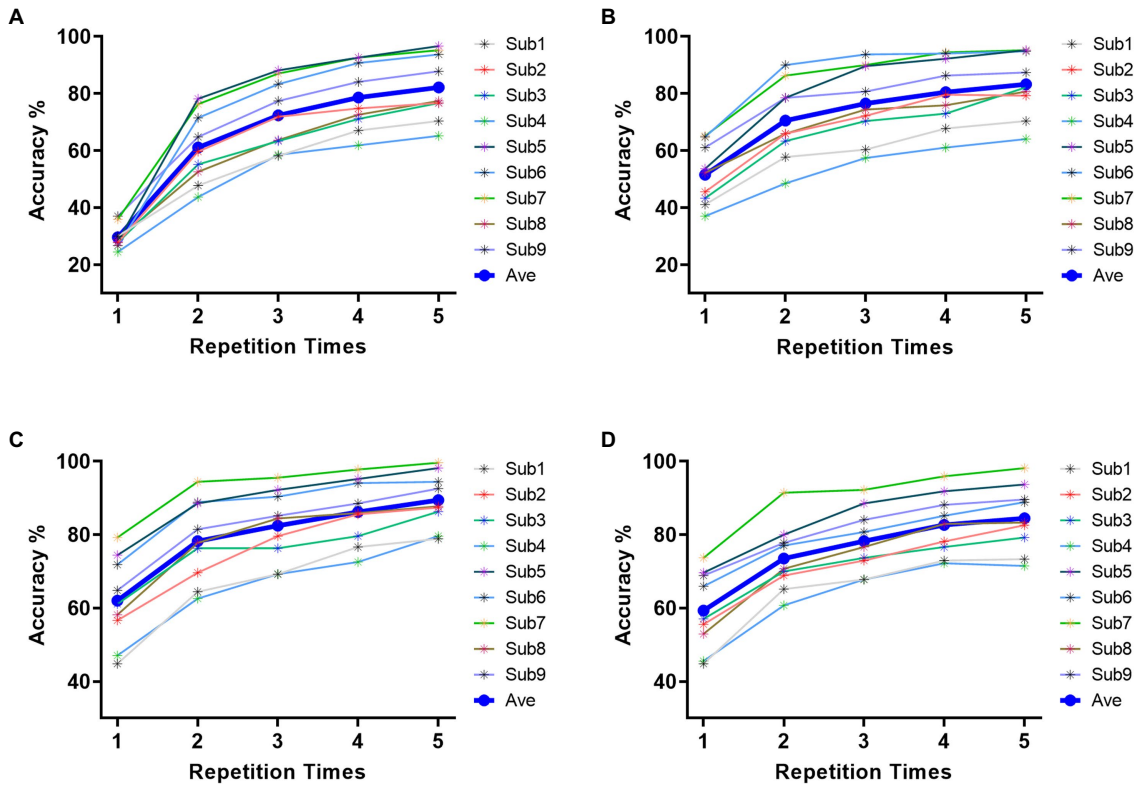


FIGURE 5
Classification accuracy across all subjects of four stimulus patterns. **(A)** classification accuracy of stimulus pattern 1. **(B)** classification accuracy of stimulus pattern 2. **(C)** classification accuracy of stimulus pattern 3. **(D)** classification accuracy of stimulus pattern 4.

subject 5 reached the highest classification accuracy among all subjects of 83.7%, 95.19%, 97.78%, 98.89%, and 99.26% for 1 to 5 repetition times. Seven of nine subjects could achieve accuracy over 90% and four subjects achieved accuracy over 95%. Figure 6B showed offline character ITR for all subjects. The consuming time for each command outputting was composed of cue (0.5 s) and stimulus duration corresponding to repetition times. The average ITRs across nine subjects were 100.39 bits/min, 99.12 bits/min, 83.31 bits/min, 73.11 bits/min, and 63.65 bits/min for 1 to 5 repetition times. Among all subjects, the maximal ITR achieved 135.97 bit/min for Sub5 for 1 repetition time. Notably, ITRs of four subjects out of nine achieved over 120 bits/min, which suggested the proposed v-BCI has the potential for high online performance.

To further evaluate the feasible of the proposed v-BCI, an online task of randomly spelling all characters was set up. For each subject, online repetition times were firstly determined according to the highest ITRs in the offline experiments. Then online test of randomly spelling all instructions twice was notified to each subject. Table 1 showed the performance of online test. It can be illustrated that all subjects performed 1–2 repeated trials for one character selection. The average online ITR could achieve 120.95 bits/min, meanwhile, the highest ITR was 177.5 bits/min (Sub5) and the lowest ITR was 69.28 bits/min (Sub4) in the experiment. The online results were similar to that of the offline experiment. ITRs of five of nine subjects achieved over 100 bits/min in the online test and all of the subjects achieved over 60 bits/min, which demonstrated the feasible of proposed v-BCI. These results demonstrated that this novel paradigm using small and few stimuli has the potential to achieve high performance BCI.

3.4. Fatigue level of subjects

Notably, we asked subjects for their opinions on the comfort level of the proposed paradigm. All of them considered the v-BCI performed friendly both in interaction comfort and performance. Table 2 showed the fatigue level (score: 1–5) by all subjects with an average score of 1.89. The highest fatigue score was 4 (Sub5) while the lowest score was 1 (Sub3, Sub4, and Sub6). These results demonstrated that the proposed paradigm has a friendly interaction mode. Compared with traditional v-BCI, the stimulus presented in our study was more natural for subjects to use, which made it possible to use out of the laboratory environment.

4. Discussion

Traditional v-BCIs such as P300-spellers (Farwell and Donchin, 1988; Jin et al., 2011) and SSVEP-BCIs (Speller, 2015; Jiao et al., 2018) have achieved high ITR performance in previous studies. However, the unfriendly interactive mode of strong and complicated stimuli hinders their practical usage in real life. To address this problem, our study proposed a novel v-BCI which only used three weak stimuli to accomplish nine instructions controlling. The average classification accuracy and ITR in the online spelling test achieved 84.56% and 120.95 bits/min across nine subjects, which suggested the proposed paradigm was a candidate for friendly v-BCI. From the perspective of practical application of the v-BCI, this study still exists something for improving the performance of the system.

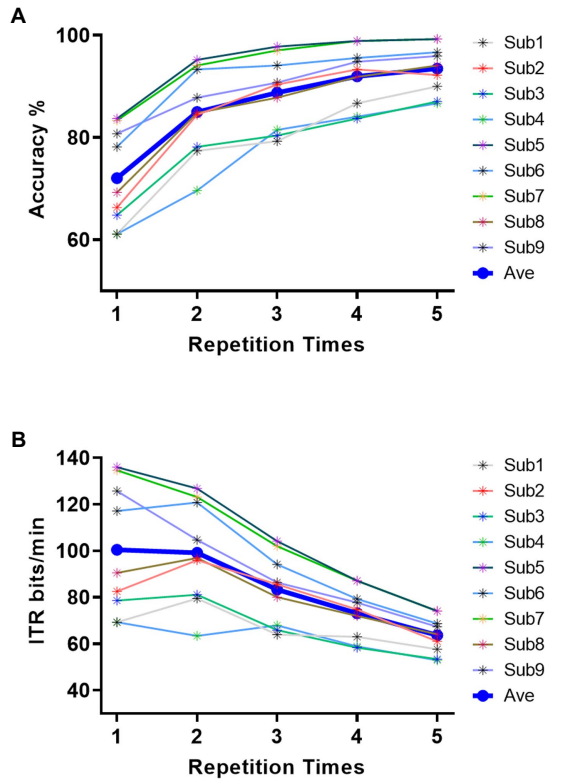


FIGURE 6
Performance for all subjects in offline experiment. **(A)** Character accuracy across all subjects. **(B)** ITR across all subjects.

TABLE 1 Performance of online spelling.

	Repetition times	Accuracy %	ITR bits/min
Sub1	2 (0.5 + 0.8 s)	83.33	92.41
Sub2	2 (0.5 + 0.8 s)	83.33	92.41
Sub3	2 (0.5 + 0.8 s)	94.44	122.88
Sub4	1 (0.5 + 0.4 s)	61.11	69.28
Sub5	1 (0.5 + 0.4 s)	94.44	177.5
Sub6	2 (0.5 + 0.8 s)	100	146.3
Sub7	1 (0.5 + 0.4 s)	88.89	156
Sub8	2 (0.5 + 0.8 s)	83.33	133.48
Sub9	1 (0.5 + 0.4 s)	72.22	98.29
Ave	–	84.56	120.95
Std	–	12.04	35.43

TABLE 2 Fatigue level of the proposed paradigm.

Subjects	Fatigue level score
Sub1	2
Sub2	2
Sub3	1
Sub4	1
Sub5	4
Sub6	1
Sub7	2
Sub8	2
Sub9	2
Ave	1.89
Std	0.93

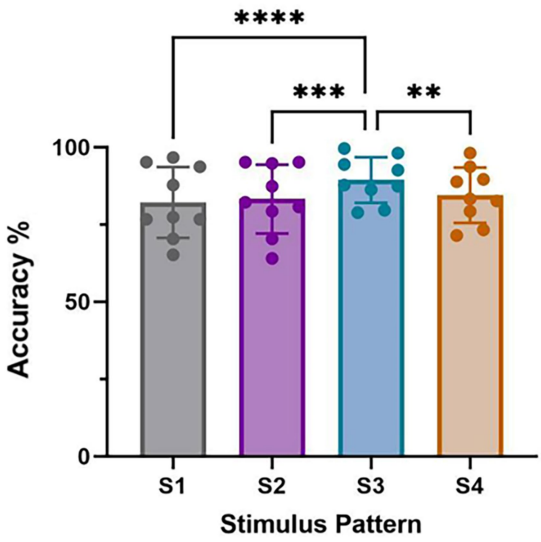


FIGURE 7
Classification accuracy of nine characters for four stimulus patterns at 5 repetitions. (****p<0.0001, ***p<0.001, **p<0.01).

On the one hand, paradigm designing was novel and effective, which could be further optimized in future studies. We have demonstrated the feasibility of character identification in each stimulus pattern and then employed these results for joint decisions. We investigated the performance for four stimulus patterns at 5 repetitions in Figure 5 and performed one-way repeated measures ANOVAs on classification accuracies among four stimulus patterns. As shown in Figure 7, the results manifested that there was a significant difference in character accuracy among different stimulus patterns (F (3, 24) =13.52, p < 0.0001). The accuracy of S3 was significantly higher than S1 (p < 0.0001), S2 (p < 0.001), and S4 (p < 0.01), meanwhile, among S1, S2, and S4, there was no significance of classification accuracy. According to these results, we could consider the superiority of stimulus pattern 3 in paradigm designing. Moreover, although there was no significance between S4 and the other patterns, the accuracy of it was still higher than S1 and S2. This may be attributed to the difference between left and right hemispheres of the brain is higher than that between upper and lower hemispheres (Zhou et al., 2021).

On the other hand, the classification method aimed at this paradigm achieved high performance in online spelling. However, this method still has limitations when applied in complex environments. Specifically, when the number of system instructions becomes larger, the decoding strategy for each character in every situation has low efficiency because of constructing more templates and spatial filters. Thus, decreasing the number of decoding templates becomes significant. For this paradigm using characteristic of spatial features, the relative position and distance

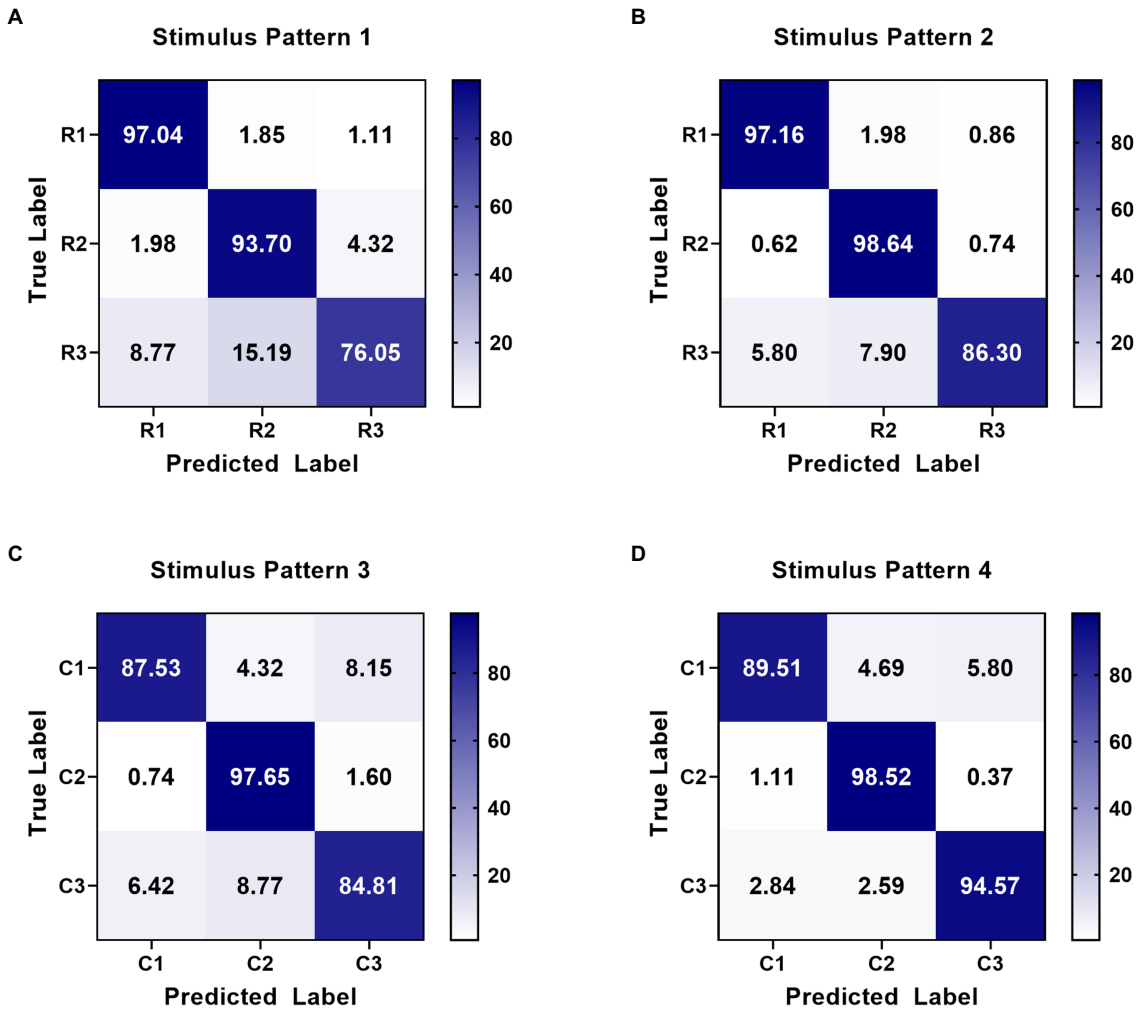


FIGURE 8
Confusion matrices of row-column classification at 5 repetitions for four stimulus patterns. **(A)** Confusion matrix for row identification (R1, R2, and R3) in stimulus pattern 1. **(B)** Confusion matrix for row identification (R1, R2, and R3) in stimulus pattern 2. **(C)** Confusion matrix for column identification (C1, C2, and C3) in stimulus pattern 3. **(D)** Confusion matrix for column identification (C1, C2, and C3) in stimulus pattern 4.

between visual stimuli and the gazed character are two vital aspects of paradigm designing (Zhou et al., 2021). According to the analysis of spatial features in Figure 4, distinct spatial features of EEG were evoked by four stimulus patterns. The spatial responses of the gazed character located in rows or columns under four stimulus patterns were similar. Thus, we could use some instructions to joint construct one template according to their relative position and distance between stimuli and instruction under different stimulus patterns. For stimulus pattern 1 and 2, the visual stimuli divided instructions into three rows (A, B, C), (D, E, F), and (G, H, I), which performed similar spatial responses in the analysis of spatial features. We tag these rows as R1, R2, and R3. For stimulus pattern 3 and pattern 4, the visual stimuli divided instructions into three columns (A, D, G), (B, E, H) and (C, F, I), which were tagged as C1, C2, and C3. We first investigated the divisibility of row or column sequences under four stimulus patterns. As shown in Figure 8, the classification results showed variability between rows or between columns. For identification of rows in S1 and S2, R3 performed easily mixable with R1 and R2. This was possibly due to the position of the gazed character located underneath the stimuli, which was consistent with previous studies (Chen et al., 2018; Zhou et al., 2021). For identification of columns in S3 and S4, the average accuracies were lower than rows, however, all identification accuracies of rows and columns were much higher than the random level for three-target classification (33%), which demonstrated the feasibility of employing row/column identification for one target recognition. Based on these results, each instruction of this 3×3 character matrix could be determined by decoding rows or columns after four times of stimulus using four stimulus patterns, which was similar to the classical classification strategy used in P300-speller. The average accuracies across nine participants were 65.43%, 77.08%, 82.67%, 86.26%, and 88.64% for 1 to 5 repetitions using this new algorithm. The highest accuracy was 97.78% for Sub5 and Sub7 using 1 repetition of stimulus trial. To further evaluate the performance of this algorithm, Figure 9 shows the offline classification of two algorithms for recognition of each character (algorithm 1) vs. recognition of row-column (algorithm 1). One-way repeated measures ANOVAs were performed on classification accuracies between two algorithms. The results revealed that algorithm 1 performed better than algorithm 2 in each

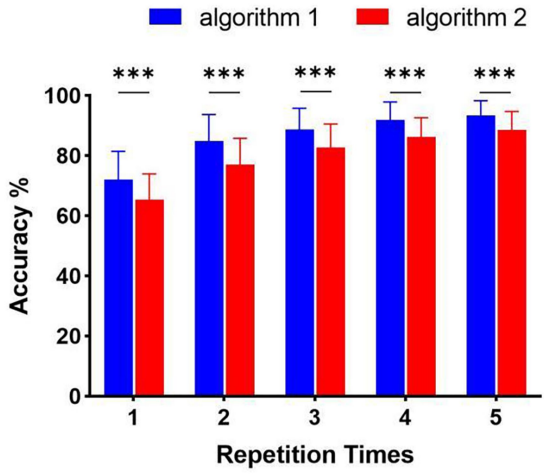


FIGURE 9
Offline classification accuracy of algorithm 1 (recognition of character-by-character) versus algorithm 2 (recognition of row-column) (***p<0.001).

repetition time ($p < 0.001$). However, although the classification of algorithm 2 was lower, it can also achieve 88.64% at 5 repetitions, which showed great performance as well. In addition, algorithm 2 could significantly decrease the number of templates from 9 to 3, which showed great potential in a complex system with large instructions.

In addition to the advantages for above two aspects, this "weak" and "small amount" stimulus-based v-BCI also has a wider application prospect due to its friendliness of interaction, such as application of games based on BCIs. Meanwhile, the proposed BCI system could achieve high ITR performance in order to ensure that the game interaction is smooth.

5. Conclusion

This study proposed a novel v-BCI paradigm using weak and small number of stimuli to accomplish nine instructions controlling. The weak and few stimuli would reduce visual burden and experimental training time, meanwhile, it could also evoke ERPs with characteristics of both spatial and temporal features. A template-matching method based on DSPs was employed to recognize ERPs containing the intention of users. Results of offline and online experiments across nine subjects showed that the average accuracy of offline experiment was 93.46% and the highest online ITR achieved 177.5 bits/min, which demonstrated the feasibility to implement such a friendly v-BCI using this novel paradigm. Furthermore, the proposed paradigm achieved higher ITR than traditional ones using ERPs as the controlled signal, which showed its great potential of being widely used in various fields.

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.

Ethics statement

The studies involving human participants were reviewed and approved by Research Ethics Committee of Tianjin University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

XX and RG designed the study and wrote the manuscript. XZ collected the relevant literatures. MX and DM reviewed and edited the manuscript. All authors read and approved the submitted manuscript.

Funding

This work was supported by the STI 2030—Major Projects 2022ZD0210200, National Natural Science Foundation of China (Nos. 62106170, 62122059, 81925020, and 62006014), and Introduce Innovative Teams of 2021 "New High School 20 Items" Project (2021GXRC071).

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

Chen, J., Li, Z., Hong, B., Maye, A., Engel, A. K., and Zhang, D. (2018). A single-stimulus, multitarget BCI based on retinotopic mapping of motion-onset VEPs. *IEEE Trans. Biomed. Eng.* 66, 464–470. doi: 10.1109/TBME.2018.2849102

Chen, X., Wang, Y., Nakanishi, M., Gao, X., Jung, T.-P., and Gao, S. (2015). High-speed spelling with a noninvasive brain–computer interface. *Proc. Natl. Acad. Sci.* 112, E6058–E6067. doi: 10.1073/pnas.1508080112

Chen, J., Zhang, D., Engel, A. K., Gong, Q., and Maye, A. (2017). Application of a single-flicker online SSVEP BCI for spatial navigation. *PLoS One* 12:e0178385. doi: 10.1371/journal.pone.0178385

Duda, R. O., and Hart, P. E. (2006). *Pattern classification*. Hoboken, NJ: John Wiley & Sons.

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern classification. 2nd Edn*, Hoboken, NJ: John Wiley & Sons. *58*, 16.

Farwell, L. A., and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr. Clin. Neurophysiol.* 70, 510–523. doi: 10.1016/0013-4694(88)90149-6

Han, J., Xu, M., Xiao, X., Yi, W., Jung, T.-P., and Ming, D. (2023). A high-speed hybrid brain-computer interface with more than 200 targets. *J. Neural Eng.* 20:016025. doi: 10.1088/1741-2552/acb105

Jiao, Y., Zhang, Y., Wang, Y., Wang, B., Jin, J., and Wang, X. (2018). A novel multilayer correlation maximization model for improving CCA-based frequency recognition in SSVEP brain–computer interface. *Int. J. Neural Syst.* 28:1750039. doi: 10.1142/S0129065717500393

Jin, J., Allison, B. Z., Sellers, E. W., Brunner, C., Horki, P., Wang, X., et al. (2011). An adaptive P300-based control system. *J. Neural Eng.* 8:036006. doi: 10.1088/1741-2560/8/3/036006

Lebedev, M. A., and Nicolelis, M. A. (2006). Brain–machine interfaces: past, present and future. *Trends Neurosci.* 29, 536–546. doi: 10.1016/j.tins.2006.07.004

Liao, X., Yao, D., Wu, D., and Li, C. (2007). Combining spatial filters for the classification of single-trial EEG in a finger movement task. *IEEE Trans. Biomed. Eng.* 54, 821–831. doi: 10.1109/TBME.2006.889206

Miao, Y., Yin, E., Allison, B. Z., Zhang, Y., Chen, Y., Dong, Y., et al. (2020). An ERP-based BCI with peripheral stimuli: validation with ALS patients. *Cogn. Neurodyn.* 14, 21–33. doi: 10.1007/s11571-019-09541-0

Nakanishi, M., Wang, Y., Chen, X., Wang, Y.-T., Gao, X., and Jung, T.-P. (2017). Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis. *IEEE Trans. Biomed. Eng.* 65, 104–112. doi: 10.1109/TBME.2017.2694818

Speller, I. (2015). A dynamically optimized SSVEP brain-computer. *I.E.E.E. Trans. Biomed. Eng.* 62, 1447–1456. doi: 10.1109/TBME.2014.2320948

Wolpaw, J. R. (2014). The BCI endeavor and the mission of this new journal. *Brain Comput. Interfaces* 1, 2–4. doi: 10.1080/2326263X.2014.884740

Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., et al. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE Trans. Rehabil. Eng.* 8, 164–173. doi: 10.1109/TRE.2000.847807

Wurtz, R. H., and Kandel, E. R. (2000). Central visual pathways. *Principles Neural Sci.* 4, 523–545.

Xu, M., Qi, H., Wan, B., Yin, T., Liu, Z., and Ming, D. (2013). A hybrid BCI speller paradigm combining P300 potential and the SSVEP blocking feature. *J. Neural Eng.* 10:026001. doi: 10.1088/1741-2560/10/2/026001

Xu, M., Xiao, X., Wang, Y., Qi, H., Jung, T.-P., and Ming, D. (2018). A brain–computer interface based on miniature-event-related potentials induced by very small lateral visual stimuli. *IEEE Trans. Biomed. Eng.* 65, 1166–1175. doi: 10.1109/TBME.2018.2799661

Yoshimura, N., Itakura, N., and Fazel, R. (2011). Usability of transient VEPs in BCIs. *Recent Adv. Brain Comput. Interface Syst.*, 119–134. doi: 10.5772/14171

Zhou, X., Xu, M., Xiao, X., Wang, Y., Jung, T.-P., and Ming, D. (2021). Detection of fixation points using a small visual landmark for brain–computer interfaces. *J. Neural Eng.* 18:046098. doi: 10.1088/1741-2552/ac0b51

frontiers | Frontiers in Neuroscience

TYPE Methods
PUBLISHED 30 June 2023
DOI 10.3389/fnins.2023.1213035

OPEN ACCESS

EDITED BY
Peng Xu,
University of Electronic Science and Technology of China, China

REVIEWED BY
Dakun Lai,
University of Electronic Science and Technology of China, China
Yakang Dai,
Chinese Academy of Sciences, China
M. Van Hulle,
KU Leuven, Belgium

*CORRESPONDENCE
Yuanhao Li
✉ li.y.ay@m.titech.ac.jp

RECEIVED 27 April 2023
ACCEPTED 15 June 2023
PUBLISHED 30 June 2023

CITATION
Li Y, Chen B, Wang G, Yoshimura N and Koike Y (2023) Partial maximum correntropy regression for robust electrocorticography decoding. *Front. Neurosci.* 17:1213035. doi: 10.3389/fnins.2023.1213035

Partial maximum correntropy regression for robust electrocorticography decoding

Yuanhao Li[1]*, Badong Chen[2], Gang Wang[3], Natsue Yoshimura[4] and Yasuharu Koike[1]

[1]Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Japan, [2]Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China, [3]Key Laboratory of Biomedical Information Engineering of Ministry of Education, Xi'an Jiaotong University, Xi'an, China, [4]School of Computing, Tokyo Institute of Technology, Yokohama, Japan

The *Partial Least Square Regression* (PLSR) method has shown admirable competence for predicting continuous variables from inter-correlated electrocorticography signals in the brain-computer interface. However, PLSR is essentially formulated with the least square criterion, thus, being considerably prone to the performance deterioration caused by the brain recording noises. To address this problem, this study aims to propose a new robust variant for PLSR. To this end, the maximum correntropy criterion (MCC) is utilized to propose a new robust implementation of PLSR, called *Partial Maximum Correntropy Regression* (PMCR). The half-quadratic optimization is utilized to calculate the robust projectors for the dimensionality reduction, and the regression coefficients are optimized by a fixed-point optimization method. The proposed PMCR is evaluated with a synthetic example and a public electrocorticography dataset under three performance indicators. For the synthetic example, PMCR realized better prediction results compared with the other existing methods. PMCR could also abstract valid information with a limited number of decomposition factors in a noisy regression scenario. For the electrocorticography dataset, PMCR achieved superior decoding performance in most cases, and also realized the minimal neurophysiological pattern deterioration with the interference of the noises. The experimental results demonstrate that, the proposed PMCR could outperform the existing methods in a noisy, inter-correlated, and high-dimensional decoding task. PMCR could alleviate the performance degradation caused by the adverse noises and ameliorate the electrocorticography decoding robustness for the brain-computer interface.

KEYWORDS

brain-computer interface, partial least square regression, maximum correntropy, robustness, electrocorticography decoding

1. Introduction

Brain-computer interface (BCI) has been conceived as a promising technology that translates cerebral recordings generated by cortical neurons into appropriate commands for controlling neuroprosthetic devices (Wolpaw et al., 2002). The capability of BCI for repairing or reproducing sensory-motor functions has been increasingly intensified by recent scientific and technological advances (Donoghue, 2002; Mussa-Ivaldi and Miller, 2003; Lebedev and Nicolelis, 2006). The non-invasive recordings, especially electroencephalogram (EEG) and magnetoencephalogram (MEG), are widely exploited to structure BCI systems

due to their ease of use and satisfactory temporal resolution, whereas the non-invasive BCI systems could be limited in their capabilities and customarily require considerable training (Amiri et al., 2013). Invasive single-unit activities and local field potentials commonly provide better decoding performance, which suffer pessimistic long-term stability, however, due to capriciousness in the recorded neuronal-ensembles (Chestek et al., 2007). A sophisticated alternative which exhibits higher signal amplitudes than EEG while presents superior long-term stability compared with invasive modalities, is the semi-invasive electrocorticography (ECoG) (Buzsáki et al., 2012). Numerous studies in recent years have investigated the potentials of ECoG signal for decoding motions (Levine et al., 2000; Leuthardt et al., 2004; Chin et al., 2007; Pistohl et al., 2008; Ball et al., 2009b; Chao et al., 2010; Shimoda et al., 2012). The serviceability of ECoG signal for online practice have also been demonstrated in Leuthardt et al. (2004, 2006), Schalk et al. (2008).

To accomplish the inter-correlated and potentially high-dimensional ECoG decoding tasks, the *partial least square regression* (PLSR) algorithm has been widely utilized to predict continuous variables from ECoG signals as well as various improved versions in the last decade (Chao et al., 2010; Eliseyev et al., 2011, 2012, 2017; Shimoda et al., 2012; Zhao et al., 2012, 2013; Eliseyev and Aksenova, 2016; Foodeh et al., 2020). Chao et al. (2010) successfully predicted the three-dimensional continuous hand trajectories of two monkeys during asynchronous food-reaching tasks from time-frequency features of subdural ECoG signals by PLSR algorithm. They further showed the admirable prediction capability of PLSR in an epidural ECoG study (Shimoda et al., 2012). Recently, different strategies have been investigated to improve the decoding performance of PLSR. For instance, multi-way PLSR algorithms have been proposed as a generalization for tensor analysis in the ECoG decoding tasks (Bro, 1996; Shimoda et al., 2012; Zhao et al., 2013; Eliseyev et al., 2017). Moreover, regularization technique has been used to penalize the objective function with an extra regularization term to achieve desirable prediction (Eliseyev et al., 2012; Eliseyev and Aksenova, 2016; Foodeh et al., 2020). Although the PLSR algorithm was initially developed for econometrics and chemometrics (Wold, 1966), it has emerged as a popular method for neural imaging and decoding (Krishnan et al., 2011; Zhao et al., 2014).

PLSR solves a regression problem primarily with dimensionality reduction on both explanatory matrix (input) and response matrix (output), in which the dimensionality-reduced samples (commonly called as *latent variables*) for respective sets exhibit maximal correlation, thus structuring association from input variables to output variables. Nevertheless, the conventional PLSR and most existing variants are in essence formulated by the least square criterion, which assigns superfluous importance to the deviated noises. On the other hand, although ECoG signal usually exhibits a relatively higher signal-to-noise ratio (SNR) than the non-invasive EEG recording, previous studies have revealed that ECoG is also prone to be contaminated by physiological artifacts with pronounced amplitudes (Otsubo et al., 2008; Ball et al., 2009a). As a result, PLSR could be incompetent for noisy ECoG decoding tasks due to subnormal robustness.

The present study aims to propose a novel robust version for PLSR through introducing the *maximum correntropy criterion* (MCC) to replace the conventional least square criterion, which was proposed in the *information theoretic learning* (ITL) (Principe, 2010), and has achieved the state-of-the-art robust approaches in different tasks, including regression (Liu et al., 2007; Chen and Príncipe, 2012; Feng et al., 2015), classification (Singh et al., 2014; Ren and Yang, 2018), principal component analysis (He et al., 2011), and feature extraction (Dong et al., 2017). Recently, a rudimentary implementation of the MCC in the PLSR algorithm has been investigated in Mou et al. (2018), where MCC was employed in the process of dimensionality reduction. However, the proposed algorithm in Mou et al. (2018) may be limited in some respects. First, except for the MCC-based dimensionality reduction, it remains acquiring the regression relations under the least square criterion. Second, it only considers the dimensionality reduction for the explanatory matrix. Consequently, one has to calculate the regression coefficients separately for each dimension of the response matrix, which means it could be inadequate for multivariate response prediction.

By comparison, the present study aims to realize a more comprehensive implementation of the MCC framework in PLSR. The main contributions of this study are summarized as follows.

1) We reformulate PLSR thoroughly with the MCC framework, that not only the dimensionality reduction, but also the regression relations between the different variables are established by the MCC framework.
2) Both the explanatory matrix (input) and the response matrix (output) are treated with MCC-based dimensionality reduction. As a result, the proposed algorithm is adequate for multivariate response prediction.
3) We utilize Gaussian kernel functions with individual kernel bandwidths for different reconstruction errors and prediction errors. In addition, each kernel bandwidth value could be calculated from the corresponding set of errors directly.

The remainder of this paper is organized as follows. Section 2 introduces the conventional PLSR method as well as the regularized versions. Section 3 gives a brief introduction about MCC and the rudimentary MCC-based PLSR algorithm. Section 4 presents the reformulation of PLSR with the MCC framework, proposing the *partial maximum correntropy regression* (PMCR) algorithm. Section 5 evaluates the proposed method on synthetic and real ECoG datasets, respectively. Some discussions about the proposed method are given in Section 6. Finally, this paper is concluded in Section 7. To facilitate the presentation of this paper, the main notations are listed in Table 1.

2. Partial least square regression

2.1. Conventional PLSR

Consider the data set with the explanatory matrix $\mathbf{X} \in \mathbb{R}^{L\times N}$ and the response matrix $\mathbf{Y} \in \mathbb{R}^{L\times M}$, in which N and M denote the respective numbers of dimension, while L is the

TABLE 1 Main notations.

Notation	Description
L	Number of observations/samples
N	Dimension of explanatory matrix (input)
M	Dimension of response matrix (output)
S	Optimal number of decomposition factors
s	Current index of decomposition factor
$\mathbf{X}$	Original explanatory matrix (input)
$\mathbf{Y}$	Original response matrix (output)
$\hat{\mathbf{Y}}$	Prediction of response matrix
$\mathbf{X}_s$	Residual matrix of $\mathbf{X}$ in s-th factor
$\mathbf{Y}_s$	Residual matrix of $\mathbf{Y}$ in s-th factor
$\mathbf{x}_s^l$	l-th observation in $\mathbf{X}_s$
$\mathbf{y}_s^l$	l-th observation in $\mathbf{Y}_s$
$\mathbf{w}_s$	Dimensionality-reduction projector for $\mathbf{X}_s$
$\mathbf{c}_s$	Dimensionality-reduction projector for $\mathbf{Y}_s$
$\mathbf{t}_s$	Input latent variables in s-th factor
$\mathbf{u}_s$	Output latent variables in s-th factor
$\mathbf{p}_s$	Loading vector in s-th factor
b_s	Regression coefficient between $\mathbf{t}_s$ and $\mathbf{u}_s$
$g_\sigma(\cdot)$	Gaussian kernel function with kernel bandwidth σ

number of observations. PLSR is an iterative regression method which implements dimensionality reduction and decomposition on explanatory and response matrices simultaneously for S iterations, so that they could be expressed by

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T,\ \mathbf{Y} = \mathbf{T}\mathbf{B}\mathbf{C}^T \tag{1}$$

where $\mathbf{T} = [\mathbf{t}_1, .., \mathbf{t}_S] \in \mathbb{R}^{L\times S}$ and $\mathbf{P} = [\mathbf{p}_1, .., \mathbf{p}_S] \in \mathbb{R}^{N\times S}$ are the latent variables and loading vectors for $\mathbf{X}$, respectively. $\mathbf{C} = [\mathbf{c}_1, .., \mathbf{c}_S] \in \mathbb{R}^{M\times S}$ is the loading vectors of $\mathbf{Y}$, and $\mathbf{B} = \mathrm{diag}(b_1, .., b_S) \in \mathbb{R}^{S\times S}$ is a diagonal matrix. For dimensionality reduction, in the s-th iteration with residual matrices $\mathbf{X}_s$ and $\mathbf{Y}_s$, the covariance between the latent variables $\mathbf{t}_s = \mathbf{X}_s\mathbf{w}_s$ and $\mathbf{u}_s = \mathbf{Y}_s\mathbf{c}_s$ are maximized by

$$\max_{\|\mathbf{w}_s\|_2=\|\mathbf{c}_s\|_2=1} \mathbf{t}_s^T\mathbf{u}_s = \mathbf{w}_s^T\mathbf{X}_s^T\mathbf{Y}_s\mathbf{c}_s \tag{2}$$

in which $\mathbf{w}_s \in \mathbb{R}^N$ and $\mathbf{c}_s \in \mathbb{R}^M$ are utilized for dimensionality reduction on $\mathbf{X}_s$ and $\mathbf{Y}_s$, respectively. $\mathbf{u}_s$ is the latent variable for $\mathbf{Y}_s$. $\|\cdot\|_2$ denotes the L_2-norm. After obtaining the latent variables $\mathbf{t}_s$ and $\mathbf{u}_s$, the loading vector $\mathbf{p}_s$ and the relation from $\mathbf{t}_s$ to $\mathbf{u}_s$ with the scalar b_s are founded by the least square criterion

$$\min_{\mathbf{p}_s} \|\mathbf{X}_s - \mathbf{t}_s\mathbf{p}_s^T\|_2^2 \ \Rightarrow\ \mathbf{p}_s = \mathbf{X}_s^T\mathbf{t}_s/(\mathbf{t}_s^T\mathbf{t}_s) \tag{3}$$

$$\min_{b_s} \|\mathbf{u}_s - \mathbf{t}_s b_s\|_2^2 \ \Rightarrow\ b_s = \mathbf{u}_s^T\mathbf{t}_s/(\mathbf{t}_s^T\mathbf{t}_s) \tag{4}$$

The residual matrices are updated by $\mathbf{X}_{s+1} = \mathbf{X}_s - \mathbf{t}_s\mathbf{p}_s^T$ and $\mathbf{Y}_{s+1} = \mathbf{Y}_s - b_s\mathbf{t}_s\mathbf{c}_s^T$. S is usually selected by cross validation. Eventually, the prediction from $\mathbf{X}$ to $\mathbf{Y}$ is structured by

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{H} \tag{5}$$

where $\mathbf{H} = \mathbf{P}^{T+}\mathbf{B}\mathbf{C}^T \in \mathbb{R}^{N\times M}$, and $\mathbf{P}^{T+}$ is the pseudo-inverse of $\mathbf{P}^T$. $\hat{\mathbf{Y}}$ denotes the prediction for $\mathbf{Y}$.

Maximizing the covariance between latent variables Eq. (2) could be rewritten as (Barker and Rayens, 2003).

$$\min_{\|\mathbf{w}_s\|_2=\|\mathbf{c}_s\|_2=1} \sum_{l=1}^{L} \begin{pmatrix} \|\mathbf{x}_s^l - \mathbf{x}_s^l\mathbf{w}_s\mathbf{w}_s^T\|^2 \\ +\|\mathbf{y}_s^l - \mathbf{y}_s^l\mathbf{c}_s\mathbf{c}_s^T\|^2 \\ +\|\mathbf{x}_s^l\mathbf{w}_s - \mathbf{y}_s^l\mathbf{c}_s\|^2 \end{pmatrix} \tag{6}$$

where $\mathbf{x}_s^l$ and $\mathbf{y}_s^l$ denote the l-th samples in $\mathbf{X}_s$ and $\mathbf{Y}_s$, respectively. One can observe that, PLSR employs the least square criterion not only to obtain the regression relations in Eqs. (3, 4), but for the projectors $\mathbf{w}_s$ and $\mathbf{c}_s$ as well. In Eq. (6), the first and second terms are the reconstruction errors for input and output, respectively. The third term denotes the prediction error for the l-th latent variables. Since each step for PLSR is based on the least square criterion, the prediction from input to output could be seriously deteriorated by noises.

2.2. Regularized PLSR

Regularization technique has been popularly employed to ameliorate the decoding performance of the PLSR algorithm. For example, L_1-regularization on the projectors was employed so as to acquire sparse projectors, conducting the feature selection simultaneously (Eliseyev et al., 2012). The authors further extended their study in Eliseyev and Aksenova (2016), in which Sobolev-norm and polynomial penalization were introduced into PLSR algorithm to strengthen the smoothness of the predicted response. Recently, the state-of-the-art regularized PLSR was proposed by utilizing L_2-regularization to find the regression relation between the latent variables $\mathbf{t}_s$ and $\mathbf{u}_s$, so as to reduce the over-fitting risk of each latent variable on the desired response (Foodeh et al., 2020). In particular, for each decomposition factor, the scalar b_s is acquired with an individual regularization parameter λ_s as

$$\min_{b_s} \|\mathbf{u}_s - \mathbf{t}_s b_s\|_2^2 + \lambda_s b_s^2 \ \Rightarrow\ b_s = \mathbf{u}_s^T\mathbf{t}_s/(\mathbf{t}_s^T\mathbf{t}_s + \lambda_s) \tag{7}$$

Experimental results in Foodeh et al. (2020) showed that, the regularization technique in Eq. (7) can achieve better ECoG decoding performance than regularizing the projectors.

Nevertheless, the regularized PLSR variants remain formulated based on the non-robust least square criterion, as a result, being still prone to suffering the performance deterioration caused by the adverse noises.

3. Maximum correntropy criterion

3.1. Maximum correntropy criterion

The correntropy concept was developed in the field of ITL as a generalized correlation function of random processes

(Santamaría et al., 2006), which measures the similarity and interaction between two vectors in a kernel space. Correntropy associates with the information potential of quadratic Renyi's entropy (Liu et al., 2007), where the data's probability density function (PDF) is estimated by the Parzen's window method (Parzen, 1962; Silverman, 1986). The correntropy which evaluates the similarity between two arbitrary variables A and B, is defined by

$$\mathcal{V}(A,B) = E[k(A-B)] \tag{8}$$

in which $k(\cdot)$ is a kernel function satisfying the Mercer's theory and $E[\cdot]$ is the expectation operator. In the practical application, one calculates the correntropy with L observations by the following empirical estimation

$$\hat{\mathcal{V}}(A,B) = \frac{1}{L}\sum_{l=1}^{L} k(a_l - b_l) \tag{9}$$

where the Gaussian kernel function $g_\sigma(x) \triangleq \exp(-x^2/2\sigma^2)$ with kernel bandwidth σ is widely used for the kernel function $k(\cdot)$, thus leading to

$$\hat{\mathcal{V}}(A,B) = \frac{1}{L}\sum_{l=1}^{L} g_\sigma(a_l - b_l) = \frac{1}{L}\sum_{l=1}^{L} \exp(-\frac{(a_l-b_l)^2}{2\sigma^2}) \tag{10}$$

Maximizing the correntropy Eq. (10), called as the *maximum correntropy criterion* (MCC), exhibits numerous advantages. Correntropy is essentially a local similarity measure, which is chiefly determined along $A = B$, i.e. zero-value error. Consequently, the effect of large error caused by adverse noise is alleviated, leading to superior robustness. Additionally, correntropy could extract sufficient information from observations, since it considers all the even moments of errors (Liu et al., 2007). It also relates closely to the m-estimation, which can be regarded as a robust formulation of Welsch m-estimator (Huber, 2004).

3.2. MCC-PLSR

Recently, a rudimentary MCC-based PLSR variant has been investigated in Mou et al. (2018), named as MCC-PLSR. For a univariate output, according to Mou et al. (2018), the dimensionality reduction Eq. (2) could be rewritten as

$$\max_{\|\mathbf{w}_s\|_2=1} \mathbf{w}_s^T \mathbf{X}_s^T \mathbf{Y}_s \mathbf{Y}_s^T \mathbf{X}_s \mathbf{w}_s \tag{11}$$

which aims to maximize the quadratic covariance. Mou et al. (2018) utilized a similar proposition as in the MCC-based principal component analysis (He et al., 2011), proposing the following objective function

$$\max_{\|\mathbf{w}_s\|_2=1} \sum_{l=1}^{L} g_\sigma(\sqrt{\mathbf{y}_s^{lT}\mathbf{x}_s^l\mathbf{x}_s^{lT}\mathbf{y}_s^l - \mathbf{y}_s^{lT}\mathbf{x}_s^l\mathbf{w}_s\mathbf{w}_s^T\mathbf{x}_s^{lT}\mathbf{y}_s^l}) \tag{12}$$

from which one can calculate the robust projector $\mathbf{w}_s$. Then, one obtains the latent variables by $\mathbf{t}_s = \mathbf{X}_s\mathbf{w}_s$, and acquires other model parameters similarly as in Eqs. (3-5).

Despite the robust implementation of the projector $\mathbf{w}_s$ in Eq. (12), the above-described MCC-PLSR algorithm could be inadequate for the following reasons. First, except for the calculation of $\mathbf{w}_s$, the other model parameters are still acquired under the least square criterion. Second, dimensionality reduction is not considered for the output matrix. As a result, the prediction performance for multivariate response could be limited. In addition, MCC-PLSR is prone to suffering excessive computation time, since one has to obtain the prediction model $\hat{\mathbf{Y}} = \mathbf{XH}$ for each dimension of the response matrix separately.

4. Partial maximum correntropy regression

In this section, we present a comprehensive reformulation of PLSR with the MCC framework. Compared with the existing MCC-PLSR, our proposed method aims to acquire each model parameter under the MCC. In addition, the generalization for multivariate response prediction is taken into account in this study. The detailed mathematical derivations of the proposed method are given as follows, in which the subscript s denoting the s-th decomposition factor is omitted for the purpose of simplicity.

Substituting the least quadratic reconstruction errors and prediction errors in the conventional PLSR Eq. (6) with the maximum correntropy yields

$$\max_{\|\mathbf{w}\|_2=\|\mathbf{c}\|_2=1} \sum_{l=1}^{L} \begin{pmatrix} g_{\sigma_x}(\mathbf{x}^l - \mathbf{x}^l\mathbf{w}\mathbf{w}^T) \\ +g_{\sigma_y}(\mathbf{y}^l - \mathbf{y}^l\mathbf{c}\mathbf{c}^T) \\ +g_{\sigma_r}(\mathbf{x}^l\mathbf{w} - \mathbf{y}^l\mathbf{c}) \end{pmatrix} \tag{13}$$

where σ_x, σ_y, and σ_r denote the Gaussian kernel bandwidths for $\mathbf{X}$-reconstruction errors, $\mathbf{Y}$-reconstruction errors, and the prediction errors, respectively.

Then, one can transform the vectors $(\mathbf{x}^l - \mathbf{x}^l\mathbf{w}\mathbf{w}^T)$ and $(\mathbf{y}^l - \mathbf{y}^l\mathbf{c}\mathbf{c}^T)$ into scalars, provided that the two projectors $\mathbf{w}$ and $\mathbf{c}$ are unit-length vectors, i.e. $\mathbf{w}^T\mathbf{w} = \mathbf{c}^T\mathbf{c} = 1$,

$$\begin{aligned} \sqrt{\|\mathbf{x}^l - \mathbf{x}^l\mathbf{w}\mathbf{w}^T\|^2} &= \sqrt{\mathbf{x}^l\mathbf{x}^{lT} - \mathbf{x}^l\mathbf{w}\mathbf{w}^T\mathbf{x}^{lT}} \\ \sqrt{\|\mathbf{y}^l - \mathbf{y}^l\mathbf{c}\mathbf{c}^T\|^2} &= \sqrt{\mathbf{y}^l\mathbf{y}^{lT} - \mathbf{y}^l\mathbf{c}\mathbf{c}^T\mathbf{y}^{lT}} \end{aligned} \tag{14}$$

Subsequently, one obtains the following optimization problem to acquire the projectors

$$\max_{\|\mathbf{w}\|_2=\|\mathbf{c}\|_2=1} \sum_{l=1}^{L} \begin{pmatrix} g_{\sigma_x}(\sqrt{\mathbf{x}^l\mathbf{x}^{lT} - \mathbf{x}^l\mathbf{w}\mathbf{w}^T\mathbf{x}^{lT}}) \\ +g_{\sigma_y}(\sqrt{\mathbf{y}^l\mathbf{y}^{lT} - \mathbf{y}^l\mathbf{c}\mathbf{c}^T\mathbf{y}^{lT}}) \\ +g_{\sigma_r}(\mathbf{x}^l\mathbf{w} - \mathbf{y}^l\mathbf{c}) \end{pmatrix} \tag{15}$$

After obtaining $\mathbf{w}$ and $\mathbf{c}$, one could calculate the latent variables as in the conventional PLSR by $\mathbf{t} = \mathbf{Xw}$ and $\mathbf{u} = \mathbf{Yc}$. We then calculate the loading vector $\mathbf{p}$ and the regression coefficient b under the MCC by

$$\max_{\mathbf{p}} \sum_{l=1}^{L} g_{\sigma_p}(\mathbf{x}^l - \mathbf{t}^l\mathbf{p}^T) \tag{16}$$

$$\max_{b} \sum_{l=1}^{L} g_{\sigma_b}(\mathbf{u}^l - \mathbf{t}^l b) \tag{17}$$

in which $\mathbf{t}^l$ and $\mathbf{u}^l$ denote the l-th elements for the latent variables $\mathbf{t}$ and $\mathbf{u}$, respectively. σ_p and σ_b denote the corresponding Gaussian kernel bandwidths. The residual matrices are then updated similarly as PLSR.

One repeats such procedures for the optimal number of factors and collects the acquired vectors from each iteration to organize the matrices $\mathbf{T}$, $\mathbf{P}$, $\mathbf{B}$, and $\mathbf{C}$, as in the original PLSR. Ultimately, the predicted response $\hat{\mathbf{Y}}$ can be obtained from $\mathbf{X}$ by the regression relationship Eq. (5). The above-mentioned PLSR variant which is comprehensively reformulated based on the MCC, is named as *partial maximum correntropy regression* (PMCR). In what follows, we discuss in detail about the optimization, convergence analysis, and determination of hyper-parameters with regard to the proposed PMCR algorithm.

4.1. Optimization

Three optimization problems Eqs. (15, 16, 17) need to be addressed in PMCR. We first consider Eq. (15) for the calculation of the projectors $\mathbf{w}$ and $\mathbf{c}$. Based on the *half-quadratic* (HQ) optimization method (Ren and Yang, 2018), Eq. (15) could be rewritten as

$$\max_{\substack{\|\mathbf{w}\|_2=\\ \|\mathbf{c}\|_2=1}} \sum_{l=1}^{L} \begin{pmatrix} \sup\{\alpha_l \frac{\mathbf{x}^l\mathbf{x}^{lT}-\mathbf{x}^l\mathbf{w}\mathbf{w}^T\mathbf{x}^{lT}}{2\sigma_x^2} - \varphi(\alpha_l)\} \\ +\sup\{\beta_l \frac{\mathbf{y}^l\mathbf{y}^{lT}-\mathbf{y}^l\mathbf{c}\mathbf{c}^T\mathbf{y}^{lT}}{2\sigma_y^2} - \varphi(\beta_l)\} \\ +\sup\{\gamma_l \frac{(\mathbf{x}^l\mathbf{w}-\mathbf{y}^l\mathbf{c})^2}{2\sigma_r^2} - \varphi(\gamma_l)\} \end{pmatrix} \tag{18}$$

where $\varphi(\cdot)$ is a convex conjugated function of $g(\cdot)$, and $\{\alpha_l\}_{l=1}^L$, $\{\beta_l\}_{l=1}^L$, and $\{\gamma_l\}_{l=1}^L$ denote three sets of introduced auxiliaries, respectively. Thus, we can conclude that optimizing Eq. (15) is equivalent to updating $(\alpha_l, \beta_l, \gamma_l)$ and $(\mathbf{w}, \mathbf{c})$ alternately by

$$\max_{\substack{\|\mathbf{w}\|_2=\\ \|\mathbf{c}\|_2=1,\\ \alpha_l,\beta_l,\gamma_l}} J \triangleq \sum_{l=1}^{L} \begin{pmatrix} \alpha_l \frac{\mathbf{x}^l\mathbf{x}^{lT}-\mathbf{x}^l\mathbf{w}\mathbf{w}^T\mathbf{x}^{lT}}{2\sigma_x^2} - \varphi(\alpha_l) \\ +\beta_l \frac{\mathbf{y}^l\mathbf{y}^{lT}-\mathbf{y}^l\mathbf{c}\mathbf{c}^T\mathbf{y}^{lT}}{2\sigma_y^2} - \varphi(\beta_l) \\ +\gamma_l \frac{(\mathbf{x}^l\mathbf{w}-\mathbf{y}^l\mathbf{c})^2}{2\sigma_r^2} - \varphi(\gamma_l) \end{pmatrix} \tag{19}$$

Since the HQ optimization is an iterative process, we denote the k-th HQ iteration with the subscript k. First, according to the HQ technique (Ren and Yang, 2018), we update the auxiliaries with the current projectors $(\mathbf{w}_k, \mathbf{c}_k)$ by

$$\begin{aligned} \alpha_{l,k+1} &= -\exp\left(-\frac{\mathbf{x}^l\mathbf{x}^{lT} - \mathbf{x}^l\mathbf{w}_k\mathbf{w}_k^T\mathbf{x}^{lT}}{2\sigma_x^2}\right) \\ \beta_{l,k+1} &= -\exp\left(-\frac{\mathbf{y}^l\mathbf{y}^{lT} - \mathbf{y}^l\mathbf{c}_k\mathbf{c}_k^T\mathbf{y}^{lT}}{2\sigma_y^2}\right) \\ \gamma_{l,k+1} &= -\exp\left(-\frac{(\mathbf{x}^l\mathbf{w}_k - \mathbf{y}^l\mathbf{c}_k)^2}{2\sigma_r^2}\right) \\ (l &= 1, .., L) \end{aligned} \tag{20}$$

Then, to optimize the projectors, we rewrite Eq. (19) by collecting the terms of projectors and omitting the auxiliaries as

$$\max_{\substack{\|\mathbf{w}\|_2=\\ \|\mathbf{c}\|_2=1}} J_p \triangleq \sum_{l=1}^{L} \begin{pmatrix} (\frac{\gamma_l}{2\sigma_r^2} - \frac{\alpha_l}{2\sigma_x^2})\mathbf{x}^l\mathbf{w}\mathbf{w}^T\mathbf{x}^{lT} \\ +(\frac{\gamma_l}{2\sigma_r^2} - \frac{\beta_l}{2\sigma_y^2})\mathbf{y}^l\mathbf{c}\mathbf{c}^T\mathbf{y}^{lT} \\ -\frac{\gamma_l}{\sigma_r^2}\mathbf{x}^l\mathbf{w}\mathbf{c}^T\mathbf{y}^{lT} \end{pmatrix} \tag{21}$$

which is a quadratic optimization issue constrained by nonlinear conditions. To accomplish Eq. (21), there exist enormous solutions in the literature, such as the sequential quadratic programming (SQP) which has been widely utilized for nonlinear programming problems (Fletcher, 2013).

After one obtains the projectors $\mathbf{w}$ and $\mathbf{c}$, the latent variables are computed by $\mathbf{t} = \mathbf{X}\mathbf{w}$ and $\mathbf{u} = \mathbf{Y}\mathbf{c}$. Then, Eqs. (16, 17) can be solved by the following iterative fixed-point optimization method with fast convergence (Chen et al., 2015).

$$\mathbf{p} = \mathbf{X}^T\Psi_\mathbf{p}\mathbf{t}/(\mathbf{t}^T\Psi_\mathbf{p}\mathbf{t}) \tag{22}$$

$$b = \mathbf{u}^T\Psi_b\mathbf{t}/(\mathbf{t}^T\Psi_b\mathbf{t}) \tag{23}$$

where $\Psi_\mathbf{p}$ and Ψ_b are $L \times L$ diagonal matrices with the diagonal elements $(\Psi_\mathbf{p})_{l,l} = g_{\sigma_p}(\mathbf{x}^l - \mathbf{t}^l\mathbf{p}^T)$ and $(\Psi_b)_{l,l} = g_{\sigma_b}(\mathbf{u}^l - \mathbf{t}^l b)$, respectively. Since $\Psi_\mathbf{p}$ and Ψ_b are dependent on the current solutions $\mathbf{p}$ and b, the updates in Eqs. (22, 23) are fixed-point equations which will require multiple iterations (Chen et al., 2015). The comprehensive procedures for PMCR are summarized in Algorithm 1.

Input: matrices of explanation $\mathbf{X}$ and response $\mathbf{Y}$; number of factors S; a small positive value ς
Output: prediction model $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{H}$
initialize $\mathbf{X}_1 = \mathbf{X}$ and $\mathbf{Y}_1 = \mathbf{Y}$;
for $s = 1, 2, .., S$ **do**
  initialize the projectors by the conventional PLSR;
  initialize *converged* = FALSE;
  repeat
    auxiliary-step: update $(\alpha_l, \beta_l, \gamma_l)$ with (20);
    projector-step: update $(\mathbf{w}_s, \mathbf{c}_s)$ with (21);
    if the difference of the objective function (15) is smaller than ς **then**
      converged = TRUE
    end if
  until *converged* == TRUE
  compute latent variables $\mathbf{t}_s = \mathbf{X}_s\mathbf{w}_s$ and $\mathbf{u}_s = \mathbf{Y}_s\mathbf{c}_s$;
  compute $\mathbf{p}_s$ and b_s by the fixed-point method (22)(23);
  update the residual matrices $\mathbf{X}_{s+1} = \mathbf{X}_s - \mathbf{t}_s\mathbf{p}_s^T$ and $\mathbf{Y}_{s+1} = \mathbf{Y}_s - b_s\mathbf{t}_s\mathbf{c}_s^T$;
end for
organize the matrices $\mathbf{T} = [\mathbf{t}_1, .., \mathbf{t}_S]$, $\mathbf{P} = [\mathbf{p}_1, .., \mathbf{p}_S]$, $\mathbf{B} = \text{diag}(b_1, .., b_S)$, and $\mathbf{C} = [\mathbf{c}_1, .., \mathbf{c}_S]$;
compute $\mathbf{H} = \mathbf{P}^{T+}\mathbf{B}\mathbf{C}^T$

Algorithm 1. Partial maximum correntropy regression

4.2. Convergence analysis

For the regression relations $\mathbf{p}$ and b, one could find the detailed convergence analysis in Chen et al. (2015). We mainly consider the convergence of the projectors $\mathbf{w}$ and $\mathbf{c}$ in the optimization problem (15). Because correntropy is in nature an m-estimator (Liu et al., 2007), the local optimums of Eq. (15) will be close sufficiently to the global optimum, which has been proved in a recent theoretical study (Loh and Wainwright, 2015). Therefore, we prove that Eq. (15) will converge to a local optimum with the HQ optimization method.

Proposition 1. If we have $J_p(\mathbf{w}_k, \mathbf{c}_k) \leqslant J_p(\mathbf{w}_{k+1}, \mathbf{c}_{k+1})$ by fixing $(\alpha_l, \beta_l, \gamma_l) = (\alpha_{l,k+1}, \beta_{l,k+1}, \gamma_{l,k+1})$, the optimization problem (Eq. 15) will converge to a local optimum.

Proof: The convergence is proved as

$$\begin{aligned} &J(\mathbf{w}_k, \mathbf{c}_k, \alpha_{l,k}, \beta_{l,k}, \gamma_{l,k}) \\ \leqslant &J(\mathbf{w}_k, \mathbf{c}_k, \alpha_{l,k+1}, \beta_{l,k+1}, \gamma_{l,k+1}) \\ \leqslant &J(\mathbf{w}_{k+1}, \mathbf{c}_{k+1}, \alpha_{l,k+1}, \beta_{l,k+1}, \gamma_{l,k+1}) \end{aligned} \tag{24}$$

in which the first inequality is guaranteed by the HQ mechanism (Ren and Yang, 2018), and the second inequality arises from the assumption of the present proposition.

One can observe that, to guarantee the convergence of Eq. (15), it is unnecessary to attain the strict maximum of Eq. (21) at each projector-step in Algorithm 1. On the contrary, so long as the updated projectors lead to a larger objective function J_p at each projector-step, Eq. (15) will converge to a local optimum. This reveals great convenience in practice, that one only needs a few SQP iterations for projector-step. One could finish the projector-step once confirming the increase on J_p, thus accelerating the convergence.

4.3. Hyper-parameter determination

There exist five Gaussian kernel bandwidths σ_x, σ_y, σ_r, σ_p, and σ_b, respectively, to be determined in practice. In the literature, an effective method to estimate a proper kernel bandwidth for probability density estimation, named as *Silverman's rule*, was proposed in Silverman (1986). Denoting the current set of errors as E with L observations, the kernel bandwidth is computed

$$\sigma^2 = 1.06 \times \min\{\sigma_E, \frac{R}{1.34}\} \times (L)^{-1/5} \tag{25}$$

in which σ_E is the standard deviation of the L errors, and R denotes the interquartile range.

5. Experiments

In this section, we assessed the proposed PMCR algorithm on a synthetic dataset and a real ECoG dataset, respectively, comparing it with the existing PLSR methods. Specifically, we compared PMCR to the following methods: the conventional PLSR, the state-of-the-art regularized PLSR (RPLSR) (Foodeh et al., 2020) described in Eq. (7), and the rudimentary MCC-PLSR (Mou et al., 2018) described in Section 3.2. For a evenhanded comparison, each algorithm used an identical number of factors, which was selected by the conventional PLSR in five-fold cross-validation. The maximal number of factors was set as 100.

Considering the performance indicators for the evaluation, we used three typical measures in regression tasks: i) Pearson's correlation coefficient (r)

$$r = \frac{Cov(\hat{\mathbf{Y}}, \mathbf{Y})}{\sqrt{Var(\hat{\mathbf{Y}})Var(\mathbf{Y})}} \tag{26}$$

where $Cov(\cdot,\cdot)$ and $Var(\cdot)$ denote the covariance and variance, respectively, and ii) root mean squared error (RMSE) which is computed by

$$\text{RMSE} = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\|\hat{\mathbf{y}}_l - \mathbf{y}_l\|^2} \tag{27}$$

in which $\hat{\mathbf{y}}_l$ and $\mathbf{y}_l$ denote the l-th observations for the prediction $\hat{\mathbf{Y}}$ and the target $\mathbf{Y}$, respectively, and iii) mean absolute error (MAE) which represents the average L_1-norm distance

$$\text{MAE} = \frac{1}{L}\sum_{l=1}^{L}\|\hat{\mathbf{y}}_l - \mathbf{y}_l\| \tag{28}$$

To compare the robustness between different algorithms, only contaminating the training samples by noises with isolating testing data from contamination is an extensively approved and implemented method in the literature, as advised in Zhu and Wu (2004). Accordingly, only the training sample would suffer the adverse contamination in the following experiments.

5.1. Synthetic dataset

5.1.1. Dataset description

First, we considered an inter-correlated, high-dimensional, and noisy synthetic example, in which various PLSR methods were assessed with different levels of contamination. Randomly, we generated 300 i.i.d.[1] latent variables $\mathbf{t} \sim U(0,1)$ for training, and 300 i.i.d. latent variables $\mathbf{t} \sim U(0,1)$ for testing, in which U denotes the uniform distribution, and the dimension of $\mathbf{t}$ was set as 20. We generated the hypothesis from the latent variable to the explanatory and response matrices then. Specifically, we randomly generated the transformation matrices with arbitrary values, which were subject to the standard normal distribution. The latent variables $\mathbf{t}$ were multiplied with a 20×500 transformation matrix, resulting in a 300×500 explanatory matrix for input. Similarly, we used a 20×3 transformation matrix to acquire a 300×3 response matrix for output. Accordingly, we predicted the multivariate responses from 500-dimensional explanatory variables with 300 training samples, and evaluated the prediction performance on the other 300 testing samples.

Considering the contamination for the synthetic dataset, we supposed the explanatory matrix to be contaminated, because the

1 Independent and identically distributed.

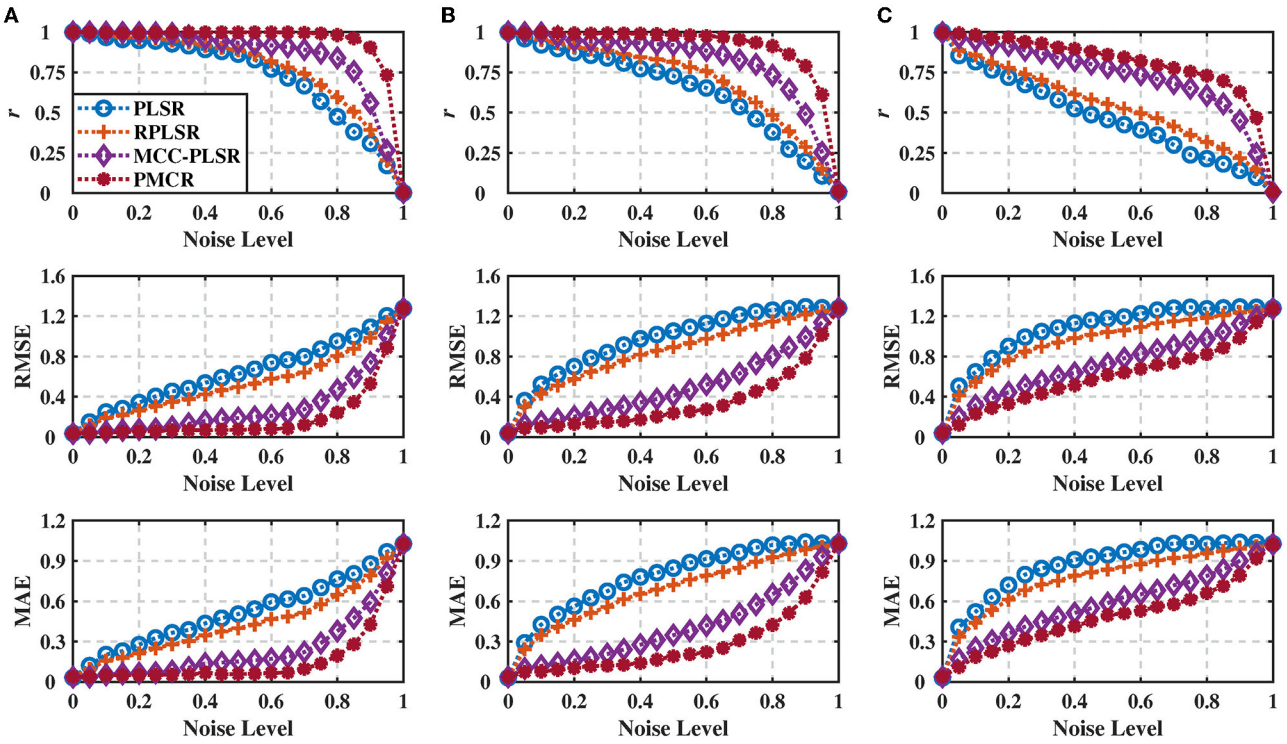


FIGURE 1
Regression performance indicators of the inter-correlated, high-dimensional, and contaminated synthetic dataset under different noise standard deviations with noise levels from 0 to 1.0. **(A)** Noise standard deviation = 30, **(B)** noise standard deviation = 100, and **(C)** noise standard deviation = 300. The performance indicators were acquired from 100 Monte-Carlo repetitive trials and averaged across three dimensions of the output. The proposed PMCR algorithm realized better performance than the existing PLSR algorithms consistently for *r*, RMSE, and MAE, in particular when the training set was contaminated considerably.

adverse noises mainly happen to the brain recordings, which are usually used as the explanatory in the BCI system. Therefore, a certain proportion (from 0 to 1.0 with a step 0.05) of training samples were randomly selected with equal probability, the inputs of which were then replaced by noises with large amplitude. For the distribution of the noise, we utilized a zero-mean Gaussian distribution with large standard deviation to imitate outliers, where 30, 100, and 300 were used, respectively.

5.1.2. Results

We evaluated the various PLSR methods with 100 Monte-Carlo repetitive trials, and present the results in Figure 1, where the results were averaged across three dimensions of the output. One could observe from Figure 1 that, for all the three different noise standard deviations, the proposed PMCR algorithm achieved superior prediction performance compared with the other existing methods consistently for *r*, RMSE, and MAE, respectively, in particular when the training set suffered considerable contamination.

The number of factors S plays a vital role in PLSR methods, representing the iteration numbers to decompose the input and output matrices. Since it usually causes a notable effect on the results, additionally, we evaluated the performance with respect to the number of factors for each method. To this end, we utilized the noise standard deviation 100 under three different noise levels, 0.2, 0.5, and 0.8, respectively. The prediction results for each method are presented in Figure 2 with 100 repetitive trials, with respect to the number of decomposition factors. One could perceive that, not only the proposed PMCR eventually achieved superior regression performance with the optimal number of factors, but also it realized rather commendable performance with a small number of factors. For example, when the noise level was equal to 0.5, the proposed PMCR achieved its optimal performance with no more than 20 factors. By comparison, for the other methods, when the number of factors was larger than 20, their performances remained promoting significantly. One can also observe a similar result in the other two noise levels. This suggests that, PMCR could abstract substantial information with a rather small number of factors from training samples in a noisy regression task.

5.2. ECoG dataset

To further demonstrate the superior robustness of the PMCR algorithm, we evaluated the various PLSR algorithms with the following practical brain decoding task. In this subsection, we used the publicly available Neurotycho ECoG dataset[2] which was initially proposed in Shimoda et al. (2012).

2 Available online at http://neurotycho.org/epidural-ecog-food-tracking-task.

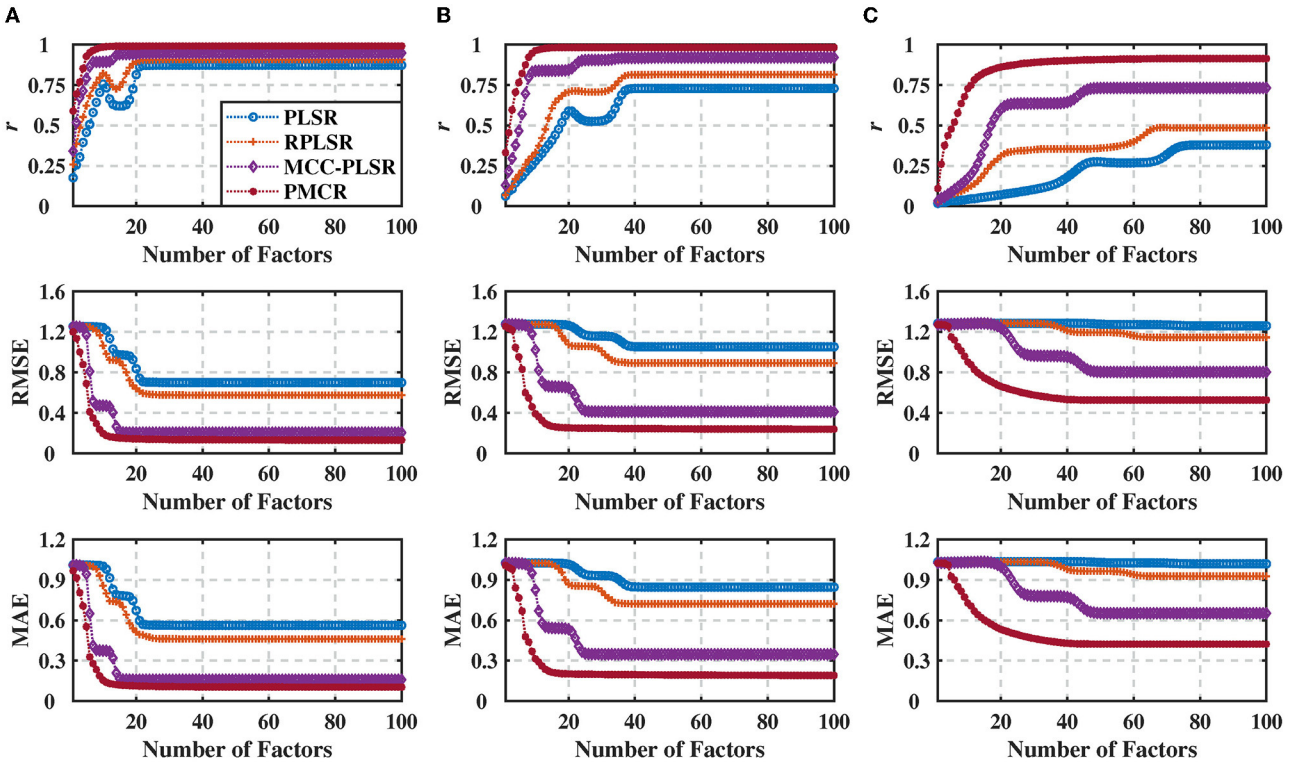


FIGURE 2
Regression performance indicators of the synthetic dataset with noise standard deviation being 100 under three different noise levels with the number of factors increasing from 1 to 100. **(A)** Noise level = 0.2, **(B)** noise level = 0.5, and **(C)** noise level = 0.8. The performance indicators were obtained from 100 repetitive trials and averaged across three dimensions of the output. The proposed PMCR algorithm not only acquired better prediction results than the other algorithms ultimately with the optimal number of factors, but also achieved admirable regression performance with a small number of factors.

5.2.1. Dataset description

Two Japanese macaques, denoted by Monkey B and C, respectively, were commanded to track foods with the right hands, during which the continuous three-dimensional trajectories of right hands with a sampling rate of 120 Hz were recorded by an optical motion capture instrument. For both Monkey B and C, ten recording sessions were performed, where each recording session lasted 15 minutes. The two macaques were in advance implanted with customized 64-channel ECoG electrodes on the contralateral (left) hemisphere, which covered the regions from the prefrontal cortex to the parietal cortex. ECoG signals were recorded simultaneously during each session with a sampling rate of 1,000 Hz. In accordance with Shimoda et al. (2012), for each recording session, the data of the first ten minutes was used to train a prediction model, while the data of the remaining five minutes was used to evaluate the prediction performance of the trained model. The schemes of the experiments and ECoG electrodes are shown in Figures 3A, B, respectively.

5.2.2. Decoding paradigm

For the feature extraction, we used an identical offline decoding paradigm as in Shimoda et al. (2012). Initially, ECoG signals were preprocessed with a tenth-order Butterworth bandpass filter with cutoff frequencies from 1 to 400 Hz, and then re-referenced by the common average referencing (CAR) method. The three-dimensional trajectories of the right wrist were down-sampled to 10 Hz, thus, leading to 9,000 samples in one session (10 Hz × 60 sec × 15 min). The three-dimensional position of time t was predicted from the ECoG signals during the previous one second. To extract the features of ECoG signals, we utilized the time-frequency representation. For the time t, the ECoG signals at each electrode from t - 1.1 s to t were processed by Morlet wavelet transformation. Ten center frequencies ranging from 10 to 120 Hz with equal spacing on the logarithmic scale were considered for the wavelet transformation, overlaying the frequency bands which are most relevant to motion tasks (Shimoda et al., 2012). The time- frequency scalogram was then resampled at ten temporal lags with a 0.1 s gap (t - 1 s, t - 0.9 s,..., t - 0.1 s). Thus, the input of each sample exhibited a 6,400-dimensional vector (64 channels × 10 frequencies × 10 temporal lags), and the output was the three-dimensional position of the right hand. Hence, we trained a regression model with 6,000 samples (the first ten minutes) to predict the three-dimensional output from the 6,400-dimensional input, and evaluated the algorithms with other 3,000 testing samples (the remaining five minutes). The illustrative diagrams for ECoG decoding are summarized in Figure 3C.

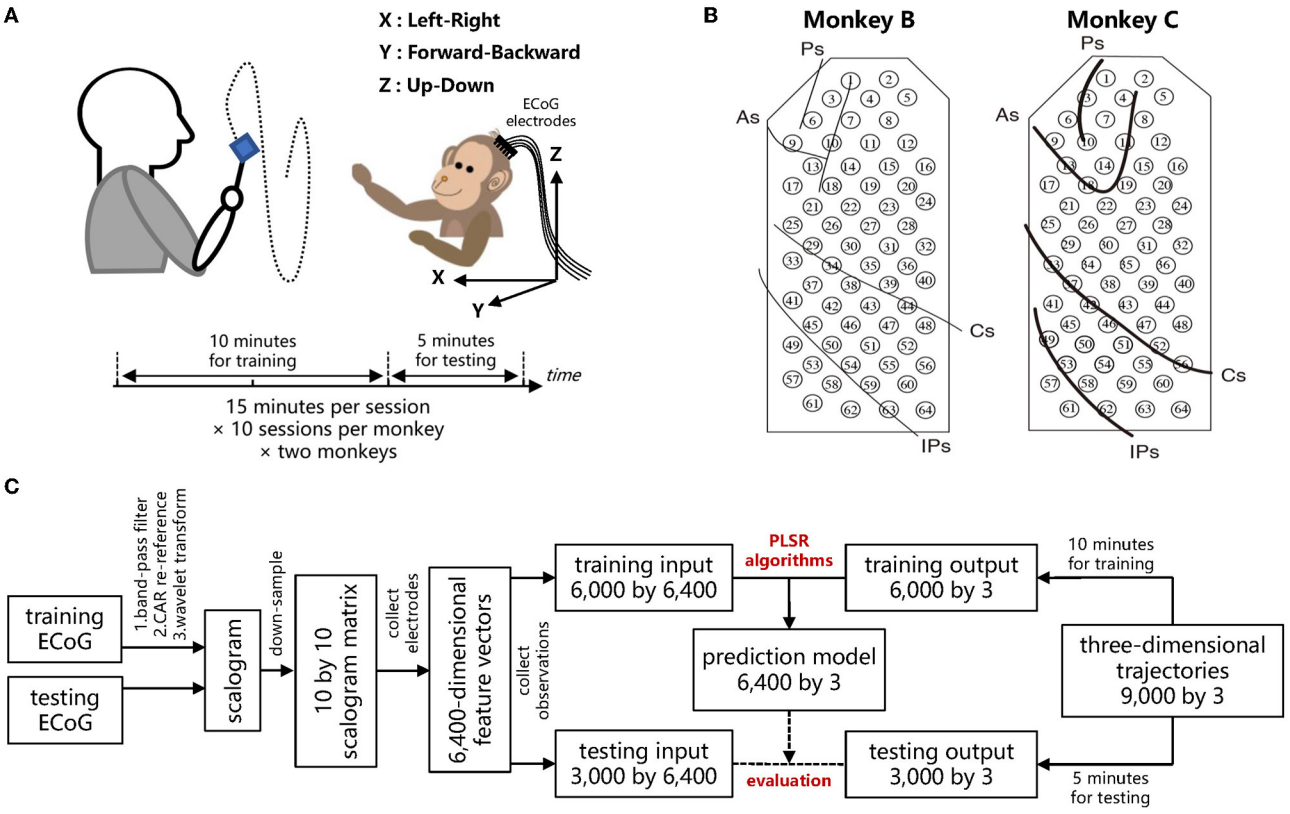


FIGURE 3
Experimental protocol of the Neurotycho ECoG dataset and decoding paradigm to evaluate the robustness of the different PLSR algorithms. **(A)** The macaque retrieved foods in a three-dimensional random location, during which the body-centered coordinates of the right wrists and the ECoG signals were recorded simultaneously. **(B)** Both Monkey B and C were implanted with 64-channel epidural ECoG electrodes on the contralateral (left) hemisphere, overlaying the regions from the prefrontal cortex to the parietal cortex. Ps: principal sulcus, As: arcuate sulcus, Cs: central sulcus, IPs: intraparietal sulcus. **(A, B)** Were reproduced from Shimoda et al. (2012), which provides the details of this public dataset. **(C)** Decoding diagram from ECoG signals to three-dimensional trajectories. The training ECoG signals are contaminated to assess the robustness of different algorithms.

5.2.3. Contamination

To evaluate the robustness of different algorithms in the practical ECoG decoding task, the ECoG signals were artificially contaminated by outlier to simulate the detrimental artifact. To be specific, we stochastically selected three certain proportions, 0 (no contamination), 10^{-3}, and 10^{-2}, of the training ECoG samplings and corrupted them with outliers which were subject to the zero-mean Gaussian distribution with the variance 50 times that of the signals for the corresponding channel. As stated in Ball et al. (2009a), the blink-related artifacts were remarkably found in ECoG signal that exhibited much larger amplitudes than a normal ECoG recording. Hence, we used the above-mentioned approach to artificially generate adverse artifacts, so as to contaminate the ECoG signals. This method has been widely utilized in the literature to deteriorate the brain signals for evaluating the robustness of different algorithms (Wang et al., 2011; Chen et al., 2018).

Note that, for this ECoG dataset, the 'Noise Level' signifies the ratio of the contaminated ECoG samplings in the entirety which is different from the ratio of the deteriorated samples in the 6,000 training samples. The ratio of the affected training samples can be evidently larger than the indicated noise level, since one contaminated ECoG sampling could deteriorate several time windows in feature extraction. For example, when the noise level is denoted as 10^{-3}, the deteriorated proportion of the training set is (0.6645 ± 0.0089). Furthermore, we illustrate how the noise would influence the time-frequency feature in Figure 4. One could obviously perceive the heavy-tailed characteristic on the feature noises, which is in particular intractable for the least square criterion. In addition, the effects of high-frequency band are more prominent, due to the property of impulsive noise.

5.2.4. Spatio-spectro-temporal pattern

Studying how the spatio-spectro-temporal weights in the regression model contribute to the entirety can help investigate the neurophysiological pattern. The element of the trained prediction model $\mathbf{H}$ can be denoted by $h_{\text{ch,freq,temp}}$, which corresponds to the ECoG electrode "ch," the frequency "freq," and the temporal lag "temp." Thus, one could calculate the spatio-spectro-temporal contributions by the ratio between the summation of absolute values of each domain and the summation of absolute values of the entire model

$$W_c(\text{ch}) = \frac{\sum_{\text{freq}} \sum_{\text{temp}} |h_{\text{ch,freq,temp}}|}{\sum_{\text{ch}} \sum_{\text{freq}} \sum_{\text{temp}} |h_{\text{ch,freq,temp}}|} \tag{29}$$

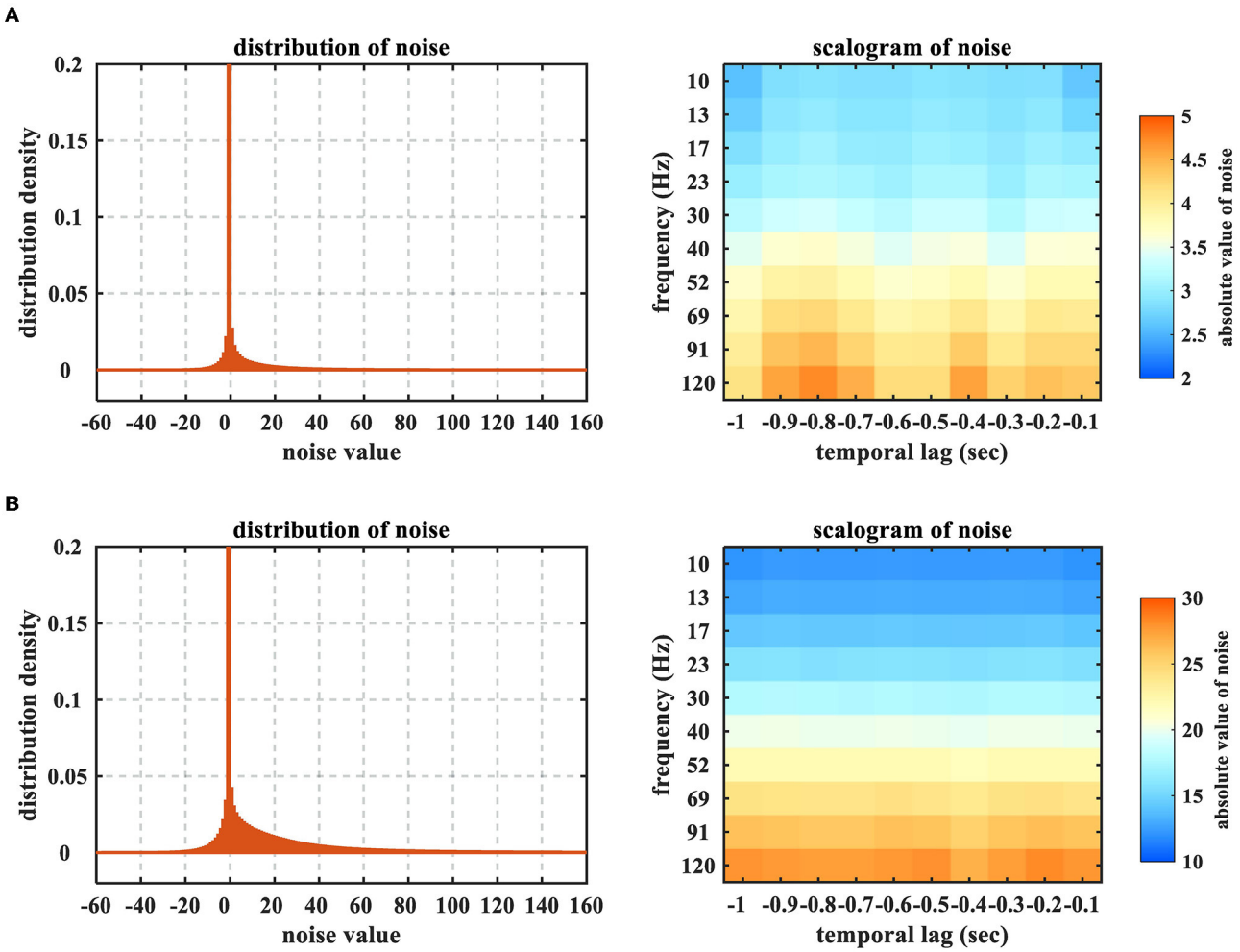


FIGURE 4
Distributions and scalograms of the time-frequency feature noises resulting from the ECoG sampling contamination. **(A)** Noise level = 10^{-3} (the deteriorated proportion of training set = 0.6645 ± 0.0089), **(B)** Noise level = 10^{-2} (the deteriorated proportion of training set ≈ 1). The time-frequency feature noises were calculated by subtracting the training datasets which were obtained from acoustic and contaminated ECoG signals, respectively. The distributions were averaged by 20 sessions of Monkey B and C, while the scalograms were averaged across all electrodes. The peaks of distributions are truncated to emphasize the heavy-tailed characteristic.

$$W_f(\text{freq}) = \frac{\sum_{\text{ch}} \sum_{\text{temp}} |h_{\text{ch,freq,temp}}|}{\sum_{\text{ch}} \sum_{\text{freq}} \sum_{\text{temp}} |h_{\text{ch,freq,temp}}|} \tag{30}$$

$$W_t(\text{temp}) = \frac{\sum_{\text{ch}} \sum_{\text{freq}} |h_{\text{ch,freq,temp}}|}{\sum_{\text{ch}} \sum_{\text{freq}} \sum_{\text{temp}} |h_{\text{ch,freq,temp}}|} \tag{31}$$

where $W_c(\text{ch})$, $W_f(\text{freq})$, and $W_t(\text{temp})$ denote the contributions of the ECoG electrode "ch," the frequency "freq," and the temporal lag "temp," respectively.

5.2.5. Results

First, we assessed the different algorithms with the uncontaminated ECoG signals. Accordingly, when the noise level was zero, the average performance indicators were obtained by the acoustic 20 sessions (Monkey B and C). Then we contaminated each session with 5 repetitive trials. Hence, for every noise level, each algorithm was evaluated for 100 times (20 sessions $\times$ 5 repetitive trials). In Table 2, we present the performance indicators for each algorithm with the noise levels 0, 10^{-3}, and 10^{-2}, respectively. In each row of a specific condition, the optimal result is marked in bold. Moreover, the other results are marked with $(*)$ if there exists statistically significant difference between the current result and the optimal result under each condition. One observes in Table 2 that, the proposed PMCR realized the optimal prediction results consistently, except for the Y-axis under noise level 0. In most cases, PMCR outperformed the other methods with statistically significant difference. One can observe that, when the noise level was 0, PMCR achieved better results than the other algorithms for X-axis and Z-axis. One major reason is, in the acoustic sessions, the motion-related artifacts have been considerably found in the ECoG signals (Shimoda et al., 2012), which further demonstrates the necessity of utilizing PMCR in the practical ECoG decoding tasks.

Furthermore, we studied how the neurophysiological patterns for different algorithms were influenced by the sampling noises.

TABLE 2 Performance indicators of each algorithm on the Neurotycho ECoG dataset under three noise levels 0, 10^{-3}, and 10^{-2}, respectively.

X-position						
Algorithm			PLSR	RPLSR	MCC-PLSR	PMCR
Noise Level	0	r	0.4378 ± 0.0933*	0.4550 ± 0.0925*	0.4598 ± 0.0942*	**0.4679 ± 0.0947**
		RMSE	0.9287 ± 0.0810*	0.9037 ± 0.0653*	0.8954 ± 0.0809*	**0.8835 ± 0.0786**
		MAE	0.7026 ± 0.0640*	0.6872 ± 0.0530*	0.6749 ± 0.0628*	**0.6658 ± 0.0651**
	10^{-3}	r	0.3334 ± 0.1165*	0.3558 ± 0.1132*	0.3684 ± 0.1127*	**0.3873 ± 0.1274**
		RMSE	0.9729 ± 0.0652*	0.9543 ± 0.0648*	0.9397 ± 0.0728*	**0.9276 ± 0.0705**
		MAE	0.7291 ± 0.0756*	0.7174 ± 0.0689*	0.7092 ± 0.0786*	**0.6987 ± 0.0759**
	10^{-2}	r	0.1524 ± 0.1399*	0.1713 ± 0.1353*	0.1926 ± 0.1342*	**0.2238 ± 0.1382**
		RMSE	1.0249 ± 0.1105*	1.0022 ± 0.1097*	0.9845 ± 0.1129*	**0.9681 ± 0.1094**
		MAE	0.7655 ± 0.1428*	0.7485 ± 0.1383*	0.7396 ± 0.1392*	**0.7246 ± 0.1397**
Y-position						
Algorithm			PLSR	RPLSR	MCC-PLSR	PMCR
Noise Level	0	r	0.5426 ± 0.1019*	**0.5582 ± 0.1026**	0.5547 ± 0.1017	0.5549 ± 0.1022
		RMSE	0.8483 ± 0.0969*	**0.8198 ± 0.0951**	0.8246 ± 0.0948	0.8233 ± 0.0952
		MAE	0.6487 ± 0.0762*	**0.6304 ± 0.0796**	0.6362 ± 0.0744	0.6358 ± 0.0759
	10^{-3}	r	0.4114 ± 0.1309 *	0.4284 ± 0.1285*	0.4425 ± 0.1302*	**0.4602 ± 0.1296**
		RMSE	0.9188 ± 0.0963 *	0.8962 ± 0.0958*	0.8795 ± 0.0979*	**0.8608 ± 0.1002**
		MAE	0.6960 ± 0.1007*	0.6849 ± 0.1014*	0.6631 ± 0.0983*	**0.6539 ± 0.1021**
	10^{-2}	r	0.2084 ± 0.1514*	0.2206 ± 0.1489*	0.2593 ± 0.1502*	**0.2723 ± 0.1537**
		RMSE	0.9781 ± 0.1143*	0.9542 ± 0.1117*	0.9306 ± 0.1159	**0.9294 ± 0.1146**
		MAE	0.7354 ± 0.1028*	0.7173 ± 0.1077*	0.7086 ± 0.1105	**0.7043 ± 0.1042**
Z-position						
Algorithm			PLSR	RPLSR	MCC-PLSR	PMCR
Noise Level	0	r	0.6320 ± 0.0324*	0.6395 ± 0.0328*	0.6482 ± 0.0359	**0.6504 ± 0.0372**
		RMSE	0.7968 ± 0.0281*	0.7814 ± 0.0293*	0.7747 ± 0.0296*	**0.7628 ± 0.0275**
		MAE	0.6181 ± 0.0222*	0.6102 ± 0.0280*	0.6055 ± 0.0241	**0.5989 ± 0.0265**
	10^{-3}	r	0.4875 ± 0.0708*	0.4935 ± 0.0701*	0.5158 ± 0.0857*	**0.5259 ± 0.0814**
		RMSE	0.9272 ± 0.0712*	0.9129 ± 0.0682*	0.8958 ± 0.0742*	**0.8834 ± 0.0738**
		MAE	0.6932 ± 0.0800*	0.6894 ± 0.0814*	0.6804 ± 0.0852*	**0.6645 ± 0.0782**
	10^{-2}	r	0.2399 ± 0.1185*	0.2456 ± 0.1173*	0.2615 ± 0.1148*	**0.2803 ± 0.1186**
		RMSE	1.0168 ± 0.0804*	0.9917 ± 0.0785*	0.9605 ± 0.0842*	**0.9485 ± 0.0809**
		MAE	0.7532 ± 0.0883*	0.7429 ± 0.0892*	0.7208 ± 0.0893	**0.7146 ± 0.0887**

The results are given in mean ± deviation, where the optimal results under each condition are marked in bold. The proposed PMCR realized the optimal results consistently, except for the Y-position under the noise level 0. For each result, *is marked if there exists statistically significant difference between the indicated one and the optimal result in the corresponding condition, according to a paired t-test ($p < 0.05$).

We show the differences between the spatial, the spectral, and the temporal contributions which were acquired from the acoustic and the contaminated sessions (under the noise level 10^{-3}), respectively, in Figure 5. The regression model concerning Monkey B's Z-position was used here. We also quantified the effects by computing the summation of the absolute values of the difference between the patterns that were attained from the acoustic and the contaminated sessions, respectively. To be specific, we illustrate $\sum |W_c(\text{ch}) - W'_c(\text{ch})|$, $\sum |W_f(\text{freq}) - W'_f(\text{freq})|$, and $\sum |W_t(\text{temp}) - W'_t(\text{temp})|$ for the spatial, the spectral, and the temporal patterns, respectively. $W_c(\text{ch})$, $W_f(\text{freq})$, and $W_t(\text{temp})$ were obtained by the acoustic sessions, while $W'_c(\text{ch})$, $W'_f(\text{freq})$, and $W'_t(\text{temp})$ were obtained from the contaminated sessions. One can observe from Figure 5 that, the proposed PMCR algorithm realized the minimal deterioration for the pattern of each domain. This further demonstrates the robustness of PMCR in noisy ECoG decoding tasks.

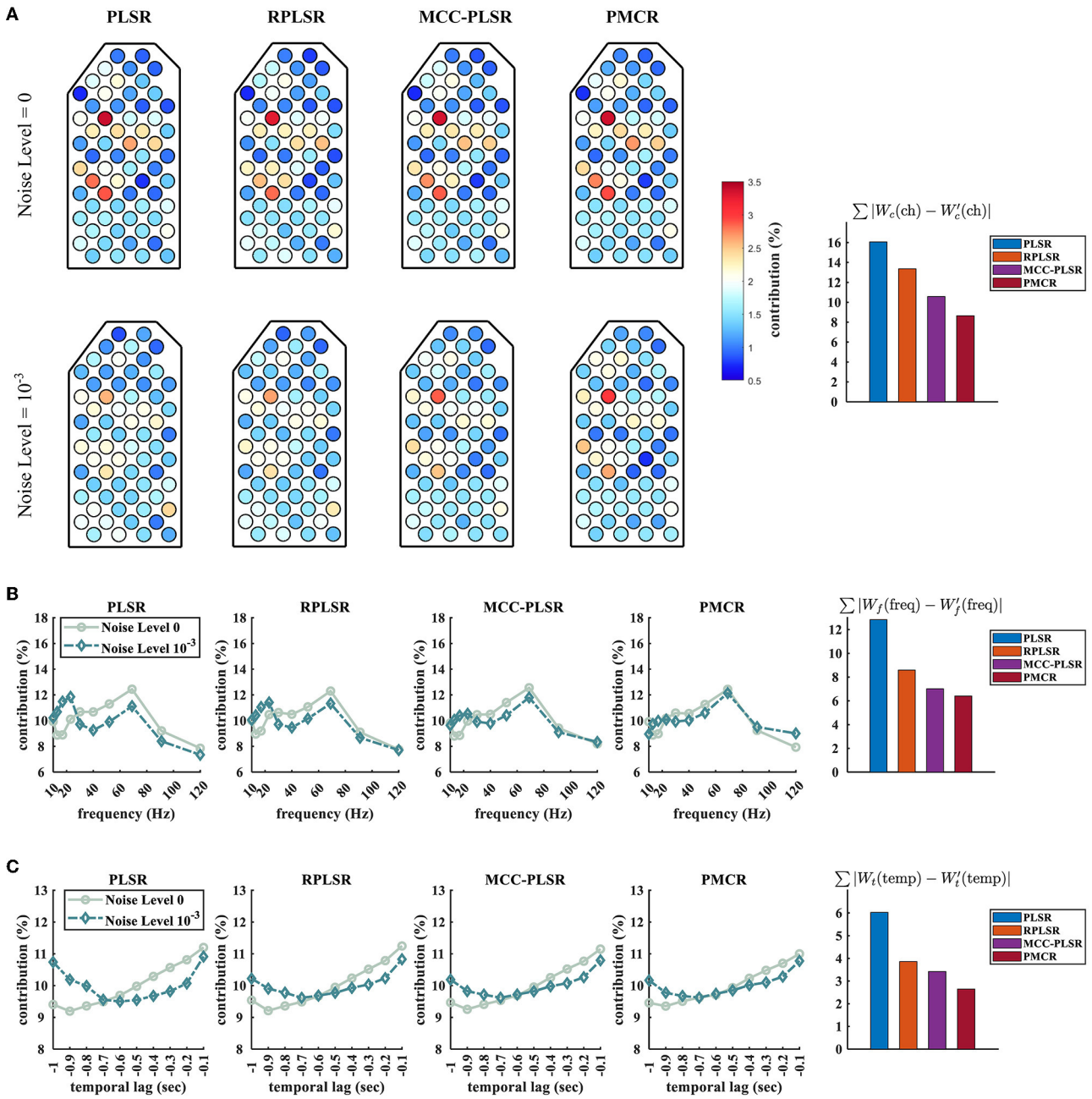


FIGURE 5
Spatio-spectro-temporal contributions of the prediction model for Monkey B's Z-position under noise levels 0 and 10^{-3}. **(A)** Spatial patterns, **(B)** spectral patterns, and **(C)** temporal patterns. For each domain, the quantitative deterioration is calculated by the absolute value summation of the difference between the original and the deteriorated patterns. The original patterns W_c(ch), W_f(freq), and W_t(temp) were averaged across the 10 acoustic sessions of Monkey B, while the deteriorated patterns W'_c(ch), W'_f(freq), and W'_t(temp) were averaged across 50 trials (10 sessions of Monkey B × 5 repetitive trials). The proposed PMCR achieved the minimal deterioration for each domain.

6. Discussion

6.1. Proposed method

In the present study, we aimed to propose a new robust version for PLSR using the MCC framework, which is named as PMCR. Similarly as the existing PLSR methods, the proposed PMCR decomposes the explanatory matrix (input) and the response matrix (output) iteratively for S decomposition factors. The crucial differences of the proposed PMCR are stated in what follows. First, the objective function regarding the projectors $\mathbf{w}_s$ and $\mathbf{c}_s$ in Eq. (15) could be considered as a generalized formulation of the conventional PLSR (Eq. 6), and would be closely related to the calculation in MCC-PLSR (Eq. 12) under specific conditions. As has been proved in Liu et al. (2007), maximizing the correntropy between two variables, when the kernel bandwidth tends to infinity,

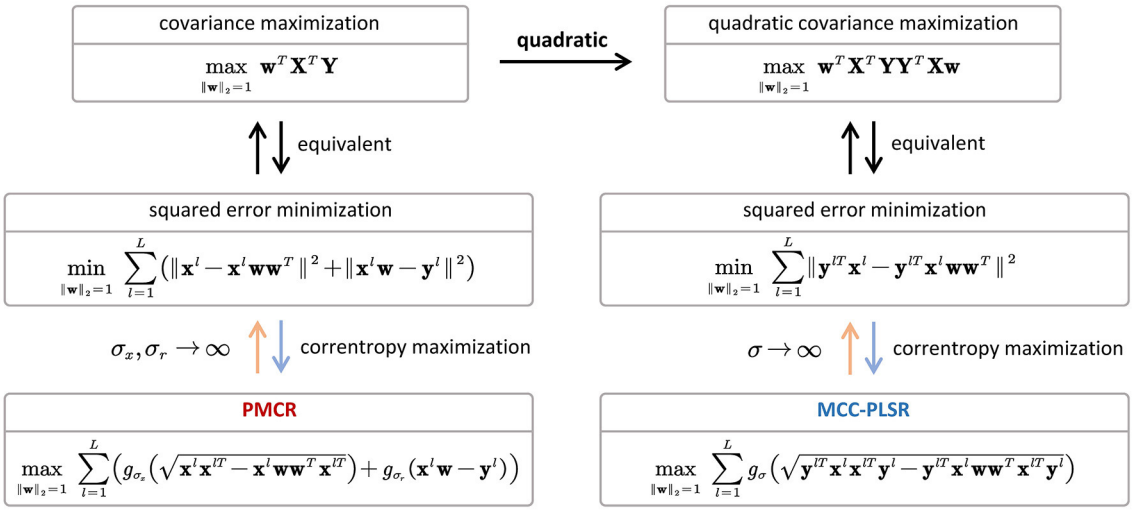


FIGURE 6
The connection between PMCR and MCC-PLSR for a univariate response.

is equal to minimizing their quadratic Euclidean distance. Hence, if we suppose $\sigma_x, \sigma_y, \sigma_r \to \infty$, the projector calculation of PMCR will degenerate to the conventional PLSR. Then, we consider the differences between MCC-PLSR and the proposed PMCR. For a univariate response, the projector $\mathbf{c}$ for dimensionality reduction regarding the response could be ignored. Thus, we can rewrite the dimensionality reduction in PMCR (Eq. 15) as

$$\max_{\|\mathbf{w}\|_2=1} \sum_{l=1}^{L} \left(g_{\sigma_x}(\sqrt{\mathbf{x}^l\mathbf{x}^{lT}-\mathbf{x}^l\mathbf{w}\mathbf{w}^T\mathbf{x}^{lT}})+g_{\sigma_r}(\mathbf{x}^l\mathbf{w}-\mathbf{y}^l)\right) \quad (32)$$

which could be regarded as a generalized form for the quadratic error minimization (Liu et al., 2007).

$$\min_{\|\mathbf{w}\|_2=1} \sum_{l=1}^{L} \left(\|\mathbf{x}^l-\mathbf{x}^l\mathbf{w}\mathbf{w}^T\|^2+\|\mathbf{x}^l\mathbf{w}-\mathbf{y}^l\|^2\right) \Leftrightarrow \max_{\|\mathbf{w}\|_2=1} \mathbf{w}^T\mathbf{X}^T\mathbf{Y} \quad (33)$$

which is essentially equal to the conventional PLSR for univariate output. By comparison, MCC-PLSR adopts the MCC framework for the quadratic covariance (Eq. 11), which can be written as Mou et al. (2018)

$$\min_{\|\mathbf{w}\|_2=1} \sum_{l=1}^{L} \|\mathbf{y}^{lT}\mathbf{x}^l-\mathbf{y}^{lT}\mathbf{x}^l\mathbf{w}\mathbf{w}^T\|^2 \Leftrightarrow \max_{\|\mathbf{w}\|_2=1} \mathbf{w}^T\mathbf{X}^T\mathbf{Y}\mathbf{Y}^T\mathbf{X}\mathbf{w} \quad (34)$$

which is the special case of MCC-PLSR when the kernel bandwidth in Eq. (12) tends to infinity. Thus, the connection between PMCR (Eq. 15) and MCC-PLSR (Eq. 12) could be illustrated as in Figure 6. One can observe that, the starting points of PMCR and MCC-PLSR are different. The proposed PMCR begins from the original covariance maximization, whereas MCC-PLSR was proposed from the quadratic covariance. Therefore, we argue that our proposed PMCR is a more rational robust implementation for PLSR. Moreover, note that we give the above discussion under the premise of a univariate output, which is only a special case of degradation for our proposed PMCR. One the other hand, considering the calculations of the loading vector $\mathbf{p}_s$ and the regression coefficient b_s, the proposed PMCR employs the MCC (Eq. 16, 17), whereas the conventional PLSR and MCC-PLSR utilize the least square criterion. As mentioned above, Eqs. (16, 17) can be also regarded as generalized forms of square error minimization. In summary, the proposed PMCR is more generalized than the conventional PLSR and MCC-PLSR.

In addition, we would like to discuss the advantages and disadvantages of the proposed PMCR algorithm. The essential benefit of utilizing the PMCR algorithm in a noisy ECoG decoding task is the conspicuous robustness with respect to the noises, which was demonstrated with extensive experiments in Section 5. Further, mathematically, the proposed PMCR algorithm is more generalized than the conventional PLSR and MCC-PLSR. As was mentioned above, the conventional PLSR and MCC-PLSR could be regarded as special cases of the proposed PMCR under specific conditions concerning the kernel bandwidths. In particular, compared with MCC-PLSR, the proposed PMCR takes into account the dimensionality reduction for the response matrix. As a result, PMCR could realize better prediction performance for multivariate response. Moreover, PMCR could be further implemented with regularization techniques and extended to the multi-way scenario, which would be discussed in the following subsections. However, PMCR might suffer the performance degradation resulting from inadequate kernel bandwidths that are calculated by the *Silverman's rule* (Eq. 25). Although the experimental results in this study verified empirically that, the proposed PMCR could perform efficiently with the kernel bandwidths acquired by Eq. (25), it may be difficult to guarantee that the *Silverman's rule* can always provide adequate kernel bandwidths. Hence, we would like to investigate a better way to determine the kernel bandwidths with solid theoretical guarantees in our future works. In addition, our proposed PMCR is effective to deal with outliers, while it may be inadequate for multi-modal-distributed noise because MCC utilizes only one kernel function for each reconstruction error. To address this issue, it is promising to use *minimum error entropy*

(MEE) to reformulate PLSR, another popular learning criterion in ITL (Principe, 2010). MEE employs multiple kernel functions for each reconstruction error, so that it can realize satisfactory robustness with respect to multi-modal-distributed noise, which has realized robust neural decoding algorithms (Chen et al., 2018; Li et al., 2021).

6.2. PMCR with regularization

One should additionally note that, the PMCR was proposed by reformulating the conventional PLSR algorithm with using the robust MCC, instead of the mediocre least square criterion. Hence, the proposed PMCR exhibits the supplementary potential for further performance improvements with regularization techniques, as well as in the existing regularized PLSR methods. For example, L_1-regularization could be utilized in Eq. (15) to encourage sparse and robust projectors. In addition, if one requires better smoothness on the predicted output, polynomial or Sobolev-norm penalization could be utilized in PMCR. Moreover, L_2-regularization could be utilized for Eq. (17) to decrease the over-fitting risk considering the regression scalar b_s, similarly as Eq. (7) (Foodeh et al., 2020). In the literature, MCC-based algorithms with regularization have been widely investigated. For instance, a robust version of sparse representation classifier (SRC) for face recognition was developed by employing L_1-regularization on the MCC-based SRC objective function (He et al., 2010).

6.3. Extension to multi-way application

The multi-way PLSR establishes the regression relationship between tensor variables with dimensionality reduction by tensor factorization technique. In the literature, the multi-way PLSR was usually reported to achieve better decoding capability than the generic PLSR algorithm in the brain decoding task, where the spatio-spectro-temporal feature is organized with the tensor form. Essentially, the multi-way PLSR decomposes the input and output under the least square criterion by minimizing the Frobenius-norm (Kolda and Bader, 2009). Therefore, the multi-way PLSR is prone to the performance deterioration caused by noises as well.

The proposed PMCR method treats the regression problem of matrix, i.e. two-way variable. Extending the PMCR algorithm to multi-way application could probably improve the prediction performance further, which would be investigated in our future works. Promisingly, MCC has been demonstrated effective for tensor variable analysis in a recent study (Zhang et al., 2016).

7. Conclusion

This paper proposed a new robust variant for the PLSR algorithm by reformulating the non-robust least square criterion with the sophisticated MCC framework. The proposed robust objective functions can be effectively optimized by half-quadratic and fixed-point optimization methods. Extensive experimental results with the synthetic dataset and Neurotycho epidural ECoG dataset demonstrate that, the proposed PMCR can outperform the existing PLSR algorithms, revealing promising robustness for high-dimensional and noisy ECoG decoding tasks.

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

YL and BC contributed to conceptualization of the study and developed the algorithm derivation. YL and GW operated the experiments and analyzed the results. YL wrote the first draft of the manuscript. NY and YK directed the study and guided the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant 19H05728, in part by the Japan Science and Technology Agency (JST) Support for the Pioneering Research Initiated by Next Generation (SPRING) under Grant JPMJSP2106, and in part by the National Natural Science Foundation of China under Grants U21A20485 and 61976175.

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

References

Amiri, S., Fazel-Rezai, R., and Asadpour, V. (2013). "A review of hybrid brain-computer interface systems," in *Advances in Human-Computer Interaction 2013*. doi: 10.1155/2013/187024

Ball, T., Kern, M., Mutschler, I., Aertsen, A., and Schulze-Bonhage, A. (2009a). Signal quality of simultaneously recorded invasive and non-invasive EEG. *Neuroimage* 46, 708–716. doi: 10.1016/j.neuroimage.2009.02.028

Ball, T., Schulze-Bonhage, A., Aertsen, A., and Mehring, C. (2009b). Differential representation of arm movement direction in relation to cortical anatomy and function. *J. Neur. Eng.* 6, 016006. doi: 10.1088/1741-2560/6/1/016006

Barker, M., and Rayens, W. (2003). Partial least squares for discrimination. *J. Chemometr.* 17, 166–173. doi: 10.1002/cem.785

Bro, R. (1996). Multiway calibration: multilinear PLS. *J. Chemometr.* 10, 47–61. doi: 10.1002/(SICI)1099-128X(199601)10:147::AID-CEM4003.0.CO;2-C

Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents–EEG, ECOG, LFP, and spikes. *Nat. Rev. Neurosci.* 13, 407–420. doi: 10.1038/nrn3241

Chao, Z. C., Nagasaka, Y., and Fujii, N. (2010). Long-term asynchronous decoding of arm motion using electrocorticographic signals in monkey. *Front. Neuroeng.* 3, 3. doi: 10.3389/fneng.2010.00003

Chen, B., Li, Y., Dong, J., Lu, N., and Qin, J. (2018). Common spatial patterns based on the quantized minimum error entropy criterion. *IEEE Trans. Syst. Man Cybern.* 50, 4557–4568. doi: 10.1109/TSMC.2018.2855106

Chen, B., and Príncipe, J. C. (2012). Maximum correntropy estimation is a smoothed map estimation. *IEEE Signal Proc. Lett.* 19, 491–494. doi: 10.1109/LSP.2012.2204435

Chen, B., Wang, J., Zhao, H., Zheng, N., and Principe, J. C. (2015). Convergence of a fixed-point algorithm under maximum correntropy criterion. *IEEE Signal Proces. Lett.* 22, 1723–1727. doi: 10.1109/LSP.2015.2428713

Chestek, C. A., Batista, A. P., Santhanam, G., Byron, M. Y., Afshar, A., Cunningham, J. P., et al. (2007). Single-neuron stability during repeated reaching in macaque premotor cortex. *J. Neurosci.* 27, 10742–10750. doi: 10.1523/JNEUROSCI.0959-07.2007

Chin, C. M., Popovic, M. R., Thrasher, A., Cameron, T., Lozano, A., and Chen, R. (2007). Identification of arm movements using correlation of electrocorticographic spectral components and kinematic recordings. *J. Neur. Eng.* 4, 146. doi: 10.1088/1741-2560/4/2/014

Dong, J., Chen, B., Lu, N., Wang, H., and Zheng, N. (2017). "Correntropy induced metric based common spatial patterns," in *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)* (IEEE) 1–6. doi: 10.1109/MLSP.2017.8168132

Donoghue, J. P. (2002). Connecting cortex to machines: recent advances in brain interfaces. *Nat. Neurosci.* 5, 1085–1088. doi: 10.1038/nn947

Eliseyev, A., and Aksenova, T. (2016). Penalized multi-way partial least squares for smooth trajectory decoding from electrocorticographic (ecog) recording. *PLoS ONE* 11, e0154878. doi: 10.1371/journal.pone.0154878

Eliseyev, A., Auboiroux, V., Costecalde, T., Langar, L., Charvet, G., Mestais, C., et al. (2017). Recursive exponentially weighted n-way partial least squares regression with recursive-validation of hyper-parameters in brain-computer interface applications. *Sci. Rep.* 7, 1–15. doi: 10.1038/s41598-017-16579-9

Eliseyev, A., Moro, C., Costecalde, T., Torres, N., Gharbi, S., Mestais, C., et al. (2011). Iterative n-way partial least squares for a binary self-paced brain-computer interface in freely moving animals. *J. Neur. Eng.* 8, 046012. doi: 10.1088/1741-2560/8/4/046012

Eliseyev, A., Moro, C., Faber, J., Wyss, A., Torres, N., Mestais, C., et al. (2012). L1-penalized n-way pls for subset of electrodes selection in bci experiments. *J. Neur. Eng.* 9, 045010. doi: 10.1088/1741-2560/9/4/045010

Feng, Y., Huang, X., Shi, L., Yang, Y., and Suykens, J. A. (2015). Learning with the maximum correntropy criterion induced losses for regression. *J. Mach. Learn. Res.*, 16, 993–1034.

Fletcher, R. (2013). *Practical Methods of Optimization*. New York: John Wiley Sons.

Foodeh, R., Ebadollahi, S., and Daliri, M. R. (2020). Regularized partial least square regression for continuous decoding in brain-computer interfaces. *Neuroinformatics* 18, 465–477. doi: 10.1007/s12021-020-09455-x

He, R., Hu, B.-G., Zheng, W.-S., and Kong, X.-W. (2011). Robust principal component analysis based on maximum correntropy criterion. *IEEE Trans. Image Proc.* 20, 1485–1494. doi: 10.1109/TIP.2010.2103949

He, R., Zheng, W.-S., and Hu, B.-G. (2010). Maximum correntropy criterion for robust face recognition. *IEEE Trans. Patt. Analy. Mach. Intell.* 33, 1561–1576. doi: 10.1109/TPAMI.2010.220

Huber, P. J. (2004). *Robust Statistics*, vol. 523. Oxford: John Wiley Sons.

Kolda, T. G., and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* 51, 455–500. doi: 10.1137/07070111X

Krishnan, A., Williams, L. J., McIntosh, A. R., and Abdi, H. (2011). Partial least squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage* 56, 455–475. doi: 10.1016/j.neuroimage.2010.07.034

Lebedev, M. A., and Nicolelis, M. A. (2006). Brain-machine interfaces: past, present and future. *TRENDS Neurosci.* 29, 536–546. doi: 10.1016/j.tins.2006.07.004

Leuthardt, E. C., Miller, K. J., Schalk, G., Rao, R. P., and Ojemann, J. G. (2006). Electrocorticography-based brain computer interface-the seattle experience. *IEEE Trans. Neur. Syst. Rehabilit. Eng.* 14, 194–198. doi: 10.1109/TNSRE.2006.875536

Leuthardt, E. C., Schalk, G., Wolpaw, J. R., Ojemann, J. G., and Moran, D. W. (2004). A brain-computer interface using electrocorticographic signals in humans. *J. Neur. Eng.* 1, 63. doi: 10.1088/1741-2560/1/2/001

Levine, S. P., Huggins, J. E., BeMent, S. L., Kushwaha, R. K., Schuh, L. A., Rohde, M. M., et al. (2000). A direct brain interface based on event-related potentials. *IEEE Trans. Rehabil. Eng.* 8, 180–185. doi: 10.1109/86.847809

Li, Y., Chen, B., Yoshimura, N., and Koike, Y. (2021). Restricted minimum error entropy criterion for robust classification. *IEEE Trans. Neur. Netw. Learn. Syst.* 33, 6599–6612. doi: 10.1109/TNNLS.2021.3082571

Liu, W., Pokharel, P. P., and Principe, J. C. (2007). Correntropy: Properties and applications in non-gaussian signal processing. *IEEE Trans. Signal Process.* 55, 5286–5298. doi: 10.1109/TSP.2007.896065

Loh, P.-L., and Wainwright, M. J. (2015). Regularized m-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *J. Mach. Learn. Res.* 16, 559–616.

Mou, Y., Zhou, L., Chen, W., Fan, J., and Zhao, X. (2018). Maximum correntropy criterion partial least squares. *Optik* 165, 137–147. doi: 10.1016/j.ijleo.2017.12.126

Mussa-Ivaldi, F. A., and Miller, L. E. (2003). Brain-machine interfaces: computational demands and clinical needs meet basic neuroscience. *TRENDS Neurosci.* 26, 329–334. doi: 10.1016/S0166-2236(03)00121-8

Otsubo, H., Ochi, A., Imai, K., Akiyama, T., Fujimoto, A., Go, C., et al. (2008). High-frequency oscillations of ictal muscle activity and epileptogenic discharges on intracranial eeg in a temporal lobe epilepsy patient. *Clin. Neurophysiol.* 119, 862–868. doi: 10.1016/j.clinph.2007.12.014

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Mathem. Stat.* 33, 1065–1076. doi: 10.1214/aoms/1177704472

Pistohl, T., Ball, T., Schulze-Bonhage, A., Aertsen, A., and Mehring, C. (2008). Prediction of arm movement trajectories from ecog-recordings in humans. *J. Neurosci. Methods* 167, 105–114. doi: 10.1016/j.jneumeth.2007.10.001

Principe, J. C. (2010). *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*. New York: Springer Science Business Media. doi: 10.1007/978-1-4419-1570-2

Ren, Z., and Yang, L. (2018). Correntropy-based robust extreme learning machine for classification. *Neurocomputing* 313, 74–84. doi: 10.1016/j.neucom.2018.05.100

Santamaría, I., Pokharel, P. P., and Principe, J. C. (2006). Generalized correlation function: definition, properties, and application to blind equalization. *IEEE Trans. Signal Proces.* 54, 2187–2197. doi: 10.1109/TSP.2006.872524

Schalk, G., Miller, K. J., Anderson, N. R., Wilson, J. A., Smyth, M. D., Ojemann, J. G., et al. (2008). Two-dimensional movement control using electrocorticographic signals in humans. *J. Neur. Eng.* 5, 75. doi: 10.1088/1741-2560/5/1/008

Shimoda, K., Nagasaka, Y., Chao, Z. C., and Fujii, N. (2012). Decoding continuous three-dimensional hand trajectories from epidural electrocorticographic signals in japanese macaques. *J. Neur. Eng.* 9, 036015. doi: 10.1088/1741-2560/9/3/036015

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis, volume 26*. New York: CRC press.

Singh, A., Pokharel, R., and Principe, J. (2014). The c-loss function for pattern classification. *Patt. Recogn.* 47, 441–453. doi: 10.1016/j.patcog.2013.07.017

Wang, H., Tang, Q., and Zheng, W. (2011). L1-norm-based common spatial patterns. *IEEE Trans. Biomed. Eng.* 59, 653–662. doi: 10.1109/TBME.2011.2177523

Wold, H. (1966). "Estimation of principal components and related models by iterative least squares," in *Multivariate analysis*, eds. P. R. Krishnajah (NewYork: Academic Press) 391–420.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3

Zhang, M., Gao, Y., Sun, C., La Salle, J., and Liang, J. (2016). "Robust tensor factorization using maximum correntropy criterion," in *2016 23rd International Conference on Pattern Recognition (ICPR)* (IEEE) 4184–4189.

Zhao, Q., Caiafa, C. F., Mandic, D. P., Chao, Z. C., Nagasaka, Y., Fujii, N., et al. (2012). Higher order partial least squares (hopls): A generalized multilinear regression method. *IEEE Trans. Patt. Analy. Mach. Intell.* 35, 1660–1673. doi: 10.1109/TPAMI.2012.254

Zhao, Q., Zhang, L., and Cichocki, A. (2014). Multilinear and nonlinear generalizations of partial least squares: an overview of recent advances. *Wiley Interdisc. Rev.* 4, 104–115. doi: 10.1002/widm.1120

Zhao, Q., Zhou, G., Adalı, T., Zhang, L., and Cichocki, A. (2013). "Kernel-based tensor partial least squares for reconstruction of limb movements," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE) 3577–3581. doi: 10.1109/ICASSP.2013.6638324

Zhu, X., and Wu, X. (2004). Class noise vs. attribute noise: A quantitative study. *Artif. Intell. Rev.* 22, 177–210. doi: 10.1007/s10462-004-0751-8

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 19 July 2023
DOI 10.3389/fnins.2023.1176344

OPEN ACCESS

EDITED BY
Minpeng Xu,
Tianjin University, China

REVIEWED BY
Huiguang He,
Chinese Academy of Sciences (CAS), China
Feng Wan,
University of Macau, China

*CORRESPONDENCE
Lirong Zheng
✉ zhenglirong@gdiist.cn
Yuxiang Huan
✉ yxhuan@gdiist.cn
Yijun Wang
✉ wangyj@semi.ac.cn

RECEIVED 28 February 2023
ACCEPTED 26 June 2023
PUBLISHED 19 July 2023

CITATION
Huang Y, Huan Y, Zou Z, Pei W, Gao X, Wang Y and Zheng L (2023) A wearable group-synchronized EEG system for multi-subject brain–computer interfaces. *Front. Neurosci.* 17:1176344. doi: 10.3389/fnins.2023.1176344

A wearable group-synchronized EEG system for multi-subject brain–computer interfaces

Yong Huang[1,2], Yuxiang Huan[2]*, Zhuo Zou[3], Weihua Pei[4], Xiaorong Gao[5], Yijun Wang[4]* and Lirong Zheng[1,2]*

[1]School of Biomedical Engineering, Southern Medical University, Guangzhou, China, [2]Brain-Inspired Computing Laboratory, Guangdong Institute of Intelligence Science and Technology, Hengqin, China, [3]School of Information Science and Technology, Fudan University, Shanghai, China, [4]Institute of Semiconductors, Chinese Academy of Sciences (CAS), Beijing, China, [5]Department of Biomedical Engineering, Tsinghua University, Beijing, China

Objective: The multi-subject brain–computer interface (mBCI) is becoming a key tool for the analysis of group behaviors. It is necessary to adopt a neural recording system for collaborative brain signal acquisition, which is usually in the form of a fixed wire.

Approach: In this study, we designed a wireless group-synchronized neural recording system that supports real-time mBCI and event-related potential (ERP) analysis. This system uses a wireless synchronizer to broadcast events to multiple wearable EEG amplifiers. The simultaneously received broadcast signals are marked in data packets to achieve real-time event correlation analysis of multiple targets in a group.

Main results: To evaluate the performance of the proposed real-time group-synchronized neural recording system, we conducted collaborative signal sampling on 10 wireless mBCI devices. The average signal correlation reached 99.8%, the amplitude of average noise was 0.87 μV, and the average common mode rejection ratio (CMRR) reached 109.02 dB. The minimum synchronization error is 237 μs. We also tested the system in real-time processing of the steady-state visual-evoked potential (SSVEP) ranging from 8 to 15.8 Hz. Under 40 target stimulators, with 2 s data length, the average information transfer rate (ITR) reached 150 ± 20 bits/min, and the highest reached 260 bits/min, which was comparable to the marketing leading EEG system (the average: 150 ± 15 bits/min; the highest: 280 bits/min). The accuracy of target recognition in 2 s was 98%, similar to that of the Synamps2 (99%), but a higher signal-to-noise ratio (SNR) of 5.08 dB was achieved. We designed a group EEG cognitive experiment; to verify, this system can be used in noisy settings.

Significance: The evaluation results revealed that the proposed real-time group-synchronized neural recording system is a high-performance tool for real-time mBCI research. It is an enabler for a wide range of future applications in collaborative intelligence, cognitive neurology, and rehabilitation.

KEYWORDS

group-synchronized, collaborative intelligence, multi-subject brain–computer interface, hyperscanning, wearable, real-time group-synchronized, real-time

1. Introduction

In recent years, mBCI and collaborative intelligence have gained great attention in the field of brain science (Czeszumski et al., 2020; Gao et al., 2021). Moreover, implementing a collaborative acquisition system, as a key tool for collaborative intelligence, is a fundamental problem in generalized BCI (Babiloni and Astolfi, 2014; Perdikis et al., 2020; Zhang et al., 2020; Bhattacharyya et al., 2021). Collaborative adaptive learning of AI requires human–human and human–machine collaboration in a synchronous manner with real-time access to relevant event information (Shenoy et al., 2014). Therefore, humans and machines can cooperate in an adaptive (Müller et al., 2017) and dynamic and effective way (van den Bosch et al., 2019). There are many examples of how the multi-brain works, for example, studies on predicting marital relationships through neural synchronization in multiple brains (Li et al., 2022) and effective interaction between teachers and students (Maksimenko et al., 2018). These investigations assist researchers in comprehending social cognition (Konvalinka and Roepstorff, 2012) and exploring the concept of the "Social Brain" (Minagawa et al., 2018). In addition, some collaborative approaches have demonstrated an mBCI that fused event-related potential (ERP) data for collective decision-making (Wang and Jung, 2011).

To support the development of collaborative mBCI applications, there is an increasing demand for an integrated system that encompasses multiple-target signal acquisition, synchronized triggering, and user-friendly configuration. When designing this system, we should consider the following: First, in non-laboratory mBCI experiments, external noise can interfere with the acquisition of system data, leading to a reduction in the signal-to-noise ratio (SNR) of EEG data and a decrease in data reliability. The desired mBCI system should provide high-quality EEG signals to guarantee its reliability and robustness when facing diverse ambient noise. Second, the traditional fixed-linked EEG devices are connected by cables to form a multi-subject acquisition system (Barraza et al., 2019). This fixed-linked system imposes limitations on subjects' range of motion and activities, consequently affecting their overall user experience and restricting the wider application of mBCI. Therefore, wireless-connected mBCI systems will be more favorable and offer significant advantages in future applications. Third, event-triggered synchronous signal acquisition is crucial for mBCI systems, as EEG signals gathered from different BCI devices require strict synchronization for further correlation analysis. For example, hybrid EEG and EMG synchronous acquisition (Artoni et al., 2018) and event-related potential (ERP) mechanisms with up to 100 classifications (Xu et al., 2020) have been implemented. This research highly requires low latency, synchronized phase (Xu et al., 2018), and time alignment of the EEG data. Time-space synchronization necessitates a few milliseconds or even <1 ms (Luck, 2014). Moreover, asynchronism may lead to incorrect estimates, such as time-domain correlation (Bowyer, 2016) and imaginary part correlation (Ayrolles et al., 2020). Traditional synchronization methods employ a wired trigger box as an event source for distributing synchronization signals, and some methods employ multi-device timestamps. Typically, there are two common synchronization methods: wired hardware synchronization and software synchronization. Hardware synchronization involves inputting a synchronization signal into a digital port (David Hairston et al., 2014). Pulse signals are typically generated by serial/parallel ports or sensors. For instance, audio signals are employed for synchronization (Pérez et al., 2021), and clock signals are sent to two wired acquisition devices (Chuang et al., 2021). On the other hand, software synchronization relies on programs that do offline calibration by aligning the data from multiple sources with timestamps in some protocols, such as the LSL (lab streaming layer) framework (Reis et al., 2014) or video frame synchronization (Raghavan et al., 2018).

In this study, we developed a wearable real-time group-synchronized EEG acquisition system to overcome the aforementioned challenges of an mBCI system. The proposed system integrates a light-based event trigger, wireless EEG acquisition devices, an analysis system, and an ERP stimuli system. Up to 10 wireless EEG acquisition devices can be group-synchronized by the event trigger, freeing the limitations of subjects' range of motion and activities. In addition, we optimized the wireless communication channels and the data packet protocol. The EEG acquisition subsystem achieved an average noise amplitude of 0.87 μV, a CMRR of 109.02 dB, a higher SNR, and a comparable ITR with the Synamps2 EEG system from the market-leading company Neuroscan. Finally, the effectiveness of the proposed mBCI system was verified in a multi-subject cognitive experiment, demonstrating its potential in research on social interaction and decision-making in cognitive neurology.

2. Materials

2.1. System architecture

The architecture of the proposed mBCI system is illustrated in Figure 1. The mBCI hardware system features a wireless trigger for data synchronization, a Wi-Fi router for wireless network connectivity, 10 wearable compact wireless amplifiers (A1–A10) at a 1 kHz sampling rate and a host computer for recording and analysis purposes. The system utilizes the light as the trigger source to send events simultaneously to all the wireless EEG amplifiers. EEG Ag/AgCl electrodes (PO6, PO4, POz, PO3, PO5, O1, Oz, and O2), along with one reference and one ground, are linked to the forehead area, as shown in Figure 1. Furthermore, we designed a group EEG cognitive experiment to verify that this system is effective and can be used in noisy settings.

The stimulus display shows real-time calculation results, stimulus signals, and optical synchronization signals, which are generated by a steady stimulus-producing host computer. A light sensor trigger is employed instead of alternative sensors, such as audio, to reduce latency and efficiency. The synchronizer processes the trigger signal and uses a dedicated 2.4G channel to transmit the output to the amplifiers, which differs from the channel between the recorder analysis server and the client A_x. The interference may occur due to adjacent channel interference from the neighbor's wireless local area network channels in the 802.11 band devices. To ensure minimal interference between different channel signals, we processed the individual channels. The protocol of each device enables simultaneous marking of the received trigger signals (to

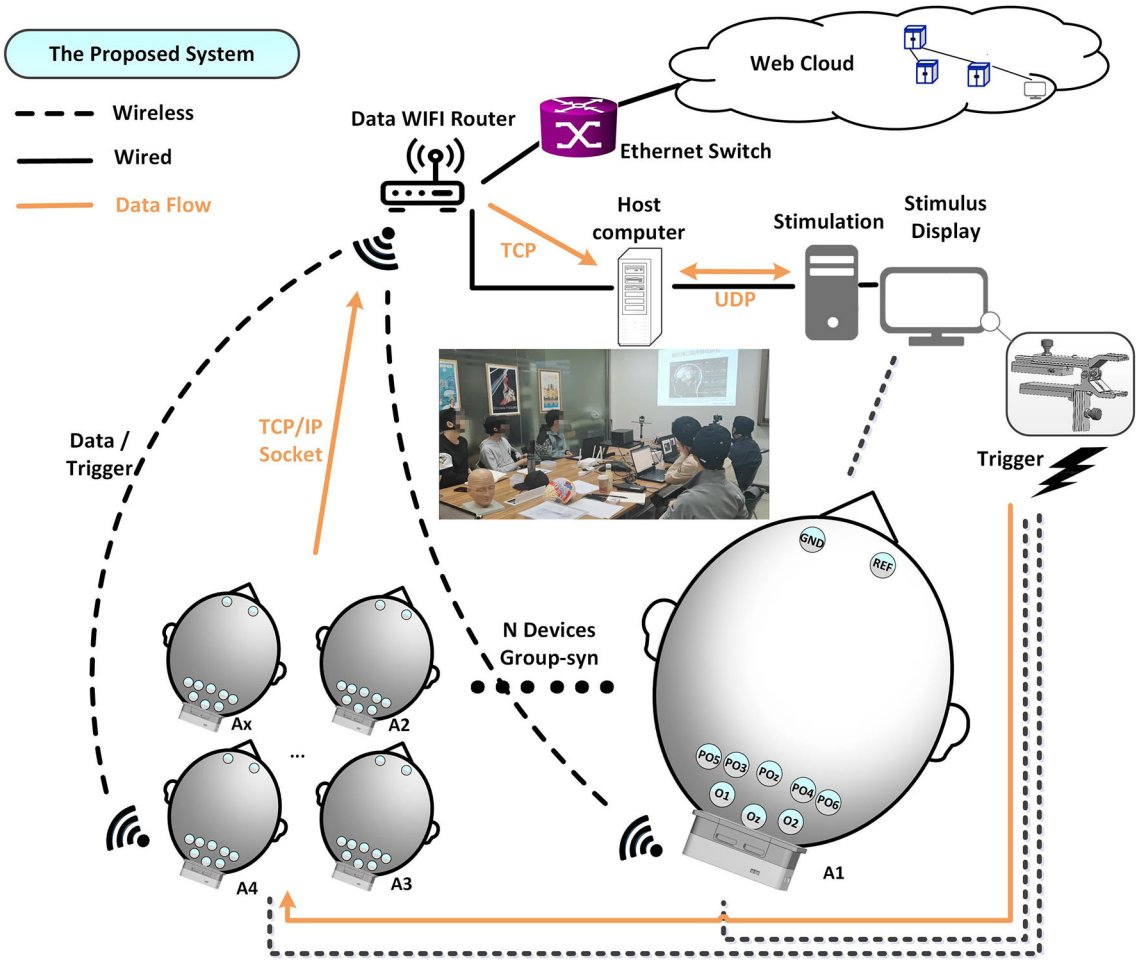


FIGURE 1
The proposed mBCI system implemented on 10 amplifiers.

indicate different events or sequences), as shown in Figure 2A(a) red box.

Raw EEG data packed together with the synchronized trigger event are wirelessly sent to the recording host computer (acquisition server) via TCP/IP socket. The host computer receives data packets from the devices and parses them according to the predefined data protocol. The recorder and analysis perform processing functions, such as filtering, data saving, and real-time decoding.

2.2. Software architecture

The software architecture is divided into four parts: trigger, recorder, analysis, and display of stimuli. The trigger is independent of the acquisition system and does not require an added hardware trigger box for connectivity. To ensure complete data are transmitted wirelessly, a TCP connection links the acquisition and recording software. Different algorithmic microservices acquire potential data from different devices over different ports when the data need to be simultaneously analyzed. TCP is used where reliable transmission is necessary at the transport layer, whereas UDP is used for communication where high-speed transmission and real-time performance are required. Because the quantity of data is minimal and resides under wired local area networks or local hosts, we used a UDP connection between display and analysis software. The UDP packet header is 8 bytes, with relatively low overhead compared with the 20-byte packet header in TCP. In addition, using UDP packets allow for lower connection latency and network traffic by reducing the three-way handshake. In this way, the results can be timely feedback to the display. Using a 40 ms packet transfer (35 bytes/ms, 1,400 bytes per packet), we maintained each maximum packet size within the maximum transmission unit (MTU) range of 1,500 bytes, thereby facilitating optimal TCP and UDP transfers. In brief, TCP is more intricate, with a higher volume of header data, which guarantees wireless communication reliability. Conversely, UDP saves network traffic by eliminating the requirement for packet loss retransmission, resulting in improved real-time performance.

We designed the application layer protocol, as shown in Figure 2A(a). A frame header (48 59 3C, 3 bytes) is used to locate different nodes. An index (from 00 00 to FF, 2 bytes) is used by the node to verify the data integrity. Each channel of EEG data consists of 3 bytes, while an additional 3 bytes are allocated for trigger data,

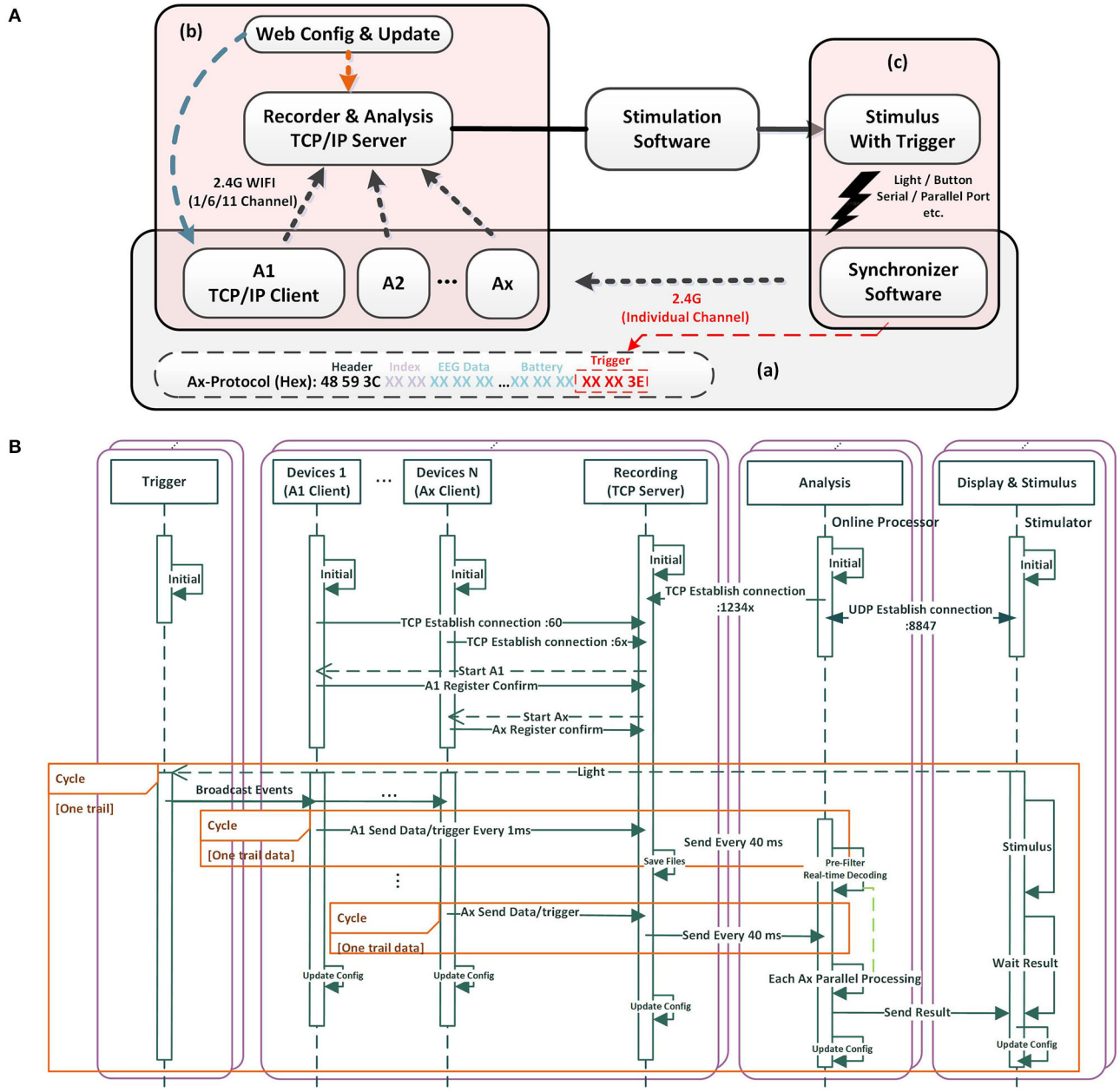


FIGURE 2
The mBCI system software architecture and processing flow. **(A)** Software featured with individual connection, run-time configuration and custom protocol. **(B)** The software processing flow.

which serves as event number for different events. Similarly, 3 bytes are assigned for battery information. Each packet contains 35 bytes, with a transmission rate of 1,400 bytes every 40 ms. The entire system can be run-time configured by updating the JSON file that defines the IP, port, channel, sample rate, and other variables. This feature allows for easy configuration updates while the system is running. In addition, the server and client A_x can be configured and updated by internet users, allowing upload data or reporting of results to some network infrastructure, as shown in Figure 2A(b). Based on the task design and stimulation software, the stimulus with trigger can be light, audio, button, serial/parallel port, and so on, as shown in Figure 2A(c).

According to the software processing flow (refer to Figure 2B), we prioritized server and client operations to secure efficient acquisition and recording. We designed the recording part as a TCP server. Once turned on, each device (client A_x) automatically connects with the server, registers, and awaits the start acquisition command. Upon receiving the command, the client A_x sends the EEG data to the server. To further enhance data reliability, the recording unit opens an additional server port, awaiting a TCP connection from the analysis component. Once initiated, the analysis unit connects with the recording unit and establishes a UDP connection with the stimulus display. Additionally,

the analysis part can operate independently without any online requirements.

To efficiently synchronize data, the display stimulus unit emits light. When detecting a sudden flash, the light trigger sends the synchronization signals to the devices through an independent channel. The synchronization signals can be marked in the data protocol packets of each device. Client A_x sends data/trigger to the recording TCP server every 1 ms. The recording TCP server sends to the analysis unit after packetizing every 40 ms. This process is repeated in subsequent rounds based on the experimental paradigm design.

3. Methods and experiments

We used a three-step method to evaluate the system's performance. The first step involved evaluation of hardware performance, the second involved evaluating ERP signal qualities, and the last involved conducting group cognition experiments.

3.1. Evaluation of hardware performance

We evaluated the hardware's performance based on three key parameters: data correlation coefficient, CMRR, and noise. In noisy settings, external industrial frequency and wireless interferences often cause common mode noise, which can disturb the acquired device data. In cases where physiological signals are weak, CMRR serves as a critical metric for demonstrating the ability to suppress common mode signals.

3.1.1. Synchronization test

We used Pearson's correlation coefficient, as described by Rodgers and Nicewander (1988), to estimate the correlation between two waves in our study. The coefficient is defined as the quotient of covariance and standard deviation between two variables:

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E\left[(X-\mu_X)(Y-\mu_Y)\right]}{\sigma_X \sigma_Y} \quad (1)$$

The abovementioned equation defines the overall correlation coefficient. For multi-channel data, we established the mutual covariance matrix using this equation.

A light-trigger method was applied to test synchronization strategies by employing a signal generator and wireless trigger unit. Specifically, the signal generator sent out a stimulus signal, 10-Hz sine with 10 mVpp, as shown in Figure 3A. The signals were sent to each amplifier using a high-performance cable and lasted for 120 min. Throughout the continuous operation, we recorded and monitored the stability of signals using GUI software and inspected the output waveform of 10 devices every half an hour from an oscilloscope. Furthermore, the spacing between trigger occurrences was measured.

The trigger-sender contains an optical sensor. The wireless trigger-sender simultaneously sends the signal to the trigger-receiver of each EEG device, as shown in Figure 3A. In the experiment, the signal source and the 10 devices were positioned on opposite sides.

Time differences were measured to estimate interference caused by the spatial transmission between the trigger sender and receiver unit, as well as to examine time delays between each transceiver protocol. The end-to-end delay is the delay between the two probes, as shown in Figure 3B, including the static delay and the dynamic delay. We defined two types of delays, as shown in Figure 3C. The minimum synchronization error is the minimum value of the dynamic delay among the values obtained from repeated experiments.

3.1.2. CMRR test

The CMRR indicates the rejection ability of the common mode signal in the differential amplifier. The calculation method is as follows:

$$CMRR = 10 \times \log_{10}\left(\frac{V_d}{V_{cm}}\right)^2 = 20 \times \log_{10}\left(\frac{V_d}{V_{cm}}\right) \quad (2)$$

where V_d represents the voltage amplification factor of the differential mode signal and V_{cm} represents the voltage amplification factor of the common mode signal. To measure the root mean square of the input noise, we short-circuited each input channel with the reference electrode. We employed this approach in the following experiments of hardware evaluation. To carry out the CMRR test, we connected all input channels and REF port to the positive output of the waveform generators, while connecting GND to the negative output. The waveform generators transmitted a 10-Hz sine wave with 500 mVpp through high-performance cable. We calculated the noise by shorting each input channel and the REF port with no signal.

We proposed the novel mBCI system, each wearable compact amplifier weighs 56 ± 4 g, and the size is 59.3 × 47.4 × 22.7 mm. The average noise amplitude is 0.87 μVrms @2–45 Hz, and the average CMRR of all the tested devices except abnormal A4 is 109.03 dB @10 Hz, as shown in Figure 4. This wearable compact system allows a 10 m distance between users by hardware-based synchronization among 10 users. The receiver in each amplifier receives triggers/markers from the wireless trigger sender.

Typically, the noise evaluated in such cases is the input reference noise (Bolatkale et al., 2014). To overcome the motion and the high-frequency thermal noise of the components, we applied a filter that allowed signals from 2 to 45 Hz. One of the devices (A4) had a negligible deficit.

3.2. Functionality evaluation

For the functionality of the system, we selected the SSVEP method as the key indicator. SSVEP is capable of detecting the periodic synchronization of the brain with an external flickering visual stimulus delivered at a fixed frequency, making it an ideal tool for assessing synchronization

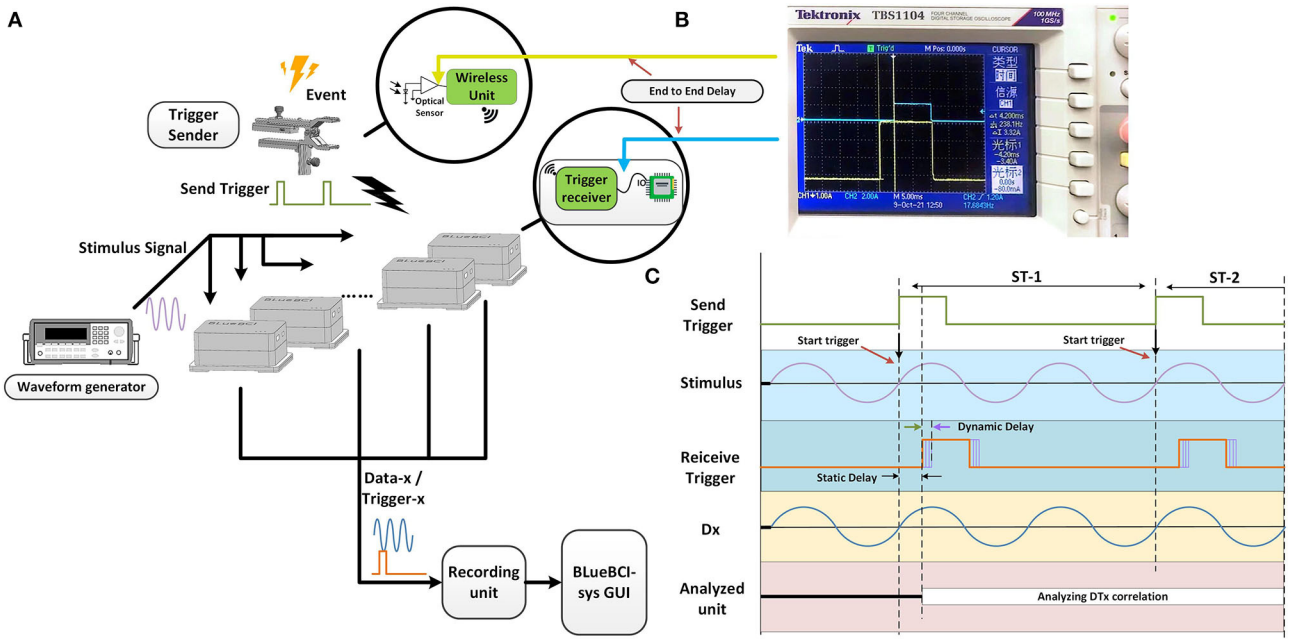


FIGURE 3
The mBCI system synchronization test. **(A)** The architecture of synchronization test. **(B)** The test of end-to-end delay. **(C)** The process of synchronization test and two types of the delays defined.

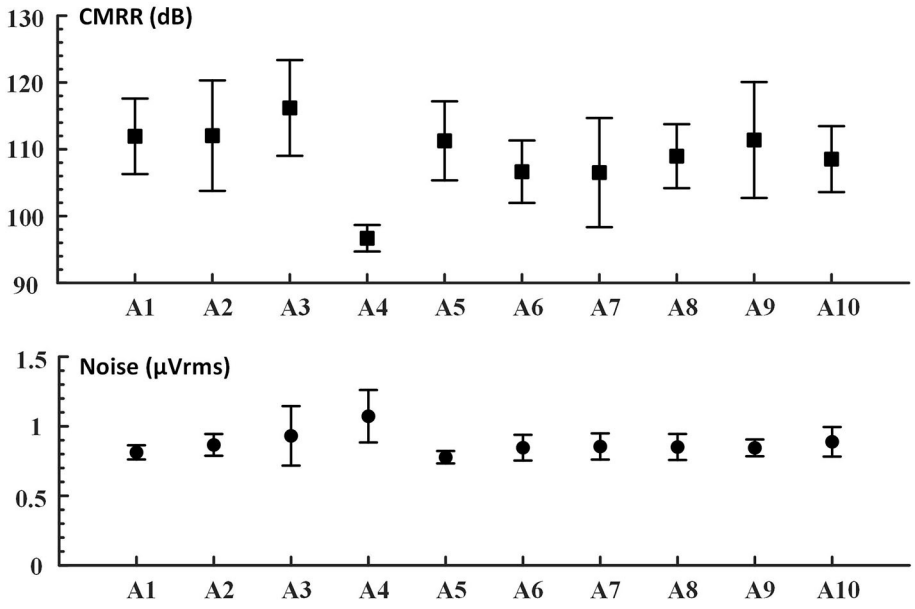


FIGURE 4
The device CMRR (dB) @10 Hz; noise (μVrms) @2–45 Hz. In this study, based on the mBCI 10 devices (A1–A10), the bandpass filters (2–45 Hz) for noise calculation from original EEG signals were in Chebyshev type I filter order. Evaluations were performed at 1 KHz sampling rates.

accuracy and reliability. We assessed both the phase and frequency accuracy of EEG data recorded at the millisecond level. To compare the proposed system with the Synamps2 EEG system from the market-leading company Neuroscan, Inc., we designed both forty-targets and single-target SSVEP spelling experiments.

3.2.1. SSVEP method

The method was previously described for visual spellers using the sampled sinusoidal stimulus method in a monitor stimulus (Manyakov et al., 2013; Chen et al., 2014). The modulation of the screen brightness represents a stimulus sequence corresponding to the frequency f.

$$s(f,i) = \frac{1}{2} \times \left\{1 + sin\left[2\pi f\left(i/Refresh\ Rate\right)\right]\right\} \quad (3)$$

where *sin*() generates a sine wave and i represents the frame index in the stimulus sequence. represents the Refresh Rate of the screen (the monitor or display).

We followed the method (Chen et al., 2015a,b; Wong et al., 2020) and applied filter bank canonical correlation analysis (FBCCA). This method is widely used to detect the frequency of SSVEP. The SNR and classification analysis can be used to evaluate SSVEP data (Chen et al., 2015a,b; Liu et al., 2020; Ladouce et al., 2022). In this paper, the SNR can be defined as follows:

$$SNR = 20\log_{10}\frac{y(f)}{\frac{1}{2n}\cdot\sum_{k=1}^{n=4} y(f-\Delta f\cdot k)+y(f+\Delta f\cdot k)} \quad (4)$$

where $y(f)$ represents the spectrum calculated by fast Fourier transform, and Δf represents the frequency resolution.

The recognition accuracy and ITR were defined (Chen et al., 2015a,b; Wang et al., 2017). The ITR represents the output information per second or minute. The calculation formula is as follows:

$$ITR = 60\cdot\left(log_2 \mathrm{N} + Plog_2 P + (1-P)log_2\frac{1-P}{N-1}\right)/T \quad (5)$$

The N is defined as the number of commands that can be output by the system. The accuracy of target recognition (P) affects the feasibility and reliability of the BCI communication system. Single-target selection time (T) is often defined as the time required for the BCI system to output a single command. This study refers to visual gaze duration in SSVEP experiments.

3.2.2. SSVEP experiments

During SSVEP experimentation, recordings of data were independently conducted using both the proposed device and Synamps2 in separate sessions. Initially, subjects wore a wet electrode EEG cap and the impedance was ensured <20 k. The experiment sequence was as follows: initial preparation -> System A -> System B -> Rest -> System B -> System A -> End. In the crossover experiment, we randomly selected the first system. To ensure that the participants remained in a good state throughout the experiment, the total duration could not exceed 90 min. The same stimulation screen, stimulus unit, EEG cap, and trigger sender shown in Figure 5 were used during data acquisition, while different systems were switched via a hardware connector.

In the forty-targets and single-target SSVEP spelling test, EEG data were recorded at a 1,000 Hz sampling rate, using the reference electrode, the ground electrode at the position shown in Figure 5. The Synamps2 amplifier sends raw data/trigger to the recording unit (another host computer) using a wire connection. The BLueBCI amplifier wirelessly sends it to the recording unit via TCP/IP socket. The device in the proposed system is named as BLueBCI.

3.2.2.1. Forty-targets SSVEP online analysis

We designed a forty-targets SSVEP spelling board scenario with a display frequency ranging from 8–15.8 Hz, as shown in Figure 6. The scheme of SSVEP trail is shown in Figure 7.

The frequency value of each character in the matrix can be represented:

$$f(k_x,k_y) = f_0 + \Delta f\times[(k_y-1)+(k_x-1)\times 10], \quad k_x\in[1\ \ 4],\quad k_y\in[1\ \ 10] \quad (6)$$

where k_x and k_y represents the row index and column index, respectively. In this study, f_0 was 8 Hz and Δf was 0.2 Hz.

We collected all forty-targets SSVEP data from seven healthy subjects (four males and three females, aged 27 ± 5 years). All participants were either students or staff members from the university and were situated 100 cm away from a monitor during the experiment. The participants had a normal or corrected-to-normal vision and had signed consent papers.

3.2.2.2. Single-target SSVEP offline analysis

The forty-targets SSVEP experiment features certain limitations. First, peripheral scintillations are aliased, which affects the subjects' vision and reduces the SNR. This may lead to inaccuracies in the SNR measurements. Second, the low scintillation frequency results in insufficient response at high frequencies. These factors constrain the extent to which the system's performance can be fully evaluated.

A supplementary experiment was carried out. We designed a single-target SSVEP spelling board scenario with frequencies of 12 Hz and 30 Hz. The experimental procedure was similar to the forty-targets experiment. The target was set in a circular area positioned at the center of the screen. As there was only one target, we removed the cue period from each trail. Blocks spanned 75 s and comprised 15 trials, with each trial lasting 4 s of stimulation followed by 1 s of rest.

All single-target SSVEP data were acquired from nine healthy subjects (one male and eight females with an age of 26 ± 1 years). All participants were either students or staff members from the university and were situated 100 cm from the monitor during the experiment. Participants had a normal or corrected-to-normal vision and had signed consent papers.

3.3. System evaluation in group cognition task

Mental fatigue was associated with increased power in theta (θ) and parietal alpha (α) EEG rhythms. Sleepiness is typically characterized by an increase in theta and alpha activity, with a decrease in the beta band (Balandong et al., 2018).

In cognitive science, frequency domain features are widely used to assess mental fatigue or sleepiness (Eoh et al., 2005). Brain rhythms are generally divided into five sub-bands: δ : 0.5–4 Hz; θ : 4–8 Hz; α : 8–13 Hz; β : 13–25 Hz; and γ : 25–40 Hz. One of the classic formulas is as follows:

$$F_1 = \frac{E_\theta + E_\alpha}{E_\beta} \quad (7)$$

where the total frequency band power:

$$E = \sum_{n=1}^{N_{FFT}}\left(\frac{F(n)}{N_{FFT}}\right)^2 \quad (8)$$

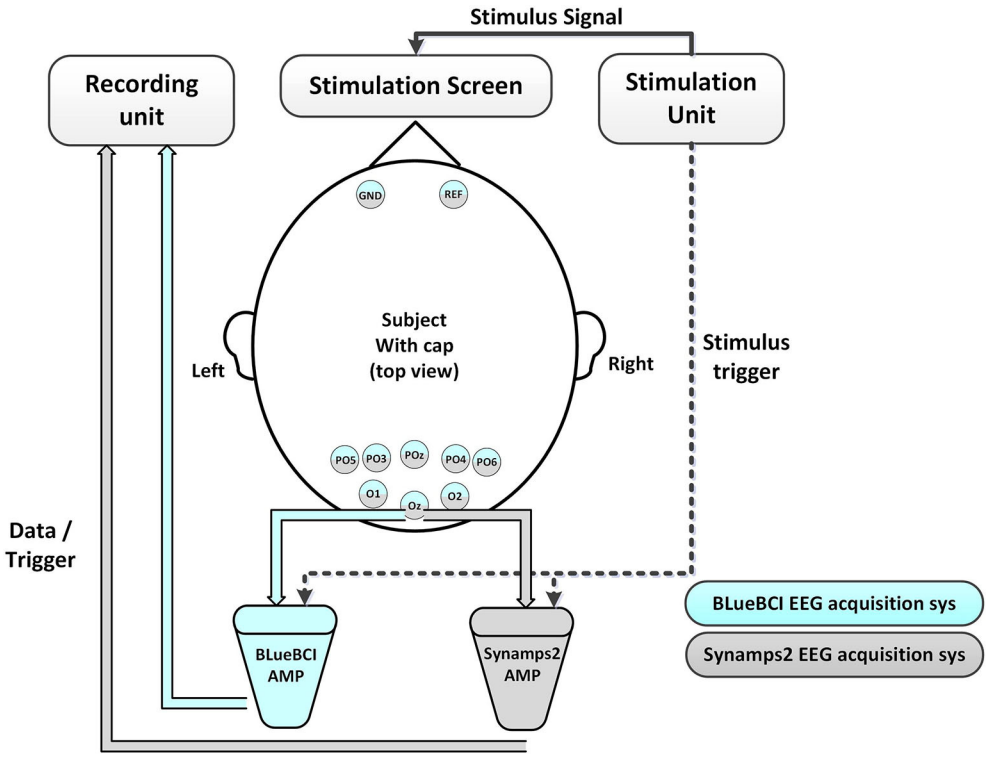


FIGURE 5
The comparison scheme between BLueBCI and Synamps2 amplifier.

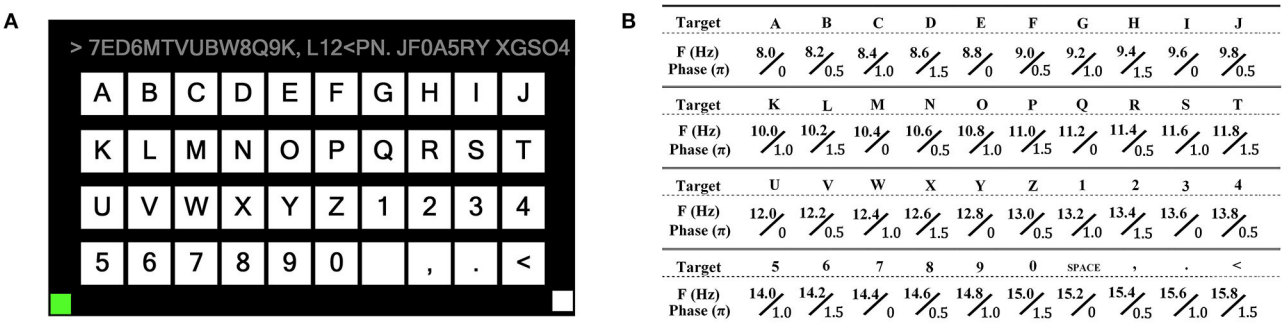


FIGURE 6
Forty targets of BCI spelling board. **(A)** A board layout with 26 letters, 10 numbers, and four non-alphanumeric keys (space, comma, dot, and backspace) arranged in four rows and 10 columns. The upper is used to display the input characters. **(B)** Encoding the frequency and initial phase of each target using joint frequency and phase modulation (Chen et al., 2014).

where $F(n)$ denotes the results of the signal $X(n)$ at frequency n.

Quick and reliable signal acquisition is vital for large-scale applications, especially for the classroom cognitive application. To achieve this, we used dry electrodes to collect EEG data from the occipital area in noisy settings. In this group cognition test, the data were acquired from 10 healthy subjects, consisting of nine students and one teacher. The mBCI scenario is shown in Figure 8.

Estimating the mBCI application test was constructed. The teaching process was divided into four parts. This process was as follows: Initially, all subjects prepared EEG cap for 10 min before class, Period 1 (Guide 2 min), Period 2 (Online teaching 27 min), and Period 3 (Offline teaching 49 min), as shown in Figure 8. There was no break from 10:00 to 11:20 AM. Before the online teaching, 2 min of guide (1-0-Contrast-G) was used as the control group. During Period 2, the process included closing the eyes for 2 min (1-1-close), opening the eyes for 2 min (1-1-open), playing the teaching video at 1x speed for 10 min (1-1-oT10), playing the video at 2x speed for 5 min (1-2-oT20), playing the video at 1.5x speed for 6 min (1-3-oT15), and then playing the video at 1x speed for 2 min (1-4-oT10). During Period 3, the process included opening the eyes for 2 min (2-0-open), closing the eyes for 2 min (2-0-close), followed by offline teaching for 45 min (2-1-1).

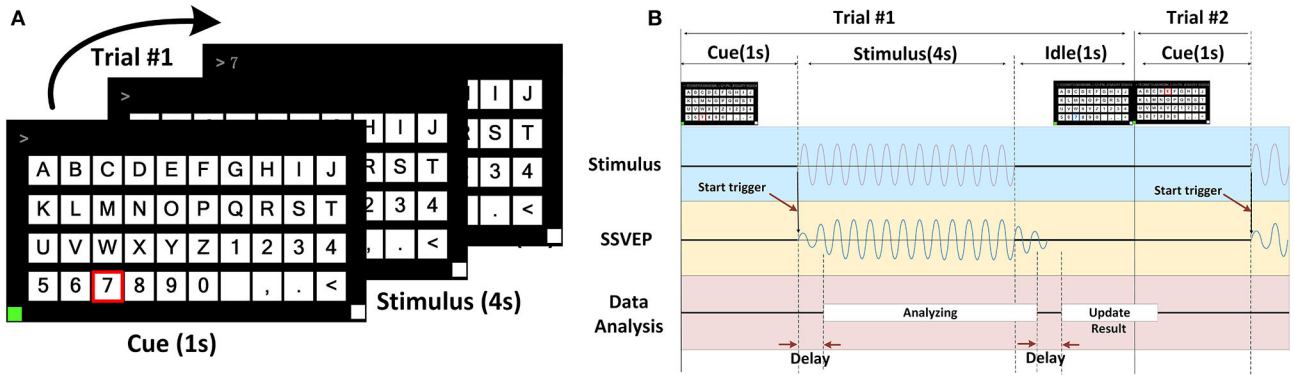


FIGURE 7
The scheme of SSVEP trail. **(A)** Each block consisted of 40 trails. Each trial was divided into three periods: cue, stimulus, and idle period. The cue was given 1 s (a red box around the target). The subjects followed the cue target, and then, all the targets started flashing at the same time for 4 s. At the end of each trial, the participants had a 1-s idle period. **(B)** Trigger started at the end of the cue and the beginning of the stimulus. During the 4-s flashing, subjects were asked to avoid eye-blinking while flashing. The real-time SSVEP EEG data were analyzed. During the 1-s idle period, the EEG feedback was analyzed using the FBCCA method, and then, the updating result was presented in the input field above, and then, the next trial was followed.

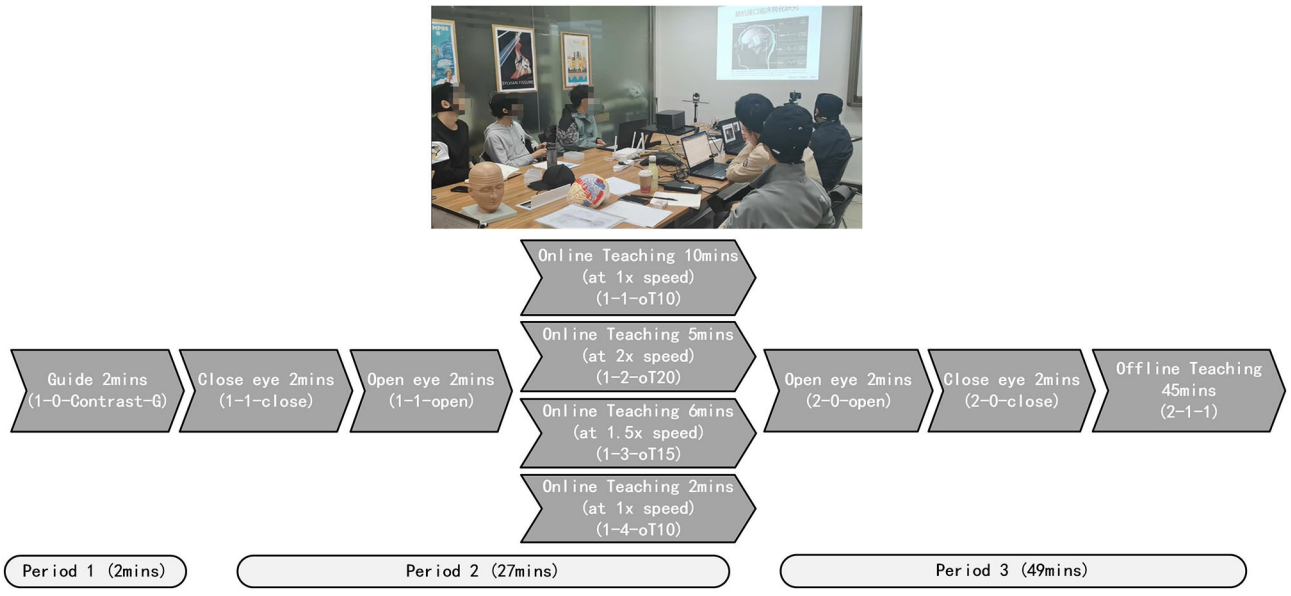


FIGURE 8
The mBCI scenario for education application.

4. Results and discussion

4.1. Comparison of synchronization performance

We delivered a light input from an optical sensor and used an oscilloscope to detect two types of delays. One oscilloscope probe is used to test the output of the optical sensor, and the other probe is used to test the TTL signal output from the trigger-receiver unit before entering the microcontroller unit IO port, as shown in Figure 3B.

We measured the time difference between the probes using an oscilloscope. We found a static delay of ~4 ms and a dynamic delay of roughly 0.9 ms over the course of the 30-min recording period, as shown in Figure 3B. The static delay and dynamic delay observed at the edges of the square wave indicated that time drifted from synchronous sampling, as shown in Figure 3B.

This result suggests that the data recorded by the 10 amplifiers were not precisely aligned. The static delay variation is <1 μs, which is almost constant over time. The static delay can be subtracted from the recorded data, while the dynamic delays can potentially lead to errors in the analysis of brain activity below the 1 ms level.

Furthermore, to assess the synchronization performance, we conducted an experiment to calculate the dynamic delay between the signals. We took the last 50,000 points (1,000 Hz by 50 s) with the trigger and used the Butterworth 50 Hz notch filter. After aligning the trigger, we analyzed the correlation of 40,000 points

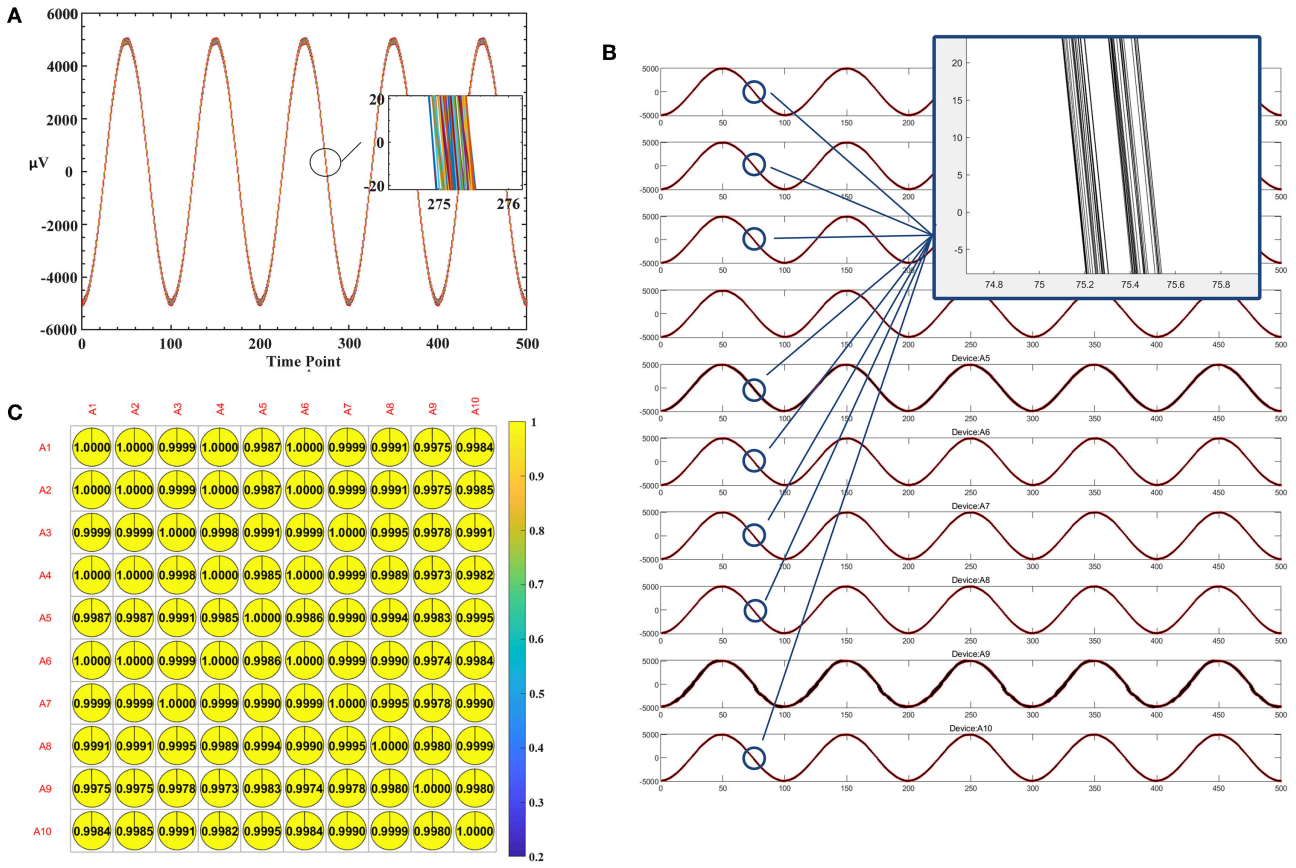


FIGURE 9
Estimating dynamic delay and correlation. **(A)** The plot of 80 consecutive segments of A1 first channel. **(B)** The plot of eight-channel signals from 10 devices. **(C)** The mutual covariances of each device's first channel.

data. To assess the cumulative phase error of individual device, we have analyzed the signals of one channel. We split the 40,000 points of A1 first channel into 80 consecutive segments (one segment every five cycles, 1 k sample rate, 100 points per cycle). These 80 segments' points are drawn in Figure 9A. The enlarged part of the figure shows that the deviation is in the range of 0.5 ms. To evaluate the phase shift between multiple devices, the signals from 8 channels of 10 devices were analyzed. The signals of 8 channels from 10 devices were plotted simultaneously in Figure 9B. This showed the waves are well synchronized and have a 0.5 ms dynamic delay. The mutual covariances of each device's first channel were shown in Figure 9C. The average correlation reached 99.93%.

After 30 repeated experiments, the minimum synchronization error was 237 μs and the average was 0.9 ms. The causes of the minimum synchronization error would refer to the cumulative "time delay" of the electronic system in the wireless transceiver process, including the baseband protocol resolution time delay, the crystal clock difference of all the subsystems, and micro-control-unit command sequence time difference.

Among all EEG systems, Emotiv is the most commonly used wearable system in research studies (Roy et al., 2019), whereas Brain Products, EGI, BIOSEMI, and g.Tec are the most frequently used desktop systems in hyperscanning studies (Barraza et al., 2019).

In particular, for hyperscanning setup types, a wireless wearable system can overcome limitations on subjects' range of motion and adapt to the experimental paradigm design (Xu et al., 2021). While a wired fixed-linked system is widely used, it can be quite inconvenient and time-consuming when conducting hyperscanning experiments. In addition, it is complicated to acquire simultaneous data with multiple devices. It suggests that wireless wearable setups with a moderate number of channels (8/16/32) can be the most suitable for the mBCI system.

According to our literature review, some related work for the mBCI application are listed in Table 1. Several wired trigger boxes are able to distribute a SYNC signal to different devices simultaneously, performing a hardware-based trigger synchronization method, such as, g.TRIGbox (by g.Tec), USB2 Receiver (by BIOSEMI), and Clock Sync box (by EGI). The systems from Brain Products, EGI and BIOSEMI adopt parallel interfaces of the host computer for hardware-based trigger synchronization since the hardware interrupt level of the parallel interface has a high priority. This can result in a faster response of processing, such as the device from g.Tec, which reaches a synchronization delay of 51.22 ms. Synchronization delay of a wireless system from Emotive adopting audio/video data is 162.69, while the proposed system achieves a far smaller synchronization delay of <1 ms by exploiting light signal and customized data protocol.

TABLE 1 Comparison between the proposed and other commercial EEG acquisition system.

Devices	Brain products[a]	EGI[a]	BIOSEMI[a]	g.Tec g.USBamp[b,c]	Emotiv EPOC[c,d]	The proposed
Hyperscanning setup types	Fixed-linked	Fixed-linked	Fixed-linked	Fixed-linked	Wireless wearable	Wireless wearable
Channels number	32/64	128	64	16	14	8/16/32
Device number	2	2	2	4	9	10
Synchronization delay	Not mentioned	Not mentioned	Not mentioned	51.22 ms	162.69 ms	<1 ms
Trigger synchronization method	Wired TTL Software: LSL protocol	Wired TTL Clock Sync box	Wired USB receiver	Wired USB (g.TRIGbox) Software: LSL protocol	Wireless audio/video data Software: LSL protocol	Wireless light signal Software: customized data protocol

[a]By Barraza et al. (2019).
[b]By Bilucaglia et al. (2020).
[c]By Wang et al. (2019).
[d]By Poulsen et al. (2017).

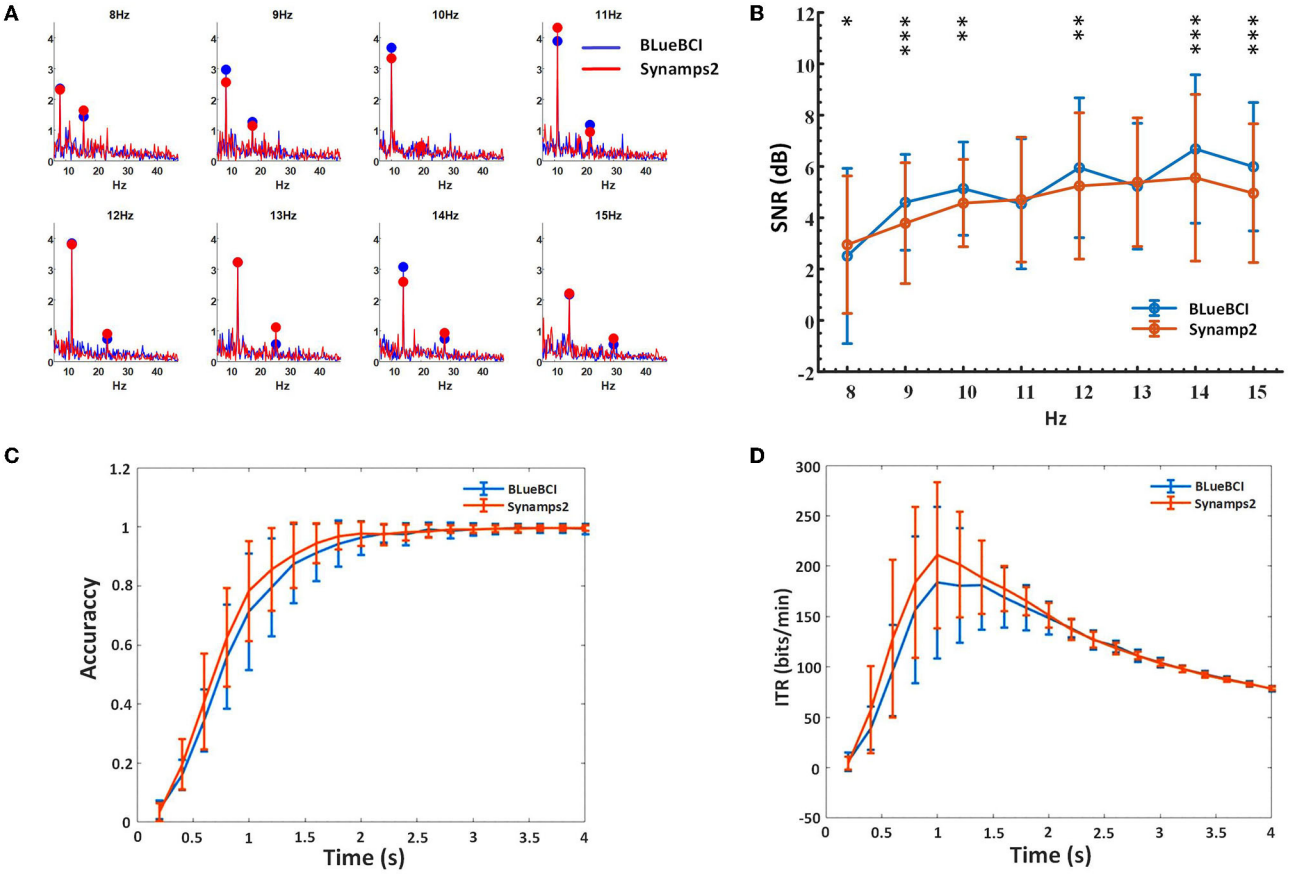


FIGURE 10
Key performance comparison in the online forty-targets SSVEP experiments. **(A)** Spectrum diagram. **(B)** SNR based on the FBCCA method. The asterisks indicate significant difference by paired t-tests (*$p < 0.8$, **$p < 0.6$, ***$p < 0.4$). **(C)** Classification accuracy. **(D)** ITR. We estimated by data lengths ranging from 0.2 to 4 s, with 0.2-s intervals.

4.2. SSVEP performance comparisons

In this section, we present a comparative result of the mBCI devices with temporal and spatial analyses. Each participant had individual REF and GND channels. The channels were connected to the individual amplifiers. In the acquisition software, the EEG data with the triggers can be viewed separately and stored in different files.

The raw SSVEP data underwent several filtering steps. Firstly, we extracted the 4-s stimulated data by the trigger event (the

10.3389/fnins.2023.1176344

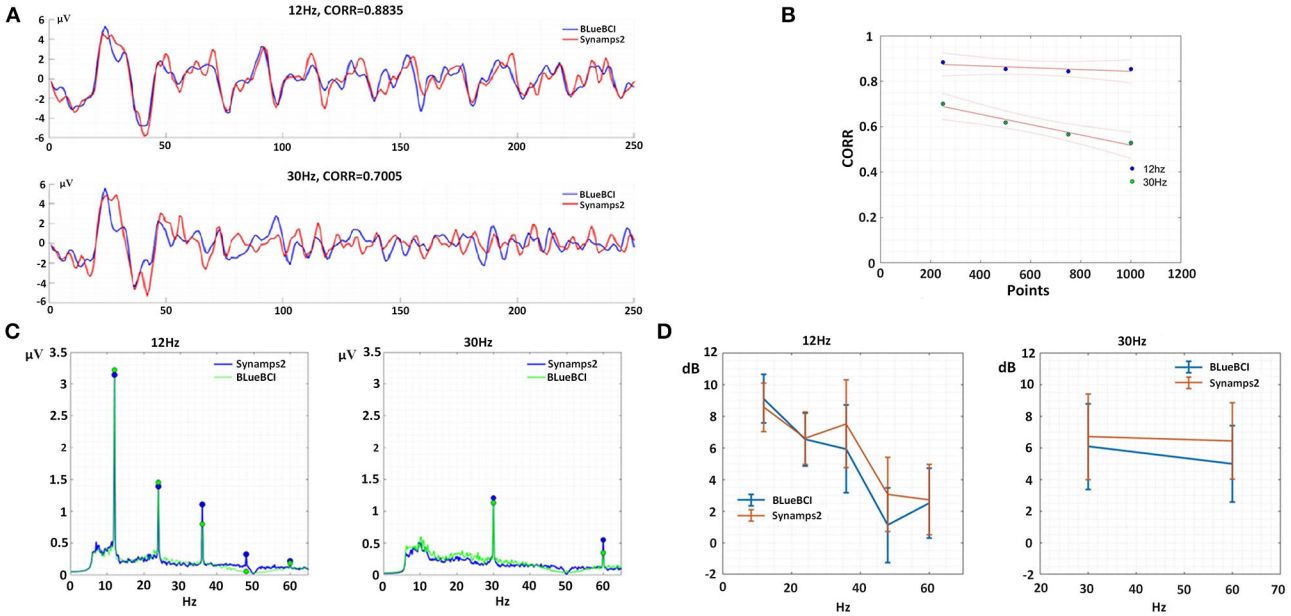


FIGURE 11
Typical temporal and spatial characteristics. **(A)** Temporal analysis at 250 points per second. **(B)** Unary linear fitting graph of similarity and 1,000 sampling points. **(C)** Spectral analysis, left: fundamental (12 Hz) to fifth harmonic (24, 36, 48, and 60 Hz), right: fundamental (30 Hz) to second harmonic (60 Hz). **(D)** SNR analysis for single-target SSVEP.

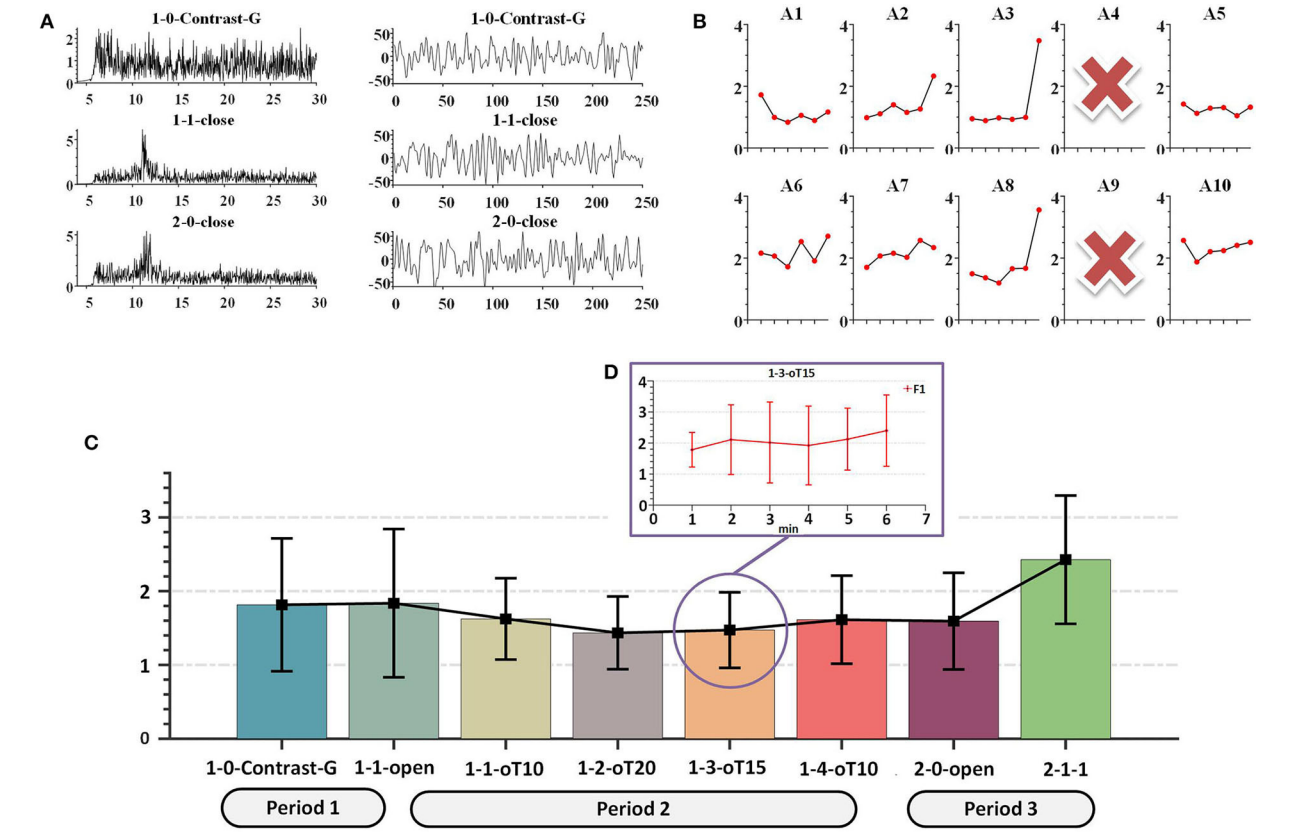


FIGURE 12
Typical fatigue cognition characteristics. **(A)** Temporal, spatial diagram. **(B)** Each subject F1 During 1-1-oT10 to 2-1-1. **(C)** F1 diagram During 1-0-Contrast-G to 2-1-1, the black line represents the collaborative result. **(D)** During 1-3-oT15, the F1 value with the interval 1 min.

sampling rate of both devices was 1,000 Hz, and 4-s data were 4,000 points per trial). Secondly, any trail data with an amplitude of >100 μV were removed, after which the data were downsampled to 250 Hz. Thirdly, the data were processed by Chebyshev type I bandpass filter, which has a stopband of 3–6 Hz, a passband of 6–65 Hz, and a stopband of 65–75 Hz. This was to eliminate environmental noise. Following data preprocessing, the EEG signal was stored as three-dimensional data with channel × time points × blocks.

4.2.1. Forty-targets SSVEP

In the forty-targets SSVEP experiment, we focused on the data resulting from the stimulus frequencies of 8, 9, 10, 11, 12, 13, 14, and 15 Hz. The comparison spectrum diagram plot was drawn between the two devices, as shown in Figure 10A. The comparison data between one device of the mBCI system and the Synamps2 device showed obvious SSVEP fundamental frequency and harmonic response.

We compared 11 sets of BLueBCI data and 13 sets of Synamps2 data by analyzing frequencies of 8, 9, 10, 11, 12, 13, 14, and 15 Hz. Paired *t*-tests showed that there was no significant difference in classification accuracy and ITR between the mBCI system device and the Synamps2 device ($p = 0.820$ and 0.656, respectively). However, *t*-tests of SNR showed there was some difference between them, as depicted by the asterisks in Figure 10B.

The results show that the average SNR of BLueBCI (5.08 ± 2.03 dB) is higher than that of Synamps2 (4.66 ± 1.76 dB), as shown in Figure 10B. Based on the 0.2-2 s data, the resulting peak accuracy for BLueBCI is slightly lower than that for Synamps2. With 2 s data length, the accuracy of target recognition was 98%, similar to that of Synamps2 (99%), as shown in Figure 10C. Based on the 0.2-2.2 s data, the ITR of BLueBCI was lower than that of Synamps2. But the ITR performance of the BLueBCI and the Synamps2 tends to be identical after 2.2 s. With 2 s data length, the average ITR reached 150 ± 20 bits/min, and the highest reached 260 bits/min (data length: 1 s), which was comparable to Synamps2 (the average: 150 ± 15 bits/min, the highest: 280 bits/min), as shown in Figure 10D.

4.2.2. Single-target SSVEP

In the single-target SSVEP experiment, we obtained averaged data for each device at each frequency from all subjects to eliminate time-space and phase differences caused by multiple subjects. The averaged data were plotted in Figure 11A. Linear fitting was conducted to evaluate the similarity and sampling points as shown in Figure 11B. Firstly, in terms of 12 Hz, the similarity remained nearly constant with an average of 86% as sampling points increased. Secondly, in terms of 30 Hz, the similarity showed a significant negative correlation with the increase in sampling points. This observation suggests that the similarity gradually decreased with increasing frequency, possibly due to insufficient data or multi-subject inconsistent time-space and phase response to a high-frequency stimulus. Our results imply that this approach can be useful for screening individuals for efficient interaction and remain a potential area for further study.

Among the nine subjects, 20 sets of 12 Hz-target and 21 sets of 30 Hz-target were collected by Synamps2; 19 sets of 12 Hz-target and 19 sets of 30 Hz-target were collected by BLueBCI. In Figure 11C, the data by BLueBCI (green line) and Synamps2 (blue line) showed obvious fundamental frequency and harmonic responses, among which the 4th and 5th harmonic responses in Synamps2 12-Hz stimulation were obvious, while the BLueBCI exhibited better performance. However, in the 12-Hz stimulus, the fundamental and second harmonic responses for BLueBCI were higher than that for Synamps2. In the 30-Hz stimulus, the fundamental frequency and second harmonic frequency were lower than that of Synamps2. We compared frequency responses SNR of 12 Hz and 30 Hz, respectively, as shown in Figure 11D. The 12 Hz and 24 Hz SNR of BLueBCI were higher ($p = 0.230$ and 0.847, respectively) than Synamps2. There was no significant difference in the fifth frequency (60 Hz, $p = 0.856$). However, for the 36 Hz and 48 Hz SNR for BLueBCI were significantly lower ($p = 0.048$ and 0.009, respectively) than Synamps2. For the 30-Hz stimulus, the fundamental frequency and the second harmonic response (60 Hz) SNR for Synamps2 were higher ($p = 0.137$ and 0.065, respectively) than those of BLueBCI. This implied that there were some minor defects in acquisition and processing. Simulated IIR filtering was carried out (notch 50 Hz) on the raw Synamps2 data, which revealed a gap in the 36–60 Hz region of the spectrum, confirming the same phenomenon detected in the BLueBCI data. This approach can be a method for assessing hardware performance and be a guide for further optimization and system design.

4.3. Group cognition result and discussion

The contrast group data were obtained from the following periods: 1-0-Contrast-G, 1-1-close, and 2-0-close. Both FFT and temporal characteristics were evident in periods 1-1-close and 2-0-close, as depicted in Figure 12A, indicating the validity of the raw data from the quick and simple wearable mBCI system. During the 1-0-Contrast-G period, the fatigue value was used in the performance comparison between online and offline teaching.

The data collected by the A4 and A9 amplifiers were removed due to hardware defects during the evaluation. The data were then downsampled to 250 Hz and filtered by a 50-Hz notch and Chebyshev type I bandpass filter with a passband from 6 to 65 Hz. First, in the hardware performance results, we found that the CMRR was abnormally low and the noise was higher than other devices, as shown in Figure 4 (the device CMRR). Second, due to the abnormal movement of the caps during the experiment, the low and abnormal high-frequency data were filtered.

As shown in Figure 12D, the overall F1 score increased over the 6-min period, indicating a rise in fatigue cognition. Additionally, during the 1-1-oT10 to 2-2-1 stage, there were variations in fatigue levels among subjects, as depicted in Figure 12B. Paired *t*-tests revealed that subjects A1, A5, and A10 showed a significant difference ($p < 0.03$) in fatigue levels compared with the collaborative result, while the fatigue levels of subjects A2, A3, and A8 were consistent ($p > 0.22$) with the collaborative result. Moreover, there was a significant difference between 1-1-oT10 and 2-1-1 periods ($p = 0.04$). The collaborative results showed that the

fatigue level decreased during the initial stage of 1-1-open to 1-1-oT20 but subsequently increased from 1-1-oT20 to 2-2-1, as shown in Figure 12C. It revealed an apparent trend of increasing the fatigue level across multiple subjects with personality differences. These outcomes were consistent with the previous fatigue detection research (Eoh et al., 2005).

In future, this wearable mBCI system can be utilized for real-time and rapid multi-subject EEG recording with synchronous collaborative computing. As demonstrated in this multi-subject experiment, it serves as a basic tool for exploring cognitive neuroscience or other multi-subject applications.

5. Conclusion

The mBCI system forms the basis for group-cognitive applications. When acquiring brain signals from multiple subjects, it is essential to deploy a wearable, user-friendly, reliable, and sturdy neural recording system with high-performance and synchronization abilities. This cutting-edge wearable mBCI system combines inputs from up to 10 users. First, in terms of SSVEP performance, it results in a higher SNR than NeuroScan Synamps2, with comparable ITR and accuracy. Second, it leverages millisecond-parallel neuro-recording and offers superior portability than other hyperscanning systems. Moreover, the mBCI signal correlation attains 99.8%, with minimal synchronization errors (237 μs). Regarding hardware performance, the average noise amplitude is 0.87 μV, and the average CMRR reaches 109.02 dB. Each wearable compact device weighs just 56 ± 4 g and measures a mere 59.3 × 47.4 × 22.7 mm. In evaluating its suitability for multi-subject teaching applications, preparation required <10 min. Group-cognitive assessment findings not only reveal individual variations but also offer insights into group EEG fatigue cognitive neurology.

Evaluation results indicate that the proposed mBCI system is a highly efficient tool for real-time research and the system will facilitate various applications in the fields of swarm intelligence and cognitive neurology.

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 human participants were reviewed and approved by the Institution Review Board of Tsinghua University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

YoH, YW, and LZ: study conception and design. YoH: data collection and visualization. YW, WP, and XG: experiment validation. YW, ZZ, and LZ: writing—review and editing. YoH and YuH: analysis and interpretation of results. LZ: supervision and funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Key-Area Research and Development Program of Guangdong Province under grant: 2021B0909060002, the National Key R&D Program of China under grant: 2017YFA0205903, the National Natural Science Foundation of China under grant: 62071447, the Beijing Science and Technology Program under grant: Z201100004420015, and the Strategic Priority Research Program of Chinese Academy of Sciences under grant: XDB32040200.

Acknowledgments

The authors would like to thank the subjects who participated in this study and also thank Weijie Zhu, Cheng Jin, Longjun Fang, Li Zheng, Lu Jiang, Jiawei Xu, and Ning Ma for their help in data collection, visualization, and review.

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

References

Artoni, F., Barsotti, A., Guanziroli, E., Micera, S., Landi, A., and Molteni, F. (2018). Effective synchronization of EEG and EMG for mobile brain/body imaging in clinical settings. *Front. Hum. Neurosci.* 11, 652. doi: 10.3389/fnhum.2017.00652

Ayrolles, A., Brun, F., Chen, A., Djalovski, A., Beauxis, Y., Delorme, R., et al. (2020). HyPyP: a Hyperscanning Python Pipeline for inter-brain connectivity analysis. *Soc. Cogn. Affect. Neurosci.* 16, 72–83. doi: 10.31234/osf.io/x5apu

Babiloni, F., and Astolfi, L. (2014). Social neuroscience and hyperscanning techniques: past, present and future. *Neurosci. Biobehav. Rev.* 44, 76–93. doi: 10.1016/j.neubiorev.2012.07.006

Balandong, R. P., Ahmad, R. F., Mohamad Saad, M. N., and Malik, A. S. (2018). A review on EEG-based automatic sleepiness detection systems for driver. *IEEE Access* 6, 22908–22919. doi: 10.1109/ACCESS.2018.2811723

Barraza, A., Dumas, G., Liu, H., Blanco-Gomez, G., van den Heuvel, M. I., Baart, M., et al. (2019). Implementing EEG hyperscanning setups. *MethodsX* 6, 428–436. doi: 10.1016/j.mex.2019.02.021

Bhattacharyya, S., Valeriani, D., Cinel, C., Citi, L., and Poli, R. (2021). Anytime collaborative brain–computer interfaces for enhancing perceptual group decision-making. *Sci. Rep.* 11, 96434. doi: 10.1038/s41598-021-96434-0

Bilucaglia, M., Masi, R., Stanislao, G. D., Laureanti, R., Fici, A., Circi, R., et al. (2020). ESB: a low-cost EEG Synchronization Box. *HardwareX* 8, e00125. doi: 10.1016/j.ohx.2020.e00125

Bolatkale, M., Breems, L. J., and Makinwa, K. A. A. (2014). *High Speed and Wide Bandwidth Delta-Sigma ADCs*. Cham: Springer International Publishing.

Bowyer, S. M. (2016). Coherence a measure of the brain networks: past and present. *Neuropsychiatr. Electrophysiol.* 2, 7. doi: 10.1186/s40810-015-0015-7

Chen, X., Wang, Y., Gao, S., Jung, T. P., and Gao, X. (2015a). Filter Bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface. *J. Neural Eng.* 12, e046008. doi: 10.1088/1741-2560/12/4/046008

Chen, X., Wang, Y., Nakanishi, M., Gao, X., Jung, T.-P., and Gao, S. (2015b). High-speed spelling with a noninvasive brain–computer interface. *Proc. Natl. Acad. Sci. U. S. A.* 112, 1508080. doi: 10.1073/pnas.1508080112

Chen, X., Wang, Y., Nakanishi, M., Jung, T. P., and Gao, X. (2014). "Hybrid frequency and phase coding for a high-speed SSVEP-based BCI speller," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. (Chicago, IL: IEEE). doi: 10.1109/EMBC.2014.6944499

Chuang, C. H., Lu, S. W., Chao, Y. P., Peng, H., Hsu, H. C., Hung, C. C., et al. (2021). Near-zero phase-lag hyperscanning in a novel wireless EEG system. *J. Neural Eng.* 18, e066010. doi: 10.1088/1741-2552/ac33e6

Czeszumski, A., Eustergerling, S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., et al. (2020). Hyperscanning: a valid method to study neural inter-brain underpinnings of social interaction. *Front. Hum. Neurosci.* 14, 39. doi: 10.3389/fnhum.2020.00039

David Hairston, W., Whitaker, K. W., Ries, A. J., Vettel, J. M., Cortney Bradford, J., Kerick, S. E., et al. (2014). Usability of four commercially-oriented EEG systems. *J. Neural Eng.* 11, e046018. doi: 10.1088/1741-2560/11/4/046018

Eoh, H. J., Chung, M. K., and Kim, S. H. (2005). Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. *Int. J. Indus. Ergon.* 35, 307–320. doi: 10.1016/j.ergon.2004.09.006

Gao, X., Wang, Y., Chen, X., and Gao, S. (2021). Interface, interaction, and intelligence in generalized brain–computer interfaces. *Trends Cogn. Sci.* 25, 671–684. doi: 10.1016/j.tics.2021.04.003

Konvalinka, I., and Roepstorff, A. (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Front. Hum. Neurosci.* 6, 215. doi: 10.3389/fnhum.2012.00215

Ladouce, S., Darmet, L., Torre Tresols, J. J., Velut, S., Ferraro, G., and Dehais, F. (2022). Improving user experience of SSVEP BCI through low amplitude depth and high frequency stimuli design. *Sci. Rep.* 12, 8865. doi: 10.1038/s41598-022-12733-0

Li, L., Huang, X., Xiao, J., Zheng, Q., Shan, X., He, C., et al. (2022). Neural synchronization predicts marital satisfaction. *Proc. Natl. Acad. Sci. U. S. A.* 119, 2515. doi: 10.1073/pnas.2202515119

Liu, B., Huang, X., Wang, Y., Chen, X., and Gao, X. (2020). BETA: a large benchmark database toward SSVEP-BCI application. *Front. Neurosci.* 14, 627. doi: 10.3389/fnins.2020.00627

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT.

Maksimenko, V. A., Hramov, A. E., Frolov, N. S., Lüttjohann, A., Nedaivozov, V. O., Grubov, V. V., et al. (2018). Increasing human performance by sharing cognitive load using brain-to-brain interface. *Front. Neurosci.* 12, 949. doi: 10.3389/fnins.2018.00949

Manyakov, N. V., Chumerin, N., Robben, A., Combaz, A., van Vliet, M., and Van Hulle, M. M. (2013). Sampled sinusoidal stimulation profile and multichannel fuzzy logic classification for monitor-based phase-coded SSVEP brain–computer interfacing. *J. Neural Eng.* 10, e036011. doi: 10.1088/1741-2560/10/3/036011

Minagawa, Y., Xu, M., and Morimoto, S. (2018). Toward interactive social neuroscience: neuroimaging real-world interactions in various populations. *Japan. Psychol. Res.* 60, 196–224. doi: 10.1111/jpr.12207

Müller, J. S., Vidaurre, C., Schreuder, M., Meinecke, F. C., Von Bünau, A., and Müller, K. R. (2017). A mathematical model for the two-learners problem. *J. Neural Eng.* 14, e036005. doi: 10.1088/1741-2552/aa620b

Perdikis, S., and Millan, J., del R. (2020). Brain-machine interfaces: a tale of two learners. *IEEE Syst. Man Cybernet. Magazine* 6, 12–19. doi: 10.1109/MSMC.2019.2958200

Pérez, A., Monahan,:J., and Lambon Ralph, M. A. (2021). Joint recording of EEG and audio signals in hyperscanning and pseudo-hyperscanning experiments. *MethodsX* 8, 101347. doi: 10.1016/j.mex.2021.101347

Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C., and Hansen, L. K. (2017). EEG in the classroom: synchronised neural recordings during video presentation. *Sci. Rep.* 7, 43916. doi: 10.1038/srep43916

Raghavan, A., Wilson, A., Wend, C., Alexander, A., Habela, C., and Nauen, D. (2018). Open-source system for millisecond-synchronized continuous video-EEG. *Epilepsy Res.* 145, 27–30. doi: 10.1016/j.eplepsyres.2018.05.010

Reis,:M. R., Hebenstreit, F., Gabsteiger, F., von Tscharner, V., and Lochmann, M. (2014). Methodological aspects of EEG and body dynamics measurements during motion. *Front. Hum. Neurosci.* 8, 156. doi: 10.3389/fnhum.2014.00156

Rodgers, J. L., and Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *Am. Statistician* 42, 59. doi: 10.2307/2685263

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., and Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *J. Neural Eng.* 2019, ab260c. doi: 10.1088/1741-2552/ab260c

Shenoy, A., Krishna, V., and Carmena Jose, M. (2014). Combining decoder design and neural adaptation in brain-machine interfaces. *Neuron* 84, 665–680. doi: 10.1016/j.neuron.2014.08.038

van den Bosch, K., Schoonderwoerd, T., Blankendaal, R., and Neerincx, M. (2019). Six challenges for human-ai co-learning. *Adapt. Instruct. Syst.* 45, 572–589. doi: 10.1007/978-3-030-22341-0_45

Wang, Y., Chen, X., Gao, X., and Gao, S. (2017). A benchmark dataset for SSVEP-based brain–computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1746–1752. doi: 10.1109/TNSRE.2016.2627556

Wang, Y., and Jung, T. P. (2011). A collaborative brain-computer interface for improving human performance. *PLoS ONE* 6, e20422. doi: 10.1371/journal.pone.0020422

Wang, Y., Markham, C., and Deegan, C. (2019). "Assessing the time synchronisation of EEG systems," in *2019 30th Irish Signals and Systems Conference (ISSC)* (Maynooth). doi: 10.1109/ISSC.2019.8904947

Wong, C. M., Wang, B., Wang, Z., Lao, K. F., Rosa, A., and Wan, F. (2020). Spatial filtering in SSVEP-based BCIs: unified framework and new improvements. *IEEE Trans. Biomed. Eng.* 67, 3057–3072. doi: 10.1109/TBME.2020.2975552

Xu, M., Han, J., Wang, Y., Jung, T. P., and Ming, D. (2020). Implementing over 100 command codes for a high-speed hybrid brain-computer interface using concurrent P300 and SSVEP features. *IEEE Trans. Biomed. Eng.* 67, 3073–3082. doi: 10.1109/TBME.2020.2975614

Xu, M., He, F., Jung, T. P., Gu, X., and Ming, D. (2021). Current challenges for the practical application of electroencephalography-based brain–computer interfaces. *Engineering* 7, 1710–1712. doi: 10.1016/j.eng.2021.09.011

Xu, M., Xiao, X., Wang, Y., Qi, H., Jung, T. P., and Ming, D. (2018). A brain–computer interface based on miniature-event-related potentials induced by very small lateral visual stimuli. *IEEE Trans. Biomed. Eng.* 65, 1166–1175. doi: 10.1109/TBME.2018.2799661

Zhang, S., Yoshida, W., Mano, H., Yanagisawa, T., Mancini, F., Shibata, K., et al. (2020). Pain control by co-adaptive learning in a brain-machine interface. *Curr. Biol.* 30, 66. doi: 10.1016/j.cub.2020.07.066

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 29 August 2023
DOI 10.3389/fnins.2023.1180471

OPEN ACCESS

EDITED BY
Josef Faller,
Optios, United States

REVIEWED BY
Leonhard Schreiner,
G.Tec Medical Engineering GmbH, Austria
Andreea Ioana Sburlea,
University of Groningen, Netherlands

*CORRESPONDENCE
Kun Wang
✉ flora_wk@tju.edu.cn

†These authors have contributed equally to this work

RECEIVED 06 March 2023
ACCEPTED 26 July 2023
PUBLISHED 29 August 2023

CITATION
Liu C, You J, Wang K, Zhang S, Huang Y, Xu M and Ming D (2023) Decoding the EEG patterns induced by sequential finger movement for brain-computer interfaces. *Front. Neurosci.* 17:1180471. doi: 10.3389/fnins.2023.1180471

Decoding the EEG patterns induced by sequential finger movement for brain-computer interfaces

Chang Liu[1†], Jia You[2†], Kun Wang[1*], Shanshan Zhang[2], Yining Huang[1], Minpeng Xu[1,2,3] and Dong Ming[1,2]

[1]Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin, China, [2]School of Precision Instruments and Optoelectronics Engineering, Tianjin University, Tianjin, China, [3]International School for Optoelectronic Engineering, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China

Objective: In recent years, motor imagery-based brain–computer interfaces (MI-BCIs) have developed rapidly due to their great potential in neurological rehabilitation. However, the controllable instruction set limits its application in daily life. To extend the instruction set, we proposed a novel movement-intention encoding paradigm based on sequential finger movement.

Approach: Ten subjects participated in the offline experiment. During the experiment, they were required to press a key sequentially [i.e., Left→Left (LL), Right→Right (RR), Left→Right (LR), and Right→Left (RL)] using the left or right index finger at about 1 s intervals under an auditory prompt of 1 Hz. The movement-related cortical potential (MRCP) and event-related desynchronization (ERD) features were used to investigate the electroencephalography (EEG) variation induced by the sequential finger movement tasks. Twelve subjects participated in an online experiment to verify the feasibility of the proposed paradigm.

Main results: As a result, both the MRCP and ERD features showed the specific temporal–spatial EEG patterns of different sequential finger movement tasks. For the offline experiment, the average classification accuracy of the four tasks was 71.69%, with the highest accuracy of 79.26%. For the online experiment, the average accuracies were 83.33% and 82.71% for LL-versus-RR and LR-versus-RL, respectively.

Significance: This paper demonstrated the feasibility of the proposed sequential finger movement paradigm through offline and online experiments. This study would be helpful for optimizing the encoding method of motor-related EEG information and providing a promising approach to extending the instruction set of the movement intention-based BCIs.

KEYWORDS

electroencephalography, sequential finger movements, movement related cortical potentials, event-related desynchronization, brain-computer interface

1. Introduction

Brain-computer interfaces (BCIs) are the direct communication pathways through which users can interact with the external world utilizing brain activities (Wolpaw et al., 2002; Chaudhary et al., 2016; Coogan and He, 2018; Xu et al., 2021). Over the last few decades, advances in disciplines such as neuroscience and engineering have introduced the BCI as a promising tool

for neurorehabilitation and neurophysiology research (Robinson and Vinod, 2016). The BCIs based on decoding motor-related neural activities can be used to drive functional electrical stimulation, intelligent prosthesis, or mechanical exoskeletons, which have important research value for the rehabilitation, replacement, and enhancement of motor functions (Birbaumer et al., 2008; Sebastian-Romagosa et al., 2021; Hernandez-Rojas et al., 2022; Ju et al., 2022; Wang et al., 2022). The motor-related neural activity of the brain can be induced by actual movement or motor imagery (MI). In the existing research, MI-based BCI is the most commonly used research paradigm (Pfurtscheller and Neuper, 2001; Wolpaw et al., 2002).

Currently, electroencephalography (EEG) has become the most widely used monitoring means of BCI due to its non-invasiveness, relatively low cost, and high time resolution (Park et al., 2012; Xu et al., 2018; Meng et al., 2020; He et al., 2022). Movement-related cortical potentials (MRCP) and sensorimotor oscillatory EEG activity (event-related desynchronization/synchronization—ERD/ERS) provide complementary information on the associated motor activity (Savić et al., 2020). Many studies have focused on detecting the pre-motor state of the upper limbs using EEG correlates such as MRCP or ERD (Sburlea et al., 2015). ERD/ERS is a particular time-locked EEG feature for MI tasks, which represents decreases and increases of power in alpha or beta bands. The alpha and beta frequency bands of ERD can be found over the corresponding sensorimotor areas of the brain when people imagine the movements of their limbs (Kosei et al., 2014; Peng et al., 2020; Dai et al., 2022). Jackson et al. found that motor execution shared similar mechanisms with MI. Motor execution can also induce the ERD/ERS features as MI tasks. In addition, movement-related cortical potentials (MRCPs) can be found during the processes of movement. MRCP is one kind of event-related potential (ERP), which is a time and phase-locked feature. Actual movement can evoke more significant MRCP features than MI tasks. Based on the similarity of neural oscillatory patterns of MI and motor execution, we could develop new paradigms and algorithms for movement-intention decoding through actual movement experiments (Jackson et al., 2003; Sochůrková et al., 2006; Katsumata et al., 2009; Sandhya et al., 2014).

Great progress has been made with the MI-BCI technique in recent years, but it still faces many research challenges. The quantity and classification accuracy of controllable instruction sets cannot meet the needs of users to complete most daily life actions (Qiu et al., 2021). So far, most studies have involved only four simple body MI tasks (i.e., left hand, right hand, foot, and tongue movements), with limited alternative paradigms (Townsend et al., 2006; Yang et al., 2009; Xygonakis et al., 2018). To solve the limitations of the instruction sets of MI-BCI, there have been studies on the decoding of complex limb and sequential limb-movement imagination tasks (Zhou et al., 2010; Doud et al., 2011; Yi et al., 2013). Hsu et al. designed a MI normal form of left and right leg steps and proposed a filter bank common space pattern (FBCSP) combined with fuzzy support vector machine type-II method, which achieved 86.25% recognition accuracy on eight subjects (Hsu et al., 2017). However, the existing MI tasks not only increase the operational complexity of the experiment but also make the output time of a single instruction longer, which reduces the decoding efficiency to a certain extent. Therefore, it is necessary to propose a new movement intention encoding paradigm to shorten the time of single instructions and ensure good classification performance at the same time.

As mentioned above, both ERD and MRCP are time-locked features. In addition, they have specific spatial distribution patterns for different limb movements or imagination tasks. Hence, the sequential limb movement paradigms can effectively combine the time-frequency and spatial domains' movement-related information, which are promising methods to extend the BCI instruction set and enhance the specificity of different task-induced EEG features. Yi et al. designed a sequential compound limb MI paradigm with a mean classification accuracy of 74.14%, while the time of one trial was 6 s (Weibo et al., 2016). Many studies have analyzed the brain activation mechanism of imagining movements of a single limb sequence. It has been found that the effect of learning movement sequences by imagining movements is similar to that of performing the same movement sequences, and the changes in brain activity between the two are consistent (Zhang et al., 2011; Hardwick et al., 2018; Wang et al., 2019; Zhang Q. et al., 2019). Recently, we investigated how data length affected the classification of repeated keystroke tasks with the index finger and found that single-trial EEG induced by the repeated finger movements had good separability (Zhang S, et al., 2019).

Therefore, we proposed a sequential finger movement paradigm for BCI, which was expected to expand the instruction set and shorten the time of single instructions. From the perspective of the time-frequency-spatial domain, this paper analyzed the neural oscillations patterns induced by sequential movement tasks. MRCP and ERD features were extracted effectively based on the common spatial filtering algorithm, such as discriminative canonical pattern matching (DCPM) (Xu et al., 2018; Wang et al., 2020) and filter bank common spatial pattern (FBCSP) (Chin et al., 2008; Ang et al., 2011; Sun et al., 2022). Mutual information analysis was used to select features. Both an offline and an online experiment were carried out to verify the feasibility of the proposed paradigm.

2. Materials and methods

2.1. Participants

A total of twenty-two subjects (eight males and fourteen females, aged 22–24 years old, all right-handed) participated in the experiments of this study. Among them, ten healthy subjects participated in the offline experiment to analyze the EEG features of sequential finger movement, and twelve subjects participated in the online experiment to evaluate the effectiveness of the proposed paradigm. None of the subjects had a history of neurological disease or movement disorders. The subjects were informed of the experimental procedure and received a letter of acceptance before the study. The study was approved by the ethical committee of Tianjin University.

2.2. Design of the experimental paradigm

2.2.1. Offline experiment

During the experiments, the subjects sat quietly in front of a monitor. Their arms were flat on the table and their left and right index fingers were on the keyboard "Z" and "1," respectively. The display background color was gray to avoid visual stimulation caused by a screen refresh. We tried to make the prompts as small as possible to help subjects focus on the middle of the screen, thus minimizing the eye movement artifacts of

subjects during the recording. Before the formal experiments, the subjects were required to practice pressing keys at one-second intervals under an auditory prompt of 1 Hz. The 1 Hz-auditory cues were always present in formal experiments as background sounds.

The flow chart of a single experiment trial is shown in Figure 1A. At the beginning of each trial, a white circle appeared in the center of the screen for 1 s to inform the subject that the trial was about to start. After the white circle disappeared, a text prompt appeared. Participants were asked to press the key using the corresponding [i.e., Left→Left (LL), Right→Right (RR), Left→Right (LR), or Right→Left (RL)] index finger in their own time. Subjects were not required to respond immediately to the text prompt. They could decide when to press the button for the first time. For example, if 'Right → Left' appeared, the subjects were reminded to press the right-hand key first, and then press the left-hand index finger after an interval of 1 s. There was a 2 s rest period after the subjects completed the keystrokes. During this time, the text prompt remained unchanged. Each participant performed 10 blocks of experiments and each session included 60 trials. Each sequential finger movement task occurred 15 times at random. For each subject, a total of 600 trials (150 trials for each task) were recorded. Trials with wrong key presses or key presses separated by more than 2 s were discarded.

2.2.2. Online experiment

To verify the feasibility of the left and right sequential finger movement paradigm, we performed the online experiment. The online experiment consisted of eight blocks. The procedure of blocks 1 ~ 6 was the same as that of the offline experiment. For each subject, a total of 360 trials (90 trials for each task) were recorded. Two two-class classifiers, i.e., LL-versus-RR and LR-versus-RL, were built using the data from the blocks 1 ~ 6. Session 7 contained 40 trials (20 trials of LL and 20 trials of RR, randomly sorted). The timeline of one trial of session 7 was the same as the offline experiment. During session 7, voice feedback containing the classification result was presented to the subject after the second keystroke of each trial. Session 8 also included 40 trials and had voice feedback following finger movement during each trial. Compared with session 7, session 8 performed 20 trials of LR and 20 trials of RL randomly.

The SVM classifier used in the online experiments was obtained based on offline data training. During the online experiments, each data segment was input to the SVM classifier for classification after pre-processing to extract feature values. The output of the SVM classifier included the predicted category label and its probability score, and we directly used the predicted label as the result output to control the corresponding speech feedback.

2.3. Signal recording

In this study, we used a Neuroscan SynAmps2 amplifier to obtain the original EEG signal. The EEG acquisition and amplification device used in this study, manufactured by Compumedics Neuroscan, included a 64-lead EEG cap, a SynAmps2 amplifier, and scan 4.5 software. Sixty Ag/AgCl scalp electrodes were placed according to the international 10–20 system (Figure 1B). The acquisition system referenced the data to the nose, and the prefrontal lobe served as ground. Some skin preparation was required before measurement. If there was dirt or excessive hair on the skin where the electrode was to be placed, the skin should be pre-cleaned or shaved. The sampling rate of EEG signal was 1,000 Hz and the notch filter of 50 Hz was used to eliminate the power frequency interference. We fully checked for bad channels and bad trials (incorrect keystrokes and keystrokes with more than 2 s between them). Bad channels and bad trials were removed if they existed.

2.4. Data processing and analysis

Independent component analysis (ICA) is a common blind source separation method in the case of multiple source signals and unknown transmission channel parameters. It functions by observing the signal to estimate the source signal, so as to recover the source signal. Observed signal X (t) = {x1 (t), x2 (t), …, xn (t)} by the source signal S (t) = {s1 (t), s2 (t), …, sn (t)} is obtained by an unknown mixed matrix A, namely, X= AS.ICA is to solve the mixing matrix W when S and A are unknown. At the same time, the estimate Y of the source signal S is separated from the observed signal by W. The prerequisite for ICA is that the number of observed signals is not less than the number of source signals (Song et al., 2022). In this experiment, the influence of eye movement can be seen according to the EMG signal. Therefore, we conducted the ICA process. We used EEGLAB to perform ICA processing on the EEG data and eliminate various artifacts, such as eye movements and blinking. We chose the Runica algorithm for ICA processing. In addition, each subject's data underwent different bad segment removal operations. To ensure the validity of the ICA processing, we visually inspected each subject's data and determined the components to be removed based on the EEG waveform and timeline. The number of components removed for each subject was not fixed, but generally ranged from 5 to 10.

In this study, we mainly analyzed MRCPs and ERD features to compare the differences among four different sequence movement-induced patterns of the offline experiment. Since the MRCP potential is a low frequency time-domain waveform signal, we down-sampled the raw EEG data to 16 Hz. Then the data were low-pass filtered at 0–3 Hz using a 4th-order zero-phase Butterworth filter to preserve the low-frequency components of the EEG signal. Common average reference (CAR) was used to improve the signal-to-noise ratio. In this study, we defined the moment of the first keystroke as 0 s and epoched the data from-2 s ~ 3 s for MRCP analysis. The paired t-test was used for statistical analysis of whole subjects between two different sequential finger movement tasks across all time points. To observe the spatial patterns of the four tasks, we calculated the averaged amplitude of all subjects on-150 ms and 850 ms at each channel and plotted the mean topographical distribution across all subjects.

For the ERD analysis, the original signal was down-sampled to 200 Hz and CAR was also applied to it. Then, the signal was bandpass filtered to 4 ~ 30 Hz. The short-time Fourier transform (STFT) of the Hanning window, which has 256 sampling points, was used to calculate the event-related spectral perturbation ERSP between the time range of −1.5 s to 2.5 s for each movement task. We also defined the moment of the first keystroke as 0 and the baseline was the mean of the data ranging from −1.5 s to −1 s. We used the mean ERSP values of all subjects from electrodes C3, Cz, and C4 to compare the time-frequency variation among the four sequential finger movement tasks. Additionally, the averaged alpha band ERSP values of each keystroke in one trial were calculated to analyze the topographical distribution of ERD features. The calculation formula is as follows:

$$ERSP = ERSP_{\text{original}} - ERSP_{\text{baseline}}$$

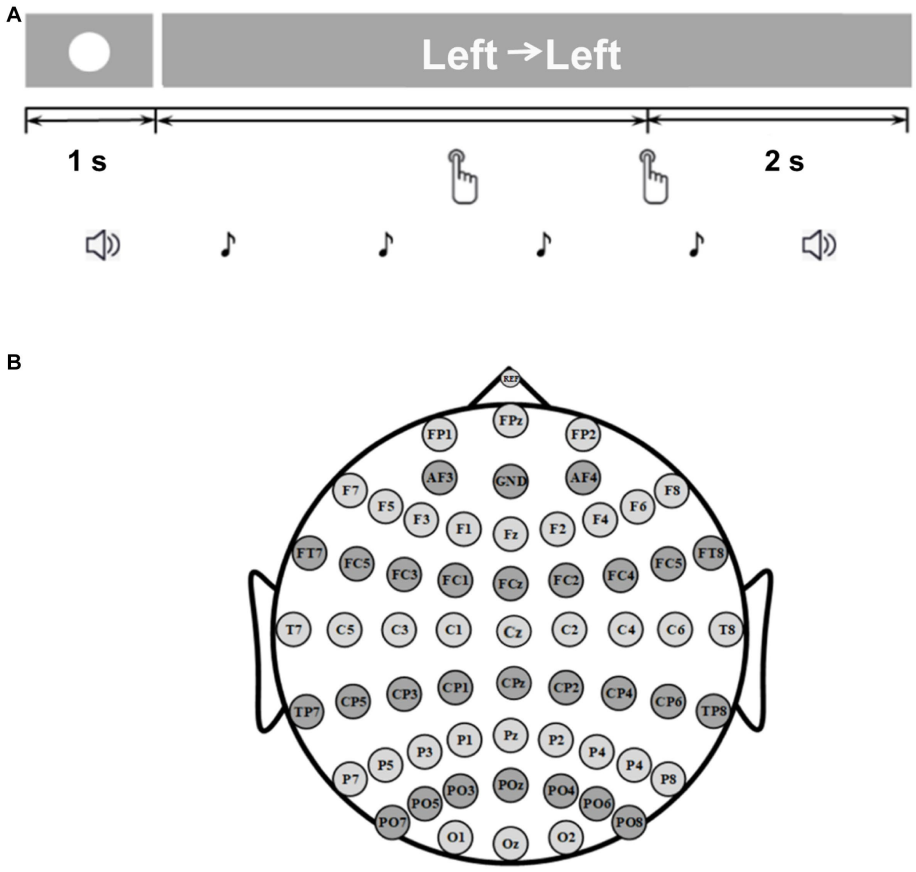


FIGURE 1
(A) The timeline of one trial of the experimental paradigm. **(B)** Locations of the electrodes.

To obtain a higher classification accuracy for single EEG recognition, we needed to utilize some spatial filtering methods to extract both MRCP and ERD features induced by the sequential finger movements. MRCP is a low-frequency waveform feature. Our previous work showed discriminative canonical pattern matching (DCPM) has superiority for MRCP feature extraction. DCPM consists of three major parts: (1) the construction of discriminative spatial patterns (DSPs); (2) the construction of CCA patterns; and (3) pattern matching (Xu et al., 2018). Canonical correlation analysis (CCA) is a multivariate statistical analysis method that uses the correlation between synthetic variable pairs to reflect the overall correlation between two groups of indicators. The CCA algorithm can be used to project the spatially filtered data into a new space and calculate the correlation to reflect the overall correlation of the two groups of indicators (Ma et al., 2022). In addition, other effective feature extraction methods for low-frequency waveform features should also be investigated, such as task-related component analysis (TRCA) (Birbaumer et al., 2008; Nakanishi et al., 2018; Sun et al., 2021), a spatial filtering method for task-dependent component analysis, where the weight coefficients are optimized to maximize the inter-trial covariance of time-locked data. The goal of TRCA is to take task-related constituent parts out from multiple time series that are linearly weighted (Tanaka et al., 2013). For ERD patterns, the filter bank common spatial pattern (FBCSP) was intended to independently select the appropriate frequency bands for feature extracting, which is a popular and effective method (Ang et al., 2012; Chu et al., 2021). The FBCSP method is the optimization of classical spatial filtering in the frequency domain. The effects of different feature selection methods are studied, and the best individual features based on mutual information are used to obtain the selection method with relatively higher classification accuracy (Yong and Wonz, 2019). Hence, we used DCPM and TRCA to extract the MRCP features and used FBCSP to extract the ERD features. Then, we selected the features based on mutual information (Zhang S. et al., 2019).

Before feature extraction, we down-sampled the raw data to 200 Hz first. For each keystroke, we epoched data from 0.5 s ahead of the key stroke and 1.5 s after the key stroke to process, i.e., −0.5 ~ 1.5 s. Different band-pass filters were used for MRCPs and ERD characterization. For MRCPs, we used a band-pass filter (1 to 8 Hz) to filter the data, and then used DCPM and TRCA, respectively, to extract the features. For the ERD features extraction, three crucial characteristic frequency bands, 4 ~ 8 Hz (theta band), 8 ~ 13 Hz (alpha band), and 13 ~ 30 Hz (beta band), were selected for band-pass filtering, and CSP features were extracted, respectively. As the eigenvectors of spatial filters are in descending orders, we selected the first two dimensions for DCPM and the first three dimensions for TRCA and FBCSP. After spatial filtering, we obtained 56 features (16 of DCPM, 4 of TRCA, and 36 of FBCSP) for each trial. In the FBCSP

method, the three-dimensional eigenvectors of each CSP filter were selected in three frequency bands for CSP spatial filtering. Therefore, the FBCSP characteristic dimension of the four classifications was 4*3*3 = 36. Furthermore, in order to reduce the characteristic dimension, the mutual information between features and labels was calculated, and pattern recognition was carried out by combining the optimal selection features. The data from the training set was used for feature selection (Aldehim and Wang, 2015). A linear support vector machine (SVM) was used to build the classifier with the help of the famous software package LIBSVM (Yang et al., 2015; Bhatnagar et al., 2016; Dhiman and Priyanka Saini, 2018). We selected the default SVM type and set the penalty factor C to 1. For the offline experiment, we used 10-fold cross-validation to calculate the classification accuracies. For the online experiment, the DCPM, TRCA, and CSP spatial filters were established using EEG data from blocks 1 to 6. We selected 10 features using the mutual information analysis for each subject. Then the online linear SVM classification models were built. All programs were compiled and run on the MATLAB (Matlab used the 2022Rb version of MathWorks) platform. The LL-versus-RR and LR-versus-RL classifiers were applied in the online experiment, respectively. During the online experiment, the EEG data was continuously transmitted to the MATLAB data processing module in real time. The program continuously detected the labels, and then analyzed and processed the data according to the labels. We provided visual feedback during the 2 s break after the second keystroke of each trial, which allowed participants to receive immediate information on their performance. The speech feedback was performed in each trial of the post-processing phase of the data processing program for each trial phase executed, which lasted 100 ms. Finally, the recognition results were fed back to the subjects through voice feedback.

3. Results

3.1. EEG patterns of sequential finger movement

We first analyzed the MRCP and ERD patterns induced by sequential finger movement from the offline experiment. The top of Figure 2 shows the average waveforms of MRCPs across all participants of four sequential finger movements at channels C3 and C4. It is obvious that the potentials decreased before the movement onset for both the left and right finger movements, especially for the initial finger. We found that the initial tasks with the left finger, i.e., LL and LR, induced more negative potential on channel C4. On the contrary, right-hand initial finger movement tasks induced more negative potential on C3. This phenomenon also coincided with the contralateral activation of the cortical activity in hand functional areas. For the second sub-action, only the LR and RL tasks showed similar MRCP patterns. In addition, the negative potential peak of the initial keystroke action was lower than that of the non-initial keystroke action.

The bottom of the Figure 2 shows the topography of the average MRCP at −150 ms and 850 ms of the 60 channels. We observed that the channels with the negative waveforms were distributed over the primary motor area and the supplementary motor area. The phenomenon of contralateral dominance could be clearly observed from the topography. MRCP-related negativity induced by the different sub-action tasks (LR and RL) was more pronounced than the repeated sub-action tasks (LL and RR). For the LR and RL tasks, there were completely opposite spatial distributions at −150 ms and 850 ms. Thus, the time-spatial differences of MRCP could be used for classification.

To further investigate the differences between the different sub-action tasks (LR and RL) and the repeated sub-action tasks (LL and RR), the average MRCP potentials between the different sub-actions and the repeated sub-action tasks were, respectively, drawn and analyzed using the paired t-test, as shown in Figure 3. The grey area is the time period with significant difference between the two types of sequential finger movements-induced potential amplitudes. As can be seen from the figure, when the initial sub-action was left finger movement, Bereitschaftspotential (BP) induced by LR and RL tasks were significantly more obvious than those induced by LL task on C3 and C4 channels. Similarly, it could be seen that BP amplitudes before LR and RL tasks were significantly larger than those before RR tasks except RR-versus-RL at C3. These results showed that, compared with simple sequential movement, complex sequential movement might induce stronger MRCP patterns.

Figure 4A shows the average time-frequency graph of four types of index finger sequence movements of 10 subjects in the offline experiment at key channels C3, Cz, and C4. In the figure, it can be observed that all the four movement tasks could induce obvious ERD phenomena in the theta, alpha, and beta bands. Notably, the intensity of ERD patterns varied over time. The ERD phenomenon in theta and alpha bands mainly occurred within 1 s before keystroke, which was of high intensity and involved a wide range. However, there was no significant difference among the three channels. For the repeated sub-action tasks (LR and RL), we found distinct contralateral hemispheric dominance, which was not obvious for the different sub-action tasks (LR and RL). Figure 4B shows the mean alpha band EEG power topography among four different finger movement tasks. It can be seen from the brain topographic map that the LL and RR tasks could activate the motor functional areas of both hands, which showed obvious contralateral dominance. In addition, the ERD intensity induced by the initial sub-action was greater than that of the second sub-action.

3.2. Classification performance of offline experiment

For the classification of the data in the offline experiment, the optimal filter dimension and characteristic dimension were selected. Figure 5A shows the classification accuracy results of the four sequential finger movement tasks. The mean classification accuracy of the four classes was 71.69%, which was much higher than the random level of 25%. The highest accuracy was 79.26% and the lowest accuracy was 54.91%.

We also calculated the confusion matrix under the optimal feature dimension of the four categories, as shown in Figure 5B. Each row represents the true label and each column represents the output result. The figure shows the proportion of each type of task divided into four different results by the classifier. It shows that the percentage of classification errors varied from task to task. The distribution of the four action task features in the classifier was not irregular but a regular distribution in a certain projection direction. At the same time, it can

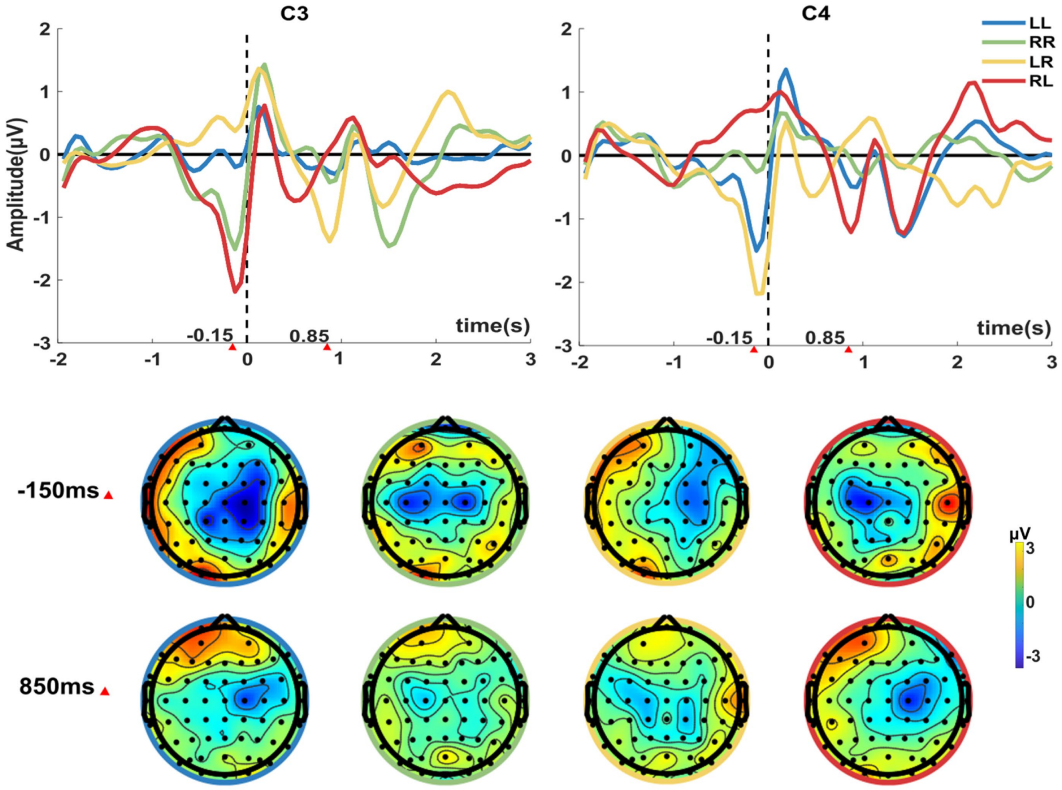


FIGURE 2
The average MRCPs across all participants of the four sequential finger movements displayed for channels C3 and C4. Time 0 corresponds to the time of the first keystroke. At the bottom of the figure is the spatial distribution of the average MRCP at −150 ms and 850 ms of the 60 channels corresponding to the four-movement tasks. LL (Left→Left), RR (Right→Right), LR (Left→Right), and RL (Right→Left) are used to depict the four tasks, respectively.

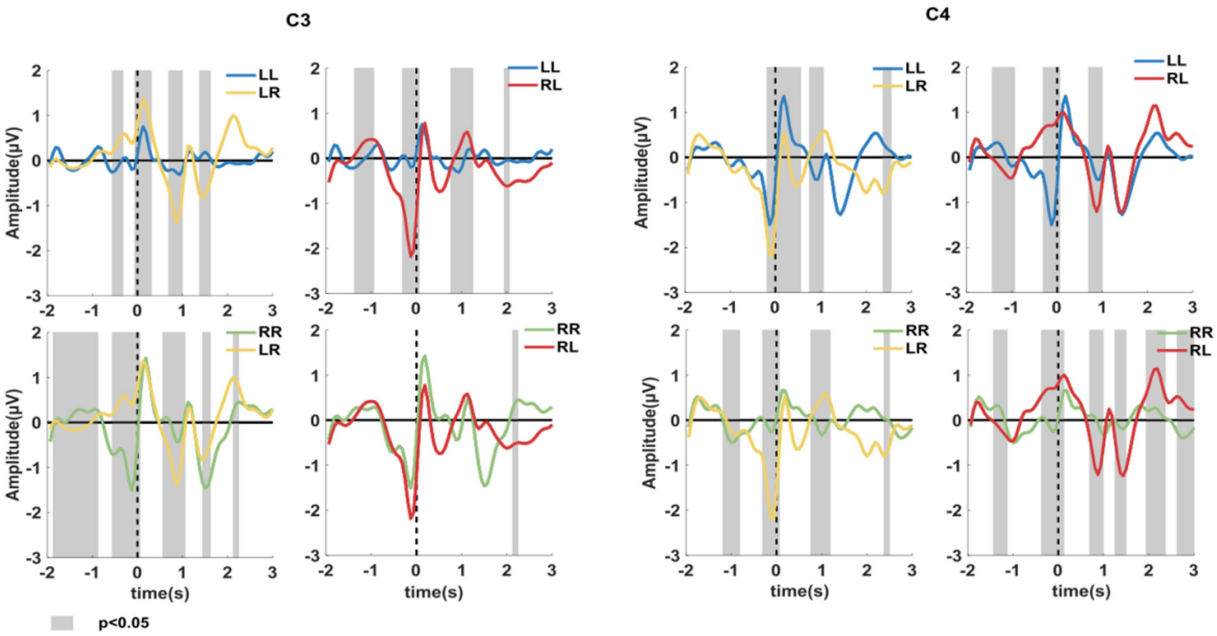


FIGURE 3
Average MRCP between the different sub-action tasks (LL and RR) and repeated sub-action tasks (LR and RL) at channel C3 and C4. The gray area is the time period with significant difference between the potential amplitudes induced by the two movement tasks ($p < 0.05$, paired t-test).

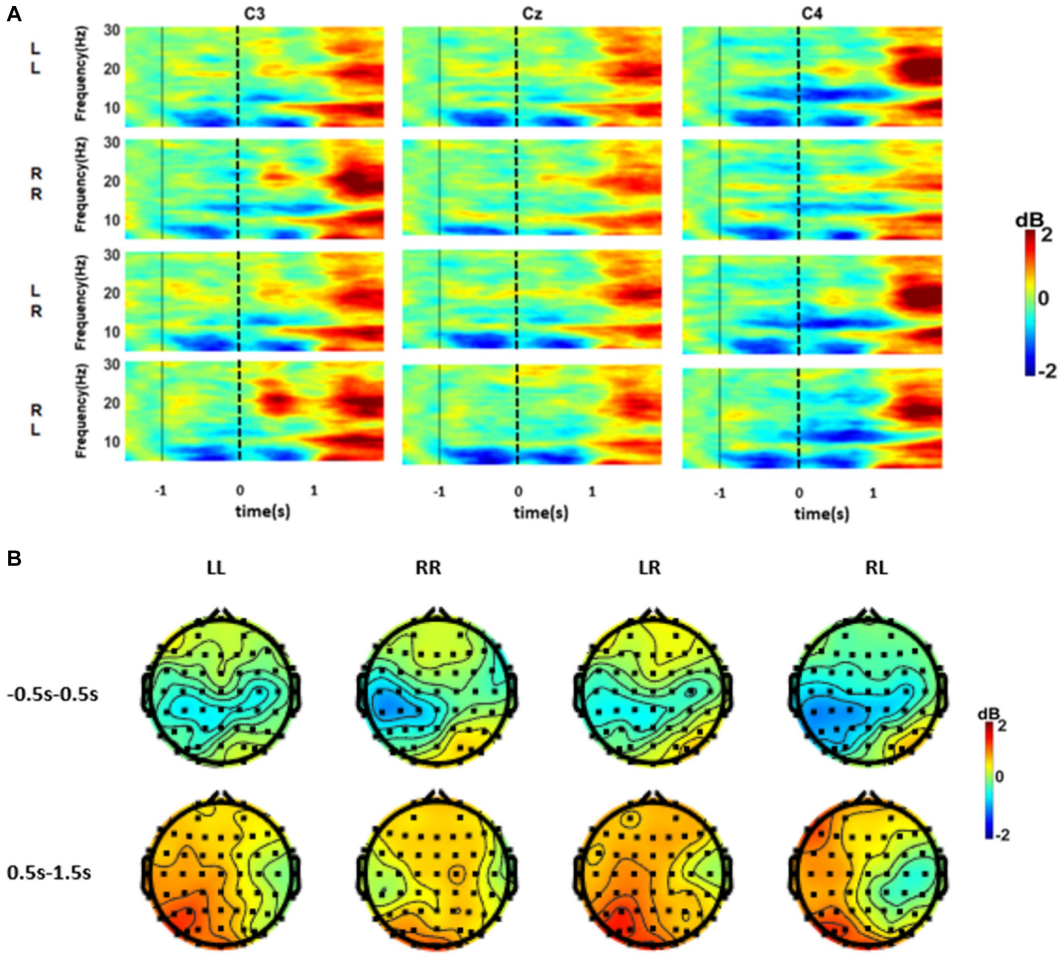


FIGURE 4
(A) Average time-frequency maps of the four finger movement tasks at channels C3, Cz, and C4. Time 0 corresponds to the time of the first keystroke. Blue is ERD, and red is ERS. **(B)** The average 8–13 Hz ERSP topography of the four movement tasks. Among them, −0.5 ~ 0.5 s and 0.5 ~ 1.5 s correspond to the first and second sub-actions, respectively.

be seen that the four types of sequential finger movement tasks had different classification difficulties. The two types of tasks with two different sub-actions were easier to distinguish than the sequences with the same sub-actions. When subjects performed RL or LR tasks, the false recognition of RR had a higher occurrence rate than the LL task.

3.3. Classification performance of online experiment

The online experimental results of 12 subjects are shown in Figure 6 and Table 1. The online recognition accuracies were 83.33 and 82.71% for LL-versus-RR and LR-versus-RL, respectively. The classification results of S4, S6, and S11 in the two types of online experiments all reached more than 90%, which proved the feasibility of the sequential finger movement paradigm proposed in this study. However, the classification accuracies of S2, S5, S7, S9, and S10 decreased significantly compared with the offline model. This was caused by the overfitting of the model. Due to the non-stationarity of MI-EEG signals, there may be significant differences in EEG features between the training and testing datasets. Therefore, a classification model constructed through the training set may not adapt well to the testing set.

4. Discussion

This paper explored expanding the instruction set for movement intention-related BCIs. This paper showed that the sequential movement of the left and right fingers could induce the distinguishable MRCP and ERD features containing time-frequency-spatial movement-related information. In our previous study, we combined the MRCP and ERD to decode the pre-movement EEG patterns of left and right finger movement and obtained a satisfactory performance (Wang et al., 2020). However, that study could only be used for binary classification. In this study, the cortical activation of the left and right index finger keystrokes were predominated contralateral, which was consistent with the results of some other studies (Zang et al., 2003; Francesco et al., 2005; Bian et al., 2022). In addition, we also found the EEG spatial patterns were rhythmically changed with the rhythmically sequential tasks (Figures 2, 4). Hence, sequential finger movement

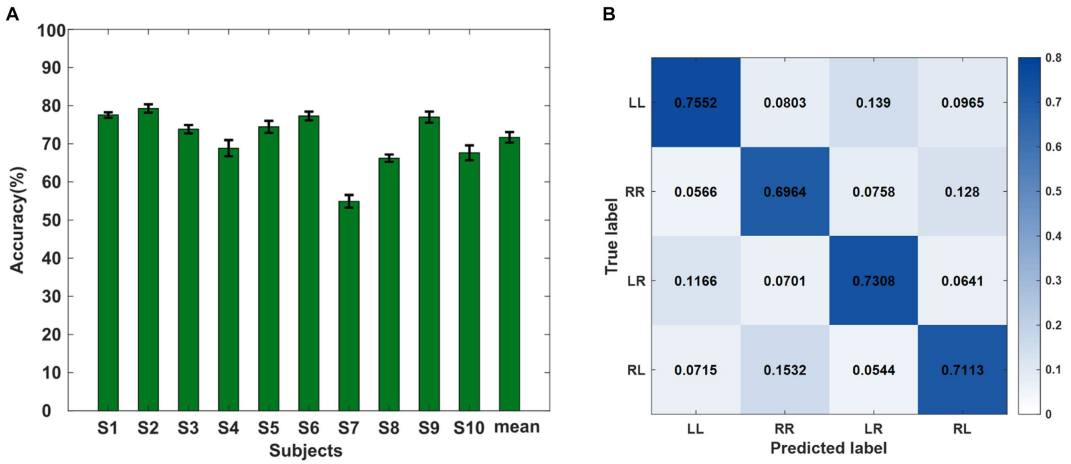


FIGURE 5
(A) Ten-fold classification accuracies (%) for all subjects of the offline experiment. **(B)** Confusion matrix of the average recognition accuracies of the four kinds of sequence finger movements of all subjects. Each row represents the true label and each column represents the predicted label.

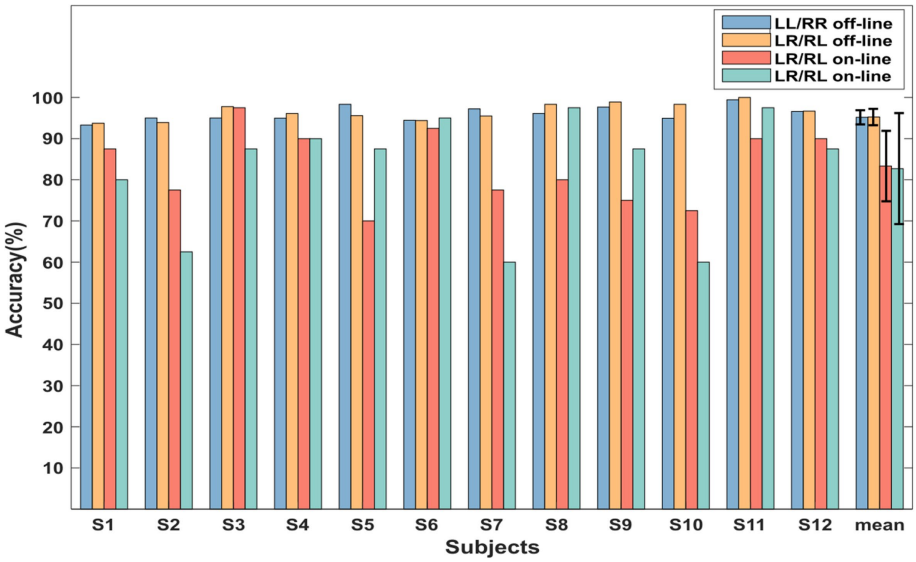


FIGURE 6
The offline and online experimental classification accuracies (%) of all subjects.

adds information encoding in the time domain, which is meaningful to extend the instruction of movement intentions.

In Figure 2, we can see the negative potential peak of the initial keystroke action was obviously lower than that of the non-initial keystroke action, especially for the repeated sub-action tasks (LL and RR). The reasons for this phenomenon might come from two aspects. On the one hand, we epoched all trials using the label of the first keystroke. Although we set a 1 Hz background sound cue during the experiment, the time between the second keystroke and the first keystroke had a certain error compared with 1 s. Hence, for the non-initial keystroke, the negative potential may not be so pronounced after calculating the mean wave because the data were not perfectly aligned. On the other hand, compared with the different sub-action tasks (LR and RL), the subjects were more familiar with the non-initial keystroke action due to it being the same as the initial action for the repeated sub-action tasks, which might have resulted in less activation of the brain cortex. Jancke et al. showed that repeated practice of an action has an effect on motor cortex activation, and familiar action-induced ERD features were reduced (Jancke et al., 2006). This phenomenon may be similar to the repetitive inhibitory effect of steady-state visual-evoked potentials (Xu et al., 2022).

In Figure 4, the onset time of ERD in the alpha band is earlier than that in the beta band. Some studies indicated that the amplitude of alpha-band oscillations significantly decreased over the motor regions that began in the motor preparation stage, which implied that the alpha rhythm was more relevant to motor planning/programming (Kajihara et al., 2015). The rhythmically sequential movement we proposed in this paper is the more

TABLE 1 Offline and online classification results of two types of keystrokes of 12 subjects (%).

Subject	L→L versus R→R		L→R versus R→L	
	Off-line	On-line	Off-line	On-line
S1	93.30	87.50	93.73	80.00
S2	95.00	77.50	93.89	62.50
S3	95.00	97.50	97.78	87.50
S4	94.97	90.00	96.11	90.00
S5	98.33	70.00	95.56	87.50
S6	94.44	92.50	94.38	95.00
S7	97.22	77.50	95.49	60.00
S8	96.11	80.00	98.33	97.50
S9	97.65	75.00	98.89	87.50
S10	94.97	72.50	98.33	60.00
S11	99.44	90.00	100.00	97.50
S12	96.60	90.00	96.67	87.50
Mean	95.17	83.33	95.24	82.71
Std.	1.72	8.56	1.99	13.48

sophisticated motor control. Hence, the cortex associated with motor planning and related advanced cognitive activities would be active before movement.

We conducted behavioral analysis. The time difference between the two keystrokes was analyzed statistically. The results showed that there was no significant difference between the two keystrokes under different tasks. The results are shown in Figure 7.

The classification results showed that the LL-versus-RR and LR-versus-RL are the two classification models with better performance. This is mainly due to the difference between the preceding sub-actions and the following sub-actions of the above two models. Secondly, we found that the classification models with the different initial sub-action (LL-versus-RL or RR-versus-LR) had better performance than that with different non-initial sub-actions (LL-versus-LR or RR-versus-RL). This indicated that the initial sub-action can provide more classification information than the non-initial sub-action in the sequential movement paradigm. In this paper, although the mean four-classification accuracy of the offline experiment has shown divisibility, it still does not meet the needs of external device control for everyday BCIs. Therefore, the online experiment mainly focused on the two best-performing binary models. The existing work is still in the preliminary stage. In the future, the classification algorithm should be further optimized and improved to realize the high classification accuracy meeting everyday BCI use with a large instruction set. Considering the similarity in EEG patterns between MI and real movement, how to transfer the sequential movement paradigm to sequential MI is also a problem worth exploring in the future.

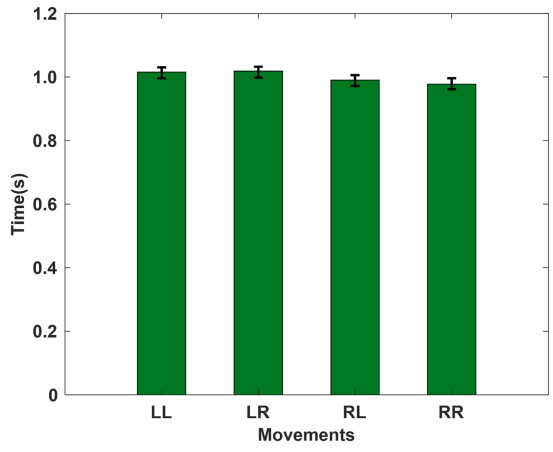


FIGURE 7
The average time difference between subjects' first and second keystrokes for each sequential action task.

The difference between the results of the online and offline experiments may be due to the non-linear and non-stationary characteristics of the EEG signals. The offline results were obtained by 10-fold cross validation calculation, and the training data were close to the test data so that it had a high similarity. In the online experiment, the training data used to build the classification model were completely separated from the test data. Consequently, the online experimental classification effect of subjects whose EEG signals changed greatly over time was poor.

The experimental paradigm we are using now is the motor execution of the subject performing a real action. Subsequently, we will use transfer learning to make subjects realize brain-computer interactive control through motor imagination. For the subsequent application of motor imagination, it can be used to help patients with rehabilitation, assistance, etc.

However, when we switch from motor execution to motor imagination, the EEG signal might be weakened. In addition, there could be some problems such as inaccurate time labels. Wu et al. studied the problems of applying transfer learning to brain-computer interfaces and how to solve them (Wu et al., 2022).

5. Conclusion

This paper demonstrated the feasibility of the proposed sequential finger movement paradigm, which had a satisfactory performance on recognition. The spatial distributions of both MRCPs and ERD were varied regularly with the different finger movements. In general, this study proposed a promising encoding method of movement intention to improve the discriminated information dimension of EEG patterns, which might provide a new idea and theoretical basis for effectively expanding the command set of movement intention-related BCIs.

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 human participants were reviewed and approved by the Research Ethics Committee of Tianjin University. The

patients/participants provided their written informed consent to participate in this study.

Author contributions

CL, JY, KW, MX, SZ, YH, and DM conceived the study. CL, JY, and KW designed and conducted the experiments. CL and SZ performed data analyses. CL and KW wrote and edited the initial draft. MX, KW, and DM performed proofreading and the finalizing of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the STI 2030—Major Projects 2022ZD0208900, National Natural Science Foundation of China (No. 62206198, 62122059, and 81925020), and Introduce Innovative Teams of 2021 "New High School 20 Items" Project (2021GXRC071).

Acknowledgments

The authors sincerely thank all participants for their voluntary participation.

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

Aldehim, G., and Wang, W. (2015). Determining appropriate approaches for using data in feature selection. *Int. J. Mach. Learn. Cybern.* 8, 915–928. doi: 10.1007/s13042-015-0469-8

Ang, K.K., Chin, Z.Y., Zhang, H., and Guan, C., (2011). "Filter Bank Common Spatial Pattern (FBCSP) Algorithm Using Online Adaptive and Semi-Supervised Learning," in *International Joint Conference on Neural Networks (IJCNN)*. IEEE; 392–396.

Ang, K. K., Chin, Z. Y., Wang, C., Guan, C., and Zhang, H. (2012). Filter Bank Common Spatial Pattern Algorithm on BCI Competition IV Datasets 2a and 2b. Frontiers in neuroscience.

Bhatnagar, V., Yede, N., Keram, R. S., and Chaurasiya, R. K. (2016). "A Modified Approach to Ensemble of SVM for P300 based Brain Computer Interface." in *International Conference on Advances in Human Machine Interaction (HMI).*

Bian, Y., Zhao, L., and Sun, Y. (2022). Research on feature modulation and classification performance of ASMI-BCI. *J Electr Meas Instrum* 36, 224–230. doi: 10.13382/j.jemi.B2104479

Birbaumer, N., Murguialday, A. R., and Cohen, L. (2008). Brain-computer interface in paralysis. *Curr. Opin. Neurol.* 21, 634–638. doi: 10.1097/WCO.0b013e328315ee2d

Chaudhary, U., Birbaumer, N., and Ramos-Murguialday, A. (2016). Brain-computer interfaces for communication and rehabilitation. *Nat. Rev. Neurol.* 12, 513–525. doi: 10.1038/nrneurol.2016.113

Chin, Z. Y., Ang, K. K., and Guan, C. (2008). Multiclass voluntary facial expression classification based on filter Bank common spatial pattern. *Ann Int Conf IEEE Eng Me Biol Soc* 2008, 1005–1008. doi: 10.1109/IEMBS.2008.4649325

Chu, Y., Zhu, B., Zhao, X., and Zhao, Y. (2021). Convolutional neural network based on temporal-spatial feature learning for motor imagery electroencephalogram signal decoding. *J. Bio. Eng.*

Coogan, C. G., and He, B. (2018). Brain-computer interface control in a virtual reality environment and applications for the internet of things. *IEEE Access* 6, 10840–10849. doi: 10.1109/ACCESS.2018.2809453

Dai, W. H., Chen, J., Xie, P., Li, G. Q., and Wang, D. P. (2022). Knowledge mapping analysis of motor related cortical potentials. *Chin. J. Biomed. Eng.* 41, 360–369. doi: 10.3969/j.issn.0258-8021.2022.03.012

Dhiman, R., and Priyanka Saini, J. S. (2018). "Motor imagery classification from human EEG signatures." *International Journal of Biomedical Engineering and Technology.*

Doud, A. J., Lucas, J. P., Pisansky, M. T., and He, B. (2011). Continuous three-dimensional control of a virtual helicopter using a motor imagery based brain-computer interface. *PLoS One* 6:e26322. doi: 10.1371/journal.pone.0026322

Francesco, D. R., Sabrina, P., Teresa, A., and Donatella, S. (2005). Effect of practice on brain activity: an investigation in top-level rifle shooters. *Med. Sci. Sports Exerc.* 37, 1586–1593. doi: 10.1249/01.mss.0000177458.71676.0d

Hardwick, R. M., Caspers, S., Eickhoff, S. B., and Swinnen, S. P. (2018). Neural correlates of action: comparing meta-analyses of imagery, observation, and execution. *Neurosci. Biobehav. Rev.* 94, 31–44. doi: 10.1016/j.neubiorev.2018.08.003

He, F., Dong, B., Han, J., Li, Y., Xu, M., and Ming, D. (2022). Advances in application of game brain-computer Interface based on ElectroEncephaloGram. *J Electr Inf Technol* 44, 415–423. doi: 10.11999/JEIT211337

Hernandez-Rojas, L. G., Cantillo-Negrete, J., Mendoza-Montoya, O., Carino-Escobar, R. I., Leyva-Martinez, I., Aguirre-Guemez, A. V., et al. (2022). Brain-computer Interface controlled functional electrical stimulation: evaluation with healthy subjects and spinal cord injury patients. *IEEE Access* 10, 46834–46852. doi: 10.1109/access.2022.3170906

Hsu, W. C., Lin, L. F., Chou, C. W., Hsiao, Y. T., and Liu, Y. H. (2017). EEG classification of imaginary lower limb stepping movements based on fuzzy support vector machine with kernel-induced membership function. *Int J Fuzzy Syst* 19, 566–579. doi: 10.1007/s40815-016-0259-9

Jackson, P. L., Lafleur, M. F., Malouin, F., Richards, C. L., and Doyon, J. (2003). Functional cerebral reorganization following motor sequence learning through mental practice with motor imagery. *NeuroImage* 20, 1171–1180. doi: 10.1016/s1053-8119(03)00369-0

Jancke, L., Lutz, K., and Koeneke, S. (2006). Converging evidence of ERD/ERS and BOLD responses in motor control research. *Prog. Brain Res.* 159, 261–271. doi: 10.1016/S0079-6123(06)59018-1

Ju, J., Feleke, A. G., Luo, L., and Fan, X. (2022). Recognition of drivers' hard and soft braking intentions based on hybrid brain-computer interfaces. *Cyborg Bionic Syst* 2022:2022. doi: 10.34133/2022/9847652

Kajihara, T., Anwar, M. N., Kawasaki, M., Mizuno, Y., Nakazawa, K., and Kitajo, K. (2015). Neural dynamics in motor preparation: From phase-mediated global computation to amplitude-mediated local computation. *Neu.* doi: 10.1016/j.neuroimage.2015.05.032

Katsumata, H., Suzuki, K., Tanaka, T., and Imanaka, K. (2009). The involvement of cognitive processing in a perceptual-motor process examined with EEG time-frequency analysis. *Clin. Neurophysiol.* 120, 484–496. doi: 10.1016/j.clinph.2008.11.024

Kosei, N., Midori, S., Yohei, T., Yoshikatsu, H., and Toshiyuki, K. (2014). Modulation of event-related desynchronization during kinematic and kinetic hand movements. *J. Neuroeng. Rehabil.* 11:90. doi: 10.1186/1743-0003-11-90

Ma, P., Dong, C., Lin, R., Ma, S., Jia, T., Chen, X., et al. (2022). A classification algorithm of an SSVEP brain-computer interface based on CCA fusion wavelet coefficients [J]. *J. Neurosci. Methods* 371:109502. doi: 10.1016/j.jneumeth.2022.109502

Meng, J., Xu, M., Wang, K., Meng, Q., Han, J., Xiao, X., et al. (2020). Separable EEG features induced by timing prediction for active brain-computer interfaces. *Sensors (Basel)* 20:3588. doi: 10.3390/s20123588

Nakanishi, M., Wang, Y. T., and Jung, T. P. (2018). Transferring shared responses across electrode montages for facilitating calibration in high-speed brain spellers. *Annu. Int. Conf. IEEE. Eng. Med. Biol. Soc.* 2018, 89–92. doi: 10.1109/EMBC.2018.8512269

Park, C., Looney, D., Ur Rehman, N., Ahrabian, A., and Mandic, D. P. (2012). Classification of motor imagery BCI using multivariate empirical mode decomposition. *IEEE Trans Neural Syst Rehabil Eng* 21, 10–22. doi: 10.1109/TNSRE.2012.2229296

Peng, Z., Yue, Z., Xiao, D., Xuejun, J., Linying, Z., Hongzhi, Q., et al. (2020). Effect of biological rhythm disorder lnduced by sleep restriction on brain alertness. *Space Med. Med. Eng.* 33, 504–511. doi: 10.16289/j.cnki.1002-0837.2020.06.006

Pfurtscheller, G., and Neuper, C. (2001). Motor imagery and direct brain-computer communication. *Proce IEEE* 89, 1123–1134. doi: 10.1109/5.939829

Qiu, S., Zhang, Y., Wu, C., Ma, X., Wei, W., and He, H. (2021). Research progress of brain-computer interface technology based on fine motor imaging. *Artif. Intell.* 6, 40–50. doi: 10.16453/j.cnki.ISSN2096-5036.2021.06.005

Robinson, N., and Vinod, A. P. (2016). Noninvasive brain-computer Interface: decoding arm movement kinematics and motor control. *IEEE Systems, Man, and Cybernetics Magazine* 2, 4–16. doi: 10.1109/msmc.2016.2576638

Sandhya, B., Shendkar, C., and Mahadevappa, M., (2014). "Single Channel Event Related (De) Synchronization (ERD/ERS) Analysis of Motor Execution in Stroke Affected Foot Drop Subjects", in: International Conference on Medical Imaging, M-Health and Emerging Communication Systems (Med Com), 325–328.

Savić, A. M., Lontis, E. R., Natalie, M.-K., and Popović, M. B. (2020). Dynamics of movement-related cortical potentials and sensorimotor oscillations during palmar grasp movements. *Eur. J. Neurosci.* 51, 1962–1970. doi: 10.1111/ejn.14629

Sburlea, A. I., Montesano, L., and Minguez, J. (2015). Continuous detection of the self-initiated walking pre-movement state from EEG correlates without session-to-session recalibration. *J. Neural Eng.* 12:036007. doi: 10.1088/1741-2560/12/3/036007

Sebastian-Romagosa, M., Cho, W., Ortner, R., and Guger, C. (2021). "Brain Computer Interface Treatment for Gait Rehabilitation of Stroke Patients – Preliminary Results", in: 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC).

Sochůrková, D. M., Rektor, I., Jurák, P., and Stancák, A. (2006). Intracerebral recording of cortical activity related to self-paced voluntary movements: a Bereitschaftspotential and event-related desynchronization/synchronization. *SEEG study.* 173, 637–649. doi: 10.1007/s00221-006-0407-9

Song, H., Song, H., Xu, S., Liu, G., Liu, J., and Xiog, P. (2022). Automatic removal algorithm of electrooculographic artifacts in non-invasive brain-computer interface based on independent component analysis. *J. Biomed. Eng.* 39, 1074–1081. doi: 10.7507/1001-5515.202111060

Sun, J., Jung, T. P., Xiao, X., Meng, J., Xu, M., and Ming, D. (2021). Classification algorithms of error-related potentials in brain-computer interface. *J. Biomed. Eng.* 38, 463–472. doi: 10.7507/1001-5515.202012013

Sun, Y.F., Zhang, H.L., Lu, Y.F., and Xue, Y. (2022). "EEG Signal Classification Using Shallow FBCSP Conv Net with a New Cropping Strategy", in: 15th International Conference on Brain Informatics (BI), 359–368.

Tanaka, H, Katura, T, and Sato, H. (2013). *Task-related component analysis for functional neuroimaging and application to near-infrared spectroscopy data[J]*. NeuroImage.

Townsend, G., Graimann, B., and Pfurtscheller, G. (2006). A comparison of common spatial patterns with complex band power features in a four-class BCI experiment. *IEEE Trans. Biomed. Eng.* 53, 642–651. doi: 10.1109/tbme.2006.870237

Wang, Z., Cao, C., Chen, L., Gu, B., Liu, S., Xu, M., et al. (2022). Multimodal neural response and effect assessment during a BCI-based neurofeedback training after stroke. *Front. Neurosci.* 16:884420. doi: 10.3389/fnins.2022.884420

Wang, K., Xu, M., Wang, Y., Zhang, S., Chen, L., and Ming, D. (2020). Enhance decoding of pre-movement EEG patterns for brain-computer interfaces. *J. Neural Eng.* 17:016033. doi: 10.1088/1741-2552/ab598f

Wang, L., Zhang, Y., Zhang, J., Sang, L., Li, P., Yan, R., et al. (2019). Aging changes effective connectivity of motor networks during motor execution and motor imagery. *Front. Aging Neurosci.* 11:312. doi: 10.3389/fnagi.2019.00312

Weibo, Y., Shuang, Q., Kun, W., Hongzhi, Q., Feng, H., Peng, Z., et al. (2016). EEG oscillatory patterns and classification of sequential compound limb motor imagery. *J. Neuroeng. Rehabil.* 13:11. doi: 10.1186/s12984-016-0119-8

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3

Wu, D. R., Xu, Y. F., and Lu, B. L. (2022). Transfer learning for EEG-based brain-computer interfaces: a review of Progress made since 2016. *IEEE Trans Cogn Dev Syst* 14, 4–19. doi: 10.1109/TCDS.2020.3007453

Xu, M., He, F., Jung, T.-P., Gu, X., and Ming, D. (2021). Current challenges for the practical application of electroencephalography-based brain–computer interfaces. *Engineering* 7, 1710–1712. doi: 10.1016/j.eng.2021.09.011

Xu, M., Wu, Q., Xiong, W., Xiao, X., and Ming, D. (2022). Research on encoding and decoding algorithms for medium/high-frequency SSVEP-based brain-computer Interface. *J. Signal Process.* 38, 1881–1891. doi: 10.16798/j.issn.1003-0530.2022.09.011

Xu, M., Xiao, X., Wang, Y., Qi, H., Jung, T. P., and Ming, D. (2018). A brain-computer Interface based on miniature-event-related potentials induced by very small lateral visual stimuli. *I.E.E.E. Trans. Biomed. Eng.* 65, 1166–1175. doi: 10.1109/TBME.2018.2799661

Xygonakis, I., Athanasiou, A., Pandria, N., Kugiumtzis, D., and Bamidis, P. D. (2018). Decoding motor imagery through common spatial pattern filters at the EEG source space. *Comput. Intell. Neurosci.* 2018, 1–10. doi: 10.1155/2018/7957408

Yang, C. Z., Keng, A. K., Chuanchu, W., Cuntai, G., and Haihong, Z. (2009). "Multi-class filter bank common spatial pattern for four-class motor imagery BCI." in *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference 2009*.

Yang, Q., Zhang, Z., Leng, Y., Yang, Y., and Ge, S. (2015). *Phase space reconstruction for improvement of classification in few-channel BCI systems.12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*.

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.* 10:106. doi: 10.1186/1743-0003-10-106

Yong, K. P., and Wonz, C. (2019). Selective feature generation method based on time domain parameters and correlation coefficients for filter-Bank-CSP BCI systems. *Sensors* 19:3769. doi: 10.3390/s19173769

Zang, Y., Jia, F., Weng, X., Li, E., Cui, S., Wang, Y., et al. (2003). Functional organization of the primary motor cortex characterized by event-related fMRI during movement preparation and execution. *Neurosci. Lett.* 337, 69–72. doi: 10.1016/s0304-3940(02)01236-3

Zhang, S., Wang, K., Xu, M., Wang, Z., Chen, L., Wang, F., et al. (2019). Analysis and Classification for Single-Trial EEG Induced by Sequential Finger movements. in: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2019.

Zhang, H., Yao, L., and Long, Z. (2011). "The Functional Alterations Associated with Motor Imagery Training: A Comparison between Motor Execution and Motor Imagery of Sequential Finger Tapping", in: *Conference on Medical Imaging 2011- Biomedical Applications in Molecular, Structural, and Functional Imaging*.

Zhang, Q., Zhang, P., and Song, L. (2019). Brain activation of elite race walkers in action observation, motor imagery, and motor execution tasks: a pilot study. *Front. Hum. Neurosci.* 13:80. doi: 10.3389/fnhum.2019.00080

Zhou, Z., Wan, B., and Ming, D. (2010). A novel technique for phase synchrony measurement from the complex motor imaginary potential of combined body and limb action. *J. Neural Eng.*

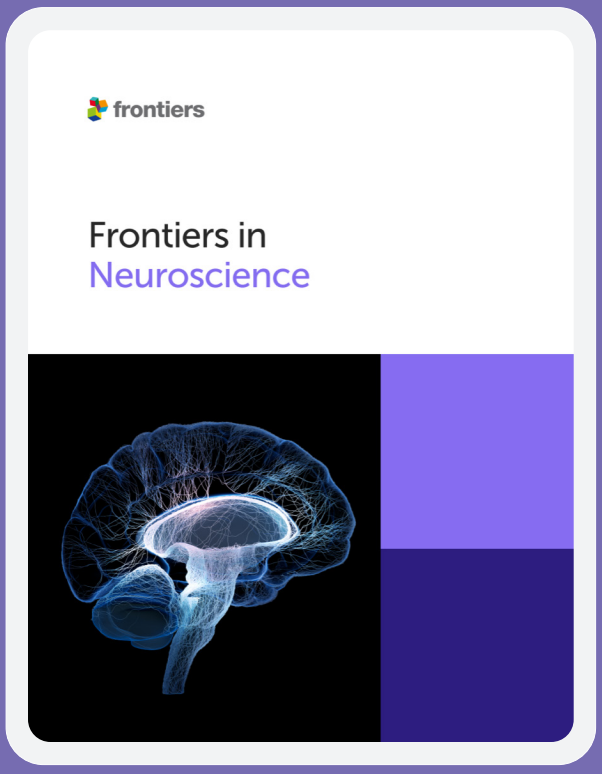
Frontiers in Neuroscience

Provides a holistic understanding of brain function from genes to behavior

Part of the most cited neuroscience journal series which explores the brain - from the new eras of causation and anatomical neurosciences to neuroeconomics and neuroenergetics.

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