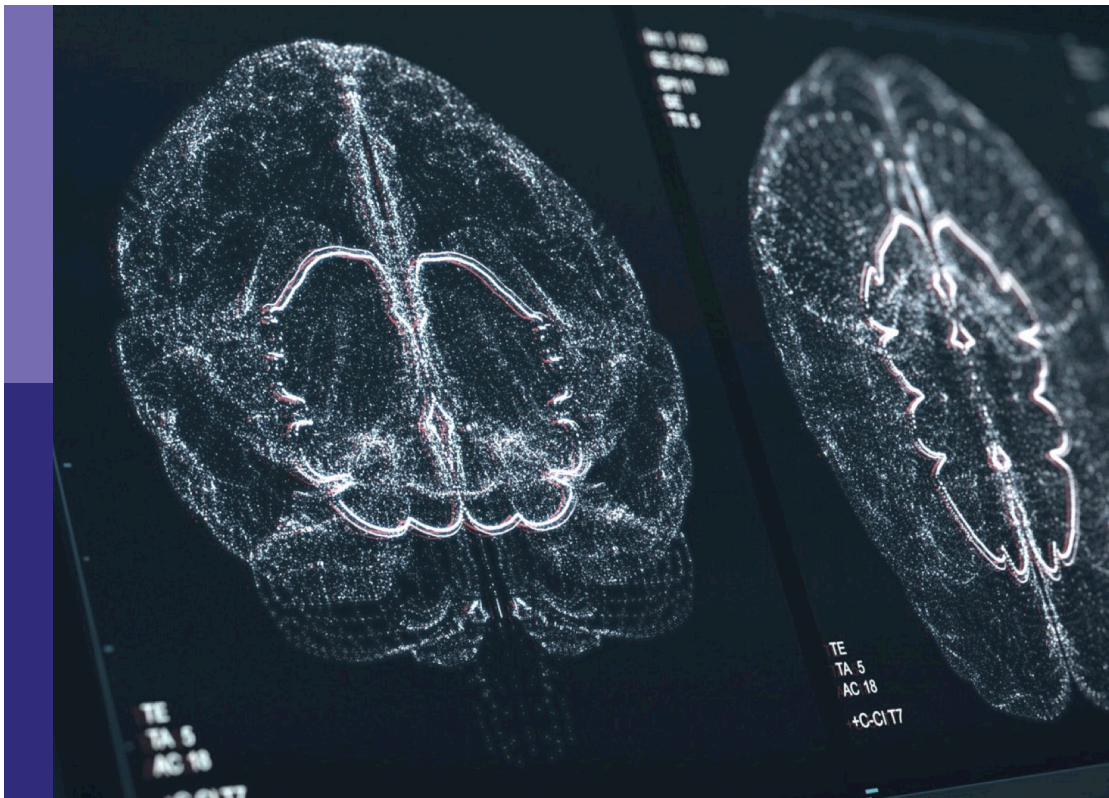
frontiers | Research Topics

# Brain-computer interface and its applications

**Edited by**
Duo Chen, Ke Liu, Jiayang Guo, Luzheng Bi and Jing Xiang

**Published in**
Frontiers in Neuroinformatics
Frontiers in Neurorobotics



ISSN 1664-8714
ISBN 978-2-83251-628-7
DOI 10.3389/978-2-83251-628-7

## 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

# Brain-computer interface and its applications

**Topic editors**

Duo Chen — Nanjing University of Chinese Medicine, China
Ke Liu — Chongqing University of Posts and Telecommunications, China
Jiayang Guo — Xiamen University, China
Luzheng Bi — Beijing Institute of Technology, China
Jing Xiang — Cincinnati Children's Hospital Medical Center, United States

**Citation**

Chen, D., Liu, K., Guo, J., Bi, L., Xiang, J., eds. (2023). *Brain-computer interface and its applications*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-628-7

# Table of contents

frontiers | Frontiers in Neurorobotics

TYPE Editorial
PUBLISHED 09 February 2023
DOI 10.3389/fnbot.2023.1140508



OPEN ACCESS

EDITED AND REVIEWED BY
Alois C. Knoll,
Technical University of Munich, Germany

*CORRESPONDENCE
Duo Chen
380013@njucm.edu.cn

RECEIVED 09 January 2023
ACCEPTED 19 January 2023
PUBLISHED 09 February 2023

CITATION
Chen D, Liu K, Guo J, Bi L and Xiang J (2023) Editorial: Brain-computer interface and its applications. *Front. Neurorobot.* 17:1140508. doi: 10.3389/fnbot.2023.1140508



# Editorial: Brain-computer interface and its applications

Duo Chen[1]*, Ke Liu[2], Jiayang Guo[3], Luzheng Bi[4] and Jing Xiang[5]

[1]School of Artificial Intelligence and Information Technology, Nanjing University of Chinese Medicine, Nanjing, China, [2]School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China, [3]National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen, China, [4]Institute of Mechatronic Systems, Beijing Institute of Technology, Beijing, China, [5]Magnetoencephalography (MEG) Center, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

KEYWORDS

**brain-computer interface (BCI), deep learning, machine learning, artificial intelligence, Electroencephalogram (EEG)**

Editorial on the Research Topic
Brain-computer interface and its applications

Currently, brain-computer interface (BCI) is the research focus and hotspot in the field of neuroscience. Related technologies are widely used in various scenarios such as clinical use, rehabilitation, engineering, and daily life. The BCI uses different brain signals, recording methods, and signal-processing algorithms to build a link between the brain and external software/hardware platforms. With the development of hardware (e.g., BCI chip, wearable device) and algorithms (e.g., machine learning, deep learning), BCI is becoming more practical and stable.

We publish this Research Topic to collect the latest research worldwide in BCI. Researchers from all over the world actively participate and contributed a lot of manuscripts. After carefully and professionally reviewing all submissions, 14 high-quality manuscripts are accepted.

In this topic, several of the contributions focus on the use of deep learning in EEG decoding for BCI, among which convolutional neural network (CNN) is the most widely used. Zhang et al. propose a Multi-Scale 3D-CNN Approach for EEG-based Identity Authentication. The experimental results show that the classification performance of the proposed framework is excellent, and the multi-scale convolution method is effective to extract high-quality identity characteristics across feature domains. Qiu et al. use Electroencephalogram (EEG) and Functional Near Infra-red Spectroscopy (fNIRS) to track the brain activities evoked by neutral and preferred music. The authors conclude that music can promote brain activities, especially in the prefrontal lobe with preferred music. Deng et al. propose SparNet, a CNN composed of five parallel convolutional filters and the Squeeze-and-Excitation Networks (SENet), to learn EEG space-frequency domain characteristics and distinguish between depressive and normal control. The computational results indicate that SparNet achieves a sensitivity of 95.07%, and a specificity of 93.66%. Wang and Cerf combine common spatial pattern (CSP) features and radial basis function neural network (RBFNN) to classify motor imagery EEGs. The algorithm provided high variability within- and across-subjects in EEG-based action decoding. The computational accuracies are higher or close to 90% on two datasets, i.e., BCI competition IV-2a, and -2b. Chen L. et al. transform the EEG signals into symmetric positive definite (SPD) matrices and captures the features of SPD matrices by using CNN. Meta-transfer-learning (MTL) is used to avoid the time-consuming calibration. Chen G. et al. explore the feasibility of an audio-assisted visual BCI speller and a deep learning-based single-trial event-related potentials (ERP) decoding strategy. A spatial-temporal attention-based CNN (STA-CNN) is proposed to recognize the single-trial ERP components. The average classification accuracy of STA-CNN is 77.7% in the EEG dataset recorded from 10 subjects.

Other studies use traditional machine learning and statistical methods for BCI applications in certain modes. Xu et al. use a simulated driving platform with an EEG data collection system for the evaluation of human trust in autonomous vehicles. The graphic theoretical analysis illustrates how human trust varies in EEG under semi-autonomous or fully autonomous driving modes. Wang J. et al. use non-linear dynamics of EEG Signals to Classify primary hand movement intent under opposite hand movement. Their experimental results show significant differences in movement-related cortical potentials between hand movement directions under an opposite hand movement. The results may lay a foundation for the future development of EEG-based human augmentation systems for both the handicapped and the healthy. Massé et al. investigate EEG-based alarm detection in the flight simulator. Cognitive fatigue and cognitive load are manipulated to trigger inattentional deafness, and brain activity is recorded via EEG. The results show that alarm omission and alarm detection can be classified based on time-frequency analysis of EEG. Triana-Guzman et al. combine filter bank CSP (FBCSP) and regularized linear discriminant analysis (RLDA) for decoding EEG rhythms offline and online during motor imagery for standing and sitting. The mean accuracy is higher than 80% in offline analysis, and higher than 90% in online experiments. Song et al. propose a rehabilitative motor imagery BCI system that focuses on rejecting false positive (FP) detection in stroke rehabilitation. A two-phase classifier is used to reject the FP. The algorithm achieved 71.76% selectivity and 13.70% FP rate by using only four EEG channels in the patient group with stroke. Kwon et al. propose a new hybrid visual stimuli for steady-state visual evoked potential (SSVEP)-based BCI, which incorporate various periodic motions into conventional flickering stimuli (FS) or pattern reversal stimuli (PRS). Results demonstrate that FS with sine-wave periodic motion and PRS with square-wave periodic motion could effectively improve the BCI performances compared to conventional FS and PRS.

Finally, two studies provide clinical and human factor applications of magnetoencephalography (MEG) and EEG. Wang Y. et al. classify the MEG data of patients with complex partial seizures (CPS) or simple partial seizures (SPS), using support vector machine. The algorithm obtained a classification accuracy higher than 80%. Cui et al. review the music-emotion recognition and analysis based on EEG signals.

Overall, we hope that this topic can provide some references and novel ideas for researchers in BCI. It should be emphasized that for such a rapidly developing research field, the work that has been done so far is only a drop in the ocean. The manuscripts we collect this time can only be a small leaf in the Amazon rainforest. For BCI, there is still a big gap between the current research and the actual use. Things that need to be solved go far beyond.

## Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## 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

frontiers in Neurorobotics

ORIGINAL RESEARCH
published: 31 January 2022
doi: 10.3389/fnbot.2022.823435

# Multi-Modal Integration of EEG-fNIRS for Characterization of Brain Activity Evoked by Preferred Music

*Lina Qiu[1†], Yongshi Zhong[1†], Qiuyou Xie[2†], Zhipeng He[1], Xiaoyun Wang[3], Yingyue Chen[3], Chang'an A. Zhan[4*] and Jiahui Pan[1*]*

[1] School of Software, South China Normal University, Guangzhou, China, [2] Department of Rehabilitation Medicine, Zhujiang Hospital, Southern Medical University, Guangzhou, China, [3] Guangdong Work Injury Rehabilitation Hospital, Guangzhou, China, [4] School of Biomedical Engineering, Southern Medical University, Guangzhou, China

Music can effectively improve people's emotions, and has now become an effective auxiliary treatment method in modern medicine. With the rapid development of neuroimaging, the relationship between music and brain function has attracted much attention. In this study, we proposed an integrated framework of multi-modal electroencephalogram (EEG) and functional near infrared spectroscopy (fNIRS) from data collection to data analysis to explore the effects of music (especially personal preferred music) on brain activity. During the experiment, each subject was listening to two different kinds of music, namely personal preferred music and neutral music. In analyzing the synchronization signals of EEG and fNIRS, we found that music promotes the activity of the brain (especially the prefrontal lobe), and the activation induced by preferred music is stronger than that of neutral music. For the multi-modal features of EEG and fNIRS, we proposed an improved Normalized-ReliefF method to fuse and optimize them and found that it can effectively improve the accuracy of distinguishing between the brain activity evoked by preferred music and neutral music (up to 98.38%). Our work provides an objective reference based on neuroimaging for the research and application of personalized music therapy.

**Keywords: multi-modality, electroencephalogram (EEG), functional near-infrared spectroscopy (fNIRS), brain activity, preferred music**

OPEN ACCESS

***Edited by:***
*Ke Liu,*
*Chongqing University of Posts and Telecommunications, China*

***Reviewed by:***
*Fali Li,*
*University of Electronic Science and Technology of China, China*
*Peiyang Li,*
*Chongqing University of Posts and Telecommunications, China*

***Correspondence:**
*Jiahui Pan*
*panjiahui@m.scnu.edu.cn*
*Chang'an A. Zhan*
*changan.zhan@gmail.com*

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

***Received:*** *27 November 2021*
***Accepted:*** *07 January 2022*
***Published:*** *31 January 2022*

***Citation:***
*Qiu L, Zhong Y, Xie Q, He Z, Wang X, Chen Y, Zhan CA and Pan J (2022) Multi-Modal Integration of EEG-fNIRS for Characterization of Brain Activity Evoked by Preferred Music. Front. Neurorobot. 16:823435. doi: 10.3389/fnbot.2022.823435*

## INTRODUCTION

Music is the reproduction of the sound of nature that combines science and art. It can not only be used as a form of entertainment to improve people's quality of life (Murrock and Higgins, 2010; Niet et al., 2010; Witte et al., 2020; Chen et al., 2021), but also as a treatment to cure some neurological diseases, such as Alzheimer's disease (Reschke-Hernández et al., 2020), stoke (Sarkamo et al., 2008), disorders of consciousness (DOC) (Carrière et al., 2020), Parkinson's disease (Alfredo, 2015), depression (Chen et al., 2021) and autism (James et al., 2015). For example, Reschke-Hernández et al. (2020) used familiar music as a stimulus to emphasize the effect of familiar music in inducing the emotions of patients with Alzheimer's disease. Sarkamo et al. (2008) showed that for stroke patients, listening music enhances cognitive recovery and mood after middle cerebral artery stroke. In their review, Rollnik and Eckart (2014) also concluded that listening to music has the effect of

awakening and improving mood in patients with impaired consciousness, and can even be used as a means of communicating with patients.

With the rapid development of neuroimaging technology, more and more studies have focused on exploring the relationship between music and its effects on the brain. Previous studies have shown that music affects people not only psychological, but also has a positive impact on cognitive development of the brain, including memory, learning and attention (Franco Jarava, 2018). Bennet and Bennet (2008) found that listening to music helps keep brain neurons active and vigorous and synapses intact. Gui et al. (2019) described the use of fMRI to analyze the brain activity of depressed patients and healthy people under positive and negative music stimulation, and found that their regions of interest (ROI) characteristics are quite different. According to the functional near-infrared spectroscopy (fNIRS) indicator of the prefrontal cortex activation, Zheng et al. (2020) showed that soothing music cause low motivational intensity emotion and uplifting music cause high motivational intensity. In particular, music therapy with personality characteristics helps to make people feel relaxed and improve the mood, behavior and prognosis of patients (Zheng et al., 2020). Jagiello et al. (2019) used electroencephalogram (EEG) to measure the brain responses to familiar vs. unfamiliar music and found that fragments of familiar and unfamiliar music can be quickly distinguished in the brain.

Different states of the brain understand music in different ways and stimulate specific areas of the brain. Different genres of music have been shown to have different effects on brain function, such as significant differences in the activation of the prefrontal lobe were studied with classical and techno music (Bigliassi et al., 2015). Previous studies have shown that music with personal meaning may be more easily perceivable than background music or "relaxing" music, and it is more beneficial to people (Gerdner, 2000; Geethanjali et al., 2018; Jagiello et al., 2019). The findings of Greenberg et al. (2015) showed that individual differences in musical preferences may be related to brain activity and structure. Koelsch (2015) mentioned that because music can evoke people's memory, listening to familiar music can make the brain area which is responsible for memory function responds accordingly and induces people's emotions. Geethanjali et al. (2018) found specific activation and increased functional connectivity after listening to Indian music and Indian music can increased the positive affective scores. However, there are relatively few studies on the effect of personal preferred music on brain activity. What is the characteristic pattern of the brain activity induced by personal preferred music and whether it can promote our brain activity better than other audio stimulus are still unclear.

The current exploration of the relationship between music and brain activity mainly uses single-modality neuroimaging technology, such as fMRI, Positron Emission Computed Tomography (PET) and Magnetoencephalography (MEG) (Blood et al., 1999; Jared et al., 2018; Gui et al., 2019; Carrière et al., 2020). However, previous studies have proved that combining multi-modal imaging technology can effectively use the complementary information of different technologies to overcome the basic limitations of individual modalities, and provide more comprehensive and richer brain information than single-modality imaging technology (Cicalese et al., 2020). Among the commonly used non-invasive neuroimaging technologies, EEG and fNIRS are relatively portable, flexible and inexpensive, and have a wider range of possible applications. EEG can capture the macro-time dynamics of brain electrical activity by recording neuron firing (Pan et al., 2020), and fNIRS estimates brain hemodynamic fluctuations through spectral measurement (Chincarini et al., 2020). These two technologies reflect different aspects of brain neural activity. In addition, EEG measurement has high time resolution but poor stability, while fNIRS has higher spatial resolution and good anti-interference but lower time resolution (He et al., 2020). Therefore, the multi-modal brain imaging system that combining both EEG and fNIRS can simultaneously obtain high-temporal-spatial resolution information, and dynamically observe the information processing processes of the cerebral cortex from the two dimensions of neuroelectric activity and hemodynamics. This is undoubtedly a better strategy for exploring brain activity. Li et al. (2020a) developed an EEG-informed-fNIRS analysis framework to investigate the neuro-correlate between neuronal activity and cerebral hemodynamics by identifying specific EEG rhythmic modulations which contribute to the improvement of the fNIRS based General Linear Model (GLM) analysis. Putze et al. (2013) used EEG and fNIRS to distinguish and detect visual and auditory stimuli processing. They extracted the time-domain and frequency-domain features of EEG and the mean feature of fNIRS and then used the classifier to classify them. They concluded that the fusion of different features of different modal signals has more advantages than the classification accuracy of single-modality features. The multi-modal EEG-fNIRS can provide richer brain activity information than a single-modality. Inspired by the above research, we tried to integrate the multi-modal brain imaging technology of EEG and fNIRS to explore brain activity evoked by personal preferred music.

However, EEG and fNIRS signals are two different brain signals, and the multi-modal integration of EEG-fNIRS is still a challenge in the field of multi-modal research. The commonly used multi-modal integration methods mainly include three strategies: data-level fusion, feature-level fusion, and decision-level fusion. Among them, the feature-level fusion strategy with relatively good effect and relatively simple processing has attracted more attention (Qi et al., 2018). However, most of the current researches using this strategy simply use feature vector splicing to fuse multi-modal features (Hubert et al., 2017). Although the splicing method is simple, it does not consider the correlation and difference between different modalities, and it is difficult to utilize the multi-modal information of EEG-fNIRS fully and effectively. An effective multi-modal integration method can further improve the performance of the EEG-fNIRS system (Khan and Hasan, 2020).

In this work, we propose a multi-modal integration framework of EEG-fNIRS from data collection to data analysis to explore the influence of personal preferred music on brain activity. Under the stimuli of personal preferred music and neutral music, we synchronously collected the brain signals of

the two modalities, and combined their features to explore the brain activities evoked by music. Meanwhile, we employed an improved Normalized-ReliefF algorithm to fuse and optimize the multi-modal features from the two brain signals, which effectively improves the accuracy of distinguishing brain activity caused by preferred music and neutral music. The main contributions of this work are as follows:

(1) Based on the complementarity of EEG and fNIRS, multi-modal data of EEG and fNIRS were simultaneously collected to explore the relationship between music (especially favorite music) and brain activity from different perspectives, which provides imaging-based evidence for the clinical application of personalized music therapy.
(2) A multi-modal integration method of EEG-fNIRS is proposed. We first normalized the multi-modal features from EEG and fNIRS, and then developed an improved ReliefF algorithm to perform feature selection on these multi-dimensional features, and finally fused these features together to realize the full utilization and effective fusion of EEG and fNIRS information.

# MATERIALS AND METHODS

## Subjects

Nine right-handed volunteers (five males and four females, with an average age of 31.25 years) with no history of neurological, psychiatric or other brain-related dis-ease participated in this study. No subjects reported damage to the auditory channel or received professional musical education. Before the start of the experiment, each subject was fully informed of the experimental purpose and methods, and provided writ-ten informed consent prior to the start of the experiment.

## Paradigm Design

Participants' personal preferred and neutral music was used as the experimental stimulus in the present study. Before the experiment, we conducted a questionnaire survey for each participant, asking them to provide one of their personal favorite music, and choosing one of the four unfamous relax music (e.g., soft music) we provided as the neutral music stimulus. All music has lyrics. During the experiment, music was played outside via mobile phones, and the playback volume of each song was basically the same. In order to minimize the interference of environmental noise, we kept the experimental environment as quiet as possible. At the beginning of the experiment, the subject was asked to close their eyes and sit awake in a comfortable chair. Short beeps were emitted at the beginning and end of music playback to indicate the beginning and end of the music stimulus. At the beginning of the measurement, the subjects were asked to stay relaxed for 2 min, and then performed neutral music stimulus task, that is, to continuously listen to a piece of neutral music (about 4 min). After the stimulus task of neutral music, the subjects were asked to stay relaxed for 2 min, and then performed the stimulus task of personal preferred music, that is, listen to a continuous music that the subjects are favorite (about 4 min). After the stimulus task of personal preferred music, the subjects remained at rest for 2 min before ending the experiment. The specific process of the experimental paradigm is shown in **Figure 1**.

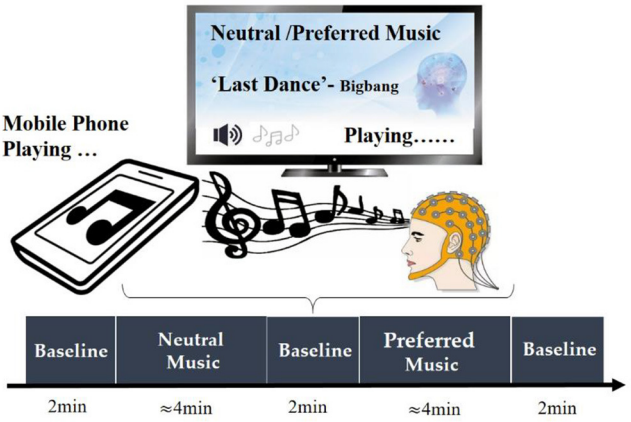



**FIGURE 1 |** The paradigm design of this experiment.

## Data Collection

In this study, we first used EEG technology alone to collect the brain signals of four subject listening to neutral music and their personal preferred music. After preliminary analysis confirmed that the two kinds of music did have an effect on brain activity, we then used EEG and fNIRS technology to simultaneously collect data from five other subjects. EEG signals were collected using a BrainAmp DC EEG recording system (Brain Products GmbH, Germany). The electrode placement follows the international 10–20 convention of a 32-channel cap, and the signal was recorded at a sampling rate of 500 Hz. The fNIRS signals were recorded using a multi-channel NIRScout system (NIRx Medizintechnik GmbH, Germany) at a sampling rate of 3.91 Hz. The source-detector distance was fixed at 3 cm, and a total of 44 measurement channels. The EEG-fNIRS acquisition equipment and the EEG and fNIRS channel locations is shown in **Figure 2**. During the data collection, we used a computer and E-Prime software to form a signal prompting device. We connected the mobile phone to the computer via a USB data cable, and the computer controls the mobile phone to play music, so as to achieve the purpose of controlling the signal prompt device and the mobile phone to play music synchronously. The signal prompting device sends a trigger signal to BrainAmp DC EEG recording system and NIRScout system at the same time through the parallel port. BrainAmp DC EEG recording system and NIRScout system, respectively, amplify these two kinds of brain signals. Finally, the computer simultaneously recorded the brain signals and the event markers processed by the two system.

## Data Processing

### EEG-fNIRS Data Preprocessing

EEG preprocessing was performed using EEGLAB software (v2021.0). Data was first re-referenced to a common-average reference and then filtered from 0.5 to 50 Hz. In order to maintain

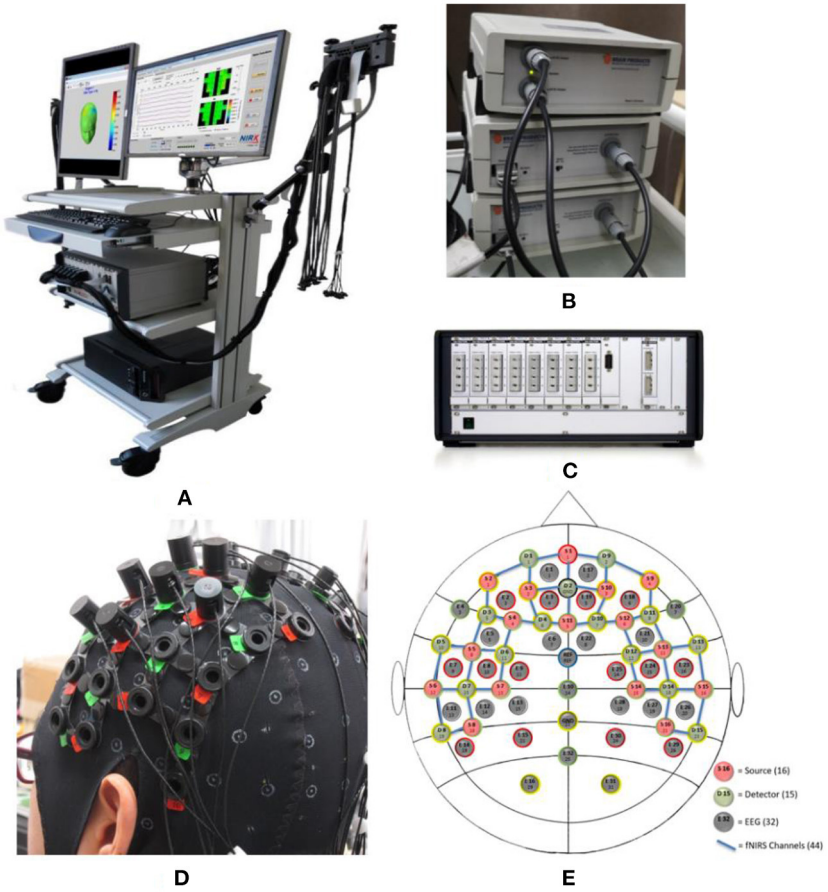



**FIGURE 2 |** Acquisition equipment and measuring cap. **(A)** View of the acquisition equipment; **(B)** Acquisition amplifier of EEG; **(C)** Laser light source of fNIRS; **(D)** View of measuring cap; and **(E)** Channel location of EEG and fNIRS. Where the red circles represent the 16 light sources of fNIRS, the green circles are 15 detesctors of fNIRS, the blue lines are 44 channels of fNIRS, and the gray circles represent 32 EEG electrodes.

the consistency of the data of each subject, we kept the 200-s climax part of the two pieces of music, so the EEG data of each piece of music were segmented to form a period of data from 5 to 205 s after the start of the stimulus. Then, a baseline correction was performed on the segmented data of each stimulus. Finally, independent component analysis (ICA) was used to remove the ocular artifact for each subject.

fNIRS preprocessing was performed using nirsLAB software (v201904). First, a 4th order Butterworth band-pass filter with cut-off frequencies of 0.01–0.1 Hz, was applied to remove artifacts such as those originated from heartbeats (∼1 Hz), venous pressure waves due to respiration (∼0.2 Hz) and arterial pressure oscillations (Mayer waves ∼0.1 Hz). Then, the concentration changes of oxyhemoglobin (*HbO*) and deoxyhemoglobin (*Hb*) were computed according to the Modified Beer-Lambert Law. In fNIRS, *HbO* and *Hb* are the parameters that can indirectly reflect the neural activity of the brain, and are often used in fNIRS data analysis (Yücel et al., 2021). Then, the baseline correction of each channel of the fNIRS signals were performed by subtracting the mean value of the 10 s baseline signal before the start of the stimulus from the signal during each stimulus task. Finally, the fNIRS data of each piece of music were also segmented to form a period of data from 5 to 205 s after the start of the stimuli.

## Feature Extraction

(a) **EEG signals:** We extracted time domain, frequency domain, time-frequency domain and spatial domain features to analyze the characteristics of brain activity.

- **Time Domain Features**
  Regarding the feature extraction in time domain, and we used centralized statistical methods to represent the time series of EEG (Li et al., 2020b). These time-domain statistic features include:
  **Mean:**

$$\mu_s = \frac{1}{N}\sum_{i=1}^{N} S(i) \tag{1}$$

  **Standard deviation:**

$$\sigma_s^2 = \frac{1}{N}\sum_{i=1}^{N} [S(i) - \mu_s]^2 \tag{2}$$

**Mean of the 1st difference absolute value:**

$$\delta_s = \frac{1}{N-1}\sum_{i=1}^{N-1} |S(i+1) - S(i)| \quad (3)$$

**Mean of the normalized 1st difference absolute value:**

$$\overline{\delta_s} = \frac{1}{N-1}\sum_{i=1}^{N-1} \left|\bar{S}(i+1) - \bar{S}(i)\right| = \frac{\delta_s}{\sigma_s} \quad (4)$$

**Mean of the 2nd difference absolute value:**

$$\gamma_s = \frac{1}{N-2}\sum_{i=1}^{N-2} |S(i+2) - S(i)| \quad (5)$$

**Mean of the normalized 2nd difference absolute value:**

$$\bar{\gamma}_s = \frac{1}{N-2}\sum_{i=1}^{N-2} \left|\bar{S}(i+2) - \bar{S}(i)\right| = \frac{\gamma_s}{\sigma_s} \quad (6)$$

where $i$ is the sampling point, $S_i$ represents the EEG signals, $N$ is the number of samples. Then, we put the above statistical features of the signals into a vector as follows:

$$FV_{statistic} = \left[\mu_s, \sigma_s^2, \delta_s, \overline{\delta_s}, \gamma_s, \bar{\gamma}_s\right] \quad (7)$$

- **Frequency Domain Features**

For frequency domain features, we extract the two most typical features: PSD feature (Åkerstedt and Gillberg, 1986) and DE feature (Duan et al., 2013).

**PSD** is a measure of the mean square value of a random variable, and it is the average power dimension per unit frequency. The average power of the signal can be obtained by integrating the power spectrum in the frequency domain. We used the periodogram method (Meziani et al., 2019) to obtain the power spectral density, and we calculated the PSD of the five frequency bands [δ (0.5–3 Hz), θ (4–7 Hz), α (8–13 Hz), β (14–30 Hz), and γ (30–50 Hz)]. Periodogram is a simple and popular method of spectrum estimation, which is based on Discrete Fourier Transform (DFT):

$$F[k] = \sum_{n=0}^{N-1} x[n]\, e^{-\frac{j2\pi kn}{N}} \quad (8)$$

where $Fs$ is the sampling rate of the EEG signals, $j$ and $\pi$ are constants, and $N$ is the number of samples, n is the sampling point, $k = 0, 1, 2...N-1$. We can get the periodic diagram of a discrete-time signal $x[n]$, $n = 1, 2, ..., N$ with a sampling rate of $Fs$ is calculated as:

$$p(f) = \frac{1}{NF_s}\left|F[k]\sum_{k=1}^{N} w[k]\right|^2 \quad (9)$$

where $f = kFs/N$ and $p(f)$ is the PSD feature of the EEG.

**DE** is the generalized form of Shannon entropy (ShEn) $S_{en} = \sum_{i=1}^{n} p(s_i)\log_a \frac{1}{p(s_i)} = -\sum_{i=1}^{n} p(s_i)\log_a p(s_i)$ on continuous variables:

$$DE = \int_a^b S_{en}dx = -\int_a^b p(s_i)\log\left[p(s_i)\right]ds_i \quad (10)$$

where $p(s_i)$ represents the probability density function of continuous information, $[a, b]$ represents the interval of information value.

- **Time-Frequency domain**

**Wavelet Entropy** is the entropy value calculated by the wavelet transform of the signal according to its probability distribution (Quiroga et al., 2001). Shannon's theory of entropy provides a useful tool for analyzing and comparing probability distributions. The calculation formula of wavelet entropy is as follows:

$$S_{WT} = S_{WT}(p) = -\sum_i p_i \ln\left[p_i\right] \quad (11)$$

where $p$ represents the energy intensity ratio of a certain signal.

- **Spatial Features**

The spatial features of EEG signals generally refer to the combined features obtained by comparing the signal features of left-right symmetrical area by using the spatial position information of the EEG electrodes. The spatial domain features are based on the principle that different brain activity states have different activation levels in different areas of the brain. The spatial domain features are calculated as follows: Time domain, frequency domain, and time-frequency domain features of the EEG signals of each channel are used as preliminary features, and then the left-right symmetric electrode feature combination is used as the final spatial domain features. The placement of the 32 EEG electrodes used in our experiment corresponds to the position on the international 10–20 convention. The Cz and Pz in the middle position are removed, and the remaining 30 channel electrodes from 15 symmetrical left-right electrode (AFp1-AFp2, AFF1h-AFF2h, AFF5h-AFF8h, F7-F8, FFC5h-FFC8h, FFC1h-FFC2h, FCC3h-FCC4h, FCC5h-FCC8h, FTT7h-FTT8h, CCP3h-CCP6h, CCP5h-CCP6h, TTP7h-TTP8h, CPP3h-CPP4h, TPP7h-TPP8h, PO3-PO4, a total of 15 pairs of electrodes).

**RASM** refers to the ratio of the eigenvalues of the symmetrical electrode pair on the left and right (Li et al., 2020b). We used the symmetrical electrode feature values ($FX_{L-EEG}$, $FX_{R-EEG}$) of the left and right brain regions to obtain RASM features. The calculation formula of RASM is as (12):

$$F_{RASM} = \frac{FX_{L-EEG}}{FX_{R-EEG}} \quad (12)$$

(b) **fNIRS signals:** We analyzed the features of the fNIRS signals from the perspective of time domain (Naseer and Hong, 2015), considering the features of the change of *HbO* concentration and the change of *Hb* concentration (we denoted it as and), including their statistic features and features based on GLM.

- **Statistic Features**
  We extracted the mean and variance the *HbO* and *Hb* signals of 44 channels of the fNIRS signals as statistical features. The calculation formulas of the two statistic features are shown in (13) and (14):

$$\mu_f = \frac{\sum_{n=t_1}^{t_2} |x(n)|}{f_s(t_2 - t_1)} \tag{13}$$

$$\sigma_f^2 = \frac{\sum_{i=t_1}^{t_2} \left[x(n)' - \mu_f\right]^2}{f_s(t_2 - t_1)} \tag{14}$$

  where $x(n)$ are the and signals, $t_1$ and $t_2$ ($t_2 > t_1$) are two time points, $\mu_f$ and $\sigma_f^2$ are the mean and the variance value of the fNIRS signals in the time periods $t_1$ and $t_2$. We combine the above two statistical features into a feature vector as the statistical feature $SF_f$ of the fNIRS signals.

$$SF_f = \left[\mu_f, \sigma_f^2\right] \tag{15}$$

- **Feature based on GLM**
  When the brain activation changes, *HbO* usually exhibits an approximately linear trend. Using GLM, the *B* value representing the degree of activation of each channel can be calculated to detect the activated channel. The calculation formula of GLM of fNIRS signals is:

$$Y = BX + E \tag{16}$$

  In formula (15), $Y$ is the preprocessed *HbO* data or *Hb* data of the fNIRS signals as the dependent variable, and $X$ is a design matrix. $E$ is a residual matrix that obeys a normal distribution, and $B$ is a matrix with estimated parameters. Based on formula (16), each item of the matrix $Y$ is $y_{ij}$, $i = 1, 2, ..., N$ represents the number of time points of data acquisition, and $j = 1, 2, ..., N$ represents the number of channels. That is, $y_{ij}$ is the *HbO* data or *Hb* data collected by the $j^{th}$ channel at the $i^{th}$ time point. Therefore, when each item of $Y$ is $y_{ij}$, we can get the calculation formula of $y_{ij}$ as:

$$y_{ij} = x_{i1}\beta_{1j} + x_{i2}\beta_{2j} + \ldots + x_{ik}\beta_{kj} + \varepsilon_{ij} \tag{17}$$

  Then we can transform formula (17) into:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \ldots & x_{mn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{18}$$

  Finally, we can obtain the feature *B* value of fNIRS signals by the least square method:

$$B = \left(X'X\right)^{-1}X'Y = [\beta_1\beta_2 \ldots \beta_k]^T \tag{19}$$

## Feature Fusion

Since EEG data and fNIRS data are two different brain signals, they are quite different in principle and acquisition mechanism. When combining these two different types of data, simple feature splicing often leads to poor performance of the machine learning algorithm. Therefore, it is very necessary to normalize the features of the two before fusion and perform feature selection after fusion. Based on this, we proposed a multi-modal feature-level fusion method, the Normalized-ReliefF method. Due to the amplitude and dimension of the various features of the two brain signals are different, in order to more effectively fuse them, we used the normalization algorithm of formula (20) to modulate all the features so that they were scaled to the range of 0 to 1. The normalized features were fused into a new multi-modal feature vector $Mul_{Feat}$, as in formula (21).

$$Feat_i' = \frac{Feat_i - min(Feat_i)}{\max(Feat_i) - min(Feat_i)} \tag{20}$$

$$Mul_{Feat} = [Feat_1' \ldots Feat_i'] \tag{21}$$

Considering that there may be redundant information between these between these multi-modal features, it is often difficult to manually extract complementary and non-redundant information. Therefore, we further used the ReliefF algorithm to optimize the selection of features to achieve more efficient fusion of the two signals. ReliefF is based on the ability of features to distinguish close samples of each class, and evaluates the features by assigning different weights to the features. The larger the feature weight, the more helpful it is to distinguish the categories. When the correlation between the feature and the classification is extremely low, the weight of the feature will be very small, even close to 0. The feature weight may be negative, which means that the distance between similar neighboring samples is greater than the distance between different types of neighboring samples, that is, the feature has a negative impact on classification. For the sample set $Q$, a sample S is randomly selected each time, and then k neighboring samples $NH$ of $S$ are searched for in the same sample set of $S$, and k neighboring samples $NM$ are searched for each sample set of a different category from $S$. Iteratively update the weight $\omega(x)$ of each feature, and the update formula is:

$$\omega(x)' = \frac{\sum_{j=1}^{k} diff\left(X, S, NH_j\right) + \sum_{C \neq Cl(s)} \left[\frac{P(C)}{1 - P\left[Cl(S)\right]} * \sum_{j=1}^{k} diff[X, S, NM(C)_j]\right]}{m * k} \tag{22}$$

$$\omega(x) = \omega(x) - \omega(x)' \tag{23}$$

$$diff\left(X, S, S'\right) = \begin{cases} \frac{|S[X]-S'[X]|}{\max(X)-min(X)}, & where\ X\ is\ continuous \\ 0, & where\ X\ is\ discrete\ and\ S[X] = S'[X] \\ 1, & where\ X\ is\ discrete\ and\ S[X] \neq S'[X] \end{cases} \tag{24}$$

In the formulas (22) and (24), $m$ represents the number of iterations, $NH_j$ defines the $j^{th}$ nearest neighbor sample of the same class, and $NM(C)_j$ is the $j^{th}$ nearest neighbor sample of a different class of class $C$ samples. $P(C)$ represents the probability of the $C^{th}$ target, $Cl(S)$ refers to the category to which sample $S$ belongs, $diff(X, S, S')$ is the distance between sample $S$ and $S'$ with respect to feature $X$.

### Classification

The time domain, frequency domain and spatial domain features are extracted from the EEG signals of each subject, and the statistical features and features based GLM are extracted from fNIRS. Based on these features of EEG and fNIRS, we used six classifiers to classify subjects' brain activity induced by personal preferred and neutral music. The six classifiers we used are support vector machines (SVM), k-nearest neighbor (KNN), Random Forest, AdaBooting, Naive Bayesian and discriminant analysis classifiers (DAC). In order to better perform multi-modal signal fusion, we modulated the signal characteristics of these two different modalities into signals with the same sampling rate (1 Hz) through sample packaging before classification. Furthermore, we performed a 5-fold cross-validation for each classification in order to avoid the phenomenon of false high accuracy.

**SVM** is probably one of the most popular and watched machine learning algorithms. The hyperplane is the line that divides the input variable space. In SVM, the hyperplane is selected to best separate the points in the input variable space from their class (level 0 or level 1). The SVM learning algorithm finds the coefficients that make the hyperplane to best separate the classes.

**KNN** is a commonly used statistical classification method, which can be used not only for regression or linear classification, but also for non-linear classification. It can achieve high classification accuracy, has no assumptions about the data, and is not sensitive to outlier.

**Random Forest** is a classifier that contains multiple decision trees in machine learning, and the output category is determined by the modal of the category output by the individual tree.

**Naive Bayesian** is a series of simple probability classifiers based on the use of Bayes' theorem under the assumption of strong independence between features. The classifier model assigns class labels represented by feature values to the problem instances, and the class labels are taken from a limited set.

**AdaBoosting** is a kind of ensemble method classifier. Each time this method uses bootstrap sampling to construct a tree, it increases the sampling weight for the misjudged observations based on the results of the previous tree, so that the next tree can be more representative of the misjudged observations.

The basic idea of **DAC** is to project high-dimensional pattern samples into the best discriminant vector space to achieve the effect of extracting classification information and compressing the dimension of the feature space. After projection, it is ensured that the model samples have the largest inter-class distance and the smallest intra-class distance in the new subspace, that is, the model has the best separability in the space. Therefore, it is an effective feature extraction method. Using this method can maximize the inter-class scatter matrix of the pattern samples after projection, and at the same time minimize the intra-class scatter matrix.

In summary, as shown in the overall framework of multi-modal EEG-fNIRS integration in **Figure 3**, we integrated EEG and fNIRS from data collection to data analysis to explore the characteristics of music on brain activity.

## RESULTS

For EEG signals recorded by nine subjects while listening to neutral music and their personal preferred music, we calculated the DE values of the five common frequency bands (i.e., δ: 0.5–3 Hz, θ: 4–7 Hz, α: 8–13 Hz, β: 14–30 Hz, and γ: 30–50 Hz) of all channels and distribution of all nine subjects in the five frequency bands induced by personal preferred music and neutral music. **Figure 4A** is the average DE distribution diagram of personal preferred music, **Figure 4B** is the average DE distribution diagram of neutral music, and **Figure 4C** is the difference diagram of the average DE distribution of personal preferred music and neutral music (personal preferred music minus neutral music). It can be seen from **Figure 4** that when the subjects listened to personal preferred music (A) and neutral music (B), the brain had a similar response pattern. The specific performance is as follows: firstly, the two kinds of music in each frequency band show a similar DE distribution pattern; secondly, personal preferred music and neutral music mainly activate the prefrontal lobe (especially the right frontal lobe) and the occipital lobe; finally, in the same brain area, the activation of the lower frequency band (e.g., δ and θ) is often stronger than that of the higher frequency band (e.g., β and γ). Moreover, we analyzed the differences in brain activity induced by the two music by subtracting the average DE of neutral music (B) from the average DE of personal preference music (A), as shown in **Figure 4C**. It can be seen that personal preferred music activates our brain more in right prefrontal, occipital and right temporal regions, and this difference is more obvious in the δ frequency bands.

In this study, due to the limited number of optodes of the fNIRS equipment, which cannot cover the entire human head, we used the high spatial resolution of fNIRS to distribute the 44 channels mainly in the prefrontal and left and right temporal lobes to collect signals. For the fNIRS signals recorded by five subjects while listening to neutral music and their personal preferred music, we calculated the changes in *HbO* concentrations in all 44 channels of each subject. We averaged the *HbO* concentration of all subjects and compared the *HbO* concentration during music listening with the *HbO* in the resting state. We found that compared with the resting state, personal preferred music and neutral music significantly enhanced brain activity. As shown in **Figure 5**, *HbO* concentration indicates that

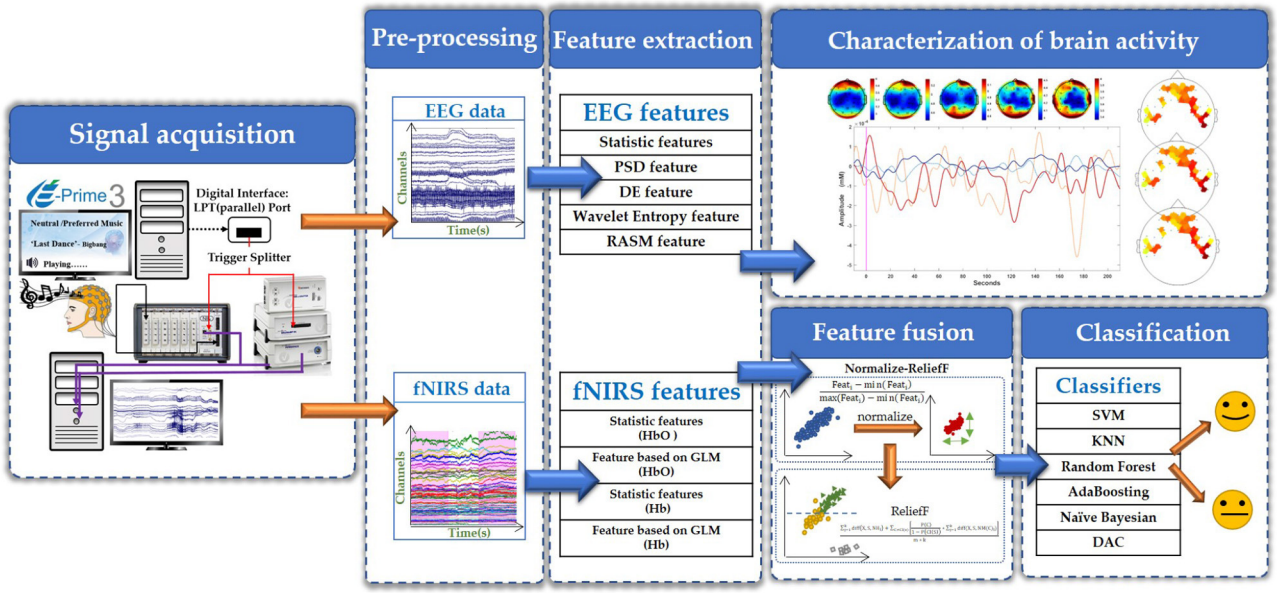



**FIGURE 3 |** The overall framework of EEG-fNIRS multi-modal integration.

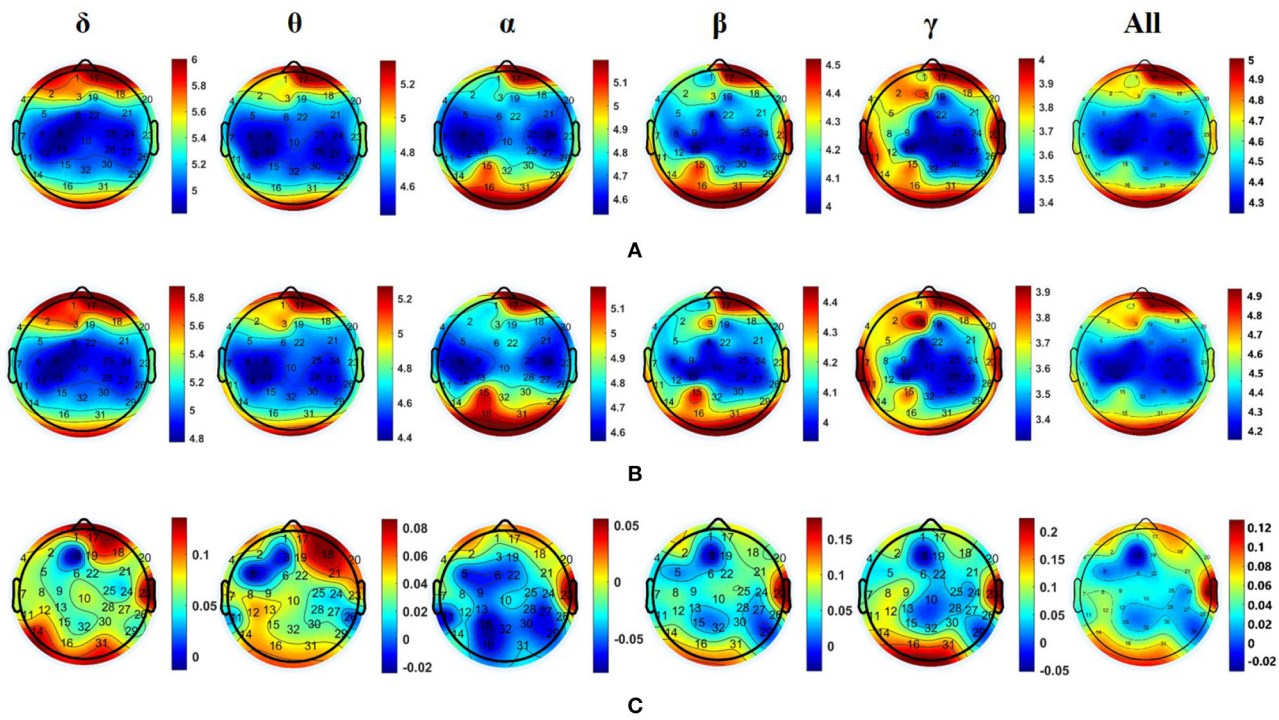



**FIGURE 4 |** Averaged DE distributions of all nine subjects in the five frequency bands induced by personal preferred music **(A)** and neutral music **(B)**, and their DE difference distribution [i.e., personal preference minus neutral music, as shown in **(C)**] in the five frequency bands. Where δ: 0.5–3 Hz; θ: 4–7 Hz; α: 8–13 Hz; β: 14–30 Hz; γ: 30–50 Hz; All (0.5–50 Hz).

the brain response patterns induced by personal preferred music (A) and neutral music (B) are also similar, both of which are significantly ($p < 0.05$) activates the prefrontal lobe and part of the right temporal lobe activation. Moreover, the brain activation induced by personal preferred music is stronger than that of neutral music, and their differences are mainly manifested in the

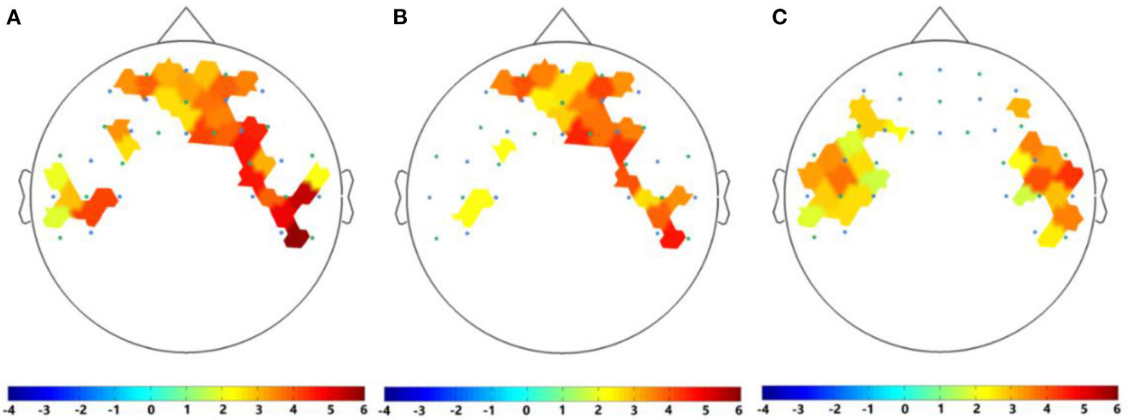



**FIGURE 5 |** *HbO* of contrasts of **(A)** personal preferred music stimulus vs. baseline, **(B)** neutral music stimulus vs. baseline, and **(C)** personal preferred music vs. neutral music. The area shown in the figure represents a statistically significant difference ($p < 0.05$). Where the blue circles represent the light sources of fNIRS and the green circles are the detectors of fNIRS. The darker the red, the more significant the difference and the smaller the *p*-value.

**TABLE 1 |** Classification accuracy of different classifiers based on EEG different features.

| | DE | PSD | Statistic | Wavelet entropy | RASM | Combined features | Averaged accuracy |
|---|---|---|---|---|---|---|---|
| SVM | 93.94% | 94.32% | 96.60% | 49.06% | 83.12% | 97.17% | 85.70% |
| KNN | 80.59% | 86.49% | 91.00% | 52.01% | 70.34% | 88.01% | 78.07% |
| Random forest | 92.12% | 94.91% | 95.69% | 52.17% | 87.41% | 97.63% | 86.66% |
| AdaBoosting | 92.12% | 94.91% | 96.35% | 50.77% | 86.55% | 97.94% | 86.44% |
| Naive bayesian | 78.49% | 70.71% | 80.81% | 51.67% | 82.01% | 84.78% | 74.75% |
| DAC | 91.29% | 90.11% | 92.42% | 49.12% | 85.35% | 95.14% | 83.91% |
| Averaged accuracy | 88.09% | 88.58% | 92.14% | 50.80% | 82.4%7 | 93.45% | |

bilateral temporal lobes, as shown in **Figure 5C**. In this study, *t*-test2 was used for significance analysis. The brain areas shown in **Figure 5** are statistically significant ($p < 0.05$).

In order to further explore the differences in brain activity induced by personal preferred and neutral music, we used six classifiers (i.e., SVM, KNN, Random Forest, AdaBoosting, Naïve Bayesian, and DAC) to classify the features of EEG and fNIRS under the personal preferred and neutral music. For EEG, the features used for classification include DE, PSD, Statistic, Wavelet Entropy, RASM and all EEG fused features. The classification results based on EEG features is shown in **Table 1**. It can be seen from **Table 1** that most of the classifiers can effectively distinguish whether the subject is listening to personal preferred or neutral music based on the characteristics of brain activity. Among multiple features, the combined features have the best average classification effect in all classifiers, with an averaged accuracy of 93.45%. Among the six classifiers, the Random Forest classifier has the highest average classification accuracy, with an averaged accuracy of 86.66%.

For fNIRS, we converted it into *HbO* and *Hb* through the modified Beer Lambert's law, then divided the data and extracted the statistical feature values and the feature values based on GLM of *HbO* and *Hb* per second. After that, we encapsulated the above four features at one sample per second, and statistic feature has a dimension of 400*280 and the feature based on GLM has a dimension of 400*140. We also used the above six classifiers to classify the four features of which are the statistic features and the feature based on GLM of *HbO* and *Hb*. As shown in **Table 2**, the accuracy of the statistical feature based on *Hb* classification under the KNN classifier is the highest, reaching 91.39%, and the statistic feature are better than the classification effect of the features extracted based on GLM.

Based on the multiple features of the above two modalities of EEG and fNIRS, we then used a Normalized-ReliefF method for fusion processing and classification. From **Table 3**, we can see that compared with features based only on EEG and only based on fNIRS, the classification accuracy of multi-modal features based on the fusion of the Normalized-ReliefF method have been significantly improved, with an average accuracy rate of 93.38%.To further demonstrate the effectiveness of the fusion method of Normalized-ReliefF, we also conducted a set of comparative experiments, that is, comparing classification performance of the direct feature splicing method, the normalization algorithm only, the ReliefF algorithm only, and the Normalized-ReliefF method. As shown in **Table 3**, the average classification accuracy of the six classifiers based on the direct feature splicing method is 85.25%, while the classification accuracy of multi-modal fusion based only on the

**TABLE 2 |** Classification accuracy of different classifiers based on fNIRS different features.

| | Statistic (*HbO*) | GLM (*HbO*) | Statistic (*Hb*) | GLM (*Hb*) | Averaged accuracy |
|---|---|---|---|---|---|
| SVM | 80.91% | 55.64% | 83.35% | 54.46% | 68.59% |
| KNN | 88.14% | 54.46% | 91.39% | 58.25% | 73.06% |
| Random forest | 88.86% | 53.82% | 91.19% | 55.31% | 72.30% |
| AdaBoosting | 88.82% | 57.67% | 91.65% | 53.06% | 72.80% |
| Naive bayesian | 67.31% | 63.33% | 69.21% | 55.98% | 63.96% |
| DAC | 79.85% | 55.78% | 82.06% | 53.67% | 67.84% |
| Averaged accuracy | 82.32% | 56.78% | 84.81% | 55.12% | |

**TABLE 3 |** Classification accuracy of different classifiers based on the EEG and fNIRS fusion feature.

| | Splicing | Only normalization | Only ReliefF | Normalized-reliefF |
|---|---|---|---|---|
| SVM | 87.24% | 96.84% | 92.81% | 97.72% |
| KNN | 72.38% | 91.55% | 74.61% | 92.12% |
| Random forest | 91.63% | 97.87% | 98.04% | 98.38% |
| AdaBoosting | 90.22% | 94.41% | 94.39% | 95.79% |
| Naive bayesian | 79.43% | 79.91% | 80.83% | 82.90% |
| DAC | 90.62% | 95.26% | 95.72% | 96.79% |
| Accuracy averaged | 85.25% | 92.64% | 89.40% | 93.68% |

normalization algorithm is 92.64%, which is 7.39% higher than the classification accuracy based on the direct feature splicing method. In the multi-modal fusion classification experiment based only on the ReliefF algorithm, the average accuracy of the six classifications was 89.40%, which is 4.15% higher than the direct feature splicing method. Experimental results prove that these two algorithms are more effective than the direct feature splicing method. The Normalized-ReliefF algorithm, which combines the normalization and ReliefF algorithm, has the highest average classification accuracy, which is 8.43% higher than the direct stitching method, 1.02% higher than the normalized algorithm only, and 4.28% higher than the ReliefF algorithm only. These results prove that the Normalized-ReliefF method we proposed is effective for the multi-modal fusion of EEG-fNIRS. In the comparative experiment, we also calculated the running time required to classify a single subject in each classifier in four methods (i.e., direct splicing method, only the normalize algorithm, only the ReliefF algorithm, and the Normalized-ReliefF method).

## DISCUSSION

Many studies have shown that music not only has a positive effect on people's emotions, but the pleasure of music also has a positive effect on the brain's activity response. To better understand the effect of personal preferred music on brain activity, in this study, we combined EEG and fNIRS technology to simultaneously measure the brain activity of healthy subjects when listening to neutral music and their preferred music, and used the Normalized-ReliefF method to identify the characteristics of the brain activity evoked by the two types of music. Previous work exploring the influence of music on brain activity was mostly based on a single-modality, and our work is to integrate two different modalities, EEG and fNIRS, so as to obtain more abundant brain information from the two aspects of neuroelectric signals and cerebral hemodynamic signals. Furthermore, most of the previous feature fusion studies were to fuse multiple features of a single-modality such as EEG signals (Nguyen et al., 2018; Li et al., 2019; Hua et al., 2021). Our study focuses on the feature fusion of two different brain signals (EEG and fNIRS), which are different in principle, acquisition mechanism, and signal amplitude. Therefore, we combined normalization and ReliefF algorithms in the fusion strategy. Normalization is mainly used to modulate features from different modes to the same scale. The ReliefF algorithm is mainly used to optimize and select high-dimensional multi-scale features from two modalities.

From the results of EEG and fNIRS analysis, we can draw a conclusion that both personal preferred and neutral music can enhance brain activity and have similar activation patterns. The activated area is mainly distributed in the prefrontal lobe. The functions of the prefrontal lobe are mainly related to emotion, cognition and memory, and is also related to reward system of human brain (Rouault et al., 2019). There's evidence showing that music is closely connected to the stimulation of neurons and executive function of the prefrontal cortex. There may be integration of music and autobiographical memory in the medial prefrontal cortex, facilitating retrieval of personally salient episodic memories when listening to familiar musical excerpts (Janata, 2009).

In the EEG analysis, we also found that the occipital lobe was also partially activated when the subjects listened to music. Since we did not collect fNIRS data from the occipital lobe,

it was impossible to compare with EEG data. The activation of the occipital lobe may due to the fact that this experiment required the subjects to close their eyes during the experiment. Participants did experiments with their eyes closed, the activation of the occipital lobes of the brain will be significantly enhanced in the α band. Barry and Blasio (2017) studied the changes in the EEG of the elderly and young people in the resting state with the eyes open and closed, and found that people have a stronger α response in the occipital lobe when the eyes are closed than when the eyes are open. Our fNIRS results show that in addition to the prefrontal lobe, the temporal lobe is also partially activated, especially the right temporal lobe, which is consistent with a previous study (Alfredson et al., 2004). The functions of temporal lobe is mainly related to hearing, memory and mental activity (Bougeard and Fischer, 2002). There is evidence that music can also activate the temporal lobe area (Alfredson et al., 2004; Hosseini and Hosseini, 2019). For the right hand, the right temporal lobe is its non-dominant hemispherical temporal lobe. Its main function is to recognize high-level neural activities such as memory, association, and comparison. Music have been demonstrated has a positive effect on cognitive development of the brain, including memory, learning and attention (Franco Jarava, 2018). Pan et al. (2018) projected the DE and common spatial pattern (CSP) features of happy and sad emotions onto the scalp for subjects and confirmed that different emotional states will produce different responses in brain regions. At the same time, they also found that during positive emotion processing, the neural patterns had significantly higher brain responses at the temporal lobes.

How to integrate information of different modalities has always been a difficult problem in the field of multi-modal research. In this study, we proposed an EEG-fNIRS multi-modal integration framework from data collection to data analysis to explore the effect of personal preference music on brain activity. First, we used a synchronous trigger device in data acquisition to integrate EEG and fNIRS into an EEG-fNIRS synchronous acquisition system, achieving the synchronous recording of two different brain signals. In data analysis, in order to make full use of the information of the two modalities, we proposed an improve Normalized-ReliefF method to fuse and optimize multi-modal features. Most of the current studies are to splice the features of different modalities directly into the classifier without any processing (Al-Shargie et al., 2016; Hong et al., 2018; Cicalese et al., 2020). Although this splicing method can also achieve higher accuracy than single-modalities, it does not consider that the types and scales of features from different modes are different, and there is also information redundancy between them. Obviously, this processing method cannot make full and effective use of multi-modal information and truly realize the superiority of multi-modal systems.

The Normalized-ReliefF method we proposed can effectively solve the problem of feature fusion of EEG and fNIRS. We first normalized all the feature data of EEG and fNIRS to modulate them to the same scale. Then, the ReliefF method was used to perform feature selection on multiple features from different modalities and different dimensions to achieve the effect of removing redundant information and reducing dimensionality at the same time. The normalization process can not only evaluate the features of the two modalities of EEG and fNIRS on the same scale, but also shorten the data training time to a certain extent. The ReliefF algorithm is a feature weighting algorithm, which assigns different weights to features according to the correlation of each feature and category, and features with a weight less than a certain threshold will be removed The running time of the ReliefF algorithm increases linearly with the increase in the number of samples and the number of original features, so the running efficiency is relatively high (Stamate et al., 2018; Kshirsagar and Kumar, 2021; Satapathy and Loganathan, 2021). In our previous work (Pan et al., 2021), we also confirmed the effectiveness of the ReliefF algorithm for EEG-based multi-modal feature extraction, and it can achieve the effect of feature optimization and feature dimensionality reduction to a certain extent. The Normalized-ReliefF method we proposed can effectively solve the problem of fusion of EEG and fNIRS features.

In order to further verify the effectiveness of the Normalized-ReliefF method, we conducted a set of comparative ablation experiments. By comparing the classification performance of direct feature splicing method, only the normalized algorithm, only the ReliefF algorithm and the Normalized-ReliefF, we found that the Normalized-ReliefF method can effectively improve the average accuracy of the classification of brain activities induced by preferred and neutral music. In terms of calculation time, the Normalized-ReliefF method is also more efficient than the other three fusion methods. This fully proves the efficiency of the Normalized-ReliefF method in the fusion of the two modal features of EEG and fNIRS. Furthermore, our experimental results also verify that the multi-modal features based on EEG-fNIRS have better classification performance than signal-modal features based only on EEG and only on fNIRS, because the combination of the two modalities can provide richer brain activity information.

In the exploration of the effect of preferred music on brain activity in this work, we proposed an EEG-fNIRS integration framework to systematically integrate EEG-fNIRS multi-modal signals from data collection to data analysis. The results of multi-modal analysis found that personally preferred music promotes brain activity, and our proposed fusion method, Normalized-ReliefF, effectively improves the recognition accuracy of brain activity induced by different music. Our work may provide neuroimaging-based references for the research and application of personalized music therapy. Taking patients with DOC as an example, patients with DOC can be divided into MCS and VS patients (Xie et al., 2018). At present, the clinical diagnosis and assessment of consciousness in patients with DOC mainly rely on behavioral scales, but the misdiagnosis rate of this method is as high as 37–43% (Hirschberg and Giacino, 2011). It is very necessary to make an objective and effective diagnosis of the consciousness of patients with consciousness disorders. Our work has confirmed that personal preferred music can promote the brain. In other words, personal preferred music cannot only be used as a brain stimulus, but also as a treatment method for patients. In view of the current difficulties in diagnosing and treating DOC patients in clinical practice, we can use the patients' personal preferred music as a stimulus, and explore

the characteristics of brain activity induced by music in patients with different states of consciousness by using multi-modal imaging technology.

However, our study still has some limitations. First, our sample size is relatively small. Secondly, our experimental paradigm only listened to preferred music and neutral music once, and did not conduct multiple rounds of repeated task. Finally, we only analyze the activation analysis, not the analysis of the brain functional connectivity. In the future, we will increase samples, improve the experimental design and data analysis methods, and include patients with neurological diseases (such as DOC patients) into the scope of subjects to compare the effects of personal preferred music on healthy subjects and patients' brain functions, and truly realize music therapy.

## CONCLUSION

In this study, we proposed the integration framework of the two modalities of EEG and fNIRS, including data collection and various data processing and analysis. We used this multi-modal integration framework to explore the characteristics of brain activity induced by personal preferred and neutral music, and found that music can enhance the brain activity, especially the prefrontal lobe, and personal preferred music activated their brain more than neutral music. We also proposed an improved Normalized-ReliefF algorithm to fuse and optimize multiple features of two different physiological signals EEG and fNIRS to identify the characteristics of brain activity induced by personal preference music and neutral music. We found that using the Normalized-ReliefF algorithm is more effective than the method of simple multi-feature vector splicing of the two modalities for classification. We also found that the classification accuracy rates obtained by using fusion features based on EEG-fNIRS are higher than the classification accuracy rates obtained by EEG-based features and fNIRS-based features, which proves that multi-modal brain imaging can provide better classification performance than single-modality. Our work can provide an objective reference based on neuroimaging for the research and application of personalized music therapy.

## 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 Ethics Committee of South China Normal University. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

LQ, YZ, and JP: conceptualization and writing-review and editing. LQ, YZ, and ZH: methodology. YZ: software, data curation, and writing-original draft preparation. LQ, ZH, XW, YC, and CZ: validation. LQ, YZ, ZH, and JP: formal analysis. LQ, QX, ZH, XW, YC, CZ, and JP: investigation. LQ, YZ, QX, XW, YC, and CZ: resources. LQ, YZ, ZH, XW, YC, and JP: visualization. LQ and JP: supervision and funding acquisition. LQ, QX, and CZ: project administration. All authors have read and agreed to the published version of the manuscript.

## FUNDING

This work was supported by the Guangdong Basic and Applied Basic Research Foundation under grant 2019A1515110388, the Key Realm R&D Program of Guangzhou under grant 202007030005, and the National Natural Science Foundation of China under grants 62076103, 81974154, and 82171174.

## ACKNOWLEDGMENTS

We thank the reviewers for excellent criticism of the article.

## REFERENCES

Åkerstedt, T., and Gillberg, M. (1986). Sleep duration and the power spectral density of the EEG. *Electroencephalogr. Clin. Neurophysiol.* 64, 119–122. doi: 10.1016/0013-4694(86)90106-9

Alfredo, R. (2015). Music therapy interventions in Parkinson's Disease: the state-of-the-art. *Front. Neurol.* 6, 185. doi: 10.3389/fneur.2015.00185

Alfredson, B. B., Risberg, J., Hagberg, B., and Gustafson, L. (2004). Right temporal lobe activation when listening to emotionally significant music. *Appl. Neuropsychol.* 11, 161–166. doi: 10.1207/s15324826an1103_4

Al-Shargie, F. M., Tang, T. B., Badruddin, N., Dass, S. C., and Kiguchi, M. (2016). "Mental Stress Assessment Based on Feature Level Fusion of fNIRS and EEG Signals," in *2016 6th International Conference on Intelligent and Advanced Systems (ICIAS) - 6th International Conference on Intelligent and Advanced Systems* (Kuala Lumpur), 1–5.

Barry, R. J., and Blasio, F. D. (2017). EEG differences between eyes-closed and eyes-open resting remain in healthy ageing. *Biol. Psychol.* 129, 293. doi: 10.1016/j.biopsycho.2017.09.010

Bennet, A., and Bennet, D. H. (2008). The human knowledge system: music and brain coherence. *Vine* 38, 277–295. doi: 10.1108/03055720810904817

Bigliassi, M., León-Domínguez, U., and Altimari, L. R. (2015). How does the prefrontal cortex "listen" to classical and techno music? A functional near-infrared spectroscopy (fNIRS) study. *Psychol. Neurosci. Methods* 8, 246–256. doi: 10.1037/h0101064

Blood, A. J., Zatorre, R. J., Bermudez, P., and Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nat. Neurosci.* 2, 382–387. doi: 10.1038/7299

Bougeard, R., and Fischer, C. (2002). The role of the temporal pole in auditory processing. *Epileptic Disord.* 4(Suppl. 1), S29. doi: 10.7227/CE.78.1.7

Carrière, M., Larroque, S. K., Charlotte, C., Bahri, M. A., and Heine, L. (2020). An echo of consciousness: brain function during preferred music. *Brain Connect.* 10:385–395. doi: 10.1089/brain.2020.0744

Chen, X., Wei, Q., Jing, R., and Fan, Y. (2021). Effects of music therapy on cancer-related fatigue, anxiety, and depression in patients with digestive tumors: a protocol for systematic review and meta-analysis. *Medicine* 100, e25681. doi: 10.1097/MD.0000000000025681

Chincarini, M., Costa, E. D., Qiu, L., Spinelli, L., and Torricelli, A. (2020). Reliability of fNIRS for noninvasive monitoring of brain function and emotion in sheep. *Sci. Rep.* 10, 14726. doi: 10.1038/s41598-020-71704-5

Cicalese, P. A., Li, R., Ahmadi, M. B., Wang, C., Francis, J. T., Selvaraj, S., et al. (2020). An EEG-fNIRS hybridization technique in the four-class classification of alzheimer's disease. *Neurosci. Methods* 336. doi: 10.1016/j.jneumeth.2020.108618

Duan, R. N., Zhu, J. Y., and Lu, B. L. (2013). "Differential entropy feature for EEG-based emotion classification," in *International IEEE/EMBS Conference on Neural Engineering* (Seattle, WA), 81–84.

Franco Jarava, J. (2018). *Preferred Music Effect on Human Brain Using Functional Near-Infrared Spectroscopy*. Barcelona: Universitat Politècnica de Catalunya.

Geethanjali, B., Adalarasu, K., Mohan, J., and Seshadri, N. (2018). Music induced brain functional connectivity using EEG sensors: a study on indian music. *IEEE Sens. J.* 19:1. doi: 10.1109/JSEN.2018.2873402

Gerdner, L.A. (2000). Effects of individualized versus classical "relaxation" music on the frequency of agitation in elderly persons with Alzheimer's disease and related disorders. *Int. Psychogeriatr.* 12, 49–65. doi: 10.1017/S1041610200006190

Greenberg, D. M., Baron-Cohen, S., Stillwell, D. J., Kosinski, M., and Rentfrow, P. J. (2015). Musical preferences are linked to cognitive styles. *PLoS ONE* 10:e0131151. doi: 10.1371/journal.pone.0131151

Gui, R., Chen, T., and Nie, H. (2019). The impact of emotional music on active ROI in patients with depression based on deep learning: a task-state fMRI study. *Comp. Intell. Neurosci.* 2019, 1–14. doi: 10.1155/2019/5850830

He, Z., Li, Z., Yang, F., Wang, L., and Pan, J. (2020). Advances in multimodal emotion recognition based on brain-computer interfaces. *Brain Sci.* 10, 687. doi: 10.3390/brainsci10100687

Hirschberg, R., and Giacino, J. T. (2011). The vegetative and minimally conscious states: diagnosis, prognosis and treatment. *Neurol. Clin.* 29, 773–786. doi: 10.1016/j.ncl.2011.07.009

Hong, K. S., Khan, M. J., and Hong, M. J. (2018). Feature extraction and classification methods for hybrid fNIRS-EEG brain-computer interfaces. *Front. Hum. Neurosci.* 12, 246. doi: 10.3389/fnhum.2018.00246

Hosseini, E., and Hosseini, S. A. (2019). Therapeutic effects of music: a review. *RepHealth Care* 4, 1–13.

Hua, Y., Zhong, X., Zhang, B., Yin, Z., and Zhang, J. (2021). Manifold feature fusion with dynamical feature selection for cross-subject emotion recognition. *Brain Sci.* 11, 1392. doi: 10.3390/brainsci11111392

Hubert, B., Rishabh, G., and Falk, T. H. (2017). Mental task evaluation for hybrid NIRS-EEG brain-computer interfaces. *Comp Intell.* 2017, 1–24. doi: 10.1155/2017/3524208

Jagiello, R., Pomper, U., Yoneya, M., Zhao, S., and Chait, M. (2019). Rapid brain responses to familiar vs. unfamiliar music – an EEG and pupillometry study. *Sci. Rep.* 9:15570. doi: 10.1038/s41598-019-51759-9

James, R., Sigafoos, J., Green, V. A., Lancioni, G. E., O'Reilly, M. F., Lang, R., et al. (2015). Music therapy for individuals with autism spectrum disorder: a systematic review. *Rev. J. Autism Dev. Disord.* 2, 39–54. doi: 10.1007/s40489-014-0035-4

Janata, P. (2009). The neural architecture of music-evoked autobiographical memories. *Cerebral Cortex* 19, 2579–2594. doi: 10.1093/cercor/bhp008

Jared, B., Yuya, T., Shinya, K., and Koichi, Y. (2018). Spectral-spatial differentiation of brain activity during mental imagery of improvisational music performance using MEG. *Front. Hum. Neurosci.* 12, 156. doi: 10.3389/fnhum.2018.00156

Khan, M. U., and Hasan, M. A. H. (2020). Hybrid EEG-fNIRS BCI fusion using multi-resolution singular value decomposition (MSVD). *Front. Hum. Neurosci.* 14, 599802. doi: 10.3389/fnhum.2020.599802

Koelsch, S. (2015). Music-evoked emotions: principles, brain correlates, and implications for therapy. *Ann. N. Y. Acad. Sci.* 1337, 193–201. doi: 10.1111/nyas.12684

Kshirsagar, D., and Kumar, S. (2021). An efficient feature reduction method for the detection of DoS attack. *ICT Exp.* 7:371–375. doi: 10.1016/j.icte.2020.12.006

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 Transact. Biomed. Eng.* 1:2869–2881. doi: 10.1109/TBME.2019.2897651

Li, R., Zhao, C., Wang, C., Wang, J., and Zhang, Y. (2020a). Enhancing fNIRS analysis using EEG rhythmic signatures: an EEG-informed fNIRS analysis study. *IEEE Transact. Biomed. Eng.* 67, 2789–2797. doi: 10.1109/TBME.2020.2971679

Li, Z., Qiu, L., Li, R., He, Z., Xiao, J., Liang, Y., et al. (2020b). Enhancing BCI-based emotion recognition using an improved particle swarm optimization for feature selection. *Sensors* 20:3028. doi: 10.3390/s20113028

Meziani, A., Djouani, K. D., Medkour, T., and Chibani, A. (2019). A Lasso quantile periodogram based feature extraction for EEG-based motor imagery. *J. Neurosci. Methods* 328, 108434. doi: 10.1016/j.jneumeth.2019.108434

Murrock, C. J., and Higgins, P. A. (2010). The theory of music, mood and movement to improve health outcomes. *J. Adv. Nurs.* 65, 2249–2257. doi: 10.1111/j.1365-2648.2009.05108.x

Naseer, N., and Hong, K. S. (2015). fNIRS-based brain-computer interfaces: a review. *Front. Hum. Neurosci.* 9, 3. doi: 10.3389/fnhum.2015.00003

Nguyen, D., Nguyen, K., Sridharan, S., Dean, D., and Fookes, C. (2018). Deep spatio-temporal feature fusion with compact bilinear pooling for multimodal emotion recognition. *Comp. Vis. Image Understand.* 174, 33–42. doi: 10.1016/j.cviu.2018.06.005

Niet, G. D., Tiemens, B., Lendemeijer, B., and Hutschemaekers, G. (2010). Music-assisted relaxation to improve sleep quality: meta-analysis. *J. Adv. Nurs.* 65, 1356–1364. doi: 10.1111/j.1365-2648.2009.04982.x

Pan, J., Xie, Q., Huang, H., He, Y., Sun, Y., Yu, R., et al. (2018). Emotion-related consciousness detection in patients with disorders of consciousness through an EEG-based BCI system. *Front. Hum. Neurosci.* 12, 198. doi: 10.3389/fnhum.2018.00198

Pan, J., Xie, Q., Qin, P., Chen, Y., He, Y., Huang, H., et al. (2020). Prognosis for patients with cognitive motor dissociation identified by brain-computer interface. *Brain* 143, 1–13. doi: 10.1093/brain/awaa026

Pan, J., Zhang, J., Wang, F., Liu, W., Huang, H., Tang, W., et al. (2021). Automatic sleep staging based on EEG-EOG signals for depression detection. *Intell. Automat. Soft Comp.* 28, 53–71. doi: 10.32604/iasc.2021.015970

Putze, F., Hesslinger, S., Tse, C. Y., Huang, Y. Y., and Schultz, T. (2013). Hybrid fNIRS-EEG based classification of auditory and visual perception processes. *Front. Neurosci.* 8, 373. doi: 10.3389/fnins.2014.00373

Qi, S., Calhoun, V. D., Erp, T. V., Bustillo, J., Damaraju, E., Turner, J. A., et al. (2018). Multimodal fusion with reference: searching for joint neuromarkers of working memory deficits in schizophrenia. *IEEE Trans. Med. Imaging* 37, 1. doi: 10.1109/TMI.2017.2725306

Quiroga, R. Q., Rosso, O. A., Başar, E., Schürmann, M. (2001). Wavelet entropy in event-related potentials: a new method shows ordering of EEG oscillations. *Biol. Cybern.* 84, 291–299. doi: 10.1007/s004220000212

Reschke-Hernández, A. E., Belfi, A. M., Guzmán-Vélez, E., and Tranel, D. (2020). Hooked on a feeling: influence of brief exposure to familiar music on feelings of emotion in individuals with Alzheimer's Disease. *J. Alzheimers Dis.* 78, 1019–1031. doi: 10.3233/JAD-200889

Rollnik, J. D., and Eckart, A. (2014). Music in disorders of consciousness. *Front. Neurosci.* 8, 190. doi: 10.3389/fnins.2014.00190

Rouault, M., Drugowitsch, J., and Koechlin, E. (2019). Prefrontal mechanisms combining rewards and beliefs in human decision-making. *Nat. Commun.* 10:301. doi: 10.1038/s41467-018-08121-w

Sarkamo, T., Tervaniemi, M., Laitinen, S., Forsblom, A., Soinila, S., Mikkonen, M., et al. (2008). Music listening enhances cognitive recovery and mood after middle cerebral artery stroke. *Brain Sci.* 131 (Pt 3), 866–876. doi: 10.1093/brain/awn013

Satapathy, S. K., and Loganathan, D. (2021). A study of human sleep stage classification based on dual channels of EEG signal using machine learning techniques. *SN Comp. Sci.* 2:157. doi: 10.1007/s42979-021-00528-5

Stamate, D., Alghambdi, W., Ogg, J., Hoile, R., and Murtagh, F. (2018). "A machine learning framework for predicting dementia and mild cognitive impairment," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)* (Orlando, FL), 671–678.

Witte, M. D., Pinho, S., Stams, G. J., Moonen, X., and Hooren, S. V. (2020). Music therapy for stress reduction: a systematic review and meta-analysis. *Health Psychol. Rev.* 14, 1–26. doi: 10.1080/17437199.2020.1846580

Xie, Q., Pan, J., Chen, Y., He, Y., Ni, X., Zhang, J., et al. (2018). A gaze-independent audiovisual brain-computer Interface for detecting

awareness of patients with disorders of consciousness. *BMC Neurol.* 18:144. doi: 10.1186/s12883-018-1144-y

Yücel, M., Lühmann, A., Scholkmann, F., Gervain, J., and Wolf, M. J. N. (2021). Best practices for fNIRS publications. *Neurophotonics*. 8. doi: 10.1117/1.NPh.8.1.012101

Zheng, M., Lin, H., and Chen, F. (2020). "An fNIRS study on the effect of music style on cognitive activities," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) in Conjunction With the 43rd Annual Conference of the Canadian Medical and Biological Engineering Society*.

**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 Neurorobotics

ORIGINAL RESEARCH
published: 28 April 2022
doi: 10.3389/fnbot.2022.845127

# Using Non-linear Dynamics of EEG Signals to Classify Primary Hand Movement Intent Under Opposite Hand Movement

*Jiarong Wang, Luzheng Bi and Weijie Fei**

*School of Mechanical Engineering, Beijing Institute of Technology, Beijing, China*

Decoding human hand movement from electroencephalograms (EEG) signals is essential for developing an active human augmentation system. Although existing studies have contributed much to decoding single-hand movement direction from EEG signals, decoding primary hand movement direction under the opposite hand movement condition remains open. In this paper, we investigated the neural signatures of the primary hand movement direction from EEG signals under the opposite hand movement and developed a novel decoding method based on non-linear dynamics parameters of movement-related cortical potentials (MRCPs). Experimental results showed significant differences in MRCPs between hand movement directions under an opposite hand movement. Furthermore, the proposed method performed well with an average binary decoding accuracy of 89.48 ± 5.92% under the condition of the opposite hand movement. This study may lay a foundation for the future development of EEG-based human augmentation systems for upper limbs impaired patients and healthy people and open a new avenue to decode other hand movement parameters (e.g., velocity and position) from EEG signals.

**Keywords: EEG, hand movement decoding, human augmentation, human factors, human-machine interaction**

OPEN ACCESS

***Edited by:***
*Ganesh R. Naik,*
*Flinders University, Australia*

***Reviewed by:***
*Baoguo Xu,*
*Southeast University, China*
*Wanzeng Kong,*
*Hangzhou Dianzi University, China*
*Erwei Yin,*
*Tianjin Artificial Intelligence Innovation Center (TAIIC), China*

***Correspondence:**
*Weijie Fei*
*3120170241@bit.edu.cn*

***Received:*** *29 December 2021*
***Accepted:*** *29 March 2022*
***Published:*** *28 April 2022*

***Citation:***
*Wang J, Bi L and Fei W (2022) Using Non-linear Dynamics of EEG Signals to Classify Primary Hand Movement Intent Under Opposite Hand Movement. Front. Neurorobot. 16:845127. doi: 10.3389/fnbot.2022.845127*

## INTRODUCTION

Human augmentation refers to using assistive devices and technologies to help people overstep human motor, perception, and cognition limitations. The applications of human augmentation have shown diversity, including but not limited to prostheses (Kvansakul et al., 2020), exoskeleton (Chen et al., 2017; Yandell et al., 2019), and augmented reality (Kansaku et al., 2010; Chen et al., 2021). For human augmentation systems, providing active assistance instead of passive assistance according to human intention is of high value. Fusing human intention into augmentation technology makes it possible to establish a more intelligent, flexible, and user-friendly system.

Brain-computer interfaces (BCIs) have been the essential tool to detect human intention with the development of neuroscience. BCIs could translate human intention from neural signals directly. Among various brain signal recording methods, electroencephalogram (EEG) is more practical for human augmentation because it is non-invasive, cheap, and convenient to use. Over the past decades, numerous studies have been focused on using EEG signals to decode human intention and develop a body augmentation system, e.g., P300 speller (Farwell and Donchin, 1988), steady-state visually evoked potential-based

BCI systems (Gao et al., 2021), e.g., exoskeleton control (Kwak et al., 2015) and motor rehabilitation (Zhao et al., 2016), motor imagery (MI)-based mobile wheelchair (He et al., 2016; Zhang et al., 2016), movement-related cortical potential (MRCP)-based robotic arm control (Schwarz et al., 2020a,b). Compared with evoked potentials-based BCIs (Yin et al., 2015a,b) and MI-based BCIs (Pei et al., 2022), MRCP-based BCIs do not rely on external evoked stimuli (such as P300) and repetitive imagination (such as MI). It can decode human intention from natural movement execution and provide a more realistic application scene for human augmentation.

MRCP-based hand (or arm) movement intention decoding is an important branch of MRCP-based movement intention decoding. Existing studies on upper limb movement intention decoding include movement parameters decoding (e.g., direction (Robinson et al., 2013; Chouhan et al., 2018), position (Hammon et al., 2008; Sosnik and Zheng, 2021), velocity (Robinson et al., 2013; Ubeda et al., 2017; Korik et al., 2018), acceleration (Bradberry et al., 2009), handgrip force (Haddix et al., 2021) and movement type recognition (Ofner et al., 2019). Reviewing existing studies about upper limb movement decoding, we find that most existing studies are concentrated on single hand (or arm) movement decoding. However, for the practical application of human augmentation, both-hand movement is common. Considering this issue, in 2020, Schwarz et al. (2020) first used the low-frequency EEG features to discriminate unimanual and bimanual daily reach-and-grasp movement types and achieved a multi-class classification accuracy of 38.6% for a combination of one rest and six movement types. Furthermore, to put the single-hand and both-hand movement intention decoding from EEG signals into an active human augmentation system, in 2020, we investigated the neural signatures and classification of single-hand and both-hand movement directions, and the 6-class classification achieved a peak accuracy of 70.29% (Wang et al., 2021).

It should be noted that both studies by Schwarz et al. (2020) and Wang et al. (2021) are focused on the discrimination of single-hand and both-hand movement. However, it is not enough to discriminate single-hand movement from both-hand movement. In many both-hand movement cases, we value the primary hand movement (e.g., the movement direction, velocity, or trajectory of single right hand) instead of whether we move one hand or both hands. Thus, it is necessary to decode the primary hand movement under the opposite hand movement condition. To solve the problem, in this study, we stride the first step by investigating the decoding of the movement direction of the primary hand (i.e., right hand in this paper) from EEG signals recorded during the opposite hand (i.e., left hand in this paper) movement. Notably, in this paper, we define the movement condition with the opposite hand movement as "W-OHM."

The contribution of this paper is that it is the first work to investigate the neural signatures and decoding of primary hand movement direction from EEG signals under the opposite hand movement and propose a novel decoding method based on non-linear dynamics parameters of MRCPs. This work not only can lay a foundation for the future development of BCI-based human augmentation systems for upper limbs impaired patients and healthy people, but it also may open a new avenue to decode other hand movement parameters (e.g., velocity and position) from EEG signals.

The remainder of the paper is structured as follows: section Methods introduces the methods. Section Results shows the results. Section Discussion and Conclusion presents the discussion, limitations of our work, and future work.

# METHODS

## Experimental Paradigm and Procedure

We recruited and measured 14 participants (one female), aged between 22 and 27 years. They reported having normal vision and no brain diseases. According to the Hand-Dominance-Test, they were all confirmed to be right-handed (Bryden, 1977). The study adhered to the principles of the 2013 Declaration of Helsinki. The research was approved by the local research ethics committee. All data were recorded at the IHMS Lab of the School of Mechanical Engineering, Beijing Institute of Technology, China. Subjects were seated on a chair in a room free of noise and electromagnetic interference. In front of them, there was a monitor for experimental instructions. **Figure 1** shows the experimental protocol.

Considering that all subjects were right-handed, we regarded the right-hand movement as the primary movement to be decoded and the left-hand movement as the opposite hand movement. For the primary movement task, all subjects were required to move their right hands in right or left directions. We preliminarily set the opposite hand movement in the vertical directions rather than horizontal directions. All subjects were asked to move their left hands in forward and backward directions. The movement of both hands was restricted in the horizontal plane parallel to the desktop. We defined the movement of right-hand in the right or left direction as "*R*" or "*L*" and the movement of left-hand in the forward or backward direction as "*F*" or "*B*." As shown in **Figure 1A**, on the monitor, two solid blocks colored as red and green correspond to the movement cue of left and right hands, respectively. When one trial was initiated, the red block would randomly appear in the *F* or *B* directions, and the green block would randomly appear in the *L* or *R* directions. That means, after 0 s (movement-cue onset), subjects were indicated for the movement directions and prepared for the movement. At the fourth second, both blocks changed from the solid into hollow, which were regarded as go-cue. Immediately, subjects were required to move both hands from the initial center to target positions appointed by green and red blocks. The movement tasks must be completed before the 7th second. After the 7th second, both hands were required to move back to the initial center position. At the 11th second, one trial ended.

During the experiment, the gaze of subjects was asked to fix on the screen to avoid eye movement. The experiment was composed of four sessions, including the right-hand movement in the *R* or *L* direction with the left-hand movement in the *F* or *B* direction. One session consisted of five runs, and each run consisted of 16 trials. In total, we recorded 80 trials per session. It meant that, for each combination of directions, there were 80

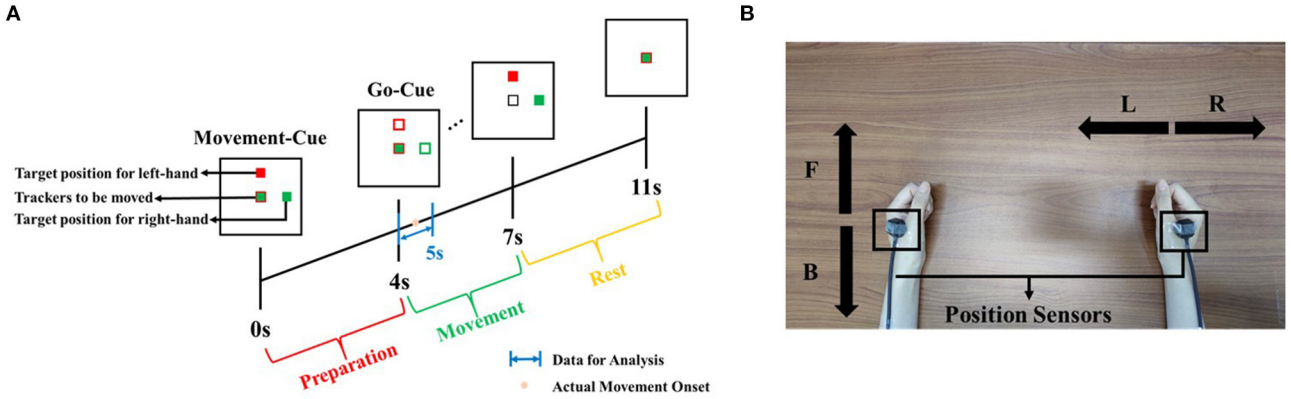



**FIGURE 1 |** Experimental protocol. **(A)** Timeline of experiment setup. Note that time 4 s refers to the Go-Cue, and time 4.5 s is the actual movement onset. **(B)** Illustration of both hand movement directions. Note that "F" and "B" refer to the movement of left-hand in the forward or backward direction, respectively, and "R" and "L" refer to the movement of right-hand in the right and left direction, respectively.

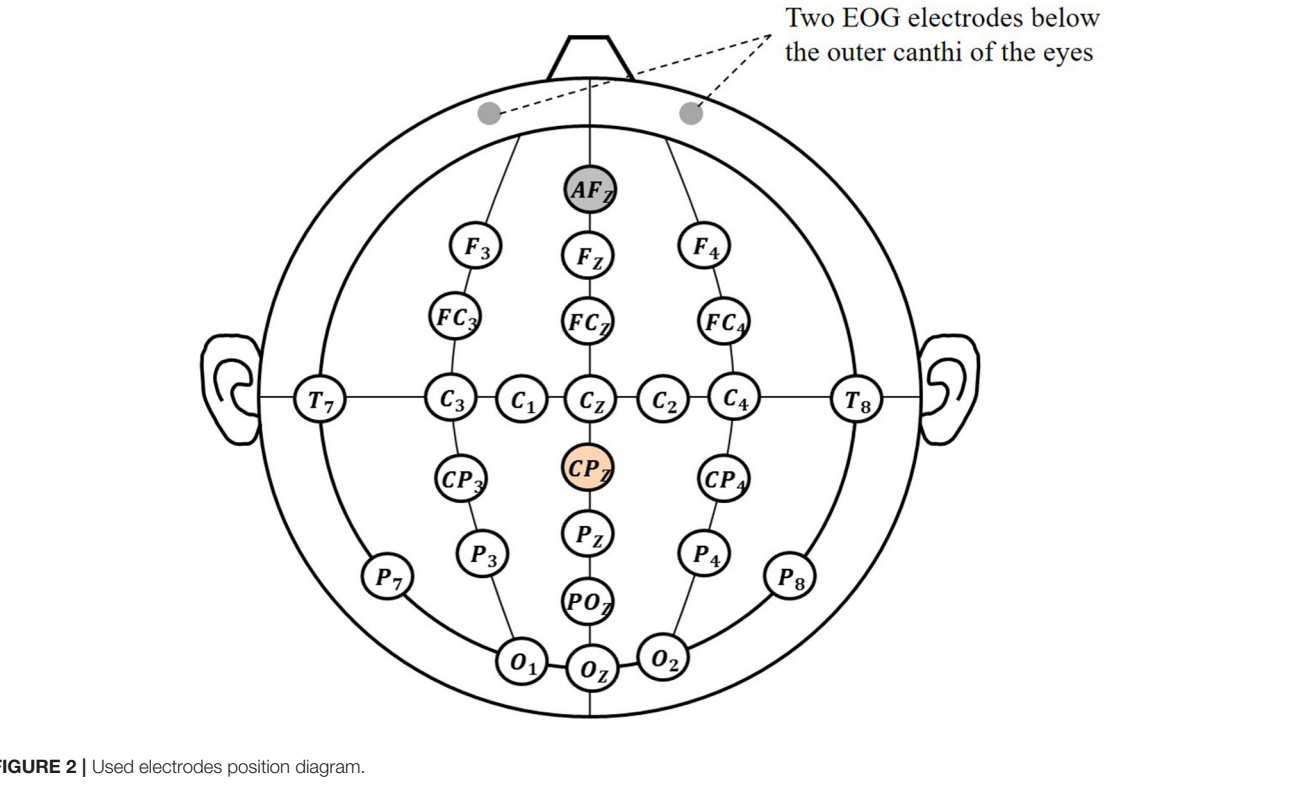



**FIGURE 2 |** Used electrodes position diagram.

trials uniformly. Between each run, the subjects were asked to perform a break of 2 min.

## Experimental Paradigm and Procedure

EEG signals were recorded by using a 64-electrode portable wireless EEG amplifier (NeuSen.W64, Neuracle, China), located at the following positions (according to the international 10–20 system): Cz, C1, C2, C3, C4, Fz, F3, F4, FCz, FC3, FC4, CP3, CP4, Oz, O1, O1, T7, T8, POz, Pz, P3, P4, P7, P8 (Wang et al., 2021), as shown in **Figure 2**. The selected electrodes involved the frontal, central, parietal, and occipital regions, which were related to the cognition, motion, perception function. The reference electrode was placed at CPz, and the ground electrode was placed at AFz. Electrooculogram (EOG) signals were recorded from two electrodes located below the outer canthi of the eyes. Two position-detecting sensors (FASTRACK) were positioned at tiger positions of both hands to track hands movement in real-time. The sampling rate of EEG signals was 1,000 Hz, and the sampling rate of position sensors was 60 Hz.

**TABLE 1 |** EEG signals decoding algorithm steps.

| EEG signals decoding algorithm | |
|---|---|
| Step 1: | Down-sample EEG signals to 100 Hz and re-reference by binaural electrodes; |
| Step 2: | Baseline correction and common average reference for EEG signals; |
| Step 3: | Eye movement artifacts rejection by independent component analysis; |
| Step 4: | Band-pass filter in [0.01, 4] Hz by using fast Fourier transform filter; |
| Step 5: | Apply z-score for EEG signals normalization; |
| Step 6: | Extract features from EEG signals by using echo state network (ESN); |
| Step 7: | Reshape ESN features to one-dimension and use PCA for feature dimension reduction; |
| Step 8: | Apply LDA for primary hand movement directions decoding. |

EEG, EOG, and position data were processed in MATLAB R2019b. The algorithm in steps for EEG signals-based primary hand movement direction decoding is listed in **Table 1**. For the signals preprocessing, EEG and EOG data were first down-sampled to 100 Hz, and each channel signal of EEG was re-referenced by subtracting the average of binaural electrodes. Baseline correction was used to eliminate the baseline drift for both EEG and EOG signals and common average reference was applied to remove common background noise for EEG signals. Eye movement artifacts were removed by using independent component analysis (ICA). The specific steps are as follows: (1) decomposing EEG signals into dependent component by applying independent component transform; (2) computing the correlation coefficients between the independent component and EOG signals; (3) rejecting the component whose correlation coefficient exceeds 0.4; (4) applying the inverse transformation on the remained component into EEG signals.

## Movement Related Cortical Potential (MRCP)

To correlate the neural activity during movement preparation and execution, MRCPs were extracted from EEG signals in the low-frequency band. After signal preprocessing, a zero-phase, 4th order Butterworth filter was used to filter EEG signals in the low-frequency band [0.01, 4] Hz. The weighted average filter was applied for electrode Cz to remove the spatial common background noise (Liu et al., 2018). To observe the difference of MRCPs for right-hand movement direction decoding, the MRCPs were obtained under condition of W-OHM for right-hand movement in *L* and *R* directions. The MRCPs under condition of W-OHM were average across all subjects.

## Feature Extraction

After signal preprocessing, a fast Fourier transform (FFT) filter was used to filter EEG signals in the frequency band [0.01, 4] Hz (Wang et al., 2010). For the primary hand movement decoding, an echo state network (ESN) was used to extract non-linear dynamics of EEG signals as the classification feature (in short ESN feature) (Sun et al., 2019).

As shown in **Figure 3A**, ESN is composed of the input layer, reservoir (hidden layer), and readout layer. The connecting matrix from the input to reservoir layers is defined as $W_{in}$. The internal connection matrix of the reservoir is sparse, and is defined as $W_{NN}$. Both $W_{in}$ and $W_{NN}$ are randomly initialized and kept invariant during network updating. The connecting matrix from the reservoir to readout layer is defined as $W_{out}$, and it is updated with the input and output data. With $W_{in}$, the ESN maps the input signals from the low-dimensional space into the high-dimensional non-linear space:

$$x(n) = f\bullet(W_{in}u(n) + W_{nn}x(n-1)), \tag{1}$$

$f\bullet$ was set to be $\tanh\bullet$ to realize the non-linearity of the network. In the high-dimensional non-linear space, the ESN model trains the $W_{out}$ by linear regression (e.g. Ridge Regression).

$$W_{out} = Y_t X^T (XX^T + \lambda_r I)^{-1}, \tag{2}$$

where $\lambda_r$ is the readout regularization coefficient.

As the core part of the ESN, the reservoir layer has the following parameters: (1) sparsely connecting with the sparse degree *c*, (2) reservoir size (i.e., the number of neurons) *NN*, (3) spectral radius $\rho$ (usually $\rho < 1$, to ensure that the effects of input and reservoir states on network vanishing after enough time), (4) the output of the reservoir layer at the current time $x(n)$, and (5) internal connection matrix $W_{NN}$ (randomly initialized and kept invariant during network updating). With the enormous and sparse reservoir layer, the ESN could capture the dynamics of a non-linear system. As mentioned in Waldert et al. (2008), EEG signals are non-steady and non-linear. From this perspective, we made a hypothesis that using the proposed method to establish the movement decoding model could obtain well-decoding performance.

In this paper, the multi-channel time-domain signals at the current time point were used as input signals, and the multi-channel time-domain signals at the next time point were used as output signals (as shown in **Figure 3B**). The output connection matrix $W_{out}$, which could reflect the non-linear dynamics of EEG signals over time, was chosen as the ESN features for decoding. In addition to the parameters c and NN that have a major influence on the ESN performance and were determined in the subsequent training (by using mesh grid search), we empirically set the residual parameters: (1) $\rho = 0.98$; (2) $x(0)$ was zero-matrix; (3) $\lambda_r = 1 \times 10^{-4}$. Besides, before encoding EEG signals to ESN features, z-score was applied for normalization, as follows

$$Z = \frac{X - \mu}{\sigma}, \tag{3}$$

where $X$ is the raw EEG signals before normalization, $\mu$ and $\sigma$ are the mean and standard deviation of EEG signals,

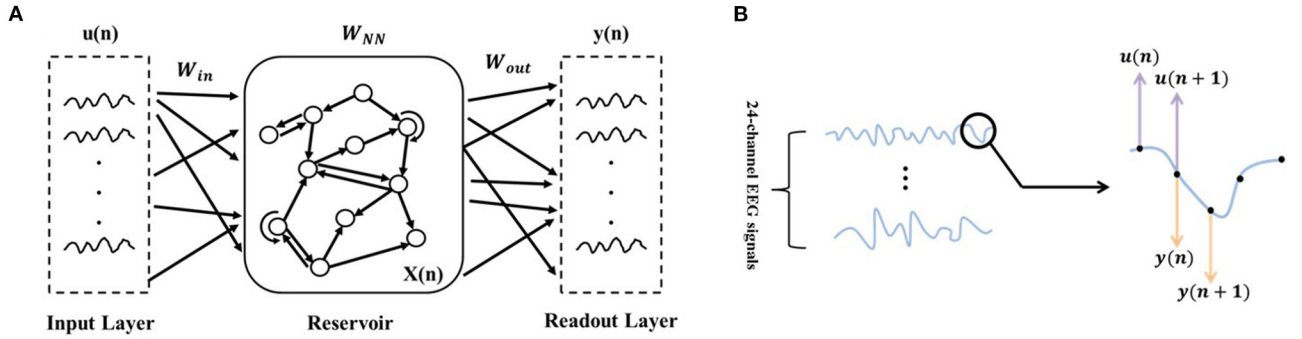



**FIGURE 3 |** ESN schematic diagram. **(A)** Network structure diagram. **(B)** Illustration of the input and output data selection based on the time series of EEG signals. u(·) refers to the input data and y(·) refers to the output data.

respectively. The original feature dimensions could be calculated by the following equation,

$$Num_F = C \bullet (NN + C + 1), \quad (4)$$

where $Num_F$ is the original feature dimension, $NN$ is the reservoir size, $C$ is the channel number. To suppress feature redundancy and accelerate computation, principal component analysis (PCA) was applied to reduce feature dimension. Choosing the principal component with the percentage above 99%, the dimension of ESN feature was reduced to 40.

## Classification

The fixed window [0, 1] s of the Go-Cue (i.e., [−0.5 0.5] s of the actual movement onset, calibrated by *FASTRACK*) was used for the primary hand movement direction decoding. Linear discriminant analysis (LDA) classifier was performed to decode the primary hand movement direction. The classification accuracy was used to measure the decoding performance, and the decoding accuracy was calculated by dividing the number of correctly classified test samples by the total number of the test samples. Mean classification accuracy was calculated by a 5 × 5 cross-validation. For the primary hand movement direction decoding under W-OHM, the classification accuracy was first calculated separately for the opposite hand movement in *F* or *B* direction and then averaged.

## Statistics

Power tables from Cohen were used to evaluate the number of participants needed to obtain a significant result (Puce et al., 2003). When 14 participants were involved in this experiment, partial eta squared ($R^2$) was calculated as 0.417 by using ANOVA in IBM SPSS Statistics 25. Effect size for F-ratios was calculated as follows:

$$f^2 = \frac{R^2}{1 - R^2} \quad (5)$$

When $f^2$ is 0.7153, the equivalent effect size d is 1.6. At the given two-tailed $\alpha = 0.05$ and the recommended power level of 80%, the number of participants needed for significant results was 9, which justified the sufficiency of subjects in our experiment.

# RESULTS

## Neural Signatures

(1) Movement related cortical potential: **Figure 4** shows the MRCPs at electrode Cz under the condition of W-OHM. Considering that the primary purpose of this study was to decode right-hand movement directions, the MRCPs associated with the right-hand movement in L and R directions were presented. The MRCPs were calculated from −1.5 to 1.5 s of the Go-Cue and averaged across all subjects. As shown in **Figure 4**, under all movement conditions, the amplitudes of MRCPs kept steady around 0 μV from −2 to 0 s, which was the movement preparation period. A positive peak was observed at around 300 ms, and after that a substantial negative shift arose and peaked at about 500 ms. The peak time of the negative shift was in agreement with the actual movement onset calibrated by FASTRACK (as labeled in **Figure 1**). For the movement under condition of W-OHM, the average negative shift maximums of the MRCPs for right-hand movements in R and L directions were −9.4153 and −10.4324 μV, respectively. By comparing the negative shift amplitudes of MRCPs between two primary hand movement directions, larger negative shift amplitude of the primary hand movement in L direction was found. However, this difference was not significant (Wilcoxon signed-rank test, $p = 0.17$). Furthermore, Wilcoxon signed-rank test showed that there was a significant difference between the MRCPs (from −1.5 to 1.5 s) associated with two primary hand movement directions ($p < 0.01$).

(2) Time-Frequency plots: **Figure 5** presents the grand average time-frequency plots in the time period [−1.5, 1.5] s of the movement cue onset across all subjects. It was seen that a prominent increment in spectral power appeared after the movement cue onset in the low frequencies of smaller than 7 Hz (especially smaller than 4 Hz), indicating that main power modulations during bimanual movement was centralized in the low frequency band, and this result was similar to the finding in Robinson et al. (2015).

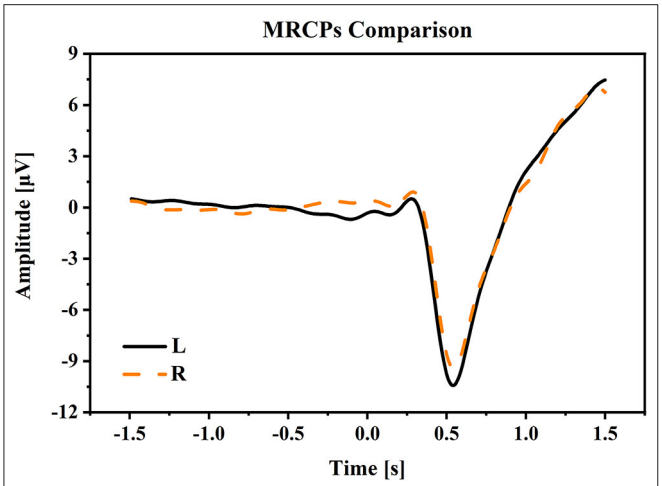



**FIGURE 4 |** The averaged MRCPs at electrode Cz during the time period [−1.5, 1.5] s of Go-Cue. "L" and "R" refer to the right-hand movement in right and left directions. Note that time 0 s is the Go-Cue, and time 0.5 s is the actual movement onset.

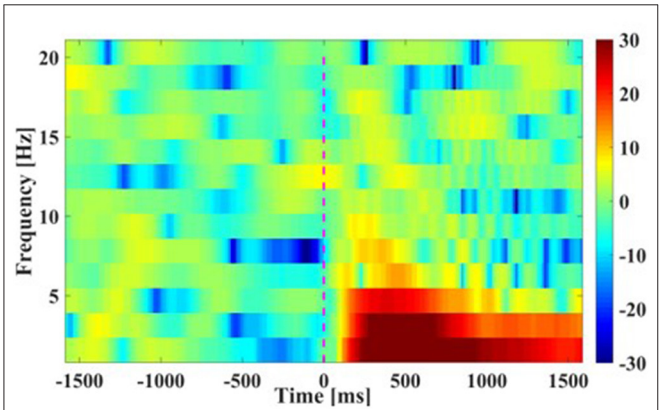



**FIGURE 5 |** Time-Frequency plots of movement. The averaged results are in frequencies [0 16] Hz at Cz channel from −1.5 to 1.5 s of the movement cue onset. Note that time 0 s is the Go-Cue, which indicates the movement execution, and time 0.5 s is the actual movement onset.

(3) Scalp topographical maps: The average EEG potential topographical distributions of the primary hand movements under the condition of W-OHM are shown in **Figure 6**. The scalp topographical maps were plotted from −1,000 to 1,500 ms with an interval of 500 ms in between. It was seen that cortical brain activities were steady from −1,000 to 0 ms with no specific modulation patterns. A significant decline in EEG potentials on central regions and a significant increment on occipital regions occurred from 0 to 500 ms. After 500 ms, peaks of these plots were centralized on central regions increasingly. Furthermore, the potential of EEG signals on temporal lobes turned into a negative shift and reached the negative maximum gradually, which was in line with the finding in Puce et al. (2003).

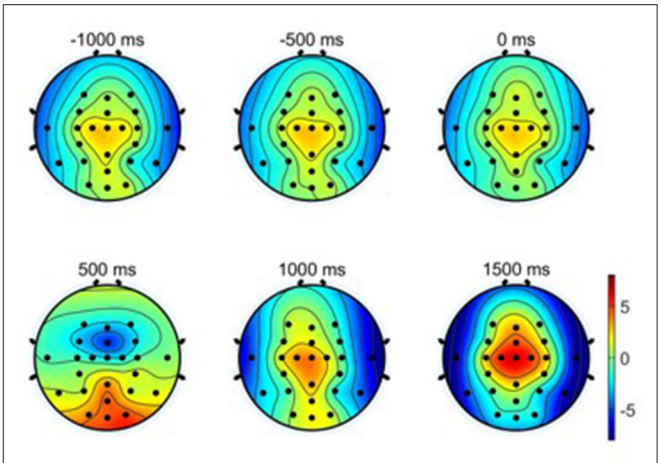



**FIGURE 6 |** Averaged EEG potential topographic maps of hand movement across all subjects. Note that time 0 s is the Go-Cue, which indicates the movement execution, and time 0.5 s is the actual movement onset.

## Parameters Selection

The parameters of reservoir sparse degree c and reservoir size *NN* were critical to the performance of the proposed decoding model. The reservoir sparse degree is related to the number of neurons activated, and the reservoir size is associated with the complexity of the proposed model. Only with befitting parameters, the proposed model could capture the dynamics of EEG signals well. In this study, we used the mesh grid search to determine well-behaved subject-specific parameters *c* and *NN*. For determining the reservoir sparse degree *c*, the step size was set to be 0.1, and the search range was in [0.1, 0.9]. For determining the reservoir size *NN*, the search set was {10, 20, 30, 40, 50, 60, 70} (Sun et al., 2019).

**Table 2** shows the subject-specific decoding accuracies and parameters selected (*NN* and *c*) by using ESN-based models under condition of W-OHM. **Figure 7** shows the example of the decoding accuracy of Subject 1 under condition of W-OHM against the reservoir sparse degree *c* and reservoir size *NN*. It was seen that, with the increase of the reservoir sparse degree *c*, the variation of decoding accuracy was slight for each reservoir size *NN*. Furthermore, with the increment of the reservoir size NN, the decoding accuracy was gradually improved and tended to be steady. The parameter combination with the best performance, i.e., "*c* 0.4, *NN* 60," was selected for Subject 1 under condition of W-OHM.

## Decoding Performance Comparison

In this study, the classification performance of the proposed ESN model was compared with two models in (Wang et al., 2021), and we named two comparison models as Model 1 and Model 2 in this study. Specifically, Model 1 used potential amplitudes of EEG signals as feature and used LDA as classifier, and Model 2 used the sum of spectral power of EEG signals as feature and used LDA as classifier. For both model 1 and model 2, no personalized parameters tuning strategy was applicable.

**TABLE 2 |** Subject-specific decoding accuracies and parameters selected (*NN* and *c*) under W-OHM by using the proposed model.

| Subject No. | acc [%] | NN | c |
|---|---|---|---|
| S1 | 91.33 | 60 | 0.4 |
| S2 | 92.44 | 70 | 0.7 |
| S3 | 96.94 | 60 | 0.5 |
| S4 | 91.81 | 50 | 0.1 |
| S5 | 89.88 | 60 | 0.7 |
| S6 | 97.69 | 70 | 0.3 |
| S7 | 96.88 | 50 | 0.7 |
| S8 | 86.56 | 70 | 0.3 |
| S9 | 87.37 | 70 | 0.5 |
| S10 | 82.75 | 60 | 0.4 |
| S11 | 89.56 | 70 | 0.7 |
| S12 | 84.25 | 70 | 0.6 |
| S13 | 75.00 | 60 | 0.4 |
| S14 | 90.19 | 70 | 0.9 |
| Mean ± Std | 89.48 ± 5.92 | 63.57 ± 7.18 | 0.51 ± 0.21 |

**TABLE 3 |** Decoding accuracies across subjects under condition of W-OHM using different kinds of models.

| Subject No. | Model 1 [%] | Model 2 [%] | Proposed Model [%] |
|---|---|---|---|
| S1 | 86.88 | 75.56 | 91.33 |
| S2 | 81.25 | 77.69 | 92.44 |
| S3 | 80.81 | 88.25 | 96.94 |
| S4 | 71.56 | 70.19 | 91.81 |
| S5 | 80.93 | 70.63 | 89.88 |
| S6 | 89.25 | 77.10 | 97.69 |
| S7 | 92.94 | 80.38 | 96.88 |
| S8 | 74.56 | 70.81 | 86.56 |
| S9 | 89.88 | 76.06 | 87.37 |
| S10 | 77.13 | 73.31 | 82.75 |
| S11 | 86.75 | 67.69 | 89.56 |
| S12 | 86.06 | 82.75 | 84.25 |
| S13 | 68.50 | 63.56 | 75.00 |
| S14 | 85.44 | 75.81 | 90.19 |
| Mean ± Std | 82.28 ± 6.98 | 74.99 ± 6.13 | 89.48 ± 5.92 |

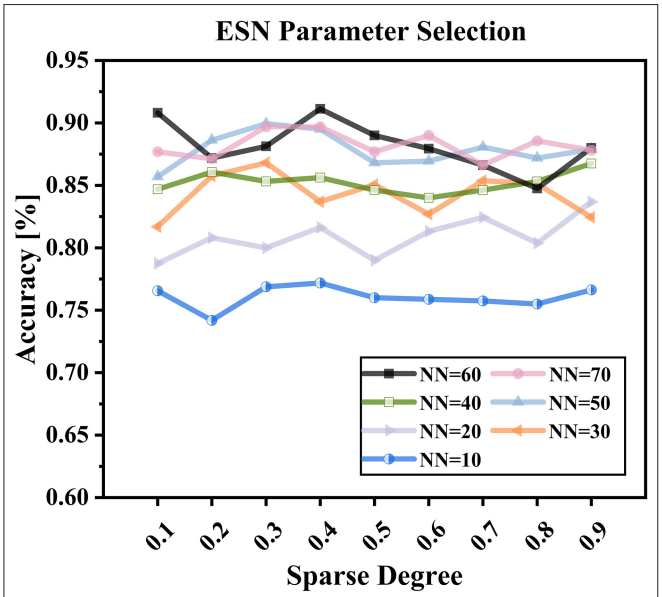



**FIGURE 7 |** ESN parameter selection with different reservoir sparse degree *c* and reservoir size *NN*.

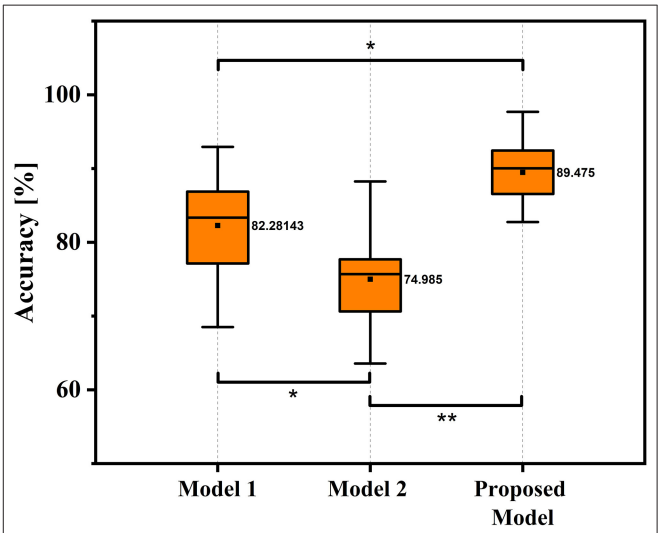



**FIGURE 8 |** Box-Plots of average decoding accuracy under W-OHM by using three kinds of models. The asterisk marks significant differences. Tukey-Kramer *post hoc* test: $^{*}p < 0.05$, $^{**}p < 0.01$.

**Table 3** shows the decoding accuracy comparison results based on different kinds of classification models under condition of W-OHM. **Figure 8** shows the decoding accuracy comparison under condition of W-OHM among three kinds of models in the box-plot form. As shown in **Table 3**, the highest average decoding accuracy was obtained when using the proposed model, and was 89.48 ± 5.92%. Correspondingly, when using Model 1 and Model 2, the decoding accuracies were 82.28 ± 6.98% and 74.99 ± 6.13%, respectively. Significant differences were found between classification models by performing one-factor analysis of variance [$F_{(2,39)} = 16.88$, $p < 0.01$]. The *post-hoc* pairwise comparison with the Tukey-Kramer method showed that there were significant differences between the Model 1 and Model 2 ($p = 0.02$), the Model 1 and the proposed model ($p = 0.02$), and the Model 2 and the proposed model ($p < 0.01$), as shown in **Figure 8**.

## Computational Time Comparison

**Table 4** shows the computational time comparison results of different decoding models. The total computational time included the sum of signal processing, feature extraction, dimensionality reduction, and classification of a single sample. For the proposed model, averaged *NN* and *c* (64 and 0.5) calculated from **Table 2** was used for calculating computational

**TABLE 4 |** Computational time of different decoding models.

| Computational Time [ms] | | |
|---|---|---|
| **Model 1** | **Model 2** | **Proposed Model** |
| 13.1 | 7.4 | 37.5 |

time. As shown in **Table 4**, the computational time of the Model 1, Model 2, and proposed decoding models was 13.1, 7.4, and 37.5 ms, respectively. It showed the feasibility of putting the proposed decoding model into real-time detection.

## DISCUSSION AND CONCLUSION

This paper explored using EEG signals to decode primary hand movement direction under the opposite hand movement. MRCPs, time-frequency plots, and scalp topographic maps were shown for neural signatures. The decoding model was built by using an ESN to extract non-linear dynamics parameters of MRCPs as decoding features. Experimental results showed that the proposed method performed well in decoding primary hand movement directions with the opposite hand movement. This paper is the first to investigate neural signatures and decoding of hand movement parameters under the opposite hand movement.

In this study, we followed the classic center-out paradigm (Robinson et al., 2013, 2015; Chouhan et al., 2018), and evolved it to both-hand center-out paradigm for the both-movement decoding. Specifically, we set the primary hand movement in L or R directions and the opposite hand movement in F or B directions. Considering that two hands often move in different directions in practical human-machine collaboration systems, we preliminarily set the primary hand and opposite hand movement in orthogonal directions. The advantages of this paradigm were that it was basic and representable for hand movement decoding and its experiment results were general and could be extended to practical hand movement decoding problem. By comparing the negative shift amplitudes of MRCPs between two primary hand movement directions, a larger negative shift amplitude of the primary hand movement in *L* direction was found (*L*: 10.4324 vs. *R*: 9.4153 $\mu$V). This result was in accord with our previous study, which indicated that the larger negative shift amplitude of MRCP might be related to the higher torque-level for the leftward motion of the right arm (Wang et al., 2021). The increment of spectrum energy was mainly centralized in the low-frequency band, which was similar to the finding in Waldert et al. (2008), which indicated that hand movement directions could be decoded from power modulations in the low-frequency band. The increment of event-related potentials (ERPs) on central regions and the decrement of ERPs on temporal regions were found in scalp topographic maps from 500 to 1,500 ms, which was in line with the findings in Puce et al. (2003) and (Wang et al., 2021), respectively.

Experimental results showed that the proposed decoding model outperformed the models used in (Wang et al., 2021 (89.48% vs. 82.28% or 74.99%). One main reason for the results is likely that the proposed method could capture more discriminable information of MRCPs for decoding hand movement direction. This ability of the proposed model may be because ESN can establish a complex non-linear dynamic system of EEG signals with a large reservoir size and complex transmission relationships between neurons and can constantly update the network parameters according to the information from the previous moment. Furthermore, compared with other neural network (e.g., convolution neural network and deep belief network), ESN, as one kind of recurrent neural network, could capture the nonstationary and nonlinear features and is good at dealing with the time sequence problem.

This work has values in at least two implications. First, the proposed method can capture more meaningful non-linear information of MRCPs for decoding hand movement direction. Thus, this work may open a new avenue to decode other hand movement parameters, such as velocity and trajectory. Second, since, for human augmentation, many tasks need to be carried out by the movement of both hands, these findings can lay a foundation for the future development and use of human augmentation systems based on hand movement decoding from EEG signals.

However, at least three limits exist in this work. First, although the proposed decoding method of primary hand movement direction under the opposite hand movement performed well, the movement intensity of the left hand was kept at a certain level. For further exploration of the decoding of primary hand movement under the opposite hand movement, different kinds and intensities of the opposite hand movement, including more natural and complex movement, should be considered. Second, like many studies in the field of using EEG signals to decode hand movement (Robinson et al., 2015; Chouhan et al., 2018; Schwarz et al., 2020), we used able-bodied subjects to investigate neural signatures and decoding of hand movement direction. However, it is unclear whether these results can be extended to persons with disabilities. Thus, more subjects, especially the target users (including the disabled), should be applied to validate these findings further. Third, in this study, all recruited participants were right-handed, and 1 female among which was recruited. Though the influence of handedness and gender were focused on in this paper, handedness and gender may be the factors that influenced the decoding of primary hand movement under opposite hand movement, which could be explored in future.

Our future work will be dedicated to solving the weaknesses mentioned above, including using more types of hand movement directions given more types and intensities of the opposite hand movements, using more subjects and even some persons with motion impairment and exploring the influence of handedness and gender on decoding.

## 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 Beijing Institute of Technology Research Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

In particular, JW, LB, and WF contributed to the design of this work. JW: methodology, validation, formal analysis, and writing—original draft. LB: conceptualization, resources, writing—review and editing, and funding acquisition. WF: software and data curation. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported in part by National Natural Science Foundation of China under Grant (No. 51975052).

## ACKNOWLEDGMENTS

The authors would like to thank all the subjects for volunteering to participate in our experiments.

## REFERENCES

Bradberry, T. J., Gentili, R. J., and Contreras-Vidal, J. L. (2009). Decoding three-dimensional hand kinematics from electroencephalographic signals. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2009, 5010–5013. doi: 10.1109/IEMBS.2009.5334606

Bryden, M. P. (1977). Measuring handedness with questionnaires. *Neuropsychologia* 15, 617–624. doi: 10.1016/0028-3932(77)90067-7

Chen, L., Chen, P., Zhao, S., Luo, Z., Chen, W., 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, S., Chen, Z., Yao, B., Zhu, X., Zhu, S., Wang, Q., et al. (2017). Adaptive robust cascade force control of 1-dof hydraulic exoskeleton for human performance augmentation. *IEEE ASME Trans. Mechatronics.* 22, 589–600. doi: 10.1109/TMECH.2016.2614987

Chouhan, T., Robinson, N., Vinod, A. P., Ang, K. K., and Guan, C. (2018). Wavlet phase-locking based binary classification of hand movement directions from EEG. *J. Neural Eng.* 15, 066008. doi: 10.1088/1741-2552/aadeed

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

Gao, Z., Dang, W., Liu, M., Guo, W., Ma, K., and Chen, G. (2021). Classification of EEG signals on VEP-based BCI systems with broad learning. *IEEE Trans. Syst. Man Cybern. Syst.* 51, 7143–7151. doi: 10.1109/TSMC.2020.2964684

Haddix, C., Al-Bakri, A. F., and Sunderam, S. (2021). Prediction of isometric handgrip force from graded event-related desynchronization of the sensorimotor rhythm. *J. Neural Eng.* 18, 056033. doi: 10.1088/1741-2552/ac23c0

Hammon, P. S., Makeig, S., Poizner, H., Todorov, E., and de Sa, V. R. (2008). Predicting reaching targets from human *EEG. IEEE Signal Process. Mag.* 25, 69–77. doi: 10.1109/msp.2008.4408443

He, L., Hu, D., Wan, M., Wen, Y., von Deneen, K. M., and Zhou, M. (2016). Common Bayesian network for classification of EEG-based multiclass motor imagery BCI. *IEEE Trans. Syst. Man Cybern. Syst.* 46, 843–854. doi: 10.1109/TSMC.2015.2450680

Kansaku, K., Hata, N., and Takano, K. (2010). My thoughts through a robot's eyes: an augmented reality-brain-machine interface. *Neurosci. Res.* 66, 219–222. doi: 10.1016/j.neures.2009.10.006

Korik, A., Sosnik, R., Siddique, N., and Coyle, D. (2018). Decoding imagined 3D hand movement trajectories from EEG: evidence to support the use of Mu, Beta, and low Gamma oscillations. *Front. Neurosci.* 12, 130. doi: 10.3389/fnins.2018.00130

Kvansakul, J., Hamilton, L., Ayton, L. N., McCarthy, C., and Petoe, M. A. (2020). Sensory augmentation to aid training with retinal prostheses. *J. Neural Eng.* 17, 045001. doi: 10.1088/1741-2552/ab9e1d

Kwak, N.-S., Mueller, 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

Liu, D., Chen, W., Lee, K., Chavarriaga, R., Iwane, F., Bouri, M., et al. (2018). EEG-Based lower-limb movement onset decoding: continuous classification and asynchronous detection. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 1626–1635. doi: 10.1109/TNSRE.2018.2855053

Ofner, P., Schwarz, A., Pereira, J., Wyss, D., Wildburger, R., and Mueller-Putz, G. R. (2019). Attempted arm and hand movements can be decoded from low-frequency EEG from persons with spinal cord injury. *Sci. Rep.* 9, 7134. doi: 10.1038/s41598-019-43594-9

Pei, Y., Luo, Z., Zhao, H., Xu, D., Li, W., Yan, Y., et al. (2022). A tensor-based frequency features combination method for brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 30, 465–475. doi: 10.1109/TNSRE.2021.3125386

Puce, A., Syngeniotis, A., Thompson, J. C., Abbott, D. F., Wheaton, K. J., and Castiello, U. (2003). The human temporal lobe integrates facial form and motion: evidence from fMRI and ERP studies. *Neuroimage.* 19, 861–869. doi: 10.1016/S1053-8119(03)00189-7

Robinson, N., Guan, C., and Vinod, A. P. (2015). Adaptive estimation of hand movement trajectory in an EEG based brain-computer interface system. *J. Neural Eng.* 12, 066019. doi: 10.1088/1741-2560/12/6/066019

Robinson, N., Vinod, A. P., Ang, K. K., Tee, K. P., and Guan, C. T. (2013). EEG-based classification of fast and slow hand movements using wavelet-CSP algorithm. *IEEE Trans. Biomed. Eng.* 60, 2123–2132. doi: 10.1109/TBME.2013.2248153

Schwarz, A., Hoeller, M. K., Pereira, J., Ofner, P., and Mueller-Putz, G. R. (2020a). Decoding hand movements from human EEG to control a robotic arm in a simulation environment. *J. Neural Eng.* 17, 036010. doi: 10.1088/1741-2552/ab882e

Schwarz, A., Pereira, J., Kobler, R., and Muller-Putz, G. R. (2020b). Unimanual and bimanual reach-and-grasp actions can be decoded from human EEG. *IEEE Trans. Biomed. Eng.* 67, 1684–1695. doi: 10.1109/TBME.2019.2942974

Sosnik, R., and Zheng, L. (2021). Reconstruction of hand, elbow and shoulder actual and imagined trajectories in 3D space using EEG current source dipoles. *J. Neural Eng.* 18, 056011. doi: 10.1088/1741-2552/abf0d7

Sun, L., Jin, B., Yang, H., Tong, J., Liu, C., and Xiong, H. (2019). Unsupervised EEG feature extraction based on echo state network. *Inf. Sci.* 475, 1–17. doi: 10.1016/j.ins.2018.09.057

Ubeda, A., Azorin, J. M., Chavarriaga, R., and Millan, J., d. R. (2017). Classification of upper limb center-out reaching tasks by means of EEG-based continuous decoding techniques. *J. Neuroeng. Rehabil.* 14, 9. doi: 10.1186/s12984-017-0219-0

Waldert, S., Preissl, H., Demandt, E., Braun, C., Birbaumer, N., Aertsen, A., et al. (2008). Hand movement direction decoded from MEG and EEG. *J. Neurosci.* 28, 1000–1008. doi: 10.1523/JNEUROSCI.5171-07.2008

Wang, J., Bi, L., Fei, W., and Guan, C. (2021). Decoding single-hand and both-hand movement directions from noninvasive neural signals. *IEEE Trans. Biomed. Eng.* 68, 1932–1940. doi: 10.1109/TBME.2020.3034112

Wang, S., Inkol, R., Rajan, S., and Patenaude, F. (2010). Detection of narrow-band signals through the FFT and polyphase FFT filter banks: noncoherent versus coherent integration. *IEEE Trans. Instrum. Meas.* 59, 1424–1438. doi: 10.1109/TIM.2009.2038294

Yandell, M. B., Tacca, J. R., and Zelik, K. E. (2019). Design of a low profile, unpowered ankle exoskeleton that fits under clothes: overcoming practical

barriers to widespread societal adoption. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 712–723. doi: 10.1109/TNSRE.2019.2904924

Yin, E., Zeyl, T., Saab, R., Chau, T., Hu, D., and Zhou, Z. (2015a). A hybrid brain-computer interface based on the fusion of P300 and SSVEP scores. *IEEE Trans. Neural Syst. Rehabil. Eng.* 23, 693–701. doi: 10.1109/TNSRE.2015.2403270

Yin, E., Zhou, Z., Jiang, J., Yu, Y., and Hu, D. (2015b). A dynamically optimized SSVEP Brain-Computer Interface (BCI) speller. *IEEE Trans. Biomed. Eng.* 62, 1447–1456. doi: 10.1109/TBME.2014.2320948

Zhang, R., Li, Y., Yan, Y., Zhang, H., Wu, S., Yu, T., et al. (2016). Control of a wheelchair in an indoor environment based on a brain-computer interface and automated navigation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 24, 128–139. doi: 10.1109/TNSRE.2015.2439298

Zhao, X., Chu, Y., Han, J., and Zhang, Z. (2016). SSVEP-based brain-computer interface controlled functional electrical stimulation system for upper extremity rehabilitation. *IEEE Trans. Syst. Man Cybern. Syst.* 46, 947–956. doi: 10.1109/TSMC.2016.2523762

**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.

frontiers | Frontiers in Neuroinformatics

ORIGINAL RESEARCH
published: 02 June 2022
doi: 10.3389/fninf.2022.914823

# SparNet: A Convolutional Neural Network for EEG Space-Frequency Feature Learning and Depression Discrimination

*Xin Deng, Xufeng Fan*, Xiangwei Lv and Kaiwei Sun*

*Key Laboratory of Data Engineering and Visual Computing, College of Computer Science and Technology, Chongqing University of Posts and Telecommunication, Chongqing, China*

Depression affects many people around the world today and is considered a global problem. Electroencephalogram (EEG) measurement is an appropriate way to understand the underlying mechanisms of major depressive disorder (MDD) to distinguish depression from normal control. With the development of deep learning methods, many researchers have adopted deep learning models to improve the classification accuracy of depression recognition. However, there are few studies on designing convolution filters for spatial and frequency domain feature learning in different brain regions. In this study, SparNet, a convolutional neural network composed of five parallel convolutional filters and the SENet, is proposed to learn EEG space-frequency domain characteristics and distinguish between depressive and normal control. The model is trained and tested by the cross-validation method of subject division. The results show that SparNet achieves a sensitivity of 95.07%, a specificity of 93.66%, and an accuracy of 94.37% in classification. Therefore, our results can conclude that the proposed SparNet model is effective in detecting depression using EEG signals. It also indicates that the combination of spatial information and frequency domain information is an effective way to identify patients with depression.

**Keywords: SENet, SparNet, space-frequency domain characteristics, depression, EEG**

OPEN ACCESS

***Edited by:***
*Duo Chen,*
*Nanjing University of Chinese Medicine, China*

***Reviewed by:***
*Zhenhu Liang,*
*Yanshan University, China*
*Mengfan Li,*
*Hebei University of Technology, China*

****Correspondence:***
*Xufeng Fan*
*xufengfan_c@163.com*

***Received:*** *07 April 2022*
***Accepted:*** *06 May 2022*
***Published:*** *02 June 2022*

***Citation:***
*Deng X, Fan X, Lv X and Sun K (2022) SparNet: A Convolutional Neural Network for EEG Space-Frequency Feature Learning and Depression Discrimination. Front. Neuroinform. 16:914823. doi: 10.3389/fninf.2022.914823*

## 1. INTRODUCTION

Major depressive disorder (MDD, also known as unipolar depression) is a physical disease of the brain, also known as a mental health disorder. It mainly affects the process of thought, behavior, and mood, and also can lead to the loss of interest and energy, interpersonal relationships, and job performance. According to the statistics of the World Health Organization, more than 300 million people in the world suffer from depression, and about 800,000 people die of depression every year (Belmaker and Agam, 2008; Olesen et al., 2012; Whiteford et al., 2013). Early and accurate diagnosis of depression is crucial for patients who need timely clinical treatment.

Most of the previous diagnoses of depression are based on the questionnaires as a judgment and screening tool. One of the major drawbacks of this method is that it requires experienced doctors. Therefore, finding a suitable and effective way to detect depression is an emerging research area. Currently, various physiological measurement tools are developing rapidly, such as functional magnetic resonance imaging (fMRI), electroencephalogram (EEG), and positron emission tomography (PET). Many studies attempt to

measure psychological data and develop auxiliary diagnostic methods in clinical practice (Van Der Stelt and Belger, 2007; Michel and Murray, 2012; Olbrich and Arns, 2013; Kerestes et al., 2014; de la Salle et al., 2016). The quantitative measurement of EEG signals is a neuroimaging technique with obvious practical advantages because it does not involve invasive manipulation, and it is easy to manage, well-tolerated and relatively inexpensive. In addition, the prevalence and persistence of depressive symptoms make scalp recording EEG an appropriate method to understand the underlying mechanisms of depression.

Most of the existing deep learning methods take the original data of EEG signals or transform them into frequency-domain signals as input, thus losing the spatial features between multiple brain regions and the whole brain. Liao et al. (2017) and Jiang et al. (2021) showed the effectiveness of spatial information in distinguishing depression. Considering that the irregular EEG network is one of the possible physiological symptoms of depression, the EEG activity has spatial characteristics originating from different brain regions, so the spatial EEG characteristics extracted from different brain regions can be used to identify depression. Cai et al. (2018) also found that the features in the frequency domain were more likely to distinguish patients than those in the time domain. By fusing these two features of comprehensive mixed information, it is expected to achieve richer and more accurate identification of depression. Although it is possible to explore the spatial information of the brain based on the internal structure of neural networks, there are seldom studies that have explored the spatial information of the various brain regions in depression based on the whole brain structure from EEG signals. The purpose of this study is to combine the spatial information and the frequency domain information of each brain region, and integrates them with deep learning.

The main contributions of this study are as follows. First, phase space reconstruction is used to denoise the EEG signals in the time domain and smooth the feature in the frequency domain. Second, a new model called SparNet is proposed to capture more specific information about depression in this study. It is a parallel convolutional network used to extract the features of different brain regions, and the attention mechanism module is added to the network. Third, the channels for each brain region are selected to explore the local spatial-frequency domain features. The frequency domain and the spatial features of each brain region are combined by the multi-layer parallel convolutional filter. By adding the attention mechanism, our deep learning model can assign the weights to different channels in the local brain region and also to different contribution degrees in the global brain region.

## 2. RELATED STUDY

All approaches to depression identification fall into two broad categories: those based on the manual features and those based on the raw data. Hosseinifard et al. (2013) used a large EEG recording dataset of 90 subjects (45 normal subjects and 45 depressed subjects) and found that in non-linear features, the correlation dimension is a powerful feature for analyzing EEG signals and identifying the depressed and the non-depressed subjects. Cai et al. (2018) through three-electrode channel acquisition, found that features were mainly concentrated in the frequency domain, and achieved the best accuracy of 79.27% with KNN. Their recent study, Cai et al. (2020b) compared KNN, DT, and SVM on the same data set. Their KNN model achieved the highest accuracy of 89.98%. Liao et al. (2017) proposed a nuclear feature filtering group common space Mode (KEFB-CSP) based on the scalp EEG signals. The signals were decomposed into each frequency band and then the spatial features were extracted by the CSP algorithm. Mumtaz et al. (2017) conducted the time-frequency decomposition of an EEG data and constructed EEG data matrix. Compared with other time-frequency methods such as STFT and EMD, the wavelet analysis has the highest classification accuracy of 87.5%. Mahato and Paul (2020) found that the average theta asymmetry of normal people was higher than that of patients with depressive people. In SVM, the classification accuracy of alpha2 and theta asymmetric combination is 88.33%. Peng et al. (2019) collected 128 electrodes of the subjects, and the research results showed that depression would affect the brain activity of almost the entire cerebral cortex, and the accuracy of 92.73% was achieved by using SVM and full frequency band features. Sun et al. (2020a) found that there were far more functional connections within hemispheres than between hemispheres. High frequency parietal occipital lobe plays an important role in depression recognition. They Sun et al. (2020b) achieved the highest classification accuracy of 82.31% by using the ReliefF feature selection method and LR classifier on the same data set. They further indicated that the functional connection feature plays an important role in depression recognition. Jiang et al. (2021) proposed an effective EEG based spatial classification detection method for depression, task-related common spatial pattern (TCSP), which significantly improved the accuracy of depression classification by using spatial information.

Although feature extraction and machine learning can effectively identify patients with depression, manual feature extraction and selection are required, which is time-consuming and laborious. There are many studies that use raw EEG data or pre-processed data as model input. Zhang et al. (2020) extracted the temporal and spatial characteristics of EEG signals by 1DCNN and added the population attention mechanism. They suggested that the combination of EEG signals and demographic factors could be better for patients with depression. Fan et al. (2020) combined CNN and LSTM to better extract time and space information. Ke et al. (2020) designed a lightweight CNN model for the online identification of patients with depression. Wan et al. (2020) proposed a convolutional neural network HybridEEGNet composed of two parallel lines for learning synchronization and regional EEG features. Seal et al. (2021) found that the right extreme value of the subjects with depression was significant, while the left extreme value of normal subjects was significant. Sharma et al. (2021) proposed a computer-aided (CAD) hybrid neural network based on EEG, that used CNN for time learning and LSTM architecture for sequence learning.

In the detection of depression, the spatial and frequency domains are two important factors, but there are no suitable neural networks to combine them together. Therefore, we propose the SpatNet neural network to combine the two features to improve the detection of depression.

# 3. MATERIALS AND METHODS

## 3.1. Participants

The dataset used in this study is from Lanzhou University. The dataset (Cai et al., 2020a) mainly includes data from patients with depression and the normal control group. Prior to the experiment, all participants signed the written informed consent. The consent and study design were approved by the Local Biomedical Research Ethics Committee of Lanzhou University Second Hospital in accordance with the World Medical Association rules. There were 48 participants, including 24 patients with depression (13 men and 11 women; 16–56 years old) and 24 healthy controls (17 men and 7 women; 18–55 years old). All patients with depression underwent a structured MINI interview, which met the diagnostic criteria for depression in the DSM-IV-based Diagnostic and Statistical Manual of Mental Disorders (DSM). Patients with MDD were selected based on their PHQ-9 (Kroenke and Spitzer, 2002), GAD-7 (Spitzer et al., 2006), and PSQI scores. Participants should be between 18 and 55 years old and have primary or higher education. For depression, inclusion criteria were MINI meeting the diagnostic criteria, patient health Questionnaire (PHQ-9) score of 5 or greater, and no psychotropic medication in the past 2 weeks.

## 3.2. Data Acquisition and Preprocessing

Subjects were asked to stay awake and still, and to reduce head and body movements and eye movements to reduce EMG and EOG, respectively, which record a 5-min resting state of closed eyes. EEG signal acquisition equipment is 128-channel HydroCel Geodesic Sensor Net. The sampling rate is 250 Hz. The reference electrode is Cz. The skin impedance of each electrode channel is kept below 70 kΩ.

The resting EEG signals are further processed using MATLAB 2021b. In the first step, the infinite impulse response digital filter IIR is used to perform 1–40 Hz band-pass filtering on the signal, and the order of the filter is set as 6. The filter can eliminate the "baseline drift" caused by low frequency noise and electrical interference from the 50 Hz-line noise. In the second step, independent principal component analysis is used to remove the EOG and EMG. Meanwhile, the integrity of channel recording signals is checked. If the invalid channels are detected, spherical interpolation is used for interpolation. The EEG signals of depressed patients and normal people are shown in **Figure 1**. In the third step, the processed time-domain signal is decomposed and reconstructed to remove the noise. A phase space reconstruction of a time-domain signal is decomposed into three signals, and then the new signal is reconstructed through the least square interpolation. The fourth step is to keep the same sample size between the subjects. The sliding window with the 2s non-overlap method is adopted for sectioning (Siuly et al., 2015), and the sample size of each subject is 148*2s. In the process of beginning and ending the experiment, the interference of brain electricity would be relatively large, so we discarded the first sample and the last sample.

## 3.3. The Time-Domain Denoising

### 3.3.1. Phase Space Reconstruction

Compared with normal EEG signals (Knott et al., 2001; Puthankattil and Joseph, 2012; Sharma et al., 2018), inhibited EEG signals have stronger stability and lower complexity. Hence, the phase space reconstruction is used to better analyze the complexity and non-stationary behavior of normal and depressed EEG signals in two-dimensional space, and the unstable noises are removed. In Sharma and Pachori (2015), Bhattacharyya and Pachori (2017), the two-dimensional diagrams of EEG signals have been used for seizure detection. The phase space reconstruction was first proposed in 1980. There are mainly two methods for phase space reconstruction: the derivative reconstruction method and the coordinate delay reconstruction method. The data set is collected in the resting state, so the signal is generally stable. The current signal feature can be

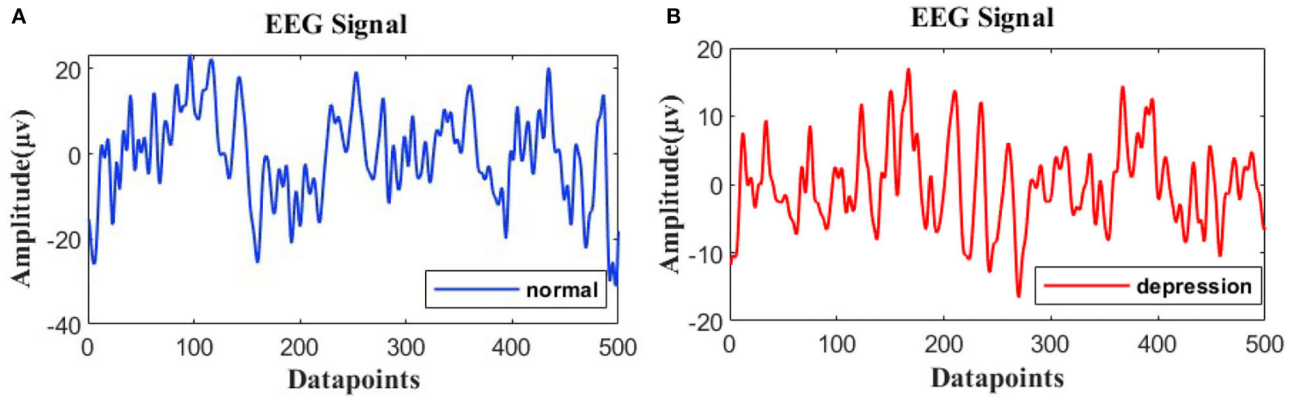



**FIGURE 1 |** Electroencephalogram (EEG) signals from **(A)** normal and **(B)** depressed subjects.

predicted by the previous signal feature and the following feature. Based on this characteristic, the coordinate delay reconstruction method is finally adopted. The generation of the PSR requires the determination of delay time t and embedding dimension d, which can be obtained by mutual information (MI) (Roulston, 1999; Bradley and Kantz, 2015) and false nearest Neighbor (FNN) (Bradley and Kantz, 2015), respectively. Supposing the time series is $x(i)$: $i = 1, 2, \ldots, n$, the d dimensional phase space vector is constructed by different delay time t of one-dimensional time series $x(i)$, as shown in Equation (1).

$$y(i) = (x(i), \ldots, x(i + (d-1)t)), 1 \leq i \leq n - (d-1)t \quad (1)$$

The initial application of the PRS is in the chaotic time series. According to Takens' embedding theorem (Stark et al., 1997; Muldoon et al., 1998; Kukavica and Robinson, 2004), we can reconstruct a phase space from one-dimensional chaotic time series that is the same as that of the prime motorial system in the sense of topology. According to this principle, the EEG signals as the general time series can reconstruct a phase space topologically identical to the original signal by determining the delay time t and d dimensions of the phase space. Since the EEG signal is collected by the subject in the resting state with eyes closed, the brain activity is relatively stable, and the characteristics of the current signal can be determined by the characteristics of the front and rear signals. If there are interference noises at this time, the original signal will have a mutation in the waveform. We can determine and correct the reconstructed phase space.

### 3.3.2. Least Square Fitting

After the original EEG signal is reconstructed in the phase space, the reconstruction is selected as a three-dimensional phase space vector. Then, the linear least square method by Kiers (1997) is used to fit each point inside, because the EEG signal in the resting state would not have mutations. Thus, the current signal point could be fitted according to the characteristics of the signals before and after. Given n points $(x_i, y_i), i = 1, 2, 3, \ldots, n$. $x_i$ is not the same, as shown in Equation (2).

$$f(x_i) = a_1 r_1(x_i) + a_2 r_2(x_i) + \cdots + a_m r_m(x_i) \quad (2)$$

$f(x)$ is closest to all the data points. We assume that the current data point is $i$ (unknown) and all known points before and after are taken to determine the fitting function. Where, the steps to determine the coefficient $a_k$ are as follows: First, the error function of the fitting curve and the original curve is Equation (3).

$$J(a_1, a_2, \ldots, a_m) = \sum_{i=1}^{n} \delta^2 = \sum_{i=1}^{n} [f(x_i - y_i)]^2 \quad (3)$$

For $a_1, a_2, \ldots, a_m$ minimizes $J$, use the necessary extreme conditions: $\partial J / \partial a_k = 0$ $(k = 1, \ldots, m)$ to get the linear Equation (4).

$$\sum_{i=1}^{n} r_j(x_i)[\sum_{k=1}^{m} a_k r_k(x_i) - y_i] = 0, (j = 1, \ldots, m) \quad (4)$$

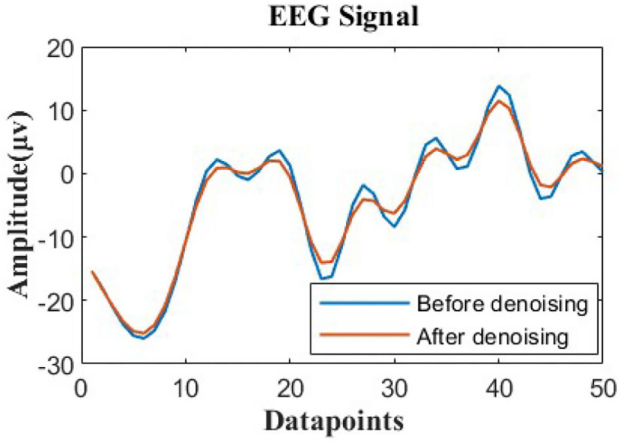



**FIGURE 2 |** Phase space reconstruction and the least square method were used to detect EEG signals before and after denoising in patients with depression.

make $R = \begin{pmatrix} r_1(x_1) & \cdots & r_m(x_1) \\ \vdots & \ddots & \vdots \\ r_1(x_n) & \cdots & r_m(x_n) \end{pmatrix}$, $A = [a_1, \ldots, a_m]^T$, $Y = [y_1, \ldots, y_n]^T$, so we can rewrite this equation as Equation (5).

$$R^T R A = R^T Y \quad (5)$$

Thus, when the equation satisfies $r_1(x), \ldots, r_m(x)$ linearly independent, $R$ column full rank, $R^T R$ invertible, there is a unique solution Equation (6).

$$A = (R^T R)^{-1} R^T Y \quad (6)$$

For the selection of function $r_k(x)$, we use the polynomial curve to better fit the EEG signal. **Figure 2** shows the comparison before and after denoising EEG signals of patients with depression by using phase space reconstruction and the least square method.

## 3.4. Characteristics of Feature Smoothing

Pham et al. (2015) proposed the importance of feature smoothing for emotional EEG classification. The denoised EEG signal is converted into the frequency domain signal by the fast Fourier transform. The frequency domain signal is used as the input of the neural network. Each sample is segmented according to the time sliding window. Therefore, the transformed frequency domain signals are also smoothed according to 148 time points (each sample size). The EEG signals are collected by the subjects with their eyes closed, so the changes in the EEG signals in the frequency domain are not particularly obvious. Equations (7) and (8) are used to calculate the difference between the signals at each time point and the mean value of the signals.

$$\delta_j^i = |y_j^i - E(Y)^i|, 1 \leq i \leq 40, 1 \leq j \leq 148 \quad (7)$$

$$E(Y)^i = \frac{1}{J} \sum_{j=1}^{J} y_j^i \quad (8)$$

Where $Y$ represents the amplitude, $i$ represents the frequency, and $j$ represents the time point. $E(Y)$ is the mean of a sample of a subject at frequency $i$. $\delta_j^i$ is the error between each sample point and the mean.

$$\delta_j^i - 3\sqrt{\frac{1}{J}\sum_{j=1}^{J}(y_j^i - E(Y)^i)^2} > 0 \tag{9}$$

Mark $y_j^i$ as an outlier if the value of $y_j^i$ differs from the mean by more than three standard deviations as shown in Equation (9). After the outliers are detected, the $y_j^i$ sample points marked before are smoothed by the method of before-and-after mean interpolation. In this way, the influence of abnormal features can be further cleaned up, while maintaining the overall trend of frequency domain features. At the same time, it can reduce the risk of over-fitting during the training of the neural network.

## 3.5. Attention Module

The attention mechanism is a special structure embedded in a machine learning model, which is used to automatically learn and calculate the contribution of input data to output data. The attention mechanism is a signal processing mechanism discovered by some scientists in the study of human vision. Some practitioners in the field of artificial intelligence have introduced this mechanism into some models. At present, the attention mechanism has become one of the most widely used "components" in the field of deep learning, especially in the field of natural language processing. The classic ones are BERT, Transformer (Devlin et al., 2018; Wang et al., 2019), and other models or structures that are highly exposed in the past 2 years. In this article, we adopt the SENet (Squeeze-and-Excitation Networks) (Hu et al., 2018) module incorporated with the channel attention mechanism.

In our proposed network model, the EEG signals are converted into frequency domain features and then used as the input of the SENet module. The SENet module is originally used to process the two-dimensional images, in this article, we use it in one-dimensional signal processing. It can mainly use the global information to selectively emphasize the information features and suppress the less useful features by assigning different weight values to each channel. This is a combinatorial function of five consecutive operations: channel global average pooling (Lin et al., 2013), complete connection layer, Relu, complete connection layer, and finally Sigmoid. The sigmoid activation plays an important role as the channel weights that adapt the input specific descriptors. Due to the fully connected layer and pooling layer, the number of parameters and the computation load increased slightly. The unique structure of this extrusion and excitation network, shown in **Figure 3**, can be used with any standard architecture.

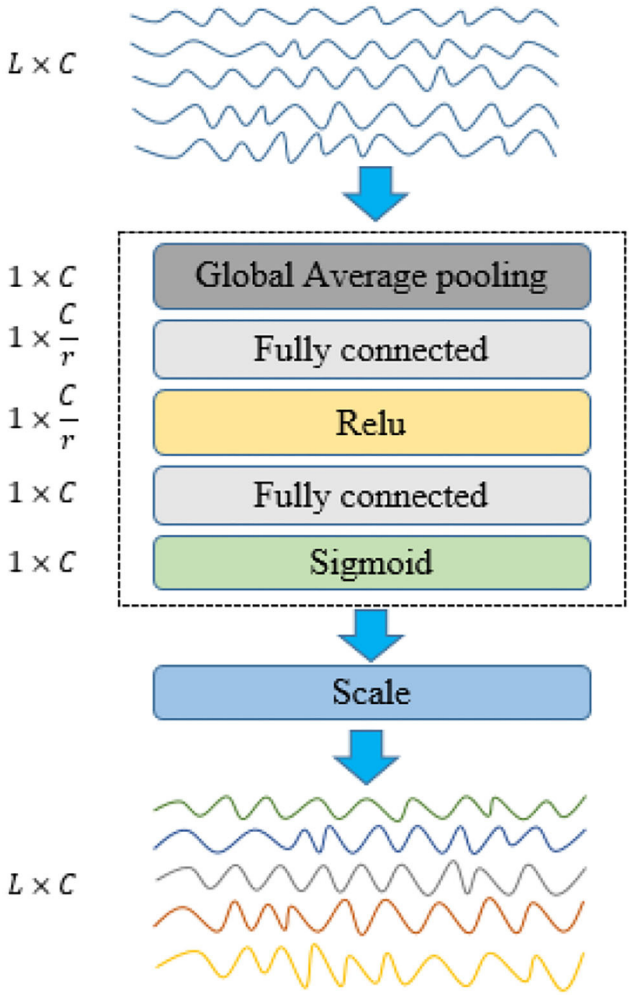



**FIGURE 3 |** Structure diagram of SENet network applied to one dimensional EEG signal. The weight of each channel of EEG signal before processing is equal, so the color of the channel is the same. The different channels after processing were assigned different weights, with different colors representing this change.

### 3.5.1. Squeeze

SENet implements compression operations through the global average pooling to generate channel statistics. Where $Z \in R^C$, and the $k^{th}$ element $z_k$ of $Z$ is calculated by Equation (10).

$$z_k = F_{sq}(u_k) = \frac{1}{L}\sum_{i=1}^{L} u_k(i), k = 1, 2, \cdots, C \tag{10}$$

where $F_{sq}(\cdot)$ is the compression operation, and $u_k$ is the feature on the $k^th$ channel. $C$ is the total number of channels.

### 3.5.2. Excitation

The excitation operations help to capture channel dependencies and greatly reduce the number of parameters and calculations. The excitation part is mainly composed of two fully connected

layers and two activation functions, which can be written as Equation (11).

$$S = F_{ex}(Z, W) = \sigma(\gamma(Z, W)) = \sigma(W_2\delta(W_1 Z)) \quad (11)$$

where $S = s_1, s_2, \ldots, s_C, s_k \in R^L (k = 1, 2, \ldots, L)$. $F_{ex}(\cdot)$ is the excitation operation. $W_1 \in R^{\frac{C}{r} \times C}$, $W_2 \in R^{C \times \frac{C}{r}}$, $W_1$, and $W_2$ are the weights of the two fully connected layers used for dimension reduction and dimension enhancement. $Z$ is the fully connected input after global average pooling. $r$ is a hyperparametric ratio, which can change the capacity and calculation cost. $\delta(x)$ is the activation function Relu used to prevent the gradient from disappearing (Gu, 2017). $\sigma(x) = \frac{1}{(1+e^{-x})}$ is a sigmoid function. Equation (12) is used to calculate the final output $\tilde{x_k} (k = 1, 2, \cdots, C)$. The output is obtained by multiplying the input channels by their respective weights.

$$\tilde{x_k} = F_{scale}(u_k, s_k) = u_k \cdot s_k \quad (12)$$

where $\tilde{x_k} \in R^L$ refers to the multiplication above the channel. $s_k$is the processed channel weight. $u_k$ is the original eigenvector.

## 4. PROPOSED DEEP LEARNING SCHEME

Convolutional Neural Network is a special type of neural network which is widely used in image processing and classification tasks. It is a state-of-the-art deep learning method consisting of many stacked convolutional layers. The network consists of a convolution layer, pool layer, and final complete connection layer. The EEG signals are one-dimensional time series signals. After converting them into frequency domain signals, multi-channel one-dimensional frequency domain signals are input. Therefore, 1DCNN is used in the convolutional neural network. The features in the frequency domain can be fully combined with the spatial information between channels.

### 4.1. Convolution Layer

At the convolution layer, 1DCNN carries out the convolution operation on the local area of the input signal to generate the corresponding one-dimensional feature map. Different convolution kernels extract different features from the input signal respectively. Each convolution kernel detects the specific features of all positions on the input feature graph to achieve the weight allocation on the same input feature graph. The characteristics of the local connectivity and weight sharing effectively can reduce the complexity of the network and the

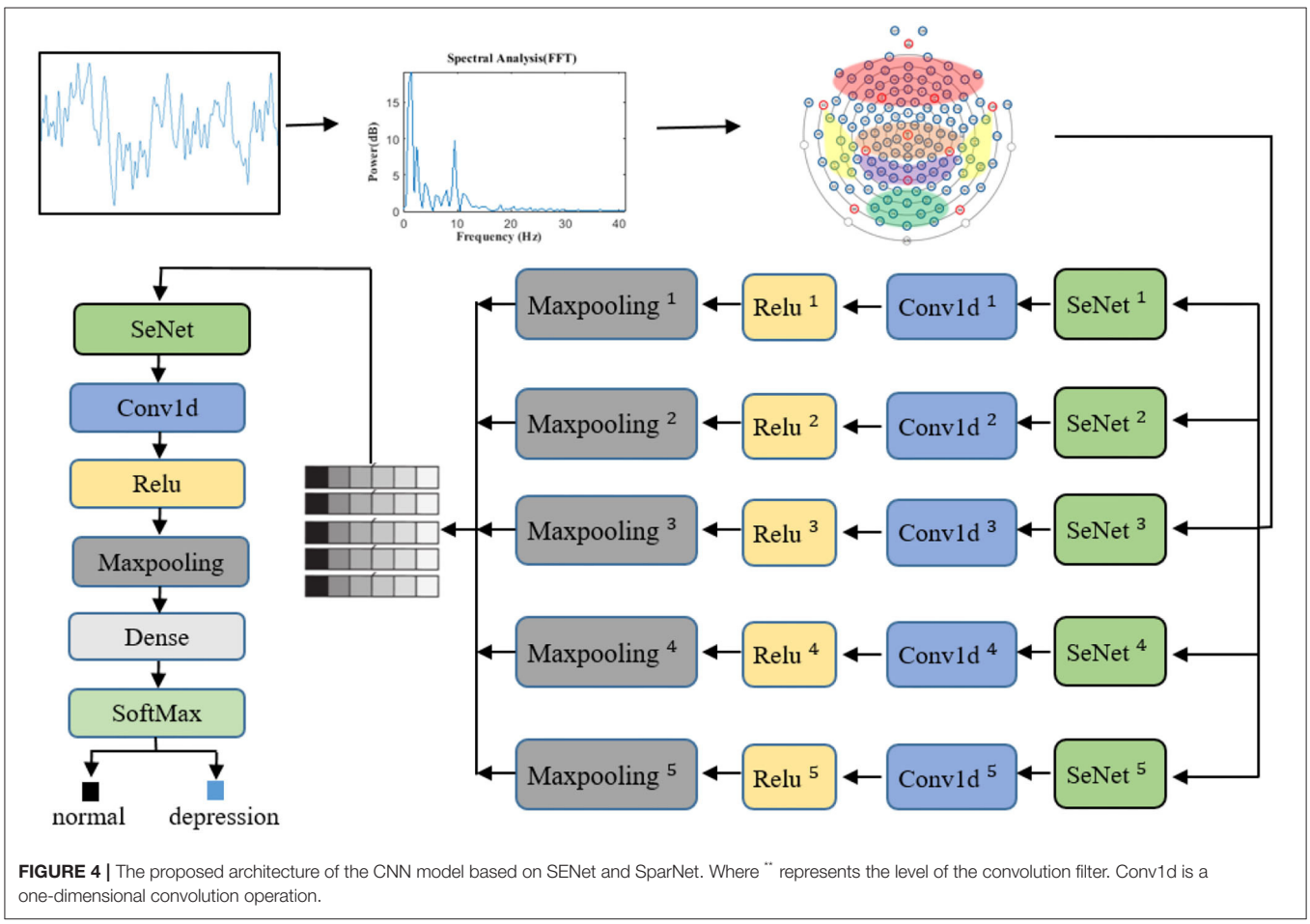



**FIGURE 4 |** The proposed architecture of the CNN model based on SENet and SparNet. Where ** represents the level of the convolution filter. Conv1d is a one-dimensional convolution operation.

number of training parameters. If the $L$ layer is the convolution layer, Equation (13) of the one-dimensional convolution layer is

$$x_j^l = f(\sum_{i=1}^{M} x_i^{l-1} * k_{ij}^l + b_j^l) \tag{13}$$

where $k$ represents the convolution kernel, $j$ represents the number of convolution kernels, and $M$ represents the number of channels of the input $x(l-1)$ of the upper layer. $b$ is the offset of the convolution kernel, and * stands for the convolution operation. $f(\cdot)$ is the activation function.

## 4.2. Activation Layer

What is done in the convolution layer of the upper layer is to process the input features in the way of convolution, i.e., to assign a weight to each pixel point. This operation is linear. But for samples, they are non-linear separable instead of linear separable. Thus, we add the activation function Relu Equation (14) here.

$$f(x) = max(0, x) \tag{14}$$

As a non-linear factor, the activation function added to the model can make the model more expressive and better fit the data.

## 4.3. Pooling Layer

The downsampling stage is after the convolution layer and the number of feature graphs increases. This leads to the expansion of the data dimension, which is not conducive to calculation. Therefore, the average pooling or maximum pooling is used to process each feature map in this stage. The average pool is calculated according to the size of the predetermined pool window, and the maximum pool method selects the maximum parameter within the predetermined window range as the output value. In our study, the maximum pooling operation is adopted. The pooling kernel size is 1*2, the step size is 1, and there is no filling.

## 4.4. Connection Layer

After passing through the convolution layer, the data scale is channel×features, and the feature dimension needs to be straightened into one dimension. At this point, the full connection layer of the node is connected with all neuron nodes output from the feature mapping of the previous layer, and the activation function is softmax function. If the final pooling layer is l+1 and output to the full connection layer, then the output of the full connection layer is Equation (15).

$$h(x) = f(w^{l+1} \cdot x^{l+1} + b^{l+1}) \tag{15}$$

where $w$ represents the weight of each feature, and $b$ represents the offset. $F(\cdot)$ represents the activation function.

## 4.5. Loss Function

It is used to calculate the error between the classification prediction label and the actual label. The classification cross entropy is used as the loss function, and the probability distribution is compared with the real distribution. $L_1, L_2, \ldots,$ and $L_T$ are represented by the one-hot encoding strategy. The loss function can be calculated as Equation (16).

$$Loss = -\sum_{i=1}^{T}\sum_{j=1}^{M} L_{i,j} * logp_{i,j} \tag{16}$$

where $T$ is the number of verification data samples, $M$ is the number of classes, $p_{i,j}$ is the predicted value obtained from the fully connected layer, and $L_{i,j}$ is the true value.

**TABLE 1 |** Detailed information about the proposed SparNet deep model.

| No | Names of layers | Kernel size | Parameters of layers |
|---|---|---|---|
| 0 | Input | — | — |
| 1 | SENet$^{(1-5)}$ | — | Reduction = 2 |
| 2 | Conv1D$^{(1)(3-5)}$ | 12*5 | Stride = 1 |
| | Conv1D$^{(2)}$ | 25*5 | Activation = Relu |
| 3 | MaxPooling1D$^{(1-5)}$ | 2 | Stride = 2 |
| 4 | SENet | — | Reduction = 2 |
| 5 | Conv1D | 5*2 | Stride = 1, Activation = Relu |
| 6 | MaxPooling1D | 2 | Stride = 2 |
| 7 | Dense | — | Neurons = 2 |
| 8 | Softmax | — | — |

*The ** represents the level of the convolution filter: SENet$^{(1-5)}$ represents the SENet of the 1st to 5th layers.*

**TABLE 2 |** Channels in different brain regions are selected.

| | |
|---|---|
| C | E36-E104,E30-E105,E41-E103,E37-E87,E42-E93,E47-E98 |
| F | E19-E4,E22-E9,E24-E124,E27-E123,E32-E1,E33-E128 |
| O | E70-E83,E71-E76,E69-E89,E74-E82,E73-E83,E75,E81 |
| P | E52-E92,E60-E85,E51-E97,E67-E77,E59-E91,E72,E62 |
| T | E58-E96,E45-E108,E114-E44,E100-E46,E102-E57,E50-E101 |

## 4.6. SparNet

In view of the advantages of CNN, one-dimensional CNN is used to extract the spatial frequency features of EEG signals. In the proposed SparNet network architecture, each layer is directly connected to each other in feedforward mode. In **Figure 4**, the SparNet network consists of six sub-CNNs. Each sub-CNN consists of a SENet, Conv1d, Relu, and Maxpooling. Five of them are at the same level to form a network of parallel structures to operate the brain region, and the last convolutional operation performs a processing operation on the global brain. Finally, there is a fully connected layer output result. The SENet module is added to each sub-network to increase the attention mechanism between channels and brain regions. **Table 1** details each layer of the proposed SparNet network and the parameters of each layer.

First of all, the same sample size is kept between the subjects. We use the sliding window 2s non-coincidence method to slice, and the sample size of each subject is 148*2s. In this study, since the recognition features of depression patients are mainly in the frequency domain signals, we manually extracted the features in this step. The EEG signals are transferred from the time domain

to the frequency domain (1–40 Hz) by a fast Fourier transform, and the frequency domain signal is used as the input of the neural network. The channels are sorted out before entering the first layer of the model. The channels are divided into five regions according to the brain regions. At this point, the data size of each brain region is channeled *40. Before entering the convolution layer, the SENet module will adjust the weight of each channel, and then the space frequency domain features of each brain region will be extracted through convolution operation and maximum pooling. The features of five brain regions will be spliced to form a feature scale of 5*features, where 5 represents five brain regions. The SENet module then weights each brain region based on global features. After the convolution operation, the output of the full connection layer is entered. **Figure 4** shows the whole signal processing process of the network.

## 4.7. Characteristics in the Space Frequency Domain

Peng et al. (2019) and Jiang et al. (2021) showed that the EEG signals of depressed patients had better feature representation in the frequency domain, where absolute power and relative power in the frequency domain were of great help in identifying patients with depression. Moreover, the expression of EEG signals in the power spectrum of patients with depression is obviously different from that of healthy people. Stark et al. (1997) and Lin et al. (2013) showed that in the spatial domain of channels, the brain functional connections of patients with depression were different from those of normal people, and the accuracy of classification of patients with depression was significantly improved.

Therefore, this article aims to combine frequency domain and spatial domain to effectively identify patients with depression

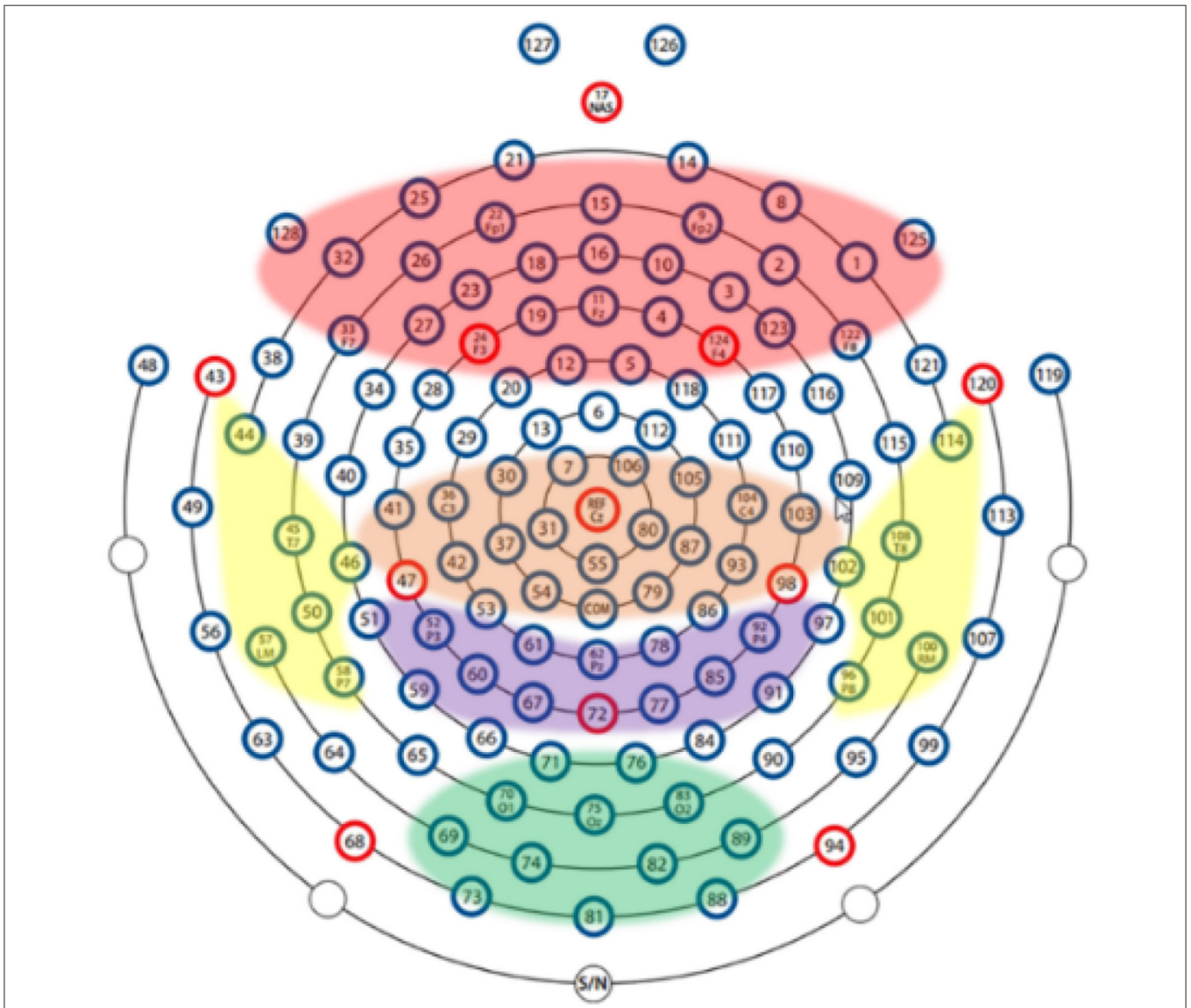



**FIGURE 5 |** The division of the brain between different brain regions.

in space-frequency domain. Based on the characteristics of the brain, it is divided into five brain regions. First, the local space-frequency domain is explored, and then the feature extraction of the global space-frequency domain is carried out. The brain regions are shown in **Table 2**.

In order to explore the influence of various brain regions on patients with depression, we divide the brain into five parts as shown in **Figure 5**: central region, frontal lobe region, occipital lobe region, parietal lobe region, and temporal lobe region. The first layer of the SparNet network has five parallel sub-CNNs corresponding to the five brain regions. Due to the tightness between brain regions, we initially select 12 channels in each brain region to carry out space frequency domain characteristics of brain regions. In order to accurately locate the channels in each brain region, we obtained the channel location information from Luu and Ferree (2005) by referring to the 128-channel HydroCel Geodesic Sensor Net device. Each channel contains spatial information about the brain. Additionally, in order to avoid the influence of multiple brain regions on a single channel, we select 12 channels distributed in the central location of each brain region without selecting channels at the edge of the brain region. The channels are then added individually according to each brain region's contribution to the recognition of depression. The selection of preliminary channels is shown in **Table 2**.

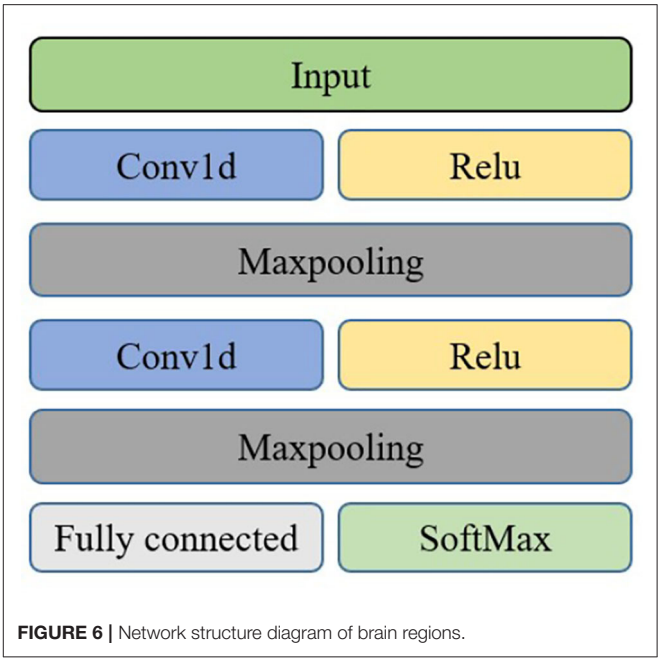



**FIGURE 6 |** Network structure diagram of brain regions.

# 5. RESULTS

## 5.1. Evaluation

There are 48 subjects in total. In order to better generalize the model, we adopt dataset division among subjects. A total of 24 sets are obtained by combining a depressed patient with healthy control, using the one-subject cross-validation (LOSOCV) to assess the generalization ability of each classification model.

In this study, accuracy, sensitivity, and accuracy based on the confusion matrix are used as the performance evaluation indexes. Sensitivity (recall rate) is defined as the percentage of patients with MDD predicted in all MDD patients (TP+ FN), and precision is defined as the percentage of healthy controls predicted in all healthy controls (TP+FP). Accuracy is defined as the percentage of correctly classified patients with MDD and healthy controls. F1 index takes into account both model accuracy and recall rate and is defined as the harmonic mean of model accuracy and recall rate.

$$Accurary = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{18}$$

$$Precision = \frac{TP}{TP + FP} \tag{19}$$

$$F_1 Score = 2 \times \frac{Precision \times Sensitivity}{Precision \times Sensitivity} \tag{20}$$

The Receiver Operating characteristic (ROC) curve is used for evaluation. The ROC curve is widely used in binary classification evaluation, which evaluates sensitivity and specificity against several thresholds.

## 5.2. Partial Results

The brain is divided into five regions: frontal, central, parietal, occipital, and temporal. In order to more effectively identify people with depression, we explored the importance of different brain regions. The importance of 12 channels in each brain region is evaluated. 1DCNN is performed on each of the five brain regions, and the evaluation index is the average accuracy after each fold. The neural network model of brain regions is shown in **Figure 6**.

Each subject is included in the test set by a 24-time retention cross-validation method for each brain region. The accuracy of each brain region is shown in **Figure 7**.

As can be seen from the results, the frontal lobe has a greater contribution to the identification of patients with depression compared with other brain regions, which is consistent with the results of Jiang et al. (2021), indicating the feasibility of the frontal lobe channel detection for depression. Wan et al. (2020) also prefer the frontal lobe channel in terms of channel selection. Depression patients with low moods will lead to emotional differences from normal people. Since the frontal lobe is the main brain area for emotional processing, the frontal lobe is more important for the identification of depression patients. Thus, in order to better explore the characteristics of the global space frequency domain, the number of channels in the frontal lobe is increased to 25 channels, E19-E4, E22-E9, E24-E124, E27-E123, E32-E1, E33-E128, E15, E11, E1 - E32, E27 - E123, E13 - E112, E29 - E111, E28-117, the E6, E121 - E38, E34 - E36, combined with the cerebral cortex region occupied by the frontal lobe.

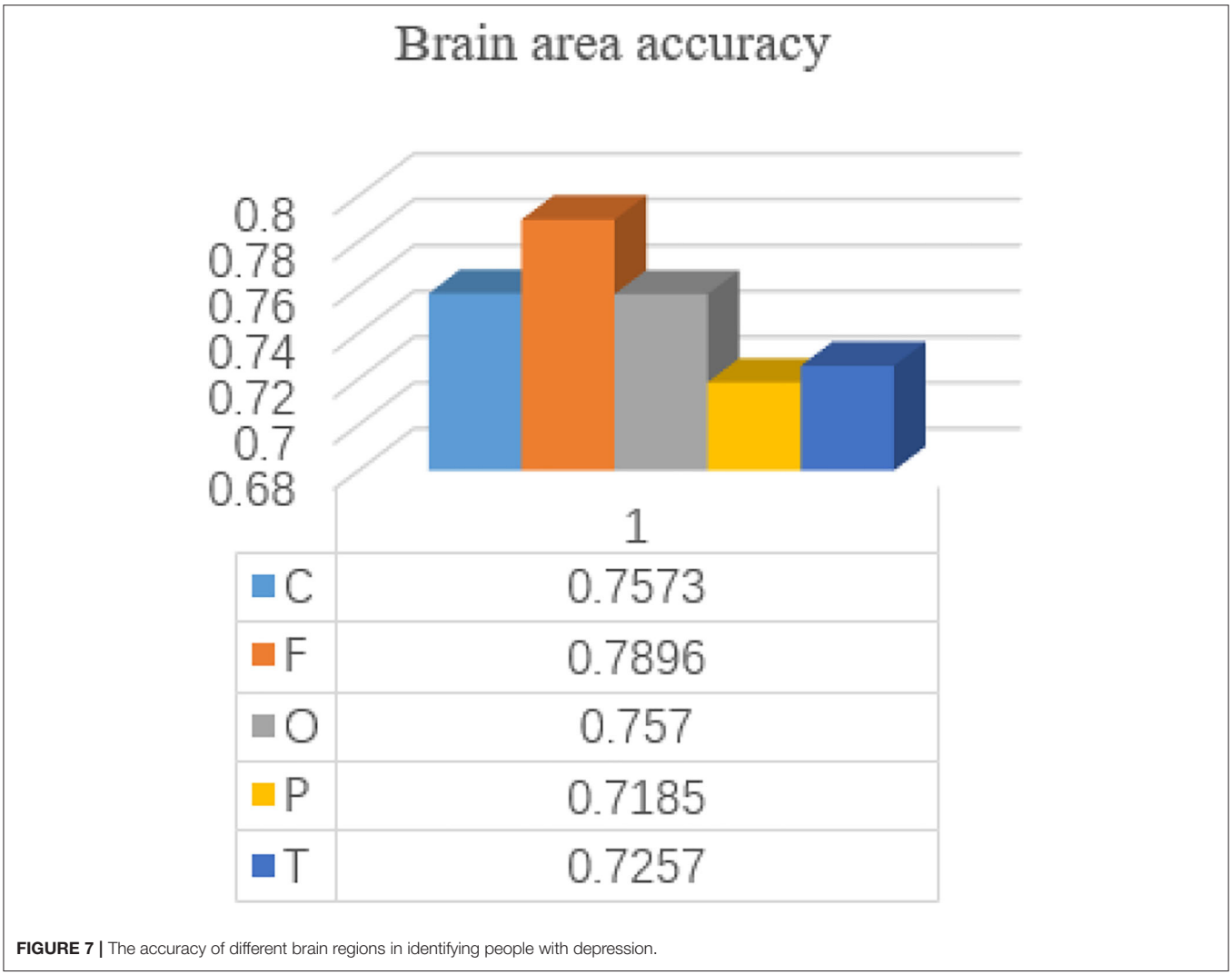



**FIGURE 7 |** The accuracy of different brain regions in identifying people with depression.

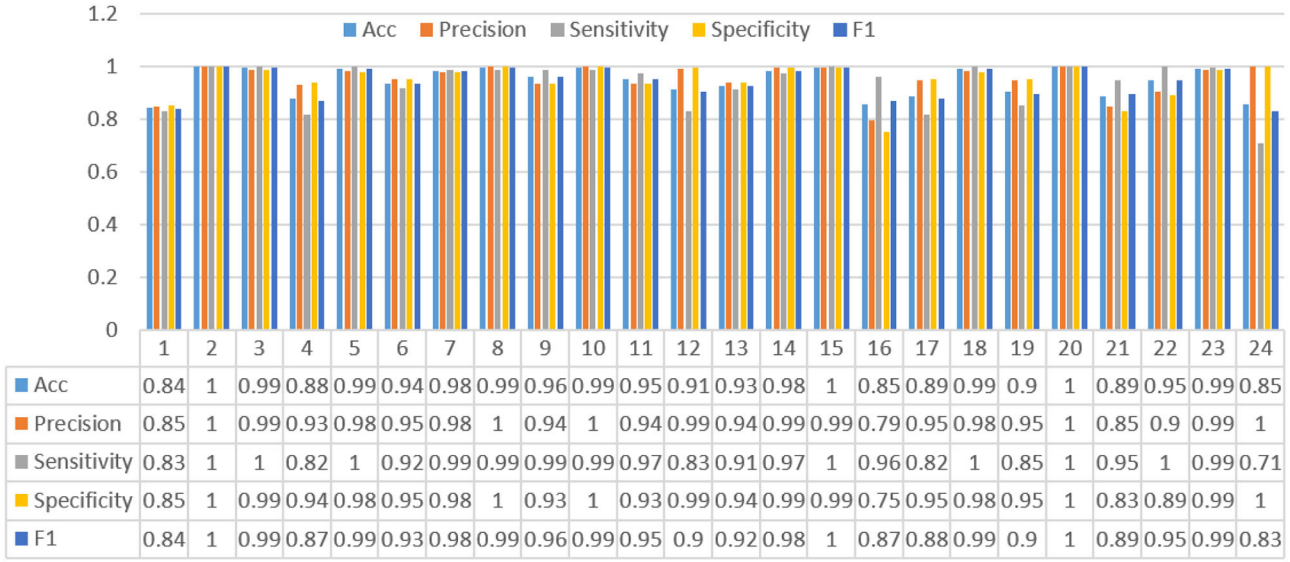
## 5.3. Global Results

The software Matlab 2020b is used to preprocess the data, and the EEG data are segmented with a 2s window length. After that, the time domain signal is converted to the frequency domain as the characteristic input of the model. The space frequency domain features of the EEG signal will be extracted by a two-layer convolution operation. In addition, in the convolution operation, each channel will be adjusted according to the parameters of a convolution kernel, and will not be affected by the parameters of other channels. Therefore, the channel grouping order has no significant impact on performance.

For training the model, the batch size is 8, and each network is trained with 50 epochs. The cross entropy is selected as the loss function. In the optimization stage, the RMS algorithm is selected to obtain better results and a shorter running time. If the loss function is verified not to improve after 10 consecutive epochs, the early stop criterion is used. **Figure 8** shows the calculation results for each fold. The ROC and AUC values of the above methods are shown in **Figure 9** for each fold. Due to the differences between the subjects, the minimum AUC area obtained is 0.829, and there are 16-fold AUC areas over 0.95, which also demonstrates the effectiveness of the model. The results of cross-validation are statistically analyzed. The mean values of F1, Acc, Precision, Sensitivity, Specificity are 0.947, 0.953, 0.937, 0.951, and 0.942, respectively. The standard deviation of F1, Acc, Precision, Sensitivity, Specificity are 0.054, 0.056, 0.082, 0.063, 0.056, respectively.

The results of 24-folds are combined and analyzed, and the evaluation indexes obtained are shown in **Table 3**, in which the accuracy rate reaches 94.37%. At the same time, the ROC and AUC values of the above methods are shown for the overall results, including the ROC and AUC values of the individual categories of normal people and depressed patients, as shown in **Figure 10**. The AUC area of the overall data is 0.9682. All the evaluation indicators decrease when the input characteristics in SparNet are time domain signals. The results are shown in **Table 4**, named "With Time Domain". Thus, the frequency domain feature is more effective than the time domain feature.

For the proposed model, in order to verify the effectiveness of the attention mechanism in identifying patients with depression,




| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Acc | 0.84 | 1 | 0.99 | 0.88 | 0.99 | 0.94 | 0.98 | 0.99 | 0.96 | 0.99 | 0.95 | 0.91 | 0.93 | 0.98 | 1 | 0.85 | 0.89 | 0.99 | 0.9 | 1 | 0.89 | 0.95 | 0.99 | 0.85 |
| Precision | 0.85 | 1 | 0.99 | 0.93 | 0.98 | 0.95 | 0.98 | 1 | 0.94 | 1 | 0.94 | 0.99 | 0.94 | 0.99 | 0.99 | 0.79 | 0.95 | 0.98 | 0.95 | 1 | 0.85 | 0.9 | 0.99 | 1 |
| Sensitivity | 0.83 | 1 | 1 | 0.82 | 1 | 0.92 | 0.99 | 0.99 | 0.99 | 0.99 | 0.97 | 0.83 | 0.91 | 0.97 | 1 | 0.96 | 0.82 | 1 | 0.85 | 1 | 0.95 | 1 | 0.99 | 0.71 |
| Specificity | 0.85 | 1 | 0.99 | 0.94 | 0.98 | 0.95 | 0.98 | 1 | 0.93 | 1 | 0.93 | 0.99 | 0.94 | 0.99 | 0.99 | 0.75 | 0.95 | 0.98 | 0.95 | 1 | 0.83 | 0.89 | 0.99 | 1 |
| F1 | 0.84 | 1 | 0.99 | 0.87 | 0.99 | 0.93 | 0.98 | 0.99 | 0.96 | 0.99 | 0.95 | 0.9 | 0.92 | 0.98 | 1 | 0.87 | 0.88 | 0.99 | 0.9 | 1 | 0.89 | 0.95 | 0.99 | 0.83 |

**FIGURE 8 |** Twenty-four-fold cross-validation results including accuracy, precision, sensitivity, specificity, and F1 evaluation indicators.

six SENet modules are removed from the model and the comparison before and after the results are shown in **Table 4**.

It can be seen from the results that adding an attention mechanism into the model can improve the learning ability of the model. The accuracy rate of depression identification is improved by 3.13%, and all the evaluation indicators are also improved.

In this article, after converting the EEG signals into the frequency domain, the smoothing operation is performed on the frequency domain features. The comparative experiments are conducted before and after feature smoothing, and the experimental results are shown in **Table 4**.

It can be seen from the results that the accuracy is improved by 4 % using the smoothed features as the input. Furthermore, we can observe that all the evaluation indicators are improved at the same time. The calculation cost of the network is calculated. The parameter number of the SparNet network is 8,725, and the network parameter without the SENet module is 455. The number of network parameters without smooth is 8,725, mainly to illustrate the necessity of feature smoothing for neural networks. The SparNet parameters are mainly contributed by the SENet module.

## 6. DISCUSSION

Screening for depression is very important for early diagnosis and treatment. However, the previous diagnosis of depression is confined to a manual questionnaire survey and feature extraction, which is subject to many limitations. For example, a questionnaire survey required experienced doctors, while feature extraction required a lot of manpower to find the characteristics of relevant indicators. The deep learning methods can overcome this limitation and can be used anywhere without highly trained experts. In this study, we use feature smoothing and deep learning for the automatic detection of patients with MDD and healthy controls with good performance. The attention mechanism is combined with 1DCNN, and the spatial-frequency domain features are extracted by brain regions. The accuracy rate of our model reaches 94.37%.

The main innovation of this study is to make full use of the spatial and frequency domain characteristics of the brain. We also try to smooth the input of the neural network. Experimental results show that the frequency domain characteristics of the input smoothing processing can effectively improve the identification accuracy of patients. In addition, due to the characteristics of the convolutional neural network, the brain is divided into different regions for feature extraction of parallel structures.

It can be seen from the final results that adding the SE module into the 1DCNN neural network can make the model have higher accuracy. This is because the convolutional layer provides a powerful feature fusion technology, although the weights between channels are unified by default. The SE module can better highlight the importance of different channels. At the same time, although there has been a feature of converting the EEG signal into the frequency domain as the input, the segmentation of brain regions according to the brain's structure can make the spatial information of brain regions better interpreted. The results for each brain region and the final results for the whole brain region show that although each brain region is helpful in identifying depression, the characteristics of the whole brain are better. It could be thought that a large amount of space-frequency information is lost for a single brain region and cannot get a significant effect.

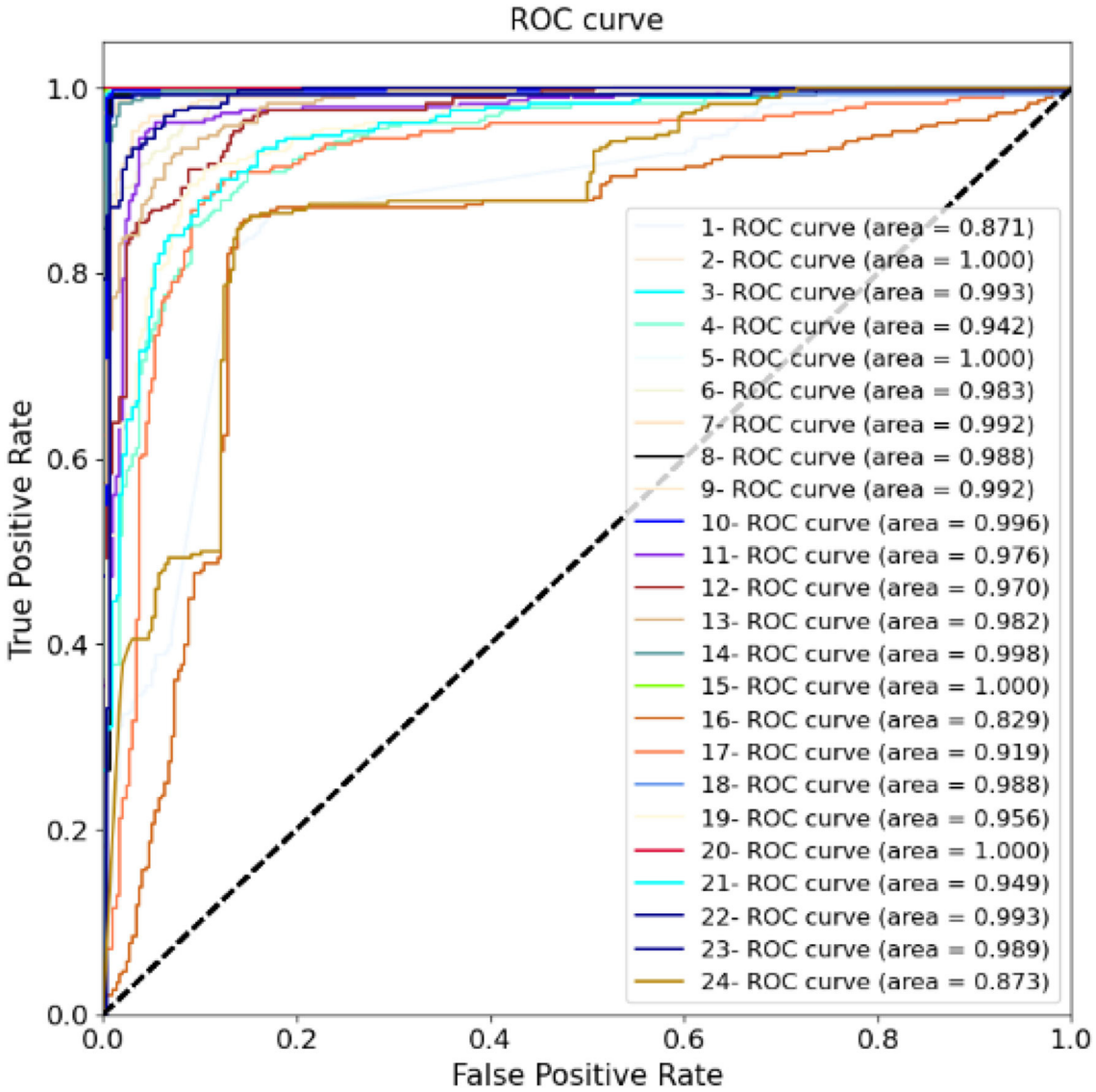



**FIGURE 9 |** Twenty-four-fold cross-verified ROC curves and areas.

**TABLE 3 |** Performance values obtained when testing the SparNet model using EEG data.

| | Class | Predicted depression | Normal | Acc | Precision | Sensitivity | Specificity | F1 |
|---|---|---|---|---|---|---|---|---|
| Actual | Depression | 3,327 | 225 | 0.9437 | 0.9375 | 0.9507 | 0.9366 | 0.9440 |
| | Normal | 175 | 3,377 | | 0.9500 | 0.9366 | 0.9507 | 0.9432 |

**Table 5** is a comparison of results on the same data set. It can be seen that the accuracy of this study is higher than that of traditional machine learning methods for linear and non-linear extraction. Additionally, compared with Peng et al. (2019), we use fewer channels to achieve better accuracy and obtain the best results in this dataset in automated detection of patients with depression and healthy controls, which could provide better solutions for future clinical applications. Compared with Zhang et al. (2020), the combination of multi-parallel 1DCNN and SE modules in this study tests the importance of space-frequency information and other advantages. However, the main shortcoming of this study is the data size to train the network, which can be overcome by simplifying the deep model. Our future goals

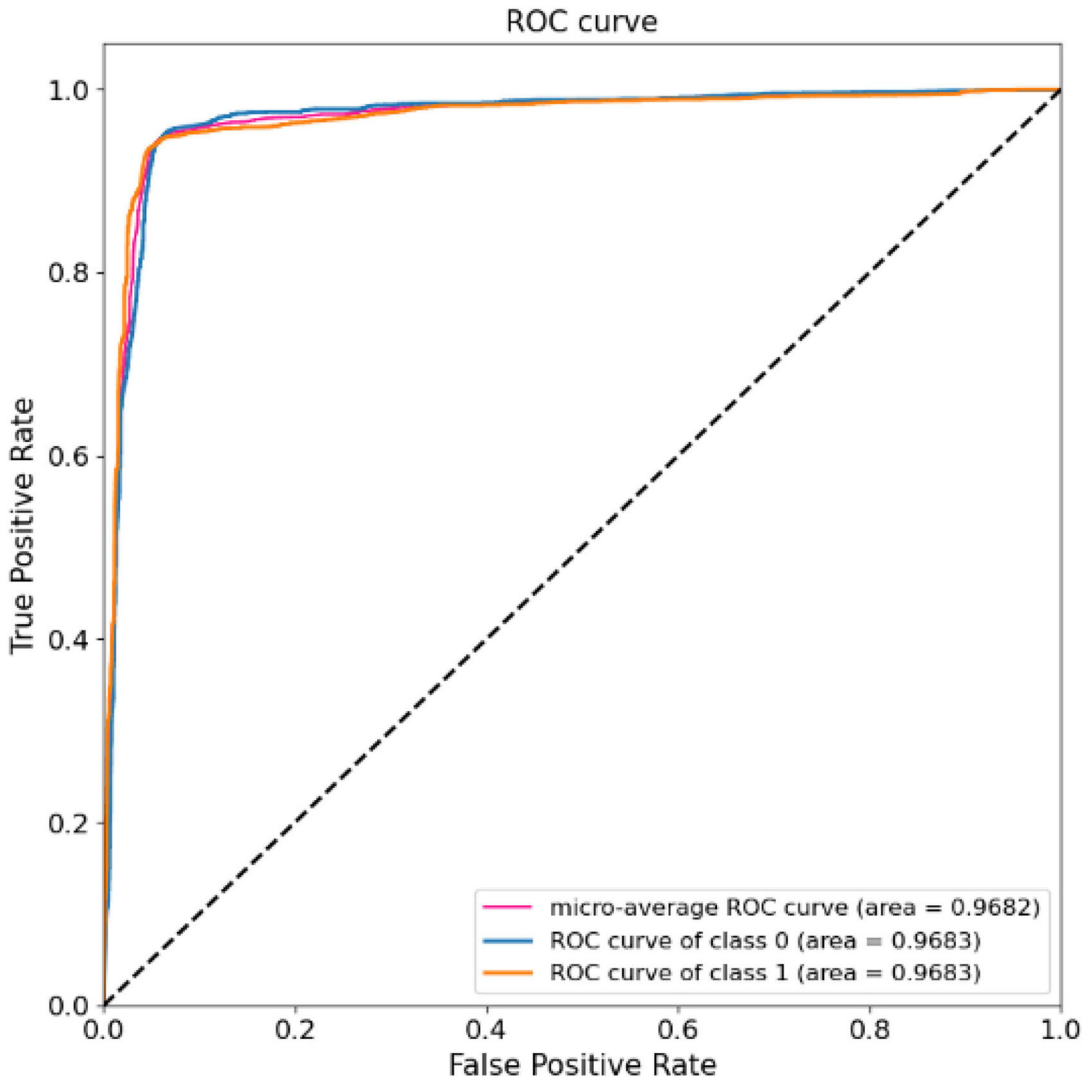



**FIGURE 10 |** Receiver Operating characteristic (ROC) curves and area of overall EEG data tests.

**TABLE 4 |** Comparison of results before and after adding SENet module.

| | Acc | Precision | Sensitivity | Specificity | F1 |
|---|---|---|---|---|---|
| SparNet | 0.9437 | 0.9375 | 0.9507 | 0.9366 | 0.9440 |
| With time domain | 0.8851 | 0.9189 | 0.8607 | 0.9104 | 0.9130 |
| Without the SENet | 0.9124 | 0.8953 | 0.9341 | 0.8907 | 0.9143 |
| Without smooth | 0.9037 | 0.9099 | 0.9042 | 0.9104 | 0.9070 |

**TABLE 5 |** The confounding matrix and evaluation index are used to compare the classification results of SparNet and correlation methods in the same data set.

| References | Methods | Channels number | Classification methods | Accuracy (%) |
|---|---|---|---|---|
| (Sun et al., 2020b) | Linear features, non-linear features, PLI | 16 Channels | ReliefF, LR | 82.31 |
| (Peng et al., 2019) | PLI | 128 Channels | Kendall rank correlation coefficient+SVM | 92.73 |
| (Sun et al., 2020a) | Linear features, non-linear features, PLI, network measures | 128 Channels | C4.5, BFDT, LR | 84.18 |
| (Zhang et al., 2020) | Time domain feature | 3 Channels | 1DCNN | 75.29 |
| SparNet | Frequency domain feature | 73 Channels | 1DCNN | **94.37** |

*The bold value is the accuracy derived from the SparNet.*

are to further expand the experimental space, collect more samples, and apply the developed methods to other types of EEG data.

## 7. CONCLUSION

The main study in this article is using the novel neural network called SparNet based on the EEG signals to identify depressed people or not. First, the denoising method of the phase space reconstruction is used to denoise and clean the data. Second, the input features are smoothed before the frequency domain features are input into the model. Third, a new model called SparNet is proposed to extract the space-frequency domain features of the local brain regions and the whole-brain. Finally, the cooperating of the attentional mechanisms to the model improves the identification accuracy of the patients with depression. Compared with other methods, the proposed model can obtain a better classification performance. From the results of the local brain regions, it can be seen that the frontal lobe plays a better role in the identification of patients with depression. From the results of the global brain region, it can be seen that the combination of the spatial features and the frequency domain features can effectively improve the accuracy of depression identification. The combination of features of the different brain regions may be the focus of future research. The methods and findings of this study may contribute to the wider application of the diagnosis of deep depression in clinical applications and neuroscience research.

## 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

XD and XF: conceptualization and writing—review and editing. XD, XF, and XL: methodology. XF: software, data curation, and writing—original draft preparation. XD, XF, XL, and KS: validation and investigation. All authors have read and agreed to the published version of the manuscript.

## FUNDING

This study was supported in part by the Natural Science Foundation of Chongqing under Grant cstc2020jcyj-msxmX0284, in part by the Scientific and Technological Research Program of Chongqing Municipal Education Commission under Grant KJQN202000625, in part by the National Natural Science Foundation of China under Grants 61806033, 61703065, and 62101084, and in part by the Key Industry Core Technology Innovation Project of CQ under Grant cstc2017zdcy-zdyfX0012.

## ACKNOWLEDGMENTS

We thank the reviewers for their excellent criticism of the article.

## REFERENCES

Belmaker, R. H., and Agam, G. (2008). Major depressive disorder. *N. Engl. J. Med.* 358, 55–68. doi: 10.1056/NEJMra073096

Bhattacharyya, A., and Pachori, R. B. (2017). A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform. *IEEE Trans. Biomed. Eng.* 64, 2003–2015. doi: 10.1109/TBME.2017.2650259

Bradley, E., and Kantz, H. (2015). Nonlinear time-series analysis revisited. *Chaos Interdiscipl. J. Nonlinear Sci.* 25, 097610. doi: 10.1063/1.4917289

Cai, H., Gao, Y., Sun, S., Li, N., Tian, F., Xiao, H., et al. (2020a). Modma Dataset: A Multi-modal Open Dataset for Mental-disorder Analysis. New York, NY: Cornell University.

Cai, H., Han, J., Chen, Y., Sha, X., Wang, Z., Hu, B., et al. (2018). A pervasive approach to EEG-based depression detection. *Complexity* 2018, 5238028. doi: 10.1155/2018/5238028

Cai, H., Qu, Z., Li, Z., Zhang, Y., Hu, X., and Hu, B. (2020b). Feature-level fusion approaches based on multimodal EEG data for depression recognition. *Inform. Fusion* 59, 127–138. doi: 10.1016/j.inffus.2020.01.008

de la Salle, S., Choueiry, J., Shah, D., Bowers, H., McIntosh, J., Ilivitsky, V., et al. (2016). Effects of ketamine on resting-state EEG activity and their relationship to perceptual/dissociative symptoms in healthy humans. *Front. Pharmacol.* 7, 348. doi: 10.3389/fphar.2016.00348

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Minneapolis, MN.

Fan, Y., Yu, R., Li, J., Zhu, J., and Li, X. (2020). "EEG-based mild depression recognition using multi-kernel convolutional and spatial-temporal feature," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (Seoul: IEEE), 1777–1784. doi: 10.1109/BIBM49941.2020.9313499

Gu, D. (2017). *3D densely connected convolutional network for the recognition of human shopping actions* (Ph.D. thesis). Université d'Ottawa/University of Ottawa, Ottawa, ON, Canada.

Hosseinifard, B., Moradi, M. H., and Rostami, R. (2013). Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. *Comput. Methods Prog. Biomed.* 109, 339–345. doi: 10.1016/j.cmpb.2012.10.008

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7132–7141. doi: 10.1109/CVPR.2018.00745

Jiang, C., Li, Y., Tang, Y., and Guan, C. (2021). Enhancing EEG-based classification of depression patients using spatial information. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 566–575. doi: 10.1109/TNSRE.2021.3059429

Ke, H., Chen, D., Shah, T., Liu, X., Zhang, X., Zhang, L., et al. (2020). Cloud-aided online EEG classification system for brain healthcare: a case study of depression evaluation with a lightweight CNN. *Softw. Pract. Exp.* 50, 596–610. doi: 10.1002/spe.2668

Kerestes, R., Davey, C. G., Stephanou, K., Whittle, S., and Harrison, B. J. (2014). Functional brain imaging studies of youth depression: a systematic review. *Neuroimage Clin.* 4, 209–231. doi: 10.1016/j.nicl.2013.11.009

Kiers, H. A. (1997). Weighted least squares fitting using ordinary least squares algorithms. *Psychometrika* 62, 251–266. doi: 10.1007/BF02295279

Knott, V., Mahoney, C., Kennedy, S., and Evans, K. (2001). EEG power, frequency, asymmetry and coherence in male depression. *Psychiatry Res. Neuroimaging* 106, 123–140. doi: 10.1016/S0925-4927(00)00080-9

Kroenke, K., and Spitzer, R. L. (2002). The PHQ-9: a new depression diagnostic and severity measure. *Psychiatr. Ann.* 32, 509–515. doi: 10.3928/0048-5713-20020901-06

Kukavica, I., and Robinson, J. C. (2004). Distinguishing smooth functions by a finite number of point values, and a version of the takens embedding theorem. *Phys. D Nonlinear Phenomena* 196, 45–66. doi: 10.1016/j.physd.2004.04.004

Liao, S.-C., Wu, C.-T., Huang, H.-C., Cheng, W.-T., and Liu, Y.-H. (2017). Major depression detection from EEG signals using kernel eigen-filter-bank common spatial patterns. *Sensors* 17, 1385. doi: 10.3390/s17061385

Lin, M., Chen, Q., and Yan, S. (2013). *Network in Network*. New York, NY: Cornell University.

Luu, P., and Ferree, T. (2005). *Determination of the HydroCel Geodesic Sensor Nets? Average Electrode Positions and Their 10-10 International Equivalents*. Electrical Geodesics, Inc.

Mahato, S., and Paul, S. (2020). Classification of depression patients and normal subjects based on electroencephalogram (EEG) signal using alpha power and theta asymmetry. *J. Med. Syst*. 44, 1–8. doi: 10.1007/s10916-019-1486-z

Michel, C. M., and Murray, M. M. (2012). Towards the utilization of EEG as a brain imaging tool. *Neuroimage* 61, 371–385. doi: 10.1016/j.neuroimage.2011.12.039

Muldoon, M. R., Broomhead, D. S., Huke, J. P., and Hegger, R. (1998). Delay embedding in the presence of dynamical noise. *Dyn. Stabil. Syst*. 13, 175–186. doi: 10.1080/02681119808806259

Mumtaz, W., Xia, L., Mohd Yasin, M. A., Azhar Ali, S. S., and Malik, A. S. (2017). A wavelet-based technique to predict treatment outcome for major depressive disorder. *PLoS ONE* 12, e0171409. doi: 10.1371/journal.pone.0171409

Olbrich, S., and Arns, M. (2013). EEG biomarkers in major depressive disorder: discriminative power and prediction of treatment response. *Int. Rev. Psychiatry* 25, 604–618. doi: 10.3109/09540261.2013.816269

Olesen, J., Gustavsson, A., Svensson, M., Wittchen, H.-U., Jönsson, B., CDBE2010 Study Group, et al. (2012). The economic cost of brain disorders in Europe. *Eur. J. Neurol*. 19, 155–162. doi: 10.1111/j.1468-1331.2011.03590.x

Peng, H., Xia, C., Wang, Z., Zhu, J., Zhang, X., Sun, S., et al. (2019). Multivariate pattern analysis of EEG-based functional connectivity: a study on the identification of depression. *IEEE Access* 7, 92630–92641. doi: 10.1109/ACCESS.2019.2927121

Pham, T. D., Tran, D., Ma, W., and Tran, N. T. (2015). "Enhancing performance of EEG-based emotion recognition systems using feature smoothing," in *International Conference on Neural Information Processing* (Switzerland: Springer), 95–102. doi: 10.1007/978-3-319-26561-2_12

Puthankattil, S. D., and Joseph, P. K. (2012). Classification of EEG signals in normal and depression conditions by ANN using RWE and signal entropy. *J. Mech. Med. Biol*. 12, 1240019. doi: 10.1142/S0219519412400192

Roulston, M. S. (1999). Estimating the errors on measured entropy and mutual information. *Phys. D Nonlinear Phenomena* 125, 285–294. doi: 10.1016/S0167-2789(98)00269-3

Seal, A., Bajpai, R., Agnihotri, J., Yazidi, A., Herrera-Viedma, E., and Krejcar, O. (2021). Deprnet: a deep convolution neural network framework for detecting depression using EEG. *IEEE Trans. Instrument. Meas*. 70, 1–13. doi: 10.1109/TIM.2021.3053999

Sharma, G., Parashar, A., and Joshi, A. M. (2021). DepHNN: a novel hybrid neural network for electroencephalogram (EEG)-based screening of depression. *Biomed. Signal Process. Control* 66, 102393. doi: 10.1016/j.bspc.2020.102393

Sharma, M., Achuth, P., Deb, D., Puthankattil, S. D., and Acharya, U. R. (2018). An automated diagnosis of depression using three-channel bandwidth-duration localized wavelet filter bank with EEG signals. *Cogn. Syst. Res*. 52, 508–520. doi: 10.1016/j.cogsys.2018.07.010

Sharma, R., and Pachori, R. B. (2015). Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions. *Expert Syst. Appl*. 42, 1106–1117. doi: 10.1016/j.eswa.2014.08.030

Siuly, S., Kabir, E., Wang, H., and Zhang, Y. (2015). Exploring sampling in the detection of multicategory EEG signals. *Comput. Math. Methods Med*. 2015, 576437. doi: 10.1155/2015/576437

Spitzer, R. L., Kroenke, K., Williams, J. B., and Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the gad-7. *Arch. Internal Med*. 166, 1092–1097. doi: 10.1001/archinte.166.10.1092

Stark, J., Broomhead, D. S., Davies, M. E., and Huke, J. (1997). Takens embedding theorems for forced and stochastic systems. *Nonlinear Anal. Theory Methods Appl*. 30, 5303–5314. doi: 10.1016/S0362-546X(96)00149-6

Sun, S., Chen, H., Shao, X., Liu, L., Li, X., and Hu, B. (2020a). "EEG based depression recognition by combining functional brain network and traditional biomarkers," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (Seoul: IEEE), 2074–2081. doi: 10.1109/BIBM49941.2020.9313270

Sun, S., Li, J., Chen, H., Gong, T., Li, X., and Hu, B. (2020b). *A Study of Resting-state EEG Biomarkers for Depression Recognition*. New York, NY: Cornell University.

Van Der Stelt, O., and Belger, A. (2007). Application of electroencephalography to the study of cognitive and brain functions in schizophrenia. *Schizophr. Bull*. 33, 955–970. doi: 10.1093/schbul/sbm016

Wan, Z., Huang, J., Zhang, H., Zhou, H., Yang, J., and Zhong, N. (2020). HybridEEGNet: a convolutional neural network for EEG feature learning and depression discrimination. *IEEE Access* 8, 30332–30342. doi: 10.1109/ACCESS.2020.2971656

Wang, Z., Ma, Y., Liu, Z., and Tang, J. (2019). *R-transformer: Recurrent Neural Network Enhanced Transformer*. New York, NY: Cornell University.

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., et al. (2013). Global burden of disease attributable to mental and substance use disorders: findings from the global burden of disease study 2010. *Lancet* 382, 1575–1586. doi: 10.1016/S0140-6736(13)61611-6

Zhang, X., Li, J., Hou, K., Hu, B., Shen, J., and Pan, J. (2020). "EEG-based depression detection using convolutional neural network with demographic attention mechanism," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (Montreal, QC: IEEE), 128–133. doi: 10.1109/EMBC44109.2020.9175956

**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 Neurorobotics

ORIGINAL RESEARCH
published: 16 June 2022
doi: 10.3389/fnbot.2022.901765

# EEG Identity Authentication in Multi-Domain Features: A Multi-Scale 3D-CNN Approach

*Rongkai Zhang*[1], *Ying Zeng*[1,2], *Li Tong*[1], *Jun Shu*[1], *Runnan Lu*[1], *Zhongrui Li*[1], *Kai Yang*[1] *and Bin Yan*[1]*

[1] *Henan Key Laboratory of Imaging and Intelligent Processing, People's Liberation Army (PLA) Strategic Support Force Information Engineering University, Zhengzhou, China,* [2] *Key Laboratory for Neuro Information of Ministry of Education, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China*

Electroencephalogram (EEG) authentication has become a research hotspot in the field of information security due to its advantages of living, internal, and anti-stress. However, the performance of identity authentication system is limited by the inherent attributes of EEG, such as low SNR, low stability, and strong randomness. Researchers generally believe that the in-depth fusion of features can improve the performance of identity authentication and have explored among various feature domains. This experiment invited 70 subjects to participate in the EEG identity authentication task, and the experimental materials were visual stimuli of the self and non-self-names. This paper proposes an innovative EEG authentication framework, including efficient three-dimensional representation of EEG signals, multi-scale convolution structure, and the combination of multiple authentication strategies. In this work, individual EEG signals are converted into spatial–temporal–frequency domain three-dimensional forms to provide multi-angle mixed feature representation. Then, the individual identity features are extracted by the various convolution kernel of multi-scale vision, and the strategy of combining multiple convolution kernels is explored. The results show that the small-size and long-shape convolution kernel is suitable for ERP tasks, which can obtain better convergence and accuracy. The experimental results show that the classification performance of the proposed framework is excellent, and the multi-scale convolution method is effective to extract high-quality identity characteristics across feature domains. The results show that the branch number matches the EEG component number can obtain the excellent cost performance. In addition, this paper explores the network training performance for multi-scale module combination strategy and provides reference for deep network construction strategy of EEG signal processing.

**Keywords: EEG, identity authentication, multi-scale, 3D-CNN, ERP**

OPEN ACCESS

***Edited by:***
*Duo Chen,*
*Nanjing University of Chinese Medicine, China*

***Reviewed by:***
*Meiyan Xu,*
*Minnan Normal University, China*
*Xugang Xi,*
*Hangzhou Dianzi University, China*

***Correspondence:**
*Bin Yan*
*ybspace@hotmail.com*

***Received:*** *22 March 2022*
***Accepted:*** *19 May 2022*
***Published:*** *16 June 2022*

***Citation:***
*Zhang R, Zeng Y, Tong L, Shu J, Lu R, Li Z, Yang K and Yan B (2022) EEG Identity Authentication in Multi-Domain Features: A Multi-Scale 3D-CNN Approach. Front. Neurorobot. 16:901765. doi: 10.3389/fnbot.2022.901765*

## INTRODUCTION

Electroencephalogram (EEG) identity authentication extracts the neural activity pattern of the user's brain and applies the EEG signal as a biological feature to the individual identity recognition system. Modern brain science research shows (Nakamura et al., 2017) that people brains have not only the structural differences determined by genes such as fingerprints and faces, but also

functional differences in memory, personality, and thinking patterns. EEG has unique advantages in the field of biometric identification, such as *in vivo*, stress resistance, and internality. As a high-level security authentication method, EEG has become a research hotspot. A large number of studies (Marcel and Millan, 2007; Alariki et al., 2018) have verified the possibility of EEG identifying individuals. Individuals show significant individual differences under both evoked (Rathi et al., 2020) and spontaneous (Thomas and Vinod, 2017) tasks.

However, the randomness and weak signal-to-noise ratio of EEG signals restrained the performance of identity authentication, and the current mainstream feature extraction methods have limited improvement in classification performance. It is an urgent problem to improve the performance of EEG identity authentication by constructing fusion features to comprehensively characterize individual EEG features.

Researchers explored a variety of feature fusion methods. In the field of traditional machine learning (Palaniappan and Mandic, 2007), the sensor domain and spatial domain features were spliced, and the fused EEG features effectively improved the classification accuracy. In addition (Arvind et al., 2012), multiple feature domains were arranged and combined. Linear combination (Zhang et al., 2021) was applied in the feature space. Yuan et al. (2018) focused on the time series information of EEG to extract the features through the contextual semantics of EEG, and the proposed method improved the performance of epileptic seizure detection. The above studies have improved the performance of EEG data features.

In this paper, an innovative 3D input representation and multi-scale vision CNN framework is proposed for EEG identity authentication task, which effectively integrates spatial, temporal, and frequency domains in EEG features. The 3D representation of EEG signals comprehensively analyzes the individual's brain identity features from multiple perspectives, and the hybrid mode across feature domains is more integrated and diverse. The multi-scale convolution kernel extracts the brain identity information from different views and characterizes the individual identity by analyzing the individual characteristics of EEG signals from different fields of vision. Furthermore, we discuss the performance and combining strategies of multiple convolution kernels. The proposed multi-scale CNN framework of mixed features significantly improves the performance of EEG authentication. On the feature level, the mixed features integrate the multi-angle information of the research object. On the network level, multi-scale networks further expand the advantages of parallel feature extraction. The proposed method has broad application prospects in dynamic timing information processing, such as video data processing, satellite image analysis, gait recognition, and so on.

The main contributions of this paper are as follows:

(1) An efficient 3D EEG data fusion representation method is proposed to characterize EEG features from multiple perspectives in the spatial, temporal, and frequency domains.
(2) The convolution kernel scale and combination strategy suitable for EEG data were explored, and the rectangular convolution kernel that prefers temporal features has better EEG extraction performance.
(3) The depth of multi-scale modules and the branch number are studied. The branch number should match the number of EEG components, and the shallow multi-branch structure has high application cost performance.

# BACKGROUND

There were some common methods of individual identity in spatial, time, and frequency domains. For example, event-related potential (ERP) components were extracted in time domain (Zeng et al., 2018), power spectral density (PSD) energy of each frequency band in frequency domain (Harshit et al., 2016), and common spatial pattern (CSP) method in spatial domain (Jayarathne et al., 2016). The research results showed that the classification performance of mixed feature domain concatenation was better than single feature domain, and feature fusion methods relying on researchers' experience have been rapidly developed. The above research provided a valuable reference, but the fusion degree of feature domain and feature diversity is still needed to be improved.

In addition to relying on empirical combination features, with the gradual rise of end-to-end learning mode of deep learning framework, automatic and non-linear feature fusion methods based on neural network were widely used, and the efficiency and performance of fusion features were better than traditional empirical methods. Specifically, after the application of multi-scale convolution model in two-dimensional image processing, the features of different scale fields were automatically extracted and fused by parallel structure. We summarized that the latest research improves the input form to three-dimensional, and the convolution neural network (CNN) performed rich, comprehensive, and complementary feature domain fusion. The multi-scale convolution network in the field of EEG had a lot of exploration. The application scenarios included emotion recognition, fatigue driving (Cho and Hwang, 2020), epilepsy prediction (Ozcan and Erturk, 2019), and motor imagery (Zhang et al., 2019) tasks.

## EEG Presentation Form

For 3D CNN, several methods were designed to convert scalp EEG signals into 3D presentation forms. Common dimensions were assigned to the main features of two dimensions and the auxiliary features of one dimension. Zhao et al. (2019a) demonstrated that in the motor imagery task, the strategy of electrode combination was studied. According to the electrode position, it was mapped to a $3 \times 3$ or $4 \times 4$ two-dimensional plane to form spatial features, and another dimension shows temporal features. In Deng et al. (2021) research, a two-dimensional $9 \times 9$ spatial representation matrix was designed for 32-channel electrodes, and the other dimension was the temporal domain signal of EEG array. Zhang et al. (2019) combined 2D temporal features and 1D frequency feature in the motor imagery task to form a three-dimensional feature domain, representing the time-frequency feature tensor of EEG. In the study of psychological load (Kwak et al., 2020), EEG signals were presented as 2D spatial

information and 1D frequency information and focused on the rhythm characteristics of EEG frequency band.

For different cognitive tasks, 3D feature combination focuses on two kinds of temporal, frequency, and spatial domains and abandons a feature domain as a cost, which is not fully utilized. The rich temporal, frequency, and spatial features of EEG data have not been fully excavated, and more efficient 3D feature domain fusion method remains to be explored.

## Multi-Scale Convolution Module

The three-dimensional convolution model applied in the field of EEG is still dominated by single convolution kernel and serial stack architecture. Inadequate advanced studies have found that the performance of parallel multi-scale convolution structure was better than serial mode, and the parallel structure of multi-scale is beneficial to capture the characteristics of different scales. The inception module (Lee et al., 2020) in the field of image was used in the task of brain control manipulator to obtain better stability and accuracy in decoding motion intention. In addition, Zhao et al. (2019b) research designed small receptive field network (SRF), medium receptive field network (MRF), and large receptive field network (LRF), using multi-scale three-dimensional dimensions ($2 \times 2 \times 1$, $2 \times 2 \times 3$, $2 \times 2 \times 5$) to extract EEG features.

The combination strategy of convolution kernel continues the experience in the field of image, without improving the characteristics of EEG. Moreover, insufficient study researched on the performance of different convolution sizes in the EEG field, and the editing and parameter setting were based on the researcher experience. The above research shows that the fusion feature domain of EEG identity authentication is not sufficient and diverse, and the performance of single convolution kernel and multi-scale combination strategy needs to be studied.

# METHODS

## Database

As shown in **Figure 1**, the EEG authentication task (Zhang et al., 2022) was to distinguish between the self and non-self-names. The oddball paradigm randomly showed 500 names, including 100 subjects' own names, 150 familiar names, 300 strangers' names, and 50 blank names. The acquaintances were the names of the subjects' family members or closest friends, which were provided by the subjects before the experiment. The names of acquaintances interfere greatly with the subjects, which was used as a supplement to the control of non-self-names in the experiment. Strange names were randomly selected from the citizen name database. The subjects confirmed that each stranger was unfamiliar before the experiment. Blank name was black background and no text. To eliminate the influence of cross and mask factors on participants, the study only retained the brain response of name stimuli. In the preprocessing stage, the blank name was subtracted as the name response baseline of the brain. First, the blank names of all subjects were collected and averaged globally. Second, the global average of blank names was subtracted from the single trial of each self, familiar, and stranger name.

Electroencephalogram identity authentication experimental flow chart is shown in **Figure 2**. In the preprocessing stage, 0.1–20 Hz band-pass filter was used to filter out slow drift and high-frequency noise offline. Band-pass filter consisted of low-pass and high-pass Chebyshev filters. The low-pass Chebyshev filters (order, 2; stopband starting frequency, 20 Hz; stopband cutoff frequency, 40 Hz; attenuation in the passband, 0.5 dB; attenuation in the stopband, 5 dB) were acquired through the built-in function of MATLAB. The high-pass Chebyshev filters (order, 2; stopband starting frequency, 0.01 Hz; stopband cutoff frequency, 0.5 Hz; attenuation in the passband, 0.5 dB; attenuation in the stopband, 10 dB) were also acquired. The original EEG data were down-sampled to 256 Hz to reduce the amount of data processing and calculation. Reference electrode standardization technique (REST) was applied for re-reference, and this method was considered to have higher accuracy in traceability and brain network analysis. The independent component analysis (ICA) was used to remove eye electrical artifacts in the signal. The ICA decomposition of EEG signals was realized by Fast-ICA toolbox of MATLAB, and the method applied the default parameters. We observed the decomposition waveform and determined the electrooculogram component, and this study only excluded the component of ocular artifacts. After the electrooculogram channel was set to zero, the rest of the ICA signal was converted back to the EEG signal. The test with amplitude $>100$ μv was automatically eliminated. EEG signals were cut into a period from $-200$ to 800 ms. The baseline of EEG signals was from $-200$ to 0 ms before stimulus presentation.

We invited 70 subjects (52 men and 18 women, age range of 18–25, standard deviation 1.7 years, 70 right-handed) to participate in the experiment, all of them were college students and had normal or corrected-to-normal visual ability. None of the subjects had a history of psychiatric illness or taking psychotropic drugs. The experimental design was approved by Ethics Committee of China National Digital Switching System Engineering and Technological Research Center. Each participant filled in the informed consent before the experiment and obtained a payment after completing the experiment. Compared with the current EEG database of 20–30 people, this experiment applied the relatively sufficient data of 70 people database. A total of 70 subjects were divided into training sets (50 subjects) and test sets (20 subjects). To avoid the contingency and increase the reproducibility, the participants in the training set and the test set were randomly selected. The result was the experiment average for 100 rounds of random training set and test set. To avoid the overfitting problem of identity authentication model, the two types of data sets have no overlap to test the robustness of the method.

The collected EEG signals were seriously disturbed by noise. The valuable ERP signal amplitude at $10^{-6}$ voltage, and any subtle influence seriously interferes with EEG data. Therefore, the applied EEG analysis and processing method needs to have good robustness. The ERP amplitude of valuable information in this experiment was below 10 μv, but the magnitude of other interference was above 100 μv. The collected EEG signals were

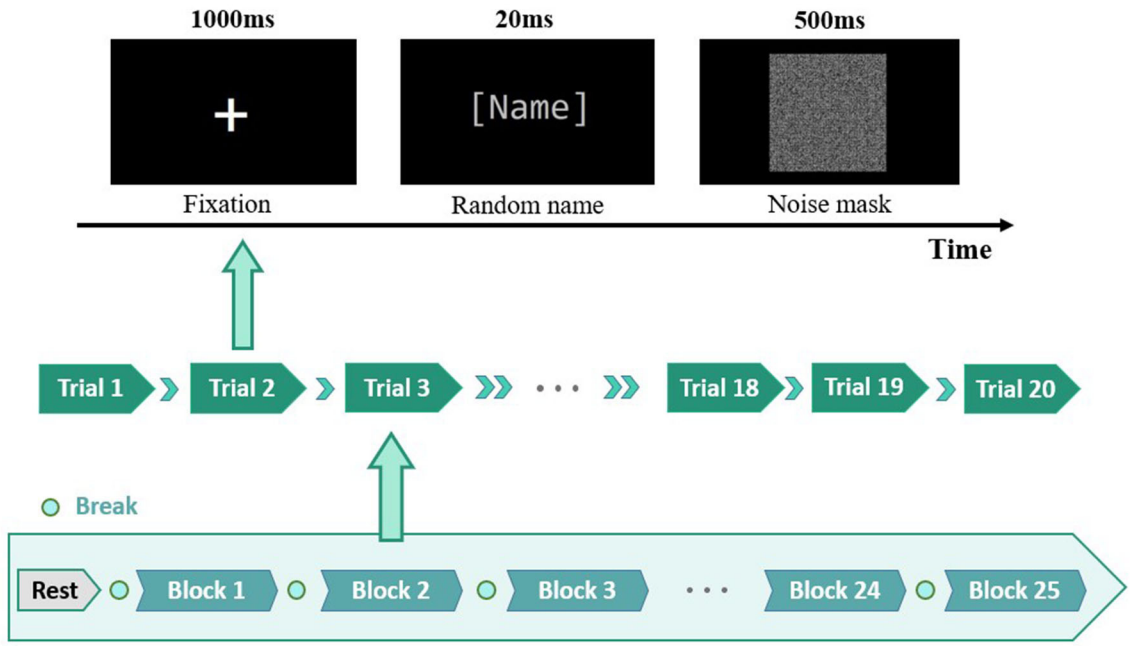



**FIGURE 1 |** Task paradigm of EEG authentication.

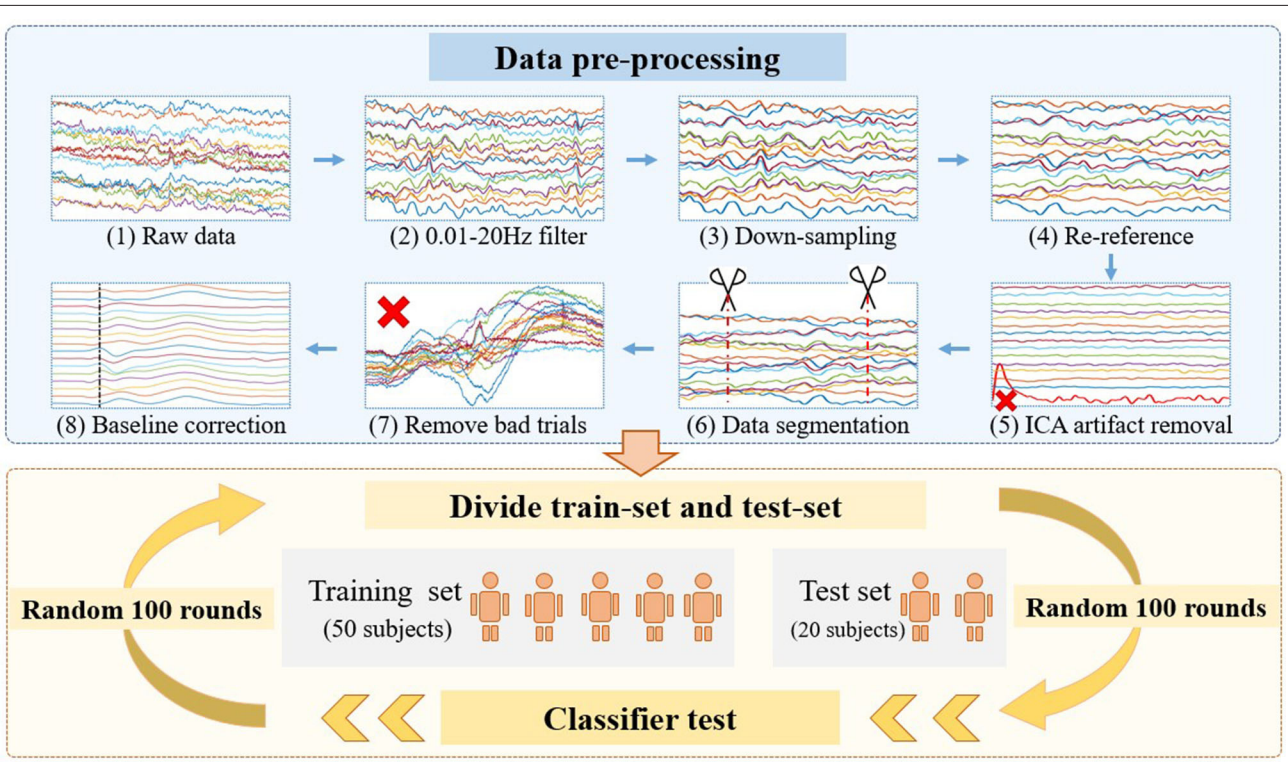



**FIGURE 2 |** EEG identity authentication experimental flow chart.

mainly in the following three categories: environmental noise, electromyographic artifact, and psychological. Environmental noise includes equipment noise, power frequency interference, electromagnetic interference, etc. The artifacts include eye movements, head movements, and other EMG signals. In addition, EEG signals are affected by physical conditions such as

mood, circadian clock, sleeping, and other psychological factors. To ensure the robustness of the proposed method, we have two special designs in our experiment. On the one hand, we collected a large-scale EEG database of 70 people, which tested the method's universality and effectiveness in the group. On the other hand, the experimental results were derived from 100 rounds randomly selected of training set and test set. The average results avoid the specificity and contingency of the proposed method.

## EEG 3D Presentation Form

To present more comprehensive and abundant EEG features, the 3D representation of EEG was assigned one dimension each for the spatial, temporal, and frequency domains. The three common feature domains were reflected in the three-dimensional representation, which fully characterizes individual identity from multiple perspectives. The neural activity recorded by EEG was shown in the time, frequency, and spatial domains. As shown in **Figure 3A**, the EEG contains P200, P300, and late negativity (LN) components, which represent the neuron activity of name stimuli. P200 shows the brain's pre-attention of visual stimuli, representing the early cognitive process. P300 is the most prominent component of EEG in this task. The presence of P300 means that the subjects' brains fully perceive name stimuli and stimulate large-scale, intense brain activity. Previous studies have found that LN belongs to post-processing of consciousness, indicating that stimulus materials are deeply processed in the brain. **Figure 3B** reveals the main energy centers at 375–625 ms in time domain and 0–10 Hz in frequency domain, which indicates that the brain's response to names is low frequency. In the spatial domain, the main energy was concentrated in the Fz, F3, F4, C3, and C4 channel. Brain activation is located in the frontal lobe and central area, which is responsible for advanced cognition and regulation of information.

The 3D transform process of EEG signals is shown in **Figure 3**. First, single-trial EEG data after preprocessing were separated according to different electrode channels. Second, the EEG signal time-frequency diagram conversion. The time-frequency diagram was obtained by the spectrum function in MATLAB. The spectrum parameters were set as follows: hamming window length 64, overlapping window length 57, and Fourier transform point 256. The size of the time-frequency graph of the processed single channel was 94 × 60, representing the time domain information of 0–800 ms and the frequency domain energy of 0–30 Hz, respectively. The stacking of 16 layers sensors formed the spatial representation. The spectrum parameter settings were constrained by the time-frequency characteristics of EEG signals and CNN input form. In terms of CNN input form, CNN was first outstanding in the field of image processing. Some studies have shown that CNN has outstanding effect on square images. Therefore, this experiment tried to generate time-frequency graphs close to square images. In terms of EEG characteristics, ERP signals showed high time information and low-frequency energy. Based on the above two points, we obtained the 94 × 60 square time-frequency diagram by setting the parameters (window length, 64; overlapping length, 57). In addition, the number of Fourier transform 256 points was the EEG sampling rate. Hamming window was widely used to reduce spectrum leakage and maintain good frequency resolution. Then, the spectrum of each channel is stacked to form a 16 × 94 × 60 three-dimensional input tensor, which represents the spatial, temporal, and frequency characteristics of EEG. Data processing decomposes EEG signals into mixed tensors of three feature domains, which is beneficial to feature fusion of neural network training.

## Multi-Scale Vision Convolution Module

The individual identity feature of EEG was mainly extracted by multi-scale visual field convolution module, and multiple modules were spliced to form the overall network architecture. The previous data were applied to multiple branch operations, each branch connecting a convolution kernel of a unique scale. The multi-scale vision convolution module is shown in **Figure 4**. The multi-scale convolution processing flow included 1 × 1 convolution layer, branch structure, feature concatenation, and pooling.

The first input layer applied dimension reduction through 1 × 1 convolution layer, which reduced network parameters and integrates local correlation. Then, the 3D tensor was padded according to the scale of convolution kernel to meet the consistency of feature size after multi-scale convolution. To prevent gradient explosions or disappearances, the 3D convolution kernel was initialized by "kaiming_uniform." Multi-branch convolution was an important structure for multi-scale feature extraction, and diversified convolution kernel combination strategy was the discussion focus in this study. Compared with the common square convolution kernel in the image field, considering the ERP signal characteristics of unbalanced information in temporal and frequency domains, a variety of rectangular convolution kernels with unequal edge lengths were explored in the study.

For the excellent multi-scale strategy [(6 × 3 × 3), (6 × 5×3), (6 × 7 × 3)], the experimental results are shown in **Figure 4**, and the 3D convolution kernels correspond to spatial × temporal × frequency domains, respectively. Multi-scale convolution kernel combination was conducive to capture EEG features from macro to detail, and multi-branch structure extracted diverse and complementary identity information. The feature output layer was regularized by batch normalization to improve the generalization ability and convergence speed of the model. The concatenate feature layer was reduced by maxpool, and the pooling kernel area decreased with the deepening of the network. The feature layer output contained comprehensive features extracted by multi-convolution kernel and transfers the three-dimensional information into the next stage.

## EEG Authentication Framework

The overall EEG authentication architecture was the stack and expansion of multi-scale convolution modules, and the number of branches and modules was flexibly combined. As shown in **Figure 5**, multi-branch structure within the module extracted diversity features, and multi-layer module stack can refine and combine the features. The multi-branch structure extracted diversity features in the basic module, and multi-layer module stacking can refine and combine the features. The end multi-scale

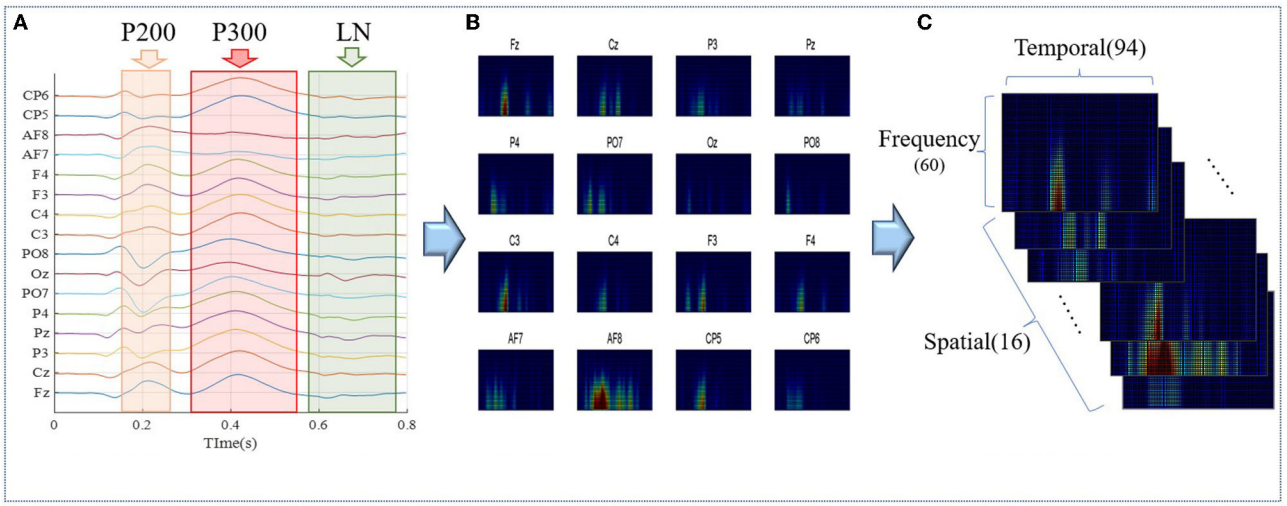



**FIGURE 3 |** Three-dimensional transformation flow chart of EEG. **(A)** EEG after preprocessing. **(B)** Time-frequency energy diagram of separated channels. **(C)** Spatial–temporal–frequency domain combinations of EEG three-dimensional tensor.

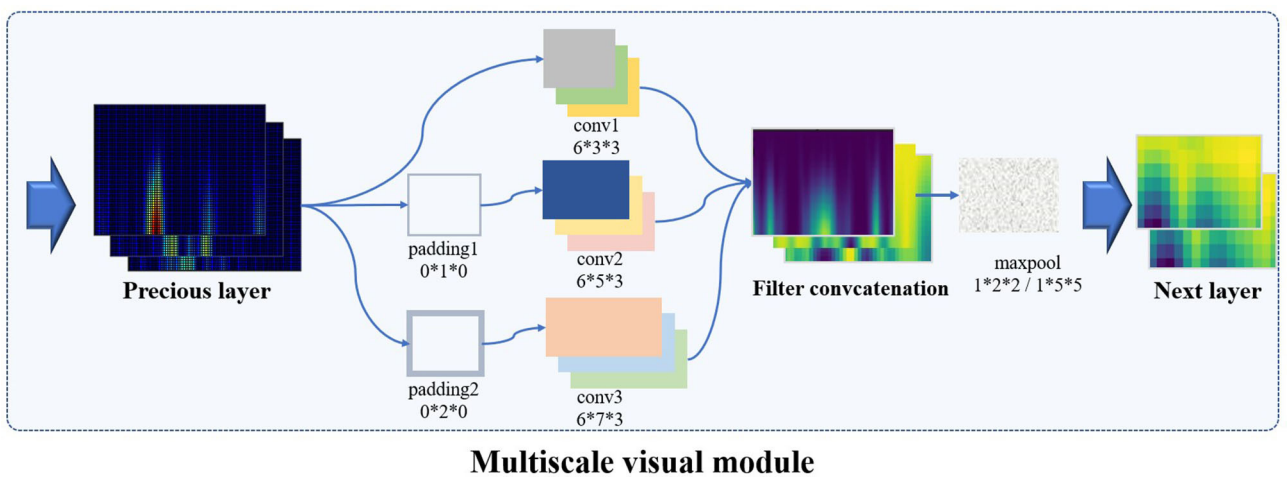



**FIGURE 4 |** The architecture of multi-scale vision convolution module.

convolution module flattens the features and then passes through a two-layer fully connected network of 768 × 32. The activation function used leaky rule ($\alpha$ =0 .01) to prevent dead neurons. To reduce the overfitting risk, dropout with a probability of 0.4 was inserted between the fully connected layers. The network parameters were set as follows: learning rate = 0.0005, epoch = 5, batch size = 50. The details of the network structure used in the experiment are shown in **Table 1**.

It should be noted that the scale convolution framework should find a balance between network performance and training cost. Since the increase in the number of modules and branch structure will lead to a huge parameters' number, branch construction and module elimination are the main methods to reduce the computational and time consumption. The network framework in the study has flexibility and diversity. The experiment analyzes the several combination strategies of multi-branch and multi-layer modules and selects the architecture scheme that balances high performance and low time consumption.

# RESULTS

## Single-Scale Kernel Performance

The study explored the convolution kernel size suitable for EEG feature extraction and provided reference for multi-scale convolution kernel combination strategy. **Figures 6**, **7** show the training loss and classification performance of 16 single-scale convolution kernels. Due to the small spatial variability of 16 sensor space, the spatial depth of the fixed convolution kernel is 6. The study focuses on the feature information in time-frequency

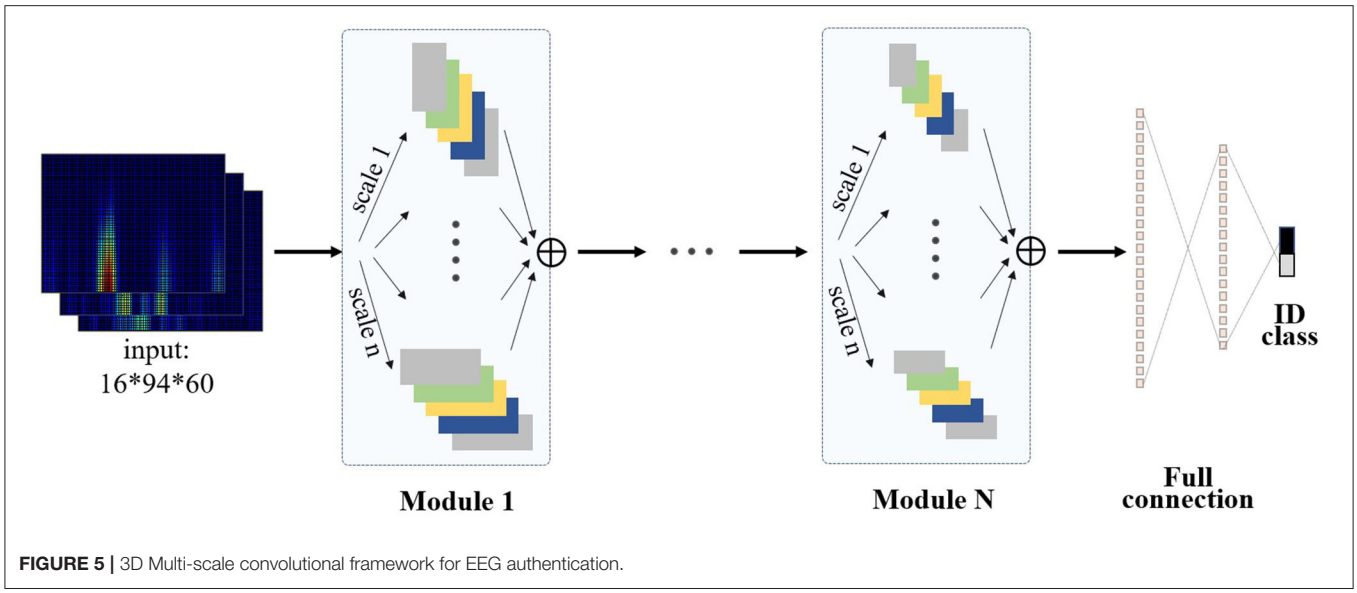



**FIGURE 5 |** 3D Multi-scale convolutional framework for EEG authentication.

domain, and the convolution kernel size in the legend is time × frequency domain pixel).

**Figure 6** shows the model training losses of convolution kernels with different scale. **Figure 6A** reveals important inflection points for convergence curves of different convolution kernels. The loss function curve shows two processes of rapid convergence and gentle decline. Model converges rapidly in the first 1,000 iterations. All convolution kernels reach the inflection point between 1,000 and 3,000 iterations, and the loss function decreases slowly after 3,000 rounds. As shown in **Figure 6B**, the top 3D kernels with the best convergence performance in time-frequency domain are 3 × 3, 5 × 3, and 7 × 3. The convolution kernels with the same area generally show that the long shape is better than the wide shape (e.g., 5 × 3 is better than 3 × 5), indicating that the attention degree of temporal characteristics mainly affects the convergence performance. The convolution kernels with different areas show that small size is better than large size (e.g., 3 × 3 is better than 7 × 7), indicating that the capture of EEG detail features by small size convolution kernel is beneficial for model training.

The classification test applied a 5-fold cross-validation method in the individual data set. A 5-fold cross-validation is a commonly used grouping test method in classification tasks, especially suitable for performance prediction of small-scale databases. The long duration of EEG experiments leads to subject discomfort and affects the data quality. The individual in single EEG experiment collected insufficient samples, and the samples' number of each type was <100. Individual model performance evaluation belongs to small-scale database testing. Therefore, we apply 5-fold cross-validation to reduce individual model overfitting. **Figure 7** shows that the classification performance of different convolution kernels is consistent with the convergence performance, and the small rectangular convolution kernel has better classification performance in ERP feature extraction. The classification performance of small-size and long kernel shape is better.

Overall, the shape of convolution kernel affects the performance of the model more than the size, and the advantage order of convergence shapes under similar sizes is long, wide, and square shape. Rectangular shape is more suitable for ERP EEG feature model training, which is different from the experience of square convolution kernel in image processing field. Based on the above single-scale convolution kernel results, the smaller-scale convolution kernel is selected in the combined multi-scale convolution kernel strategy, and the shape difference in convolution kernel is retained for in-depth analysis. At present, our known literature retrieval has not found the research of the convolution kernel size in the EEG field. This study provides suggestions for convolution kernel selection of ERP tasks, including kernel size and strategy. Innovative research finds that rectangular cores and small sizes have better performance than common specifications in ERP tasks. In the construction of network framework, the convolution kernel can be selected and combined according to the design and strategy of this paper, which will be conducive to the improvement in EEG network performance.

## 3D-CNN Performance Evaluation

After quantitative analysis of the network performance of single-scale convolution kernel, the convolution kernel with excellent single-scale performance is composed of multi-scale convolution network. Taking the small size and non-square convolution kernel as the basic branch, seven kinds of convolution kernel combination strategies are proposed, and the convolution kernel combinations of long, wide, and square shapes are tested, respectively. One-dimensional multi-scale convolution model is vector input and only has long convolution kernel base. In addition, three commonly used single-scale CNN models

**TABLE 1 |** Architecture of main neural networks for EEG authentication.

| Type | | 1D-simple | 2D-simple | 3D-simple scale-length | 1D-Multi-scale-width | 2D-Multi-scale-equal | 2D-Multi-scale-length | 2D-Multi-scale-width | 3D-Multi-scale-width | 3D-Multi-scale-equal | 3D-Multi-scale-length |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Multi-scale visual module(a) | Conv(a1)/ padding | (24) × 1 × 7/0 × 0 | (24) × 9 × 7/0 × 0 | (24) × 5 × 9 × 7/0 × 0 | (24) × 1 × 7/0 × 2 | (24) × 3 × 7/0 × 2 | (24) × 7 × 7/2 × 2 | (24) × 7 × 3/0 × 2 × 0 | (24) × 6 × 3 × 7/0 × 0 × 2 | (24) × 6 × 7 × 7/0 × 2 × 2 | (24) × 6 × 7 × 3/0 × 2 × 0 |
| | Conv(a2)/ padding | / | / | / | (24) × 1 × 5/0 × 1 | (24) × 3 × 5/0 × 1 | (24) × 5 × 5/1 × 1 | (24) × 5 × 3/0 × 1 × 0 | (24) × 6 × 3 × 5/0 × 0 × 1 | (24) × 6 × 5 × 5/0 × 1 × 1 | (24) × 6 × 5 × 3/0 × 1 × 0 |
| | Conv(a3) /padding | / | / | / | (24) × 1 × 3/0 × 0 | (24) × 3 × 3/0 × 0 | (24) × 3 × 3/0 × 0 | (24) × 3 × 3/0 × 0 | (24) × 6 × 3 × 3/0 × 0 × 0 | (24) × 6 × 3 × 3/0 × 0 × 0 | (24) × 6 × 3 × 3/0 × 0 × 0 |
| Max pool(a) | 1 × 2 | 2 × 2 | 1 × 2 × 2 | 1 × 2 | 2 × 2 | 2 × 2 | 2 × 2 | 1 × 2 × 2 | 1 × 2 × 2 | 1 × 2 × 2 | |
| Multi-scale visual module(b) | Conv(b1) /padding | (72) × 1 × 7/0 × 0 | (24) × 12 × 8/0 × 0 | (72) × 7 × 12 × 8/0 × 0 | (72) × 1 × 7/0 × 2 | (72) × 3 × 7/0 × 2 | (72) × 7 × 7/2 × 2 | (72) × 7 × 3/0 × 2 × 0 | (72) × 6 × 3 × 7/0 × 0 × 2 | (72) × 6 × 7 × 7/0 × 2 × 2 | (72) × 6 × 7 × 3/0 × 2 × 0 |
| | Conv(b2) /padding | / | / | / | (72) × 1 × 5/0 × 1 | (72) × 3 × 5/0 × 1 | (72) × 5 × 5/1 × 1 | (72) × 5 × 3/0 × 1 × 0 | (72) × 6 × 3 × 5/0 × 0 × 1 | (72) × 6 × 5 × 5/0 × 1 × 1 | (72) × 6 × 5 × 3/0 × 1 × 0 |
| | Conv(b3) /padding | / | / | / | (72) × 1 × 3/0 × 0 | (72) × 3 × 3/0 × 0 | (72) × 3 × 3/0 × 0 | (72) × 3 × 3/0 × 0 | (72) × 6 × 3 × 3/0 × 0 × 0 | (72) × 6 × 3 × 3/0 × 0 × 0 | (72) × 6 × 3 × 3/0 × 0 × 0 |
| Max pool(b) | 1 × 2 | 4 × 4 | 1 × 4 × 4 | 1 × 2 | 2 × 2 | 2 × 2 | 2 × 2 | 1 × 2 × 2 | 1 × 2 × 2 | 1 × 2 × 2 | |
| Multi-scale visual module(c) | Conv(c1) /padding | (72) × 1 × 7/0 × 0 | (24) × 6 × 4/0 × 0 | (72) × 6 × 6 × 4/0 × 0 | (72) × 1 × 7/0 × 2 | (72) × 3 × 7/0 × 2 | (72) × 7 × 7/2 × 2 | (72) × 7 × 3/2 × 0 | (72) × 6 × 3 × 7/0 × 0 × 2 | (72) × 6 × 7 × 7/0 × 2 × 2 | (72) × 6 × 7 × 3/0 × 2 × 0 |
| | Conv(c2) /padding | / | / | / | (72) × 1 × 5/0 × 1 | (72) × 3 × 5/0 × 1 | (72) × 5 × 5/1 × 1 | (72) × 5 × 3/0 × 1 × 0 | (72) × 6 × 3 × 5/0 × 0 × 1 | (72) × 6 × 5 × 5/0 × 1 × 1 | (72) × 6 × 5 × 3/0 × 1 × 0 |
| | Conv(c3) /padding | / | / | / | (72) × 1 × 3/0 × 0 | (72) × 3 × 3/0 × 0 | (72) × 3 × 3/0 × 0 | (72) × 3 × 3/0 × 0 | (72) × 6 × 3 × 3/0 × 0 × 0 | (72) × 6 × 3 × 3/0 × 0 × 0 | (72) × 6 × 3 × 3/0 × 0 × 0 |
| Max pool(c) | 1 × 6 | 2 × 2 | / | 1 × 5 | 5 × 5 | 5 × 5 | 5 × 5 | 1 × 5 × 5 | 1 × 5 × 5 | 1 × 5 × 5 | |
| Dropout | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | |
| Linear | 16 × 3 × 72 | 72 × 3 × 2 | 72 × 3 × 2 × 1 | 16 × 4 × 72 | 4 × 2 × 72 | 4 × 2 × 72 | 4 × 2 × 72 | 1 × 4 × 2 × 72 | 1 × 4 × 2 × 72 | 1 × 4 × 2 × 72 | |
| Dropout | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | |
| Linear | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | |
| Softmax | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | |

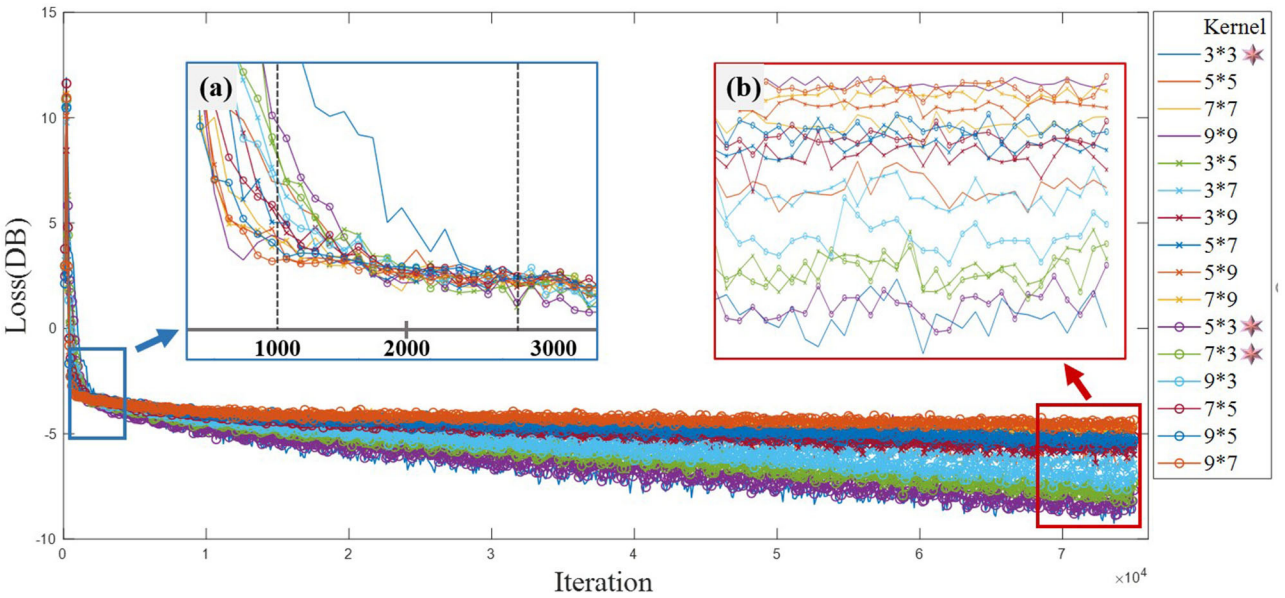



**FIGURE 6 |** Network training loss of single-scale convolution kernel. **(a)** The important inflection points of network loss convergence. The loss functions of different convolution kernels decrease rapidly before 1,000 iterations. Additionally, the loss function curves of all convolution kernels tend to be flat before 3,000 rounds. **(b)** The loss function performance of different convolution kernels after multi-round iteration. Long-shape and small-size convolution kernels are suitable for feature extraction of ERP tasks.

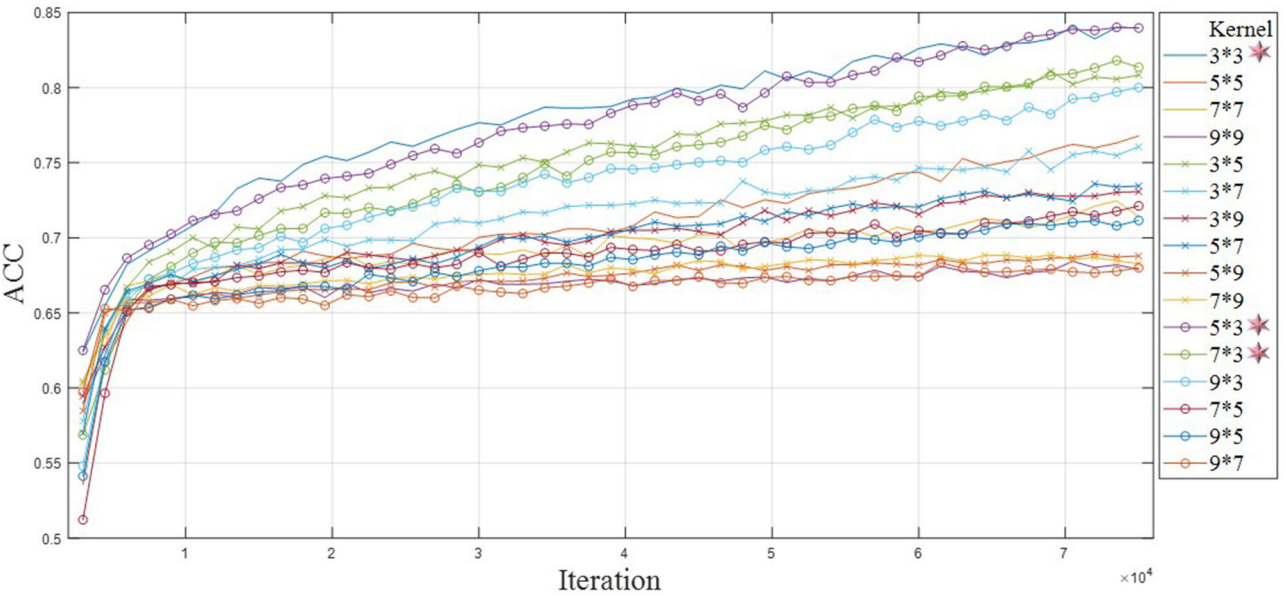



**FIGURE 7 |** Test accuracy of single-scale convolution kernel.

and seven machine learning classification are compared in the experiment. The details of the network model are shown in **Table 1**.

The results in **Figure 8** show that the classification performance of the traditional machine learning classifier is similar to that of the serial CNN model. The performance of traditional classifiers is quite different. Discriminant analysis classifier (DAC), support vector machine (SVM), and ensemble classifiers obtain higher classification accuracy, and their performance is close to the single-scale 3D-CNN network. The performance of the serial single-scale CNN model is affected by the data input dimension, and the fusion representation of

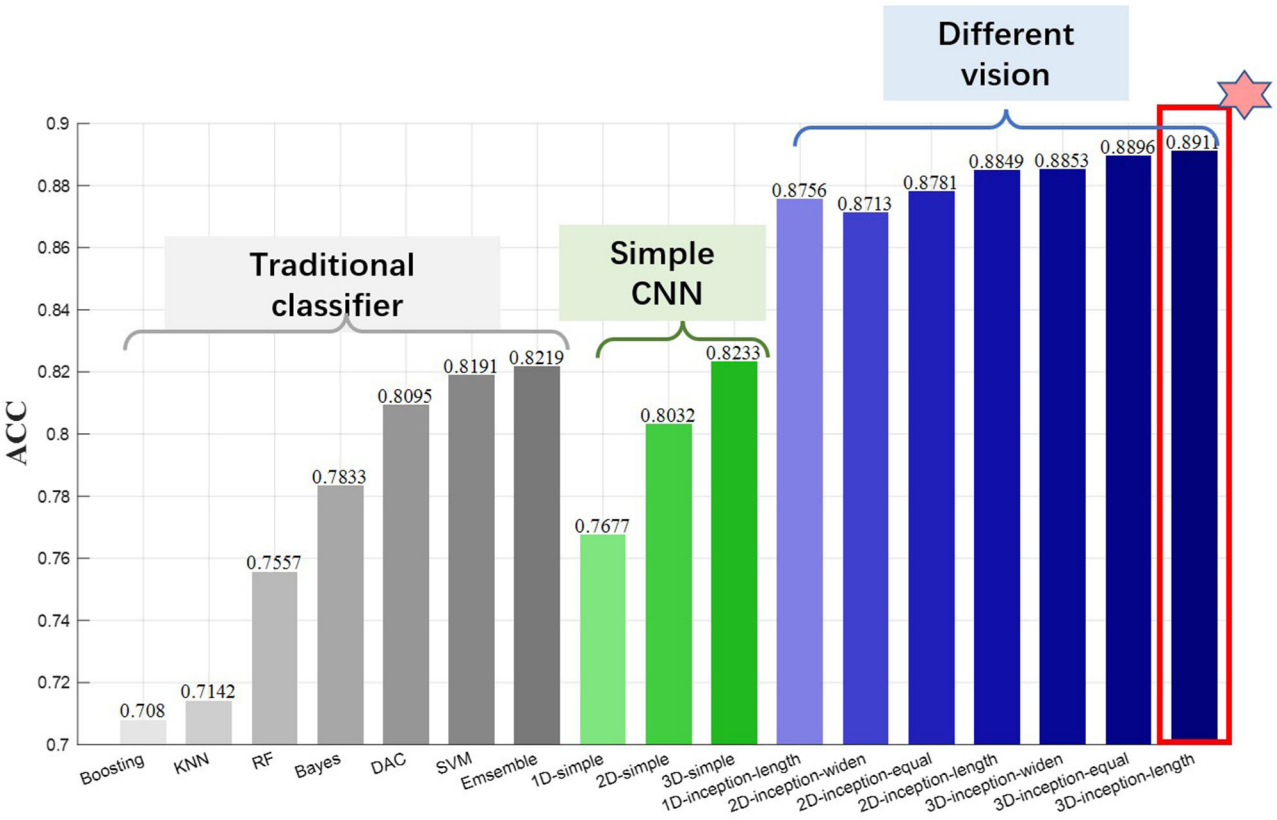



**FIGURE 8 |** Classification performance of multiple authentication strategies.

multi-feature domains effectively improves the classification performance. The 3D single-scale convolutional network achieves 82.33% of the identity authentication accuracy, which is higher than the optimal traditional machine learning algorithm.

The convolution architecture of multi-scale visual field shows outstanding classification performance, and its multi-view feature extraction method obtains more than 5% of accuracy improvement compared with single-scale serial network. In addition, the multi-scale network also shows the advantages of high-dimensional feature fusion. The fusion features of multi-feature domain can extract individual identity information more accurately, and the accuracy rate increases with the increase of feature input dimension.

Compared with the 2D network, the 3D convolutional network proposed in the study has an improved accuracy rate, indicating that the spatial features extracted from the third dimension assist the classification decision. The cognitive process of self-name stimulation is divided into several stages: visual coding, semantic understanding, decision analysis, self-awareness, advanced cognition, neural feedback, etc. Different brain functional areas are responsible for the part of the cognitive process. Therefore, EEG signals of different electrodes have different contributions. However, the commonly used 2D-CNN lack the freedom degree of EEG spatial data, which makes it difficult to refine the importance of each electrode. The 3D-CNN is flexible in the third dimension, which helps to utilize spatial information of electrodes in different brain regions. In this article, the total number of channels is 16, and the convolution kernel size in the spatial dimension is 6. The 3D convolution kernel performs sliding window operation on the third dimension, and the convolution results are rich in 3D space layout. The 3D presentation helps the convolution network to obtain multi-dimensional optimization space, which is conducive to the optimization of network parameters and the improvement in classification performance. Since each functional area of the brain is responsible for different information processing tasks, the overall realization of the task requires the cooperation of multiple brain regions and the connection of the brain network, so the three-dimensional EEG presentation form is more suitable for the EEG biological characteristics. EEG acquisition equipment often uses multiple sensors to extract the neuron discharge signals of the whole brain, and the scalp electrical signal data between multiple sensors are cross-complementary, which effectively alleviates the spatial information lack of the single sensor. The 3D network model obtains the spatial correlation of EEG through the convolution vector of the third dimension and further extracts the spatial pattern of brain activity.

In the performance of the combination strategy of convolution kernel shape, the 2D and 3D networks show that long convolution kernel is better than wide and square. It is worth noting that the multi-scale convolution model with one-dimensional input achieves 87.56% of accuracy, and its classification effect is higher than that of the two-dimensional multi-scale network with wide shape. The advantages of rectangular convolution kernel in classification test are consistent with those of single convolution kernel, indicating that the rectangular kernel shape is more suitable for the ERP task network model.

Summarizing the results of the above three groups classification models, we find that the diverse features of multi-scale convolution kernel help to improve the classification performance, and the contribution of multi-scale architecture is better than that of multi-dimensional feature representation. In addition, a variety of long (temporal domain priority) convolution kernel combinations are suitable for EEG feature extraction.

## Analysis of Module Branch and Network Depth

To explore the influence of the breadth and depth of the network on the classification performance, the architecture of multi-branch and multi-layer networks was tested. Each branch assigns the operations of a single-scale convolution kernel. The multi-scale convolution kernel is expanded on the commonly used three branches, and the commonly used depth is selected for the network layer. The 2–6 branches are combined with the 3–5 module layers architecture, and the classification test results are shown in **Figure 9**.

The classification performance of multi-scale convolutional networks increases with the increase of layers and branches, indicating that the expansion of network parameters is conducive to feature extraction and classification. When 2 branches expand to 3 branches, the accuracy increases most significantly, and the increase in the branch number is limited after 3 branches. This shows that the convolution kernel of three branches can cover the basic EEG features, and too many branches increase the repeated acquisition and redundancy of features. In addition, the classification accuracy tends to be stable when the network depth is >3 layers, and the stacking module depth has limited performance improvement. In addition, complex networks require massive data and lead to overfitting.

The increase of branches and layers is accompanied by the improvement in training parameters, and the improvement in classification performance by complex networks is limited. The results show that the network architecture with three layers and three branches has good application cost performance, and the shallow network with fewer branches can obtain prominent identity authentication results.

## Feature Visualization of Multi-Scale Kernel

To analyze the extraction effect of multi-scale model for EEG identity features, we take three-branch and three model layers as an example to analyze the attention areas of different scale convolution kernels. Each branch outputs the feature heat map after convolution operation by the own scale and normalizes the attention intensity of the feature map by z-score. **Figure 10** illustrates the feature layer extraction process of multi-scale network and shows the representative results of attention heat map.

For the results of the longitudinal single-layer feature, the multi-scale structure captures the diversity of EEG features, and the attention regions of different convolution kernels complement each other. We assume that the attention of multiple branches corresponds to the important components of EEG, such as the three convolution kernels of feature layer 1 capture the main features of ERP, respectively. Branch 1 corresponds to P300 component (a1), branch 2 corresponds to P200 component (a2), and branch 3 corresponds to early cognition and late negativity component (a3). In addition, EEG components affect convolution kernel size, the smaller convolution kernel (a1) size leads to the smaller span of captured features, and (a3) the longer convolution kernel size covers the large-scope features.

For the horizontal multi-layer feature inheritance, the bottom to the deep features represents good inheritance. The network effectively retains EEG features in layer-by-layer iteration and reduces the dimension of high-dimensional raw data to low-dimensional identity features. For example, (a1) (b1) (c1) feature maps retain the P300 component features in the conduction and convert the 2D time-frequency map of (a1) to one-dimensional vector of (c1). After compression in time domain and frequency domain, the P300 feature in (c1) still be clearly presented, and the corresponding position conforms to the temporal energy distribution. In addition, (c2) and (c3) also inherited the P200 and LN components of the bottom heat map, respectively. The results show that the multi-scale model integrates the early EEG features into the corresponding low-dimensional identity information.

The results in **Figure 10** show that the features extracted by multi-scale model are diverse, complementary, and inheritable. The feature heat map is similar to the energy distribution of EEG in time domain, and the frequency domain features are merged and compressed, which is confirmed by the temporal priority of the early kernel result. In addition, we suggest that the branch number of the multi-scale module matches the main component number of ERP and select the multi kernel strategy of the corresponding ERP component size, which helps the branch structure of the multi-scale vision to focus on the EEG characteristics suitable for the respective scales.

## Time Consumption

The practical application of BCI system considers the computational power and time consumption of model construction, and the actual system needs to establish a balance between classification accuracy and time cost. **Figure 11** shows the time-consuming training of the classification model for research and application, and the model training uses two high-performance CPU chips (Intel (R) Xeon (R) Gold 5118). The proposed model applies the large network parameters

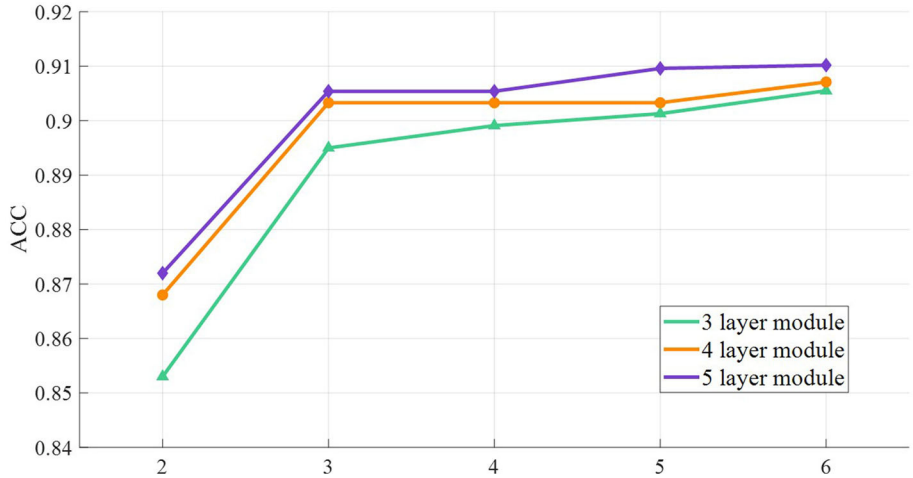



**FIGURE 9 |** Classification performance of multi-layer, multi-branch network architecture.

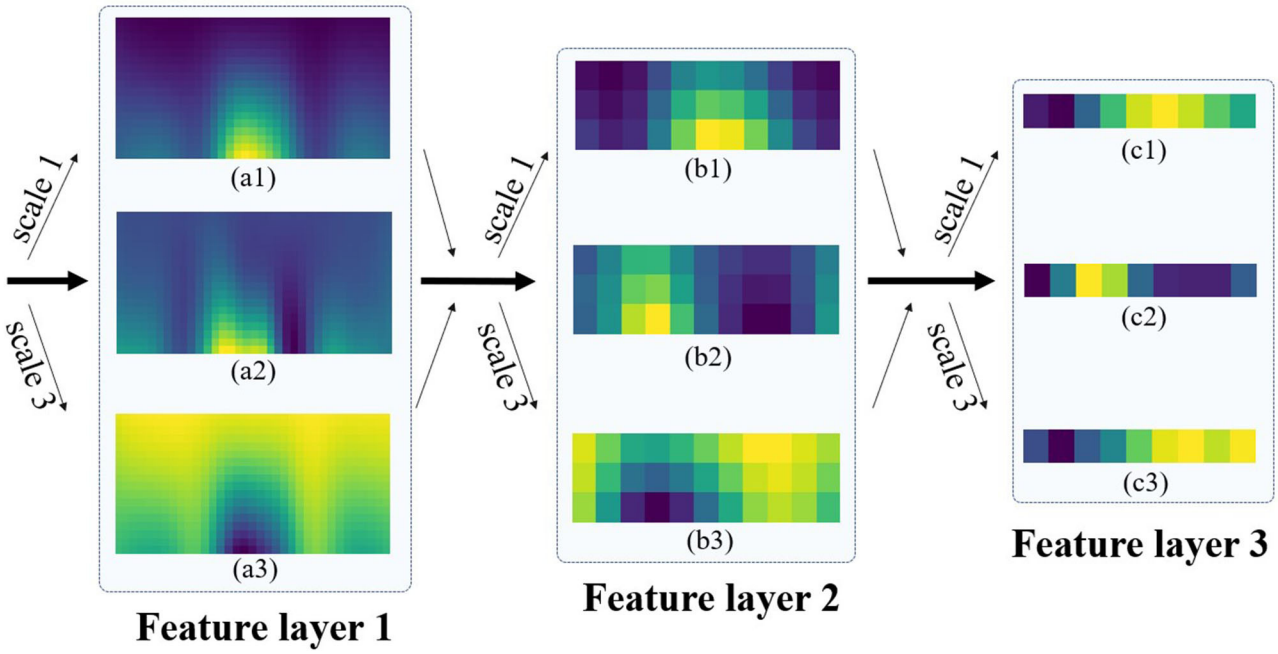



**FIGURE 10 |** Visualization of intermediate feature layer of multi-scale convolution structure.

and the large batch size, and GPU is not applied due to insufficient memory.

The results show that the training time of the machine learning method is second level, and the serial single-scale convolution network is minute level, whereas the parallel multi-scale convolution network needs hour-level training consumption. The traditional machine learning method and serial single-scale convolution network have the advantages of low parameter and obtain effective classification performance in a short time, but the feature completeness and detection accuracy still have great room for improvement. The high-precision advantage of parallel multi-scale networks needs to sacrifice the computing consumption, which is lengthy and time-consuming for ordinary equipment. In addition, the complex model requires massive EEG samples; otherwise, it may lead to non-convergence or overfitting of model training.

There are some suggestions for reducing computational complexity and time-consuming for complex model frameworks. Data migration is a common method for training model optimization. A small amount of new data was inputted in

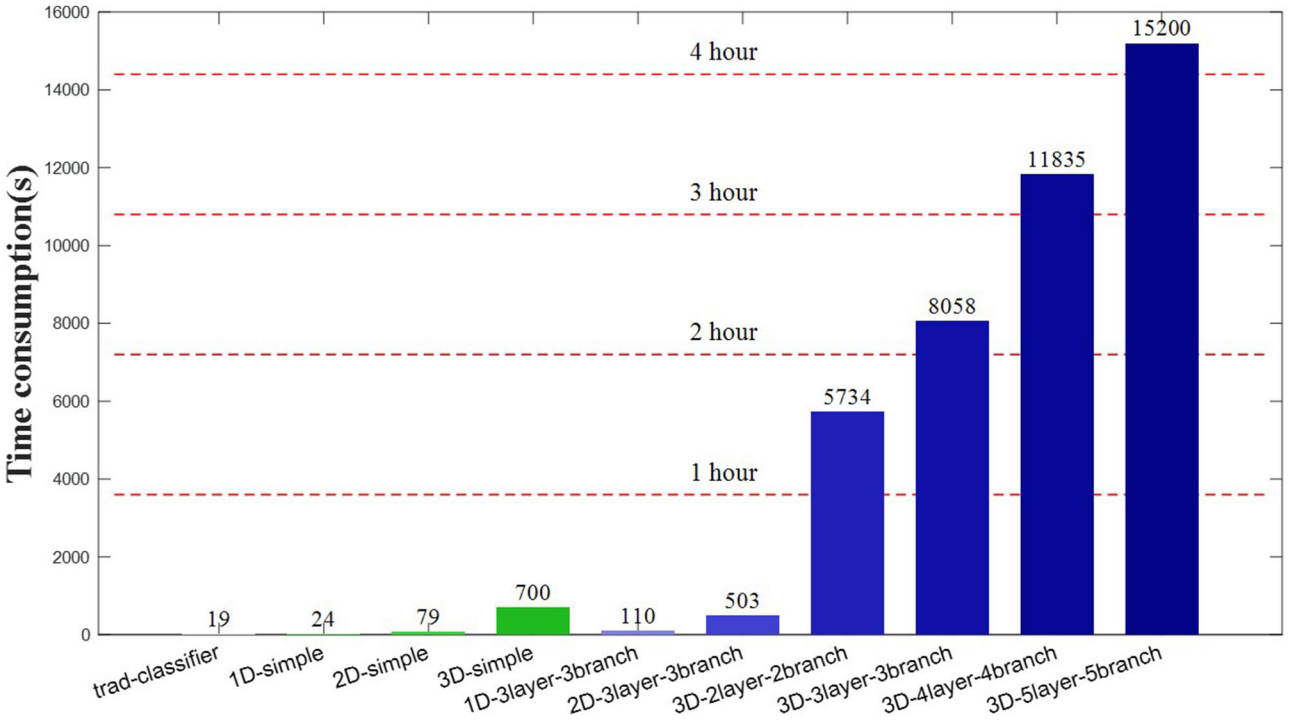



**FIGURE 11 |** Training time consumption of different EEG authentication strategies.

the completed complex model, which obtains the low fine-tuning cost and the stability of the original model. In addition, the initialized parallel multi-scale network training should select a low-dimensional convolution framework. **Figures 9**, **11** show that the multi-scale network has similar classification accuracy, and the low-dimensional framework gains higher cost performance. Moreover, the increase in the branch number leads to the decrease in training speed. The results in **Figure 9** show that the branch number should match the main component number of ERP, and the excess branch structure contributes less to the correct performance.

# DISCUSSION

## Convolution Kernel Size Analysis

The results of single-scale and multi-scale model framework show that the size and shape of convolution kernel affect the network model performance. **Figure 12** shows the kernel size corresponding to the actual EEG data. **Table 2** quantizes the time-frequency area covered by the convolution kernel. The odd number of convolution kernel size is selected as the side length. The single pixel scale of EEG time-frequency diagram represents 8.5 ms $\times$ 0.5 Hz.

The best three convolution dimensions of EEG authentication accuracy are $6 \times 3 \times 3$, $6 \times 5 \times 3$, and $6 \times 7 \times 3$, which represents the convolution vector length of spatial, temporal, and frequency domains, respectively. The performance of long convolution kernel (temporal domain priority) is prominent, indicating that the network allocates a larger area of time domain feature. The rectangular convolution kernel shape has more advantages than the square convolution kernel in the image field. This is related to the ERP characteristic of diversity temporal components and low-frequency energy. A variety of convolution kernel sizes are tried in the experiment, and the results show that the larger convolution area in the time domain can obtain good convergence and accuracy. We speculate that there is more important information in time domain, which can help the network to extract effective EEG features. Large convolution kernel area can obtain larger receptive field, which helps to fully integrate the feature information in the field of vision. In addition, the larger convolution area in time domain represents more convolution parameters, which provides more optimization space for time domain feature extraction.

Most ERP studies tend to EEG feature extraction in temporal domain, while few studies on the low-frequency energy of the oddball paradigm. This is consistent with the temporal domain orientation of the convolution kernel size in this paper, and **Figure 12** shows that ERP signal has the characteristics of high time-varying and low-frequency signals. This experiment is mainly aimed at the single-trial P300 EEG, and the performance of the long-during EEG convolution kernel still needs further research. It should be noted that the long kernel in this study corresponds to the high attention of time domain characteristics. If the input form is transformed, and the specific size should consider the actual corresponding time and frequency resolution.

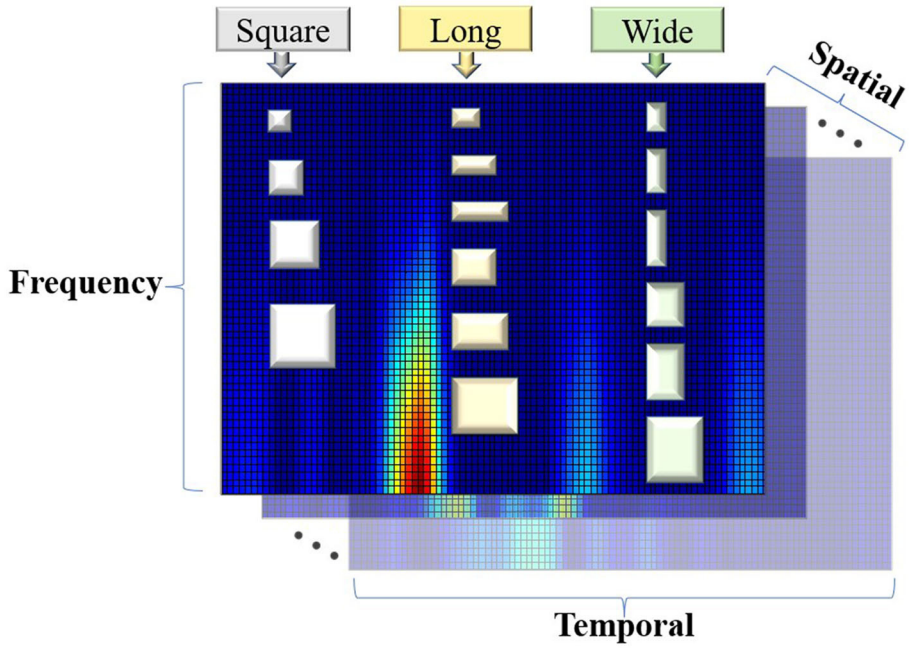



**FIGURE 12 |** The convolution kernel corresponds to the real size of EEG data.

**TABLE 2 |** Convolution kernel size corresponds to real EEG window size.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Square kernel | Kernel size | **3 × 3** | 5 × 5 | 7 × 7 | 9 × 9 | / | / |
| | Actual size | **25.5 ms × 1.5 Hz** | 42.5 ms × 2.5 Hz | 59.5 ms × 3.5 Hz | 76.5 ms × 5.5 Hz | / | / |
| Long kernel | Kernel size | **5×3** | **7×3** | 9×3 | 7×5 | 9×5 | 9×7 |
| | Actual size | **42.5 ms × 1.5 Hz** | **59.5 ms × 1.5 Hz** | 76.5 ms × 1.5 Hz | 59.5 ms × 2.5 Hz | 76.5 ms × 2.5 Hz | 76.5 ms × 3.5 Hz |
| Wide kernel | Kernel size | 3×5 | 3×7 | 3×9 | 5×7 | 5×9 | 7×9 |
| | Actual size | 25.5 ms × 2.5 Hz | 25.5 ms × 5.5 Hz | 25.5 ms × 5.5 Hz | 42.5 ms × 3.5 Hz | 42.5 ms × 5.5 Hz | 59.5 ms × 5.5 Hz |

*The bold value shows the best three convolution kernel shapes.*

For the input time-frequency diagram transformation, changing the window length and overlap length of the time-frequency diagram parameters result in unit-valued size updates. The advantage of long convolution kernel mentioned in this work is 8.5 ms × 0.5 Hz unit size, and the unit size in the new experiment needs multiple transformation on this conclusion. For the input EEG task form transformation, EEG presents different features in time-locked and long-term tasks such as resting state and emotional detection. The rich frequency features of long-term tasks may lead to convolution kernel transform shape to capture frequency information.

**Figure 6** shows that the smaller kernel size achieves better convergence. **Table 2** shows that the convolution kernels with excessively large time span completely span the important components of the ERP, making it difficult to extract the changing details of the component dynamics. In addition, the energy of ERP signal is mainly concentrated in the low-frequency below 5 Hz, and the increase in frequency domain direction size reduces the capture of low-frequency feature details. A large number of studies on deep network structure also show that the performance of small-size convolution kernel is better than that of large-size convolution kernel. The vision of multi-layer small-size convolution kernel can replace the large-size convolution kernel and obtain the advantages of small parameters and conducive to training.

## Work Limitation

In EEG applications, the proposed network framework is mainly for time-locked tasks, especially P300 capture of target stimulus. However, long-during EEG tasks such as resting state, emotional recognition, and fatigue monitoring were not discussed. The proposed framework can be modified appropriately in future research to be applied to long-during EEG tasks. Due to the different EEG components caused by diverse stimuli, the convolution kernel size and shape of long-during EEG may be different from the time-locked task, which requires extensive attempts of multi-scale strategies for actual tasks. In addition, time-locked EEG tasks analyze the meaning of network feature

layer by the P300 component localization, whereas the long-term stimulation usually lacks obvious observable components, which may lead to more difficulty in feature visualization and analysis.

In the future research of neural network systems, we will learn the advantages of excellent networks such as Vgg16, ResNet50, 101, and DarkNet53. Consistent with the strategy of stacking small convolution kernels in Vgg16 networks, this study also found excellent performance of small convolution kernels. Vgg16 focuses on the deep network of multilayers, and the proposed method applies the multi-branch breadth network. In the future network framework, the depth and breadth of the network structure can be combined to improve overall performance. ResNet applies the advantages of shortcut to alleviate the gradient disappearance problem and achieves outstanding performance in ultra-deep networks. The future network framework can apply the advantages of residual structure and insert shortcuts between multi-branch modules to achieve the deeper networks. It should be noted that insufficient EEG data may limit the performance of deep network, which should be considered in ultra-deep network training. The problem of insufficient EEG data can be solved by long-term collection or generation of artificial EEG samples. DarkNet53 is widely used in the image segmentation field, which applies the residual structure and convolution stride to replace the pooling layer. The proposed multi-scale 3D-CNN method can also simplify the network structure and reduce the time consumption by adjusting the convolution stride. The above methods inspire us with the network structure improvement in multiple perspectives. The proposed network should not only focus on multi-domain fusion and network breadth, but also improve the network depth and simplify the parameters.

## CONCLUSION

In this work, for the EEG identity authentication task, we proposed the feature domain fusion method of fusion temporal, frequency, and spatial domains. In addition, a corresponding 3D CNN framework with multi-scale model is constructed. The results show that sufficient feature fusion extracts individual EEG identity from multiple domains, which is conducive to the accurate identification of individual identity. In addition, the shape and size of convolution kernel are quantitatively analyzed according to the EEG signal characteristics. The research shows that the convergence and accuracy performance of $3 \times 3$ small-size kernel is outstanding in the time-frequency diagram, and the $3 \times 5$ and $3 \times 7$ rectangular-shape kernel with emphasis on time domain is more suitable for ERP tasks. In multi-scale architecture design, 3-layer 3-branch network matches the EEG components (P200, P300, LN) that balance the classification performance and time consumption. In summary, multi-scale convolution kernel parallelly extracts EEG identity features, which significantly improves 7% of accuracy in the identity authentication. The proposed novel network framework is suitable for EEG identity authentication, which can be transplanted and applied to various brain–computer interaction scenarios such as emotion recognition, fatigue driving, and neural rehabilitation.

## 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

RZ is mainly responsible for research design, data analysis, and manuscript writing of this study. YZ and KY are mainly responsible for research design. LT is mainly responsible for data collection and production of charts. JS is mainly responsible for contact subjects and participate in EEG data acquisition. RL is mainly responsible for experiment design and preparation. ZL is mainly responsible for experiment design and preparation, contact subjects, and participate in EEG data acquisition. BY is mainly responsible for data collection and manuscript modification. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported in part by the National Key Research and Development Plan of China under Grant 2017YFB1002502, in part by the National Natural Science Foundation of China under Grant 61701089, and in part by the Natural Science Foundation of Henan Province of China under Grant 162300410333.

## REFERENCES

Alariki, A. A., Ibrahimi, A. W., Wardak, M., and Wall, J. (2018). A review study of brian activity-based biometric authentication. *J. Comput. Sci.* 14, 173–181. doi: 10.3844/jcssp.2018.173.181

Arvind, R., Karthik, B., and Sriraam, N. (2012). Multi-feature characterization of epileptic activity for construction of an automated internet-based annotated classification. *J. Med. Syst*. 36:1155–1163. doi: 10.1007/s10916-010-9577-x

Cho, J., and Hwang, H. (2020). Spatio-temporal representation of an electoencephalogram for emotion recognition using a three-dimensional convolutional neural network. *Sensors* 20, 3491. doi: 10.3390/s20123491

Deng, X., Zhu, J., and Yang, S. (2021). *SFE-Net: EEG-based Emotion Recognition with Symmetrical Spatial Feature Extraction*. Available online at: http://ui.adsabs.harvard.edu/abs/arXiv:2104.06308 (accessed March 20, 2022). doi: 10.1145/3474085.3475403

Harshit, R. S., Thomas, K. P., Smitha, K. G., and Vinod, A. P. (2016). "Online Electroencephalogram (EEG) based biometric authentication using visual and audio stimuli," in *2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*. Available online at: http://ieeexplore.ieee.org/abstract/document/7843492/citations (accessed March 20, 2022). doi: 10.1109/IECBES.2016.7843492

Jayarathne, I., Cohen, M., and Amarakeerthi, S. (2016). "BrainID: development of an EEG-based biometric authentication system," in *IEEE IEMCON 2016*. Available online at: http://ieeexplore.ieee.org/document/7746325/references (accessed March 20, 2022). doi: 10.1109/IEMCON.2016.7746325

Kwak, Y., Kong, K., Song, W. J., Min, B. K., and Kim, S. E. (2020). Multilevel feature fusion with 3D convolutional neural network for EEG based workload estimation. *IEEE Access* 8, 16009–16021. doi: 10.1109/ACCESS.2020.2966834

Lee, D. Y., Jeong, J. H., Shim, K. H., and Kim, D. J. (2020). "Classification of upper limb movements ewline using convolutional neural network ewline with 3D inception block,"in *2020 8th International Winter Conference on Brain-Computer Interface (BCI)* (Gangwon: IEEE). doi: 10.1109/BCI48061.2020.9061671

Marcel, S., and Millan, J. (2007). Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 743–752. doi: 10.1109/TPAMI.2007.1012

Nakamura, T., Goverdovsky, V., Fellow, I. E. E. E., and Mandic, D. P. (2017). In-ear EEG biometrics for feasible and readily collectable real-world person authentication. *IEEE Trans. Inform. Forensics Sec.* 13, 648–661. doi: 10.1109/TIFS.2017.2763124

Ozcan, A. R., and Erturk, S. (2019). Seizure prediction in Scalp EEG using 3D convolutional neural networks with an image-based approach. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 2284–2293. doi: 10.1109/TNSRE.2019.2943707

Palaniappan, R., and Mandic, D. P. (2007). EEG based biometric framework for automatic identity verification. *J. Vlsi Signal Proc. Syst. Signal Image Video Technol.* 49, 243–250. doi: 10.1007/s11265-007-0078-1

Rathi, N., Singla, R., and Tiwari, S. (2020). Brain signatures perspective for high-security authentication. *Biomed. Eng. Applic. Basis Communic.* 32, 2050025. doi: 10.4015/S1016237220500258

Thomas, K. P., and Vinod, A. P. (2017). EEG-based biometric authentication using gamma band power during rest state. *Circuits Syst. Signal Process.* 37, 277–289. doi: 10.1007/s00034-017-0551-4

Yuan, Y., Xun, G., Jia, K., and Zhang, A. (2018). A multi-context learning approach for EEG epileptic seizure detection. *BMC Syst. Biol.* 12, 107. doi: 10.1186/s12918-018-0626-2

Zeng, Y., Wu, Q., Yang, K., Tong, L., Yan, B., Shu, J., et al. (2018). EEG-based identity authentication framework using face rapid serial visual presentation with optimized channels. *Sensors* . 19, 6. doi: 10.3390/s19010006

Zhang, R., By, A., Zeng, Y. A., Tong, L. A., Shu, J. A., and Bao, G. (2022). *EEG-Based Identity Authentication Using Subliminal Name Stimuli*. Available online at: http://www.sciencedirect.com/science/article/pii/S0167876021007327 (accessed March 20, 2022).

Zhang, Y., Chen, W., Lin, C. L., Pei, Z., and Chen, Z. (2021). Boosting-LDA algriothm with multi-domain feature fusion for motor imagery EEG decoding. *Biomed. Signal Process. Control* 70, 102983. doi: 10.1016/j.bspc.2021.102983

Zhang, Y., Zhang, X., Sun, H., Fan, Z., and Zhong, X. (2019). Portable brain-computer interface based on novel convolutional neural network. *Comput. Biol. Med.* 107, 248–256. doi: 10.1016/j.compbiomed.2019.02.023

Zhao, X., Zhang, H., Zhu, G., You, F., Kuang, S., and Sun, L. (2019a). A multi-branch 3D convolutional neural network for EEG-based motor imagery classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 2164–2177. doi: 10.1109/TNSRE.2019.2938295

Zhao, X., Zhang, H., Zhu, G., You, F., Kuang, S., and Sun, L. (2019b). A multi-branch 3D convolutional neural network for EEG-based motor imagery classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 2164–2177.

**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 Neuroinformatics

BRIEF RESEARCH REPORT
published: 16 June 2022
doi: 10.3389/fninf.2022.904301

# Classification of Electrophysiological Signatures With Explainable Artificial Intelligence: The Case of Alarm Detection in Flight Simulator

*Eva Massé, Olivier Bartheye and Ludovic Fabre**

*Centre de Recherche de l'Ecole de l'Air, Salon-de-Provence, France*

Relevant sounds such as alarms are sometimes involuntarily ignored, a phenomenon called inattentional deafness. This phenomenon occurs under specific conditions including high workload (i.e., multitasking) and/or cognitive fatigue. In the context of aviation, such an error can have drastic consequences on flight safety. This study uses an oddball paradigm in which participants had to detect rare sounds in an ecological context of simulated flight. Cognitive fatigue and cognitive load were manipulated to trigger inattentional deafness, and brain activity was recorded *via* electroencephalography (EEG). Our results showed that alarm omission and alarm detection can be classified based on time-frequency analysis of brain activity. We reached a maximum accuracy of 76.4% when the algorithm was trained on all participants and a maximum of 90.5%, on one participant, when the algorithm was trained individually. This method can benefit from explainable artificial intelligence to develop efficient and understandable passive brain–computer interfaces, improve flight safety by detecting such attentional failures in real time, and give appropriate feedback to pilots, according to our ambitious goal, providing them with reliable and rich human/machine interactions.

**Keywords: single-trial classification, pBCI, inattentional deafness, brain activity, ERP, explainable AI**

OPEN ACCESS

***Edited by:***
*Ke Liu,*
*Chongqing University of Posts and Telecommunications, China*

***Reviewed by:***
*Shenghong He,*
*University of Oxford, United Kingdom*
*Haixin Zhong,*
*Fudan University, China*

***Correspondence:**
*Ludovic Fabre*
*ludovic.fabre@ecole-air.fr*

***Received:*** *25 March 2022*
***Accepted:*** *22 April 2022*
***Published:*** *16 June 2022*

***Citation:***
*Massé E, Bartheye O and Fabre L (2022) Classification of Electrophysiological Signatures With Explainable Artificial Intelligence: The Case of Alarm Detection in Flight Simulator. Front. Neuroinform. 16:904301. doi: 10.3389/fninf.2022.904301*

## INTRODUCTION

Increased operational capabilities of aircraft had considerably modified the missions of pilots and introduce new problematics. For example, long periods of intense and sustained cognitive activities induce cognitive fatigue that is known to impair the performance of reasoned cognitive processing tasks over a period and also to be one of the major risks of incidents/accidents in aviation [e.g., Holtzer et al. (2010), Marcus and Rosekind (2017), Dehais et al. (2018), and Dönmez and Uslu (2018)]. In this study, we aimed at furthering our understanding of the influence of cognitive fatigue on alarm detection in order to develop passive brain–computer interfaces (pBCIs) based on explainable artificial intelligence (AI). To achieve these ends and following previous studies (Dehais et al., 2018, 2019), we asked participants to perform an alarm-detection task during

**Abbreviations:** pBCI, passive brain-computer interface; AI, artificial intelligence; ERP, event-related potential; IFBE, instruction flight before experiment; NFBE, no flight before experiment; VAS, visual analogous scale; LCL, low cognitive load; HCL, high cognitive load; SVC, support vector classification; KNN, K-nearest neighbors; LDA, linear discriminant analysis; RF, random forest.

repeated landing sessions on a flight simulator. To accentuate the presence of cognitive fatigue, we also manipulated the mental workload. We tested whether a real glider flight in instruction prior to the experiment influences performance in the alarm detection task on a flight simulator. We hypothesized that (a) cognitive fatigue impairs alarm detection as a function of the mental workload, (b) cognitive fatigue modulates electrophysiological activities, and (c) these modulations can be used as a predictor of reduced pilot's efficiency.

Previous studies have found that pilots' performance is influenced by cognitive fatigue [e.g., Dehais et al. (2018, 2019), Keller et al. (2019), Rocha and Silva (2019), Quental et al. (2021), and Rosa et al. (2021)]. Implementing pBCI or neuro-adaptive technology is a relevant approach to study cognitive fatigue and to improve flight safety (Zander et al., 2016; Arico et al., 2017; Dehais et al., 2018). For example, Dehais et al. (2018) asked participants to perform four identical traffic patterns along with a secondary auditory task (i.e., oddball paradigm) in simulated and real flight conditions. The oddball paradigm is used as an indirect index of cognitive fatigue and alarm detection and allows evaluating the P300 component as well as the main frequency bands associated with cognitive fatigue. They found that pilots more erred when reporting the number of auditory probes during the second part of the experiment than during the first part. In other words, participants' accuracy decreased with time on task. However, their small sample size did not allow them to statistically test the classification accuracies between the used features.

Empirically, previous findings showed that cognitive fatigue and mental workload have deleterious effects on stimulus-detection performance [e.g., Dehais et al. (2018, 2019)], whereas other findings showed an absence of a relationship between mental workload, cognitive fatigue, performance, and the occurrence of inattentional blindness [e.g., Bredemeier and Simons (2012), Beanland and Chan (2016), and Kreitz et al. (2016a,b)]. Unknown are the conditions under which cognitive fatigue or mental workload leads to poorer detection performance and their electrophysiological correlates. This is what we sought to know in this experiment.

The previously found attenuation of the P300 amplitudes reveals that inattentional deafness could result from an inability to automatically shift attention to the alarm that has been correctly detected or from an inability to process and recognize the warning (Giraudet et al., 2015b). However, we do not know whether event-related potentials (ERPs) and the time–frequency signal as a neural signature of inattentional deafness are good candidates as features to detect the occurrence of missed alarms.

The present experiment had two main goals. First, we investigated how alarm-detection changes associated with time on task interacted with other factors such as the cognitive workload or the type of previous activities (same task—flight instruction or different task—daily activities) and, *via* which mechanisms these factors influence performance. Second, we aimed at setting the scene to develop an EEG-based pBCI to detect alarm omissions to improve flight safety. Following previous studies on cognitive fatigue and alarm-detection tasks [e.g., Dehais et al. (2018)], participants had to perform an auditory task (i.e., oddball paradigm) during landing sessions. The mental workload was also manipulated to increase resulting cognitive fatigue. Based on previous findings that cognitive fatigue could impair performance by modulating attentional resources leaving fewer resources for tasks to perform [e.g., Chaudhuri and Behan (2004) and Holtzer et al. (2010)], two sets of hypotheses and predictions were tested in this study. The first hypothesis is that cognitive fatigue impairs alarm detection, resulting in increased alarm omissions in the fatigue group compared with the non-fatigue group and in the last landings compared with the first ones. The second hypothesis is that an efficient classification algorithm would be able to classify trials in which alarms were omitted and trials in which alarms were treated, based only on neurophysiological markers.

# MATERIALS AND METHODS

## Participants

Twenty-four male students of the Ecole de l'Air et de l'Espace (EAE) [mean age: 22.6 (2.0) years; flight experience: 75.6 (79.6) h, including 44.7 (58.9) h of glider experience; **Table 1**] were recruited. Participants were divided into two groups of 12 each based on their activity preceding the experiment: (1) Instruction Flight Before the Experiment (IFBE) group and (2) No Flight Before the Experiment (NFBE) group.[1] An informed consent was obtained from each participant prior to participation according to the Declaration of Helsinki.

## Subjective Scales

At the beginning and end of the experimental session, participants rated their subjective level of fatigue (VASf; Lee et al., 1991), sleepiness (Karolinska's Sleepiness Scale and VASs; Åkerstedt and Gillberg, 1990), and alertness (Samn-Perelli scale; Samn and Perelli, 1982).

## Tasks

### Experimental Task in a Flight Simulator

The flight simulator of the EAE, used for training young student pilots, was used to conduct the experiment based on previous studies using flight simulators [e.g., Durantin et al. (2017) and

[1]Participants in the NFBE group did not fly during the day of experiment but had daily activities such as classes or sports. They filled in a questionnaire with respect to these daily activities to inform experimenters whether they performed new and costful activities before the experiment, and this was not the case.

**TABLE 1** | Participants' characteristics.

| Characteristics | NFBE group | IFBE group | $F_{(1,22)}$ |
|---|---|---|---|
| *N* | 12 | 12 | - |
| Mean age, in years (SDs, range) | 23.2 (2.4, 21–27) | 22.1 (1.6, 20–26) | 1.75 |
| Mean flight experience (glider and plane), in hours (SDs, range) | 102.2 (97.3, 4–300) | 49.1 (47.5, 4.5–150) | 2.89 |
| Mean flight experience (glider), in hours (SDs, range) | 64.7 (59.3, 4–240) | 24.7 (41.0, 4–150) | 3.69 |

Dehais et al. (2014, 2016, 2019)]. It simulates an ASK21 glider using the X-plane 11 software allowing a 135° view of the screen. No participant reported experiencing motion sickness or dizziness, nor had their visual perception been disturbed during the simulated flight.

### Oddball Task

The auditory oddball task was coded and displayed using PsychoPy3 (Peirce, 2008). In this task, 100 pure tones, 1,000 or 1,100 Hz, at approximately 75 dB (20 dB above the ambient noise) were played, with 75% of standard sounds and 25% of target sounds. Participants had to respond to the auditory target (i.e., the alarm) by pressing a button on the joystick and ignore the frequent sounds. The frequency of the target sound was counterbalanced between participants. The intertrial interval was randomly set between 1.5 and 2.5 s to avoid anticipation and synchronization with brain rhythm (adapted from Dehais et al., 2019).

### Flight Scenario

Participants performed six successive runs, in optimal weather conditions. Each run consisted of a normal approach and landing on the grass runway of the BA701 in Salon-de-Provence and lasted approximately 3–5 min. Each run was divided into two conditions of cognitive load, namely, a low cognitive load (LCL) condition (alarm detection task during the downwind leg) and a high cognitive load (HCL) condition (alarm detection task and backward counting task during the base leg, the final, and the landing). In the backward-counting task (Sweller, 2011), they had to mentally count backward in threes from 100 (e.g., 100-97-94. . .) and pronounce the result at the end of the landing.

## Procedure

The experience took place at the end of the afternoon. First, participants completed subjective questionnaires. Second, participants were trained for 5 min to handle the simulator and for 5 min to perform the oddball task. The experimental session lasted for approximately 1 h 30 min. At the end of the experiment, participants completed again the subjective questionnaires.

## Electroencephalogram Recording

The EEG apparatus contained 32 passive electrodes (R-Net-helmet, LiveAmp-Brain Products), positioned following the 10/20 international system, recording at a 1,000 Hz sampling rate. The offline preprocessing was achieved using the MATLAB EEGlab package (Iversen and Makeig, 2014). Data were first bandpass filtered between 1 and 40 Hz, the signal was re-referenced on the average of all electrodes, and an independent component analysis was performed to reject eye and muscle artifacts using the RUNICA function of EEG lab. The signal was then segmented into 1,200 ms epochs, starting 200 ms before the stimuli. The ERPs were computed using a baseline correction with the first 200 ms of each epoch. ERP amplitude was considered as the averaged amplitude over the time period, in each trial and then averaged for each participant. P300 was considered between 400 and 650 ms, and N100 was considered between 100 and 200 ms after the stimulus onset.[2]

The time-frequency analysis was achieved using the Brain Vision Analyzer 2 software (Brain Products, version 2.2.0.7383). Data were resampled at 512 Hz, and the power spectral density was extracted for δ (1–4 Hz), θ (4–8 Hz), α (8–12 Hz), and β (12–30 Hz) and then decomposed in low-β (12–16 Hz), mid-β (16–20 Hz), and high-β (20–30 Hz) bands for each trial (i.e., each epoch of 1.2 s). We focused our analyses on the Fz, Cz, Pz, and Oz electrodes.

The first three runs were considered as the beginning of the session while the last three runs were considered as the end of the session, in the subsequent analyses.

## Analyses

Based on the previous study (Dehais et al., 2019), we focused our EEG analyses on three electrodes for ERPs and on four electrodes for spectral power, in order to cut computation time from the perspective of real-time analyses. All statistical analyses were carried out using JASP software (JASP Team, 2020). *Post-hoc* tests were carried out with the Bonferroni's correction for multiple comparisons, and a Greenhouse-Geisser correction was applied to respond to the sphericity condition when necessary.

# RESULTS

## Subjective Fatigue Evaluation

No difference was observed between the beginning and the end of the experimental task ($Fs < 1$, $ps > 0.5$) for the Visual Analogous Scale of Fatigue, the Samn-Perelli scale, and the Karolinska scale.

## Experimental Task

### Oddball Task

A 2 (group: NFBE and IFBE) × 2 (Time on task: beginning and end) × 2 (cognitive load: low and high) ANOVA with repeated measures and group as a between-subject factor was performed.

### Detection Rate

The detection rate was higher in the LCL condition than in the HCL condition (83.8 vs. 61.2%), $F_{(1,14)} = 102.92$, $p < 0.001$, $\eta_p^2 = 0.88$, and participants in the IFBE group detected more alarms than the NFBE group (79.8 vs. 63.2%), $F_{(1,14)} = 7.46$, $p = 0.016$, $\eta_p^2 = 0.35$ (**Figure 1**). No other effect was found.

Participants responded faster in the LCL condition than in the HCL condition (547 vs. 609 ms), $F_{(1,13)} = 22.66$, $p < 0.001$, and $\eta_p^2 = 0.64$. No other effect was found on reaction times.

### Electrophysiological Results

To compare electrophysiological signals between alarm detection and alarm omission, we focused our analyses on the HCL condition (participants missed more alarms in this condition). Data were analyzed with 2 (group: NFBE and IFBE) × 2 (time on task: beginning and end) × 3 (electrode: Fz, Cz, and Pz) × 2 (response: hit and miss) ANOVAs with repeated measures and group as the between-subject factor.

[2] After evaluation of a control group doing the oddball task only.

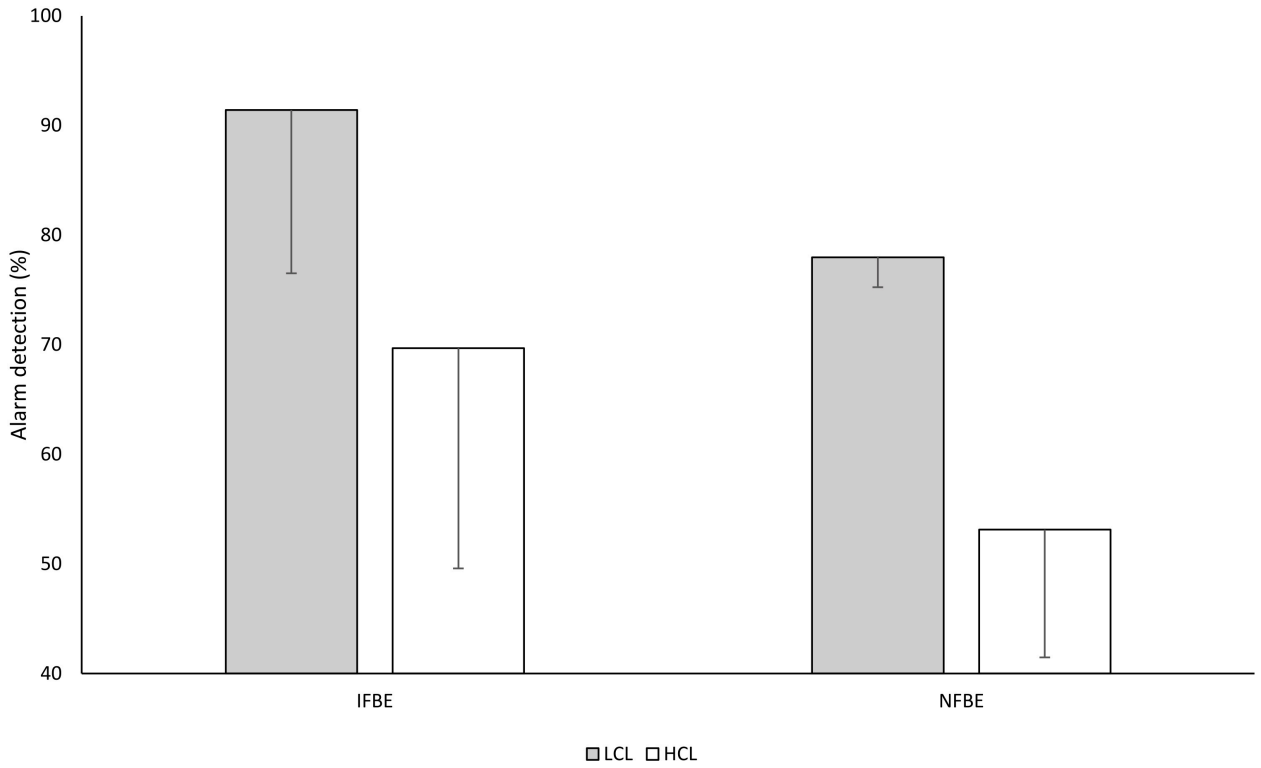



**FIGURE 1 |** Mean detection rate in the oddball task across cognitive load conditions for the two groups of pilots. LCL corresponds to the low cognitive load condition, and HCL to the high cognitive load condition. Error bars represent the standard deviation of the mean.

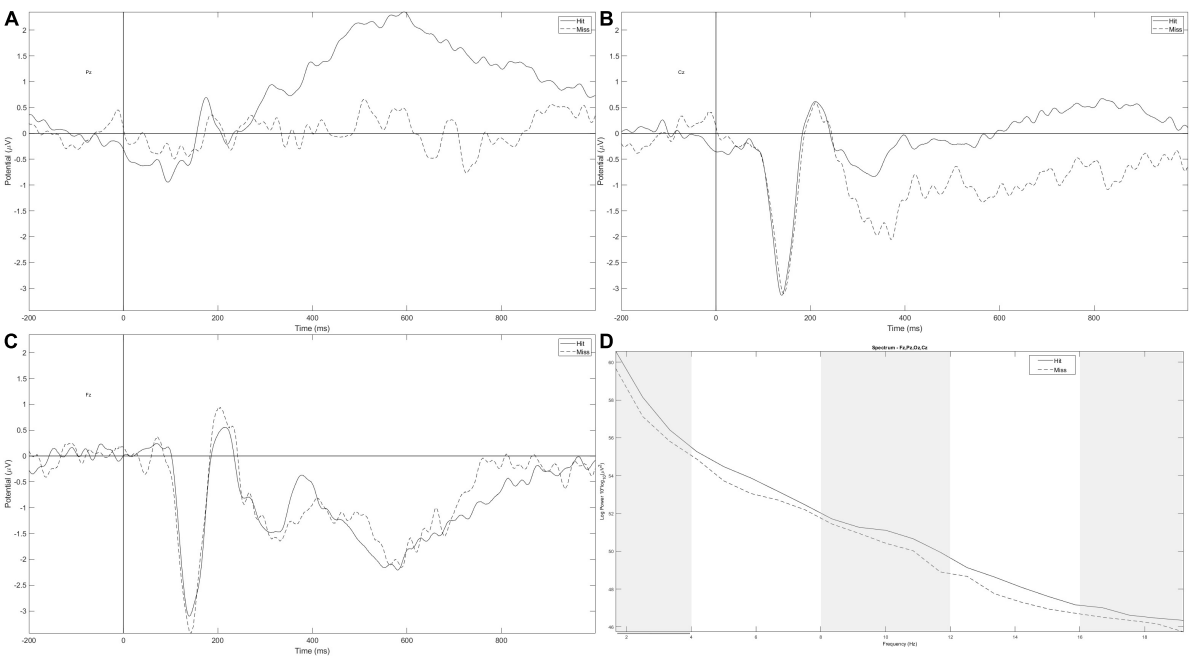



**FIGURE 2 |** Event-related potential (ERP) measured on Pz **(A)**, Cz **(B)**, and Fz **(C)** for hit (full line) and miss trials (dotted line). **(D)** Averaged spectral power on Pz, Cz, Fz, and Oz for hits (full line) and missed trials (dotted line). Gray parts correspond to frequency bands of interest (delta, alpha, and mid-beta) and gray lines correspond to a significant difference between hit and miss trials in all conditions.

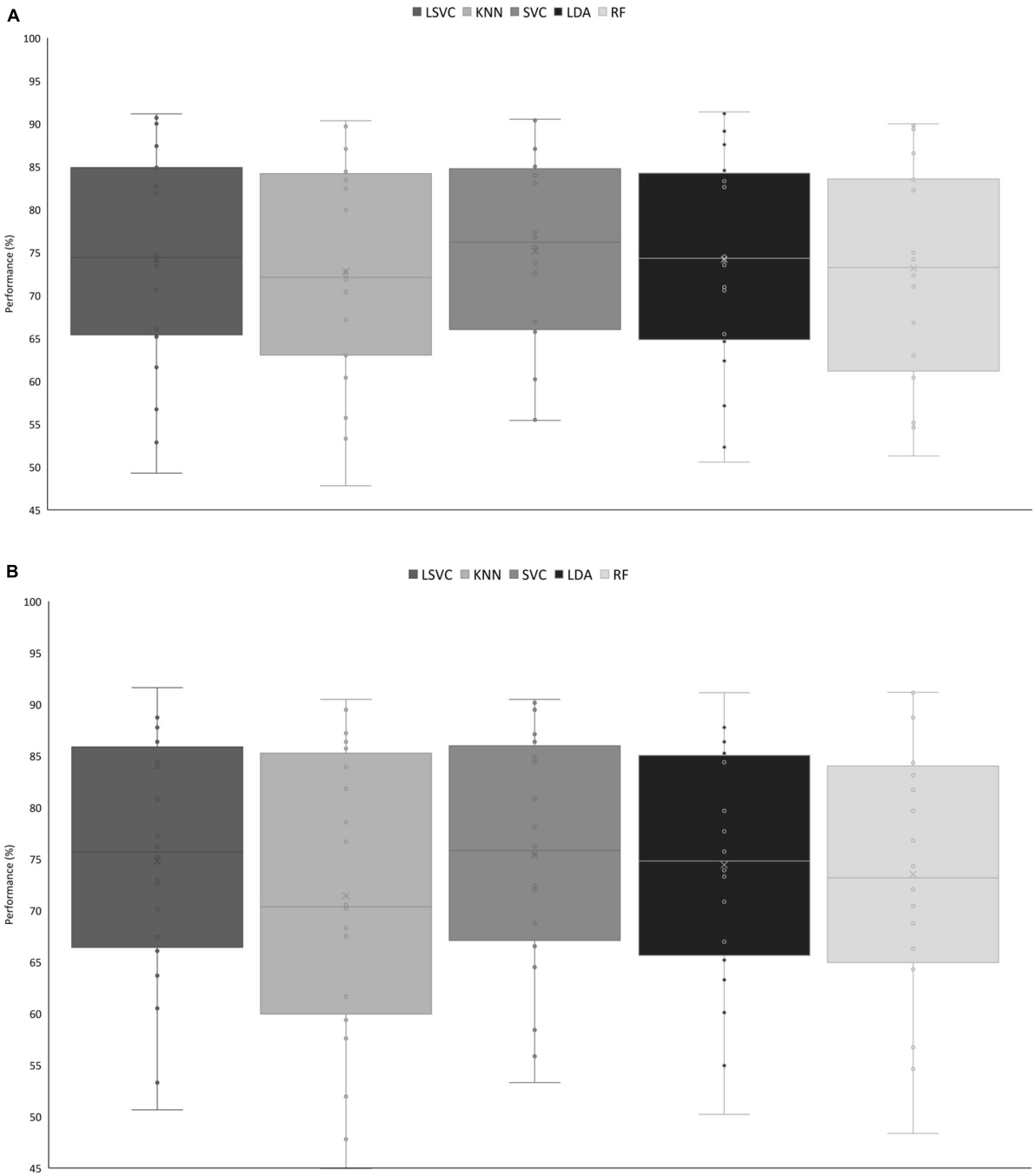



**FIGURE 3 | (A)** Mean accuracy on the training dataset across classifiers. KNN < LSVC, $t = 2.89$, $p = 0.050$; KNN < SVC, $t = -4.88$, $p < 0.001$; RF < SVC, $t = 4.12$, $p < 0.001$. **(B)** Performance of classifiers on the test set. LSVC < KNN, $t = 4.60$, $p < 0.001$; KNN < SVC, $t = -5.41$, $p < 0.001$; KNN < LDA, $t = -4.09$, $p < 0.001$. KNN, k-nearest neighbor; SVC, support vector classification; RF, random forest.

## Event-Related Potentials

The P300 amplitude varied across electrodes, $F_{(2,32)} = 13.45$, $p < 0.001$, and $\eta_p{}^2 = 0.46$. The amplitude was larger on Pz than on Cz and Fz, respectively, $t = -2.88$, $p = 0.02$ and $t = -5.17$, $p < 0.001$ (**Figures 2A–C**). Numerically, the P300 amplitude measured on Pz is reduced in miss trials compared with hit trials, but this difference did not reach significance (**Figure 2A**).

The N100 amplitude also varied across electrodes, $F_{(2,32)} = 8.57$, $p = 0.004$, and $\eta_p^2 = 0.35$, being larger on Fz and Cz compared with Pz, respectively, $t = -3.86$, $p = 0.001$ and $t = -3.198$, $p = 0.01$ (**Figures 2A–C**).

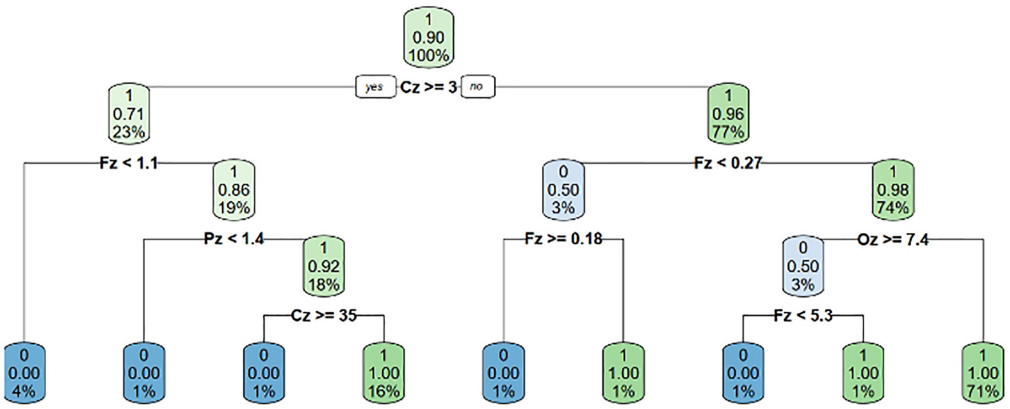



**FIGURE 4 |** Decision tree generated by the decision tree algorithm.

### δ, θ, α, and β Frequency Bands

The spectral power of the δ frequency band tended to be larger in hit trials (**Figure 2D**) compared with miss trials, $F_{(1,17)} = 3.16$, $p = 0.093$, and $\eta_p^2 = 0.16$. No other effect was found.

On the α frequency band, the significant effect of response (**Figure 2D**), $F_{(1,17)} = 5.28$, $p = 0.035$, and $\eta_p^2 = 0.24$, was qualified by the response × time on task interaction, $F_{(1,17)} = 5.28$, $p = 0.035$, and $\eta_p^2 = 0.24$. In the first three landings, the spectral power of the α frequency band was larger in hit trials compared with miss trials $t = 3.248$ and $p = 0.016$.

For the β frequency band, only the effect of the electrode was significant, $F_{(3,51)} = 4.28$, $p = 0.053$, and $\eta_p^2 = 0.20$, with a maximum on Oz compared with Fz and Cz, $t = -2.89$, $p = 0.034$ and $t = 3.15$, $p = 0.017$, respectively.

In the mid-β frequency band, *post-hoc* tests of the response × time on task × electrode × group interaction, $F_{(3,51)} = 3.36$, $p = 0.075$, and $\eta_p^2 = 0.17$, revealed that in the NFBE group, the spectral power was larger for hits than for miss trials at the beginning of the session, $t = 4.74$ and $p = 0.003$, and it was also larger in the beginning than at the end of the session, for hit trials, $t = 4.06$ and $p = 0.048$.

No effect was found on the θ frequency band.

## Single-Trial Classification

The classification pipeline was performed with the Scikit-Learn package of Python (Pedregosa et al., 2011). The first step of this process was to evaluate the performance of five classifiers [linear kernel, k-nearest neighbor (KNN), linear discriminant analysis (LDA), and random forest (RF) classifier] in participant-specific decoding of inattentional deafness, to distinguish trials in which the alarm was detected vs. trials in which alarms were omitted. Thus, classifiers were trained (80% of trials) and tested (20% of other trials) on individual pilots' electrophysiological data, and features were tested according to previous results. Accuracy values of the different algorithms were analyzed with a five [classifier: linear support vector classification (SVC), KNN, SVC, LDA, and RF] × 7 (features: δ, α, mid-β, δ and α, δ and mid-β, α and mid-β, α and δ, and mid-β) ANOVA.

The cross-validated scores obtained on the training set were first compared. The main effects of classifier, $F_{(4,76)} = 7.48$, $p < 0.001$, and $\eta_p^2 = 0.28$, and the interaction between classifier and features, $F_{(24,456)} = 2.84$, $p < 0.001$, and $\eta_p^2 = 0.13$, were significant. Across all features, the support vector machine (SVM) classifier reached the best performance of 75.2% on average (**Figure 3A**). For the SVM classifier, the most efficient configuration was the combination of the three frequency bands, with 75.9% of accuracy.

The inter-participant variability was quite high in the single-trial classification process, with accurate classification ranging from 47.1 to 90.5% across all configurations. However, generalization performance was then compared across configurations. In this analysis, the main effect of classifier was significant, $F_{(4,76)} = 8.92$, $p < 0.001$, and $\eta_p^2 = 0.32$. Across all features, the SVM classifier remained the most performant classifier on the testing dataset (**Figure 3B**).

The SVM algorithm aims at optimizing the classification accuracy and the distance between the boundary (which is a hyperplane) and each class. In fact, the algorithm is trained on the training dataset to minimize the expression of the form:

$$\left[\frac{1}{n}\sum_{i=1}^{n}\max(0,\ 1-y_i(w^T x_i - b))\right] C||w||^2$$

where $n$ is the number of data points, $w$ is the normal vector to the hyperplane, $b$ is the offset of the hyperplane from the origin, and $C$is the trade-off between correct classifications and distance separating the boundary hyperplane and each class.

For every classifier and feature, on average, the classifier performance exceeded the adjusted chance level of 61% based on Combrisson and Jerbi's recommendations (Combrisson and Jerbi, 2015) to consider the number of available trials. We reached

a maximum average performance of 76.4% (range: 57.7–90.5%) in participant-specific single-trial classification from the spectral power of δ and α frequency bands.

In a second step, data from all participants were taken altogether, and the different configurations were also tested for inter-participant classification. The main effect of classifier was significant, $F_{(4,16)} = 40.67$, $p < 0.001$, and $\eta_p^2 = 0.91$, showing that the KNN classifier is the least efficient classifier on the training set. We reached a maximum accuracy of 72.3% with the RF classifier and the combination of the three frequency bands.

## DISCUSSION

This study aimed to implement an EEG-based pBCI with explainable AI to monitor alarm detections under cognitive fatigue in aviation. Cognitive fatigue could be accentuated by the previous activities (i.e., IFBE or NFBE). Participants had to perform flying sessions with a secondary auditory alarm detection task under HCL or LCL. Our results replicate previous findings on inattentional deafness (Dehais et al., 2014; Giraudet et al., 2015a,b; Causse et al., 2016) showing that participants performed better to detect alarms under LCL conditions compared with HCL conditions. However, the difference between the P300 evoked by detected alarms and the P300 evoked by omitted alarms did not reach significance. Also, we did not find the expected effect of cognitive fatigue on alarm detection performance, potentially because our task was not sufficiently difficult to induce high cognitive fatigue in such a short time. By comparing alarm detection with respect to alarm omission, we found increased α, δ, and β (only at the beginning of the session and for the NFBE group) power. Based on these three frequency bands, we performed a single-trial classification of alarm detection or omission. The SVM reached a mean of 76.4%, which is considered sufficient for pBCIs. In fact, there is a need to detect these attentional failures in cockpits, and as our classifier overpassed the adjusted chance level (i.e., 61%), this study showed that frequency features, and more specifically d and a bands, implemented in an SVC formed an efficient tool to assess auditory alarm misperception in simulated flight conditions, with a classification process adapted to each individual pilot. However, real-time implementation of pBCI is still difficult to achieve due to the large preprocessing step that is needed before classification. The challenge in these analyses was to reduce computation time and noise related to other factors (e.g., muscle activities). Possibly, neural oscillations are also related to movement and so, the differences we found between hits and miss trials could reflect not just inattentional deafness *per se* but also a difference in behavior. The same results have already been observed in previous studies using the same protocol and interpreted as inattentional deafness (Somon et al., 2022). As our goal was to classify alarm detection vs. alarm omission, motion-related variation could be used as an effective detection marker and be a true single-trial classification tool.

Another promising direction we investigated is to exploit explainable results from classification and machine learning computations. The objective is two-fold: to enlarge the experimentation process by relaying the result of the classification with an appropriate sequence of actions as a virtuous loop and ultimately to design new doctrines based on reliable and rich human/machine interactions. Such an understandable information (numerical, symbolic, and logical) constitutes a ground cognitive support and justifies the interpretability criterion (Lundberg and Lee, 2017) providing a good level of confidence at the operational level. The initial step is to look for explainable classification methods. For instance, a decision tree delivers logical rules characterizing the criteria separating alarm omission and alarm detection. The idea is to detect abnormal behaviors by our apparatus, and from sense-making information, to apply safely decision-making later (Bartheye and Chaudron, 2019, 2020), for instance, to enable a sequence of actions to be engaged, whether these actions are automatic or not. As a use-case, one can mention the situation in a cockpit characterized by a loss of attention of the pilot and his/her inability to continue his/her current mission. That is, the operator did not consciously detect the alarm although his brain processed the signal. It is, therefore, necessary to inform the operator that he has omitted the alarm (by feedback) and to adapt the work environment with the explainable AI to help him in his task so that he comes back in the loop.

The interpretability criterion provides a good level of confidence at the operational level and leads to the choice of the best candidate machine learning model, which will not necessarily be the most efficient in terms of classification, but one which would enable a sequence of actions to be engaged at the end, whether automatic or not. This choice of machine learning methods agreeing with the interpretability criterion is strongly restricted and one can mention decision trees and to a lesser extent RFs but there are great expectations to be associated with.

To illustrate our discourse, we shall restrict the α frequency band full case study (2,855 individuals), and one can illustrate the principle on a single participant for sake of clarity (107 individuals) although the full case provides satisfactory results but obviously with more complicated formulas. The decision tree algorithm used is the Classification and Regression Trees (CART) algorithm (Breiman et al., 1984) and it provides the decision tree shown in **Figure 4**. The CART algorithm is a type of classification algorithm able to build a decision tree according to the Gini's impurity index. This index computes the degree of probability of a specific variable that is wrongly being classified when chosen randomly. It works on categorical variables and provides outcomes either be "successful" or "failure" and hence conducts binary splitting only. The R statistical language implementation is called RPART (Recursive Partitioning And Regression Trees) (Therneau, 1997) and is available in a package of the same name. The control is defined according to an integer value, the minimum number of observations that must exist in a node for which the routine will even try to compute a split (4 for 107 individuals and 40 for 2855 individuals).

Starting from a normalized form of these decision rules, we generated the appropriate code in a static context or in dynamic context. In a static context, the missed hit logical rules generated in the R statistical language are the following:

```
experiment[which(((experiment$Fz<1.050255
| (experiment$Pz<1.3769 |
experiment$Cz>=35.19693 & experiment$Pz>
=1.3769) &

experiment$Fz>=1.050255) & experiment$Cz>
=3.040795 |
(experiment$Fz>=0.17683 & experiment$Fz<
0.26921 | experiment$Fz<5.33816 &
experiment$Oz>=7.444865 & experiment$Fz>
=0.26921) &
experiment$Cz<3.040795)),]
```

which means: print out all the columns of the table `experiment` whose lines correspond to missing hits as the column `target` shows (`0` instead of `1`) and the execution of this expression gives the classification result by extracting the right lines.

| | Fz | Pz | Oz | Cz | target |
|---|---|---|---|---|---|
| 12 | 3.00544 | 1.55082 | 7.83941 | 1.64613 | 0 |
| 14 | 0.97795 | 1.66771 | 1.85181 | 4.23368 | 0 |
| 15 | 2.97765 | 0.87634 | 2.49094 | 3.12279 | 0 |
| 32 | 0.58338 | 2.46886 | 4.05271 | 3.06769 | 0 |
| 33 | 0.25340 | 1.28057 | 1.86640 | 1.40851 | 0 |
| 51 | 2.22273 | 3.67557 | 2.41245 | 61.71373 | 0 |
| 72 | 0.59377 | 0.98492 | 2.32070 | 4.86425 | 0 |

One can predict that way attention failure applying these rules regardless of the software involved (R, Python, Java, ...). One can write a computer program as a case-based analysis by executing a task once a condition identifying a missing hit situation is true. If this situation characterizes a loss of attention for the pilot and his/her inability to continue his/her current mission, the associated task corresponds to crisis management. In a dynamic context, one can reengineer completely these rules according to a simulation platform intertwining actuators and sensors to be more creative on human/machine interactions. To summarize, our contribution to that field is to post-process the measurement and the acquisition mechanisms to deliver understandable statements able to be translated into program statements contributing to the global loop in studying cognitive fatigue.

## DATA AVAILABILITY STATEMENT

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

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

EM and LF conceived and planned the experiments and contributed to the interpretation of the results. EM carried out the experiments. LF supervised the project. All authors wrote the manuscript and approved the submitted version.

## ACKNOWLEDGMENTS

We thank all the participants of l'Ecole de l'Air et de l'Espace. We also thank the reviewers for their work.

## REFERENCES

Åkerstedt, T., and Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *Int. J. Neurosci.* 52, 29–37. doi: 10.3109/00207459008994241

Arico, P., Borghini, G., Di Flumeri, G., Sciaraffa, N., Colosimo, A., and Babiloni, F. (2017). Passive BCI in operational environments: insights, recent advances, and future trends. *IEEE Trans. Biomed. Eng.* 64, 1431–1436. doi: 10.1109/TBME.2017.2694856

Bartheye, O., and Chaudron, L. (2019). "Algebraic modeling of the causal break and representation of the decision process in contextual structures," in *Computational Context the Value, Theory and Application of Context With AI*, 1st Edn. eds W. Lawless, R. Mittu, and D. Sofge (Boca Raton, FL: CRC Press), 229–252. doi: 10.1201/9780429453151

Bartheye, O., and Chaudron, L. (2020). "Human-machine sense making in context-based computational decision," in *Human-Machine Shared Contexts* eds O. Bartheye and L. Chaudron (Amsterdam: Elsevier). 385–398. doi: 10.1016/B978-0-12-820543-3.00019-5

Beanland, V., and Chan, E. H. C. (2016). The relationship between sustained inattentional blindness and working memory capacity. *Atten. Percept. Psychophys.* 78, 808–817. doi: 10.3758/s13414-015-1027-x

Bredemeier, K., and Simons, D. J. (2012). Working memory and inattentional blindness. *Psychon. Bull. Rev.* 19, 239–244. doi: 10.3758/s13423-011-0204-8

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees (Cart)*. Belmont, CA: Wadsworth.

Causse, M., Imbert, J.-P., Giraudet, L., Jouffrais, C., and Tremblay, S. (2016). The role of cognitive and perceptual loads in inattentional deafness. *Front. Hum. Neurosci.* 10:344. doi: 10.3389/fnhum.2016.00344

Chaudhuri, A., and Behan, P. O. (2004). Fatigue in neurological disorders. *Lancet* 363, 978–988. doi: 10.1016/S0140-6736(04)15794-2

Combrisson, E., and Jerbi, K. (2015). Exceeding chance level by chance: the caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *J. Neurosci. Methods* 250, 126–136. doi: 10.1016/j.jneumeth.2015.01.010

Dehais, F., Causse, M., Vachon, F., Régis, N., Menant, E., and Tremblay, S. (2014). Failure to detect critical auditory alerts in the cockpit: evidence for inattentional deafness. *Hum. Factors* 56, 631–644. doi: 10.1177/0018720813510735

Dehais, F., Duprès, A., Blum, S., Drougard, N., Scannella, S., Roy, R., et al. (2019). Monitoring Pilot's mental workload using erps and spectral power with a six-dry-electrode eeg system in real flight conditions. *Sensors* 19:1324. doi: 10.3390/s19061324

Dehais, F., Dupres, A., Di Flumeri, G., Verdiere, K., Borghini, G., Babiloni, F., et al. (2018). "Monitoring pilot's cognitive fatigue with engagement features in simulated and actual flight conditions using an hybrid fNIRS-EEG passive BCI," in *Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (Piscataway, NJ: IEEE). 544–549. doi: 10.1109/SMC.2018.00102

Dehais, F., Roy, R. N., Gateau, T., and Scannella, S. (2016). "Auditory alarm misperception in the cockpit: an EEG study of inattentional deafness," in *Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience*, Vol. 9743, eds D. D. Schmorrow and C. M. Fidopiastis (New York, NY: Springer International Publishing), 177–187. doi: 10.1007/978-3-319-39955-3_17

Dönmez, K., and Uslu, S. (2018). The relationship between flight operations and organizations in aircraft accidents; the application of the human factor analysis and classification system. *Anadolu Univ. J. Sci. Technol. A Appl. Sci. Eng.* 19, 316–333. doi: 10.18038/aubtda.348219

Durantin, G., Dehais, F., Gonthier, N., Terzibas, C., and Callan, D. E. (2017). Neural signature of inattentional deafness: neural signature of inattentional deafness. *Hum. Brain Mapp.* 38, 5440–5455. doi: 10.1002/hbm.23735

Giraudet, L., St-Louis, M.-E., Scannella, S., and Causse, M. (2015b). P300 Event-related potential as an indicator of inattentional deafness? *PLoS One* 10:e0118556. doi: 10.1371/journal.pone.0118556

Giraudet, L., Imbert, J.-P., Tremblay, S., and Causse, M. (2015a). High rate of inattentional deafness in simulated air traffic control tasks. *Procedia Manuf.* 3, 5169–5175. doi: 10.1016/j.promfg.2015.07.555

Holtzer, R., Shuman, M., Mahoney, J. R., Lipton, R., and Verghese, J. (2010). Cognitive fatigue defined in the context of attention networks. *Aging Neuropsychol. Cognit.* 18, 108–128. doi: 10.1080/13825585.2010.517826

Iversen, J. R., and Makeig, S. (2014). "MEG/EEG Data analysis using EEGLAB," in *Magnetoencephalography*, eds S. Supek and C. J. Aine (Berlin: Springer). 199–212. doi: 10.1007/978-3-642-33045-2_8

JASP Team (2020). *JASP (0.15) [Computer Software]*. Available online at: https://jasp-stats.org/

Keller, J., Mendonca, F. C., and Cutter, J. E. (2019). Collegiate aviation pilots: analyses of fatigue related decision-making scenarios. *Int. J. Aviat. Aeronautics Aerosp.* 6, 1–26. doi: 10.15394/ijaaa.2019.1360

Kreitz, C., Furley, P., Memmert, D., and Simons, D. J. (2016a). The Influence of attention set, working memory capacity, and expectations on inattentional blindness. *Perception* 45, 386–399. doi: 10.1177/0301006615614465

Kreitz, C., Furley, P., Simons, D. J., and Memmert, D. (2016b). Does working memory capacity predict cross-modally induced failures of awareness? *Conscious. Cognit.* 39, 18–27. doi: 10.1016/j.concog.2015.11.010

Lee, K. A., Hicks, G., and Nino-Murcia, G. (1991). Validity and reliability of a scale to assess fatigue. *Psychiatry Res.* 36, 291–298. doi: 10.1016/0165-1781(91)90027-M

Lundberg, S. M., and Lee, S.-I. (2017). "A Unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, Vol. 30, eds I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, et al. (Red Hook, NY: Curran Associates, Inc).

Marcus, J. H., and Rosekind, M. R. (2017). Fatigue in transportation: NTSB investigations and safety recommendations. *Inj. Prev.* 23, 232–238. doi: 10.1136/injuryprev-2015-041791

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.1080/13696998.2019.1666854

Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinformat.* 2:10. doi: 10.3389/neuro.11.010.2008

Quental, N., Rocha, J., Silva, J., Menezes, L., and Santos, J. (2021). The impact of cognitive fatigue on airline pilots performance. *J. Airline and Airport Manage.* 11:16. doi: 10.3926/jairm.183

Rocha, J. A. D. F, and Silva, J. M.D.R (2019). Pilots performance and floght safety: the case of cognitive fatigue in unpressurized aircraft cabins. *Internationla Congress of the Iberoamerican Air Transportation Reasearch Society*. Covilhã: University of Beira Interior. 1, 501–513.

Rosa, E., Gronkvist, M., Kolegard, R., Dahlstrom, N., Knez, I., Ljung, R., et al. (2021). Fatigue, emotion, and cognitive performance in simulated long-duration, single-piloted flight missions. *Aerosp. Med. Hum. Perform.* 92, 710–719. doi: 10.3357/AMHP.5798.2021

Samn, S., and Perelli, L. (1982). *Estimating Aircrew Fatigue: A Technique With Application to Airlift Operations. Brooks Air Force Base*. San Anonio (Tex.): USAF School of Aerospace Medicine. 29.

Somon, B., Giebeler, Y., Darmet, L., and Dehais, F. (2022). Benchmarking cEEGrid and solid gel-based electrodes to classify inattentional deafness in a flight simulator. *Front. Neuroergon.* 2:802486. doi: 10.3389/fnrgo.2021.802486

Sweller, J. (2011). "Cognitive load theory," in *Cognition in education*, Vol. 55, eds J. Mestre and B. H. Ross (Oxford: Academic Press). 37–76. doi: 10.1016/B978-0-12-387691-1.00002-8

Therneau, T. M. and Atkinson, E. J. (1997). *An Introduction to Recursive Partitioning Using the RPART Routines (Vol. 61*, p. 452). Mayo Foundation: Technical report.

Zander, T. O., Krol, L. R., Birbaumer, N. P., and Gramann, K. (2016). Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity. *Proc. Natl. Acad. Sci. U.S.A* 113, 14898–14903. doi: 10.1073/pnas.1605155114

**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.

frontiers | Frontiers in Neurorobotics

TYPE Hypothesis and Theory
PUBLISHED 03 August 2022
DOI 10.3389/fnbot.2022.958052



OPEN ACCESS

EDITED BY
Duo Chen,
Nanjing University of Chinese Medicine, China

REVIEWED BY
Minpeng Xu,
Tianjin University, China
Peng Zhang,
Huazhong University of Science and Technology, China

*CORRESPONDENCE
Zhuliang Yu
zlyu@scut.edu.cn
Jian Yang
yangjianxin@scut.edu.cn

RECEIVED 31 May 2022
ACCEPTED 04 July 2022
PUBLISHED 03 August 2022

CITATION
Chen L, Yu Z and Yang J (2022) SPD-CNN: A plain CNN-based model using the symmetric positive definite matrices for cross-subject EEG classification with meta-transfer-learning. *Front. Neurorobot.* 16:958052. doi: 10.3389/fnbot.2022.958052



# SPD-CNN: A plain CNN-based model using the symmetric positive definite matrices for cross-subject EEG classification with meta-transfer-learning

Lezhi Chen[1], Zhuliang Yu[1,2]* and Jian Yang[1]*

[1]College of Automation Science and Engineering, South China University of Technology, Guangzhou, China, [2]Pazhou Laboratory, Guangzhou, China

The electroencephalography (EEG) signals are easily contaminated by various artifacts and noise, which induces a domain shift in each subject and significant pattern variability among different subjects. Therefore, it hinders the improvement of EEG classification accuracy in the cross-subject learning scenario. Convolutional neural networks (CNNs) have been extensively applied to EEG-based Brain-Computer Interfaces (BCIs) by virtue of the capability of performing automatic feature extraction and classification. However, they have been mainly applied to the within-subject classification which would consume lots of time for training and calibration. Thus, it limits the further applications of CNNs in BCIs. In order to build a robust classification algorithm for a calibration-less BCI system, we propose an end-to-end model that transforms the EEG signals into symmetric positive definite (SPD) matrices and captures the features of SPD matrices by using a CNN. To avoid the time-consuming calibration and ensure the application of the proposed model, we use the meta-transfer-learning (MTL) method to learn the essential features from different subjects. We validate our model by making extensive experiments on three public motor-imagery datasets. The experimental results demonstrate the effectiveness of our proposed method in the cross-subject learning scenario.

KEYWORDS

**EEG, Motor imagery, SPD matrices, CNN, Meta-transfer-learning**

## 1. Introduction

An EEG-based Brain-Computer Interface (BCI) is a system to measure and analyze the electroencephalography (EEG) brain signal (Rao, 2013), thus enabling the communication or interaction between the brain and external environment (Kothe and Makeig, 2013). Recent research has opened up the possibility for EEG signals to apply in rehabilitation (Tariq et al., 2018), entertainment (Nijholt et al., 2008), and transportation (Göhring et al., 2013) because of the harmless, non-invasive, and inexpensive features of the EEG-BCI. Motor imagery (MI), which refers to the mental simulation of body

movements, is a famous paradigm of the EEG-BCI system (Lotze and Halsband, 2006). MI signals are widely used in the BCI system (Alamgir et al., 2010; Arvaneh et al., 2013; Jayaram et al., 2016) because of their flexibility in reflecting the bioelectrical activity of the brain. These signals attract increasing attention in rehabilitation therapy (Naseer and Hong, 2013, 2015; Hong et al., 2015).

However, due to the separation between the signal source (inside the human brain) and the detector, the EEG signals would be easily contaminated by various artifacts and noise, including muscle movements, eye blinks, heartbeats, and environmental electro-magnetic field in the applications of the BCI-system. This phenomenon induces a domain shift in each subject, even in different sessions of the same subject (Reuderink et al., 2011), and exhibits significant pattern variability between different subjects. Consequently, it hinders people from using the data generated from different subjects to improve the performance of the BCI system (Lotte and Guan, 2010) and increasingly reduces the accuracy and stability of EEG cross-subject classification. Currently, the users of the BCI-system have to provide tons of EEG-data to build a user-specific classifier so that the system can work properly. Accordingly, it greatly lengthens the time of calibration of the BCI system and heavily inhibits BCI-system development.

To overcome this problem, lots of methods are proposed to eliminate the shifting problem of data distribution among different subjects. Rodrigues et al. (2018) present a transfer Learning approach to match the statistical distributions of different sessions/subjects. This method allows the BCI systems to reuse the data from different users and reduce the calibration time. He and Wu (2019) propose a Euclidean Space Data Alignment Approach to align the time-domain EEG trials in the Euclidean Space and alleviate the domain shift between different sessions and subjects successfully. However, this kind of **data-augmentation** method normally classifies the data by the traditional geometry-aware classifiers (such as support vector machine and the minimum distance to mean classifier) (Barachant et al., 2013), which are insufficient for feature extraction. Also, it requires people to use certain prior knowledge of brain science.

With the development of machine learning, deep learning technology has been applied to extract discriminative features from EEG (Lotte et al., 2018) and many **model-based** learning algorithms have been proposed for MI-EEG cross-session/subject classification (Wu et al., 2020). Schirrmeister et al. (2017) focus on the application of different CNN architectures in EEG-MI classification and design an efficient network architecture to decode information from the EEG-MI signal. This method shows the powerful feature extraction ability of CNN and draws a great deal of attention to the applications of CNN in the BCI system. Lawhern et al. (2018) propose a brand-new compact CNN-based model called EEGNet, which contains depth-wise and separable convolutions to extract the descriptive information from EEG signal directly. This network structure is robust enough to learn a wide variety of interpretable features over a range of BCI tasks in cross-session/subject learning and gain outstanding classification performance. Fahimi et al. (2019) propose an inter-subject transfer learning framework built on top of the CNN model which is fed into three different EEG representations and transfers knowledge between different subjects thus avoiding time-consuming re-training. However, this kind of network focus on the feature extraction of EEG signal and their performances would deteriorate when the data of the user are insufficient, especially in the few-shot scenario of cross-subject learning.

In the most recent studies, meta-learning, which is a task-level learning method, has seen substantial advancements in computer vision and speech recognition recently (Vanschoren, 2018). This kind of learning method helps the neural network to extract usable features from related tasks and largely increases the generalization ability of the neural network. Li et al. (2021) use the training method called Model-Agnostic Meta-Learning (MAML) (Finn et al., 2017) and build the CNN-based classifier which combines one and two dimensional-CNN layers to improve the accuracy of the MI-EEG classification. However, these kinds of meta-learning structure are very sensitive to neural network architectures (usually shallow neural networks), which often leads to instability during training and easily induces overfitting problems. Therefore, it limits the effectiveness of meta-learning.

In consequence, given the above, an effective model that is capable of capturing essential features and a robust meta-learning method are both essential to cross-subject learning in EEG classification. The symmetric positive definite (SPD) matrices have been widely used in motor imagery EEG-based classification over the past few decades (Barachant et al., 2013; Xu et al., 2021), because of their capacity to capture informative structure from the data (Huang and Van Gool, 2017). In terms of the ability to capture input data structure, the CNN has the powerful capability of extracting features of two-dimensional matrix-shape data (LeCun et al., 1998; Krizhevsky et al., 2012) and the SPD matrices are one of the two-dimensional matrix-shape data. Therefore, Hajinoroozi et al. (2017) combine the SPD matrices of EEG data and the deep learning method and present a series of deep covariance learning models for drivers' fatigue prediction, which explore the potential of this kind of method for the application of BCI system. Inspired by this, we propose a plain CNN-based model called SPD-CNN, which transforms the EEG signal into the SPD matrices and uses a CNN with five convolutional layers to capture the features of SPD matrices. Also, we apply a cutting edge meta-updating strategy called the meta-transfer-learning (MTL) (Sun et al., 2019) which combines the advantage of transfer learning and meta-learning to extract the subject invariant features and alleviates the shifting problem

TABLE 1 Key information of the three MI-EEG datasets used for experiences.

| Dataset | Number of subjects | Trails per subject | Class | Band pass filter (Hz) | Number of electrodes channnels |
|---|---|---|---|---|---|
| BNCI2014001 | 9 | 576 | 4 | 4–250 | 22 |
| BNCI2015004 | 9 | 400 | 5 | 0.5–100 | 30 |
| Sch2017 | 14 | 1,000 | 4 | 4–250 | 128 |

TABLE 2 Table of symbols used in the article.

| Symbols | Meaning |
|---|---|
| $F(\Theta, \theta)$ | The classification network with parameter $\Theta$ and $\theta$ |
| $\Theta, \theta$ | The parameter of the feature extractor block and the classifier block |
| $\Theta^{pre}, \Theta^{meta}$ | The parameter after the pre-train phase and the meta-adaption phase |
| $f(\Theta, \theta)$ | The network after the $F(\Theta, \theta)$ is upgraded by specific task |
| $X$ | The multiple time-series of a EEG matrix |
| $C$ | The covariance matrix estimated by $X$ |
| $D_{tr}, D_{val}, D_{te}$ | Dataset $D$ for Training,Validation and Testing phase |
| $T_i$ | The specific task which is sample from the $D$ |
| $l_i$ | The loss function in task i during the inner-loop |
| $L_\phi$ | The meta loss function in meta training |

between the source domain (training subjects) and the target domain (test subjects). The major contributions of this article can be summarized as follows.

- The SPD-CNN model we proposed uses the SPD matrices of the EEG signal as descriptors to highlight the spatial information of the EEG-MI signal and reduces the diversity of EEG data characteristics of different subjects. Additionally, the proposed descriptor tremendously decreases the size of data and effectively reduces the difficulty of feature extraction.
- Using the MTL as our learning strategy helps the network extract the crucial features. In other words, it can transfer the domain knowledge between different subjects during the training process and enhance the robustness of the network in the BCI system.
- To the best of our knowledge, the network we proposed is simple to design and has fewer parameters than most networks for EEG classification currently. Therefore, it could simplify the training process tremendously and shortens the training time extremely.

The remainder of the article is organized as follows. Section 2 presents the framework of the proposed approach. Section 3 describes the experimental settings, then shows the results, and provides a comprehensive analysis. The effectiveness of the proposed SPD descriptor is discussed in Section 4 and the conclusion is summarized in Section 5.

## 2. Materials and methodology

### 2.1. Data description

We present examples with three public EEG-MI datasets which are BNCI2014001 (Tangermann et al., 2012), BNCI2015004 (Scherer et al., 2015), and Sch2017 (Schirrmeister et al., 2017).

BNCI2014001 consists of the EEG data from 9 subjects and this MI-paradigm consists of four different motor imagery tasks that the subjects are required to make the imagination of movement of the left hand, right hand, both feet and tongue. The EEG Signals are recorded with 22 electrodes at a 250 Hz sampling rate and two sessions were recorded for each subject. Each session is composed of 6 runs separated by short breaks. One run consists of 48 trials (12 for each of the four possible classes), yielding a total of 288 trials per session.

BNCI2015004 is a 30-electrode dataset obtained from 14 subjects with disability (spinal cord injury and stroke). The dataset consists of five classes of imagined movements of right-hand and feet, mental word association, mental subtraction, and spatial navigation. The EEG signals are recorded at a 250 Hz sampling rate, and two sessions were recorded for each subject. Each session consists of 8 runs, resulting in 40 trials of each class. The EEG signals were bandpass filtered 0.5–100 Hz and sampled at a rate of 256 Hz.

Sch2017 is a 128-electrode dataset obtained from 14 healthy subjects [6 women, 2 left-handed, age 27.2 $\pm$ 3.6 (mean $\pm$ std)] and this MI-paradigm consists of four different motor imagery tasks which ask subjects to make the imagination of movement of the left hand, right hand, both feet, and rest (no movement), with roughly 1,000 four-second trials of executed movements divided into 13 runs (each run consist of the approximately 1,000 trails per subject.

Three datasets mentioned above are publicly available on the "Mother of all BCI Benchmarks"(MOABB) framework (Jayaram and Barachant, 2018).In the experiment section, the subjects in the same dataset will be divided into training subjects, validation subjects, and test subjects who provide data for the training

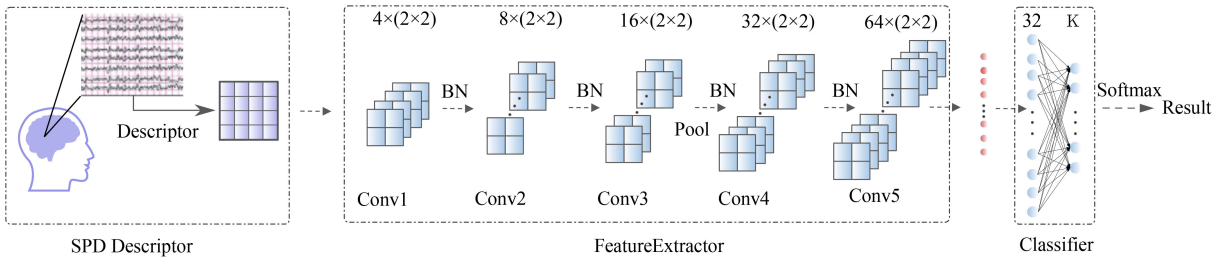



FIGURE 1
Overall visualization of the SPD-CNN architecture. It starts with an SPD descriptor to transform EEG into SPD matrices, then the matrices are encoded by Feature Extractor Block and flattened as the input data of the classification block. In the classification block, the features are passed to a two-fully connected layer and put into a soft-max classification with K units, K is the number of classes in the data.

set, validation set, and test set for the cross-subject learning experiments, respectively. More details can be seen in Table 1.

## 2.2. SPD-CNN model

Table 2 gives a brief description of the mathematical symbols that will be used in the rest of the article.

As mentioned above, we are particularly interested in the case where the SPD matrices are spatial covariance matrices, which describe the second-order statistics of zero-mean multivariate time series. We assume that the information on the power and spatial distribution of EEG sources can be coded by a covariance matrix. Therefore, the spatial covariance matrix $C$ of a T-sample realization of a zero-mean $d$-dimensional time series ($d$ being the number of electrode channels) $X \in R^{d\times T}$, is estimated as

$$C_i = \frac{1}{T} X_i^T X_i, i = 1, 2, \ldots, n \tag{1}$$

where $X_i$ is the sample from the EEG dataset $D = \{X_i, i = 1, 2, \cdots, n\}$ and n is the total amount of samples in dataset $D$.

Based on the analysis above, we develop a covariance matrix estimator called the SPD descriptor that captures not only the diversity among different electrode channels but also the statistical properties of EEG image regions. The descriptor is capable of estimating the $d \times d$ covariance matrix of the EEG features mentioned in Equation (1). Then, these matrices are normalized with the whole sample set mentioned in Equation (2) to improve the numerical stability of the model. Consequently, the network is able to focus on the critical features and accelerate the learning process (Shanker et al., 1996).

$$C_i^* = (C_i - C_{mean}) \Big/ C_{std}, i = 1, 2, \cdots, n \tag{2}$$

where $C_{mean}$, $C_{std}$ is the mean and SD of covariance matrix set $C = \{C_i, i = 1, 2, \cdots, n\}$ and $C_i^*$ is the output sample of the descriptor.

TABLE 3 Basic parameter of SPD-CNN model.

| Block | Layers | Size and Kernel | Activation |
|---|---|---|---|
| | *Conv*1 | $4 \times (2 \times 2)$ | Relu |
| | *Conv*2 | $8 \times (2 \times 2)$ | Relu |
| | *Conv*3 | $16 \times (2 \times 2)$ | Relu |
| Feature extractor | $Max - pool$ | $(2 \times 2)$ | - |
| | *Conv*4 | $32 \times (2 \times 2)$ | Relu |
| | *Conv*5 | $64 \times (2 \times 2)$ | Relu |
| Clasiifier | *FC* | $2 \times 32$ | softmax |

In the Feature Extractor block,*Conv* means a convolutional layer and *FC* represents a two-layer fully-connected network, with 32 neurons inside the hidden layer as shown in Figure 1.

After being processed by the SPD descriptor, the $d \times d$ matrices are taken into a Feature Extractor block. This block contains five convolutional layers ($Conv1, Conv2, Conv3, Conv4, Conv5$) with minimum convolution kernel ($2 \times 2$). Then the output data from Feature Extractor were flattened and taken through a classifier block with a two-layer fully-connected network (*FC*) onto the BCI outputs. A whole visualization and full description of the SPD-CNN model can be found in Figure 1 and Table 3.

## 2.3. Training structure and learning strategy

Our training structure is to help the model extract the key features through learning a better initial set of parameters from various tasks of different subjects. Hence, the network gains a fast adaption to new user tasks using only a few data. This learning strategy is based on the assumption that the EEG data from different subjects share the same representative features. These features are just masked by the effect of individual variation and wide discrepancy in the experiment

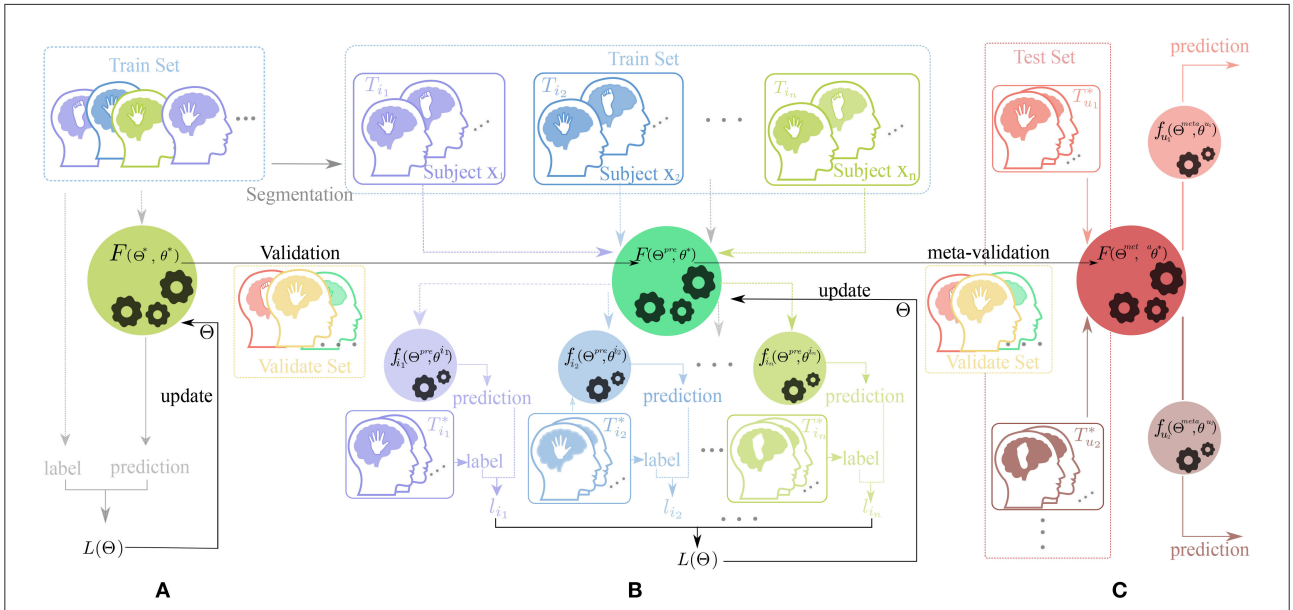



FIGURE 2
Workflow of our training framework. The dataset for training, validation, and test process is displayed on different rectangular regions with the colors blue, yellow, and red, respectfully. The picture of human heads in different colors (such as purple, blue, yellow, brown, and red) with a hand or feet inside represent the data from different subjects doing motor-imagery tasks. The black horizontal lines with a black arrow represent the change of the parameter of the neural network and the colorful vertical and horizontal lines (such as purple, blue, gray, and green) indicate the direction of data flow. In addition, the black gears in the circle represent the update process of parameters. **(A)** Pre-train phase, **(B)** Meta-update phase, and **(C)** Domain-adaption phase.

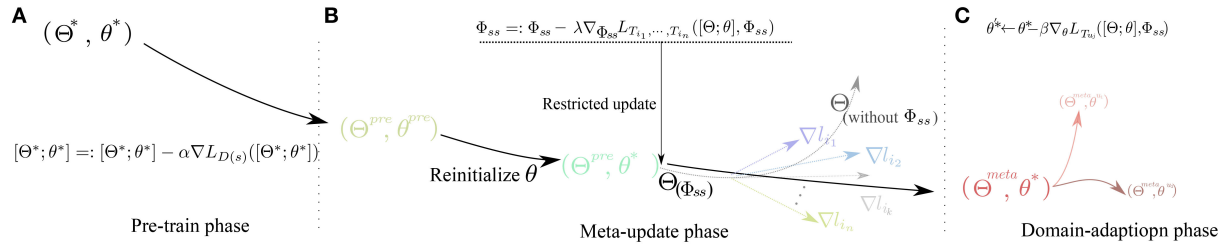



FIGURE 3
Diagram of parameters variation through the learning process in different phases. **(A)** Pre-train phase, **(B)** Meta-update phase, and **(C)** Domain-adaption phase.

environment. In this section, we illustrate the main idea of MTL and describe its application in the EEG cross-subject learning scenario.

The MTL combines the advantage of transfer learning and meta-learning structure. This training method uses the fine-tune skill and model-agnostic meta-learning (MAML) algorithm (Finn et al., 2017) with a novel constrained setting on network parameters called scaling and shifting (SS) operation to solve the overfitting problem. Hence, our training framework consists of three parts: Pre-train, Meta-updating, and Fast adaption. The whole workflow in this framework is shown in Figure 2.

As shown in Figures 2, 3, in the Pre-train phase, data of training subjects are merged randomly into a training dataset $D_{tr}$ for classifier $F(\Theta^*, \theta^*)$. The network $F$ with initialized parameter $(\Theta^*, \theta^*)$ is optimized by the traditional gradient descent method (refer to Equation 3) and gains the better initialized parameter$(\Theta^{pre}, \theta^{pre})$.

$$[\Theta^*; \theta^*] =: [\Theta^*; \theta^*] - \alpha \nabla L_{D_{tr}}([\Theta^*; \theta^*]) \tag{3}$$

where $\alpha$ is the learning rate of and $L_{D_{tr}}$ denotes the most frequently used empirical loss in machine learning like cross-entropy (Zhang and Sabuncu, 2018). This process neglects the

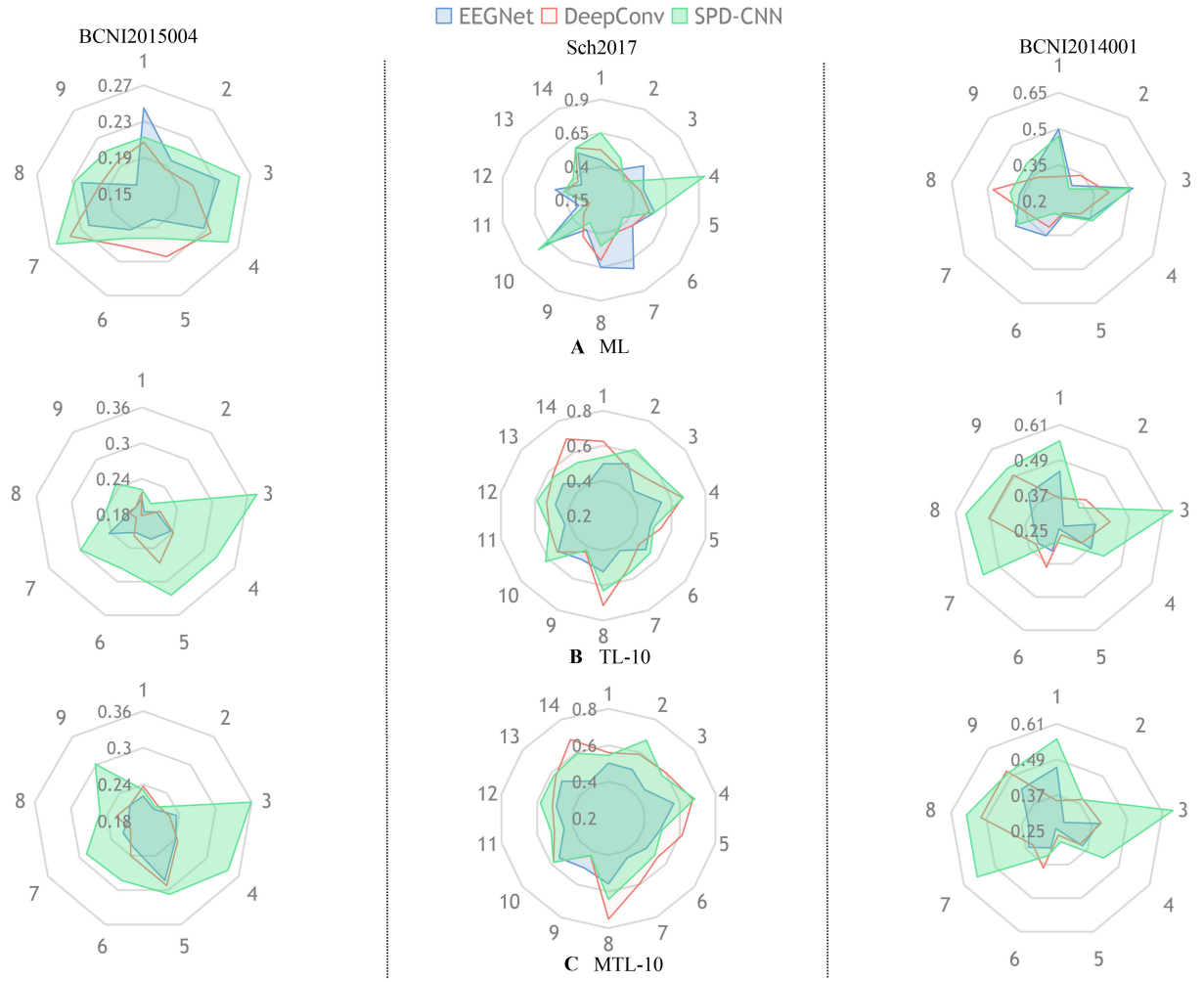



FIGURE 4
In each radar picture, every axis is assigned a variable that represents the classification accuracy of the specific subject (such as 1:subject1 and 2:subject2) and the different colors represent different network architectures (Blue:EEGNet,Red:DeepConvNet,Green:SPD-CNN).Also, the radar pictures arranged in the same column are shown the performance of experiences in the same dataset. The subgraph **(A)** represents the experiments that train the network with the ML method using zero-shot in the target domain. The subgraphs **(B,C)** represent the experiments that train the network with TL and MTL methods, respectively and fine-tune the network using 10shot on the target domain.

domain shift from different subjects and provides a rough direction for the network to upgrade the parameter.

In the meta-update phase(b), we randomly re-initialize the parameter $\theta^*$ first and use the MAML structure (Finn et al., 2017) as a training structure with constraining parameter $\Phi_{ss}$. The $\Phi_{ss}$ is updated by Equation (4) as follows,

$$\Phi_{ss} =: \Phi_{ss} - \lambda \nabla_{\Phi_{ss}} L_{T_{i_1},\cdots,T_{in}}([\Theta;\theta],\Phi_{ss}) \quad (4)$$

where $\lambda$ is the learning rate during the update process of $\Phi_{ss}$. The main idea of constraining parameter $\Phi_{ss}$ is to restrict the learning process of weight and bias in each convolutional layer, which means the weights and the biases of the same CNN layer are scaled and shifted as a whole, respectively.

To be specific, the weights $W$ in the same CNN layer will time a scaling factor $\Phi_{s_1}$ and the biases $b$ in the same CNN layer will add a shifting factor $\Phi_{s_2}$ through an update process. Assuming X is the input data, the SS operation could be expressed by Equation (5).

$$SS(X;,W,b;\Phi_{s_1},\Phi_{s_2}) = (W \odot \Phi_{s_1})X + (b + \Phi_{s_2}) \quad (5)$$

where $\odot$ denotes the element-wise multiplication (For details, refer to the article by Sun et al., 2019).

Inside the MAML learning framework, we sample the data of $j$ classes (where $j$ is the number of ways in few-shot learning) from the same training subject for a task. Therefore, each subject-specific task is seen as an independent sample of the same classification problem.

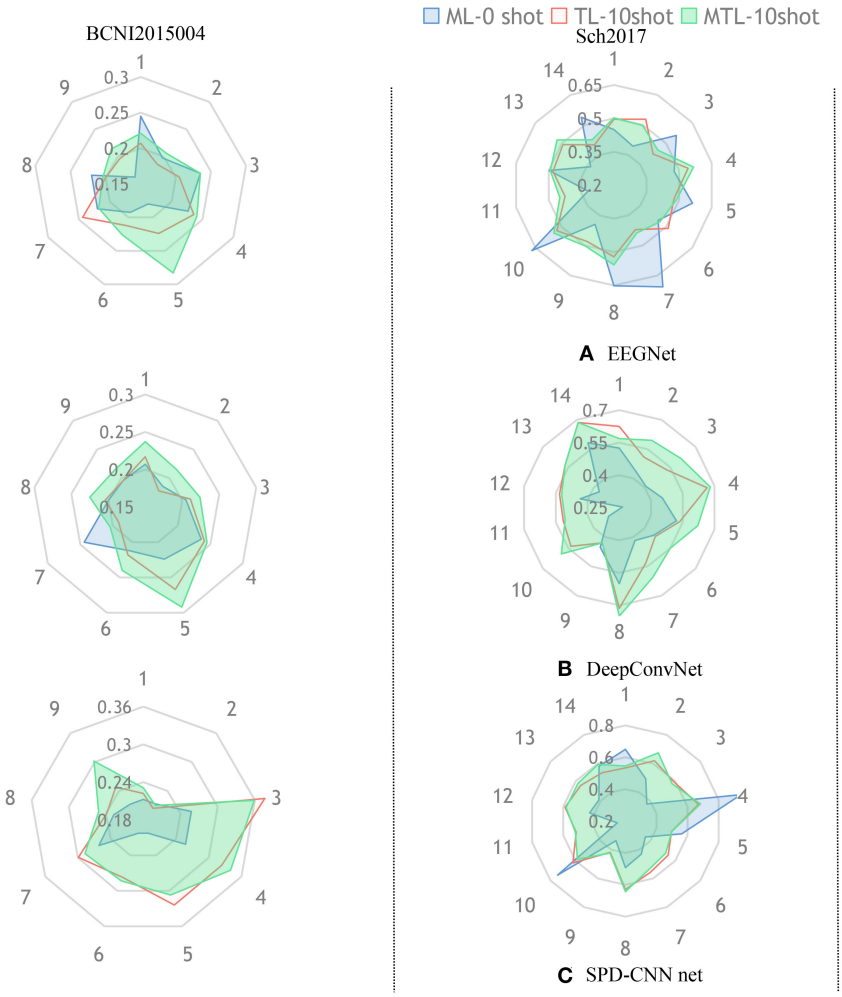



FIGURE 5
The aim of this radar picture is to show the different performances using different training methods and different colors represent different training strategies (Blue:ML,Red:TL,Green:MTL).The three subgraphs **(A–C)** represent the classification performance of the three models, respectively.

More specifically, the train set $D_{tr}$ was segmented into different training tasks $T_{i_k}$ and test tasks $T^*_{i_k}$,where $T_{i_k} \subset T = \{T_{i_1}, T_{i_2}, ...Ti_n\}$ and $T^*_{i_k} \subset T^* = \{T^*_{i_1}, T^*_{i_2}, ...Ti^*_n\}$, n being the number of tasks in meta-learning. Significantly, the data of $T_{i_k}$ and $T^*_{i_k}$ are sampled from the same training subject $x_i$ and the data of the subject-specified task ($T_{i_k}$ or $T^*_{i_k}$) are divided into training data and test data for the training process. As a result, the generalized model $F(\Theta^{pre}, \theta^*)$.

will be trained into different subject-specified networks $f(\Theta^{pre}, \theta^{i_k})$ by gradient descent method. Also, after training the $f_{i_k}(\Theta^{pre}, \theta^{i_k})$ with the training data of the test task $T^*_{i_k}$ again and calculating the loss function based on the test data of the $T^*_{i_k}$, each network $f(\Theta^{pre}, \theta^{i_k})$ would generate subject-specified loss $l_{i_k}$. After updating the parameter $\Theta^{pre}$ several learning epochs, which is guided by the meta-loss $L(\Theta)$ based on different subject-specified loss $l_{i_k}$(refer to Equation 6), the parameters $\Theta^{meta}$ with better generalization ability are selected by validate set $D_{val}$ through the meta-validation process.

$$L(\Theta)_T = \sum l_{i_k}, k = 1, 2, \cdots, n \tag{6}$$

In the domain-adaption phase(c), we fix the parameter of Feature Extractor $\Theta^{meta}$ learned from the Meta-update phase and use the Fine-tune skill to train a user-specify network $f(\Theta^{meta}, \theta^{u_j})$, which is greatly adapted to the user $u_j$ pattern. In this process, a few data of the user from the test set are used to train the $F(\Theta^{meta}, \theta^*)$ into $f(\Theta^{meta}, \theta^{u_j})$ and the parameter of the classifier block is updated by the Equation (7).

$$\theta^{*'} \leftarrow \theta - \beta \nabla_\theta L_{T_{u_j}}([\Theta^{meta}; \theta^*], \Phi_{ss}) \tag{7}$$

TABLE 4 The 4-way, 10-shot, and 5-shot classification accuracy (%) on three datasets (5-way for BNCI2015004).

| Dataset | | | | | |
|---|---|---|---|---|---|
| | | | BNCI2014001 | | |
| Method / Network | ML | TL-10 | MTL-10 | TL-5 | MTL-5 |
| EEGNet | **37.65±2.847** | 35.25±2.7 | 35.95±3.56 | 28.68±3.17 | 29.84±3.15 |
| DeepConv | 33.96±3.14 | 38.75±4.32 | 38.8±3.9 | 34.52±3.52 | 35.64±3.4 |
| **SPD-CNN** | 36.88±3.56 | **46.78±2.78** | **47.44±4.1** | **42.99±2.78** | **43.39±2.63** |
| | | | BNCI2015004 | | |
| EEGNet | 20.88±2.72 | 21.22±3.32 | 22.76±2.36 | 21.3±4.28 | 22.79±3.4 |
| DeepConv | 20.46±2.23 | 21.37±2.21 | 23.29±3.02 | 21.74±3.76 | 22.47±4.31 |
| **SPD-CNN** | **22.74±2.45** | **27.92±3.3** | **28.57±2.58** | **25.62±3.63** | **26.82±3.77** |
| | | | Sch2017 | | |
| EEGNet | 50.2±4.23 | 48.07±3.91 | 49.27±3.12 | 43.82±4.21 | 45.13±4.26 |
| DeepConv | 44.25±3.06 | 56.02±3.49 | **59.22±4.39** | 51±2.67 | **56.4±3.53** |
| **SPD-CNN** | **50.44±3.04** | **56.31±4.53** | 56.92±3.79 | **51.13±3.44** | 52.94±3.67 |

Each accuracy is averaged over all subjects and folds. (in 95% confidence level). Bold values represents best results in this set of experiments.

TABLE 5 The time complexity and scale of the dataset for different networks are compared in this table.

| Method | EEGNet | DeepConvNet | SPD-CNN |
|---|---|---|---|
| Dataset | | Training Time (minute) | |
| BNCI2014001 | 48 m | 156 m | **23m** |
| BNCI2015004 | 52 m | 179 m | **36m** |
| Sch2017 | 218 m | 418 m | **175m** |
| | | Data Size(Gigabyte) | |
| BNCI2014001 | | 0.89G | **0.016G** |
| BNCI2015004 | | 1.5G | **0.022G** |
| Sch2017 | | 30.8G | **1.7G** |

In this table, we train different network architectures using MTL, then calculate the scale of input data and the average consuming time of all folds. Bold values represents best results in this set of experiments.

where $\beta$ is the learning rate during the update process. After this phase, the network is greatly adapted to the situation of user $u_j$ and gains better prediction in the BCI system.

## 3. Experiments and results

Our experiments aim to assess whether SPD-CNN is capable of extracting the discriminative information of EEG data recorded from different subjects and evaluate the transfer capacity of our proposed learning structure in the cross-subject scenario based on the recognition accuracy in the few-shot learning framework.

### 3.1. Implementions details

We conduct normal machine learning and few-shot learning experiments on the cross-subject scenario. In these experiments, we compare SPD-CNN with two wildly used models, DeepConvNet (Lawhern et al., 2018) and EEGnet (Schirrmeister et al., 2017), which perform well on EEG classification with code publicly available. The experiments show the different performance of classification between our training strategy and the benchmark of transfer learning methods in EEG classification.

In the experiences of datasets BNCI2014001 and BNCI20150004, we choose three subjects for the validation set, two subjects as the user for the test set and all the remaining subjects for the training set randomly. This choosing process repeats 18 times, thus, producing 18 different folds. We follow the same procedure for the experiences of dataset Sch2017 except we increase the number of validate subjects to 5 and generate 28 folds.

In the few-shot scenario, we consider the 4-class classification (5-class classification for BNCI2015004), so we sample 4-class(5-class classification for BNCI2015004), 5-shot/10 shot episodes to contain 5 or 10 samples for a train episode and 10 samples (each class) for episode test.

The parameter of the network in our experiments are initialized by the normalization method from He et al. (2015) and the whole model is trained by Adam optimizer (Kingma and Ba, 2014). The learning rates $\alpha$, $\lambda$, and $\beta$ of all learning phases are initialized as 0.001 and dropped by 1% every 10 epochs. All the loss functions are the normal form of cross-entropy cause there is no sample imbalance problem in all datasets (Fatourechi et al., 2008). In the Pre-train phase, the batch size is set to 64 and the network will be trained 50 epochs in each fold. In the experiments of MTL, each task is sampled from the same subjects of all classes evenly in the meta-update phase. Furthermore, we use 60 tasks that form 12 meta-batch(5 tasks for each meta-batch) in one training update loop and choose 30 random tasks for meta-validation and meta-test. In the meta-update phase, the network will be trained 40 epochs in each fold.

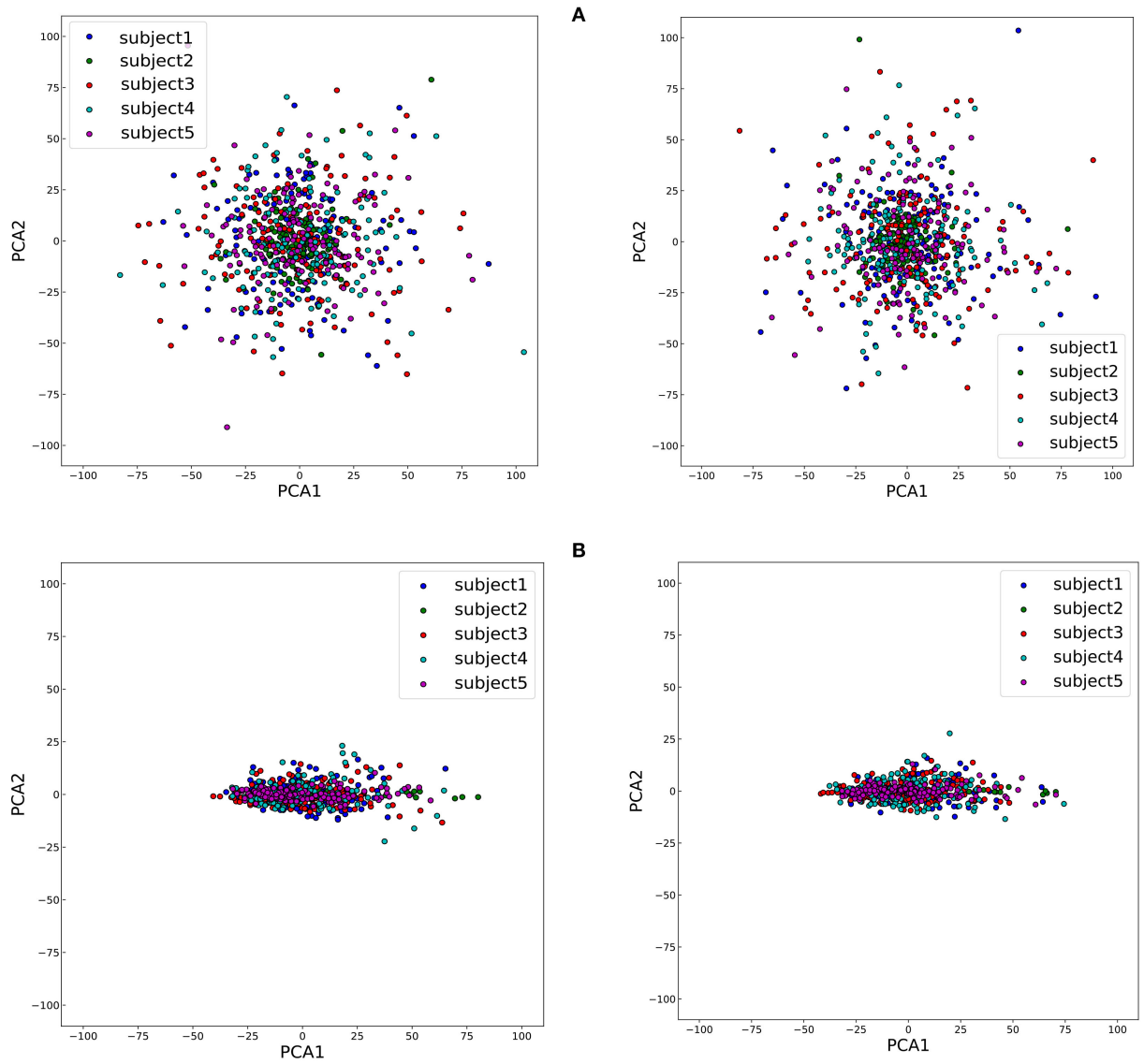



FIGURE 6
The disparity among five subjects of two classes, right hand and feet, which are on the left and right of the figure, respectively. **(A)** The visualization of sample distributions of raw EEG data. **(B)** The visualization of sample distributions of the SPD matrices.

All the models were implemented based on PyTorch (Paszke et al., 2019) and trained on a single GPU of 12 GB TITAN-Xp with Intel Xeon CPU E5-2620 v4 as CPU. More details can be found in the GitHub repository https://github.com/sabinechen/SPD-CNN-Using-Meta-Transfer-Learing-EEG-Cross-Subject-learning.

## 3.2. Experimental evaluation

To show the effectiveness of our model and learning strategy, we design some comparative experiments and ablative settings: Three networks are trained on the chosen dataset using Normal Machine Learning (ML), Transfer Learning (TL), and MTL method. In the experiments of ML, we train the networks from scratch only using the source-domain data, which is also called zero-shot.In the experiments of TL, we pre-train the networks on the source domain and fine-tune the classifier block of the networks on the target domain (5-shot and 10-shot). Figures 4, 5 provide a qualitative summary of the results for the cross-subject classification accuracy.

Figure 4 gives an overall picture of the performances obtained by training EEGNet, DeepConvNet, and SPD-CNN net on the target domains (10shot) with three learning strategies:

ML, TL, and MTL. It shows that the three networks show noticeably varying patterns in the accuracy of different subjects in cross-subject learning. The green area, which represents the performance of SPD-CNN, almost covers other different color areas. It reveals that SPD-CNN has the remarkable ability to transfer source domains (train set) to the target domain (test set) in three datasets.

Figure 5 gives an overall picture of the performances of using different learning strategies on different networks. It shows that the coverage area of MTL is more evenly distributed in all dimensions than other learning strategies in most cases, indicating that the MTL strategy performs better than the other two learning strategies in the experiments. Therefore, we can conclude that the MTL learning strategy strengthens the generalization ability and robustness of the networks.

Furthermore, we present the accuracy of different experiments and give a quantitative summary of the results in Table 4 below.

As can be seen in Table 4, there was a statistically significant difference in the performance of different models [Friedman Test $X^2_{(15)} = 16.53$, $p = 0.0002 < 0.05$, *Post-hoc* analysis with Wilcoxon signed rank tests was conducted] and our model outperforms EEGNet ($p = 0.0003 < 0.05$) and DeepConv ($p = 0.005 < 0.05$) in most cases through vertical comparison in the table.

Also, there was a statistically significant difference in the performance of different learning structures [$X^2_{(9)} = 8.67$, $p = 0.013 < 0.05$] and our learning structure (MTL-10) outperforms the traditional learning structures (TL-10: $p = 0.0039 < 0.05$, ML: $p = 0.019 < 0.05$) by a margin of 0.4–5.4% on accuracy through horizontal comparison and the improvement is much more evident when the data provided by the user for fast adaption (number of shots) is fewer in most cases. Furthermore, DeepConvNet gains much more improvement (about 3–5% in Sch2017) through MTL learning strategy than EEGNet and SPD-CNN net with shallow layers and little parameters. It suggests that the SS operation of MTL can effectively avoid the problem of "catastrophic forgetting" (Lopez-Paz and Ranzato, 2017) (It means forgetting general patterns when adapting to a specific task) and as a result, the performance advantage of large-scale CNN is unleashed thoroughly, especially facing with large-scale data.

Nevertheless, there is no free lunch, DeepConvNet required complex network design, and this kind of large-scale network architecture needs a high level of hardware, which consumes lots of time on the design and calibration in the BCI system. To be specific, the comparison of time complexity and the scale of data of neural networks are shown in Table 5. Table 5 shows that SPD-CNN has a high speed of convergence and shorter training time, which are attributed to the small-scale input data and the plain network structure with little parameter.

As described above, it can be concluded that the proposed SPD-CNN with few learnable parameters has a stronger feature extraction ability to find an approximate boundary to separate different samples from different labels, when the datasets are well described in the SPD manifold. Moreover, with the improvement coming from the MTL learning structure, the CNN-based model would rapidly adapt to the target domain with efficient usage of target data without forgetting key features learned from the source domain.

TABLE 6 The averaged Euclidean distance among different subjects of SPD matrices and EEG data in the feature space.

| Class | Euclidean distance (SPD matrices) | Euclidean distance (EEG data) |
|---|---|---|
| Right hand | 8.85 | 13.85 |
| Feet | 8.71 | 13.72 |
| Left hand | 8.90 | 13.86 |
| Tongue | 8.33 | 13.96 |

# 4. Discussion

## 4.1. Analysis of the SPD descriptor

Extensive experiments above show that the SPD matrices are capable of retaining the discriminative information of brain activity and the information can be effectively extracted by the proposed network.

To study the impact of different data descriptions in the cross-subject learning scenario, the raw EEG data and the SPD matrices of different subjects in BNCI2014001 were reduced to two dimensions by Principal Component Analysis (PCA) and all the samples from the same class were projected to this 2D feature space (Zhang et al., 2018). Consequently, the sample distributions of the different subjects could be visualized in Figure 6. Then we use averaged Euclidean distance to quantitatively measure the distance among different subjects in the feature space, and the result are shown in Table 6.

The results of Figure 6 and Table 6 revealed that the gaps in the sample distributions among different subjects were closed by transforming the EEG data into SPD matrices.

## 4.2. Limitations and future directions

Though the proposed network and learning strategy have achieved great performance in the cross-subject scenario, the limitation is still involved in the current study. For the experiments, we only validate our method on the paradigm of motor imagery and the effectiveness of our method on the other paradigm in the EEG-BCI system is still unclear. Therefore, in future studies, we will focus on the other paradigm such as Steadystate Visually Evoked Potential (SSVEP) datasets and further explore the potential of the proposed approach.

## 5. Conclusion

In this study, we represent a brand-new model to extract cognitive information from EEG data. Compared with the two famous EEG networks, which utilize different convolutional layers to learn specific filters, we transform EEG signals into SPD matrices and design a plain CNN to learn the essential features from it. Considering the shifting problem between different subjects, we use the MTL training strategies to train our model and related experiments show that our training strategy is capable of keeping the adaptation ability of the networks while significantly reducing the number of parameters to transfer. It can be concluded that our proposed model performs well in the cross-subject learning scenario.

Our contribution is part of a larger effort in the BCI learning research, intending to design robust algorithms which use the experience of deep learning in image recognition to mitigate inter-subject variability (Xu et al., 2021) and extract shared information between different subjects. Besides, it is easy to notice that we could use more complex CNN-based models, which have the powerful feature extraction ability for SPD data. Given that, the topic considered here also opens several important questions to be investigated in the future. For instance, considering the feasibility of the network to extract the characteristics of the SPD data, to determine how to design the specific network architecture for this kind of data is promising research. Furthermore, with the feature expression based on the SPD form, data formats of different experiments in the same paradigm can be unified, and it allows us to gather information from several databases and use the CNN-based model to form a more robust classifier in the future.

## Data availability statement

The data that support the findings of this study are openly available in https://github.com/NeuroTechX/moabb.

## Ethics statement

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

## Author contributions

LC developed the theoretical formalism, performed the analytic calculations, and performed the numerical simulations. ZY and JY contributed to the final version of the manuscript. ZY supervised the project. All authors contributed to the article and approved the submitted version.

## Funding

This research was supported in part by the National Natural Science Foundation of China under Grants 61836003 and 61906211 and the Major Program of - Brain Science and Brain-Like Research of the National Science and Technology Innovation 2030 under Grant 2022ZD0211700.

## Acknowledgments

We thank the associate editor and the reviewers for their useful feedback that improved this paper. We are grateful to Professor Zhenghui Gu for her supervising the project and MinLing Feng for her help with the preparation of figures in this paper.

## 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

Alamgir, M., Grosse-Wentrup, M., and Altun, Y. (2010). "Multitask learning for brain-computer interfaces," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (Sardinia: JMLR Workshop and Conference Proceedings), 17–24.

Arvaneh, M., Guan, C., Ang, K. K., and Quek, C. (2013). Eeg data space adaptation to reduce intersession nonstationarity in brain-computer interface. *Neural Comput*. 25, 2146–2171. doi: 10.1162/NECO_a_00474

Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2013). Classification of covariance matrices using a riemannian-based kernel for

bci applications. *Neurocomputing* 112, 172–178. doi: 10.1016/j.neucom.2012.12.039

Fahimi, F., Zhang, Z., Goh, W. B., Lee, T.-S., Ang, K. K., and Guan, C. (2019). Inter-subject transfer learning with an end-to-end deep convolutional neural network for eeg-based bci. *J. Neural Eng.* 16, 026007. doi: 10.1088/1741-2552/aaf3f6

Fatourechi, M., Ward, R. K., Mason, S. G., Huggins, J., Schloegl, A., and Birch, G. E. (2008). "Comparison of evaluation metrics in classification applications with imbalanced datasets," in *2008 Seventh International Conference on Machine Learning and Applications* (San Diego, CA: IEEE). 777–782.

Finn, C., Abbeel, P., and Levine, S. (2017). "Model-agnostic meta-learning for fast adaptation of deep networks," in *International Conference on Machine Learning* (Sydney, NSW: PMLR), 1126–1135.

Göhring, D., Latotzky, D., Wang, M., and Rojas, R. (2013). Semi-autonomous car control using brain computer interfaces. *Intell. Auton. Syst.* 12, 393–408. doi: 10.1007/978-3-642-33932-5_37

Hajinoroozi, M., Zhang, J. M., and Huang, Y. (2017). "Driver's fatigue prediction by deep covariance learning from eeg," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (Banff, AB: IEEE), 240–245.

He, H., and Wu, D. (2019). Transfer learning for brain-computer interfaces: a euclidean space data alignment approach. *IEEE Trans. Biomed. Eng.* 67, 399–410. doi: 10.1109/TBME.2019.2913914

He, K., Zhang, X., Ren, S., and Sun, J. (2015). "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision* (Santiago: IEEE), 1026–1034.

Hong, K.-S., Naseer, N., and Kim, Y.-H. (2015). Classification of prefrontal and motor cortex signals for three-class fnirs-bci. *Neurosci. Lett.* 587, 87–92. doi: 10.1016/j.neulet.2014.12.029

Huang, Z., and Van Gool, L. (2017). "A riemannian network for spd matrix learning," in *Thirty-First AAAI Conference on Artificial Intelligence* San Francisco, FL.

Jayaram, V., Alamgir, M., Altun, Y., Scholkopf, B., and Grosse-Wentrup, M. (2016). Transfer learning in brain-computer interfaces. *IEEE Comput. Intell. Mag.* 11, 20–31. doi: 10.1109/MCI.2015.2501545

Jayaram, V., and Barachant, A. (2018). Moabb: trustworthy algorithm benchmarking for bcis. *J. Neural Eng.* 15, 066011. doi: 10.1088/1741-2552/aadea0

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

Kothe, C. A., and Makeig, S. (2013). Bcilab: a platform for brain-computer interface development. *J. Neural Eng.* 10, 056014. doi: 10.1088/1741-2560/10/5/056014

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

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

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791

Li, D., Ortega, P., Wei, X., and Faisal, A. (2021). "Model-agnostic meta-learning for eeg motor imagery decoding in brain-computer-interfacing," in *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)* (IEEE), 527–530.

Lopez-Paz, D., and Ranzato, M. (2017). "Gradient episodic memory for continual learning," in *Advances in Neural Information Processing Systems 30 (NIPS 2017)* Long Beach, CA: CAES.

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., et al. (2018). A review of classification algorithms for eeg-based brain-computer interfaces: a 10 year update. *J. Neural Eng.* 15, 031005. doi: 10.1088/1741-2552/aab2f2

Lotte, F., and Guan, C. (2010). "Learning from other subjects helps reducing brain-computer interface calibration time," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing* (Dallas, TX: IEEE), 614–617.

Lotze, M., and Halsband, U. (2006). Motor imagery. *J. Physiol. Paris* 99, 386–395. doi: 10.1016/j.jphysparis.2006.03.012

Naseer, N., and Hong, K.-S. (2013). Classification of functional near-infrared spectroscopy signals corresponding to the right-and left-wrist motor imagery for development of a brain-computer interface. *Neurosci. Lett.* 553, 84–89. doi: 10.1016/j.neulet.2013.08.021

Naseer, N., and Hong, K.-S. (2015). Decoding answers to four-choice questions using functional near infrared spectroscopy. *J. Near Infrared Spectrosc.* 23, 23–31. doi: 10.1255/jnirs.1145

Nijholt, A., Tan, D., Allison, B., del, R., Milan, J., and Graimann, B. (2008). "Brain-computer interfaces for hci and games," in *CHI'08 Extended Abstracts on Human Factors in Computing Systems* Florence (3925–3928).

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). "Pytorch: an imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32* (Moscow: Curran Associates, Inc.), 8024–8035.

Rao, R. P. (2013). *Brain-Computer Interfacing: An Introduction*. Cambridge: Cambridge University Press.

Reuderink, B., Farquhar, J., Poel, M., and Nijholt, A. (2011). "A subject-independent brain-computer interface based on smoothed, second-order baselining," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Boston, MA: IEEE), 4600–4604.

Rodrigues, P. L. C., Jutten, C., and Congedo, M. (2018). Riemannian procrustes analysis: transfer learning for brain-computer interfaces. *IEEE Trans. Biomed. Eng.* 66, 2390–2401. doi: 10.1109/TBME.2018.2889705

Scherer, R., Faller, J., Friedrich, E. V., Opisso, E., Costa, U., Kübler, A., et al. (2015). Individually adapted imagery improves brain-computer interface performance in end-users with disability. *PLoS ONE* 10, e0123727. doi: 10.1371/journal.pone.0123727

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

Shanker, M., Hu, M. Y., and Hung, M. S. (1996). Effect of data standardization on neural network training. *Omega* 24, 385–397. doi: 10.1016/0305-0483(96)00010-2

Sun, Q., Liu, Y., Chua, T.-S., and Schiele, B. (2019). "Meta-transfer learning for few-shot learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Long Beach, CA: IEEE), 403–412.

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

Tariq, M., Trivailo, P. M., and Simic, M. (2018). Eeg-based bci control schemes for lower-limb assistive-robots. *Front. Hum. Neurosci.* 12, 312. doi: 10.3389/fnhum.2018.00312

Vanschoren, J. (2018). Meta-learning: a survey. *arXiv preprint arXiv:1810.03548*. doi: 10.48550/arXiv.1810.03548

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. Develop. Syst.* 14, 4-19. Available online at: https://arxiv.org/pdf/2004.06286.pdf

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

Zhang, P., Ma, X., Chen, L., Zhou, J., Wang, C., Li, W., et al. (2018). Decoder calibration with ultra small current sample set for intracortical brain-machine interface. *J. Neural Eng.* 15, 026019. doi: 10.1088/1741-2552/aaa8a4

Zhang, Z., and Sabuncu, M. (2018). "Generalized cross entropy loss for training deep neural networks with noisy labels," in *NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems* Montréal, QC.

frontiers | Frontiers in Neuroinformatics

TYPE Original Research
PUBLISHED 16 August 2022
DOI 10.3389/fninf.2022.907942



OPEN ACCESS

EDITED BY
Ke Liu,
Chongqing University of Posts and Telecommunications, China

REVIEWED BY
Zhengchao Xie,
South China University of Technology, China
Meiyan Xu,
Minnan Normal University, China
Jing Zhao,
University of Macau, Macao SAR, China

*CORRESPONDENCE
Anastasios Bezerianos
anastasios.bezerianos@gmail.com
Hongtao Wang
nushongtaowang@qq.com

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

RECEIVED 05 April 2022
ACCEPTED 16 June 2022
PUBLISHED 16 August 2022

CITATION
Xu T, Dragomir A, Liu X, Yin H, Wan F, Bezerianos A and Wang H (2022) An EEG study of human trust in autonomous vehicles based on graphic theoretical analysis. *Front. Neuroinform.* 16:907942. doi: 10.3389/fninf.2022.907942



# An EEG study of human trust in autonomous vehicles based on graphic theoretical analysis

Tao Xu[1†], Andrei Dragomir[2†], Xucheng Liu[3], Haojun Yin[1], Feng Wan[3], Anastasios Bezerianos[4*] and Hongtao Wang[1*]

[1]The Faculty of Intelligent Manufacturing, Wuyi University, Jiangmen, China, [2]The N1 Institute, National University of Singapore, Singapore, Singapore, [3]Department of Electrical and Computer Engineering, Faculty of Science and Technology, University of Macau, Macau, Macao SAR, China, [4]Hellenic Institute of Transport (HIT), Centre for Research and Technology Hellas (CERTH), Thessaloniki, Greece

With the development of autonomous vehicle technology, human-centered transport research will likely shift to the interaction between humans and vehicles. This study focuses on the human trust variation in autonomous vehicles (AVs) as the technology becomes increasingly intelligent. This study uses electroencephalogram data to analyze human trust in AVs during simulated driving conditions. Two driving conditions, the semi-autonomous and the autonomous, which correspond to the two highest levels of automatic driving, are used for the simulation, accompanied by various driving and car conditions. The graph theoretical analysis (GTA) is the primary method for data analysis. In semi-autonomous driving mode, the local efficiency and cluster coefficient are lower in car-normal conditions than in car-malfunction conditions with the car approaching. This finding suggests that the human brain has a strong information processing ability while facing predictable potential hazards. However, when it comes to a traffic light with a car malfunctioning under the semi-autonomous driving mode, the characteristic path length is higher for the car malfunction manifesting a weak information processing ability while facing unpredictable potential hazards. Furthermore, in fully automatic driving conditions, participants cannot do anything and need low-level brain function to take emergency actions as lower local efficiency and small worldness for car malfunction. Our results shed light on the design of the human-machine interaction and human factor engineering on the high level of an autonomous vehicle.

KEYWORDS

**trust in automation, behavioral modeling, autonomous vehicles, brain functional network, graphic theoretical analysis**

## Introduction

With progressions in automotive safety technologies, both passive and active methods such as lane-centering assistance and adaptive cruise control have contributed to a considerable reduction in traffic fatalities (Van Arem et al., 2006; Lee and Litkouhi, 2012; Milanés et al., 2013). However, the driving death toll in China is still significant and human error accounts for a large proportion (Bhalla et al., 2013). Moreover, traffic congestion is another factor that contributes to traffic accidents. Hence, autonomous vehicles (AVs), which can make optimal decisions bypassing human intervention and avoiding congested routers, have attracted much interest for a long time. In addition to improving driving safety and route planning, AVs can outperform driving efficiency. Most importantly, they can create a passion for driving in people of all ages (Choi and Ji, 2015).

The society of automotive engineering (SAE) categorizes driving automation into six levels ranging from levels 0 to 5. Levels 0 and 5 represent no automation and full automation driving, respectively, in which either the driver or the vehicle independently performs all driving tasks under all conditions. Meanwhile, from levels 1 to 4, the degree of automation increases from drive assistance to high automation. The discrimination of automation levels lies in the drivers' vigilance of the surrounding environment. Interestingly, the discrimination between levels 4 and 5 is blurred due to the condition that the vehicle performs all driving functions. Therefore, it will be interesting to investigate the impact of malfunction of the vehicle under these two degrees of automation as there will likely be a psychological difference between with and without control of a vehicle while facing an emergency for an individual. Evidence shows that the driver and the copilot exhibit different attitudes during an emergency. Therefore, the psychological difference can be regarded as the trust of humans in machines and such trust is most important to developing full automation driving with the maturity of driving automation.

Along with the increasing automation of the AV area, human-computer interaction will be fully utilized in which the human trust in AV will play a crucial role. This is because the driver is permitted to do a secondary task instead of concentrating on driving along the journey (Ma and Kaber, 2005; Carsten et al., 2012; Hergeth et al., 2016; Petersen et al., 2019). Lee and See consider trust from the organizational, sociological, interpersonal, psychological, and neurological perspectives. Trust in AVs can be defined as the human attitude toward how AVs can help achieve user goals in a situation characterized by uncertainty and vulnerability (Lee and See, 2004). They consider how the context, automation characteristics, and cognitive processes affected the appropriateness of trust. Previous studies on human trust in AV limit subjective feelings, such as a well-designed questionnaire for self-reported measurement so that it can be used to repeatedly measure the subjects' trust in the autonomous driving process (Kraus et al., 2020). Moreover, an objective assessment was confined to testing reaction time while facing an emergency (Payre et al., 2016). Then a more comprehensive method includes heart rate measurement and the grasp of eye gaze while executing commands of the driving assistant system (Petersen et al., 2019). However, such measurements can be regarded as the achievement of the delayed and filtered signal from the brain. Therefore, Seet M. et al. (2020) utilized EEG analysis, a fast and highly correlated electrophysiology measurement for trust in AVs. Nonetheless, the analysis focused more on the power spectral density and the functional connectivity graph metrics, which lacks the analysis according to scene switching. In addition, EEG analysis was also used for the trust testing scenario in which the participants in a matrix game included both human and machine counterparts (Dong et al., 2015). Two strategies (collaboration and egoism) were used. Results demonstrated that human-like cues affected neural responses related to the partner's capability. In contrast, in the egoism session, the trust level of predictive partners was reflected by a statistically significant capability effect in the midline electrodes. However, the EEG analysis was confined to the ERP amplitude of different nodes, which lacks the consideration of connectivity between nodes. The results described above suggested that the discrimination of EEG signals in human-computer interaction can be a potential candidate for the study of human trust in AV whereas a more comprehensive method that considers the brain's global or local effect should be proposed.

Recently, graph theoretical analysis (GTA) for functional connectivity networks has attracted much attention. In neuroscience, because of the intricate connections inside the brain, GTA can build a network model that contains regions of interest (nodes) and their connection (edges) to represent characteristics of the brain during different tasks. In this way, both global and local effects of the brain can be analyzed for different tasks. For example, GTA can be used for the diagnosis of degenerative disease and the analysis of working memory tasks (Langer et al., 2013; Sun et al., 2014). The advantage of GTA lies in the analysis of EEG signals in a subdivision frequency (Dai et al., 2017). Many studies have shown that the amplitude of alpha activity is negatively correlated to the number of cortical resources in performing cognitive tasks (Gevins et al., 2012; Roux et al., 2012; León-Domínguez et al., 2015). Therefore, by GTA, the human brain can be modeled as a complex network and have a small-world structure at the level of anatomical as well as functional connectivity (Stam and Reijneveld, 2007).

In this study, we will adopt GTA for the analysis of human trust in AVs. First, we introduce the simulated platform and

the well-designed experimental protocol in the method section. Then we show the results of behavior performance and the analysis based on graph theory. Finally, we discuss our results and provide a conclusion.

# Materials and methods

## Participants

Fifty healthy students aged from 21 to 35 (mean:23.6 *SD* = 3.06) are recruited for this study. They should have normal vision or corrected-to-normal vision with no history of any mental diseases. They are also forbidden to take any medications during the participation of this study. In addition, they should have sufficient driving experience and be aware of the basic traffic rules and norms. This study was approved by the Institutional Review Board (IRB) at the National University of Singapore. Written informed consent was obtained from participants before the study and monetary compensation was given for their participation.

## Driving platform

The driving platform consists of three 65-inch LCD screens for monitoring the driving scene, a driving console (Logitech G27 Racing Wheel; with a steering wheel, a pedalboard, and a gear shifter unit) for operating, and a host computer for the control of the simulation. More details about the experimental senior could be found in our previous studies (Wang et al., 2020a,b). Before the experiment, it takes about 2 min for the participant to get familiar with the platform, which guarantees the comfort of driving for individuals and well follow road rules and navigational instructions. The participant follows instructions of the platform in terms of hearing and vision assembling one abides by the road navigation during driving.

## Experimental protocol

Before starting and after finishing the experiment, participants are asked to fill out two trust questionnaires that help to analyze the initial feeling of the AV and the impact of the experimental procedure on trust, respectively (Jian et al., 2000).

The car drives in an urban area with a scene of cars and pedestrians on the road. There are two basic traffic scenarios that the driver should handle. One scenario is the traffic light (TL) in which the participant should stop the car before a junction while seeing the traffic light is red. The other scenario is a car approaching (CA) in which the participant is asked to stop the car before a junction without any traffic light to avoid collision with other cars. A trial of driving is defined as an encounter with a version of the aforementioned scenario. The intertrial interval is set to be about 1 km distance. The whole experiment consists of two stages one is the practice and the other is the driving simulation. In the practice stage, the driver is navigated to drive the car manually on a road without any junction. After 2-min of driving, the driving platform guides the driver to switch to autonomous mode by pressing a switching button. Then the car will run automatically without any possible crashes and malfunctions for another 2 min.

After the practice stage, it comes to the driving simulation stage, which consists of three modes: manual driving mode (SAE Level 0), semi-autonomous driving mode (SAE Level 4), and fully autonomous driving mode (SAE Level 5). In the manual driving mode, as is shown in **Figure 1**, participants face four junction trials, alternating between the TL and CA junctions. In the semi-autonomous mode, the trial starts with an autonomous mode which can be taken over by pressing the manual button on the steering wheel. The driver goes through eight trials arranged by the program. However, in a trial of the autonomous phase, the car condition may turn from normal to malfunction without advance notice. The malfunction is defined as the car running at the stoplight or the car still running while another car approaches the intersection. There are four trials with malfunction for both semi-autonomous driving mode and autonomous driving mode. The order of malfunction trials should fulfill the following criteria:

1. No strict trial-type alternations (or other discernable patterns) can set up the expectancy of malfunctions and unduly influence trust dynamics as the experiment progresses.
2. No more than two consecutive trials which have the same car condition (normal or malfunction). This was to avoid excessive cumulative loss of trust that would happen if there were massing of malfunctions in succession.
3. To avoid trust loss as the participant gets used to the new phase, the trial sequence must always begin with a normal trial.

At the end of each trial, the driver should guarantee the driving mode is retrieved to the autonomous driving mode. In the fully autonomous mode, the driver will repeat the drive through the routine as in the semi-autonomous mode.

## Signal processing

The flow chart of functional connectivity graph metrics was extracted after analysis of the EEG data as shown in **Figure 2**, which mainly includes parts which are EEG apparatus, EEG data pre-processing, brain functional network, and GTA.

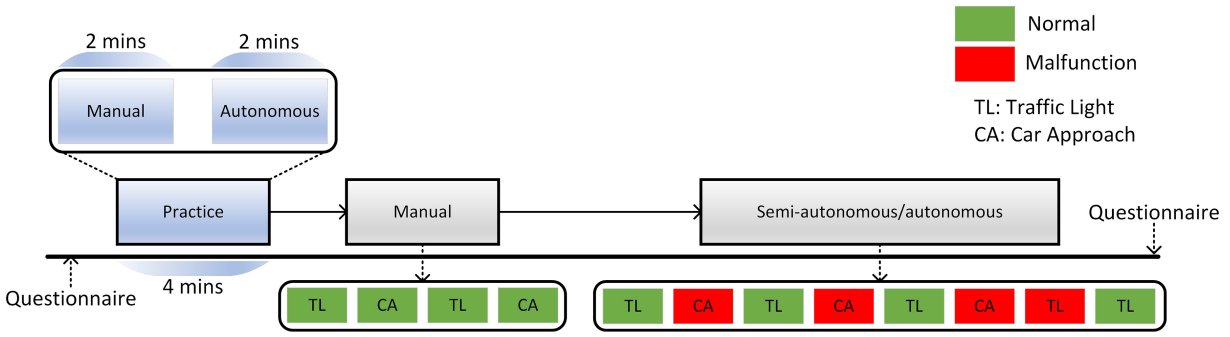



FIGURE 1
The experimental protocol of the work.

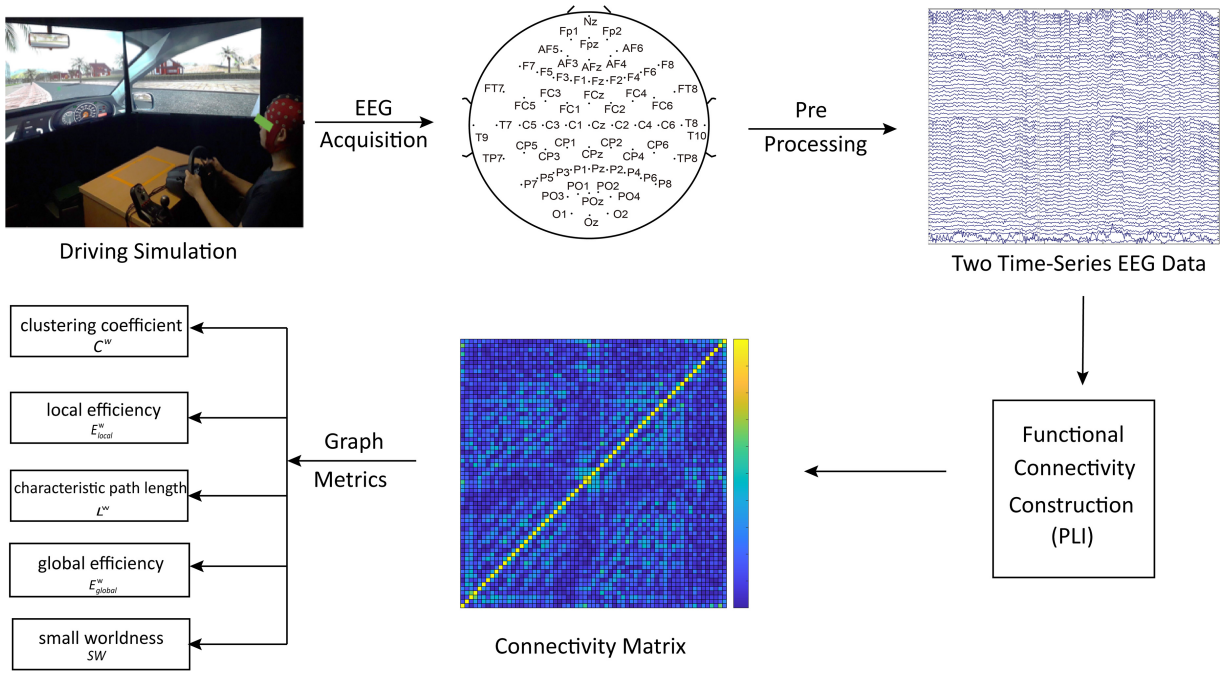



FIGURE 2
Flow chart of functional connectivity graph metrics extracted after analysis of the EEG data.

## EEG apparatus

During the simulated driving, EEG data are recorded using Waveguard$^{TM}$ caps (CA-142; ANT Neuro, Netherlands) with a sampling rate of 512 Hz and 64 Ag/AgCl electrodes whose impedance is below 15 kΩ. We also use additional electrodes to record horizontal and vertical electrooculograms (hEOG and vEOG) on both temples, as well as below and above the right eye. In particular, participants are asked to reduce unnecessary movements for the reduction of artifacts during driving.

## EEG data preprocessing

The recorded EEG signals will be resampled to 256 Hz and bandpass filtered between 0.3 and 40 Hz. Then the processed data will be re-referenced to the left and right mastoids. At the same time, we will remove ocular and muscle artifacts with automatic artifact rejection (AAR) (Sun et al., 2014). The channels that have poor contact with the scalp will be replaced with interpolated signals of neighboring channels. Then we will do data segmentation according to the trial that the driver simulated. For each trial, 2 s were selected from the onset of the traffic light turning yellow at junctions or when the first car can be seen at intersections. In semi-autonomous or fully autonomous mode, the car may turn to malfunction during such period and thus permits us to probe into drivers' cognitive states when they react to the AV. Finally, we will use independent

component analysis to filter data with only EEG signals remaining (Figure 3).

## Brain functional network

After the EEG data preprocessing, we employed phase synchronization (PS) to compute the statistical coupling for the functional connectivity construction in each frequency band. It is the same as the feature of EEG data, the PS was computed between two-time series. We employed the phase lag index (PLI) to estimate the degree of pairwise coupling. The EEG signals were divided by dividing the floating time window over the step for each band.

We employed the Hilbert transform to compute the instantaneous phase $z_i(t)$:

$$z_i(t) = s_i(t) + jHT(s_i(t)) \tag{1}$$

where $HT(s_i(t))$ is the Hilbert transform of each time series $s_i(t)$, which is estimated by:

$$HT(s_i(t)) = \frac{1}{\pi} P.V. \int_{-\infty}^{\infty} \frac{s_i(t)}{t-\tau} d\tau \tag{2}$$

In Eq.2, $P.V.$ represents Cauchy principal value. Once the phase of each time series is computed, the relative phase locking can be estimated as:

$$\Delta\varphi(t) = arg\left(\frac{z_1(t)\, z_2^*(t)}{|z_1(t)|\,|z_2(t)|}\right) \tag{3}$$

The PLI value ranges between 0 and 1 and is calculated with the following equation:

$$PLI = \left|\langle sign\varphi(t)\rangle\right| \tag{4}$$

The PLI value is defined as [0,1] with 0 representing the case where there is no phase synchronization (PS), while 1 represents the perfect phase locking between two-time series.

## Graph theoretical analysis

To delve into the unknown information in the EEG data, we employed the method of GTA after building the functional connectivity network.

There were N * N adjacency matrices ($N$ = 64 in this study) computed after building the functional connectivity network, which represents the connectivity structures of brain nodes. Because the functional connectivity network contains complex information and numerous useless combinations, the sparsity ranged from 10 to 20% with the step of 1% utilized in these networks, which is the ratio of the present connection number to preserve a real functional connection. In this way, we will transform the PLI matrix into a sparse matrix with different thresholds. The threshold is a proportion of the most important elements within the PLI matrix. We need to do such transformation 11 times with the proportion ranging from 10 to 20%. Hence, the weighted adjacency matrices were computed, which preserved the connection

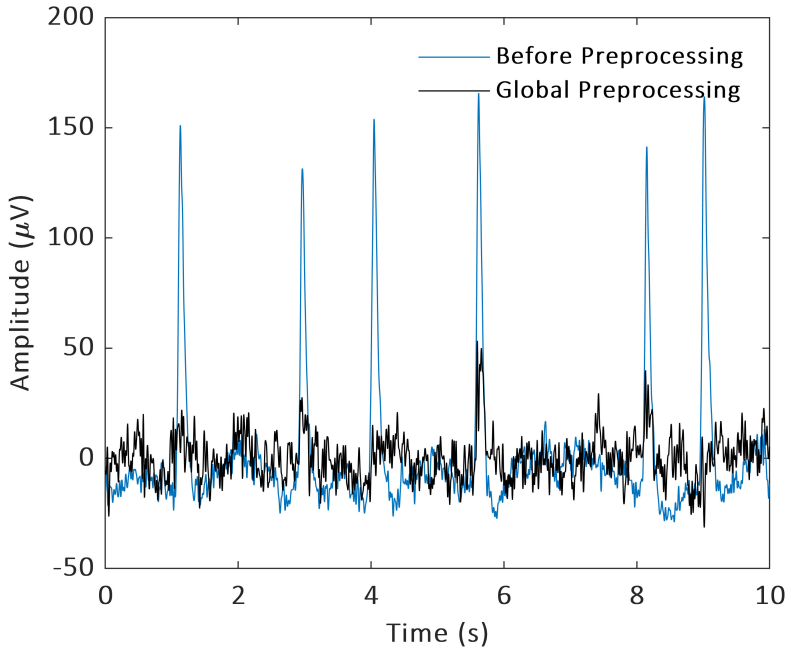



FIGURE 3
Data preprocessing.

strength of the real connections. For the graph-theoretical properties with the considered sparsity, the area under the curve of the corresponding properties was extracted as features for further study.

To quantitatively investigate the topological properties of functional connectivity between the CA condition and TL condition, we implemented GTA with the Brain Connectivity Toolbox. We characterized the graph in the aspects of local segregation including clustering coefficient ($C^w$) and local efficiency ($E^w_{local}$) and global integration including characteristic path length ($L^w$) and global efficiency ($E^w_{global}$) and small worldness ($SW$) based on the weighted adjacency matrices.

$C^w$ is the main indicator of information differentiation in complex network computing, which can measure the degree of aggregation within the brain functional network and reflect the possibility of each node being a neighbor. It is given by Equation (5):

$$C^w = \frac{1}{n}\sum_{i\in N}\frac{2t_i^w}{k_i\left(k_i-1\right)} \tag{5}$$

where $t_i^w$ is the number of connections, which is the weighted geometric mean of a triangle in the neighbored node $i$ and $k_i$ is the number of connected nodes of $i$. $C^w$ reflects that the network forms the tendency of the local loop, and the bigger the $C^w$. the more nodes connecting with $i$. $L^w$ is the mean of the shortest path length and is the path with the maximum total weight between vertices. It is given by Equation (6):

$$L^w = \frac{1}{n}\sum_{i\in N}\frac{\sum_{j\in N,\, j\neq i} d_{ij}^w}{n-1} \tag{6}$$

where $d_{ij}^w$ is the shortest path length between node $i$ and node $j$. $L^w$ is the main indicator of global integration. The shorter the path length, the stronger the functional integration and the more direct connections between brain regions. The calculation method of $SW$ is shown in Equation (7):

$$SW = \frac{C^w/C^w_{rand}}{L^w/L^w_{rand}} \tag{7}$$

where $L^w_{rand}$ and $C^w_{rand}$ is the mean of a random network of $C^w$ and $L^w$ after 100 times random circulation. They have the same degree, node, and edge with a functional connection network. $E^w_{global}$ measures the capability of global information transmission and is the inverse of the length of the shortest path. It is shown in Equation (8):

$$E^w_{global} = \frac{1}{n}\sum_{i\in N}\frac{\sum_{j\in N, j\neq i}\left(d_{ij}^w\right)^{-1}}{n-1} \tag{8}$$

$E^w_{global}$ is a measure that evaluates the efficiency of information transfer within a region of the network. It is shown in Equation (9):

$$E^w_{local}=\frac{1}{2}\sum_{i\in N}\frac{\sum_{j,h\in N, j\neq i}\left(w_{ij}w_{ih}\left[d_{jh}^w\left(N_i\right)\right]^{-1}\right)^{1/3}}{k_i\left(k_i-1\right)} \tag{9}$$

where $w_{ij}$ is the connecting weight between node $i$ and node $j$.

# Results

## Behavioral performance

The quality of the recorded data was validated by the trial-by-trial trust rating, takeover decision-making, and user preference. From the trial-by-trial trust rating, there was no difference between normal and malfunctioning trials in the semi-autonomous driving mode ($p = 0.82$). Furthermore, participants showed lower trust after car malfunction than in car normal conditions ($p < 0.001$). we also find that participants in the semi-autonomous driving mode have a stronger willingness to take over the control of the vehicle during the malfunction trials ($p < 0.001$). Finally, participants prefer the semi-autonomous driving mode to the fully autonomous driving mode ($p < 0.001$).

## Graph theoretical analysis in semi-autonomous driving mode

The graph's theoretical properties show significant local segregation of the brain function during the semi-autonomous driving condition. In the theta band and the car approaching condition, the participants show significant higher local efficiency [$F_{(1,\ 74)} = 4.848$, $p = 0.031$, $\eta^2 = 0.061$, $0.192 \pm 0.007$ vs. $0.189 \pm 0.008$] and clustering coefficient [$F_{(1,\ 74)} = 6.716$, $p = 0.012$, $\eta^2 = 0.083$, $0.128 \pm 0.009$ vs. $0.123 \pm 0.009$] in malfunction trials than normal trials (**Figure 4**), which suggests the human brain showed more efficient information processing ability when participants encounter malfunction of a vehicle.

On the contrary, in the TL condition, participants show significant higher characteristic path length [$F_{(1,\ 74)} = 6.084$, $p = 0.015$, $\eta^2 = 0.077$, $0.608 \pm 0.019$ vs. $0.599 \pm 0.010$] and lower global efficiency [$F_{(1,\ 74)} = 6.379$, $p = 0.014$, $\eta^2 = 0.079$, $0.174 \pm 0.003$ vs. $0.176 \pm 0.002$] in the theta band during the malfunction condition (**Figure 5**). Such results suggest human brains have low level of information processing ability when approaching the traffic light.

## Graph theoretical properties in full automation condition

In this paper, the graph-theoretical properties in autonomous driving conditions were evaluated. In CA condition, participants show significant lower local efficiency [$F_{(1,\ 74)} = 4.491$, $p = 0.029$, $\eta^2 = 0.063$ $0.188 \pm 0.007$ vs. $0.192 \pm 0.008$] in beta band during malfunction occurred (**Figure 6**). Furthermore, the significant lower small worldness

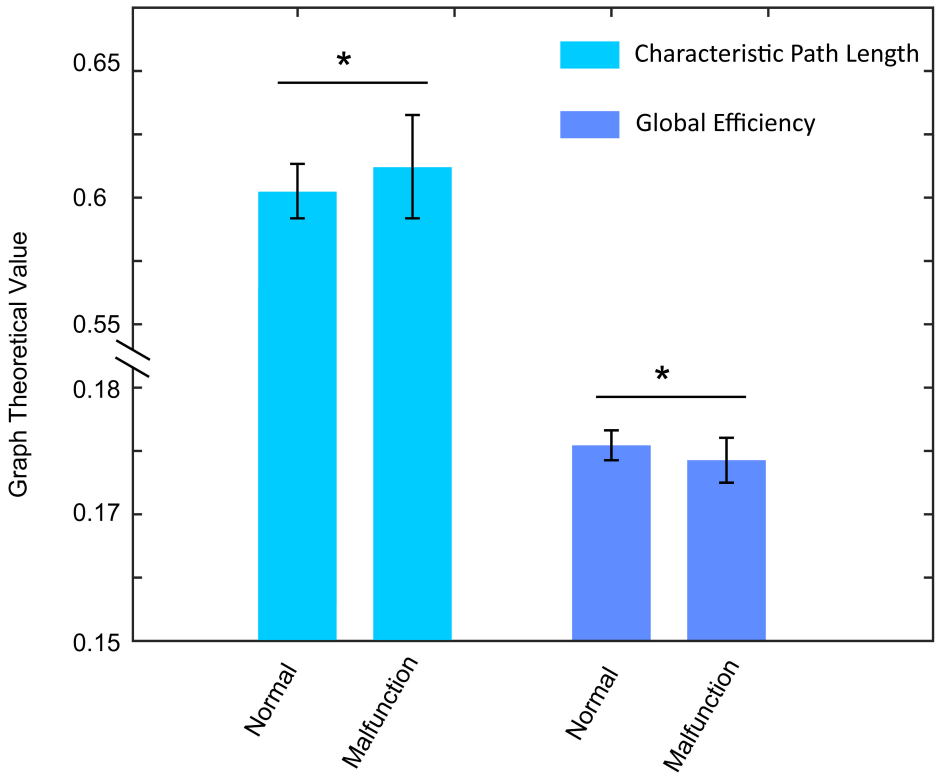



FIGURE 4
The local efficiency and clustering coefficient alterations between normal condition and malfunction condition in the CA condition. Both two indexes show significantly higher values in malfunction conditions. *$p < 0.01$.

[$F_{(1,\ 74)} = 6.982$, $p = 0.010$, $\eta^2 = 0.086$ $0.415 \pm 0.035$ vs. $0.436 \pm 0.034$] was observed which means the brain has lower information processing ability in such condition (**Figure 6**). Such results may suggest participants cannot do anything in the fully autonomous driving condition and need low-level brain function to take emergency actions.

# Discussion

## Resource identification initiative

The extent to trust in new technology always decides the speed of the development of the corresponding technology (Gefen and Straub, 2000; Gefen et al., 2003), especially in various automation that trust is a decisive factor in the acceptance of automation (Lee and Moray, 1992; Pavlou, 2003; Lee and See, 2004; Carter and Bélanger, 2005; Parasuraman et al., 2008). For example, Jong and Yong investigated the importance of trust in adopting AVs and the factors that promote people to trust AVs (Choi and Ji, 2015). In addition, digging insights into factors that build trust can encourage a better understanding of trust in a specific item (Leimeister et al., 2005). Therefore, systematically studying AV can boost us to understand ourselves more clearly. For example, Seet M.S. et al. (2020) optimize driver-vehicle trust management according to the subtypes of trust in AV. The trust can be subdivided into competence-based trust (CT) and integrity-based trust (IT) which refer to the functional capacity of AV and the integrity of AV that will not cause deliberate harm, respectively. However, most of the analysis of trust is based on the analysis of the collected questionnaire, which is more subjective and lacks a more comprehensive consideration. These questionnaires often assume some scenarios and preconditions. However, if we encounter the scene in the questionnaire, people are often at a loss in practice. Meanwhile, survey-based investigations always give out similar conclusions in which the circumstances, as well as the performance of a robot, directly affect trust, satisfaction, and frustration. It is hard to objectively elucidate how these factors influence our trust in the interaction of human-robot let alone emergent human-vehicle interaction (Castelfranchi and Falcone, 2000; Murphy et al., 2004; Hancock et al., 2011; Abd et al., 2017). Therefore, a more objective method that seeks factors that contribute to human trust in AV is needed.

In this study, we adopt the GTA for the investigation of human trust in a high level of autonomous driving, including semi-autonomous driving (SAE level 4) and autonomous driving (SAE level 5) in a simulated driving environment. The graph-theoretical properties are efficient approaches to

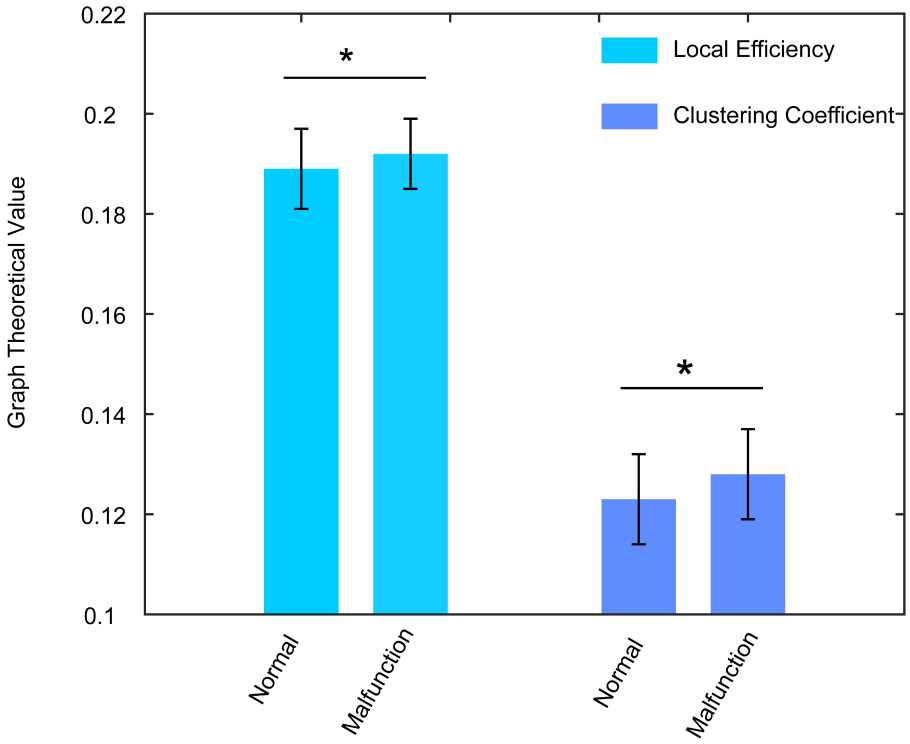



FIGURE 5
The characteristic path length and global efficiency alterations between normal condition and malfunction condition in the TL condition. Both two indexes show significantly higher values in malfunction conditions. *$p < 0.01$.

evaluating the function of the human brain. Particularly, the human brain showed local segregation and global integration of brain functions (Dai et al., 2017). On one hand, the local efficiency and the clustering coefficient are measurements of the brain's local information transmission ability. On the other hand, the characteristic path length and global efficiency measured the information spreading ability of the whole brain. The local and global properties are the critical indexes to assess the brain states of different driving conditions and reflect the trust degree during driving a vehicle.

Seet M. et al. (2020) also paid attention to the EEG-based analysis of human trust in AV. And they also concluded that a reduction in trust during full automation malfunctions. However, they focus on the analysis of brain regional influence on the driving condition. For example, they found that there was a remarkable decrease in functional segregation in the right frontal area during the fully autonomous driving mode and such regional discrimination may be related to the momentary impairment of the ability to plan logically about specific actions. In contrast, our analysis focuses more on the influence of the driving scene on brain activity.

The development of human trust in AVs can be divided into several stages. The first one is to investigate human trust once there is a malfunction of the vehicle. In previous work, Seet M. et al. (2020) used self-reported trust ratings to demonstrate the difference in human trust in normal or malfunctioning driving conditions in both high automation mode and full automation mode. They found that there is no significant difference between normal and malfunction trials in high automation driving mode whereas there is a significant difference for that in full automation mode. And drivers are prone to take over the task in high automation mode once there is a malfunction. The second stage that we need to focus on is to elucidate how the brain reacts to the different scenarios during automatic driving. In this study, we focus on the analysis of the brain reaction to CA and TL conditions with different car conditions. As is shown in **Figures 4**, **5**, there is an opposite way for the brain to process the malfunction in which it shows high local efficiency and low global efficiency for CA and TL, respectively. The opposite information-processing ability of the human brain during the CA condition and TL condition demonstrates that participants have different levels of trust during these two conditions. When the vehicle runs into a complex environment, such as an intersection without a traffic light, the participants show a low level of trust in the machine and can handle an emergency in time (higher information processing ability). However, while driving in a safe condition (TL condition), participants show a high level of trust in the machine and cannot take emergency action in time (lower information processing ability). The third stage

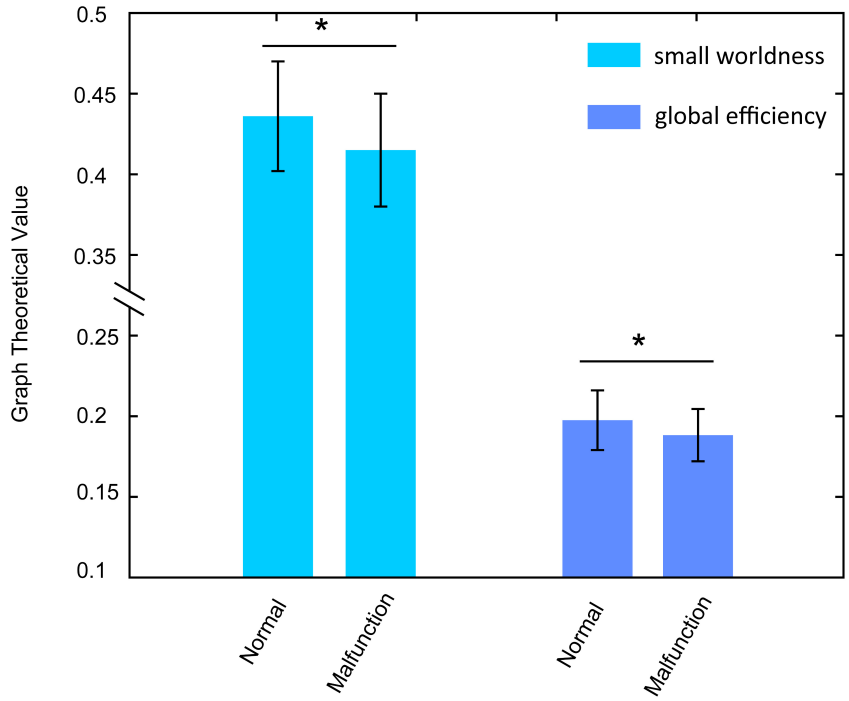



FIGURE 6
The small worldness and local efficiency alterations between normal condition and malfunction condition in the CA condition. Both two indexes show significantly lower values in malfunction conditions. *$p < 0.01$.

may lie in the facilitation of human trust in AVs. However, the premise of many studies is that the driver should observe the vehicle's performance and be ready to take over the task once there is an emergency. The secondary task is always regarded as the distraction of driving that deviates from the original intention of automatic driving. Situational awareness can help the driver to promote their trust in AVs (Miller et al., 2014). Petersen et al. (2019) changed the situational awareness with the variation of a verbal message to the driver and found that the high situational awareness condition can cause a significantly high level of trusting behavior. In the future, we can also add voice prompt with situational awareness into the experiment for the analysis of brain reaction to the AV.

Seet M. et al. (2020) also analyzed the AV malfunction on human trust. However, there are some discriminations between these two works. One of them is that the frequency band used for analysis is different. In our analysis, we focus on the theta band whereas they aim on the alpha band.

## Conclusion

In this study, a simulated driving platform with an EEG data collection system is used for the evaluation of human trust in AVs. The behavior performance shows that the driver has less trust during the fully automatic driving mode. We also used GTA to illustrate how the brain reacts to both semi-autonomous driving mode and fully autonomous driving mode. Our results have the potential to be adopted for the improvement of human trust in AV.

## 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 Institutional Review Board (IRB) at National University of Singapore. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## Author contributions

TX, AB, and HW designed the study. XL and HY wrote the code. AD carried out the experiments and analyzed the data. TX

wrote the manuscript. FW contributed to the conception of the study. HW helped draft the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by Special Projects in Key Fields Supported by the Technology Development Project of Guangdong Province (no. 2020ZDZX3018), the Special Fund for Science and Technology of Guangdong Province (no. 2020182), the Wuyi University, and Hong Kong and Macao joint Research and Development Project under Grant (2019WGALH16), the Jiangmen Science and Technology Project (2020JC01020 and 2020JC01036), Startup funds for Scientific Research of High-level Talents of Wuyi University (2019AL020), the Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging under Grant SZD201909, and the EU's H2020 Research and Innovation Programme [Proactive Safety Systems and Tools for a Constantly Upgrading Road Environment (SAFE_UP)] (no. 031019).

## Acknowledgments

We thank Wang Tian and Nida Abbasi for their assistance in equipment preparation and data collection.

## Conflict of interest

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

The reviewer JZ declared a shared affiliation with the authors XL and FW to the handling editor at the time of review.

## Publisher's note



## References

Abd, M., Gonzalez, I., Nojoumian, M., and Engeberg, E. (2017). "Impacts of robot assistant performance on human trust, satisfaction, and frustration," in *RSS: Morality and Social Trust in Autonomous Robots*, Boston.

Bhalla, K., Li, Q., Duan, L., Wang, Y., Bishai, D., and Hyder, A. A. (2013). The prevalence of speeding and drink driving in two cities in China: A mid project evaluation of ongoing road safety interventions. *Injury* 44:S49–S56. doi: 10.1016/S0020-1383(13)70213-4

Carsten, O., Lai, F. C., Barnard, Y., Jamson, A. H., and Merat, N. (2012). Control task substitution in semiautomated driving: Does it matter what aspects are automated? *Hum. Factors* 54, 747–761. doi: 10.1177/0018720812460246

Carter, L., and Bélanger, F. (2005). The utilization of e-government services: Citizen trust, innovation and acceptance factors. *Inf. Syst. J.* 15, 5–25. doi: 10.1111/j.1365-2575.2005.00183.x

Castelfranchi, C., and Falcone, R. (2000). Trust and control: A dialectic link. *Appl. Artif. Intell.* 14, 799–823. doi: 10.1080/08839510050127560

Choi, J. K., and Ji, Y. G. (2015). Investigating the importance of trust on adopting an autonomous vehicle. *Int. J. Hum. Comput. Interact.* 31, 692–702. doi: 10.1016/j.apergo.2018.04.006

Dai, Z., De Souza, J., Lim, J., Ho, P. M., Chen, Y., Li, J., et al. (2017). EEG cortical connectivity analysis of working memory reveals topological reorganization in theta and alpha bands. *Front. Hum. Neurosci.* 11:237. doi: 10.3389/fnhum.2017.00237

Dong, S.-Y., Kim, B.-K., Lee, K., and Lee, S.-Y. (2015). "A preliminary study on human trust measurements by EEG for human-machine interactions," in *Proceedings of the 3rd International Conference on Human-Agent Interaction* (New York, NY: ACM), 265–268. doi: 10.1145/2814940.2814993

Gefen, D., and Straub, D. (2000). The relative importance of perceived ease of use in IS adoption: A study of e-commerce adoption. *J. Assoc. Inf. Syst.* 8, 1–30. doi: 10.17705/1jais.00008

Gefen, D., Karahanna, E., and Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Q.* 27, 51–90. doi: 10.2307/30036519

Gevins, A., Mcevoy, L. K., Smith, M. E., Chan, C. S., Sam-Vargas, L., Baum, C., et al. (2012). Long-term and within-day variability of working memory performance and EEG in individuals. *Clin. Neurophysiol.* 123, 1291–1299. doi: 10.1016/j.clinph.2011.11.004

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., and Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Hum. Factors* 53, 517–527. doi: 10.1177/0018720811417254

Hergeth, S., Lorenz, L., Vilimek, R., and Krems, J. F. (2016). Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving. *Hum. Factors* 58, 509–519. doi: 10.1177/0018720815625744

Jian, J.-Y., Bisantz, A. M., and Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *Int. J. Cogn. Ergon.* 4, 53–71. doi: 10.1186/s13054-016-1208-6

Kraus, J., Scholz, D., Stiegemeier, D., and Baumann, M. (2020). The more you know: Trust dynamics and calibration in highly automated driving and the effects of take-overs, system malfunction, and system transparency. *Hum. Factors* 62, 718–736. doi: 10.1177/0018720819853686

Langer, N., Von Bastian, C. C., Wirz, H., Oberauer, K., and Jäncke, L. (2013). The effects of working memory training on functional brain network efficiency. *Cortex* 49, 2424–2438. doi: 10.1016/j.cortex.2013.01.008

Lee, J. D., and See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Hum. Factors* 46, 50–80. doi: 10.1518/hfes.46.1.50.30392

Lee, J., and Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics* 35, 1243–1270. doi: 10.1080/00140139208967392

Lee, J.-W., and Litkouhi, B. (2012). "A unified framework of the automated lane centering/changing control for motion smoothness adaptation," in *2012 15th International IEEE Conference on Intelligent Transportation Systems* (Piscataway: IEEE), 282–287. doi: 10.1109/ITSC.2012.6338738

Leimeister, J. M., Ebner, W., and Krcmar, H. (2005). Design, implementation, and evaluation of trust-supporting components in virtual communities for patients. *J. Manag. Inf. Syst.* 21, 101–131. doi: 10.1080/07421222.2005.11045825

León-Domínguez, U., MartíN-Rodríguez, J. F., and LeóN-CarrióN, J. (2015). Executive n-back tasks for the neuropsychological assessment of working memory. *Behav. Brain Res.* 292, 167–173. doi: 10.1016/j.bbr.2015.06.002

Ma, R., and Kaber, D. B. (2005). Situation awareness and workload in driving while using adaptive cruise control and a cell phone. *Int. J. Ind. Ergon.* 35, 939–953. doi: 10.1016/j.ergon.2005.04.002

Milanés, V., Shladover, S. E., Spring, J., Nowakowski, C., Kawazoe, H., and Nakamura, M. (2013). Cooperative adaptive cruise control in real traffic situations. *IEEE Trans. Intell. Transp. Syst.* 15, 296–305. doi: 10.1080/10962247.2020.1750505

Miller, D., Sun, A., and Ju, W. (2014). "Situation awareness with different levels of automation," in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (Piscataway: IEEE), 688–693. doi: 10.1109/SMC.2014.6973989

Murphy, R. R., Riddle, D., and Rasmussen, E. (2004). "Robot-assisted medical reachback: a survey of how medical personnel expect to interact with rescue robots. RO-MAN 2004," in *13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)* (Piscataway: IEEE), 301–306. doi: 10.1109/ROMAN.2004.1374777

Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *J. Cogn. Eng. Decis. Mak.* 2, 140–160. doi: 10.1518/155534308X284417

Pavlou, P. A. (2003). Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *Int. J. Electron. Commer.* 7, 101–134. doi: 10.1080/10864415.2003.11044275

Payre, W., Cestac, J., and Delhomme, P. (2016). Fully automated driving: Impact of trust and practice on manual control recovery. *Hum. Factors* 58, 229–241. doi: 10.1177/0018720815612319

Petersen, L., Robert, L., Yang, J., and Tilbury, D. (2019). Situational awareness, driver's trust in automated driving systems and secondary task performance. *SAE Int. J. Connect. Auton. Vehicles Forthcoming* 2, 129–141. doi: 10.4271/12-02-02-0009

Roux, F., Wibral, M., Mohr, H. M., Singer, W., and Uhlhaas, P. J. (2012). Gamma-band activity in human prefrontal cortex codes for the number of relevant items maintained in working memory. *J. Neurosci.* 32, 12411–12420. doi: 10.1523/JNEUROSCI.0421-12.2012

Seet, M., Harvy, J., Bose, R., Dragomir, A., Bezerianos, A., and Thakor, N. (2020). "Differential impact of autonomous vehicle malfunctions on human trust," in *IEEE Transactions on Intelligent Transportation Systems,* Vol. 99, 1-10.

Seet, M. S., Dragomir, A., Mathialagan, I., Ann, L. Y., Binte Zaid, Z., Ramapatna, S. L., et al. (2020). "Subtype Divergences of Trust in Autonomous Vehicles: Towards Optimisation of Driver–Vehicle Trust Management," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)* (Piscataway: IEEE), 1–6. doi: 10.1109/ITSC45102.2020.9294495

Stam, C. J., and Reijneveld, J. C. (2007). Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomed. Phys.* 1:3. doi: 10.1186/1753-4631-1-3

Sun, Y., Yin, Q., Fang, R., Yan, X., Wang, Y., Bezerianos, A., et al. (2014). Disrupted functional brain connectivity and its association to structural connectivity in amnestic mild cognitive impairment and Alzheimer's disease. *PLoS One* 9:e96505. doi: 10.1371/journal.pone.0096505

Van Arem, B., Van Driel, C. J., and Visser, R. (2006). The impact of cooperative adaptive cruise control on traffic-flow characteristics. *IEEE Trans. Intell. Transp. Syst.* 7, 429–436. doi: 10.1016/j.jsr.2020.09.012

Wang, H., Liu, X., Hu, H., Wan, F., and Jung, T. P. (2020a). Dynamic Reorganization of Functional Connectivity Unmasks Fatigue Related Performance Declines in Simulated Driving. *IEEE Trans. Neural Syst.* 28, 1790–1799. doi: 10.1109/TNSRE.2020.2999599

Wang, H., Liu, X., Li, J., Xu, T., Bezerianos, A., Sun, Y., et al. (2020b). Driving Fatigue Recognition with Functional Connectivity Based on Phase Synchronization. *IEEE Trans. Cogn.* 99:1.

frontiers | Frontiers in Neuroinformatics

TYPE Original Research
PUBLISHED 18 August 2022
DOI 10.3389/fninf.2022.934480



OPEN ACCESS

EDITED BY
Ke Liu,
Chongqing University of Posts and Telecommunications, China

REVIEWED BY
Jinping Xu,
Shenzhen Institutes of Advanced Technology (CAS), China
Chao Wang,
Shenzhen University, China

*CORRESPONDENCE
Ting Wu
fsyy00598@njucm.edu.cn

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

RECEIVED 02 May 2022
ACCEPTED 27 June 2022
PUBLISHED 18 August 2022

CITATION
Wang Y, Li Z, Zhang Y, Long Y, Xie X and Wu T (2022) Classification of partial seizures based on functional connectivity: A MEG study with support vector machine.
*Front. Neuroinform.* 16:934480.
doi: 10.3389/fninf.2022.934480



# Classification of partial seizures based on functional connectivity: A MEG study with support vector machine

Yingwei Wang[1†], Zhongjie Li[2†], Yujin Zhang[3,4†], Yingming Long[1†], Xinyan Xie[1] and Ting Wu[1,5*]

[1]Department of Radiology, Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, China, [2]College of Intelligence and Computing, Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, Tianjin, China, [3]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, [4]Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing, China, [5]Department of Magnetoencephalography, Nanjing Brain Hospital, Affiliated to Nanjing Medical University, Nanjing, China

Temporal lobe epilepsy (TLE) is a chronic neurological disorder that is divided into two subtypes, complex partial seizures (CPS) and simple partial seizures (SPS), based on clinical phenotypes. Revealing differences among the functional networks of different types of TLE can lead to a better understanding of the symbology of epilepsy. Whereas Although most studies had focused on differences between epileptic patients and healthy controls, the neural mechanisms behind the differences in clinical representations of CPS and SPS were unclear. In the context of the era of precision, medicine makes precise classification of CPS and SPS, which is crucial. To address the above issues, we aimed to investigate the functional network differences between CPS and SPS by constructing support vector machine (SVM) models. They mainly include magnetoencephalography (MEG) data acquisition and processing, construction of functional connectivity matrix of the brain network, and the use of SVM to identify differences in the resting state functional connectivity (RSFC). The obtained results showed that classification was effective and accuracy could be up to 82.69% (training) and 81.37% (test). The differences in functional connectivity between CPS and SPS were smaller in temporal and insula. The differences between the two groups were concentrated in the parietal, occipital, frontal, and limbic systems. Loss of consciousness and behavioral disturbances in patients with CPS might be caused by abnormal functional connectivity in extratemporal regions produced by post-epileptic discharges. This study not only contributed to the understanding of the cognitive-behavioral comorbidity of epilepsy but also improved the accuracy of epilepsy classification.

KEYWORDS

**temporal lobe epilepsy, resting-state functional connectivity, MEG, machine learning, classification**

## Introduction

According to the World Health Organization, about 50 million people worldwide suffered from epilepsy (Fiest et al., 2017). In recent years, epilepsy has been proposed to be a disorder of the brain network caused by hypersynchrony of neuronal activity (Zhang et al., 2011; Richardson, 2012). Epilepsy could be divided into focal, generalized onset, and unknown onset based on clinical manifestations and electroencephalogram (EEG) in line with the new International League Against Epilepsy ILAE criteria in 2017 (Scheffer et al., 2017). Depending on whether awareness was impaired, focal seizures were classified as aware or impaired awareness seizures, which are also known as "simple partial seizure (SPS)" and "complex partial seizure (CPS)," respectively (Falco-Walter et al., 2018).

Temporal lobe epilepsy (TLE), a common neurological disorder originates in the temporal lobe (de Lanerolle et al., 1989). Seizures types in TLE primarily incorporated SPS, which can cause focal motor or selective emotional or visual changes with relatively preserved consciousness, and CPS, which has more sophisticated clinical manifestations, including epigastric paresthesia, cognitive impairment, paresthesia, and automatisms (Proposal for Revised Clinical and Electroencephalographic Classification of Epileptic Proposal for Revised Clinical Electroencephalographic Classification of Epileptic Seizures., 1981; Depaulis et al., 1997). The EEG is mostly normal in SPS, and abnormal EEG changes were transient and were generally restricted peaks or paroxysmal activity in the temporal lobe regions (Inoue et al., 2000; Janszky et al., 2004).

During the past few decades, increasing evidence have linked epileptic cognitive impairment and loss of consciousness to diffuse brain network changes (Blume, 2002; Blumenfeld et al., 2009; Xu et al., 2009; Englot et al., 2010; González et al., 2019; Hermann et al., 2021). The duration of CPS invariably exceeded 30 s, and the discharge position was deeper. Meanwhile, CPS tended to spread to the brainstem or contralateral hemisphere, resulting in extensive neurological alterations (Stayman and Abou-Khalil, 2011; Hauf et al., 2013). "Network inhibition hypothesis" was a new theory of CPS proposed by Blumenfeld (Blumenfeld et al., 2004, 2009; Guye et al., 2006), was According to the theory, focal discharges in the temporal lobe interfered with brainstem–diencephalon arousal system, and then inhibited ascending reticular activation system, which indirectly brought forth impaired cortical function and loss of consciousness (Steriade, 1970; Motelow et al., 2015). Pathophysiological studies had shown a significant increase in slow-wave activity in the frontoparietal neocortex and an enhanced rate of diffusion of fast EEG activity from the medial temporal lobe to the contralateral side during CPS compared to SPS (Englot et al., 2017). In animal studies, it was also found that blood oxygenation level-dependent in the hippocampus of TLE increased, while it decreased in the cortex and thalamus (Motelow et al., 2015). This suggested that cortical and subcortical structures are involved in regulating consciousness. Studying alterations in resting-state brain functional connectivity could be conducive to explore the mechanisms underlying cognitive dysfunction in individuals with epilepsy. Memory deficiencies in individuals with CPS were associated with compensatory increases in the hippocampus. There was also evidence that executive dysfunction was relevant to a reduced resting-state functional connectivity in the frontoparietal lobe (Park et al., 2017; Ives-Deliperi and Butler, 2021; Li et al., 2022). Language dysfunction was associated with reduced functional connectivity in the frontotemporal lobe language network. EEG researches have revealed that even if the bilateral pikes in patients with SPS and CPS originated in disparate brain regions, the discharges spread to the same area, the temporal lobe base (Sirven et al., 1996). This indicated that SPS and CPS may have shared network nodes in the temporal lobe. However, the mechanisms by which CPS brain network alterations are associated with impaired consciousness and behavioral abnormalities have not been systematically studied.

Unlike other diseases, the first symptom defined the epilepsy type, consciousness turned into a "watershed," differentiated SPS from CPS. Nevertheless, a flat dichotomy may result in neglecting the other clinical symptoms (Muayqil et al., 2018). Currently, there is a lack of direct studies on brain network alters among SPS and CPS individuals. Therefore, our study hypothesized that aberrant differences in functional connectivity of CPS and SPS were vital nodes involved in the regulation of brain network consciousness and behaviors. Furthermore, seizures were primarily self-reported by the patient, and even with EEG testing, the underreporting rate was still as high as 50% (Glauser et al., 2010; Elger and Hoppe, 2018; Verdru and Van Paesschen, 2020). Especially, when epilepsy originates in deep or non-dominant regions, the initial weak signal may not be obtained on scalp EEG (Benbadis et al., 2020). Compared to EEG, the sensitivity and specificity of magnetoencephalography (MEG) for localization of minute EEG activity were higher than that of EEG (van Mierlo et al., 2014). In contrast to MRI, even though functional brain networks were widely used in MRI (Salma et al., 2019; McKavanagh et al., 2021), the time span for information processing was only milliseconds to seconds when the brain was in a resting state. Therefore, the selection of MEG with higher localization accuracy for the analysis of functional connectivity differences was more promising for detecting subtle changes in the brain (Kakisaka et al., 2013; Nissen et al., 2018; van Klink et al., 2019).

Prior to this study, we preprocessed MEG data from patients with SPS and CPS and normal controls, and performed a source level analysis. The results of the coherence analysis showed that the functional connectivity of patients with both SPS and CPS was lower than that of normal individuals in the whole brain, while there was no significant difference between the two groups. Therefore, to address the problems of low efficiency and easy

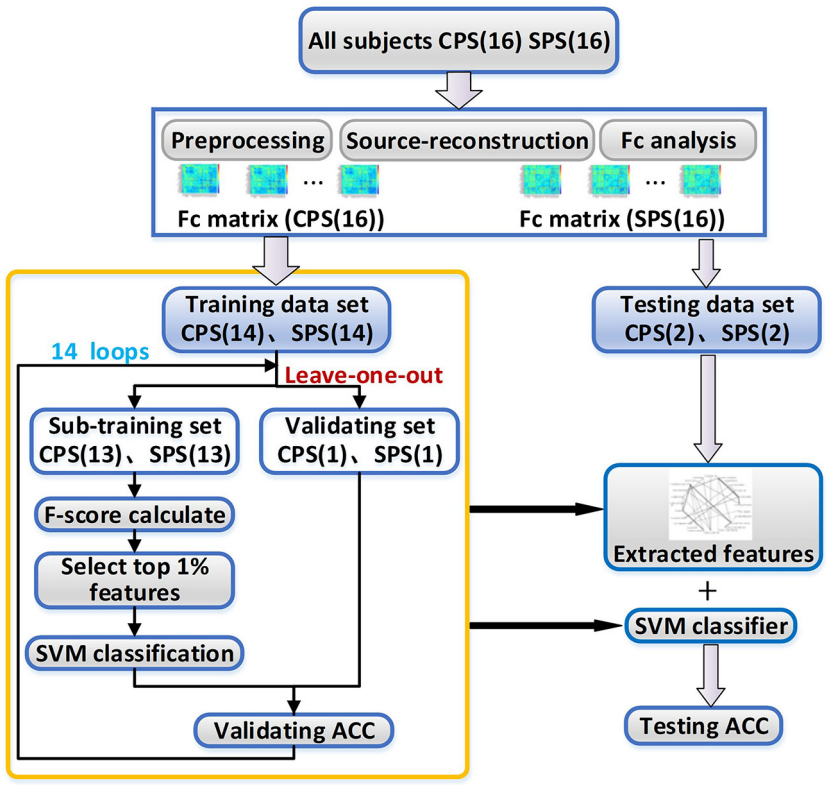



FIGURE 1
The working flowchart of the proposed framework. **CPS**, complex partial seizures; **SPS**, simple partial seizures; **SVM**, support vector machine; **ACC**, the accuracy rate.

misdiagnosis in manual identification of EEG signals, we chose machine learning to further differentiate the differences between patients with SPS and CPS (Fallahi et al., 2021). At present, machine learning has shown sound application prospects for various neurological diseases incorporating epilepsy (Craley et al., 2020; Gleichgerrcht et al., 2020; Pavel et al., 2020). (Li et al., 2018; Bharath et al., 2019). In a previous study, SVM was able to distinguish temporal lobe epilepsy from benign epilepsy in healthy controls or central temporal spikes (Jin and Chung, 2017; Sriraam and Raghu, 2017; Yang et al., 2020). In addition, the use of SVM could also distinguish between resected and unresected regions based on preoperative interictal MEG data in epileptic patients. Therefore, the present study intends to use SVM combined with two sample selection approaches to explore the differences in functional connectivity between the two groups at rest.

## Materials and methods

### Patient population

This study was part of a research program on neuroimaging in epilepsy. It consecutively recruited 40 patients with TLE between 2015 and 2020. Inclusion criteria were as follows: 1. in line with the International League Against Epilepsy (ILAE, 1989) diagnostic criteria for epilepsy; 2. partial epilepsy diagnosed by V-EEG observation and medical history; 3. the clinical and EEG characteristics were onefold SPS or CPS ; 4. individuals had no visible lesions in structural MRI images.; 5. able to cooperate with the inspection and the head movement during MEG examination was no more than 5 mm. Exclusion criteria: 1. combined with a history of generalized seizure ; 2. suffered from other types of paroxysmal illness (e.g., mental illness, severe systemic illness, etc.); 3. implants that may seriously interfere with MEG and MRI data collection (such as dentures, cochlear implants, pacemakers, etc.); 4. unable to cooperate with MEG and MRI examination. Of the 40 patients we recruited, All of them underwent MEG and MRI scans at Nanjing Brain Hospital affiliated with Nanjing Medical University. All individuals have read and signed voluntary and written informed consent for the study prior to enrollment, according to the standards set by the ethical committee of Nanjing Brain Hospital of Nanjing Medical University, which approved this study.

### MEG resting state acquisition

All studies were performed in a magnetically shielded room by using our 275-channel whole-head biomagnetometer (VSM

TABLE 1 Specific information of the 14 subsets of subjects.

| Subset | Training set | | | | | | | | | | | | | Testing set | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | |
| CPS | Su12 (102) | Su01 (73) | Su03 (78) | Su05 (81) | Su17 (33) | Su10 (34) | Su18 (37) | Su09 (40) | Su07 (72) | Su14 (84) | Su08 (117) | Su04 (125) | Su16 (144) | Su15 (169) | Su02 (75) | Su04 (42) |
| SPS | Su11 (105) | Su16 (77) | Su32 (123) | Su33 (146) | Su14 (51) | Su22 (57) | Su09 (58) | Su18 (59) | Su15 (64) | Su12 (89) | Su07 (117) | Su23 (148) | Su03 (150) | Su06 (168) | Su24 (34) | Su31 (80) |

The numbers in bracket indicating the number of FC matrices belonging to the subject. SPS, simple partial seizures. CPS, complex partial seizures.

TABLE 2 Ictal semiology in the three groups of patients,the number of patients in each group having the concerning symptoms.

| | CPS ($n = 16$) | SPS ($n = 16$) |
|---|---|---|
| Age (year) | 29.11 ± 8.63 | 23.53 ± 5.66 |
| Male (*n*, %) | 5 (31.25) | 8 (50) |
| Seizure duration (second) | 104.06±34.62 | 15.56±6.12 |
| Impairment of consciousness (*n*, %) | 16 (100) | 0 (0) |
| Oro-alimentary Automatisms (*n*, %) | 13 (81.25) | 5 (31.25) |
| Motor Automatisms (*n*, %) | 15 (93.75) | 2 (12.5) |
| Vegetative symptoms (*n*, %) | 6 (37.5) | 2 (12.5) |

MedTech, Coquitlam, BC, Canada). The full head sensor of the 275 super-conducting quantum interference device (SQUID) was used to measure the brain magnetic field in the direction perpendicular to the scalp. Three electrically active coils were placed as fiducial markers at the nasion and 1 cm anterior to the left and right tragus to measure the position of each person's head relative to the MEG sensor.

All individuals did not have seizures at least 16 h before the examination. If the frequency of seizures was less than once a week, individuals were required to stop anti-epileptic drugs for 2 days and overnight sleep deprivation (increased seizure activity). MEG recording is performed at a sample rate of 1,200 Hz for twenty 120-s recordings with the patient in the state of rest and their eyes closed to detect interictal MEG sharp waves and spikes, as well as bursts of rhythmic activity.

## Anatomical mri

MRI data were collected using a US GE Signa NV/i 1.5 T super-conducting magnetic resonance apparatus, and the head of the subject was fixed with a sponge pad. The routine anatomical MRI data were acquired to detect structural details. T1-weighted image scans were obtained, with following parameters: TR/TE = 1,750 ms/24 ms, FA = 90°, matrix = 256 × 256, FOV = 24 × 24 mm2, slice thickness = 6 mm, slice gap = 2 mm, and acquired slices = 16. Coronal T2-FLAIR-weighted image scans were also obtained, with following parameters: TR/TE = 8,400 ms/135 ms, FA = 150°, matrix = 256 × 256, FOV = 24 × 24 $mm^2$, slice thickness = 16 mm, slice gap = 2 mm and acquired slices = 16.

## Data preparation

All MEG recordings were reviewed by two experienced epileptologists at Nanjing Medical University, and the peaks of

all epileptic spikes and inter epileptic discharge were marked manually based on the MEG recordings.

### Artifact rejection and subtraction of inter-spikes relative-stable activity

All analyses were done with custom-written MATLAB (Mathworks, Natick, MA) scripts and FieldTrip (http://www.ru.nl/fcdonders/fieldtrip/). Firstly, all spikes and abnormal discharge before and after the 10 s time period were excluded. Then we short-cut the remaining segments with a non-overlapped length of 10 s and further extracted the sub-segments which are free from jump-like artifacts and muscular artifacts. To ensure the same length of the remaining sub-segments, we further removed the ones with a length $<4$ s and short-cut the others with a length of 4 s. The number of segments per subject varied from 33 to 169 at last. The total number of segments for all the individuals with CPS was 1,306, and the total number of segments for all the individuals with SPS was 1,526.

### Filtering and removing artifact

The remaining signals were low-pass-filtered at 70 Hz and high-pass-filtered at 1 Hz with notch-filtered around 50 Hz (the vertical refresh rate of the LCD projector). Furthermore, to increase the speed of the following data analysis, the data were down-sampled to 100 Hz. In addition, we removed blink artifacts and ECG noise from data by using Independent Component Analysis (ICA).

## Source reconstruction

We performed source reconstruction after preprocessing using a partial cannonical correlation/coherence (PCC) (Mukuta and Harada, 2014). Specifically, first, the subject-specific T1-weighted MR images were re-sliced and segmented to obtain brain/skull boundary by using FieldTrip. Then, we generated the individual cortical meshes with $>$ 1,00,000 vertices per hemisphere by using the Freesurfer package (version 5.3.0) (surfer.nmr.mgh.harvard.edu), and we downsampled them to 8,196 nodes in all by using MNE Suite (martinos.org/mne/stable/index.html). Next, the downsampled cortical sheet was coregistered to the sensor-based coordinate system with FieldTrip. Finally, the volume conduction for the forward model was computed by using FieldTrip's "single shell" method, and the source reconstruction was computed between 1 and 40 Hz with a partial canonical correlation/coherence (PCC) method.

## Functional connectivity analysis

For each segment time course, we calculated functional connectivity (FC) by the imaginary part of the coherency index between the time courses at every two sheets after source reconstruction. To reduce the dimensionality of the FC matrix, we applied a parcellation scheme according to the Destrieux Atlas, which consists of 74 parcels in the whole cortical surface (Destrieux et al., 2010). Therefore, each FC matrix was merged to a size of 74-by-74.

## Classification of fc matrix between cps and sps

In the next step, we tried to find the functional connectivity pairs that were different between CPS and SPS groups by using a nonlinear classification method with feature extraction and leave-one-out loop. The features which contributed well and were stable for separating the two patient groups were considered to indicate the corresponding cortical pairs having obvious distinct functional connectivity between them. The operational flowchart of the proposed framework is shown in Figure 1.

Specifically, as shown in Table 1, we divided all the subjects into 16 subsets, and each subset had one patient with CPS and one with SPS. For each subset, the number of CPS FC matrices was basically equal to the number of SPS FC matrices. We randomly selected two subsets as the testing set, and the others were considered as a training set for classification.

To improve the generalization ability of the classification model as possible as we can, we adopted a leave-one-out method for the training processing in this work. For each loop, we selected 13 subsets from the training set as sub-training set and left one subset as validating set to select the best hyper-parameters for the classification model. For the sub-training set, we selected the FC features, which have stable and great contribution for separating the CPS and SPS, as the methods introduced in "Feature extraction" section. Then, we trained SVM classifier as introduction in "Support vector machine classifier". And then, we repeated this process 14 times for all possible selection options.

After leave-one-out training, we counted the overlap rate of all appeared features among 14 times leave-one-out training loop and extracted the features with the highest ($\geq$12/14) overlap rate as a reliable feature assemblage. Finally, we extracted the corresponding feature values in the testing dataset as the input for the classifier and performed the classification as we trained. The test accuracy rate here was calculated. The method for feature extraction and classification were introduced below in detail.

## Feature extraction

To reduce the duplicate information in the FC matrix which was symmetrical, we extracted its lower triangular part and its main diagonal as the vector. Each value in the vector is a feature for classification, indicating an FC strength between a pair of cortical areas. The total number of features contained in each vector is C_74^2+74=2,775. It is important to select the most essential features using feature selection before entering the classifier (Guyon and Elissee, 2003) when the number of features is much more than the number of subjects used in classification.

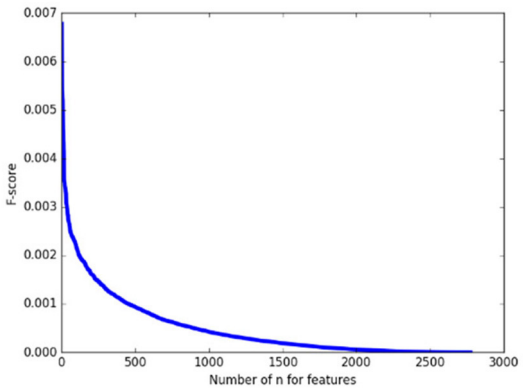



FIGURE 2
The F-score of 2,775 features.

To more effectively enhance the efficiency of computation and avoid the high-dimensional and small-sample-size problem for classification, we further measured the classification capability of each feature by the F-score method (Chen et al., 2006) and removed the features which were irrelevant or has low correlation to the classification (Guyon and Elissee, 2003). F-score is a simple and generally quite an effective technique that can measure the ability of a feature to discriminate between two classes of samples. Given a set of training samples $x_k, \mathrm{k} = 1, \cdots, m$, then the F-score of the ith feature is defined as follows:

$$F_i = \frac{\left(\bar{x}_i^{(-)} - \bar{x}_i\right)^2 + \left(\bar{x}_i^{(+)} - \bar{x}_i\right)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n_+}\left(x_{k,i}^{(+)} - \bar{x}_i^{(+)}\right)^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n_-}\left(x_{k,i}^{(-)} - \bar{x}_i^{(-)}\right)^2} \quad (1)$$

where $\bar{x}_i$ is the average of the $i^{\text{th}}$ feature in the whole data set, whereas $\bar{x}_i^{(+)}, \bar{x}_i^{(-)}$ are the average of the $i^{\text{th}}$ feature in CPS, and SPS data sets, respectively; $n_+$ and $n_-$ represent the number of samples in CPS and SPS data sets, respectively; and $x_{k,i}^{(+)}$ and $x_{k,i}^{(-)}$ indicate the $i^{\text{th}}$ feature of the $k^{\text{th}}$ sample in the two groups' data sets, respectively. F-score is positive, and the larger

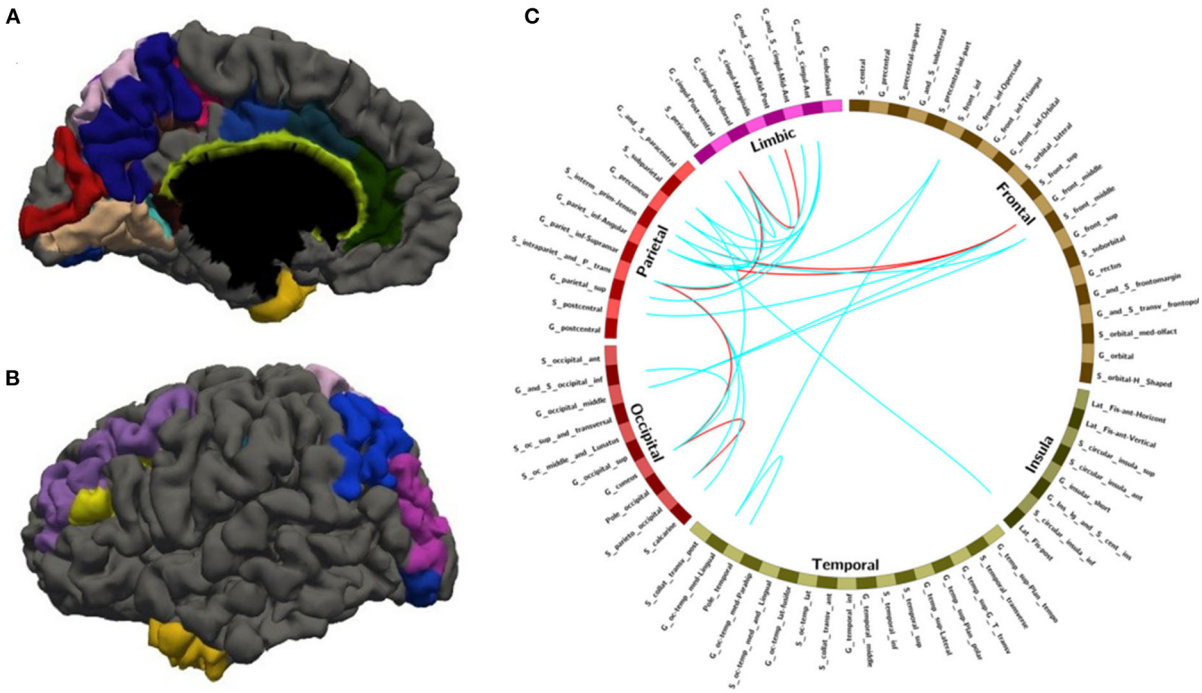



FIGURE 3
The nodes and edges involved in the FC features with high overlap rate (12/14) among training. **(A,B)** Regions involved in 28 FC features with overlapping ratios ≥12/14 across 14 training sessions. **(C)** Edges involved in 28 FC features with overlap ratio ≥12/14 in 14 training sessions.

TABLE 3 Validation accuracy for each leave-one-out training loop.

| # Subset for validation | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Acc(%) | 88.73 | 76.55 | 81.28 | 74.44 | 85.06 | 81.02 | 83.58 | 76.32 | 88.29 | 80.56 | 75.13 | 77.85 | 72.39 | 76.98 |
| Average | 79.87% | | | | | | | | | | | | | |

ACC, the accuracy rate.

TABLE 4 The nodes and edges involved in the FC features with high overlap rate (12/14) among training.

| Index | ROI 1 | ROI 2 |
|---|---|---|
| 1 | G_pariet_inf-Angular | G_front_middle |
| 2 | S_parieto_occipital | G_cuneus |
| 3 | S_interm_prim-Jensen | G_front_middle |
| 4 | S_intrapariet_and_P_trans | G_cingul-Post-ventral |
| 5 | S_intrapariet_and_P_trans | G_cuneus |
| 6 | G_cingul-Post-ventral | G_and_S_cingul-Mid-Post |
| 7 | S_pericallosal | G_pariet_inf-Angular |
| 8 | G_parietal_sup | G_and_S_cingul-Ant |
| 9 | S_front_inf | G_oc-temp_med-Lingual |
| 10 | G_occipital_middle | G_pariet_inf-Angular |
| 11 | S_intrapariet_and_P_trans | G_and_S_cingul-Ant |
| 12 | G_precuneus | G_and_S_cingul-Ant |
| 13 | G_pariet_inf-Angular | G_cingul-Post-ventral |
| 14 | S_interm_prim-Jensen | Lat_Fis-post |
| 15 | S_subparietal | G_precuneus |
| 16 | S_pericallosal | G_cingul-Post-ventral |
| 17 | S_subparietal | G_cuneus |
| 18 | G_pariet_inf-Angular | G_and_S_cingul-Mid-Ant |
| 19 | S_intrapariet_and_P_trans | S_calcarine |
| 20 | S_front_middle | G_occipital_middle |
| 21 | G_cuneus | G_and_S_occipital_inf |
| 22 | S_parieto_occipital | S_intrapariet_and_P_trans |
| 23 | S_subparietal | G_and_S_cingul-Ant |
| 24 | Pole_temporal | G_oc-temp_med-Lingual |
| 25 | S_postcentral | G_front_middle |
| 26 | S_cingul-Marginalis | G_cingul-Post-ventral |
| 27 | S_interm_prim-Jensen | S_front_middle |
| 28 | S_front_inf | G_pariet_inf-Angular |

The pairs are ranked in descending order according to the mean F-score value of each feature among 14 times training.

the F-score of a feature is the stronger its ability to distinguish two categories of samples.

As a pretest, we calculated the F-score of 2,775 features on the whole training set first and arranged them in descending order. As shown in Figure 2, it can be found that the F-score curve of 2,775 features declines rapidly and there are a large number of feature F-scores close to zero, suggesting only a very small number of features have a good contribution to classification while the others are redundant. Therefore, in the following classification processing, we extracted the features with the top 1% F–scores from the training subset as the input for classification.

TABLE 5 Validation accuracy for each leave-one-out training loop with random sample selection.

| Group | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Acc(%) | 83.25 | 81.94 | 79.27 | 78.55 | 87.21 | 84.35 | 81.13 | 84.57 | 86.88 | 82.46 | 86.12 | 81.97 | 82.48 | 84.59 |
| Average | 82.69% | | | | | | | | | | | | | |

ACC, the accuracy rate.

### Support vector machine classifier

The SVM is one type of binary supervised learning classifier that has been widely used in recent years (Ines et al., 2013) with a high ability for generalization (Vapnik, 2000). Rajpoot et al. (2015) used AP clustering combined with SVM to study functional connectivity changes in individuals with epilepsy and normal subjects. Its basic model is a linear classifier. For linearly separable data, SVM generates a separating hyperplane which separates the data with the largest margin. For linearly inseparable data, SVM can efficiently perform a non-linear one using kernel function $\Phi(x)$ (2), mapping their inputs into high-dimensional feature spaces.

$$\Phi(x) : R^n \rightarrow R^{nh} \tag{2}$$

By choosing a suitable $\Phi(x)$, the SVM constructs an optimal separating hyperplane in higher-dimensional feature space to solve the linear-inseparable problem (Cantor-Rivera et al., 2015). The kernel function mentioned above may be any of the symmetric functions that satisfies the Mercel conditions (Dhanalakshmi et al., 2009). We selected the Radial Basis Function (RBF) in this work since it was the most frequently used kernel function, and there were two parameters $C$ and $\gamma$, *which* must be preset before training the SVM classifier.

In the experiment, we implemented the scikit-learn, a machine learning tool package in python, to train the SVM classifier with parameter C of 1.2 and parameter $\gamma$ of 5, and each sample was converted to scikit-learn input data format in this work.

## Results

Five patients had obvious head movement during scanning, defined as more than a 2 mm translation. Two patients refused to participate in further research and asked to withdraw. This meant that 32 patients met the stringent inclusion criteria and were finally included in the study. Among them, 16 patients (five male, 29.1 $\pm$ 8.63 years) were suffering from CPS and 16 patients (11 male, 23.5 $\pm$ 5.66 years) from SPS. The detailed demographic and electroclinical data are summarized in Table 2.

We calculated the classification accuracies from each validating subset in the leave-one-out loop to evaluate the effectiveness of our feature extraction method and SVM classifier. The results were summarized in Table 3. For all the 14 times of training, the validation accuracies varied from 72.39 to 88.73%, of which the mean is 79.87%.

After leave-one-out training, we got 28 FC features whose overlap rate was $\geq$12/14 among 14 times training. Figure 3 and Table 4 demonstrated the nodes and edges involved in these 28 FC features. According to the mean F-score value of each feature among 14 times training, we marked the FCs with great mean F-scores (higher than 75% of the highest mean F-score value among all the features) as a red line and colored the others in blue. They were mainly located

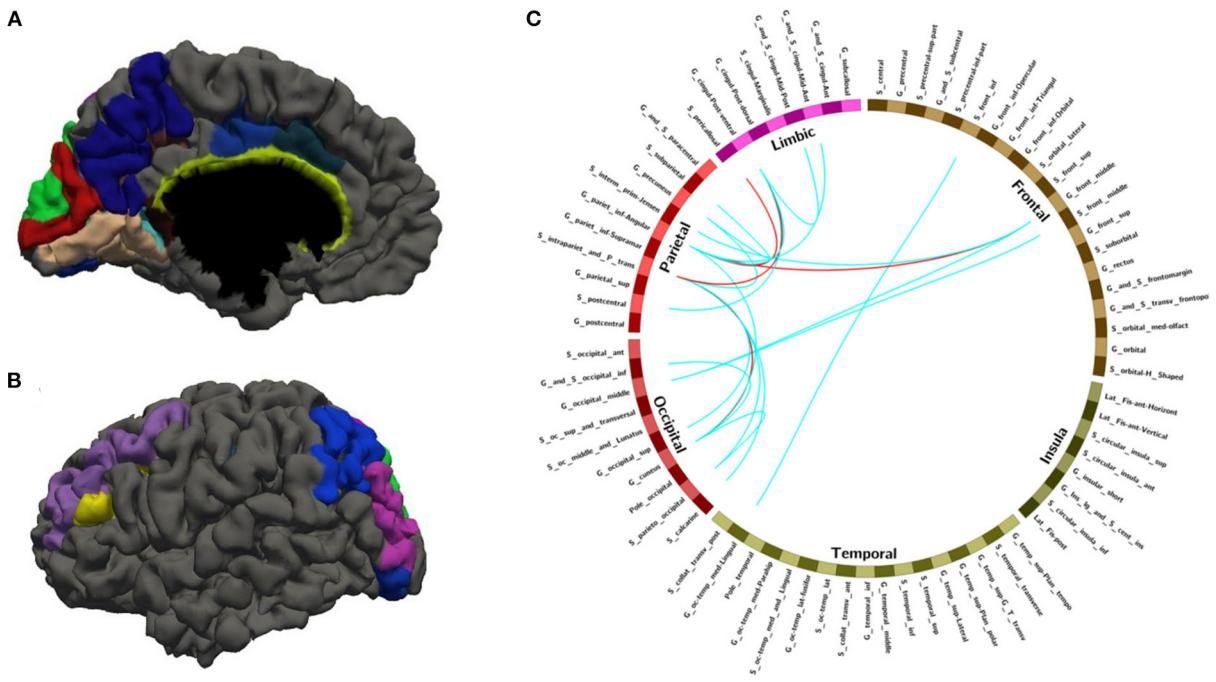



FIGURE 4
The nodes and edges involved in the FC features with a high overlap rate (12/14) among training with random sample selection. **(A,B)** Regions involved in 28 FC features with overlapping ratios ≥12/14 across 14 training sessions. **(C)** Edges involved in 28 FC features with overlap ratio ≥12/14 in 14 training sessions.

within and between the parietal, the occipital, the frontal lobe, and the limbic system. For the temporal lobe and the insula, there was less different functional connectivity between SPS and CPS individuals. Specifically, as shown in Figure 3 and Table 4, the FCs with great mean F-scores were located between the middle frontal gyrus (G_front_middle), and the angular gyrus (G_pariet_inf-Angular), and Sulcus intermedius primus (of Jensen) (S_interm_prim-Jensen); between the Cuneus (G_cuneus) and the parieto-occipital sulcus (S_parieto_occipital), and the intraparietal sulcus (S_intrapariet_and_P_trans); and between the intraparietal sulcus (S_intrapariet_and_P_trans) and the posterior-ventral part of the cingulate gyrus (vPCC, G_cingul-Post-ventral); and between the posterior-ventral part of the cingulate gyrus (vPCC, G_cingul-Post-ventral) and the middle-posterior part of the cingulate gyrus and sulcus (pMCC, G_and_S_cingul-Mid-Post).

According to these 28 FC features, we extracted the corresponding values in the testing dataset which contained 117 FC matrices for individuals with CPS and 114 FC matrices for individuals with SPS (as shown in the "Testing set" in Table 1). The test accuracy rate was 78.52%.

To further verify the stability of the results described above, we also adopted another criterion for sample selection in the training set. First, we mixed all the samples (FC matrices from all the time segments and all the subjects in the training set). For each time of the leave-one-out training procession, we randomly selected 90% of all the samples as a subset for training and the rest of 10% for validation. By calculating the F-score of each feature in the training subset, we took the features with top 1% F-score values as the input for the following SVM classifier. We also repeated this process 14 times. The validation accuracies in each leave-one-out loop are shown in Table 5, which varied from 78.55 to 87.21%, of which the mean is 82.69%. The final test accuracy rate reached upto 81.37%.

With random sample selection in leave-one-out training loop, we got 20 FC features with an overlap rate ≥12/14. Figure 4 and Table 6 demonstrated their nodes and edges. They were mainly located within and between the parietal, the occipital, and the frontal lobe, and the limbic system, similar to the results in Figure 3. For the temporal lobe and the insula, there was less different functional connectivity between SPS and CPS individuals, either. In Figure 4, we marked the FCs with great mean F-scores (higher than 75% of the highest mean F-score value among all the features) as a red line, and colored the others in blue. Specifically, the FCs with great mean F-scores were located between the intraparietal sulcus (S_intrapariet_and_P_trans) and the posterior-ventral part of the cingulate gyrus (vPCC, G_cingul-Post-ventral) and the cuneus (G_cuneus), between the angular gyrus (G_pariet_inf-Angular) and the middle frontal gyrus (G_front_middle) and pericallosal sulcus (S_pericallosal). Comparing the results obtained with the two different ways of sample selection, we can

find that there are many coincident ROIs and edges, as shown in Figure 5 and Table 6 (marked with *).

TABLE 6 The nodes and edges involved in the FC features with high overlap rate (12/14) among training with random sample selection.

| Index | ROI 1 | ROI 2 |
|---|---|---|
| 1* | S_intrapariet_and_P_trans | G_cingul-Post-ventral |
| 2* | S_intrapariet_and_P_trans | G_cuneus |
| 3* | G_pariet_inf-Angular | G_front_middle |
| 4* | S_pericallosal | G_pariet_inf-Angular |
| 5* | S_parieto_occipital | S_intrapariet_and_P_trans |
| 6* | S_front_inf | G_oc-temp_med-Lingual |
| 7 | S_intrapariet_and_P_trans | G_occipital_sup |
| 8 | G_pariet_inf-Angular | G_cuneus |
| 9* | G_pariet_inf-Angular | G_cingul-Post-ventral |
| 10* | G_occipital_middle | G_front_middle |
| 11* | S_subparietal | G_precuneus |
| 12* | G_cingul-Post-ventral | G_and_S_cingul-Mid-Post |
| 13* | S_intrapariet_and_P_trans | S_calcarine |
| 14* | S_front_middle | G_occipital_middle |
| 15* | S_parieto_occipital | G_cuneus |
| 16 | S_intrapariet_and_P_trans | G_precuneus |
| 17* | S_interm_prim-Jensen | G_front_middle |
| 18 | S_postcentral | G_cingul-Post-ventral |
| 19* | G_cuneus | G_and_S_occipital_inf |
| 20* | G_pariet_inf-Angular | G_and_S_cingul-Mid-Ant |

The pairs are ranked in descending order according to the mean F-score value of each feature among 14 times training (* after index represents the pair is overlapped in the results with two different ways of sample selection).

# Discussion

## Summary

In our pre-experiment, we found a significant decrease in functional connectivity in both groups of patients compared to healthy controls, which was consistent with previous studies (Ives-Deliperi and Butler, 2021). To date, few studies have focused on the resting state of SPS and CPS, especially the analysis of differences in brain functional connectivity between the two groups. We evaluated subjects' FC for exploring network connection markers for TLE with altered consciousness status. The synchronicity of activities between network nodes was calculated at the brain network level. A higher mean F-score value indicated a stronger activity of network nodes and a greater connectivity with other nodes. The network connection differences of CPS from SPS were distinguished by SVM. In this way, dissimilar regions in the functional connectivity of global networks in SPS and CPS were determined. The following primary results were made: (1) The SVM classification model used the optimal feature set of 28 functional connections calculated from MEG data to distinguish the CPS subjects from SPS at a mean accuracy of 81.37% (sensitivity = 81.1%; specificity = 81.54%) on test data. (2) Compared with SPS, individuals with CPS revealed a hyper-connectivity in several primary regions including intraparietal sulcus, transverse parietal sulcus of brissaud, middle frontal gyrus, callosal suleus, ventral posterior cingulate gyrus, cuneus, and inferior parietal marginal angular gyrus. By comparing the differences in FC between SPS and CPS, it was possible to explore the pathological basis of consciousness impairment and cognitive abnormalities.

## Relationship between cps network connections and consciousness

There was no significant relationship between the occurrence of impaired consciousness in CPS and the functional connectivity of the epileptic region of origin (Najm, 2018). We did find that CPS and SPS functional connectivity differences were concentrated in extratemporal lobe regions. It was mainly distributed between the parietal, occipital, frontal. and limbic systems. This implied that the occurrence of CPS was not only associated with structural damages in the temporal lobe but also with abnormal brain network connectivity in extratemporal brain regions (Englot et al., 2010). Abnormal functional connectivity in these brain regions might accelerate the outward diffusion of temporal lobe discharge (Yoo et al., 2014; Sirin et al., 2020). This may lead to individuals who showed up with the loss of consciousness (Li et al., 2020).

Our observations were consistent with the "network inhibition hypothesis." If seizures spread beyond the epileptogenic zone, consciousness may be vulnerable to impairment Bancaud et al., 1994; So, 1995; Norden and Blumenfeld, 2002; Blumenfeld et al., 2004; Englot et al., 2009, 2010. Previous studies have shown that the activation state of the frontoparietal region was associated with loss of consciousness (Untergehrer et al., 2014). CPS also had a distinct "spatial shift of slow waves", that was, after a seizure, slow waves spread from the frontal cortex to the contralateral parietal and temporal lobes. However, this phenomenon was not found in SPS individuals (Yang et al., 2012). Similar results were seen in neuroimaging. Both SPECT and fMRI detected a reduction in subcortical cerebral blood oxygen level-dependent signals in the frontoparietal region in patients with generalized epilepsy (Gotman et al., 2005; Bai et al., 2010). Similarly, individuals with other unconscious states, such as coma, anesthesia, and brain death, showed impaired functional integration of the resembled cortex (Noirhomme et al., 2010; Crone et al., 2013; Gruenbaum,

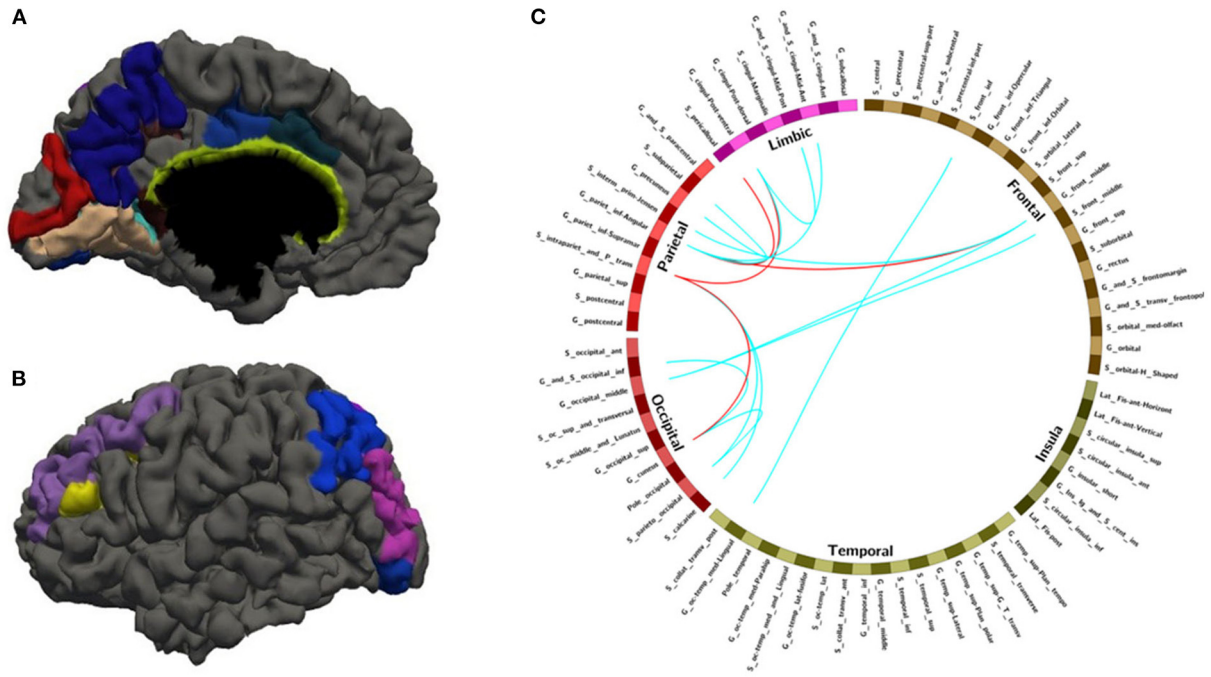



FIGURE 5
The overlapped ROIs and edges in both approaches **(A,B)** Regions involved in 28 FC features with overlapping ratios ≥12/14 across 14 training sessions. **(C)** Edges involved in 28 FC features with overlap ratio ≥12/14 in 14 training sessions. Red line: feature mean functional connections greater than the remaining 75% are marked with a red line. Blue line: no more than the remaining 75% of the feature mean functional connections are marked with blue lines.

2021). Therefore, we speculated that in TLE, the presence of abnormal functional connectivity in the frontoparietal cortex had led to impairment of consciousness. Aberrant neuronal discharge in **patients with** CPS activated significant network disturbances at a later stage, which in turn led to frontoparietal network abnormalities that clinically manifested as impaired consciousness.

## Possible causes of automatism in patients with cps

Increased epilepsy network coherence was a pathophysiology of epilepsy semiotics (Chauvel and McGonigal, 2014). Semiology depended on the interaction of epileptogenic focus and dissemination targets (Maillard et al., 2004). Automatism was considered to be one of the most common symptoms of CPS. About 75% of individuals with CPS might present with buccal and tongue movements, including smacking, swallowing, and spitting, called oral automatisms (OAAs) (Maldonado et al., 1988; Janati et al., 1990; Kramer et al., 1997).

Automatisms had been proposed to be associated with widespread cortical excitation. Several previous stimulation studies had shown that stimulation of frontal, insular, and temporal cortex or amygdala regions could induce automatisms (Maestro et al., 2008). The mechanism by which OAAs arose was the synchronous propagation of brain waves in the temporal–insular–parietal lobes, disturbing the cortical masticatory region. The abnormal cerebral cortex triggered the emergence of oral movements (Aupy et al., 2018). Similarly, individuals with preserved verbal responsiveness had lower frequencies of perfusion in ipsilateral parietal regions during interictal episodes of automatisms (Park et al., 2018). In our study, patients in the CPS group exhibited more complex clinical symptoms in addition to impaired consciousness compared to patients with SPS. There was a prominent difference in the functional connectivity in the CPS group. This further demonstrates that the clinical behavioral differences in **patients with** CPS were likely due to alterations in network connectivity between the limbic system and parietal lobes during the late stages of neuronal firing.

## Limitations

The present study has several limitations. First, since the number of features was much more than the number of subjects used in classification, we used the F-score method to measure the classification ability of each input feature so as to select the most important edges for making accurate predictions (Guyon and Elissee, 2003). However, the F-score method did not take

the mutual information between features into account (Chen et al., 2006), and it can only be used for two classifications. Once the number of classifications was more than two, the F-score method will fail. Second, an insufficient sample size of subjects was still a factor that cannot be ignored due to the limitation of the research. Due to the small sample size of individuals with simple clinical SPS or CPS, to strictly enroll the criteria, we only absorbed individuals whose timing of onset was short and the frequency of clinical seizure was low. Thus, there were only individuals who were up to standard with a single seizure form.

## Conclusions

In this research, we divided the individuals with TLE into two subtypes according to different clinical symptoms, SPS and CPS, and constructed an SVM classifier to classify the functional connectivity corresponding to these two types of individuals. Finally, a classification accuracy of 78.52% was achieved when we used the already existing SVM classifier on the testing set. To further verify the stability of the results above, we redesigned the data set allocation method and extracted the features, and finally obtained 81.37% classification accuracy. The final results showed that the nodes and edges involved in these features extracted by the above two methods had a high coincidence. Moreover, the two groups of individuals had significant differences in the connection between the parietal lobe and other brain regions. This finding may provide an idea for studying the pathogenesis of CPS or new thinking that can be used to research refractory epilepsy.

## 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 Ethics Committee of Affiliated Hospital of Nanjing University of Traditional Chinese Medicine. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

TW contributed to conceptualization, methodology, software, data curation, and writing. YW contributed to conceptualization, methodology, and supervision. ZL and YL contributed to methodology and software. YZ and YW contributed to writing the manuscript. YZ and XX contributed to conceptualization and supervision. All authors have read and agreed to the published version of the manuscript.

## Funding

This work was partially supported by the National Natural Science Foundation of China (Grant No. 82172022).

## 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

Aupy, J., Noviawaty, I., Krishnan, B., Suwankpakdee, P., Bulacio, J., Gonzalez-Martinez, J., et al. (2018). Insulo-opercular cortex generates oroalimentary automatisms in temporal seizures. *Epilepsia.* 59, 583–594. doi: 10.1111/epi.14011

Bai, X., Vestal, M., Berman, R., Negishi, M., Spann, M., Vega, C., et al. (2010). Dynamic time course of typical childhood absence seizures: EEG, behavior, and functional magnetic resonance imaging. *J. Neurosci.* 30, 5884–5893. doi: 10.1523/JNEUROSCI.5101-09.2010

Bancaud, J., Brunet-Bourgin, F., Chauvel, P., and Halgren, E. (1994). Anatomical origin of déjà vu and vivid 'memories' in human temporal lobe epilepsy. *Brain.* 117, 71–90. doi: 10.1093/brain/117.1.71

Benbadis, S. R., Beniczky, S., Bertram, E., MacIver, S., and Mosh,é, S. L. (2020). The role of EEG in patients with suspected epilepsy. *Epileptic Dis.* 22, 143–155. doi: 10.1684/epd.2020.1151

Bharath, R. D., Panda, R., Raj, J., Bhardwaj, S., Sinha, S., Chaitanya, G., et al. (2019). Machine learning identifies "rsfMRI epilepsy networks" in temporal lobe epilepsy. *Eur. Radiol.* 29, 3496–3505. doi: 10.1007/s00330-019-5997-2

Blume, W. T. (2002). Complex partial seizures. Clinical description and diagnosis. *Adv. Exp. Med. Biol.* 497, 9–18. doi: 10.1007/978-1-4615-1335-3_2

Blumenfeld, H., McNally, K. A., Vanderhill, S. D., Paige, A. L., Chung, R., Davis, K., et al. (2004). Positive and negative network correlations in temporal lobe epilepsy. *Cerebral Cortex,* 14, 892–902. doi: 10.1093/cercor/bhh048

Blumenfeld, H., Varghese, G. I., Purcaro, M. J., Motelow, J. E., Enev, M., McNally, K. A., et al. (2009). Cortical and subcortical networks in human secondarily generalized tonic-clonic seizures. *Brain.* 132, 999–1012. doi: 10.1093/brain/awp028

Cantor-Rivera, D., Khan, A. R., Goubran, M., Mirsattari, S. M., and Peters, T. M. (2015). Detection of temporal lobe epilepsy using support vector machines

in multi-parametric quantitative MR imaging. *Comput. Med. Imaging Graph.* 41, 14–28. doi: 10.1016/j.compmedimag.2014.07.002

Chauvel, P., and McGonigal, A. (2014). Emergence of semiology in epileptic seizures. *Epilepsy Behav.* 38, 94–103. doi: 10.1016/j.yebeh.2013.12.003

Chen, H., Dai, W. -L., Yang, X. -L., Gao, R., Cao, Y., Li, H., et al. (2006). Studies on the structural change of a reaction-controlled phase-transfer $\pi - C_5H_5NC_{16}H_{33}]_3\{PO_4[WO_3]_4$ catalyst during the selective oxidation of cyclopentene to glutaric acid with aqueous $H_2O_2$. *Appl. Cataly. A Gen.* 309, 62–69. doi: 10.1016/j.apcata.2006.04.037

Craley, J., Johnson, E., and Venkataraman, A. (2020). A spatio-temporal model of seizure propagation in focal epilepsy. *IEEE Trans. Med. Imaging.* 39, 1404–1418. doi: 10.1109/TMI.2019.2950252

Crone, J. S., Soddu, A., Höller, Y., Vanhaudenhuyse, A., Schurz, M., Bergmann, J., et al. (2013). Altered network properties of the fronto-parietal network and the thalamus in impaired consciousness. *NeuroImage. Clini.* 4, 240–248. doi: 10.1016/j.nicl.2013.12.005

de Lanerolle, N. C., Kim, J. H., Robbins, R. J., and Spencer, D. D. (1989). Hippocampal interneuron loss and plasticity in human temporal lobe epilepsy. *Brain Res.* 495, 387–395. doi: 10.1016/0006-8993(89)90234-5

Depaulis, A., Helfer, V., Deransart, C., and Marescaux, C. (1997). Anxiogenic-like consequences in animal models of complex partial seizures. *Neuroscience and biobehavioral reviews,* 21, 767–774. doi: 10.1016/S0149-7634(96)00060-7

Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage.* 53, 1–15. doi: 10.1016/j.neuroimage.2010.06.010

Dhanalakshmi, V., Kumar, M. A., Rekha, R. U., Kumar, C. A., and Rajendran, S. (2009). "Morphological analyzer for agglutinative languages using machine learning approaches," in *International Conference on Advances in Recent Technologies in Communication & Computing* (IEEE).

Elger, C. E., and Hoppe, C. (2018). Diagnostic challenges in epilepsy: seizure under-reporting and seizure detection. *Lancet. Neurol.* 17, 279–288. doi: 10.1016/S1474-4422(18)30038-3

Englot, D. J., D'Haese, P. F., Konrad, P. E., Jacobs, M. L., Gore, J. C., Abou-Khalil, B. W., et al. (2017). Functional connectivity disturbances of the ascending reticular activating system in temporal lobe epilepsy. *J. Neurol.* 88, 925–932. doi: 10.1136/jnnp-2017-315732

Englot, D. J., Modi, B., Mishra, A. M., DeSalvo, M., Hyder, F., and Blumenfeld, H. (2009). Cortical deactivation induced by subcortical network dysfunction in limbic seizures. *J. Neurosci.* 29, 13006–13018. doi: 10.1523/JNEUROSCI.3846-09.2009

Englot, D. J., Yang, L., Hamid, H., Danielson, N., Bai, X., Marfeo, A., et al. (2010). Impaired consciousness in temporal lobe seizures: role of cortical slow activity. *Brain.* 133, 3764–3777. doi: 10.1093/brain/awq316

Falco-Walter, J. J., Scheffer, I. E., and Fisher, R. S. (2018). The new definition and classification of seizures and epilepsy. *Epilepsy Res.* 139, 73–79. doi: 10.1016/j.eplepsyres.2017.11.015

Fallahi, A., Pooyan, M., Lotfi, N., Baniasad, F., Tapak, L., Mohammadi-Mobarakeh, N., et al. (2021). Dynamic functional connectivity in temporal lobe epilepsy: a graph theoretical and machine learning approach. *Neurol.Sci.* 42, 2379–2390. doi: 10.1007/s10072-020-04759-x

Fiest, K. M., Sauro, K. M., Wiebe, S., Patten, S. B., Kwon, C. S., Dykeman, J., et al. (2017). Prevalence and incidence of epilepsy: A systematic review and meta-analysis of international studies. *Neurology.* 88, 296–303. doi: 10.1212/WNL.0000000000003509

Glauser, T. A., Cnaan, A., Shinnar, S., Hirtz, D. G., Dlugos, D., Masur, D., et al. (2010). Ethosuximide, valproic acid, and lamotrigine in childhood absence epilepsy. *N. Engl. J. Med.* 362, 790–799. doi: 10.1056/NEJMoa0902014

Gleichgerrcht, E., Keller, S. S., Drane, D. L., Munsell, B. C., Davis, K. A., Kaestner, E., et al. (2020). Temporal lobe epilepsy surgical outcomes can be inferred based on structural connectome hubs: a machine learning study. *Ann. Neurol.* 88, 970–983. doi: 10.1002/ana.25888

González, H., Chakravorti, S., Goodale, S. E., Gupta, K., Claassen, D. O., Dawant, B., et al. (2019). Thalamic arousal network disturbances in temporal lobe epilepsy and improvement after surgery. *J. Neurol.* 90, 1109–1116. doi: 10.1136/jnnp-2019-320748

Gotman, J., Grova, C., Bagshaw, A., Kobayashi, E., Aghakhani, Y., and Dubeau, F. (2005). Generalized epileptic discharges show thalamocortical activation and suspension of the default state of the brain. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15236–15240. doi: 10.1073/pnas.0504935102

Gruenbaum, B. F. (2021). Comparison of anaesthetic- and seizure-induced states of unconsciousness: a narrative review. *Br. J. Anaesth.* 126, 219–229. doi: 10.1016/j.bja.2020.07.056

Guye, M., Régis, J., Tamura, M., Wendling, F., McGonigal, A., Chauvel, P., et al. (2006). The role of corticothalamic coupling in human temporal lobe epilepsy. *Brain.* 129, 1917–1928. doi: 10.1093/brain/awl151

Guyon, I., and Elissee, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3, 1157–1182. Available online at: https://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf

Hauf, M., Wiest, R., Schindler, K., Jann, K., Dierks, T., Strik, W., et al. (2013). Common mechanisms of auditory hallucinations-perfusion studies in epilepsy. *Psychiat. Res.* 211, 268–270. doi: 10.1016/j.pscychresns.2012.06.007

Hermann, B. P., Struck, A. F., Busch, R. M., Reyes, A., Kaestner, E., and McDonald, C. R. (2021). Neurobehavioural comorbidities of epilepsy: towards a network-based precision taxonomy. *Neurology.* 17, 731–746. doi: 10.1038/s41582-021-00555-z

Ines, H., Slim, Y., and Noureddine, E. (2013). "EEG classification using support vector machine," in *International Multi-conference on Systems* (IEEE).

Inoue, Y., Mihara, T., Matsuda, K., Tottori, T., Otsubo, T., and Yagi, K. (2000). Absence of simple partial seizure in temporal lobe epilepsy: its diagnostic and prognostic significance. *Epilepsy Res.* 38, 133–138. doi: 10.1016/S0920-1211(99)00081-9

Ives-Deliperi, V., and Butler, J. T. (2021). Mechanisms of cognitive impairment in temporal lobe epilepsy: A systematic review of resting-state functional connectivity studies. *Epilepsy Behavio.* 115, 107686. doi: 10.1016/j.yebeh.2020.107686

Janati, A., Nowack, W. J., Dorsey, S., and Chesser, M. Z. (1990). Correlative study of interictal electroencephalogram and aura in complex partial seizures. *Epilepsia.* 31, 41–46. doi: 10.1111/j.1528-1157.1990.tb05358.x

Janszky, J., Schulz, R., and Ebner, A. (2004). Simple partial seizures (isolated auras) in medial temporal lobe epilepsy. *Seizure.* 13, 247–249. doi: 10.1016/S1059-1311(03)00192-4

Jin, S. H., and Chung, C. K. (2017). Electrophysiological resting-state biomarker for diagnosing mesial temporal lobe epilepsy with hippocampal sclerosis. *Epilepsy Res.* 129, 138–145. doi: 10.1016/j.eplepsyres.2016.11.018

Kakisaka, Y., Gupta, A., Enatsu, R., Wang, Z. I., Alexopoulos, A. V., Mosher, J. C., et al. (2013). Magnetoencephalography reveals a unique neurophysiological profile of focal-onset epileptic spasms. *Tohoku J. Exp. Med.* 229, 147–151. doi: 10.1620/tjem.229.147

Kramer, U., Riviello, J. J. Jr, Carmant, L., Black, P. M., Madsen, J., and Holmes, G. L. (1997). Clinical characteristics of complex partial seizures: a temporal versus a frontal lobe onset. *Seizure.* 6, 57–61. doi: 10.1016/S1059-1311(97)80054-4

Li, R., Hu, C., Wang, L., Liu, D., Liu, D., Liao, W., et al. (2020). Disruption of functional connectivity among subcortical arousal system and cortical networks in temporal lobe epilepsy. *Brain Imaging Behav.* 14, 762–771. doi: 10.1007/s11682-018-0014-y

Li, Y., Cui, W., Luo, M., Li, K., and Wang, L. (2018). Epileptic seizure detection based on time-frequency images of EEG signals using gaussian mixture model and gray level co-occurrence matrix features. *Int. J. Neural Syst.* 28, 1850003. doi: 10.1142/S012906571850003X

Li, Z., Jiang, C., Xiang, W., Qi, Z., Gao, Q., Peng, K., et al. (2022). Brain structural connectivity sub typing in unilateral temporal lobe epilepsy. *Brain Imaging Behav.* 10.1007/s11682-022-00691-0. doi: 10.1007/s11682-022-00691-0

Maestro, I., Carreño, M., Donaire, A., Rumi,à, J., Conesa, G., Bargall,ó, N., et al. (2008). Oroalimentary automatisms induced by electrical stimulation of the fronto-opercular cortex in a patient without automotor seizures. *Epilepsy Behavior.* 13, 410–412. doi: 10.1016/j.yebeh.2008.03.013

Maillard, L., Vignal, J. P., Gavaret, M., Guye, M., Biraben, A., McGonigal, A., et al. (2004). Semiologic and electrophysiologic correlations in temporal lobe seizure subtypes. *Epilepsia.* 45, 1590–1599. doi: 10.1111/j.0013-9580.2004.09704.x

Maldonado, H. M., Delgado-Escueta, A. V., Walsh, G. O., Swartz, B. E., and Rand, R. W. (1988). Complex partial seizures of hippocampal and amygdalar origin. *Epilepsia.* 29, 420–433. doi: 10.1111/j.1528-1157.1988.tb03741.x

McKavanagh, A., Kreilkamp, B., Chen, Y., Denby, C., Bracewell, M., Das, K., et al. (2021). Altered structural brain networks in refractory and nonrefractory idiopathic generalized epilepsy. *Brain Connect.* 10.1089/brain.2021.0035. doi: 10.1089/brain.2021.0035

Motelow, J. E., Li, W., Zhan, Q., Mishra, A. M., Sachdev, R. N., Liu, G., et al. (2015). Decreased subcortical cholinergic arousal in focal seizures. *Neuron.* 85, 561–572. doi: 10.1016/j.neuron.2014.12.058

Muayqil, T. A., Alanazy, M. H., Almalak, H. M., Alsalman, H. K., Abdulfattah, F. W., Aldraihem, A. I., et al. (2018). Accuracy of seizure semiology obtained from first-time seizure witnesses. *BMC Neurol.* 18, 135. doi: 10.1186/s12883-018-1137-x

Mukuta, Y., and Harada, T. (2014). "Probabilistic partial canonical correlation analysis," in *Proceedings of the 31st International Conference on Machine Learning*, Vol. 32 (PMLR), 1449–1457.

Najm, I. M. (2018). Mapping brain networks in patients with focal epilepsy. *The Lancet.* 17, 295–297. doi: 10.1016/S1474-4422(18)30090-5

Nissen, I. A., Stam, C. J., van Straaten, E., Wottschel, V., Reijneveld, J. C., Baayen, J. C., et al. (2018). Localization of the epileptogenic zone using interictal MEG and machine learning in a large cohort of drug-resistant epilepsy patients. *Front. Neurol.* 9, 647. doi: 10.3389/fneur.2018.00647

Noirhomme, Q., Soddu, A., Lehembre, R., Vanhaudenhuyse, A., Boveroux, P., Boly, M., et al. (2010). Brain connectivity in pathological and pharmacological coma. *Front. Neurol.* 4, 160. doi: 10.3389/fnsys.2010.00160

Norden, A. D., and Blumenfeld, H. (2002). The role of subcortical structures in human epilepsy. *Epilepsy Behav.* 3, 219–231. doi: 10.1016/S1525-5050(02)00029-X

Park, C. H., Choi, Y. S., Jung, A. R., Chung, H. K., Kim, H. J., Yoo, J. H., et al. (2017). Seizure control and memory impairment are related to disrupted brain functional integration in temporal lobe epilepsy. *J. Neuropsychiatry Clin. Neurosci.* 29, 343–350. doi: 10.1176/appi.neuropsych.16100216

Park, H. R., Seong, M. J., Shon, Y. M., Joo, E. Y., Seo, D. W., and Hong, S. B. (2018). SPECT perfusion changes during ictal automatisms with preserved responsiveness in patients with right temporal lobe epilepsy. *Epilepsy Behav.* 80, 11–14. doi: 10.1016/j.yebeh.2017.12.030

Pavel, A. M., Rennie, J. M., de Vries, L. S., Blennow, M., Foran, A., Shah, D. K., et al. (2020). A machine-learning algorithm for neonatal seizure recognition: a multicentre, randomised, controlled trial. *The Lancet.* 4, 740–749. doi: 10.1016/S2352-4642(20)30239-X

Proposal for Revised Clinical and Electroencephalographic Classification of Epileptic Seizures. (1981). From the commission on classification and terminology of the International League Against Epilepsy. *Epilepsia.* 22, 489–501. doi: 10.1111/j.1528-1157.1981.tb06159.x

Rajpoot, K., Riaz, A., Majeed, W., and Rajpoot, N. (2015). Functional connectivity alterations in epilepsy from resting-state functional MRI. *PloS ONE.* 10, e0134944. doi: 10.1371/journal.pone.0134944

Richardson, M. P. (2012). Large scale brain models of epilepsy: dynamics meets connectomics. *J. Neurol. Neurosurg. Psychiatry.* 83, 1238–1248. doi: 10.1136/jnnp-2011-301944

Salma, Z., Hanen, H. K., Salma, S., Olfa, H., Nouha, F., Mariem, D., et al. (2019). Headaches and their relationships to epileptic seizures. *Epilepsy Behav.* 90, 233–237. doi: 10.1016/j.yebeh.2018.11.012

Scheffer, I. E., Berkovic, S., Capovilla, G., Connolly, M. B., French, J., Guilhoto, L., et al. (2017). ILAE classification of the epilepsies: position paper of the ILAE Commission for Classification and Terminology. *Epilepsia.* 58, 512–521. doi: 10.1111/epi.13709

Sirin, N. G., Kurt, E., Ulasoglu-Yildiz, C., Kicik, A., Bayram, A., Karaaslan, Z., et al. (2020). Functional connectivity analysis of patients with temporal lobe epilepsy displaying different ictal propagation patterns. *Epileptic Dis.* 22, 623–632. doi: 10.1684/epd.2020.1210

Sirven, J. I., Sperling, M. R., French, J. A., and O'Connor, M. J. (1996). Significance of simple partial seizures in temporal lobe epilepsy. *Epilepsia.* 37, 450–454. doi: 10.1111/j.1528-1157.1996.tb00590.x

So, E. L. (1995). Classifications and epidemiologic considerations of epileptic seizures and epilepsy. *Neuroimaging Clin. N. Am.* 5, 513–526.

Sriraam, N., and Raghu, S. (2017). Classification of focal and non focal epileptic seizures using multi-features and SVM classifier. *J. Medical Syst.* 41, 160. doi: 10.1007/s10916-017-0800-x

Stayman, A., and Abou-Khalil, B. (2011). FDG-PET in the diagnosis of complex partial status epilepticus originating from the frontal lobe. *Epilepsy Behav.* 20, 721–724. doi: 10.1016/j.yebeh.2011.02.001

Steriade, M. (1970). Ascending control of thalamic and cortical responsiveness. *Int. Rev. Neurobiol.* 12, 87–144. doi: 10.1016/S0074-7742(08)60059-8

Untergehrer, G., Jordan, D., Kochs, E. F., Ilg, R., and Schneider, G. (2014). Fronto-parietal connectivity is a non-static phenomenon with characteristic changes during unconsciousness. *PloS ONE.* 9, e87498. doi: 10.1371/journal.pone.0087498

van Klink, N., Mooij, A., Huiskamp, G., Ferrier, C., Braun, K., Hillebrand, A., et al. (2019). Simultaneous MEG and EEG to detect ripples in people with focal epilepsy. *Clin Neurophysiol.* 130, 1175–1183. doi: 10.1016/j.clinph.2019.01.027

van Mierlo, P., Papadopoulou, M., Carrette, E., Boon, P., Vandenberghe, S., Vonck, K., et al. (2014). Functional brain connectivity from EEG in epilepsy: seizure prediction and epileptogenic focus localization. *Prog. Neurobiol.* 121, 19–35. doi: 10.1016/j.pneurobio.2014.06.004

Verdru, J., and Van Paesschen, W. (2020). Wearable seizure detection devices in refractory epilepsy. *Acta Neurologica Belgica.* 120, 1271–1281. doi: 10.1007/s13760-020-01417-z

Xu, Q., Zhou, H., Wang, Y., and Huang, J. (2009). Fuzzy support vector machine for classification of EEG signals using wavelet-based features. *Med. Eng. Phys.* 31, 858–865. doi: 10.1016/j.medengphy.2009.04.005

Yang, L., Worrell, G. A., Nelson, C., Brinkmann, B., and He, B. (2012). Spectral and spatial shifts of post-ictal slow waves in temporal lobe seizures. *Brain.* 135, 3134–3143. doi: 10.1093/brain/aws221

Yang, W., Joo, M., Kim, Y., Kim, S. H., and Chung, J. M. (2020). "Hybrid machine learning scheme for classification of bects and TLE patients using EEG brain signals," in *IEEE Access* (IEEE).

Yoo, J. Y., Farooque, P., Chen, W. C., Youngblood, M. W., Zaveri, H. P., Gerrard, J. L., et al. (2014). Ictal spread of medial temporal lobe seizures with and without secondary generalization: an intracranial electroencephalography analysis. *Epilepsia.* 55, 289–295. doi: 10.1111/epi.12505

Zhang, Z., Liao, W., Chen, H., Mantini, D., Ding, J. R., Xu, Q., et al. (2011). Altered functional-structural coupling of large-scale brain networks in idiopathic generalized epilepsy. *Brain.* 134, 2912–2928. doi: 10.1093/brain/awr223

frontiers | Frontiers in Neuroinformatics

TYPE Original Research
PUBLISHED 02 September 2022
DOI 10.3389/fninf.2022.961089



OPEN ACCESS

EDITED BY
Jiayang Guo,
Xiamen University, China

REVIEWED BY
Kun Wang,
Tianjin University, China
Minpeng Xu,
Tianjin University, China
Gan Huang,
Université Catholique de Louvain, Belgium

*CORRESPONDENCE
Omar Mendoza-Montoya
omendoza83@tec.mx

RECEIVED 03 July 2022
ACCEPTED 03 August 2022
PUBLISHED 02 September 2022

CITATION
Triana-Guzman N, Orjuela-Cañon AD, Jutinico AL, Mendoza-Montoya O and Antelis JM (2022) Decoding EEG rhythms offline and online during motor imagery for standing and sitting based on a brain-computer interface. *Front. Neuroinform.* 16:961089. doi: 10.3389/fninf.2022.961089



# Decoding EEG rhythms offline and online during motor imagery for standing and sitting based on a brain-computer interface

Nayid Triana-Guzman[1], Alvaro D. Orjuela-Cañon[2], Andres L. Jutinico[3], Omar Mendoza-Montoya[4]* and Javier M. Antelis[4]

[1]Doctorado en Ciencia Aplicada, Universidad Antonio Nariño, Bogota, Colombia, [2]School of Medicine and Health Sciences, Universidad del Rosario, Bogota, Colombia, [3]Facultad de Ingeniería Mecánica, Electrónica y Biomédica, Universidad Antonio Nariño, Bogota, Colombia, [4]Tecnologico de Monterrey, Escuela de Ingeniería y Ciencias, Monterrey, Mexico

Motor imagery (MI)-based brain-computer interface (BCI) systems have shown promising advances for lower limb motor rehabilitation. The purpose of this study was to develop an MI-based BCI for the actions of standing and sitting. Thirty-two healthy subjects participated in the study using 17 active EEG electrodes. We used a combination of the filter bank common spatial pattern (FBCSP) method and the regularized linear discriminant analysis (RLDA) technique for decoding EEG rhythms offline and online during motor imagery for standing and sitting. The offline analysis indicated the classification of motor imagery and idle state provided a mean accuracy of 88.51 ± 1.43% and 85.29 ± 1.83% for the sit-to-stand and stand-to-sit transitions, respectively. The mean accuracies of the sit-to-stand and stand-to-sit online experiments were 94.69 ± 1.29% and 96.56 ± 0.83%, respectively. From these results, we believe that the MI-based BCI may be useful to future brain-controlled standing systems.

KEYWORDS

**brain-computer interface (BCI), electroencephalogram (EEG), motor imagery (MI), sit-stand, filter bank common spatial pattern (FBCSP), regularized linear discriminant analysis (RLDA), online BCI**

## 1. Introduction

A brain-computer interface (BCI) system provides a communication channel between the brain and an external device. These systems have been developed for decades and the choice of the BCI paradigm depends on the application (Lotte et al., 2018b). Among the possible strategies reported in the literature, the most successful noninvasive BCI paradigms are based on three main approaches: evoked response (P300), steady-state visually evoked potential (SSVEP), and motor imagery (MI) (Lee et al., 2019). However, research on electroencephalogram (EEG) based MI of lower limb movements toward BCI-controlled applications remains relatively scarce

(Bobrova et al., 2020; Asanza et al., 2022). Many of these studies have only been tested in offline scenarios due to the complexity of the movements and experimental setups that produce unrealistic EEG signals when compared to experimental setups in online scenarios (Rodríguez-Ugarte et al., 2017). Only a small number of studies have been conducted on standing and sitting behaviors in offline scenarios (Zhou et al., 2007; Bulea et al., 2014; Singh et al., 2017; Chaisaen et al., 2020). Therefore, we examined the use of an EEG-based BCI to decode offline and online MI information for these types of movements.

In recent decades, a wide variety of methods have been developed to decode motor imagery tasks from EEG signals in order to improve the performance of BCI systems (George et al., 2021; Singh et al., 2021). These methods include feature extraction techniques that use temporal (Rodríguez-Bermúdez and García-Laencina, 2012; Hamedi et al., 2014; Kee et al., 2017; Samuel et al., 2017), spectral (Al-Fahoum and Al-Fraihat, 2014; Oikonomou et al., 2017), and time-frequency representations (Kevric and Subasi, 2017; Gao et al., 2018; Aggarwal and Chugh, 2019; Padfield et al., 2019; Ortiz et al., 2020). Nevertheless, the usefulness of spatial filtering techniques in BCI applications has been explored for many years now, as a way to select the most discriminative features in EEG recordings for motor imagery tasks, as well as to reduce the huge dimensionality that can be present in feature spaces (Ang et al., 2012; Congedo et al., 2017; Lotte et al., 2018a; Rejer and Górski, 2018). In this sense, the common spatial pattern (CSP) method has been shown to extract discriminative information more effectively than other spatial filters such as bipolar, Laplacian, or common average reference, as well as unsupervised data-driven techniques such as independent component analysis (ICA) (Naeem et al., 2009; Ortner et al., 2015; He and Wu, 2018). While research already exists concerning the CSP method and was successfully applied (Chaisaen et al., 2020), important knowledge is still missing regarding the challenge of decoding EEG rhythms online during motor imagery tasks for standing and sitting.

Some researchers use the term "BCI illiteracy" for people unable to control a BCI (Allison and Neuper, 2010; Ahn et al., 2013). Nevertheless, the effective control threshold depends on many factors, including the BCI application and paradigm (Edlinger et al., 2015; Lee et al., 2019). For example, an accuracy level of less than 80% might be insufficient in a BCI system designed for communication. Conversely, for BCI systems intended for motor rehabilitation purposes, an accuracy above the confidence level might become sufficient (Thompson, 2018). Previous studies have focused on decoding EEG signals of left-hand and right-hand motor imagery tasks (which represented sitting down and standing up) (Noda et al., 2012; Wang et al., 2018), or SSVEP signals (in which flickering lights corresponded to the command for standing and sitting) (Kwak et al., 2017), instead of investigating the decoding of continuous EEG rhythms during motor imagery concerning standing and sitting. For these reasons, considering the applications of lower limb motor imagery and the necessity of using an online BCI for these little-researched movements, this study establishes an offline and online performance analysis of an EEG-based BCI during motor imagery tasks for standing and sitting.

In the present study, we investigated whether people could control an EEG-based BCI using motor imagery for standing and sitting movements. For this purpose, we explored two different classification scenarios: offline and online. The goal of the offline scenario was to obtain individual training sets for each participant in offline experiments to adjust and evaluate the machine learning models of the BCI (one model for sit-to-stand, one model for stand-to-sit). After training the interface, the online scenario aimed to measure the speed and accuracy of the BCI to decode EEG rhythms in real time during motor imagery tasks for standing and sitting. To our knowledge, the proposed EEG-based BCI is the first one to recognize motor imagery tasks online for standing and sitting, which is crucial for implementing brain-controlled standing technology. The filter bank common spatial pattern (FBCSP) method was used for feature extraction based on the modulation of theta wave (4–8 Hz) and sensorimotor rhythm (SMR), which includes two bands in the spectrum: alpha (8–12 Hz) and beta (12–30 Hz), which are associated with movement-related tasks in physical activity execution, motor planning, intention to move, and motor imagery (Yuan and He, 2014). To make the classification as fast and simple as possible, the regularized version of the linear discriminant analysis (RLDA) approach was used.

## 2. Materials and methods

### 2.1. Participants

The study involved 32 healthy subjects aged 19–29 years (16 women and 16 men). The mean (± standard deviation) age of the participants was 22.4 (± 2.3) years. None of the participants reported a history of neurological, musculoskeletal, or other disorders, and all had normal or corrected-to-normal vision. All participants were undergraduate students, with no academic relationship to the experimenters, and none had previous experience with EEG or BCI experiments. Before starting their experimental session, participants were duly informed of the nature of the study and instructed on the correct execution of the experiments. In addition, participants voluntarily signed an informed consent form in accordance with the experimental protocol approved by the ethics committee of the Universidad Antonio Nariño. This experimental protocol followed the standards of the Declaration of Helsinki (Association, 2013). Each subject was paid for their participation at the end of their session.

## 2.2. Electroencephalographic data recording

EEG data were obtained from 17 active wet electrodes (g.LADYbird) mounted on a g.Nautilus PRO biopotential amplification system (g.tec medical engineering GmbH, Austria) with wireless data transmission technology (see Figure 1). Electrodes were moistened with conductive gel and placed according to the international 10–20 system at the following positions around the primary motor cortex (Xu et al., 2017): *F*3, *Fz*, *F*4, *FC*5, *FC*1, *FC*2, *FC*6, *C*3, *Cz*, *C*4, *CP*5, *CP*1, *CP*2, *CP*6, *P*3, *Pz* and *P*4, with the ground (GND) electrode placed at *AFz* and the reference (REF) electrode placed in the right earlobe. EEG signals were acquired at a sampling rate of 250 Hz and digitally band-pass filtered with cutoff frequencies from 0.01 Hz to 60 Hz, using 6*th* order Butterworth filter at each electrode (Podder et al., 2014). Before starting the EEG recording, the impedance of the electrodes was verified to be below 30 kΩ using the impedance measurement tool provided by the manufacturer of the g.Nautilus PRO. Additionally, an in-house software platform developed in C++ was used to manage and control the execution of the experiment, collect EEG signals, store the data, and process them both offline and online (Copyright @ 2018 Instituto Tecnologico y de Estudios Superiores de Monterrey).

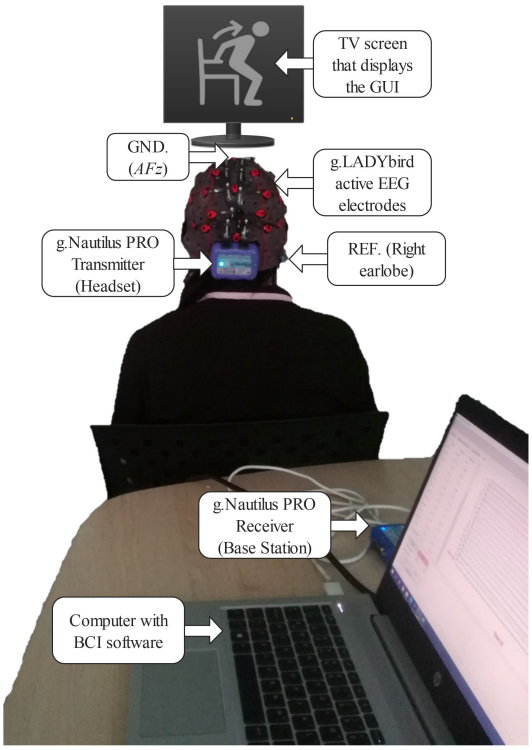



FIGURE 1
Experimental setup with a participant in front of the TV screen with the graphical user interface (GUI) to present visual cues regarding each step in the sequence of a trial. The participant is equipped with a g.Nautilus PRO system with 17 gel-based active electrodes (g.LADYbird technology) to acquire the electroencephalogram (EEG) and transmit it wirelessly from the headset (transmitter) to the receiver (base station) connected to the computer with the BCI software. REF, reference electrode; GND, ground electrode at *AFz*.

## 2.3. Experimental design

The experiments were conducted in an acoustically isolated room where only the participant and the experimenter were present. The participant was seated in a chair in a posture that was comfortable for him/her but did not affect data collection. In front of the participant, a 40-inch TV screen was placed at about 3 m, as shown in Figure 1. On this screen, a graphical user interface (GUI) displayed images that guided the participant through the experiment. Each experimental session was divided into two phases: an offline phase and an online phase.

### 2.3.1. Offline phase

The offline experiments consisted of recording participants' EEG signals during motor imagery trials for standing and sitting that were guided by the GUI presented on the TV screen (see Figure 2). Just before starting the recording of the EEG signals, the participants practiced the sequences of mental tasks that were indicated by the GUI on the TV screen. Once the recording of the EEG signals started, six offline runs were conducted in which the participants were standing in three runs and sitting in the other three runs. The participant could choose the order of the runs, and between each run, there was a break of a few minutes for the participant to avoid fatigue and boredom, recover, and prepare to continue with the recording of the next run.

In each run, the participant had to repeat a block of 30 trials of mental tasks indicated by visual cues continuously presented on the screen in a pseudo-random sequence. The temporal sequence of mental tasks performed by each participant is shown in Figure 2. Each sequence of a trial consisted of four steps:

1. Fixation: As a first step, a cross symbol appeared on the TV screen for 4 s during which the participant was asked to avoid any body movement or effort and to stay focused while looking at the symbol.
2. Action observation: In the second step, a figure appeared on the TV screen for 3 s, which the participant had to observe and perform one experimental task subsequently in the third step.
3. Imagining: In the third step, the participant had to visualize the action indicated by the figure shown in

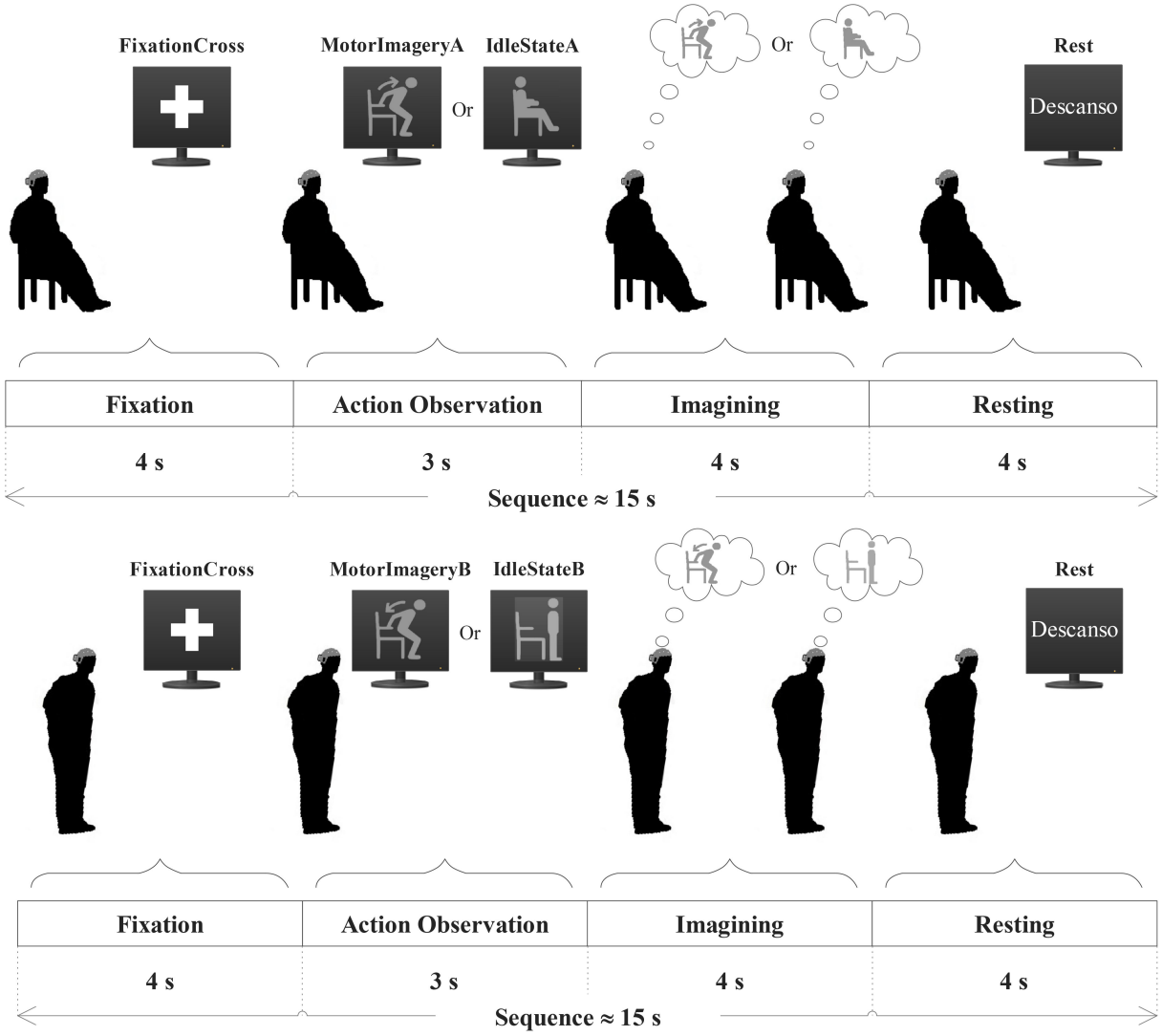



FIGURE 2
Illustration of the temporal sequence of a trial performed by the participant in the sit-to-stand experiment **(Top)** and in the stand-to-sit experiment **(Bottom)**. Each sequence consisted of four steps: fixation, action observation, imagining, and resting.

the second step and perform one experimental task for 4 s in response to the figure. For instance, the task could be sitting motionless while actively imagining the sit-to-stand movement (labeled as MotorImageryA), sitting motionless without imagining the sit-to-stand movement (labeled as IdleStateA), standing motionless while actively imagining the stand-to-sit movement (labeled as MotorImageryB), or standing motionless without imagining the stand-to-sit movement (labeled as IdleStateB).

4. Resting: Finally, in the fourth step of the sequence, the text "Descanso" (the Spanish word for "Rest") appeared on the TV screen for 4 s, instructing the participant to rest from the experimental task, blink, or move the head and body if necessary.

Participants were asked to avoid or minimize muscle effort and blinking from the first step to the third step of each sequence. For each participant, an offline experimental session was conducted for the construction of two datasets: (*A*) Sit-to-stand and (*B*) Stand-to-sit. The participant's EEG data were collected from 90 sequences of dataset *A* (45 trials of MotorImageryA tasks and 45 trials of IdleStateA tasks) and 90 sequences of dataset *B* (45 trials of MotorImageryB tasks and 45 trials of IdleStateB tasks). In total, the time duration of the offline experimental session was at least 1 h for the collection of the two EEG datasets.

#### 2.3.1.1. EEG data preprocessing

The two EEG datasets recorded from the offline experiments were independently subjected to the following automatic preprocessing pipeline (Mendoza-Montoya, 2017). First, the EEG data were split into epochs (2-s data segments of contiguous sample points of each trial) from 1 s before to 1 s after beginning the experimental task of step 3 in the sequence of a trial. Second, each epoch was separated and labeled according to one of the four experimental tasks: MotorImageryA, IdleStateA, MotorImageryB, or IdleStateB, and $E_k \subseteq \{1, 2, \ldots, 45\}$ was the subset of indices of the epochs that belong to the task $k \in \{\text{MotorImageryA, IdleStateA, MotorImageryB, IdleStateB}\}$. For each participant, the total number of epochs for each pair of experimental tasks was $n_{\text{epochs}} = 90$. Third, each EEG epoch $X = [x_e(t)]_{n_e \times n_t}$ of $n_e$ electrodes (or 17 EEG channels) and $n_t$ sample points (500 sample points per EEG epoch) was filtered using digital finite impulse response (FIR) filters with cut-off frequencies between 20–40 and 4–40 Hz. The result of this filtering step was the signal $X^{20-40}(t) = [x_e^{20-40}(t)] \in \mathbb{R}^{n_e \times n_t}$ to detect muscle artifacts and the signal $X^{4-40}(t) = [x_e^{4-40}(t)] \in \mathbb{R}^{n_e \times n_t}$ to encompass the motor-related frequency bands of the oscillatory EEG activity. Then, the peak-to-peak voltage $V_e^{\text{PP}}$, the standard deviation $\sigma_e$, and the normalized power $P_e^{\text{norm}}$ of each channel were calculated as below:

$$V_e^{\text{PP}} = max_t(x_e^{4-40}(t)) - min_t(x_e^{4-40}(t)), \tag{1}$$

$$\sigma_e = \sqrt{\frac{1}{n_t - 1} \sum_{t=1}^{n_t} (x_e^{4-40}(t) - \mu_e)^2}, \tag{2}$$

$$P_e^{\text{norm}} = \frac{\sum_{t=1}^{n_t} (x_e^{20-40}(t))^2}{\sum_{t=1}^{n_t} (x_e^{4-40}(t))^2}, \tag{3}$$

where

$$\mu_e = \frac{1}{n_t} \sum_{t=1}^{n_t} x_e^{4-40}(t). \tag{4}$$

The processed data for each experimental task from each participant contained a collection of epochs × time points × channels (45 × 500 × 17). The following exclusion criteria were applied to identify and discard noisy epochs: (*i*) Maximum peak-to-peak value $V_e^{\text{PP}}$ greater than 200 $\mu$V; (*ii*) Standard deviation amplitude $\sigma_e$ greater than 50 $\mu$V; and (*iii*) Noise to signal ratio $P_e^{\text{norm}}$ greater than 0.7. These criteria may indicate if the subject is blinking, the amplifier is saturated, the electrodes are not making good contact with the scalp, or there are some muscle artifacts, as suggested in Mendoza-Montoya (2017), Delijorge et al. (2020), and Hernandez-Rojas et al. (2022). Finally, any epoch where at least one electrode met these criteria was visually inspected to rule out noise-contaminated trials (as a double check) and labeled as an "artifact" manually. The trials with epochs labeled as "artifacts" were discarded and were not used in the subsequent analysis. Conversely, the epochs below the threshold levels passed validation and were used to investigate spatially discriminative EEG features with the filter bank common spatial pattern (FBCSP) method.

#### 2.3.1.2. EEG signals analysis

Time-frequency analysis (TFA) of EEG time series is a suitable technique to study cognitive events, such as motor imagery tasks, that induce transient power modulations of the EEG spectrum (Graimann and Pfurtscheller, 2006; Zhang, 2019). Modulations of this kind appear as a decrease (event-related desynchronization or ERD) or an increase (event-related synchronization or ERS) of spectral power at specific frequency bands (Pfurtscheller and Lopes da Silva, 1999). ERD/ERS is also known as an event-related spectral perturbation (ERSP), which measures the event-related spectral changes relative to a reference interval used as the spontaneous EEG baseline in a wide range of frequencies (Makeig, 1993). Therefore, TFA was performed on the aforementioned preprocessed trials to visualize the ERD/ERS patterns using the EEGLAB toolbox (version 2021.1) (Delorme and Makeig, 2004).

The resting stage of every trial was discarded and not considered in the present study, as it does not contain relevant EEG activity for the analysis. ERSP was computed at the frequency ranges from 4 to 30 Hz for all channels to calculate the power spectrum by applying the Morlet wavelets transform with incremental cycles (7 cycles at the lowest frequency to 14 at the highest), resulting in 200-time points $(-6.03, 3.02)$ s. The baseline reference was then taken from $-3.5$ to $-3$ s (which corresponds to the non-movement interval) at the beginning of step 3 in the sequence of each trial. Spectral power changes were averaged at each time point and normalized by baseline spectra. The significance of ERSP deviations from the baseline was analyzed using the bootstrap method ($\alpha = 0.05$) (Graimann and Pfurtscheller, 2006). Accordingly, ERSP could identify significant ERD and ERS as negative and positive spectral changes, respectively (Zhang, 2019).

#### 2.3.1.3. Feature extraction

One of the most successful algorithms in BCI research for feature extraction is the common spatial pattern (CSP) (Padfield et al., 2019). This method finds spatial filters that project EEG data into a new space in which the variances corresponding to one class are maximized while the variances of a second class are minimized (Lotte and Guan, 2011). In this study, an enhanced version of the original CSP algorithm, known as the filter bank common spatial pattern (FBCSP) algorithm (Ang et al., 2012), was implemented using a FIR filter bank of five digital band-pass filters centered on five EEG frequency bands (theta: 4–8 Hz, alpha: 8–12 Hz, low-beta: 12–16 Hz, mid-beta: 16–20 Hz, high-beta: 20–30 Hz) (Chen et al., 2018). All the filters in the filter bank were designed in the frequency domain using a Gaussian kernel with unitary gain.

The FBCSP algorithm is useful when the frequency components of the modulated signals may vary among subjects. For instance, in the motor imagery paradigm, a particular frequency of the sensorimotor rhythm is not the same for all users (Saha and Baumert, 2020). For this reason, each preprocessed epoch of a training set was filtered using the FIR filter bank in order to obtain spectrally filtered epochs $Y = [y_{e,f}(t)]_{n_e \times n_t \times n_f}$, where $n_f$ is the EEG frequency subband (4–8 Hz, 8–12 Hz, 12–16 Hz, 16–20 Hz, 20–30 Hz).

We studied the bi-class classification of two pairs of experimental tasks: MotorImageryA vs. IdleStateA for the sit-to-stand transition and MotorImageryB vs. IdleStateB for the stand-to-sit transition. The CSP algorithm was then applied to each subband and each pair of the experimental tasks. The CSP algorithm provided an $n_e \times n_e$ projection matrix $W = [w_1, w_2, \ldots, w_{n_e}]'$ for each pair of experimental tasks. This matrix was a set of subject-dependent spatial patterns, reflecting the specific activation of cortical areas during the experimental task. With the projection matrix $W$, the decomposition of an epoch $Y$ was described by $Z = WY$, where this transformation projected the variance of the filtered EEG signals of $Y$ onto the rows of $Z$ and gave rise to $n_e$ new time series. The columns of $W^{-1}$ were a set of CSPs that can be thought of as time-invariant EEG source distributions (Ortner et al., 2015).

The number of spatial filters retained was chosen as six for all subjects and training sets, as recommended in Blankertz et al. (2008b). The three first spatial filters contribute most to the variance of class one data, and the last three spatial filters contribute most to the variance of class two data. If $n_m = 6$ represents the number of spatial filters retained per frequency subband, the spectrally filtered epochs $Y$ are transformed into spatially filtered epochs $Y^{\mathrm{CSP}} = [y^{\mathrm{CSP}}_{e,f}(t)]_{n_m \times n_w \times n_t \times n_f}$, where $n_w$ is the number of time windows of $n_t$ time points for each EEG epoch. During the offline feature extraction, each EEG epoch was split into two-time windows of intervals $[-1, 0)$ s and $[0, 1)$ s, respectively, where 0 s is the onset time of step 3 in the sequence of a trial. Finally, for each frequency subband and projected channel, the BCI calculated the log-variance. This resulted in 30 features ($n_m \times n_f$) by 180 observations ($n_{\mathrm{epochs}} \times n_w$) for each pair of experimental tasks from each participant that would be used in the regularized linear discriminant analysis (RLDA) classifier. The Fisher's criterion was applied to evaluate the extracted features.

#### 2.3.1.4. Classifier

Linear classifiers have proven to be an efficient option for the detection of EEG rhythms in motor imagery paradigms for BCI applications (Oikonomou et al., 2017). In this category, linear discriminant analysis (LDA) can provide optimal results and outperform more complex classification techniques. Additionally, LDA is relatively easy to train and evaluate and requires a low computational cost to classify new observations. Therefore, two binary classification models based on LDA with regularized covariances were used to discriminate with a first model, MotorImageryA vs. IdleStateA for the sit-to-stand classification scenario, and with a second model, MotorImageryB vs. IdleStateB for the stand-to-sit classification scenario.

These types of binary-class models are highly employed in MI-based BCI applications (Lotte et al., 2018a). The proposed BCI employed the regularized linear discriminant analysis (RLDA) as a classification machine learning model to decide what class to assign to the processed data according to a linear combination of the feature vector (Fu et al., 2019). If $x$ represents a real vector of $n_c = 30$ features for an EEG epoch, the classification model evaluates the function

$$f(x) = g\left(\sum_{i=1}^{n_c} b_i x_i + d\right), \tag{5}$$

where $b = [b_1, b_2, \ldots, b_{n_c}]'$ and $d$ are the coefficients of the linear model, and $g(a)$ is a scalar function. Then, the classification model returns a label or category $l \in \{1, -1\}$ to the given observation based on the evaluation of $f(x)$. A typical approach is to use a threshold value such that values above it have the class label $l = 1$. Conversely, values below this threshold correspond to the other class label $l = -1$.

LDA finds the class label $l$ that maximizes the conditional probability $p(L = l|X = x)$ (Ng and Jordan, 2001). It assumes that the probability density functions $p(X = x|L = -1)$ and $p(X = x|L = 1)$ are both normally distributed with mean vectors $m_{-1}$, $m_1$ and covariance matrices $C_{-1}$, $C_1$. Under these assumptions, the decision rule $p(L = 1|X = x) > p(L = -1|X = x)$ is expressed as a dot product $b'x + d > 0$, where

$$b = 2C^{-1}(m_1 - m_{-1}), \tag{6}$$

$$d = \ln\left(\frac{P(L=-1)}{P(L=1)}\right) + m'_{-1}C_{-1}^{-1}m_{-1} - m'_1C_1^{-1}m_1, \tag{7}$$

and $P(L = l)$ is the probability of class label $l$. Additionally, for the automatic regularization of the LDA algorithm, the BCI uses the method proposed by Ledoit and Wolf (2004) to compute $C_{-1}$ and $C_1$ (Lotte and Guan, 2011).

The goal of the two classification models was the discrimination of the different pairs of experimental tasks to return a motor imagery state or class label $l$ that represented when the participant was imagining ($l = 1$) or not imagining ($l = -1$) a movement based on the observations or corresponding resulting features for each EEG epoch. For this purpose, the BCI incorporated two RLDA classifiers to complete the two machine learning models for the sit-to-stand and stand-to-sit transitions, respectively. These classifiers are simple and have a low computational requirement, which makes them suitable for the online BCI (Mao et al., 2017).

#### 2.3.1.5. Classification model performance

In the final steps of the offline phase, the two complete machine learning models of the study subject were evaluated for the transitioning actions: (*i*) Sit-to-stand and (*ii*) Stand-to-sit. The evaluation of these models depends on the collection of labeled data obtained from datasets *A* and *B* in the offline experiments. For this assessment, the machine learning models *i* and *ii* for each participant were independently assessed by applying a five-fold cross-validation procedure to avoid overfitting and measure generalization on each model (Berrar, 2019). In this procedure, the set of trials of *A* and *B* were randomly split into five equal-sized subsets each, respectively. For each fold, the BCI uses four subsets to train the models $m \in \{i, ii\}$. Then, the remaining subset is used to test the corresponding model *m*. This process was repeated with mutually exclusive training and test subsets until the five cross-validations were completed. The classification accuracy $acc_{m,c}$ of each class $c \in \{\text{MotorImagery}, \text{IdleState}\}$ was calculated as described below:

$$acc_{m,c} = \frac{n_{\text{correct}}}{n_{\text{total}}} \times 100\%, \tag{8}$$

where $acc_{m,c}$ is the offline accuracy, $n_{\text{total}}$ is the total number of instances of class *c*, and $n_{\text{correct}}$ is the number of instances classified correctly in class *c* by model *m*. The overall model accuracy $acc_{m,\text{overall}} = 0.5 \times (acc_{m,\text{MotorImagery}} + acc_{m,\text{IdleState}})$ and the confusion matrices were also computed.

Additionally, permutation testing was applied to assess the significance level of the overall model accuracies (Good, 2006). This test repeats the five-fold cross-validation procedure by shuffling the class labels during the training of the classifiers to compute the empirical random classification accuracy. In this methodology, the null hypothesis ($H_0$) indicates that observations of both classes are exchangeable so that any random permutation of the class labels produces similar accuracies to the obtained with the non-permuted data. The alternative hypothesis ($H_1$) is accepted when the overall model accuracy is an extreme value in the empirical distribution built with several random permutations of the labels. When the alternative hypothesis is accepted, we can say that the overall model accuracy is above the chance level.

### 2.3.2. Online phase

For each participant, the two machine learning models obtained in the offline phase were used to carry out two online experiments: (*I*) Sit-to-stand and (*II*) Stand-to-sit. Each participant was instructed to select, in no particular order, 30 sequences for experiment *I* (15 trials of MotorImageryA tasks and 15 trials of IdleStateA tasks) and 30 other sequences for experiment *II* (15 trials of MotorImageryB tasks and 15 trials of IdleStateB tasks). Each trial was unique and was generated pseudo-randomly before the experiment.

The timeline of the online sequences was indicated by the user interface in the same way as shown in Figure 2. However, the difference between the offline and online timelines was in step 3. In the third step of the online timeline, the participant performed one experimental task in response to the figure shown in step 2, and the BCI attempted to detect this task in real time for 3–15 s at the same time that provided feedback.

Technically, the online classification of single-trial EEG data could be done as in the offline phase since the trained classifiers can be applied to feature vectors calculated from an arbitrary window. However, this is likely to lead to unreliable results since those classifiers are adjusted to detect signals with a specific time related to the response (Blankertz et al., 2001). There is no guarantee that the classifier will behave similarly elsewhere. As suggested in Blankertz et al. (2001, 2008a), Syan and Harnarinesingh (2010), and Mendoza-Montoya (2017), sliding windows are usually used to increase online classification's robustness to time-shifted signals. Thus, during the online feature extraction, the data length of each epoch was 1 s, that was 250 sample points, and the BCI split them into five sliding windows (50 sample points each), resulting in five feature vectors from each trial.

The BCI processed and provided continuous visual feedback on the results obtained after classifying five consecutive time windows (50 sample points each) from one epoch (250 sample points). If one experimental task produces high MI-related activity, the BCI makes the white background of the corresponding figure look bigger (see Figure 3A). If the idle state-related activity is higher, that figure's white background looks bigger (see Figure 3B). Otherwise, both figures' backgrounds are the same size (see Figure 3C). This visual feedback notifies the participant when the BCI is detecting MI-related activity and helps to increase the MI modulations of the intended movement (Yu et al., 2015).

As shown in Figure 4, the RLDA classifier labeled each of the time windows from 50 sample points with the name of one of the experimental tasks. The label might be used directly to determine the action or control command to produce with the BCI. However, because the accuracy of the MI-based BCI is typically below 90%, the risk of executing the wrong action is high (Irimia et al., 2018). For this reason, the BCI only generates command or action signals when an experimental task has been detected several times for a few seconds. The minimum time required is about 3 s: 1 s to acquire a whole epoch of EEG signals, another second to classify the EEG data five times, and a third second to select a command or action to execute.

When the same experimental task label has been detected five consecutive times in one epoch, the BCI synchronizes the state of the GUI to produce visual feedback on the selected task, and the corresponding figure background is colored blue (see Figure 3D). On the other hand, if the BCI does not detect the same label five consecutive times, it waits for new labels and dismisses the oldest ones. If the BCI detects the requested

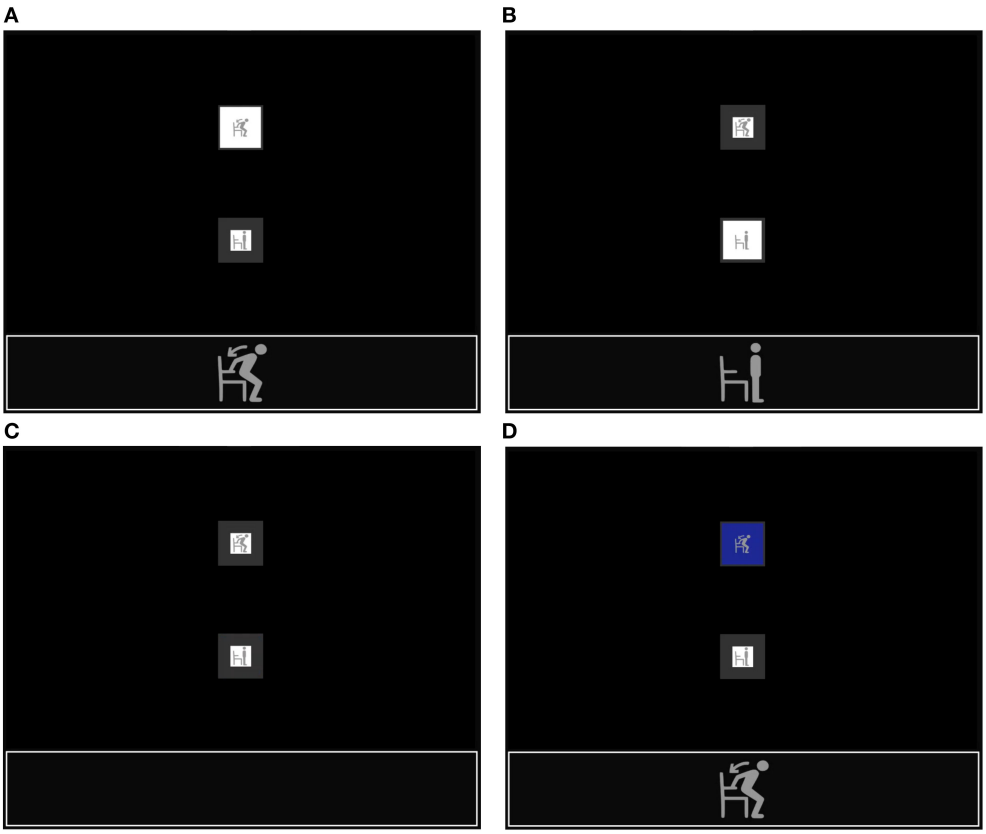



FIGURE 3
Continuous visual feedback during the online stand-to-sit experiment when: **(A)** the BCI has detected a high MI-related activity, **(B)** the BCI has detected a high idle state-related activity, **(C)** the BCI has not detected a dominant experimental task, and **(D)** the BCI has detected a dominant experimental task.

experimental task in less than 15 s, the sequence is interrupted to provide visual feedback (see Figure 3D) and continues with step 4 (see Figure 4). On the contrary, if the BCI does not recognize the experimental task and reaches the time limit of 15 s, the BCI simply continues with step 4.

#### 2.3.2.1. Online BCI evaluation

The online assessment aims to investigate the feasibility of decoding in real time the two classes (MotorImagery vs. IdleState) of the two binary machine learning models in the sit-to-stand and stand-to-sit experiments. Consequently, the online evaluation procedure was carried out for each participant independently. Two machine learning models were used for each participant to assess the feasibility of continuous detection of motor information along the online trials. To this end, the performance of the BCI was evaluated in terms of the following detection metrics that were calculated using Equations (9)–(13).

$$TPR = \frac{TP}{TP + FN}, \tag{9}$$

$$TNR = \frac{TN}{TN + FP}, \tag{10}$$

$$acc_{\text{online}} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{11}$$

$$PPV = \frac{TP}{TP + FP}, \tag{12}$$

$$NPV = \frac{TN}{TN + FN}, \tag{13}$$

where:

- Sensitivity or true positive rate (*TPR*) indicates the percentage of times that the motor imagery class was detected correctly (*TP* are the true positives and *FN* are the false negatives).
- Specificity or true negative rate (*TNR*) denotes the percentage of times that the idle state class was detected correctly (*TN* are the true negatives and *FP* are the false positives).

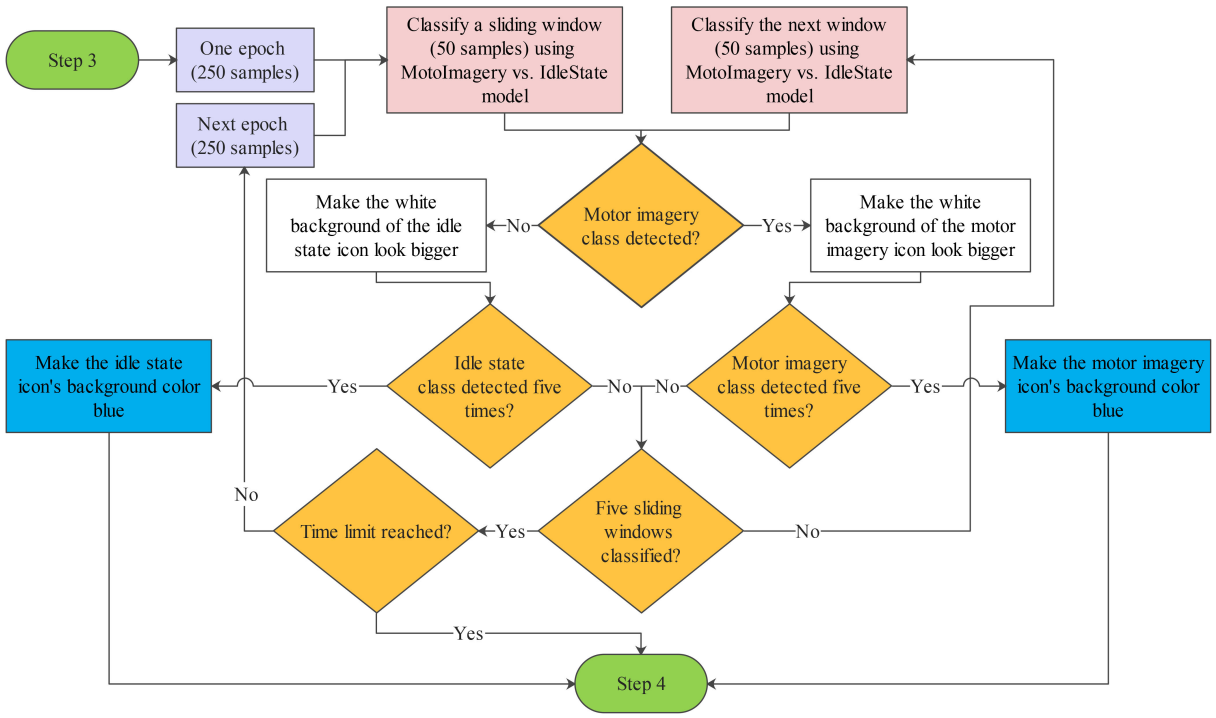



FIGURE 4
Flowchart of the online classification. In online conditions, the BCI processes each time window of the epoch and assigns it one class between motor imagery and idle state. The BCI also provides continuous visual feedback on the classifications by making the white background of the icon corresponding to the class look bigger or turn blue. The time limit to detect five consecutive times the same class and pass step 3 is 15 s.

- Accuracy ($acc_{\text{online}}$) represents the probability of correctly detecting the motor imagery and idle state classes given the total number of attempts to detect them.
- The positive predictive value (*PPV*), also called precision, is the probability that the detection of the motor imagery class is correct given the total number of times that class is detected.
- The negative predictive value (*NPV*) is the probability that the detection of the idle state class is correct given the total number of times that class is detected.

The information transfer rate (*ITR*) was also used as a performance metric for the online evaluation of the BCI. The calculation of this metric is based on the amount of information transferred per unit of time. The *ITR* was calculated for each participant in bits/min using the following formula (He et al., 2018):

$$ITR = \frac{60}{T} \times [1 + (acc_{\text{online}}) \log_2(acc_{\text{online}}) + (1 - acc_{\text{online}}) \log_2(1 - acc_{\text{online}})], \quad (14)$$

where T is the average time from task performing to task detection (detection time in seconds). Under these conditions, the maximum possible information transfer rate is 20 bits/min for each online experiment (Wolpaw et al., 2002).

## 3. Results

ERD/ERS has been studied widely as one of the brain activity markers for motor imagery tasks. Figure 5 demonstrates the grouped ERSP across 32 participants in the time-frequency (TF) plots on all electrodes and in the group-level 2-D scalp topographies during each stage of the sit-to-stand and stand-to-sit experiments (excluding the rest period). The ERSP estimates ERD/ERS from the entire duration of the trials relative to the baseline spectra from 4 to 30 Hz. All present ERSP values were significant (see Figure 5, ERD in blue, ERS in red) compared to the baseline ($\alpha = 0.05$). There was a tendency to decrease the alpha-band power for the action observation stage in all sit-to-stand and stand-to-sit trials, indicating ERD mainly in the parietal and parieto-occipital regions. Only for the motor imagery stage, the ERD sustained toward the centro-parietal and central electrode sites was found. However, this ERD was not present for the idle state in all trials of both experiments. Furthermore, we observed a significant increase

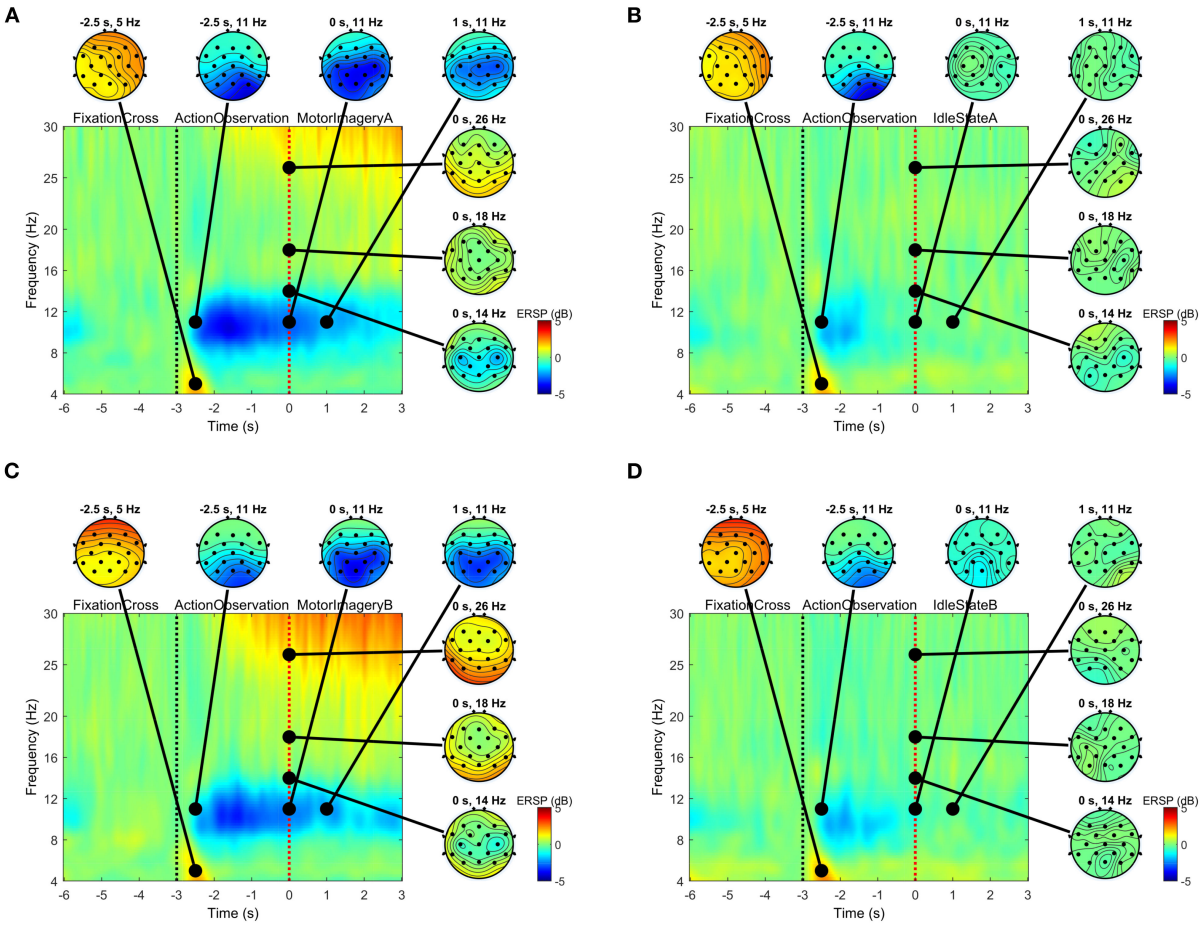



FIGURE 5
Group-level event-related spectral perturbation (ERSP) for frequencies between 4 and 30 Hz across all trials pooled for sit-to-stand (top panels) and stand-to-sit (bottom panels) experiments compared to the baseline from −3.5 to −3 s. Time-frequency (TF) plots combined the TF decompositions across all channels. Note that these plots with the 2-D scalp topographies also combined ERSP from different subjects and all present ERSP values were statistically significant compared to the baseline ($\alpha = 0.05$). The time interval $(-6, -3]$ s corresponds to the fixation stage, $(-3, 0)$ s corresponds to the action observation stage, and $[0, 3)$ s corresponds to the stage of one of the four experimental tasks: **(A)** MotorImageryA, **(B)** IdleStateA, **(C)** MotorImageryB, or **(D)** IdleStateB.

in the beta-band power, indicating ERS, in the motor imagery stage of sit-to-stand and stand-to-sit trials compared to the idle state.

The Fisher's criterion was applied to evaluate the extracted features from each participant and show the highest rank features, as reported in Table 1. The features most common to all participants for the sit-to-stand and stand-to-sit classification scenarios were the low-beta frequency band with the fifth spatial filter and the alpha frequency band with the sixth spatial filter, respectively. In both classification scenarios, the highest Fisher score values were 5.08 and 9.65 in the low-beta frequency band of participant ID P25. The lowest Fisher score values were 0.32 for the sit-to-stand classification scenario in the theta frequency band of participant ID P03 and 0.34 for the stand-to-sit classification scenario in the alpha frequency band of participant ID P13.

Box plots were used to present the distribution of the offline classification accuracy estimated with the five-fold cross-validation procedure across all participants (see Figure 6). In particular, the mean accuracies (denoted by ×) of the MotorImageryA, IdleStateA, and overall A classes were, respectively, 89.21, 87.81, and 88.51% in the sit-to-stand classification scenario. Likewise, the medians of the MotorImageryA, IdleStateA, and overall A classes were 90.58, 89.60, and 89.36%, respectively. The worst classifier performances were below 80% (2 women and 1 man), whereas 15 participants obtained classifier performances above 90% (7 women and 8 men). The best model performance was 98.49% and the worst was 58.02%.

TABLE 1 Comparison of the highest-ranking features.

| Participant ID | Sit-to-stand | | | Stand-to-sit | | |
|---|---|---|---|---|---|---|
| | Feature | | Fisher score value | Feature | | Fisher score value |
| | Rhythm | Spatial filter | | Rhythm | Spatial filter | |
| P01 | Alpha | 1 | 3.64 | High Beta | 2 | 2.03 |
| P02 | Low Beta | 4 | 1.23 | Mid Beta | 1 | 0.52 |
| P03 | Theta | 5 | 0.32 | Alpha | 4 | 0.40 |
| P04 | Low Beta | 6 | 2.61 | Alpha | 6 | 2.01 |
| P05 | High Beta | 6 | 1.65 | High Beta | 2 | 1.13 |
| P06 | Alpha | 5 | 2.71 | Alpha | 4 | 1.98 |
| P07 | Mid Beta | 2 | 1.31 | High Beta | 2 | 1.68 |
| P08 | Mid Beta | 6 | 0.58 | Low Beta | 6 | 0.54 |
| P09 | Low Beta | 5 | 2.32 | Low Beta | 5 | 1.93 |
| P10 | Alpha | 5 | 3.97 | Alpha | 6 | 1.75 |
| P11 | High Beta | 5 | 1.13 | Alpha | 6 | 1.70 |
| P12 | Mid Beta | 2 | 0.57 | Theta | 1 | 0.88 |
| P13 | Alpha | 6 | 0.67 | Alpha | 5 | 0.34 |
| P14 | Alpha | 5 | 0.87 | Alpha | 1 | 1.00 |
| P15 | Alpha | 1 | 0.79 | Alpha | 5 | 0.54 |
| P16 | Theta | 1 | 3.04 | Alpha | 2 | 3.33 |
| P17 | Mid Beta | 1 | 2.14 | Mid Beta | 1 | 1.61 |
| P18 | High Beta | 6 | 0.71 | Alpha | 5 | 0.57 |
| P19 | Low Beta | 6 | 0.84 | Alpha | 6 | 0.72 |
| P20 | Low Beta | 1 | 1.04 | Low Beta | 2 | 2.45 |
| P21 | Theta | 6 | 0.62 | Mid Beta | 6 | 0.55 |
| P22 | Low Beta | 4 | 1.00 | Theta | 1 | 0.95 |
| P23 | Low Beta | 5 | 4.82 | High Beta | 2 | 0.63 |
| P24 | Low Beta | 1 | 0.90 | High Beta | 1 | 1.15 |
| P25 | Low Beta | 5 | 5.08 | Low Beta | 6 | 9.65 |
| P26 | High Beta | 2 | 1.30 | Theta | 2 | 0.99 |
| P27 | Low Beta | 6 | 2.39 | Alpha | 5 | 1.63 |
| P28 | Alpha | 6 | 0.72 | Alpha | 1 | 1.51 |
| P29 | Low Beta | 5 | 3.10 | Low Beta | 5 | 2.16 |
| P30 | Low Beta | 5 | 0.98 | Alpha | 6 | 0.80 |
| P31 | Low Beta | 6 | 2.19 | Alpha | 6 | 1.24 |
| P32 | Theta | 6 | 1.61 | Alpha | 3 | 1.05 |

Features were ranked using Fisher's separability criterion.

In the stand-to-sit offline classification scenario, the mean accuracies of the MotorImageryB, IdleStateB, and overall B classes were, respectively, 84.99, 85.60, and 85.29%. Similarly, the medians of the MotorImageryB, IdleStateB, and overall B classes were 86.12, 87.43, and 86.83%, respectively. Additionally, 7 participants obtained classifier performances below 80% (4 women and 3 men) and 10 participants above 90% (6 women and 4 men). In this case, the best model performance was 99.44% and the worst was 56.51%.

In the permutation tests, the overall classification accuracies in 30 of 32 participants were statistically significant ($p < 0.05$, 1,000 random permutations) for the sit-to-stand and stand-to-sit classification scenarios. Only participant ID P03 presented $p$-values higher than 0.05 for both classification scenarios, which means not statistically significant. Likewise, for the stand-to-sit classification scenario, the $p$-values of participant ID P13 were higher than 0.05, not statistically significant, and indicate strong evidence for the null hypothesis. Altogether, the offline classification results showed the feasibility of recognizing the studied motor imagery tasks vs. idle state above empirical chance levels. The Supplementary material section provides the results for each participant.

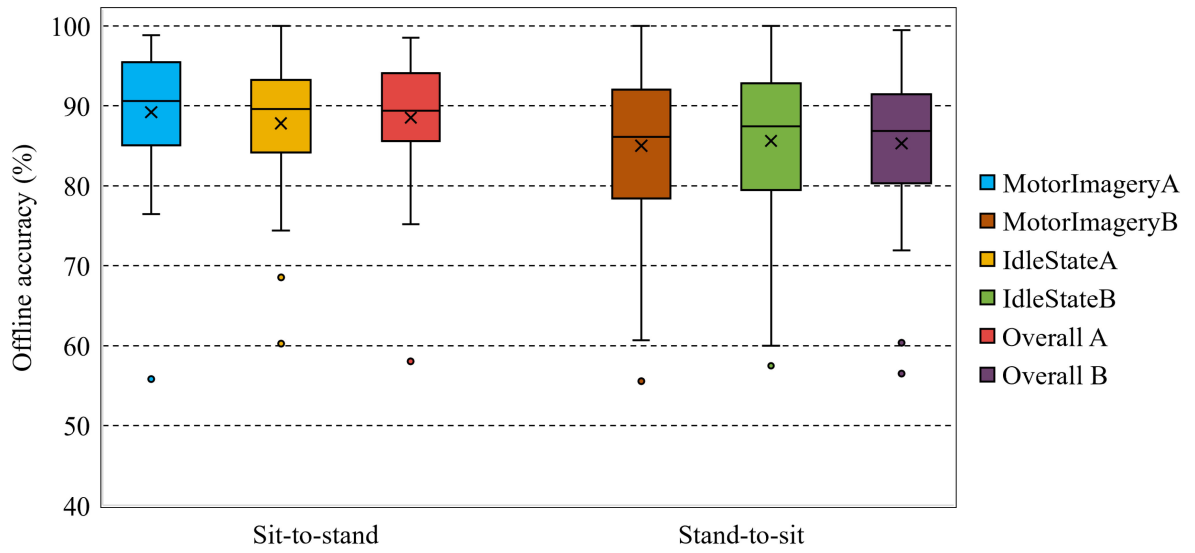



FIGURE 6
Classification accuracies (%) estimated with five-fold cross-validation across all participants for the motor imagery vs. idle state classes in the sit-to-stand and stand-to-sit offline experiments. × represents the mean. The Supplementary material section provides the result for each participant.

Figure 7 shows the results of the confusion matrices obtained in the sit-to-stand and stand-to-sit classification scenarios for the MotorImagery vs. IdleState classes. Regarding the confusion matrix in Figure 7A, corresponding to the sit-to-stand scenario, the true positive rate (*TPR*), false negative rate (*FNR*), false positive rate (*FPR*), and true negative rate (*TNR*) were, respectively, 89.2, 10.8, 12.2, and 87.8%. As for the confusion matrix in Figure 7B, for the stand-to-sit scenario, the *TPR*, *FNR*, *FPR*, and *TNR* were 85.0, 15.0, 14.4, and 85.6%, respectively. Overall, the sit-to-stand and stand-to-sit offline classification scenarios showed comparable results and a relatively balanced performance among the different classes.

The box plots in Figure 8 were used to show the characteristics of the sensitivity, precision, specificity, and negative predictive value calculated across all participants in the sit-to-stand and stand-to-sit online classification scenarios. In the sit-to-stand scenario, most of the characteristics had a mean (indicated by ×) of between 90 and 100% and a median of 100% for both the women and men groups. With respect to the stand-to-sit scenario, these characteristics had also a mean of between 90 and 100% and a median of 100%. These results indicate that models for the sit-to-stand scenario can discriminate between EEG epochs of MotorImagery vs. IdleState classes just as well as models for the stand-to-sit scenario.

Finally, the distribution of the online accuracy, detection time and ITR in the sit-to-stand and stand-to-sit online experiments are represented in Figure 9. The mean accuracies ± standard error of the sit-to-stand and stand-to-sit online experiments were 94.69 ± 1.29% and 96.56 ± 0.83%, respectively. The average detection times were 4.70 ± 0.11 s and 4.77 ± 0.16 s, and the mean ITRs were 10.12 ± 0.73 bit/min and 11.13 ± 0.72 bit/min for the sit-to-stand and stand-to-sit online experiments, respectively. The shortest detection times were 3.75 s in the sit-to-stand experiments and 3.60 s in the stand-to-sit experiments. The longest detection times were 6.15 s in the sit-to-stand experiments and 6.90 s in the stand-to-sit experiments. Likewise, the minimum ITRs were 2.57 bits per minute in the sit-to-stand experiments and 4.10 bits per minute in the stand-to-sit experiments. The maximum ITRs were 16.02 bits per minute in the sit-to-stand experiments and 16.68 bits per minute in the stand-to-sit experiments.

## 4. Discussion

In this study, we found that sitting and standing motor imagery tasks can be recognized online using an EEG-based BCI. From our findings, a high percentage of participants (above 80%) can control the motor imagery (MI)-based BCI for standing and sitting. In addition to expanding the participant sample size, using more EEG active electrodes, and improving signal processing, this study enhanced previous research by proposing a solution to the problem of online decoding of motor imagery electroencephalography (MI-EEG) signals for standing and sitting.

For the first time, decoding EEG rhythms offline and online during motor imagery tasks for standing and sitting had a satisfactory performance using a feasible BCI paradigm. Current

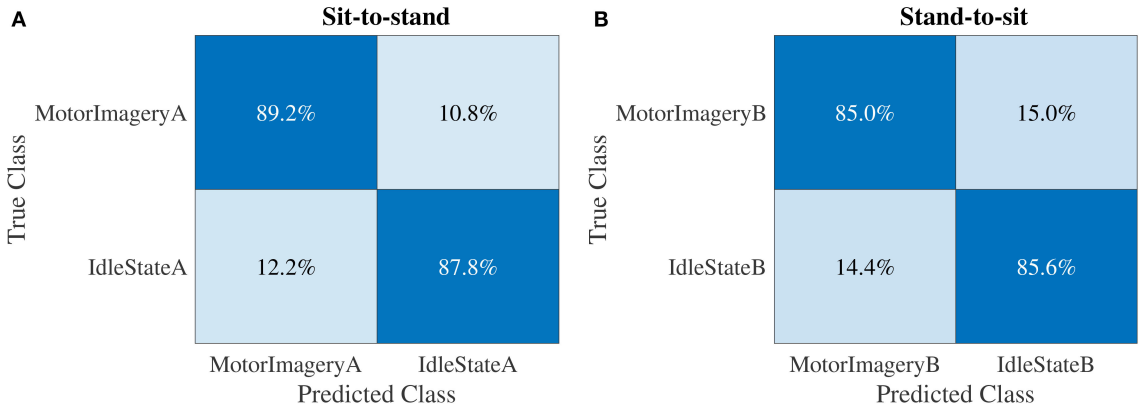



FIGURE 7
Confusion matrices obtained in the sit-to-stand **(A)** and stand-to-sit **(B)** offline classification scenarios for all participants.

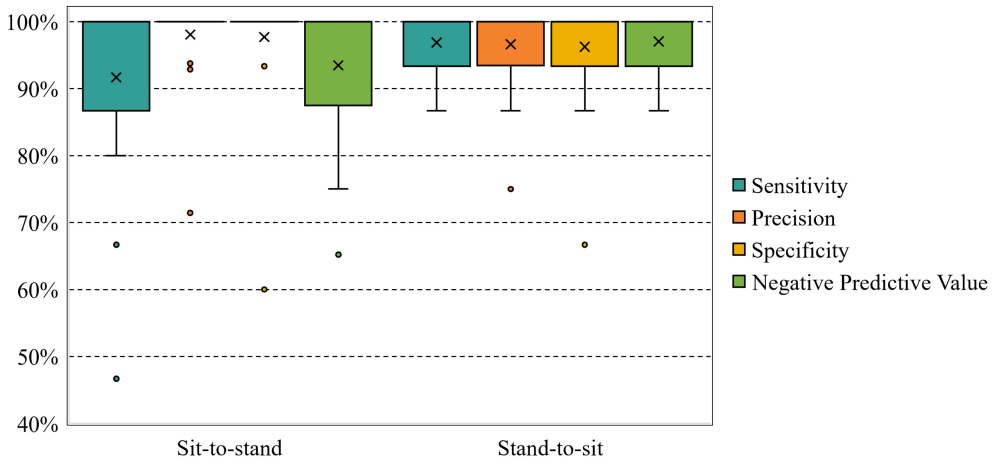



FIGURE 8
Across-all-participants distributions (%) of sensitivity, precision, specificity, and negative predictive value obtained in the sit-to-stand and stand-to-sit online experiments. × represents the mean. The Supplementary material section provides the result for each participant.

BCI paradigms that consider the complexity of shifting from sitting to standing and vice versa are usually either based on left-hand and right-hand motor imagery tasks (Noda et al., 2012; Wang et al., 2018) or SSVEP signals (Kwak et al., 2017). These BCI paradigms have proven effective in transferring information from the brain to a computer. However, they are unnatural for the brain to interact with and thus require much more cognitive resources to act as traditional human-computer interfaces for the sit-to-stand and stand-to-sit transitions. For this reason, one of the main contributions of the proposed BCI paradigm is to provide a more natural interaction between the user and the interface, which is a current challenge in the design of BCI systems (Xu et al., 2021).

Previous studies have shown the effects of task complexity on ERD/ERS rhythms and the complexity of the sit-to-stand and stand-to-sit movements (Bulea et al., 2014; Singh et al., 2017; Mashat et al., 2019; Chaisaen et al., 2020). To overcome these limitations, we used information from the idle state, a neutral condition, to facilitate the classifier's recognition of distinctive characteristics of the related motor imagery task. ERD BCIs have pursued this approach with considerable success generating brain signals that are easier to categorize (see Figure 5). This BCI paradigm with a low cognitive load could explain why our study did not show the typical BCI inefficiency (Allison and Neuper, 2010; Edlinger et al., 2015; Liu et al., 2020; Xu et al., 2021). Additionally, the frequency bands and spatial filters of the extracted features reported in Table 1 could provide the most discriminated information.

The overall classification accuracies (see Figure 6), estimated with cross-validation in the offline classification scenarios, are similar to those reported in other related BCI literature (Wang et al., 2018; Chaisaen et al., 2020). For instance, in the

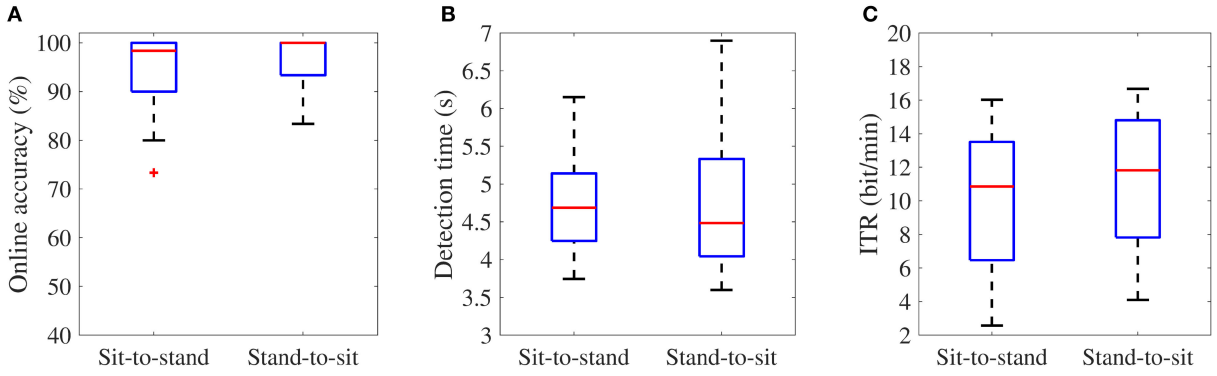



FIGURE 9
Box plots of the online accuracy **(A)**, task detection time **(B)**, and ITR values **(C)** of the BCI online experiments. The Supplementary material section provides the result for each participant.

offline analysis by Chaisaen et al. (2020), the classification of action observation (AO) and motor imagery (MI) provided the grand average accuracy ± standard error (SE) of 82.73 ± 2.54% in the stand-to-sit transition, which is lower compared to 85.29 ± 1.83% between the classification of motor imagery and idle state in this study. In the current study, the highest grand average accuracy ± SE was 88.51 ± 1.43% between the classification task of motor imagery and idle state in the sit-to-stand transition, compared to 76.14 ± 3.14% in the classification of AO and MI by Chaisaen et al. (2020). Furthermore, within the 60 min of training, 30 of 32 participants achieved an overall accuracy above chance level.

The confusion matrices (see Figure 7) showed that the trained classification models generated balanced results for the different classes. The overall accuracy results obtained in the classification models trained with the permutation testing method also confirm this situation empirically (see Supplementary material section). These results show that the classifiers were not highly biased toward any experimental task. Furthermore, sensitivity, precision, specificity, negative predictive value, and accuracy in the online phase are not measured by cross-validation as in offline experiments. Therefore, due to the heuristics described to detect one experimental task and compute the performance metrics in the online experiments, these metrics are not directly comparable between the online and offline phases.

The online classification results (see Figure 8) demonstrated the feasibility of the BCI to decode real-time EEG rhythms during the studied motor imagery tasks, whereas previous studies usually only presented classification accuracy. We also calculated sensitivity, precision, specificity, and negative predictive value to illustrate the online detection ability of the BCI to motor imagery-related potentials vs. idle state potentials. When considering all participants as a single group, the mean accuracies ± SE of the sit-to-stand and stand-to-sit online experiments were 94.69 ± 1.29% and 96.56 ± 0.83%, respectively, which are above the range of previous studies (Noda et al., 2012; Wang et al., 2018; Choi et al., 2020).

In both online classification scenarios (see Figure 9), the number of processed epochs and thresholds for motor imagery and idle state classes are customizable for each participant to improve online accuracy, detection time, and ITR of the MI-based BCI system. However, we used the same parameters for all participants, and it is essential to improve the system performance for participants who cannot achieve high detection rates. One potential strategy for improving system performance would be to modify the detection criteria in the third step of the online timeline. For instance, the classification of multiple sliding windows per trial provides a simple way to find a balance between the detection speed and the average accuracy of the system (Lee et al., 2019).

Both the offline (five-fold cross-validated data) and online (not cross-validated data) classification results demonstrated that the MI-based BCI could identify new observations of each class with high accuracy. Nevertheless, the results are similar to those reported in other MI-based BCI studies (Irimia et al., 2018; Choi et al., 2020; Gurve et al., 2020). These may be due to the high motivation of the research subjects. Additionally, the instructions provided during the subjects' training emphasized the differences between the studied motor imagery tasks with a high cognitive load vs. the no imaginary movement state (idle or rest state) with a low cognitive load to achieve a good performance. At the same time, the strategy applied to increase the online accuracy by single-trial classification using sliding windows allowed to reduce the classification errors but slowed down the detection time and ITR. Therefore, the choice of the parameters is crucial to keep a balance between the detection speed and the classification accuracy of the interface.

If we consider the response times of the present BCI system, it is not possible to implement an active fine control for a robotic device. However, by improving the response times of this new BCI, users could send commands to standing devices (like standing wheelchairs) to execute complete sit-to-stand or stand-to-sit transitions using the interface. For this reason, it is our view that the MI-based BCI system is suitable for the movements studied in this research.

One of the difficulties encountered in this study is the lack of an objective comparison between offline vs. online results. Ideally, the online classification performance should be calculated similarly to the values calculated using the cross-validation procedure of the offline phase. However, the crucial problem is to perform cross-validation with only a few observations in the online phase because it may lead to overfitting and poor generalization. Hence, cross-validation was necessary to evaluate the classification performance of the machine learning models in the offline phase. By contrast, in the online phase, the classification of multiple sliding windows per trial addressed the problem of single-trial misclassification and false positives in order to evaluate the online classification performance of the models (Mendoza-Montoya, 2017; Delijorge et al., 2020; Hernandez-Rojas et al., 2022).

The results suggest that a large population can control the EEG-based BCI and that high accuracy of above 90% can be achieved. Further research is required to establish whether people suffering from mobility impairments (who had previously been able to stand up and sit down before the impairments developed) could perform motor imagery tasks and operate the EEG-based BCI system for standing and sitting. Furthermore, more techniques for feature extraction and more machine learning models also are considered to extend the analyses. Classification techniques such as deep learning can be another alternative to analyze the problem studied here. Another interesting aspect would be to include more motor imagery tasks (e.g., three-class classification: imagining standing vs. imagining sitting vs. resting) that the system can interpret and test in more realistic environments. This system could represent the basis for modern interfaces' integration into future technologies (e.g., exoskeleton-based rehabilitation systems or brain-controlled standing wheelchairs) where the interface can be adapted to the user's specific disability.

## 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 Ethics Committee of the Universidad Antonio Nariño. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

NT-G, AO-C, and AJ conceived and developed the presented idea. NT-G, OM-M, and JA participated in the design and implementation of the study. NT-G performed the experiments, conducted the research, and responsible for writing the manuscript. AO-C, AJ, OM-M, and JA supervised NT-G research's data analysis, methods, and findings. All authors contributed to the article and approved the submitted version.

## Funding

This research was funded by Ministerio de Ciencia, Tecnología e Innovación of Colombia-Minciencias/Colciencias, contract 594-2019, project 80740-594-2019, and in part by the University of Illinois Chicago and Tecnologico de Monterrey (UIC-TEC) Seed funding Program 2021-2022.

## Acknowledgments

We would like to thank the Universidad Antonio Nariño (UAN), Universidad del Rosario, and Tecnologico de Monterrey for all their support in the development of this research. We would also like to thank the reviewers and the Journal Editor for their insightful comments. Special thanks to Noah Foster-Koth for all his efforts in editing the various versions 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



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fninf.2022.961089/full#supplementary-material

# References


Aggarwal, S., and Chugh, N. (2019). Signal processing techniques for motor imagery brain computer interface: a review. *Array* 1–2, 1–12. doi: 10.1016/j.array.2019.100003

Ahn, M., Cho, H., Ahn, S., and Jun, S. C. (2013). High theta and low alpha powers may be indicative of BCI-illiteracy in motor imagery. *PLoS ONE* 8, e80886. doi: 10.1371/journal.pone.0080886

Al-Fahoum, A. S., and Al-Fraihat, A. A. (2014). Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains. *ISRN Neurosci.* 2014, 1–7. doi: 10.1155/2014/730218

Allison, B. Z., and Neuper, C. (2010). "Could anyone use a BCI?" in *Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction, Chapter 3, 1st Edn* (London: Springer), 35–54.

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

Asanza, V., Peláez, E., Loayza, F., Lorente-Leyva, L. L., and Peluffo-Ordóñez, D. H. (2022). Identification of lower-limb motor tasks via brain-computer interfaces: a topical overview. *Sensors* 22, 1–24. doi: 10.3390/s22052028

Association, W. M. (2013). World medical association declaration of helsinki: ethical principles for medical research involving human subjects. *JAMA* 310, 2191–2194. doi: 10.1001/jama.2013.281053

Berrar, D. (2019). "Cross-validation," in *Encyclopedia of Bioinformatics and Computational Biology, Vol. 1* (Oxford: Academic Press Inc.), 542–545.

Blankertz, B., Curio, G., and Müller, K.-R. (2001). "Classifying single trial EEG: towards brain computer interfacing," in *Advances in Neural Information Processing Systems, Vol. 14* (Vancouver, BC: MIT Press), 157–164.

Blankertz, B., Losch, F., Krauledat, M., Dornhege, G., Curio, G., and Muller, K.-R. (2008a). The berlin brain-computer interface: accurate performance from first-session in BCI-naive subjects. *IEEE Trans. Biomed. Eng.* 55, 2452–2462. doi: 10.1109/TBME.2008.923152

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Muller, K.-R. (2008b). Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Process Mag.* 25, 41–56. doi: 10.1109/MSP.2008.4408441

Bobrova, E. V., Reshetnikova, V. V., Frolov, A. A., and Gerasimenko, Y. P. (2020). Use of imaginary lower limb movements to control brain-computer interface systems. *Neurosci. Behav. Physiol.* 50, 585–592. doi: 10.1007/s11055-020-00940-z

Bulea, T. C., Prasad, S., Kilicarslan, A., and Contreras-Vidal, J. L. (2014). Sitting and standing intention can be decoded from scalp EEG recorded prior to movement execution. *Front. Neurosci.* 8, 376. doi: 10.3389/fnins.2014.00376

Chaisaen, R., Autthasan, P., Mingchinda, N., Leelaarporn, P., Kunaseth, N., Tammajarung, S., et al. (2020). Decoding EEG rhythms during action observation, motor imagery, and execution for standing and sitting. *IEEE Sens J.* 20, 13776–13786. doi: 10.1109/JSEN.2020.3005968

Chen, W., Wang, S., Zhang, X., Yao, L., Yue, L., Qian, B., et al. (2018). "EEG-based motion intention recognition via multi-task RNNs," in *Proceedings of the 2018 SIAM International Conference on Data Mining (SDM)* (San Diego, CA: Society for Industrial and Applied Mathematics Publications), 279–287.

Choi, J., Kim, K. T., Jeong, J. H., Kim, L., Lee, S. J., and Kim, H. (2020). Developing a motor imagery-based real-time asynchronous hybrid BCI controller for a lower-limb exoskeleton. *Sensors* 20, 1–15. doi: 10.3390/s20247309

Congedo, M., Barachant, A., and Bhatia, R. (2017). Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review. *Brain Comput. Interfaces* 4, 155–174. doi: 10.1080/2326263X.2017.1297192

Delijorge, J., Mendoza-Montoya, O., Gordillo, J. L., Caraza, R., Martinez, H. R., and Antelis, J. M. (2020). Evaluation of a P300-based brain-machine interface for a robotic hand-orthosis control. *Front. Neurosci.* 14, 589659. doi: 10.3389/fnins.2020.589659

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

Edlinger, G., Allison, B. Z., and Guger, C. (2015). "How many people can use a BCI system?" in *Clinical Systems Neuroscience, Chapter 3* (Tokyo: Springer), 33–66.

Fu, R., Tian, Y., Bao, T., Meng, Z., and Shi, P. (2019). Improvement motor imagery EEG classification based on regularized linear discriminant analysis. *J. Med. Syst.* 43, 1–13. doi: 10.1007/s10916-019-1270-0

Gao, Z., Wang, Z., Ma, C., Dang, W., and Zhang, K. (2018). A Wavelet time-frequency representation based complex network method for characterizing brain activities underlying motor imagery signals. *IEEE Access* 6, 65796–65802. doi: 10.1109/ACCESS.2018.2876547

George, O., Smith, R., Madiraju, P., Yahyasoltani, N., and Ahamed, S. I. (2021). "Motor imagery: a review of existing techniques, challenges and potentials," in *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)* (Madrid: Institute of Electrical and Electronics Engineers Inc.), 1893–1899.

Good, P. I. (2006). *Resampling Methods: A Practical Guide to Data Analysis*. Boston, MA: Birkhauser.

Graimann, B., and Pfurtscheller, G. (2006). "Quantification and visualization of event-related changes in oscillatory brain activity in the time-frequency domain," in *Progress in Brain Research, Vol. 159, Chapter 6* (Amsterdam: Elsevier), 79–97.

Gurve, D., Delisle-Rodriguez, D., Romero-Laiseca, M., Cardoso, V., Loterio, F., Bastos, T., et al. (2020). Subject-specific EEG channel selection using non-negative matrix factorization for lower-limb motor imagery recognition. *J. Neural Eng.* 17, 1–15. doi: 10.1088/1741-2552/ab4dba

Hamedi, M., Salleh, S. H., Noor, A. M., and Mohammad-Rezazadeh, I. (2014). "Neural network-based three-class motor imagery classification using time-domain features for BCI applications," in *2014 IEEE Region 10 Symposium* (Kuala Lumpur: Institute of Electrical and Electronics Engineers Inc.), 204–207.

He, H., and Wu, D. (2018). "Spatial filtering for brain computer interfaces: a comparison between the common spatial pattern and its variant," in *2018 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)* (Qingdao: Institute of Electrical and Electronics Engineers Inc.), 1–6.

He, Y., Eguren, D., Azorín, 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, 1–15. doi: 10.1088/1741-2552/aaa8c0

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

Irimia, D. C., Ortner, R., Poboroniuc, M. S., Ignat, B. E., and Guger, C. (2018). High classification accuracy of a motor imagery based brain-computer interface for stroke rehabilitation training. *Front. Rob. AI* 5, 130. doi: 10.3389/frobt.2018.00130

Kee, C.-Y., Ponnambalam, S. G., and Loo, C.-K. (2017). Binary and multi-class motor imagery using Renyi entropy for feature extraction. *Neural Comput. Appl.* 28, 2051–2062. doi: 10.1007/s00521-016-2178-y

Kevric, J., and Subasi, A. (2017). Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system. *Biomed. Signal Process. Control* 31, 398–406. doi: 10.1016/j.bspc.2016.09.007

Kwak, N.-S., Müller, K.-R., and Lee, S.-W. (2017). A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. *PLoS ONE* 12, e172578. doi: 10.1371/journal.pone.0172578

Ledoit, O., and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* 88, 365–411. doi: 10.1016/S0047-259X(03)00096-4

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, 1–16. doi: 10.1093/gigascience/giz002

Liu, T., Huang, G., Jiang, N., Yao, L., and Zhang, Z. (2020). Reduce brain computer interface inefficiency by combining sensory motor rhythm and movement-related cortical potential features. *J. Neural Eng.* 17, 1–9. doi: 10.1088/1741-2552/ab914d

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., et al. (2018a). A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update. *J. Neural Eng.* 15, 1–28. doi: 10.1088/1741-2552/aab2f2

Lotte, F., and Guan, C. (2011). Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. *IEEE Trans. Biomed. Eng.* 58, 355–362. doi: 10.1109/TBME.2010.2082539

Lotte, F., Nam, C. S., and Nijholt, A. (2018b). "Introduction: evolution of brain-computer interfaces," in *Brain-Computer Interfaces Handbook: Technological and Theoretical Advance* (Boca Raton, FL: Taylor & Francis; CRC Press), 1–11.

Makeig, S. (1993). Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalogr. Clin. Neurophysiol.* 86, 283–293. doi: 10.1016/0013-4694(93)90110-H

Mao, X., Li, M., Li, W., Niu, L., Xian, B., Zeng, M., et al. (2017). Progress in EEG-based brain robot interaction systems xiaoqian. *Comput. Intell. Neurosci.* 2017, 1–25. doi: 10.1155/2017/1742862

Mashat, M. E. M., Lin, C.-T., and Zhang, D. (2019). Effects of task complexity on motor imagery-based brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 2178–2185. doi: 10.1109/TNSRE.2019.2936987

Mendoza-Montoya, O. (2017). *Development of a Hybrid Brain-Computer Interface for Autonomous Systems* (Doctoral dissertation). Freie Universität Berlin.

Naeem, M., Brunner, C., and Pfurtscheller, G. (2009). Dimensionality reduction and channel selection of motor imagery electroencephalographic data. *Comput. Intell. Neurosci.* 2009, 1–8. doi: 10.1155/2009/537504

Ng, A. Y., and Jordan, M. I. (2001). "On Discriminative vs. Generative classifiers: a comparison of logistic regression and naive Bayes," in *Advances in Neural Information Processing Systems, Vol. 14* (Vancouver, BC: MIT Press), 841–848.

Noda, T., Sugimoto, N., Furukawa, J., Sato, M. A., Hyon, S. H., and Morimoto, J. (2012). "Brain-controlled exoskeleton robot for BMI rehabilitation," in *2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)* (Osaka: Institute of Electrical and Electronics Engineers Inc.), 21–27.

Oikonomou, V. P., Georgiadis, K., Liaros, G., Nikolopoulos, S., and Kompatsiaris, I. (2017). "A comparison study on EEG signal processing techniques using motor imagery EEG data," in *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)* (Thessaloniki: Institute of Electrical and Electronics Engineers Inc.), 781–786.

Ortiz, M., Iáñez, E., Contreras-Vidal, J. L., and Azorín, J. M. (2020). Analysis of the EEG rhythms based on the empirical mode decomposition during motor imagery when using a lower-limb exoskeleton. A case study. *Front. Neurorob.* 14, 48. doi: 10.3389/fnbot.2020.00048

Ortner, R., Scharinger, J., Lechner, A., and Guger, C. (2015). "How many people can control a motor imagery based BCI using common spatial patterns?" in *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)* (Montpellier: Institute of Electrical and Electronics Engineers Inc.), 202–205.

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, 1–34. doi: 10.3390/s19061423

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

Podder, P., Mehedi Hasan, M., Rafiqul Islam, M., and Sayeed, M. (2014). Design and implementation of butterworth, chebyshev-I and elliptic filter for speech signal analysis. *Int. J. Comput. Appl.* 98, 12–18. doi: 10.5120/17195-7390

Rejer, I., and Górski, P. (2018). "EEG classification for MI-BCI with independent component analysis," in *Proceedings of the 10th International Conference on Computer Recognition Systems CORES 2017, Vol. 578* (Polanica Zdroj: Springer International Publishing AG), 393–402.

Rodríguez-Bermúdez and García-Laencina,Rodríguez-Bermúdez, G., and García-Laencina, P. J. (2012). Automatic and adaptive classification of electroencephalographic signals for brain computer interfaces. *J. Med. Syst.* 36, 51–63. doi: 10.1007/s10916-012-9893-4

Rodríguez-Ugarte, M., Iáñez, E., Ortíz, M., and Azorín, J. M. (2017). Personalized offline and pseudo-online BCI models to detect pedaling intent. *Front. Neuroinform.* 11, 45. doi: 10.3389/fninf.2017.00045

Saha, S., and Baumert, M. (2020). Intra- and inter-subject variability in EEG-based sensorimotor brain computer interface: a review. *Front. Comput. Neurosci.* 13, 87. doi: 10.3389/fncom.2019.00087

Samuel, O. W., Geng, Y., Li, X., and Li, G. (2017). Towards efficient decoding of multiple classes of motor imagery limb movements based on EEG spectral and time domain descriptors. *J. Med. Syst.* 41, 1–13. doi: 10.1007/s10916-017-0843-z

Singh, A., Hussain, A. A., Lal, S., and Guesgen, H. W. (2021). A comprehensive review on critical issues and possible solutions of motor imagery based electroencephalography brain-computer interface. *Sensors* 21, 1–35. doi: 10.3390/s21062173

Singh, B., Wagatsuma, H., and Natsume, K. (2017). The detection of the rise to stand movements using bereitschaftspotential from scalp electroencephalography (EEG). *SICE J. Control Meas. Syst. Integrat.* 10, 149–155. doi: 10.9746/jcmsi.10.149

Syan, C. S., and Harnarinesingh, R. E. (2010). Comparison of pre-processing and classification techniques for single-trial and multi-trial P300-based brain computer interfaces. *Am. J. Appl. Sci.* 7, 1219–1225. doi: 10.3844/ajassp.2010.1219.1225

Thompson, M. C. (2018). Critiquing the concept of BCI illiteracy. *Sci. Eng. Ethics* 25, 1217–1233. doi: 10.1007/s11948-018-0061-1

Wang, C., Wu, X., Wang, Z., and Ma, Y. (2018). Implementation of a brain-computer interface on a lower-limb exoskeleton. *IEEE Access* 6, 38524–38534. doi: 10.1109/ACCESS.2018.2853628

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

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, R., Allison, B. Z., Ortner, R., Irimia, D. C., Espinosa, A., Lechner, A., et al. (2017). "How many EEG channels are optimal for a motor imagery based BCI for stroke rehabilitation?" in *Converging Clinical and Engineering Research on Neurorehabilitation II, Vol. 15* (Segovia: Springer International Publishing), 1109–1113.

Yu, T., Xiao, J., Wang, F., Zhang, R., Gu, Z., Cichocki, A., et al. (2015). Enhanced motor imagery training using a hybrid BCI with feedback. *IEEE Trans. Biomed. Eng.* 62, 1706–1717. doi: 10.1109/TBME.2015.2402283

Yuan, H., and He, B. (2014). Brain-computer interfaces using sensorimotor rhythms: current state and future perspectives. *IEEE Trans. Biomed. Eng.* 61, 1425–1435. doi: 10.1109/TBME.2014.2312397

Zhang, Z. (2019). "Spectral and time-frequency analysis," in *EEG Signal Processing and Feature Extraction, Chapter 6* (Singapore: Springer Singapore), 89–116.

Zhou, Z.-X., Ming, D., Wan, B.-K., and Cheng, L.-L. (2007). "Event-related EEG-changes during attempted standing up task," in *2007 Joint Meeting of the 6th International Symposium on Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging* (Hangzhou: Institute of Electrical and Electronics Engineers Inc.), 66–69.

frontiers | Frontiers in Neurorobotics

TYPE Original Research
PUBLISHED 12 September 2022
DOI 10.3389/fnbot.2022.971547



OPEN ACCESS

EDITED BY
Luzheng Bi,
Beijing Institute of Technology, China

REVIEWED BY
Zhipeng He,
Sun Yat-sen University, China
Omar Mendoza Montoya,
Monterrey Institute of Technology and Higher Education (ITESM), Mexico

*CORRESPONDENCE
Jonghyun Kim
jonghyunkim@skku.edu

†These authors have contributed equally to this work

RECEIVED 17 June 2022
ACCEPTED 08 August 2022
PUBLISHED 12 September 2022

CITATION
Song M, Jeong H, Kim J, Jang S-H and Kim J (2022) An EEG-based asynchronous MI-BCI system to reduce false positives with a small number of channels for neurorehabilitation: A pilot study. *Front. Neurorobot.* 16:971547. doi: 10.3389/fnbot.2022.971547



# An EEG-based asynchronous MI-BCI system to reduce false positives with a small number of channels for neurorehabilitation: A pilot study

Minsu Song[1†], Hojun Jeong[2†], Jongbum Kim[3], Sung-Ho Jang[4] and Jonghyun Kim[2*]

[1]Department of Medical Device, Korea Institute of Machinery and Materials, Daegu, South Korea, [2]School of Mechanical Engineering, Sungkyunkwan University, Gyeonggi-do, South Korea, [3]Department of Robotics Engineering, Daegu Gyeongbuk Institute of Science and Technology, Daegu, South Korea, [4]Department of Physical Medicine and Rehabilitation, College of Medicine, Yeungnam University, Daegu, South Korea

Many studies have used motor imagery-based brain–computer interface (MI-BCI) systems for stroke rehabilitation to induce brain plasticity. However, they mainly focused on detecting motor imagery but did not consider the effect of false positive (FP) detection. The FP could be a threat to patients with stroke as it can induce wrong-directed brain plasticity that would result in adverse effects. In this study, we proposed a rehabilitative MI-BCI system that focuses on rejecting the FP. To this end, we first identified numerous electroencephalogram (EEG) signals as the causes of the FP, and based on the characteristics of the signals, we designed a novel two-phase classifier using a small number of EEG channels, including the source of the FP. Through experiments with eight healthy participants and nine patients with stroke, our proposed MI-BCI system showed 71.76% selectivity and 13.70% FP rate by using only four EEG channels in the patient group with stroke. Moreover, our system can compensate for day-to-day variations for prolonged session intervals by recalibration. The results suggest that our proposed system, a practical approach for the clinical setting, could improve the therapeutic effect of MI-BCI by reducing the adverse effect of the FP.

KEYWORDS

**brain-computer interface, brain plasticity, contamination, false positive rejection, motor imagery, neurorehabilitation**

## Introduction

Brain–computer interface (BCI) using electroencephalogram (EEG) signals is gaining significance in stroke neurorehabilitation owing to its positive effect on rehabilitation (Friehs et al., 2004; Lebedev and Nicolelis, 2006; Daly and Wolpaw, 2008; Grosse-Wentrup et al., 2011; Young and Tolentino, 2011; Bai et al., 2020). Rehabilitative BCI systems use EEG signals to provide motor-related neurofeedback immediately after the motor intention to generate a planning execution cycle.

By repeating this cycle, brain plasticity can be induced by firing mirror neurons to reorganize the damaged neural circuits in the brain (Bennett et al., 1964; Livingston, 1966; Murphy and Corbett, 2009; Duffau, 2016; Reinkensmeyer et al., 2016; Sasmita et al., 2018). Many studies have shown that these BCI systems can improve the rehabilitation results in patients with stroke by increasing motor function after the training sessions (Friehs et al., 2004; Lebedev and Nicolelis, 2006; Daly and Wolpaw, 2008; Murphy and Corbett, 2009; Grosse-Wentrup et al., 2011; Young and Tolentino, 2011; Bai et al., 2020).

Rehabilitative BCI systems can be classified into two types: synchronous and asynchronous. The synchronous system, which detects target brain signals during a pre-defined time after a visual or sound cue is provided to the users (Pfurtscheller et al., 2003), is inappropriate for training protocols based on the activities of daily living (ADL) as users cannot freely control BCI whenever desired. In contrast, the asynchronous system, which keeps monitoring until the target brain signal is detected (Leeb et al., 2007; Diez et al., 2011; Chae et al., 2012; Kus et al., 2012), is beneficial to rehabilitative BCI as it can provide a more ADL-like experience (Aricò et al., 2020). Since the asynchronous system monitors the brain signal continuously, the feedback of the system can be provided not only in the user-intended time (true positive; TP) but also for the rest of the time (false positive; FP).

For the asynchronous system of rehabilitative BCI, event-related desynchronization (ERD), an attenuating power on certain frequency (alpha and beta) bands, is a typical feature (Pfurtscheller and Lopes Da Silva, 1999). Motor execution (ME) results in ERD; however, most patients with stroke have difficulties performing ME due to motor impairments. Hence, ERD caused by motor imagery (MI) has been regarded as an alternative to ME ERD. This is supported by the following facts. MI ERD shares almost the same activation area and frequency band if the participant performs the exact image of the desired motor task (Miller et al., 2010; Jeong et al., 2019), and motor function recovery after MI training has been reported in patients with stroke (Sun et al., 2016).

It is well known that asynchronous BCI systems are more complicated than their synchronous counterparts (Nicolas-Alonso and Gomez-Gil, 2012; Hramov et al., 2021). Moreover, MI is a complex mental task, namely, intention, tactile, proprioceptive, and visual feeling of the specific motor task (Jeannerod, 2006); thus, MI ERD generated by stroke patients with chronic motor impairments would be weak, leading the asynchronous system to become more challenging. To solve this challenge, a study used an additional electromyogram (EMG) sensor to deliver synchronous-like situations in an asynchronous system (Bhagat et al., 2016); however, this scheme can only be used by a minority of patients with stroke who can provide sufficient EMG on the limb. In contrast, many studies have used spatial pattern-based detection methods, such as spatially applied linear discriminant analysis (Lew et al., 2012; Mrachacz-Kersting et al., 2017), independent component analysis (Ahmadian et al., 2013), and common spatial patterns (Wang et al., 2005; Blankertz et al., 2008), to increase the overall accuracy of BCI based on MI (Hortal et al., 2015; Mrachacz-Kersting et al., 2017); thus, these methods could be applied to overcome the aforementioned challenges. However, it is inappropriate for use in rehabilitative BCI in the clinical environment as they require excessive EEG channels, which leads to a lack of motivation and a decrease in concentration due to fatigue. Note that there were a few attempts to reduce the number of channels; however, they were not successful owing to the significant deterioration of their accuracy (Arvaneh et al., 2011; Tam et al., 2011).

To provide rehabilitative BCI to a broad patient population and to reduce the number of EEG channels, single-channel-based MI detection, also called as brain-switch has been attempted (Müller-Putz et al., 2010; Ge et al., 2014; Camacho and Manian, 2016; Chen et al., 2017; Ko et al., 2017). However, some previous studies just focused on the increase of TP detection and it leads to an increase in FP detection (Camacho and Manian, 2016). The increased number of FPs is more dangerous than decreasing TP of MI from a rehabilitation perspective as the wrong-directed neural cycle could induce inappropriate (pathologic) brain plasticity and interfere with the improvement of MI skills (Barbero and Grosse-Wentrup, 2010; Grosse-Wentrup et al., 2011; Liu et al., 2013; Alimardani et al., 2014; Niazi et al., 2022).

An alternate approach to reduce FPs in ERD is to identify possible sources of the signals that can be confused. The possible sources can be considered non-region of interest (non-ROI) channels for MI task, whereas the region of interest (ROI) channels are interesting channels for investing the effects of the MI task, which is generally contralateral motor area (Kober et al., 2019). Some brain signals (other movement-related signals and cognitive task signals) and the EMG signals generated by eye movement, contraction of the frontalis, temporalis, and neck muscles can be formed as alpha and beta attenuation, similar to MI alpha and beta rhythm (Goncharova et al., 2003). They can be reduced by experimental instructions or easily rejected by using EOG/EMG sensors. The sensory-related signals such as visual evoked potential (VEP) and auditory evoked potential (AEP) also show ERD-like short-lasting attenuation in the alpha and beta bands in non-motor areas (Makeig, 1993; Salenius et al., 1995; Duarte et al., 2009; Toscani et al., 2010; Oppitz et al., 2015). However, the visual/auditory stimuli are generally used in BCI but are difficult to eliminate with external sensors or experimental instructions. Especially, patients with stroke lack attention and require various types of visual and auditory aids to properly concentrate on the rehabilitative BCI (Thaut and McIntosh, 2014; Loetscher and Lincoln, 2019). Therefore, identifying and rejecting these signals could minimize the expected FPs during BCI sessions.

This study addressed a detection and rejection algorithm for a fully asynchronous BCI system using MI ERD. We first identified the sensory-related signals, which could confound individual MI ERD. Then, based on the characteristics of the signals, we designed a classifier for an asynchronous BCI system to detect MI ERD and reject FPs by combining (1) a single-channel-based MI detection in ROI and (2) a non-region of interest (non-ROI) channel-based FP rejection algorithm originating from our previous work (Song et al., 2018). Through experiments with healthy participants and patients with stroke, the validity of the idea of a non-ROI channel was investigated, and the MI detection performance of the proposed classifier was evaluated using both offline simulations and online BCI sessions.

# Methods

## Classifier design for rejecting EEG contamination

The experimental protocol for MI-based BCI generally contains a calibration session before the BCI. In this study, the calibration session not only extracts training data but also screens ROI and non-ROI candidate channels. Note that, in this study, ROI indicates an EEG channel that contains the origin of our interested EEG signal (i.e., MI) and non-ROI indicates the region made of EEG channels outside of our interested area, where we define the sources that EEG contaminations occur. Along with the MI task, the session included several paradigms for screening the source of EEG contamination from sensory (visual/auditory) stimuli: (1) VEP from action observation, (2) VEP from non–motor-related various themed images, and (3) AEP from the auditory cue. The VEP from action observation represents passive action observation in a rehabilitation environment and action recollection that may occur during rest, which refers to unintended cognitive activity that unconsciously reminds the patient of exercise execution. For another VEP, the various themed images were intended to induce unwanted non–motor-related cognition tasks by showing different images for each trial, to mimic the lack of concentration of patients with stroke. The AEP represents miscellaneous auditory cues and sound originated diversions in the rehabilitation environment, which attract attention from the patient. VEPs are known to have a negative peak in the alpha band in the posterior–occipital area (Salenius et al., 1995; Toscani et al., 2010), and the AEP is known to show negative oscillations in the alpha band in the temporal and midline areas (Makeig, 1993; Duarte et al., 2009; Oppitz et al., 2015). Considering these characteristics, we designed paradigms to reveal the time-frequency patterns of EEG contamination that can be used to develop classifiers for rejecting them.

## Selection of ROI and non-ROI EEG channels

Instead of applying conventional spatial filters with many EEG channels to overcome the limitation of weak MI ERD, we used a small number of channels and algorithm following characteristic of EEG signals: the EEG signals radially flow through the scalp-like electrocortical ripple, affecting nearby electrodes (Salenius et al., 1995; Mcfarland et al., 1997). If MI ERD appears in the EEG channel located in the MI-related area (ROI channel), we could find similarly desynchronized power of signals on the nearby channels; however, the signals would be weaker than the ROI channel due to skin impedance. In contrast, if stronger power desynchronization appeared in the channel located outside of the MI-related area (non-ROI channel) when ERD is detected in the ROI channel, the detected power desynchronization can be regarded as pseudo-MI ERD, originating from the non-ROI channel due to EEG contamination. This means that the use of proper non-ROI channel information could enable the effective discrimination of pseudo-MI ERD without using many EEG channels.

To implement the approach above, ROI and non-ROI channels for individuals should be carefully selected. Figure 1A is a diagram of a workflow of channel selection, which is colored by a group of tasks that can be represented by characteristic example figures (Figures 1B,C). Figure 1B shows a characteristic example of the selection process of the ROI channel. Based on the event-related spectral perturbation (ERSP) map of the MI task, we extracted the data during the MI task in five frequency bands from 8 Hz to 28 Hz (mu and beta) with 4 Hz intervals (yellow boxes in Figures 1A,B). Note that EEGLAB functions were used to calculate ERSP and sinusoidal wavelet (short-time DFT) transform was used for the computation of spectral estimate (Delorme and Makeig, 2004). Baseline correction is applied to ERSP based on the pre-stimulus segment (−4 to −2 s from the cue). Then, the averaged ERSP in the task period for each band data was drawn into a topographical map (Figure 1B), and a channel that was closest to the area, where it showed the lowest averaged ERSP power among all five maps, was chosen as the ROI channel, the source channel of MI ERD (yellow dotted circle on the topographical map in Figure 1B). Finally, we drew the ERSP map of the chosen channel to specify the frequency band of the MI ERD (green box in Figure 1B). The ROI channel selection was double-checked by drawing the topographical map of the frequency band specified to see whether the selected channel showed the strongest ERD (Figure 1B).

The candidates for the non-ROI channel were also chosen as follows: To find the source channel of EEG contamination, the inducers of contamination, visual and auditory stimulation, were provided to the participants. Based on the determined frequency band in ROI selection, the ERSP map and topographical maps of the ROI for the paradigms were drawn. Then, the sources showing the lowest averaged ERSP values in the topographical maps were chosen as candidates for the

10.3389/fnbot.2022.971547

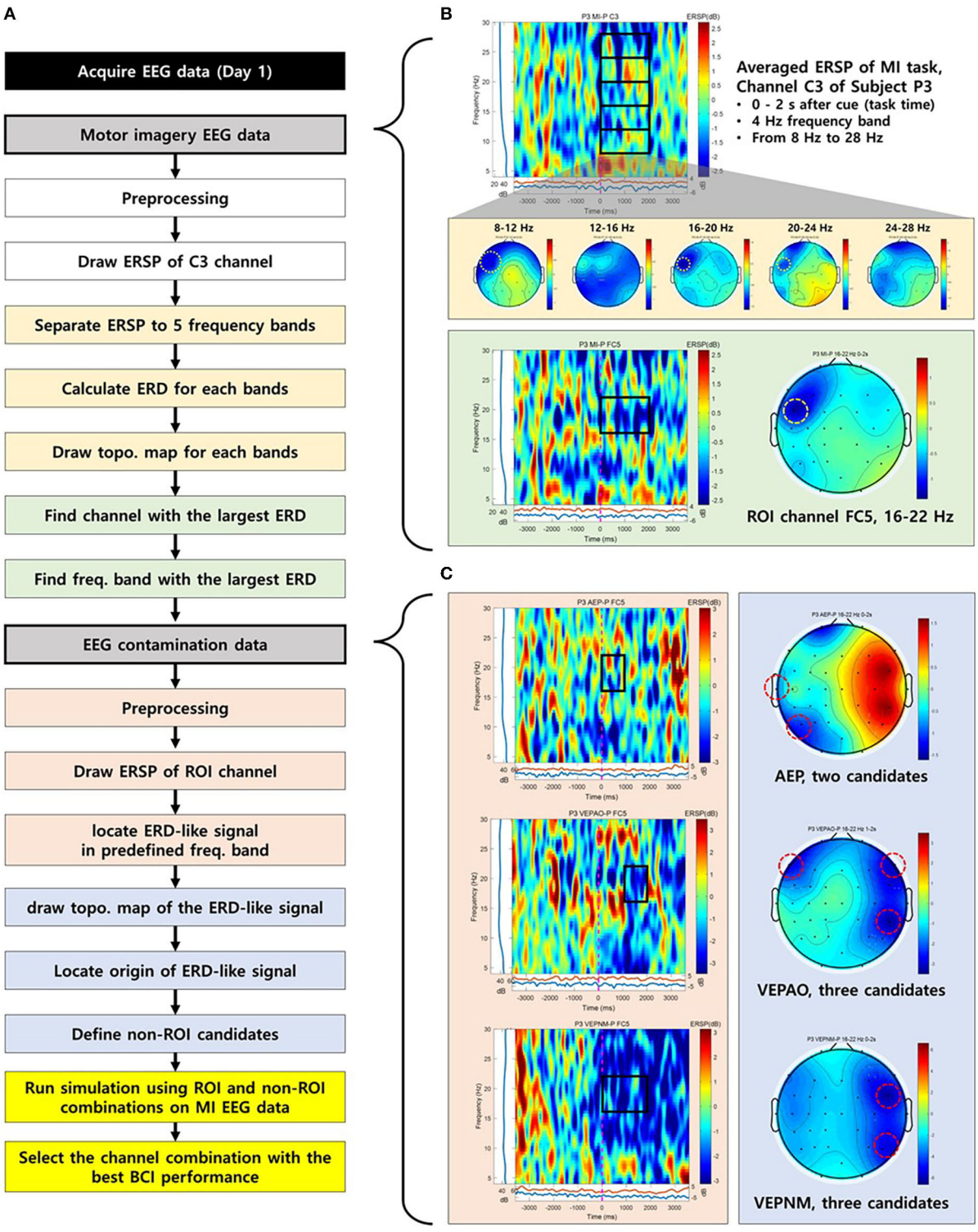



FIGURE 1
The process to select a region of interest (ROI) and non-ROI candidate channels of a representative participant (P3). **(A)** The flow chart for the process to select ROI and non-ROI candidate channels. **(B)** The process to select a ROI. The yellow dotted circle on the topographical map represents the selected ROI. **(C)** Non-ROI candidate channels. The red dotted circle on the topographical map represents the source of electroencephalogram (EEG) contamination and candidate channels for non-ROI.

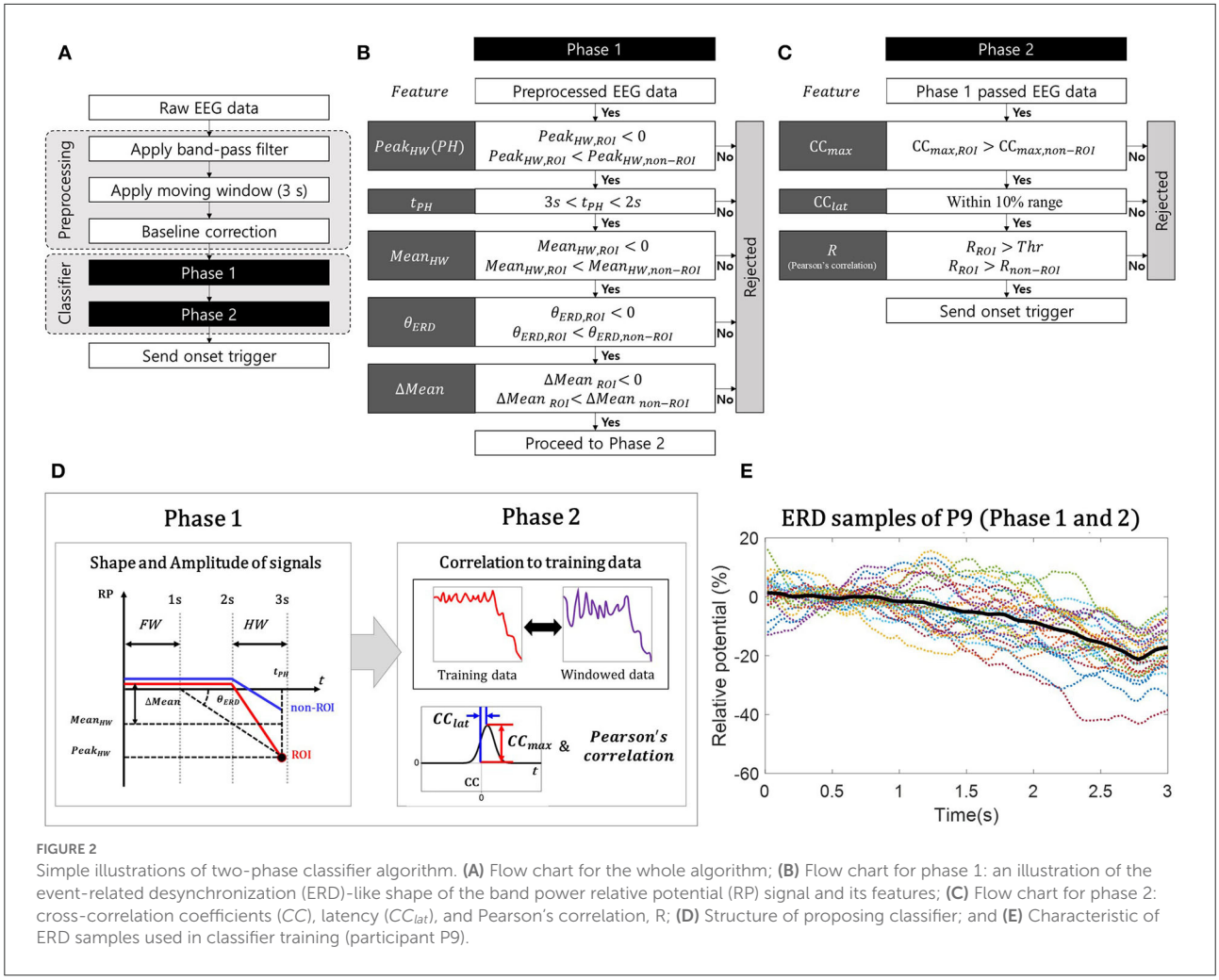



FIGURE 2
Simple illustrations of two-phase classifier algorithm. **(A)** Flow chart for the whole algorithm; **(B)** Flow chart for phase 1: an illustration of the event-related desynchronization (ERD)-like shape of the band power relative potential (RP) signal and its features; **(C)** Flow chart for phase 2: cross-correlation coefficients (*CC*), latency ($CC_{lat}$), and Pearson's correlation, R; **(D)** Structure of proposing classifier; and **(E)** Characteristic of ERD samples used in classifier training (participant P9).

non-ROI channel (Figure 1C). Offline MI-BCI simulation was performed in a proposed classifier with multiple combinations of candidates to determine the best combination with the highest FP rejection rate as a non-ROI channel.

## Classifier structure

The classifier, which contains the detection algorithm of MI ERD and rejection algorithm of pseudo-MI ERD, was designed with the following hypothesis: The ERD signals generated by EEG contamination spread-like radial waves from the non-ROI channel and affect the ROI channel as pseudo-MI ERD. Based on this hypothesis, the proposed classifier was built by comparing the non-ROI channels obtained from various EEG contamination paradigms with the ROI channels obtained from the MI paradigm. It should be noted that we validated this hypothesis using experimental data in this study. To detect the desired feature of MI ERD by rejecting pseudo-MI ERD, the proposed classifier adopted a two-phase structure, MI-ERD detection method in a single ROI channel, and FP rejection method with non-ROI channels as illustrated in Figure 2A.

We first specified the characteristics of the 3-s-long windowed signal that we wanted to detect as MI ERD. For the feature of the proposed classifier, the relative potential (RP) was used for the ERD calculation method (Pfurtscheller and Lopes Da Silva, 1999). The length of the windowed signal was 3 s, to distinguish the feature by including the signal during "rest" before the ERD begins. The desired characteristic of the signal was basically to contain the negative peak of ERD; however, we also considered that the peak is located at the hind area of the windowed signal, as illustrated in Figure 2D, to minimize its detection latency (Song et al., 2018). Based on this characteristic, the first phase was to distinguish the promising MI ERD from the incoming windowed signals using the features for describing the shape and amplitude of the signal. After dividing the signal into the front and hind areas (Figure 2D), we calculated the following features: the minimum peak in the hind area ($Peak_{HW}$) and its timing ($t_{PH}$), mean band power value of

the hind area ($Mean_{HW}$)), mean band power difference between the areas ($Mean = Mean_{FW} - Mean_{HW}$), and decline angle of ERD ($\theta_{ERD}$), as summarized in Figure 2B. The decline angle was calculated as follows:

$$\theta_{ERD} = \text{atan}\left(\frac{Peak_{HW} - Mean_{FW}}{t_{PH} - t_{FW}}\right) \quad (1)$$

where $t_{FW}$ denotes the timing of the baseline, determined as the end of the front area. Using the features of the ROI and non-ROI channels, we constructed the conditions to find the promising MI ERD, as shown in Figures 2B,C, based on the following statement originating from the hypothesis: the signal on the ROI channel would be MI ERD when it has the lowest peak ($Peak_{HW,ROI} < 0$) and average amplitude ($Mean_{HW}$) in the hind window ($3s < t_{PH} < 2s$), with larger amplitude ($Peak_{HW,ROI} < Peak_{HW,non-ROI}$, and $Mean_{HW,ROI} < Mean_{HW,non-ROI}$), and larger reduced amplitude ($Mean = Mean_{FW} - Mean_{HW}$) and deeper decline angle ($\theta_{ERD} < 0$) compared with pseudo-MI ERD on the non-ROI channels ($Mean_{ROI} < Mean_{non-ROI}$, and $\theta_{ERD,ROI} < \theta_{ERD,non-ROI}$).

In the second phase, the following correlation-related features between each windowed signal (on the ROI and non-ROI channels) and the actual MI ERD signal collected as training data were calculated: the maximum value of the cross-correlation coefficients (*CCmax*), latencies of the coefficients (*CClat*) (Lewis, 1995; Sadeghian and Moradi, 2008; Chandaka et al., 2009; Siuly and Li, 2012), and normalized Pearson's correlation coefficients (*R*) (Pearson, 1895). The similarity between the training data and windowed signal on the ROI was evaluated using *CClat* and *R* (Figure 2C). Moreover, by comparing *CCmax* and *R* from the ROI with those from the non-ROI, we checked whether the windowed signal on the ROI was more similar to the training data than the windowed signal on the non-ROI (Figure 2C). Note that the correlation features, *CCmax* and *R*, are insensitive to the magnitude of the signals. Since our algorithm relies on a relative comparison between ROI and non-ROIs, we focused on detecting the similarity with training data from calibration session rather than its magnitude.

## Participants and experimental design

The experiment comprised two sessions. In the calibration session, we measured the participant's EEG behavior when performing a targeted MI task and when exposed to different sensory stimulations. After a few days, the MI-BCI session was conducted based on the classifier that was calibrated for each participant using the data obtained from the calibration session, to evaluate the performance of the proposed MI-BCI system.

Eight healthy young adults (four men, four women, average age: 22.8 ± 4.1 years) and nine patients with stroke (seven men, two women, average age: 56.7 ± 7.9 years) who were in the chronic stage post-stroke for 124.2 ± 42.7 months were recruited in this study. All healthy participants were right-handed with no history of brain–nervous system injuries or neurological diseases. Seven patients had a hemorrhage in the left hemisphere, resulting in hemiplegia on the right upper limb, while the other two patients had the opposite. Note that three of the eight healthy participants did not participate in the MI-BCI session owing to personal reasons. With the approval of the institutional review board (DGIST-170721-HR-025-08), all participants voluntarily signed their consent after the experimental details were provided.

Figure 3 shows the experimental setup. The experiment was performed in a quiet and air-conditioned room with minimal visual artifacts blocked by partitions (Figure 3A). A custom hand exoskeleton robot was used to provide the participant's fist open/close motor feedback for the MI-BCI session (Bae et al., 2017; Lee et al., 2017) (Figure 3B). The classifier for the MI-BCI session was implemented using customized OpenVibe (Inria, France), Python, and LabVIEW (National Instruments, USA) codes.

We used a 32-channel EEG (Active Two EEG, BioSemi Co. Ltd., Netherlands), in which electrodes were attached to a 64-channel EEG cap (FLASH type EEG holder, Shimadzu Corp., Japan) based on a 10–20 system. The channel locations were widespread and densely distributed on the left motor cortex to locate the ROI channel when the participants imagined right-handed movement, as illustrated in Figure 4. For better convenience for the patient group, two channels on each temporal area (T9 and T10) were relocated to the left and medial parietal areas (P1, POz) (Figure 4).

## Protocols

### Calibration session

The participant sat on a chair with an armrest in front of the monitor leveled on the eye level (Figure 3A). Each paradigm shared the same block design as that described in Figure 3C. During the rest period, the other cognitive actions were restricted. The single beep sound was played for 0.25 s at the beginning of the task period to notify the participants. The task period of each paradigm had its event cue and certain cognitive or motor tasks for 2 s (Figure 3C). In the MI paradigm, the participants were asked to perform pure MI of opening and closing the fist for a single time after a visual cue (green circle) with closing their fist of the right hand (or affected hand for the patient group) during the task period (Figure 3C).

For the paradigm to measure AEP, the participant was asked to relax and to concentrate on the sound cue (a beep sound repeated four times at 0.5 s intervals) provided during the task period (Figure 3C). For VEP, we measured two types of VEP: VEP with action observation and non–motor-related VEP. In

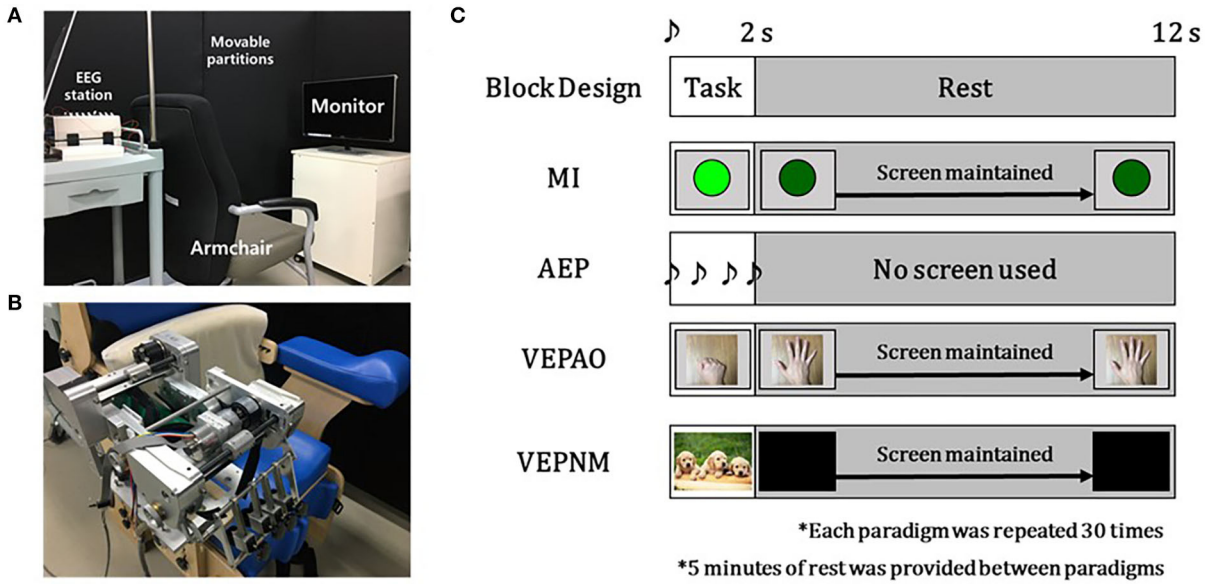



FIGURE 3
Experiment setup and block design of the Paradigms. **(A)** Experiment setup for the measurement session. **(B)** Exoskeleton hand rehabilitation robot setup for the motor imagery-based brain–computer interface (MI-BCI) session. **(C)** Block design-based experiment paradigms. (VEPAO: VEP with action observation; VEPNM: non–motor-related VEP).

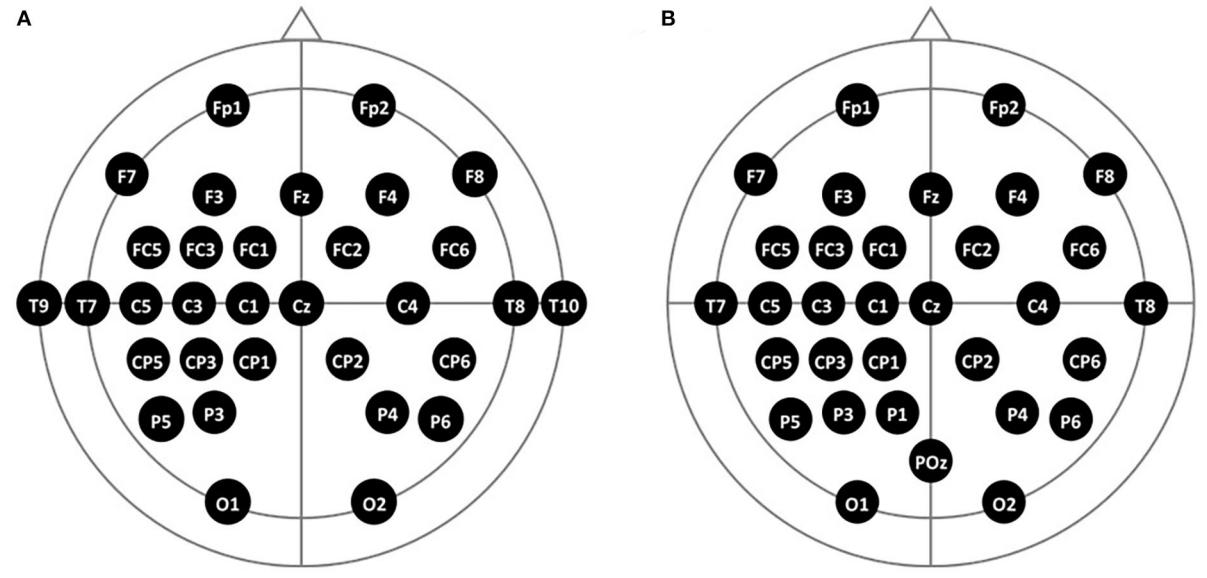



FIGURE 4
The 10–20 system-based 32-channel locations of two participant groups. **(A)** Locations of the healthy group and **(B)** locations of the stroke group.

both paradigms for the VEPs, the participants were asked to relax without any movement and to watch the monitor. In each VEP paradigm, the monitor displayed a top-view image of the fist open/close for the former VEP and random images not related to the motor task for the latter VEP (Figure 3C), which turned black during the rest period for proper relaxation.

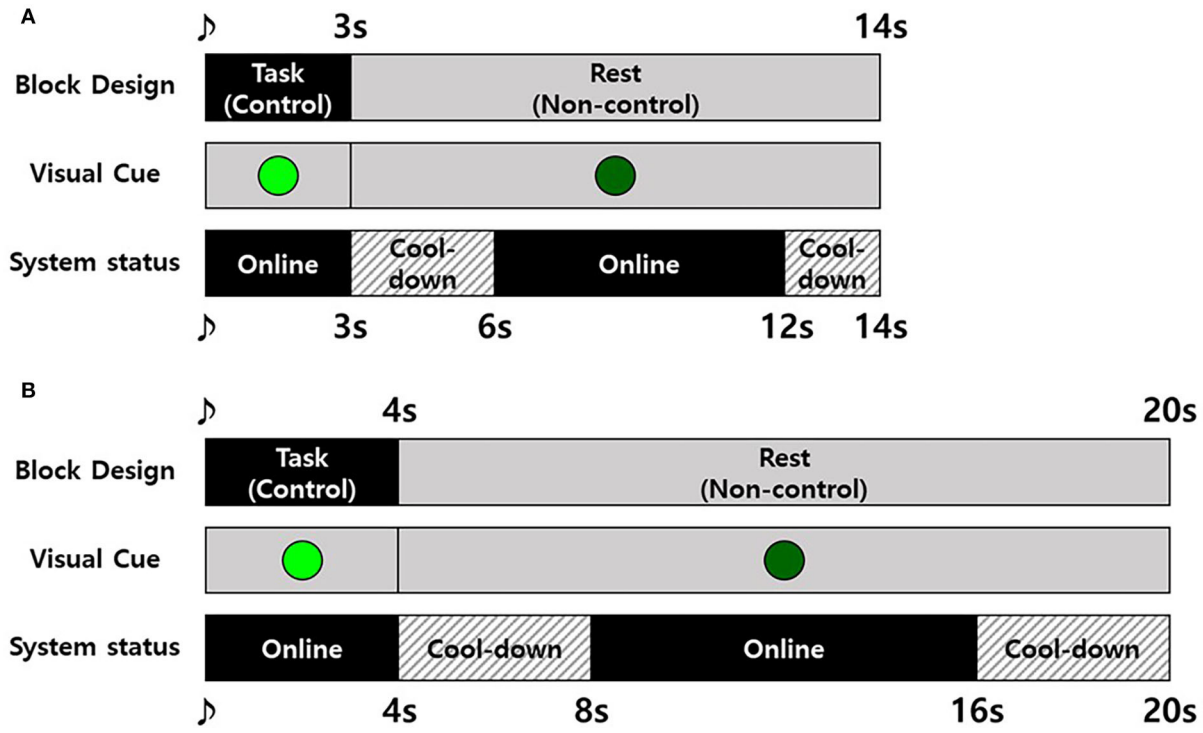



FIGURE 5
Experiment block design for the motor imagery-based brain–computer interface (MI-BCI) session. **(A)** The healthy participant group and **(B)** stroke participant group.

During the calibration session, the participants participated in paradigms in the following order: MI was performed first to avoid the influence of other EEG contamination paradigms, and following AEP, VEP with action observation, and non–motor-related VEP were performed in randomized order. The block for each paradigm was repeated 30 times for 7 min, with 5 min of rest between each paradigm.

## MI-BCI session

First, the participants were asked to perform the MI paradigm of the calibration session as practice, as the last MI was performed a few days ago. During the paradigm, the operator monitored the classifier and slightly adjusted the threshold for Pearson's product-moment correlation coefficient (PPMCC) from the second phase of the classifier to compensate for the day-to-day variation. After the MI paradigm, the hand exoskeleton robot was attached to the chair armrest, the right side for healthy participants and affected side for patients with stroke (Figure 3B), and the participants were put on the robot for the MI-BCI session.

In the MI-BCI session, participants were instructed to perform MI to open and close their fist for a single time during the task (control) period and asked to remain as calm as possible in the rest (non-control) period (Leeb et al., 2007) following block design, as shown in Figure 5. Here, we used a synchronous block design for evaluating asynchronous MI-BCI systems as there are no observable signs to confirm the execution of MI. The BCI system went to the offline state (cool-down) immediately before and after the task period (Figure 5). Except for the cool-down status, the system was online to wait for the detection of MI, and the detection resulted in movement feedback by the robot. The cool-down was used to calm the brain signals after MI and/or movement feedback. If the detection occurred during the task period, it was considered TP, and if it occurred during the rest period, it was considered FP. The number of TPs and that of FPs were counted to evaluate the performance of the proposed MI-BCI system.

The experimental block designs for the healthy and patient groups were slightly different (Figure 5) because the patients with stroke felt more difficulty performing MI and took more time to concentrate than healthy participants. To compensate for the burden increase due to a longer task period, the time of rest period and cool-down were also increased to maintain the time ratio between control, non-control, and cool-down.

## Data analysis

### Data processing

To investigate the source of EEG contamination, raw EEG data from the calibration session were epoched based on the time information of the cue (−4 to 4 s from the cue). Epoched data were normalized by subtracting the baseline, which is the averaged data from −4 to −2 s based on the cue. A baseline-corrected data epoch was used to plot the time-frequency information map and ERSP map. As mentioned in Method.1.a, five frequency bands were extracted from the ERSP map and used to draw topographical maps to select the participant-specific frequency band, ROI channel, and candidates of the non-ROI channel.

To validate our hypothesis for the proposed classifier, RP was calculated the same as the ERD calculation for the feature of the proposed classifier as follows:

$$(A - R)/R \times 100 \tag{2}$$

where $A$ is the power of filtered data and $R$ is the power of preceding baseline data (Pfurtscheller and Lopes Da Silva, 1999). To calculate the RP, 8 s data epochs (−4 to 4 s from the cue) were extracted based on the time information in the filtered data, and the band power of the epoch was normalized by the average power of the baseline data (−4 to −2 s before the task cue) (Song et al., 2018; Song and Kim, 2019). The mean and standard deviation for the peak amplitude of RP from the ROI and non-ROI channels in the three paradigms were compared by quantitative comparison. For statistical analysis, we performed a paired t-test on the peak values of ROI and non-ROI. For the non-ROI in the comparison, the amplitude of the negative peak for non-ROI candidate channels was averaged for the MI paradigm, and the non-ROI candidate with the largest peak amplitude was selected for the other EEG contamination paradigms.

After screening the frequency bands and channels, to extract the training data, raw EEG signals were resampled to 64 Hz and band-pass filtered using the determined frequency band in the ROI selection. The filtered signal was sliced to 3 s moving window, overlapping every 20 ms (50 Hz). We then applied a phase 1 classifier to each moving window. The data that fit the classifier and its lower peak existed between 0 and 2 s after the cue were selected as the training data. The average of the training data was used for the phase 2 classifier. During the MI-BCI session, online EEG signals were sliced to a 3 s moving window (50 Hz) and applied to the proposed classifier.

### Performance evaluation

We evaluated the performance of the classifier during both the calibration and MI-BCI sessions. For the calibration session, we obtained offline simulation results of the classifier, and the actual online performance of the classifier was analyzed for the MI-BCI session. Based on the number of TPs and FPs, the performance was evaluated using sensitivity (Altman and Bland, 1994; Bhagat et al., 2016), selectivity (Altman and Bland, 1994; Chae et al., 2012), FP rate (Pfurtscheller et al., 2003; Leeb et al., 2007; Chae et al., 2012; Lew et al., 2012; Liu et al., 2013; Bhagat et al., 2016; Mrachacz-Kersting et al., 2017), and FP per minute (FPM) (Li et al., 2013; Rodriguez-Ugarte et al., 2017), as follows:

$$Sensitivity = \frac{Number\ of\ TP}{Number\ of\ trials} \times 100\ (\%) \tag{3}$$

$$Selectivity = \frac{Number\ of\ TP}{Number\ of\ total\ detections} \times 100\ (\%) \tag{4}$$

$$FP\ rate = \frac{Number\ of\ FP}{Number\ of\ trials} \times 100\ (\%) \tag{5}$$

$$FP\ per\ minute = \frac{Number\ of\ FP}{Total\ elapsed\ rest\ period} \times 100\ (\%) \tag{6}$$

For the quantitative comparison between offline and online results, we applied a paired t-test on the number of TPs and FPs from offline results of the MI paradigm in the calibration session and online results of the MI-BCI session. Note that as the number of participants in the healthy group was too small for statistical analysis, the analysis was applied in the case of a stroke group ($n = 9$) and in the case of all participants who participated in the MI-BCI session ($n = 14$).

To investigate the effect of the key idea in the proposed classifier for FP rejection, the classification without non-ROI channels was simulated using MI-BCI session data of healthy participants and stroke patient groups and compared with online classification results in the same MI-BCI session by calculating the rejection rate, as follows:

$$rejection\ rate = \frac{Number\ of\ rejected\ FP}{Number\ of\ FP\ without\ non\ ROI\ channels} \times 100\ (\%) \tag{7}$$

# Results

## Calibration session

Figure 6 shows the group analysis results for three paradigms (AEP, VEP with action observation, and non–motor-related VEP) of the calibration session to screen for EEG contamination.

Figure 7 shows the comparison results of the relative potential between the ROI and non-ROI candidate channels. Figure 7A illustrates the statistical analysis of the peak ERDs for each participant group during each paradigm. This result implies that the non-ROI candidate channels show stronger ERD signals when EEG contamination occurs due to visual/auditory stimulation. It should be noted that the largest peak amplitude of the non-ROI candidate channel exceeded the amplitude of

10.3389/fnbot.2022.971547

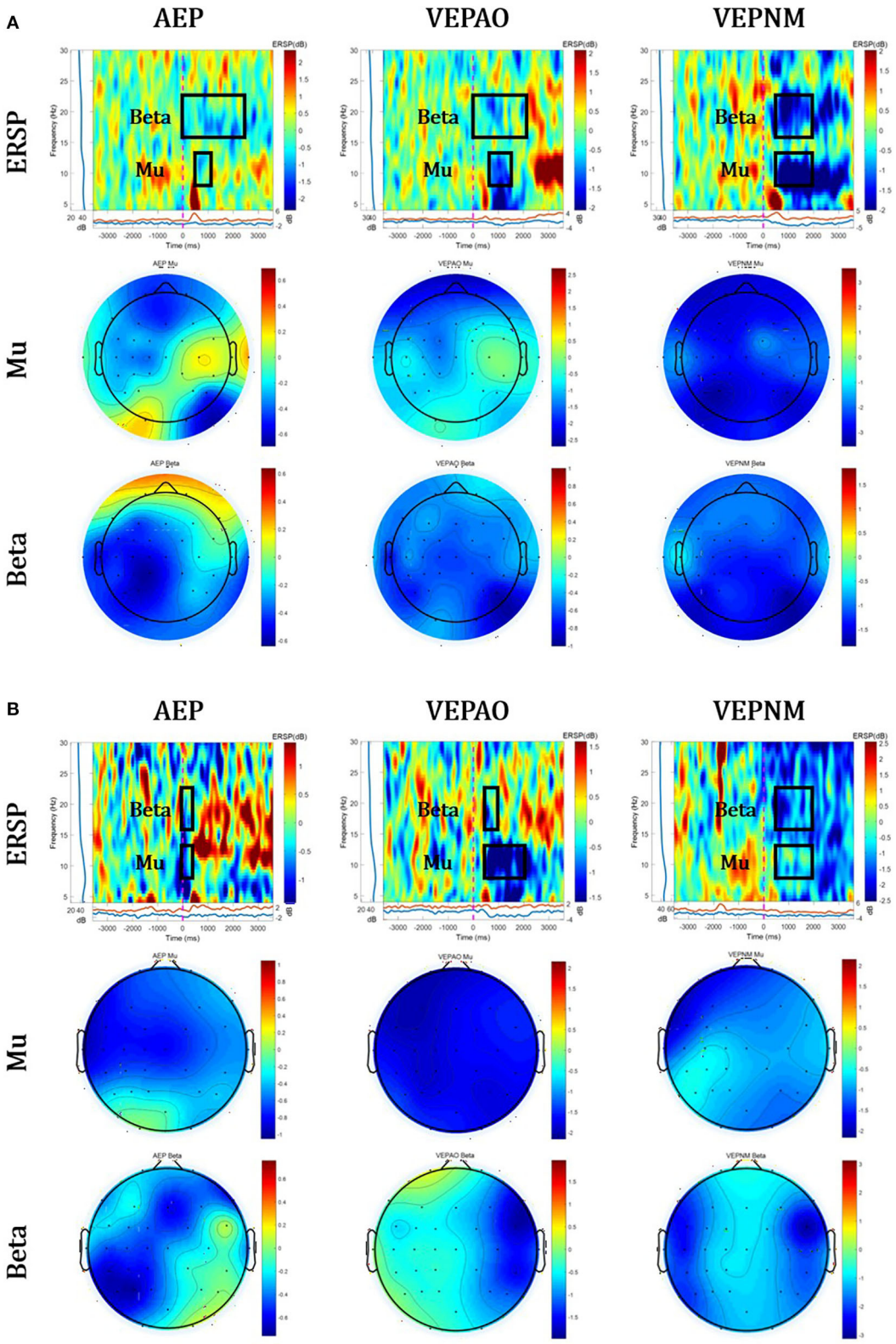



FIGURE 6
Group analysis of event-related spectral perturbation (ERSP) and its topographical map of Mu and Beta ERD. The red circles represent the average source area of EEG contaminations. **(A)** The healthy group and **(B)** stroke patient group (VEPAO: VEP with action observation; VEPNM: non–motor-related VEP).

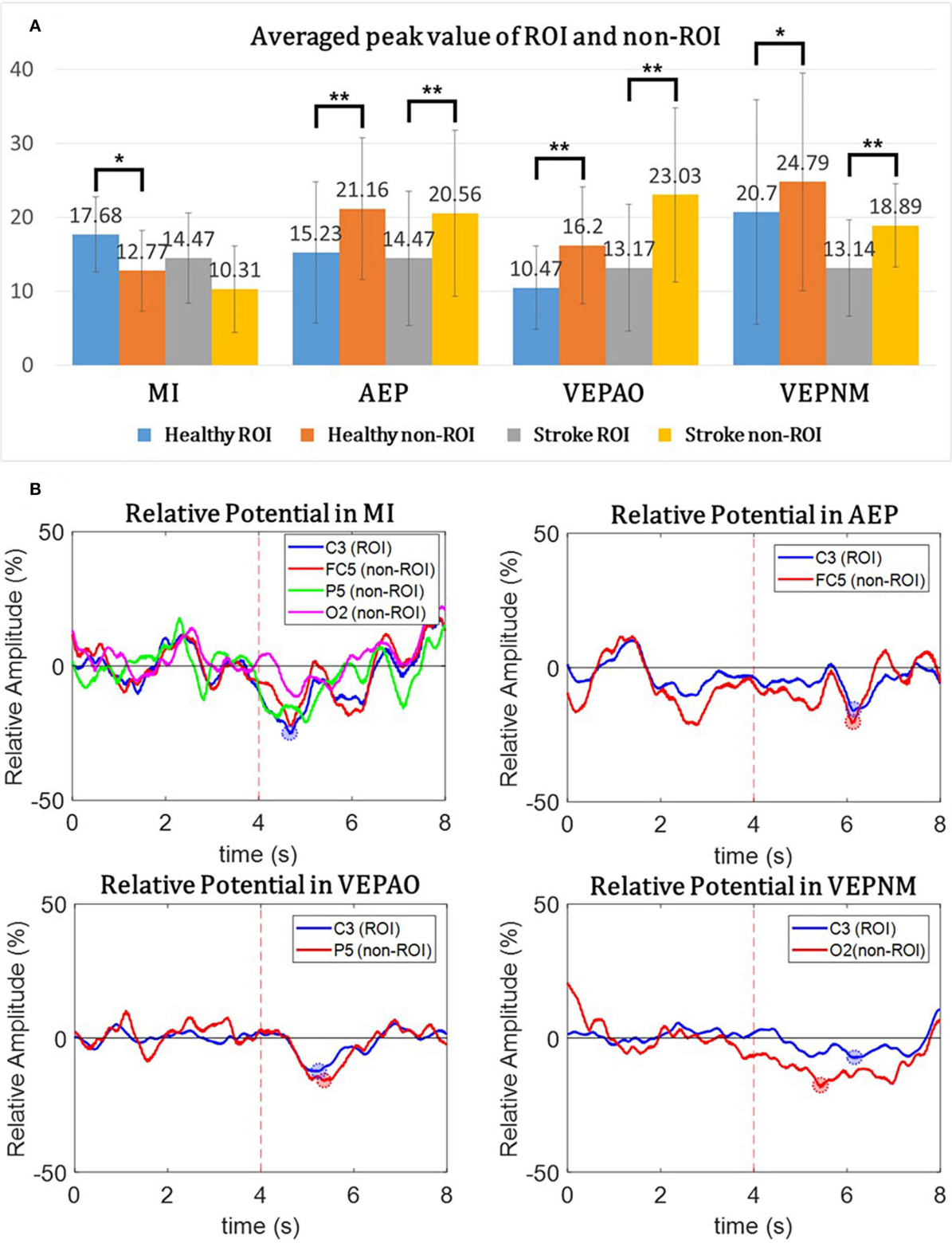



FIGURE 7
The relative potential from the region of interest (ROI) and non-ROI candidate channels, and the mean peak value of overall participants. **(A)** The average peak value of the ROI and non-ROI candidate channels for each participant group during each paradigm. The paired t-test is performed between the ROI and non-ROI candidate channels. Statistically mild significance ($p<0.08$) is described using a single asterisk (*) and statistical significance ($p < 0.05$) using a double asterisk (**). The candidates of the non-ROI channel for the MI paradigm are an average of three different non-ROI candidate channels. **(B)** The relative potential of the characteristic participant. The blue line represents the data of the ROI channels, and the red line represents the data from the non-ROI candidate channels. VEPAO, VEP with action observation; VEPNM, non–motor-related VEP.

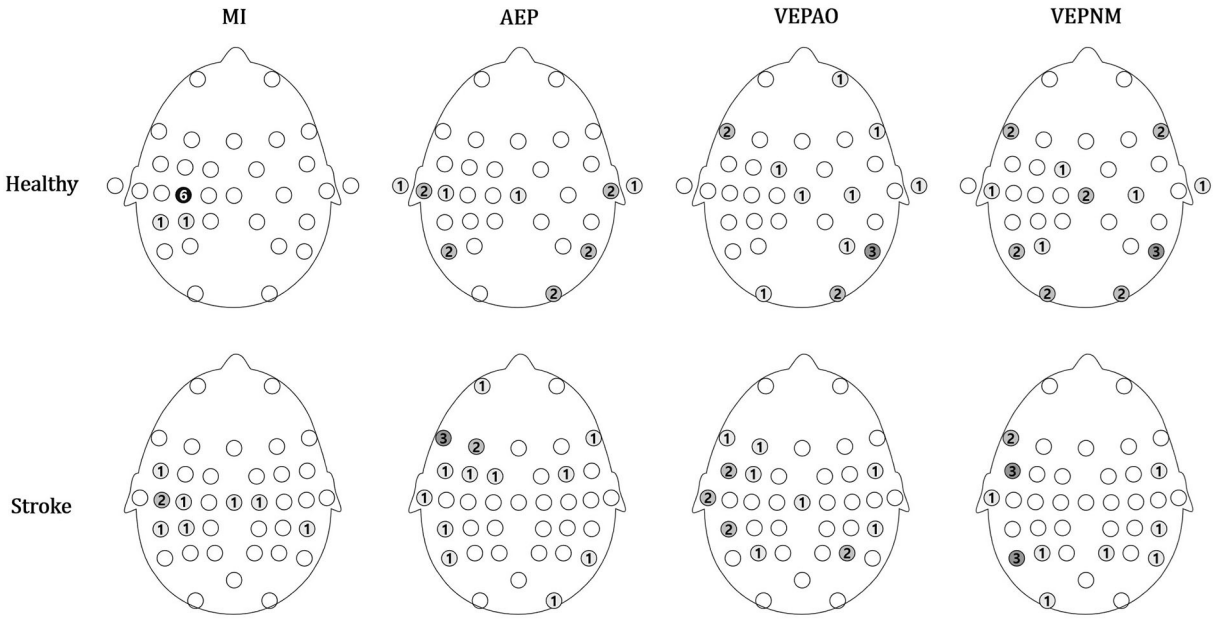



FIGURE 8
Group analysis of the topographical map for the number of regions of interest (ROI) and non-ROI candidate channels. The number of ROI channels is shown in blue, and the number of non-ROI candidate channels is shown in red. VEPAO, VEP with action observation; VEPNM, non–motor-related VEP.

the ROI channel for all paradigms except the MI paradigm in all trials and subjects, and the difference in the amplitude was statistically significant ($p<0.05$). As shown in Figure 7B, the relative potentials in the ROI channel tended to show larger ERD compared with that in the selected non-ROI candidate channels after the cue during the MI paradigm. The contamination paradigms tend vice versa.

The distributions of the chosen ROI and non-ROI candidate channels are illustrated as color maps in Figure 8. Here, the more the channels are concentrated, the darker the color. The ROI channels for the patient group were distributed in channels near the motor area (FC5, Cz, C3, C5, and CP5 for the right affected participants and C2 and CP6 for left affected participants), while most of the ROI channels were located on the motor cortex (C3 for six participants) and few were located in the somatosensory cortex (CP3 and CP5) in the healthy group (Figure 8). The candidates of the non-ROI channel for the healthy group were distributed on each diagonal end of the scalp, and those in the stroke group were mostly distributed in the left frontal (FT7 and FC5) and parietal lobes (P5).

## MI-BCI session

Tables 1, 2 describes the performance of the classifier in the offline simulation results and online MI-BCI sessions for all participants. The sensitivity was below 30% for both the healthy and stroke groups, and the FP rate was 12.67% in the healthy group and 8.52% in the stroke group. The non-ROI channels were widespread but mostly located in the temporal, parietal, and occipital lobes, as we targeted AEP and VEPs. For the patient group, the non-ROI channels were located similar to that of the healthy group; however, these channels were also located in the premotor cortex. Tables 2A,B describe the performance of the online MI-BCI session. For the healthy group, both sensitivity and selectivity increased compared with the offline simulation. For the patient group, the sensitivity increased; however, the selectivity slightly decreased.

Figure 9 describes the mean and standard deviation of parameters during the offline analysis of the MI paradigm (day 1) and online MI-BCI session (day 2). For the stroke group, the TPs showed a statistically significant increase ($p = 0.015$), and the FPs increased but were not significant ($p = 0.071$). The sensitivity also showed a significant increase as the sensitivity was dominantly related to the number of TPs. For all participants in the MI-BCI session ($n = 14$), the TPs showed a significant increase ($p = 0.007$), while the FPs did not ($p = 0.246$). The selectivity showed no significant difference between sessions for both groups ($p = 0.792$, $p = 0.359$ for the patient group and all participants' group each).

Table 3 describes the rejected number of FPs due to the non-ROI channel-based method for the healthy and stroke patient

10.3389/fnbot.2022.971547

TABLE 1 Performance of the classifier in offline simulation.

| Subject | Freq. band | ROI | Non-ROIs | TP | FP | FPR | FPM | Selec. | Sens. |
|---|---|---|---|---|---|---|---|---|---|
| **(A) Offline simulation result of healthy group** | | | | | | | | | |
| S1 | 17–21 Hz | CP5 | F7, Cz, P5 | 10 | 3 | 10 | 0.6 | 76.92 | 33.33 |
| S2 | 20–25 Hz | C3 | C4, P3, P6 | 13 | 8 | 26.67 | 1.6 | 61.90 | 43.33 |
| S3 | 9–12 Hz | C3 | F7, P3, P4 | 10 | 4 | 13.33 | 0.8 | 71.43 | 33.33 |
| S4 | 17–22 Hz | CP3 | Cz, T8, O1 | 8 | 3 | 10 | 0.6 | 72.73 | 26.67 |
| S5 | 24–28 Hz | C3 | T7, P6, O2 | 4 | 1 | 3.33 | 0.2 | 80 | 13.33 |
| S6 | 10–14 Hz | C3 | F7, T7, Cz | 8 | 2 | 6.67 | 0.4 | 66.67 | 26.67 |
| S7 | 18–22 Hz | C3 | T10, P6, O2 | 4 | 0 | 0 | 0 | 100 | 13.33 |
| S8 | 8–12 Hz | C3 | F7, FC1, P5 | 4 | 2 | 6.67 | 0.4 | 66.67 | 13.33 |
| Average | – | – | – | 7.62 | 2.87 | 9.58 | 0.58 | 72.62 | 25.42 |
| **(B) Offline simulation result of stroke group** | | | | | | | | | |
| P1 | 8–10 Hz | Cz | FC5, CP5, P5 | 8 | 5 | 16.67 | 1 | 61.54 | 26.67 |
| P2 | 8–12 Hz | C5 | P3, P4, F7 | 9 | 4 | 13.33 | 0.8 | 69.23 | 30 |
| P3 | 16–22 Hz | FC5 | F7, FC6, CP6 | 8 | 1 | 3.33 | 0.2 | 88.89 | 26.67 |
| P4 | 10–12 Hz | C3 | FC5, P5, O2 | 6 | 3 | 10 | 0.6 | 66.67 | 20 |
| P5 | 8–10 Hz | Cz | F7, T7, P6 | 9 | 3 | 10 | 0.6 | 75 | 30 |
| P6 | 19–22 Hz | C5 | F7, F3, CP6 | 4 | 2 | 6.67 | 0.4 | 66.67 | 13.33 |
| P7 | 8–12 Hz | CP6(L) | FC4, P5, O1 | 4 | 2 | 6.67 | 0.4 | 66.67 | 13.33 |
| P8 | 14–17 Hz | CP5 | F3, FC5, Cz | 10 | 2 | 6.67 | 0.4 | 83.33 | 33.33 |
| P9 | 17–23 Hz | C2 (L) | FC5, FC1, P5 | 4 | 1 | 3.33 | 0.2 | 80 | 13.33 |
| Average | – | – | – | 6.88 | 2.55 | 8.52 | 0.51 | 72.94 | 22.96 |

TP, true positives; FP, false positives; FPR, false positive ratio; FPM, false positive per minute.

TABLE 2 Performance online MI-BCI (A) in the healthy group and (B) in the stroke patients' group.

| Subject | *Session interval (days) | TP | FP | FPR | FPM | Selec. | Sens. |
|---|---|---|---|---|---|---|---|
| **(A) Online result of healthy group** | | | | | | | |
| S1 | 40 | 10 | 4 | 13.33 | 1.33 | 71.43 | 33.33 |
| S2 | 29 | 15 | 7 | 23.33 | 2.33 | 68.18 | 50 |
| S3 | 22 | 10 | 3 | 10 | 1 | 76.92 | 33.33 |
| S4 | 23 | 7 | 1 | 3.33 | 0.33 | 87.5 | 23.33 |
| S5 | 18 | 10 | 0 | 0 | 0 | 100 | 33.33 |
| Average | 26.4 ± 8.6 | 10.4 | 3 | 10 | 1 | 77.61 | 34.67 |
| **(B) Online result of stroke group** | | | | | | | |
| P1 | 14 | 14 | 6 | 20 | 1.5 | 70 | 46.67 |
| P2 | 14 | 9 | 3 | 10 | 0.75 | 75 | 30 |
| P3 | 14 | 10 | 3 | 10 | 0.75 | 76.92 | 33.33 |
| P4 | 20 | 6 | 2 | 6.67 | 0.5 | 75 | 20 |
| P5 | 14 | 8 | 3 | 10 | 0.75 | 72.72 | 26.67 |
| P6 | 7 | 8 | 3 | 10 | 0.75 | 72.72 | 26.67 |
| P7 | 8 | 12 | 7 | 23.3 | 1.75 | 63.16 | 40 |
| P8 | 10 | 15 | 4 | 13.33 | 1 | 78.95 | 50 |
| P9 | 5 | 12 | 6 | 20 | 1.5 | 66.67 | 40 |
| Average | 11.8 ± 4.7 | 10.44 | 4.11 | 13.70 | 1.03 | 71.76 | 34.81 |

TP, true positives; FP, false positives; FPR, false positive ratio; FPM, false positive per minute.
*Session interval indicates the interval between calibration session and online MI-BCI session.

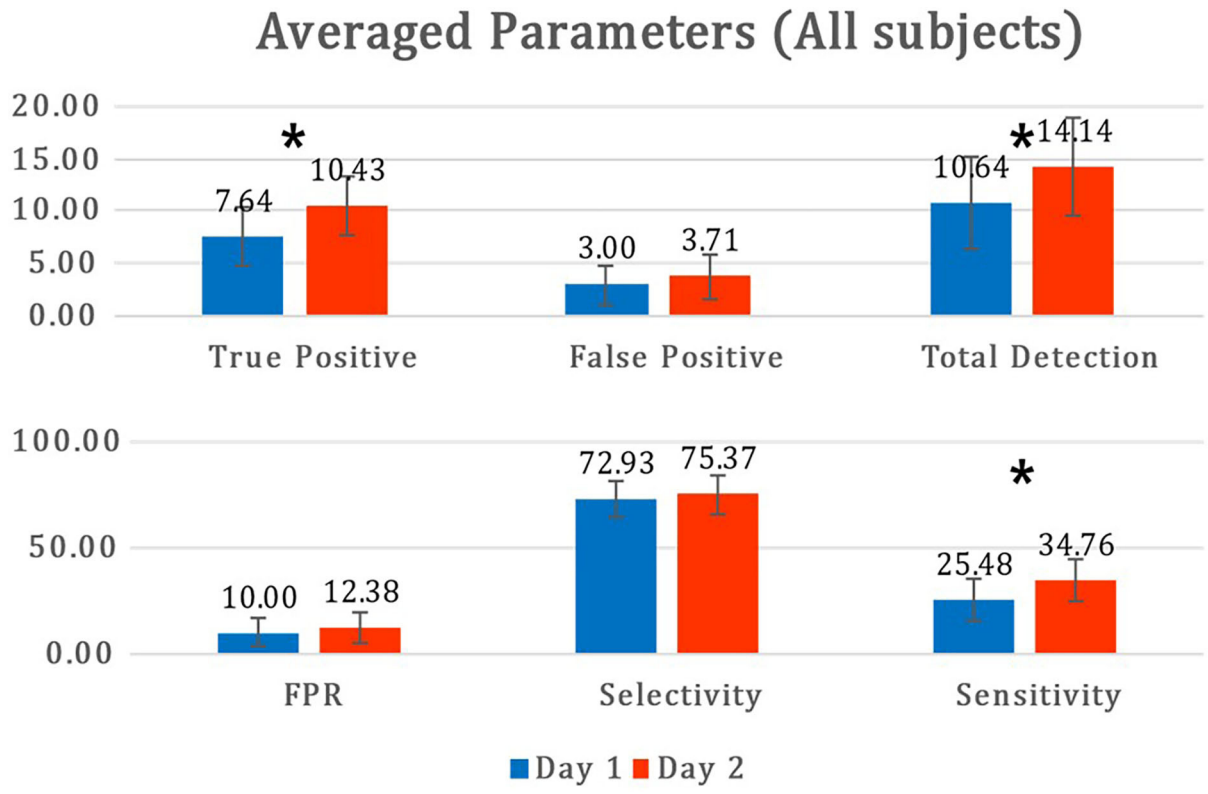



FIGURE 9
Mean and standard deviations of parameters between day 1 and day 2. The asterisk (*) is statistically significant ($p<0.05$) between days. FPR, false positive ratio.

group. The most rejected FPs were 21, which is 75% of the FPs on P7. The highest and lowest rejection rates were 100 and 53.33%, respectively. The total rejection rate was 76.04% for all FPs.

## Discussion

This study aimed to reduce FPs during rehabilitative MI-BCI that could result in wrong-directed brain plasticity. To this end, we proposed a classifier that contains single-channel-based MI detection and FP rejection using non-ROI channels.

As shown in the ERSP and band power of Figures 6, 7, EEG contamination elements (AEP, VEP with action observation, and non–motor-related VEP) affect the mu and beta bands in the motor area. The candidates of the non-ROI channel show larger amplitudes than the ROI channel when EEG contamination occurs due to visual or auditory stimulations. This means that the contamination elements originating from the non-ROI candidates can result in desynchronization at the ROI channel, and it could be detected as FP in the ROI. Despite the desynchronization at the ROI, there were significant band power differences between the ROI and non-ROI candidates (Figure 7). Therefore, it is feasible to find and reject contamination based on power differences.

Since the sources of the contamination elements on the scalp were differently distributed, it is essential to identify the sources through individual calibration (Figure 8). The MI signals of patients with stroke were also distributed around the motor area, which explains why finding a participant-specific ROI channel is an important task to improve MI-BCI performance for clinical application.

The experimental results suggest that our proposed MI-BCI system has a good FP rejection performance online, with a rejection rate of over 75%. As shown in Figure 9, both TP and FP tend to increase on day 2, compared with day 1. However, the mean selectivity did not show a significant difference and even increased slightly. This implies that our proposed algorithm is robust as it rejected a certain ratio of FP despite a significant session interval (day-to-day variation), even a month.

Our method consists of two phases of classifiers: (1) hand-crafted detection algorithm, and (2) correlation-based detection algorithm. The design intention was to use the phase 1 algorithm on reducing the number of sample windows and provides a synchronous-like state for assisting the phase 2 algorithm. To

TABLE 3 The number of false positives before and after applying the non-ROI channel-based method of the proposed classifier, and FP rejection in the online MI-BCI session.

| Subject | TP | FP (pre) | FP (post) | Rejected | Rejection rate (%) |
|---|---|---|---|---|---|
| S1 | 10 | 13 | 4 | 9 | 69.23 |
| S2 | 15 | 15 | 7 | 8 | 53.33 |
| S3 | 10 | 10 | 3 | 7 | 70 |
| S4 | 7 | 3 | 1 | 2 | 66.67 |
| S5 | 10 | 2 | 0 | 2 | 100 |
| P1 | 14 | 14 | 6 | 8 | 57.14 |
| P2 | 9 | 20 | 3 | 17 | 85 |
| P3 | 10 | 17 | 3 | 14 | 82.35 |
| P4 | 6 | 18 | 2 | 16 | 88.89 |
| P5 | 8 | 16 | 3 | 13 | 81.25 |
| P6 | 8 | 18 | 3 | 15 | 83.33 |
| P7 | 12 | 28 | 7 | 21 | 75 |
| P8 | 15 | 20 | 4 | 16 | 80 |
| P9 | 12 | 23 | 6 | 17 | 73.91 |
| Average | 146 | 217 | 52 | 165 | 76.04 |

TP, true positive; FP (pre), false positive before applying non-ROI channel-based method; FP (post), false positive after applying non-ROI channel-based method.

evaluate if the purpose was fulfilled, we performed an offline performance test of the phase 1 algorithm combined with non–ROI-based false positive rejection using day 1 MI data (Table 4). The results show relatively high sensitivity near 75% for both subject groups, with selectivity below but near 50%. These numbers indicate that the phase 1 algorithm combined with non-ROI technique provides a 50% chance of distinguishing true and false positives for the phase 2 algorithm with 25% of data loss.

Several studies have applied BCI systems to patients with stroke (Hortal et al., 2015; Bhagat et al., 2016; Mrachacz-Kersting et al., 2017; Miladinović et al., 2020; Niazi et al., 2022). Table 5 compares the proposed method with existing studies. The main difference is that our system relies on a hand-crafted feature classifier, which is discriminated approach compared to spatial pattern-based machine-learning methods. Our method is originated from single-channel-based MI ERD detection, which cannot apply any spatial pattern-based machine-learning approach, but can only rely on the time-frequency aspect of the signal. Since our target signal has been clearly justified and it follows with the neurophysiological agreement throughout many studies (Pfurtscheller and Lopes Da Silva, 1999; Pfurtscheller et al., 2003; Kus et al., 2012; Nicolas-Alonso and Gomez-Gil, 2012; Sun et al., 2016; Jeong et al., 2019). We decided to use its nature to design features and algorithms without leaving them to machine learning; since machine-learning methods depend on the amount of training data, they are inappropriate to induce a decision rule like the proposed one, which consists of a large number of required features, out of such small datasets (patient's data) (Choi et al., 2018; Lee et al., 2021).

An advantage of our method is the use of a small number of channels. It uses the smallest number (four) of channels in the MI-BCI session after a one-time calibration with 32 channels (Table 5). This can reduce the setting time of EEG for MI-BCI, which results in minimal fatigue for the patients and clinicians as well as better time efficiency of rehabilitative MI-BCI therapy. Since patients with stroke generally lose their attention and motivation easily, fatigue due to heavy EEG settings for the therapy would be critical for clinical application. All existing studies relied on spatial-based methods, such as LDA with spatial features (Blankertz et al., 2008; Lew et al., 2012) and LPP-LDA (Mrachacz-Kersting et al., 2017), Source Power Co-Modulation (Meinel et al., 2019), Spectrally Weighted Common Spatial Filter (CSP) (Wei et al., 2008), Filter Bank CSP (Park and Chung, 2019), and CSP with likelihood ratio method (Niazi et al., 2011, 2022); thus, these approaches suffer from heavy MI-BCI performance deterioration under a small number of channels (Arvaneh et al., 2011; Tam et al., 2011). Another advantage is the rare occurrence of FP during MI-BCI. For a fair comparison, we checked the false positives and experimental paradigms to calculate the FP rate (equation 5). The study by Hortal et al. (2015), Bhagat et al. (2016), Mrachacz-Kersting et al. (2017), and Miladinović et al. (2020) used the same calculation method as our study to report the FP rate. The study by Niazi et al. (2022) reported false positive per minute, true positive rate, and percentage of false positives over true positive. We inversely calculated the false positive rate using given parameters. Our

TABLE 4 Performance of the phase 1 classifier in offline simulation.

**(A) Offline simulation result of healthy group**

| Subject | TP | FP | FPR | FPM | Selec. | Sens. |
|---|---|---|---|---|---|---|
| S1 | 23 | 20 | 0.67 | 4 | 0.53 | 0.77 |
| S2 | 24 | 32 | 1.67 | 6.4 | 0.43 | 0.8 |
| S3 | 21 | 18 | 0.6 | 3.6 | 0.54 | 0.7 |
| S4 | 21 | 16 | 0.53 | 3.2 | 0.57 | 0.7 |
| S5 | 26 | 28 | 0.93 | 5.6 | 0.48 | 0.87 |
| S6 | 24 | 27 | 0.9 | 5.4 | 0.47 | 0.8 |
| S7 | 22 | 20 | 0.67 | 4 | 0.52 | 0.73 |
| S8 | 22 | 24 | 0.8 | 4.8 | 0.48 | 0.73 |
| Average | 183 | 185 | 0.77 | 4.1 | 0.50 | 0.76 |

**(B) Offline simulation result of stroke group**

| | | | | | | |
|---|---|---|---|---|---|---|
| P1 | 22 | 17 | 0.57 | 3.4 | 0.56 | 0.73 |
| P2 | 23 | 31 | 1.03 | 6.2 | 0.43 | 0.77 |
| P3 | 25 | 31 | 1.03 | 6.2 | 0.45 | 0.83 |
| P4 | 19 | 22 | 0.73 | 4.4 | 0.46 | 0.63 |
| P5 | 24 | 24 | 0.8 | 4.8 | 0.5 | 0.8 |
| P6 | 26 | 25 | 0.83 | 5 | 0.51 | 0.87 |
| P7 | 21 | 22 | 0.73 | 4.4 | 0.49 | 0.7 |
| P8 | 21 | 23 | 0.77 | 4.6 | 0.48 | 0.7 |
| P9 | 20 | 30 | 1 | 6 | 0.4 | 0.67 |
| Average | 201 | 225 | 0.83 | 5 | 0.47 | 0.74 |

TP, true positives; FP, false positives; FPR, false positive ratio; FPM, false positive per minute.

system showed a 10% FP rate in the healthy group and 13.7% in the stroke group, which is the lowest FP rate compared with other existing studies (Table 5). It should be noted that our FP rates were obtained under the longest session interval between calibration and MI-BCI (Table 5). Along with the classifier used, the paradigm design also affected the occurrence of FP. Although a short task period is a disadvantage as patients with stroke generally require a longer time for MI due to chronic motor impairments, we used a shorter task period than other existing studies as extending the task period would cause misclassification between FP and TP. Moreover, the possibility of FP increases as the rest period becomes longer; however, our total rest period is the longest. Therefore, we believe that the FP rejection performance of the proposed system outperforms other existing studies, even though our paradigm design has disadvantages for FP.

The positive and negative effects of FP remain controversial (Levine et al., 2000; Barbero and Grosse-Wentrup, 2010; Alimardani et al., 2014). The exact effect of FP has not been determined; however, some studies claim that FP could be useful for improving MI in naïve BCI users (Alimardani et al., 2014). However, the goal of rehabilitative MI-BCI systems for patients with stroke is to guide them to perform correct MI based on neurophysiology to stimulate direct brain plasticity and improve the neuro-circuits. The most effective way to achieve this goal is by applying MI-BCI asynchronously; however, in this situation, the participant and/or clinician cannot notice whether the robotic feedback comes from TP or FP, without any cue. Since the nature of the training experience dictates the nature of neural plasticity (Kleim and Jones, 2008), if the patient is repeatedly exposed to the feedback induced by FP, it might lead to inappropriate brain plasticity. Therefore, reducing and minimizing FP would be essential for MI-BCI systems for neurorehabilitation.

Many patients with stroke who participated in the experiment commented that moving the rehabilitation robot due to TPs induced the feeling of body ownership like "I was controlling the robot hand" (Altman and Bland, 1994; Botvinick and Cohen, 1998; Michielsen et al., 2010; Evans and Blanke, 2013; Liang et al., 2016; Sun et al., 2016). In contrast, they also commented that the wrong robotic feedback due to FPs caused them to lose the agency and ownership of their hand, and this was frustrating and unpleasant. These comments show that rejecting FP is important for maintaining body ownership and agency in the MI-BCI system. However, some patients experienced anxiety and loss of interest when MI detection did

TABLE 5 Comparison of the method, experiment, and results with other studies applied to patients with stroke.

| Study | Task | Feature | Method | # obtained Calibration data/time spent | Session intervals | Subjects | # of channels used | Experiment paradigm | | | Performances | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Task time (s) | Rest time (s) | T:R ratio/ total rest time per set (s) | Type of performance evaluation dataset | Sensitivity (%) | FPR (%) |
| Hortal et al. (2015) | Motor Imagery, Grasping | ERD | SVM Spatial Pattern | 304 task data 912 rest data/ 10 min. | 1 day (no interval) | 3 healthy 5 stroke | 16 | 10 | 10 | 1:1/100 | Online | H 82.9 S 45 | H 19.2break S 15.0 |
| Bhagat et al. (2016) | Motor Execution, Elbow | MRCP/EMG | SVM Spatial Pattern | 160–320 data/53 min./day (not mentioned) | 2-days of measurement 1-day calibration 2-days of the online trials (1 day interval) | 4 stroke | 60 | 15 | 5 | 1:0.33/100 | Online | Day4 62.7 Day5 67.1 | Day4 27.74 Day5 27.5 |
| Mrachacz-Kersting et al. (2017) | Motor Execution, Reaching | MRCP | LPP-LDA | 30 data/15 mins | 1 day (no interval) | 6 stroke | 9 | 4 | 7 | 1:1.75/210 | Online | 1st 68.6 2nd 68.6 | 1st 33.6 2nd 21.2 |
| Miladinović et al. (2020) | Motor Imagery, Grasping | ERD | (1) Source Power Co-Modulation*, (2) Spectrally Weighted CSP**, (3) Filter Bank CSP*** | 35–40 data per session (day), 15 sessions/ 10 mins | 1 day (no interval) | 5 stroke | 15 | 5 | 2.1–2.8 | 1:0.48–0.56 /78.75–105 | Offline | 1) 83.0 2) 83.8 3) 85.1 | 1) 16.9break 2) 15.5 3) 15.5 |
| Niazi et al. (2022) | Motor Execution, Ankle dor-siflexion | MRCP | Spatial pattern, likelihood | 50 dataset, 6–10 min/(not mentioned) | 2 day, ≥24 h | 9 stroke | 9 | 1.5 | 3–4 | 1:2–2.6/150 | Online | 82.68 | 15.25 |
| Proposed | Motor imagery, Grasping | ERD | Shape, correlation | 30 dataset/day 1:30 min. Day 2: 3 min. | 2 days (avg. 17 days interval) | 5 healthy 9 stroke | Day 1: 32 Day 2: 4 | 4 | 8 | 1:2/240 | Online | H 35 S 34.8 | H 10 S 13.70 |

*Meinel et al. (2019), **Wei et al. (2008), ***Park and Chung (2019).





10.3389/fnbot.2022.971547

not occur. This implies that low sensitivity could negatively affect MI-BCI therapy for some participants and work as an obstacle.

The proposed asynchronous MI-BCI system showed state-of-the-art FP rejection performance, while the sensitivity of the system was decreased compared with existing spatial-based approaches. The classifier design based on the characteristics of EEG contamination led our MI-BCI system to use a minimal number of channels for detecting MI ERD and for rejecting FP. Moreover, the classifier was insensitive to day-to-day variations. Therefore, we believe that the proposed system fits the conditions for practical use in clinics, fast setup time due to the small number of channels, and reliable performance owing to its insensitive day-to-day variation.

Our study mainly considered the EEG contamination on motor-related VEP, non–motor-related VEP, and AEP. However, our proposing algorithm can be applied to other EEG contaminations due to sensory stimuli, such as sound generated from medical devices and visual distractions in a rehabilitation facility, which show pseudo-MI ERD-like behavior. It is because the design of the algorithm intended to reject all pseudo-MI ERD originated from non-ROI channels which could be easily extended by screening the candidates of the non-ROI channel.

Despite many benefits, the limitation of our proposed system is its low sensitivity. The study of brain-switch on healthy subjects address that ERS-based single-channel MI detection could be achieved to a sensitivity of 59.2%, with FP rate below 10%, but sensitivity decreased to 28.4%, while ERD was used as a feature (Pfurtscheller and Solis-Escalante, 2008). This result suggests that ERD is a challenging feature compared to ERS. However, since our goal is to detect movement intention at the right timing to induce brain plasticity, ERS was inappropriate due to its delayed appearance. The averaged peak value during the MI paradigm in Figure 7A shows the low significance of ERD between ROI and non-ROI, which illustrates that some ERD from ROI channels might be rejected by the non–ROI-based classifier in some cases. This might imply that our non-ROI selection needs to be improved to consider the MI paradigm. Moreover, the second phase in the classifier, which was intended to detect samples with similar patterns to training data, might be too conservative because we only used 30 training data for each subject. In the viewpoint of inducing brain plasticity, whereas the FP-rejected asynchronous MI-BCI system induced cortical plasticity more than a typical self-paced asynchronous system with FP (Niazi et al., 2022), the correlation between sensitivity and cortical plasticity showed a negative association with significance (Jochumsen et al., 2019). The pieces of literature could illustrate that sensitivity does not significantly affect cortical activation compared to the FP rate. Nevertheless, the goal of the asynchronous MI-BCI system detects users' movement intention and gives them feedback at the proper time. Therefore, further research is needed to determine the appropriate level of sensitivity to encourage users, and it needs to be improved for better sensitivity in future. In future studies, we would like to evaluate the cortical activation difference between intensity-focused algorithms (high sensitivity, low FP) and specificity-focused algorithms (high FP) during MI-BCI training to verify the more important factor. Since the study did not evaluate the actual effect in the patients after the MI-BCI sessions, a long-term follow-up study would also become our next objective.

## Conclusion

This study aimed to develop the asynchronous MI-BCI system for neurorehabilitation use for people with stroke. To apply EEG-based BCI, we prioritized two factors: (1) small number of channels for user convenience and (2) reducing the number of FP to prevent wrong-directed brain plasticity and rehabilitation. We developed an MI ERD detection and FP rejection algorithm based on the time-frequency characteristics of MI ERD and EEG contaminations, with rippling characteristics of EEG signals. We categorized three EEG contaminations to assume as sources of FP: VEP during action observation, VEP during random images, and AEP with simple beep sound. These contaminations are easily found in the clinical rehabilitation environment, where our future system will be applied. We localized the surface source of each contamination and used a combination of those channels to reject FPs.

The designed algorithm was validated online for eight healthy subjects and nine patients with hemiplegic stroke. As a result, we showed the best FP rate compared to other asynchronous MI-BCI studies (10% for healthy subjects, 13.70% for patient subjects with stroke), while 76.04% of FP was rejected by applying a non-ROI channel method to single-channel detection-based algorithm. However, our system also showed the least sensitivity. The proposed system matched our intended objective; to reject FP conservatively. However, the sensitivity of the proposed system should be improved by further research.

## Data availability statement

The datasets presented in this article are prohibited due to privacy protection. Further inquiries can be directed to the corresponding author.

## Ethics statement

The studies involving human participants were reviewed and approved by Daegu Gyeongbuk Institute of Science and Technology. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

JhK supervised the study, conceptualized designed the study, and acquired the funding and provided the resources for the study, and finalized the manuscript. MS and HJ implemented the proposed system and designed the experiments, acquired the data, and drafted the original manuscript. MS, HJ, S-HJ, and JbK recruited subjects and prepared IRB for the experiments. MS, HJ, and JbK processed, analyzed the data from the experiments, and interpreted the results from the data. All authors read and revised the manuscript and approved the final manuscript for publication.

## Funding

This research was supported in part by the National Research Foundation of Korea (NRF) and grant funded by the Korean government (MSIP) (No. 2022 R1 A2 C1008150).

## Acknowledgments

The authors would like to thank all subjects who volunteered for the study.

## Conflict of interest

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

## 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

Ahmadian, P., Sanei, S., Ascari, L., González-Villanueva, L., and Umiltà, M. A. (2013). Constrained blind source extraction of readiness potentials from EEG. *IEEE Trans. Neural. Syst. Rehabil. Eng.* 21, 567–575. doi: 10.1109/TNSRE.2012.2227278

Alimardani, M., Nishio, S., and Ishiguro, H. (2014). Effect of biased feedback on motor imagery learning in BCI-teleoperation system. *Front. Syst. Neurosci.* 8, 52. doi: 10.3389/fnsys.2014.00052

Altman, D. G., and Bland, J. M. (1994). Statistics notes: diagnostic tests 1: sensitivity and specificity. *BMJ* 308, 1552 doi: 10.1136/bmj.308.6943.1552

Aricò, P., Sciaraffa, N., and Babiloni, F. (2020). Brain–computer interfaces: toward a daily life employment. *Brain Sci.* 10, 157. doi: 10.3390/brainsci10030157

Arvaneh, M., Guan, C., Ang, K. K., and Quek, C. (2011). Optimizing the channel selection and classification accuracy in EEG-Based BCI. *IEEE Trans. Biomed. Eng.* 58, 1865–1873. doi: 10.1109/TBME.2011.2131142

Bae, S. J., Jang, S. H., Seo, J. P., and Chang, P. H. (2017). The optimal speed for cortical activation of passive wrist movements performed by a rehabilitation robot: a functional NIRS study. *Front. Hum. Neurosci.* 2017, 194. doi: 10.3389/fnhum.2017.00194

Bai, Z., Fong, K. N. K., Zhang, J. J., Chan, J., and Ting, K. H. (2020). Immediate and long-term effects of BCI-based rehabilitation of the upper extremity after stroke: a systematic review and meta-analysis. *J. Neuroeng. Rehabil.* 17, 57. doi: 10.1186/s12984-020-00686-2

Barbero, A., and Grosse-Wentrup, M. (2010). Biased feedback in brain-computer interfaces. *J. Neuroeng. Rehabil.* 7, 34. doi: 10.1186/1743-0003-7-34

Bennett, E. L., Diamond, M. C., Krech, D., and Rosenzweig, M. R. (1964). Chemical and anatomical plasticity of brain. *Science* 146, 610–619. doi: 10.1126/science.146.3644.610

Bhagat, N. A., Venkatakrishnan, A., Abibullaev, B., Artz, E. J., Yozbatiran, N., Blank, A. A., et al. (2016). Design and optimization of an EEG-based brain machine interface (BMI) to an upper-limb exoskeleton for stroke survivors. *Front. Neurosci.* 10, 122. doi: 10.3389/fnins.2016.00122

Blankertz, B., Losch, F., Krauledat, M., Dornhege, G., Curio, G., and Müller, K. (2008). The berlin brain–computer interface: accurate performance from first-session in BCI-naive subjects. *IEEE Trans. Biomed. Eng.* 55, 2452–2462. doi: 10.1109/TBME.2008.923152

Botvinick, M., and Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*. 391, 756. doi: 10.1038/35784

Camacho, J., and Manian, V. (2016). Real-time single channel EEG motor imagery based brain computer interface. *World Autom. Congr. Proc.* 2016, 73. doi: 10.1109/WAC.2016.7582973

Chae, Y., Jeong, J., and Jo, S. (2012). Toward brain-actuated humanoid robots: asynchronous direct control using an EEG-based BCI. *IEEE Trans. Robot.* 28, 1131–1144. doi: 10.1109/TRO.2012.2201310

Chandaka, S., Chatterjee, A., and Munshi, S. (2009). Cross-correlation aided support vector machine classifier for classification of EEG signals. *Expert. Syst. Appl.* 36, 1329–1336. doi: 10.1016/j.eswa.2007.11.017

Chen, S.-C., Chen, Y.-J., Zaeni, I. A. E., and Wu, C.-M. (2017). A single-channel SSVEP-based BCI with a fuzzy feature threshold algorithm in a maze game. *Int. J. Fuzzy Syst.* 19, 553–565 doi: 10.1007/s40815-016-0289-3

Choi, S., Shin, Y. B., Kim, S.-Y., and Kim, J. (2018). A novel sensor-based assessment of lower limb spasticity in children with cerebral palsy. *J. Neuroeng. Rehabil.* 15, 45. doi: 10.1186/s12984-018-0388-5

Daly, J. J., and Wolpaw, J. R. (2008). Brain-computer interfaces in neurological rehabilitation. *Lancet Neurol.* 7, 1032–1043. doi: 10.1016/S1474-4422(08)70223-0

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

Diez, P. F., Mut, V. A., Avila Perona, E. M., and Laciar Leber, E. (2011). Asynchronous BCI control using high-frequency SSVEP. *J. Neuroeng. Rehabil.* 8, 39. doi: 10.1186/1743-0003-8-39

Duarte, J. L., Alvarenga, K., de, F., Banhara, M. R., Melo, A. D., Sás, R. M., et al. (2009). P300-long-latency auditory evoked potential in normal hearing subjects: simultaneous recording value in Fz and Cz. *Braz. J. Otorhinolaryngol.* 75, 231–236. doi: 10.1016/S1808-8694(15)30783-7

Duffau, H. (2016). "Chapter 18 - Brain Plasticity and Reorganization Before, During, and After Glioma Resection," in *Book: Glioblastoma*. edn. Elsevier, 225–236.

Evans, N., and Blanke, O. (2013). Shared electrophysiology mechanisms of body ownership and motor imagery. *NeuroImage*. 64, 216–228. doi: 10.1016/j.neuroimage.2012.09.027

Friehs, G. M., Zerris, V. A., Ojakangas, C. L., Fellows, M. R., and Donoghue, J. P. (2004). Brain–Machine and Brain–Computer Interfaces. *Stroke* 35, 2702–2705. doi: 10.1161/01.STR.0000143235.93497.03

Ge, S., Wang, R., and Yu, D. (2014). Classification of four-class motor imagery employing single-channel electroencephalography. *PLoS One*. 9, e98019. doi: 10.1371/journal.pone.0098019

Goncharova, I. I., McFarland, D. J., Vaughan, T. M., and Wolpaw, J. R. (2003). EMG contamination of EEG: Spectral and topographical characteristics. *Clin. Neurophysiol*. 114, 1580–1593. doi: 10.1016/S1388-2457(03)00093-2

Grosse-Wentrup, M., Mattia, D., and Oweiss, K. (2011). Using brain–computer interfaces to induce neural plasticity and restore function. *J. Neural Eng.* 8, 025004. doi: 10.1088/1741-2560/8/2/025004

Hortal, E., Planelles, D., Resquin, F., Climent, J. M., Azorín, J. M., and Pons, J. L. (2015). Using a brain-machine interface to control a hybrid upper limb exoskeleton during rehabilitation of patients with neurological conditions. *J. Neuroeng. Rehabil.* 12, 92. doi: 10.1186/s12984-015-0082-9

Hramov, A. E., Maksimenko, V. A., and Pisarchik, A. N. (2021). Physical principles of brain–computer interfaces and their applications for rehabilitation, robotics and control of human brain states. *Physic. Rep*. 918, 1–133. doi: 10.1016/j.physrep.2021.03.002

Jeannerod, M. (2006). *Motor Cognition: What Actions Tell the Self*. Oxford: Oxford University Press.

Jeong, H., Song, M., Oh, S., Kim, J., and Kim, J. (2019). Toward comparison of cortical activation with different motor learning methods using event-related design: EEG-fNIRS study. *Ann. Int. Conf. IEEE Eng. Med. Biol. Soc*. 2019, 6339–6342. doi: 10.1109/EMBC.2019.8857693

Jochumsen, M., Navid, M. S., Rashid, U., Haavik, H., and Niazi, I. K. (2019). EMG- Versus EEG-triggered electrical stimulation for inducing corticospinal plasticity. *IEEE Transact. Neural Syst. Rehabilitat. Eng.* 22,1901–1908. doi: 10.1109/TNSRE.2019.2932104

Kleim, J. A., and Jones, T. A. (2008). Principles of experience-dependent neural plasticity: Implications for rehabilitation after brain damage. *J. Speech Lang. Hear. Res*. 51, S225–S239. doi: 10.1044/1092-4388(2008/018)

Ko, L.-W., Ranga, S. S. K., Komarov, O., and Chen, C.-C. (2017). Development of single-channel hybrid bci system using motor imagery and SSVEP. *J. Healthc Eng.* 2017, 3789386. doi: 10.1155/2017/3789386

Kober, S. E., Grössinger, D., and Wood, G. (2019). Effects of motor imagery and visual neurofeedback on activation in the swallowing network: a real-time fMRI study. *Dysphagia* 34, 879–895. doi: 10.1007/s00455-019-09985-w

Kus, R., Valbuena, D., Zygierewicz, J., Malechka, T., Graeser, A., and Durka, P. (2012). Asynchronous BCI based on motor imagery with automated calibration and neurofeedback training. *IEEE Trans. Neural. Syst. Rehabil. Eng*. 20, 823–835. doi: 10.1109/TNSRE.2012.2214789

Lebedev, M. A., and Nicolelis, M. A. L. (2006). Brain–machine interfaces: past, present and future. *Trends Neurosci*. 29, 536–546. doi: 10.1016/j.tins.2006.07.004

Lee, D. J., Bae, S. J., Jang, S. H., and Chang, P. H. (2017). Design of a clinically relevant upper-limb exoskeleton robot for stroke patients with spasticity. *IEEE Int. Conf. Rehabil Robot*. 2017, 622–627. doi: 10.1109/ICORR.2017.8009317

Lee, S.-H., Hwang, Y.-J., Lee, H.-J., Kim, Y.-H., Ogrinc, M., Burdet, E., et al. (2021). Proof-of-concept of a sensor-based evaluation method for better sensitivity of upper-extremity motor function assessment. *Sensors* 21, 5926. doi: 10.3390/s21175926

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

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

Lew, E., Chavarriaga, R., Silvoni, S., and Millán Jdel, R. (2012). Detection of self-paced reaching movement intention from EEG signals. *Front. Neuroeng.* 5, 13. doi: 10.3389/fneng.2012.00013

Lewis, J. P. (1995). Fast normalized cross-correlation. *Proc. Vis. Interf*. 5, 120–3.

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

Liang, S., Choi, K.-S., Qin, J., Pang, W.-M., Wang, Q., and Heng, P.-A. (2016). Improving the discrimination of hand motor imagery via virtual reality based visual guidance. *Comput. Methods Program. Biomed*. 132, 63–74. doi: 10.1016/j.cmpb.2016.04.023

Liu, Y.-H., Huang, C.-W., and Hsiao, Y.-T. (2013). Controlling the false positive rate of a two-state self-paced brain-computer interface. *Conf. Proc. IEEE Int. Conf. Syst. Man Cybern*. 1476–81. doi: 10.1109/SMC.2013.255

Livingston, R. B. (1966). Brain mechanisms in conditioning and learning. *Neurosci. Res. Progr. Bull*. 4, 349–354

Loetscher, T., and Lincoln, N. B. (2019). Cognitive rehabilitation for attention deficits following stroke. *Cochr. Database Syst. Rev*. 11, CD002842. doi: 10.1002/14651858.CD002842.pub3

Makeig, S. (1993). Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalogr. Clin. Neurophysiol*. 86, 283–293. doi: 10.1016/0013-4694(93)90110-H

Mcfarland, D. J., Lefkowicz, A. T., and Wolpaw, J. R. (1997). Design and operation of an EEG-based brain-computer interface with digital signal processing technology. *Behav. Res. Methods Instrum. Comput*. 29, 337–345. doi: 10.3758/BF03200585

Meinel, A., Castaño-Candamil, S., Blankertz, B., Lotte, F., and Tangermann, M. (2019). Characterizing regularization techniques for spatial filter optimization in oscillatory EEG regression problems. *Neuroinformatics* 17, 235–251. doi: 10.1007/s12021-018-9396-7

Michielsen, M. E., Selles, R. W., van der Geest, J. N., Eckhardt, M., Yavuzer, G., Stam, H. J., et al. (2010). Motor recovery and cortical reorganization after mirror therapy in chronic stroke patients: a phase II randomized controlled trial. *Neurorehabil. Neural Rep*. 25, 223–233. doi: 10.1177/1545968310385127

Miladinović, A., Ajčević, M., Jarmolowska, J., Marusic, U., Silveri, G., Battaglini, P. P., et al. (2020). Performance of EEG Motor-Imagery based spatial filtering methods: a BCI study on stroke patients. *Procedia Comput. Sci.* 176, 2840–2848. doi: 10.1016/j.procs.2020.09.270

Miller, K. J., Schalk, G., Fetz, E. E., Den Nijs, M., Ojemann, J. G., and Rao, R. P. N. (2010). Cortical activity during motor execution, motor imagery, and imagery-based online feedback. *Proc. Natl. Acad. Sci. U S A*. 107, 4430–4435. doi: 10.1073/pnas.0913697107

Mrachacz-Kersting, N., Aliakbaryhosseinabadi, S., Pedersen, M., Jiang, N., and Farina, D. (2017). Tactile stimulation training to enhance MRCP detection in chronic stroke patients. *Lect. Notes Comput. Sci.* 17, 354–63, doi: 10.1007/978-3-319-58625-0_26

Müller-Putz, G. R., Kaiser, V., Solis-Escalante, T., and Pfurtscheller, G. (2010). Fast set-up asynchronous brain-switch based on detection of foot motor imagery in 1-channel EEG. *Med. Biol. Eng. Comput*. 48, 229–233. doi: 10.1007/s11517-009-0572-7

Murphy, T. H., and Corbett, D. (2009). Plasticity during stroke recovery: from synapse to behaviour. *Nat. Rev. Neurosci*. 10, 861–872. doi: 10.1038/nrn2735

Niazi, I. K., Jiang, N., Tiberghien, O., Nielsen, J. F., Dremstrup, K., and Farina, D. (2011). Detection of movement intention from single-trial movement-related cortical potentials. *J. Neural Eng.* 8, 066009. doi: 10.1088/1741-2560/8/6/066009

Niazi, I. K., Navid, M. S., Rashid, U., Amjad, I., Olsen, S., Haavik, H., et al. (2022). Associative cued asynchronous BCI induces cortical plasticity in stroke patients. *Annal. Clinic. Translat. Neurol.* 9, 722–733. doi: 10.1002/acn3.51551

Nicolas-Alonso, L. F., and Gomez-Gil, J. (2012). Brain Computer Interfaces: a Review. *Sensors* 12, 1211–1279. doi: 10.3390/s120201211

Oppitz, S. J., Didoné, D. D., Silva, D. D., Gois, M., Folgearini, J., Ferreira, G. C., et al. (2015). Long-latency auditory evoked potentials with verbal and nonverbal stimuli. *Braz. J. Otorhinolaryngol*. 81, 647–652. doi: 10.1016/j.bjorl.2014.10.005

Park, Y., and Chung, W. (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

Pearson, K. (1895). Notes on regression and inheritance in the case of two parents. *Proc. R Soc. Lond. B Biol. Sci*. 58, 240–242 doi: 10.1098/rspl.1895.0041

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

Pfurtscheller, G., Neuper, C., Muller, G. R., Obermaier, B., Krausz, G., Schlogl, A., et al. (2003). Graz-BCI: state of the art and clinical applications. *IEEE Transact. Neural Syst. Rehabilit. Eng.* 11, 1–4. doi: 10.1109/TNSRE.2003.814454

Pfurtscheller, G., and Solis-Escalante, T. (2008). Could the beta rebound in the EEG be suitable to realize a "brain switch"? *Clin Neurophysiol*. 120, 24–29. doi: 10.1016/j.clinph.2008.09.027

Reinkensmeyer, D. J., Burdet, E., Casadio, M., Krakauer, J. W., Kwakkel, G., and Lang, C. E. (2016). Computational neurorehabilitation: modeling plasticity and learning to predict recovery. *J. Neuroeng. Rehabil*. 13, 42. doi: 10.1007/978-3-319-28603-7

Rodriguez-Ugarte, M., Iáñez, E., Ortíz, M., and Azorín, J. M. (2017). Personalized offline and pseudo-online BCI models to detect pedaling intent. *Front. Neuroinform*. 11, 45. doi: 10.3389/fninf.2017.00045

Sadeghian, E. B., and Moradi, M. H. (2008). Fractal dimension for detection of ERD/ERS patterns in asynchronous brain computer interface. *Int. Cong. Bioinform. Biomed. Eng*. 2008, 560–563. doi: 10.1109/ICBBE.2008.136

Salenius, S., Kajola, M., Thompson, W. L., Kosslyn, S., and Hari, R. (1995). Reactivity of magnetic parieto-occipital alpha rhythm during visual imagery. *Electroencephalogr. Clin. Neurophysiol*. 95, 453–462. doi: 10.1016/0013-4694(95)00155-7

Sasmita, A. O., Kuruvilla, J., and Ling, A. P. K. (2018). Harnessing neuroplasticity: modern approaches and clinical future. *Int. J. Neurosci*. 128, 1061–1077. doi: 10.1080/00207454.2018.1466781

Siuly, S., and Li, Y. (2012). Improving the separability of motor imagery EEG signals using a cross correlation-based least square support vector machine for brain-computer interface. *IEEE Trans. Neural. Syst. Rehabil. Eng*. 20, 526–538. doi: 10.1109/TNSRE.2012.2184838

Song, M., and Kim, J. (2019). A paradigm to enhance motor imagery using rubber hand illusion induced by visuo-tactile stimulus. *IEEE Trans. Neural Syst. Rehabil. Eng*. 27, 477–486. doi: 10.1109/TNSRE.2019.2895029

Song, M., Oh, S., Jeong, H., Kim, J. B., and Kim, J. H. (2018). A novel movement intention detection method for neurorehabilitation brain-computer interface system. *Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern*. 2018, 1016–1021. doi: 10.1109/SMC.2018.00181

Sun, Y., Wei, W., Luo, Z., Gan, H., and Hu, X. (2016). Improving motor imagery practice with synchronous action observation in stroke patients. *Top Stroke Rehabil*. 23, 245–253. doi: 10.1080/10749357.2016.1141472

Tam, W., Ke, Z., and Tong, K. (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. *Ann. Int. Conf. IEEE Eng. Med. Biol Soc*. 2011, 6344–6347. doi: 10.1109/IEMBS.2011.6091566

Thaut, M. H., and McIntosh, G. C. (2014). Neurologic music therapy in stroke rehabilitation. *Curr. Phys. Med. Rehabil. Rep*. 14, 2106–13 doi: 10.1007/s40141-014-0049-y

Toscani, M., Marzi, T., Righi, S., Viggiano, M. P., and Baldassi, S. (2010). Alpha waves: a neural signature of visual suppression. *Exp. Brain Res*. 207, 213–219. doi: 10.1007/s00221-010-2444-7

Wang, Y., Gao, S., and Gao, X. (2005). Common spatial pattern method for channel selelction in motor imagery based brain-computer interface. *Conf. Proc. IEEE Eng. Med. Biol. Soc*. 2005, 5392–5395. doi: 10.1109/IEMBS.2005.1615701

Wei, W., Xiaorong, G., Bo, H., and Shangkai, G. (2008). Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL). *IEEE Transact. Biomed. Eng*. 55, 1733–1743. doi: 10.1109/TBME.2008.919125

Young, J. A., and Tolentino, M. (2011). Neuroplasticity and its applications for rehabilitation. *Am. J. Therapeutic*. 18, 70–80, doi: 10.1097/MJT.0b013e3181e0f1a4

frontiers | Frontiers in Neuroinformatics

TYPE Original Research
PUBLISHED 21 September 2022
DOI 10.3389/fninf.2022.997068



OPEN ACCESS

EDITED BY
Luzheng Bi,
Beijing Institute of Technology, China

REVIEWED BY
Sung-Phil Kim,
Ulsan National Institute of Science and Technology, South Korea
Yingchun Zhang,
University of Houston, United States
Shangen Zhang,
University of Science and Technology Beijing, China

*CORRESPONDENCE
Chang-Hwan Im
ich@hanyang.ac.kr

RECEIVED 18 July 2022
ACCEPTED 30 August 2022
PUBLISHED 21 September 2022

CITATION
Kwon J, Hwang J, Nam H and Im C-H (2022) Novel hybrid visual stimuli incorporating periodic motions into conventional flickering or pattern-reversal visual stimuli for steady-state visual evoked potential-based brain-computer interfaces.
*Front. Neuroinform.* 16:997068.
doi: 10.3389/fninf.2022.997068



# Novel hybrid visual stimuli incorporating periodic motions into conventional flickering or pattern-reversal visual stimuli for steady-state visual evoked potential-based brain-computer interfaces

Jinuk Kwon[1,2], Jihun Hwang[2], Hyerin Nam[3] and Chang-Hwan Im[1,2,3,4]*

[1]Department of Biomedical Engineering, Hanyang University, Seoul, South Korea, [2]Department of Electronic Engineering, Hanyang University, Seoul, South Korea, [3]Department of Artificial Intelligence, Hanyang University, Seoul, South Korea, [4]Department of HY-KIST Bio-Convergence, Hanyang University, Seoul, South Korea

In this study, we proposed a new type of hybrid visual stimuli for steady-state visual evoked potential (SSVEP)-based brain-computer interfaces (BCIs), which incorporate various periodic motions into conventional flickering stimuli (FS) or pattern reversal stimuli (PRS). Furthermore, we investigated optimal periodic motions for each FS and PRS to enhance the performance of SSVEP-based BCIs. Periodic motions were implemented by changing the size of the stimulus according to four different temporal functions denoted by none, square, triangular, and sine, yielding a total of eight hybrid visual stimuli. Additionally, we developed the extended version of filter bank canonical correlation analysis (FBCCA), which is a state-of-the-art training-free classification algorithm for SSVEP-based BCIs, to enhance the classification accuracy for PRS-based hybrid visual stimuli. Twenty healthy individuals participated in the SSVEP-based BCI experiment to discriminate four visual stimuli with different frequencies. An average classification accuracy and information transfer rate (ITR) were evaluated to compare the performances of SSVEP-based BCIs for different hybrid visual stimuli. Additionally, the user's visual fatigue for each of the hybrid visual stimuli was also evaluated. As the result, for FS, the highest performances were reported when the periodic motion of the sine waveform was incorporated for all window sizes except for 3 s. For PRS, the periodic motion of the square waveform showed the highest classification accuracies for all tested window sizes. A significant statistical difference in the performance between the two best stimuli was not observed. The averaged fatigue scores were reported to be $5.3 \pm 2.05$ and $4.05 \pm 1.28$ for FS with sine-wave periodic motion and PRS with square-wave periodic motion, respectively. Consequently, our results demonstrated that FS with sine-wave periodic motion and PRS with square-wave periodic motion could effectively improve the BCI performances compared to conventional FS and PRS. In

addition, thanks to its low visual fatigue, PRS with square-wave periodic motion can be regarded as the most appropriate visual stimulus for the long-term use of SSVEP-based BCIs, particularly for window sizes equal to or larger than 2 s.

KEYWORDS

**brain-computer interfaces (BCIs), steady-state visual evoked potential (SSVEP), steady-state motion visual evoked potential (SSMVEP), hybrid visual stimulus, periodic motion**

## Introduction

Brain-computer interfaces (BCIs) are promising alternative communication technologies that have been generally developed for people who suffer from neuromuscular disorders or physical disabilities such as spinal cord injury, amyotrophic lateral sclerosis, and locked-in syndrome (Daly and Wolpaw, 2008). BCIs have provided new non-muscular communication channels that allowed for interaction between a user and the external environment. A variety of non-invasive brain imaging modalities have been employed to record brain activities in the field of BCIs. For example, functional magnetic resonance imaging (fMRI), magnetoencephalography, and functional near-infrared spectroscopy have been successfully employed to implement BCIs. In addition, electroencephalography (EEG) is another representative non-invasive neuroimaging modality that has been the most intensively studied owing to its advantages over the other modalities, such as high temporal resolution, affordability, and portability (Dai et al., 2020; Zhang et al., 2021).

In the EEG-based BCIs, the user performs certain mental tasks according to paradigms designed for eliciting task-related neural activities. Motor imagery, event-related potential, P300, and auditory steady-state response are popular paradigms employed to implement EEG-based BCIs (Lotte et al., 2018; Abiri et al., 2019). Steady-state visual evoked potential (SSVEP) is also one of the most promising EEG-based BCI paradigms, which has attracted increased interest from BCI researchers in recent decades (Waytowich et al., 2018). SSVEPs are periodic brain activities evoked in response to the presentation of visual stimulus flickering or pattern-reversing at a specific temporal frequency. SSVEP signals are entrained at the fundamental and harmonic frequencies of the visual stimulus and are well-known to be mainly observed in the occipital region of the brain over a wide range of 1–90 Hz (Herrmann, 2001; Choi et al., 2019a). SSVEP-based BCIs interpret the user's intention by detecting the visual stimulus that the user gazed at based on these characteristics and have various advantages over the other paradigms, such as high information transfer rate (ITR), excellent stability, and little training requirement (Zhang et al., 2020; Kim and Im, 2021). Thanks to these advantages, SSVEP-based BCIs have been successfully applied to various applications including mental speller (Nakanishi et al., 2018), assistive technology for patients (Perera et al., 2016), online home appliance control (Kim et al., 2019), and hands-free controllers for virtual reality (VR) (Armengol-Urpi and Sarma, 2018) or augmented reality (AR) (Arpaia et al., 2021).

In general, two types of visual stimuli have been employed to evoke SSVEPs: (1) flickering stimulus (FS) and (2) pattern-reversal stimulus (PRS) (Bieger et al., 2010; Zhu et al., 2010). FS is the visual stimulus that modulates the color or luminance of the stimulus at a specific frequency. Flickering single graphics in the form of squares or circles rendered on an LCD monitor is the representative FS used to elicit the SSVEPs. PRS evokes SSVEP responses by alternating the patterns of the visual stimuli (e.g., checkerboard or line boxes) at a constant frequency. Based on these visual stimuli, a number of studies have been conducted to improve the performance of SSVEP-based BCIs, examples of which include optimization of stimulus parameters such as spatial frequency of PRS, stimulation frequencies, colors, and waveform of FS (Bieger et al., 2010; Teng et al., 2011; Duszyk et al., 2014; Jukiewicz and Cysewska-Sobusiak, 2016; Chen et al., 2019). Recently, Choi et al. (2019b) and Park et al. (2019) proposed a novel type of visual stimulus called grow/shrink stimulus (GSS) to improve the performance of SSVEP-based BCI in AR and VR environments, respectively. GSS was implemented by incorporating a periodic motion into FS to concurrently evoke SSVEP and steady-state motion visual evoked potential (SSMVEP), inspired by previous studies that reported that the periodic motion-based visual stimuli could elicit SSMVEP (Xie et al., 2012; Yan et al., 2017). GSS has shown a higher BCI performance compared to the conventional PRS or FS in both VR and AR environments. However, no study has been conducted on the performance of GSS-like visual stimuli for SSVEP-based BCI when the LCD monitor is used as a rendering device. Furthermore, the effect of the motion parameters (i.e., the waveform of the temporal motion dynamics) on the BCI performances has not been investigated. Indeed, the investigation of the performance of various GSS-like visual stimuli with the LCD monitor environment is important because most SSVEP-based BCI studies employ the

LCD monitor to present the visual stimuli (Ge et al., 2019; Chen et al., 2021; Xu et al., 2021). In addition, to the best of our knowledge, hybrid visual stimuli that consolidate PRS with periodic motions have never been proposed in previous studies.

In this study, we proposed novel hybrid visual stimuli that consolidate the conventional PRS with periodic motions and further investigated the effect of waveforms of the periodic motions for hybrid visual stimuli based on either FS or PRS on the performance of SSVEP-based BCIs. As for the periodic motions, the stimulus size was changed according to four different waveforms: none (no change in the size), square (changing size in a binary manner), triangular (linearly increasing and decreasing size), and sine (changing size with a sinusoidal waveform), resulting in a total of eight different hybrid visual stimuli (i.e., FS and PRS each with four periodic motions). We evaluated two crucial factors for the practical use of SSVEP-based BCIs: (1) BCI performances and (2) visual fatigue, for each visual stimulus, with 20 healthy participants. A filter bank canonical correlation analysis (FBCCA) algorithm, which is a state-of-the-art training-free algorithm for SSVEP-based BCIs was employed to evaluate the performances of SSVEP-based BCIs in terms of classification accuracy and information transfer rate (ITR). Moreover, an extended version of FBCCA, named subharmonic-FBCCA (sFBCCA) was developed for the SSVEP-based BCIs with PRS-based hybrid visual stimuli.

# Methods

## Participants

A total of 20 healthy adults (10 males, aged 23.7 $\pm$ 3.5 years) with normal or corrected-to-normal vision participated in the experiments. None of the reported any serious history of neurological, psychiatric, or other severe diseases that could otherwise influence the experimental results. All participants were informed of the detailed experimental procedure and provided written consent before the experiment. This study and the experimental paradigm were approved by the Institutional Review Board Committee of Hanyang University, Republic of Korea (IRB No. HYU-202006-004-03) according to the Declaration of Helsinki.

## Visual stimuli

The visual stimuli were developed with the Unity 3D engine (Unity Technologies ApS, San Francisco, CA, USA). Based on previous GSS studies (Choi et al., 2019b; Park et al., 2019), all stimuli were designed in a star shape, with a base size of 7 cm (5.7°) to increase the visibility of periodic motions. The background color was set to gray. Both FS and PRS changed the color or reversed the patterns with the periodic square waveform according to the results of previous studies that reported that square-wave FS exhibited significantly higher classification accuracy than FS of other waveforms (Teng et al., 2011; Chen et al., 2019). The periodic motions were implemented by varying the size of visual stimuli according to four different types of waveforms: none (no change in the size), square (changing size in a binary manner), triangular (linearly increasing and decreasing size), and sine (changing size with a sinusoidal waveform) waveforms with a modulation ratio of 33% compared to the base size (i.e., the radius of each stimulus was changed from 0.67 to 1.33 when the radius of the base stimulus was assumed to be one). The conventional visual stimuli of FS and PRS were combined with four periodic motions, resulting in eight hybrid visual stimuli. Hereinafter, none, square, triangular, and sine waveforms are referred to as None, Square, Triangular, and Sine, respectively, and each hybrid visual stimuli are referred to as FS-None, FS-Square, FS-Triangular, FS-Sine, PRS-None, PRS-Square, PRS-Triangular, and PRS-Sine. Note that FS-None and PRS-None were the same as the conventional FS and PRS with the base size. Figure 1 illustrates the examples of the hybrid visual stimuli when the stimulation frequency was set to 6 Hz. Blue circles indicate the stimulus size presented to the participants considering the refresh rate of the LCD monitor (= 60 Hz). It is worthwhile noting that the most important difference between FS and PRS is that FS elicits SSVEP responses at the number of full cycles (i.e., two reversals) per second, whereas PRS evokes SSVEP responses at the number of reversals per second (Zhu et al., 2010). Therefore, the stimulation frequencies of periodic motions for PRS were set to be half of those for FS, which were considered as subharmonics of the stimulation frequencies in the further analysis.

## Experimental paradigm

The participants sat 70 cm away from a 27-inch LCD monitor with a resolution of 1920 x 1080 pixels and the 60 Hz refresh rate. The experiment consisted of eight sessions corresponding to each hybrid visual stimuli and the order of the sessions was randomized for each participant. Each session was composed of 20 trials (5 trials $\times$ 4 stimuli), each of which consisted of the visual cue of 3 s and the stimulation time of 5 s. The red bar was presented under the target stimulus during visual cue period in a randomized order. The timing sequence of a single trial is shown in Figure 2. The stimulation frequencies of four stimuli were determined as 6, 6.67, 7.5, and 10 Hz considering the refresh rate of the LCD monitor. In each trial, the participants were instructed to focus their attention on the target stimulus among four simultaneously flickering stimuli without eye blinks and body movements during the stimulation time. At the end of each session, the participants evaluated the

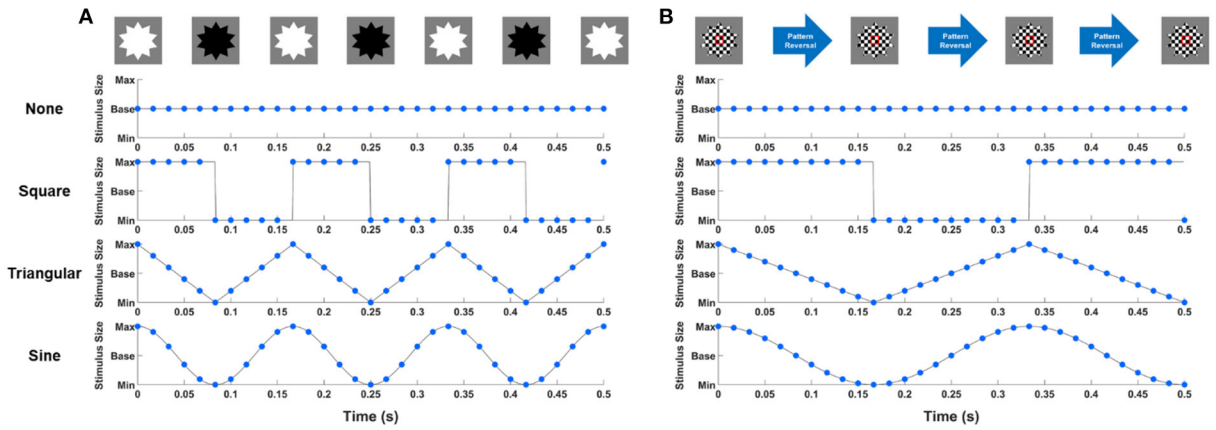



FIGURE 1
Examples of **(A)** FS-based hybrid visual stimuli and **(B)** PRS-based hybrid visual stimuli when the stimulation frequency was 6 Hz. Blue circles indicate the stimulus size presented to participants considering the refresh rate of the LCD monitor.

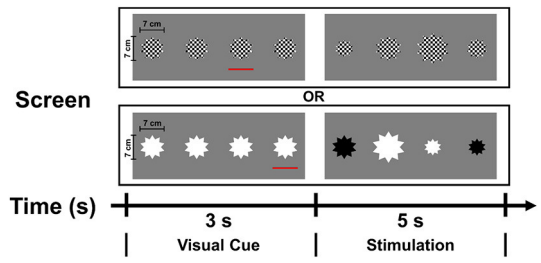



FIGURE 2
The timing sequence of a single trial. Each trial consisted of the visual cue of 3 s and the stimulation time of 5 s.

visual fatigue score for each hybrid visual stimulus in the range of 1–10 (1, low fatigue; 10, high fatigue).

## Data recording and pre-processing

EEG data were recorded from eight scalp electrodes (O1, Oz, O2, PO7, PO3, POz, PO4, and PO8) using a commercial EEG system (BioSemi Active Two; Biosemi, Amsterdam, The Netherlands) at a sampling rate of 2,048 Hz. A CMS active electrode and a DRL passive electrode were used to form a feedback loop for the amplifier reference (Park et al., 2019). MATLAB 2020b (Mathworks; Natick, MA) was used to analyze the EEG data, and the functions implemented in the BBCI toolbox (https://github.com/bbci/bbci_public) were employed. The raw EEG data were down-sampled to 256 Hz to reduce the computational cost and then bandpass-filtered using a sixth-order zero-phase Butterworth filter with cutoff frequencies of 2 and 54 Hz. Considering a latency delay in the visual pathway, the EEG data were segmented into epochs from 0.135 to 0.135+$w$ s with respect to the task onset time (0 s), where $w$ indicates the window size used for SSVEP detection (Rabiul Islam et al., 2017).

## Classification methods

### Canonical correlation analysis

CCA is a multivariate statistical method used to measure the underlying correlation between two sets of multidimensional variables, $X \in R^{d_x \times N_s}$ and $Y \in R^{d_y \times N_s}$ where, $N_s$ is the number of sample points and $d_x$ and $d_y$ indicate the dimension of $X$ and $Y$, respectively (Nakanishi et al., 2015). Considering their linear combinations $x = X^T W_X$ and $y = Y^T W_Y$, CCA finds a pair of weight vectors $W_X \in R^{d_x \times 1}$ and $W_Y \in R^{d_y \times 1}$ that maximize Pearson's correlation coefficients between $x$ and $y$ using the following equation:

$$\max_{W_x, W_y} \rho(x, y) = \frac{E\left[W_X^T X Y^T W_Y\right]}{\sqrt{E\left[W_X^T X X^T W_X\right] E\left[W_Y^T Y Y^T W_Y\right]}}. \quad (1)$$

Here, $T$ denotes the transpose operation. The maximum correlation coefficient with respect to $W_X$ and $W_Y$ is called the "CCA coefficient."

For SSVEP detection, the CCA coefficients, $\rho_f$, between multichannel EEG signals, $X \in R^{N_c \times N_s}$, and the reference signals for each stimulus frequency, $Y_f \in R^{2N_h \times N_s}$, were evaluated and the frequency with the largest CCA coefficient was classified as the target frequency, as follows:

$$f_{target} = \max_f \rho_f, \quad f = f_1, f_2, \ldots, f_K. \quad (2)$$

Here, $K$ is the number of stimulus frequencies presented to the participants.

The reference signal for each stimulus frequency ($Y_f$) was set as

$$Y_f = \begin{bmatrix} \sin\left(2\pi fn\right) \\ \cos\left(2\pi fn\right) \\ \vdots \\ \sin\left(2\pi N_h fn\right) \\ \cos\left(2\pi N_h fn\right) \end{bmatrix}, \; n = \frac{1}{f_s}, \frac{2}{f_s}, \ldots, \frac{N_s}{f_s}, \tag{3}$$

where, $f$ is the stimulus frequency. In this study, $N_c$ and $f_s$ denote the number of channels and sampling frequency, which were set to 8 and 256, respectively. $N_h$ represents the number of harmonics, which was set to 5 according to previous studies (Chen et al., 2015).

## Filter bank CCA

FBCCA combines CCA with filter bank analysis to extract the discriminative information in the harmonic components (Chen et al., 2015). The filter bank is applied to decompose EEG data into multiple sub-band data, and CCA coefficients are evaluated for each sub-band. The weighted sums of the squared sub-band CCA coefficients for each stimulus frequency are calculated using the following equations:

$$\rho_f = \sum_{m=1}^{N_m} w\left(m\right) \cdot \left(\rho_f^m\right)^2, \tag{4}$$

$$w\left(m\right) = m^{-a} + b, \tag{5}$$

where, $N_m$ is the number of subbands, $m$ is the index of the subbands, and $\rho_f^m$ denotes the CCA coefficient of sub-band $m$. The target frequency is determined in the same manner as in CCA. According to previous studies (Chen et al., 2015; Zhao et al., 2020), the following parameters were set: $a = 1.25$, $b = 0.25$, and $N_m = 5$. The filter bank for five sub-bands was designed with lower and upper cutoff frequencies of 4–52, 8–52, 12–52, 16–52, and 20–52 Hz, respectively (Chen et al., 2015). In this study, FBCCA was employed to identify SSVEPs because it is generally regarded as the best available algorithm, yielding the highest classification accuracy without the need for training sessions (Zerafa et al., 2018; Liu et al., 2020).

## Subharmonic FBCCA (sFBCCA)

In this study, we proposed an extended version of FBCCA, named subharmonic-FBCCA (sFBCCA), to utilize the information in the subharmonic component, elicited by periodic motions for PRS. In sFBCCA, the reference signal was expanded to include the subharmonic component as follows:

$$Y_f = \begin{bmatrix} \sin\left(\pi fn\right) \\ \cos\left(\pi fn\right) \\ \sin\left(2\pi fn\right) \\ \cos\left(2\pi fn\right) \\ \vdots \\ \sin\left(2\pi N_h fn\right) \\ \cos\left(2\pi N_h fn\right) \end{bmatrix}, \; n = \frac{1}{f_s}, \frac{2}{f_s}, \ldots, \frac{N_s}{f_s}. \tag{6}$$

In addition, the equation for the weighted sums of the squared sub-band CCA coefficients is extended as the following equations:

$$\rho_f = \sum_{m=1}^{N_m} w\left(m\right) \cdot \left(\rho_f^m\right)^2 + w_{sub} \cdot \left(\rho_{f_{sub}}\right)^2, \tag{7}$$

$$w\left(m\right) = m^{-a} + b, \tag{8}$$

$$w_{sub} = {m_{sub}}^{-a} + b, \tag{9}$$

where, $m_{sub}$ represents the index of the subharmonic, set to 0.5, in this study. The bandpass filter for the subharmonic component was designed with lower and upper cutoff frequencies of 1–52 Hz. sFBCCA was employed to classify SSVEPs for hybrid visual stimuli of PRS-Square, PRS-Triangular, and PRS-Sine.

# Information transfer rate

In addition to the classification accuracy, ITR (bits per minute) has been widely employed as a metric to assess the performance of the BCI system (Wolpaw et al., 2002). The ITR was evaluated using the following equation:

$$ITR = \frac{60}{T}\left\{log_p 2N + plog_p 2p + \left(1-p\right) log_p 2\left(\frac{1-p}{N-1}\right)\right\}, \tag{10}$$

Where, $T$ denotes the window size (in seconds), $N$ indicates the number of classes, and $p$ represents the classification accuracy. In the present study, the $N$ value was 4.

# Statistical analysis

Statistical analyses were also performed using MATLAB 2020b (MathWorks; Natick, MA, USA). The non-parametric method was employed because the normality criterion was not satisfied owing to the small sample size. Friedman test was conducted to verify if there were significant differences among the BCI performances. Wilcoxon signed-rank test with the false discovery rates (FDRs) correction for multiple comparisons was performed for *post-hoc* analyses.

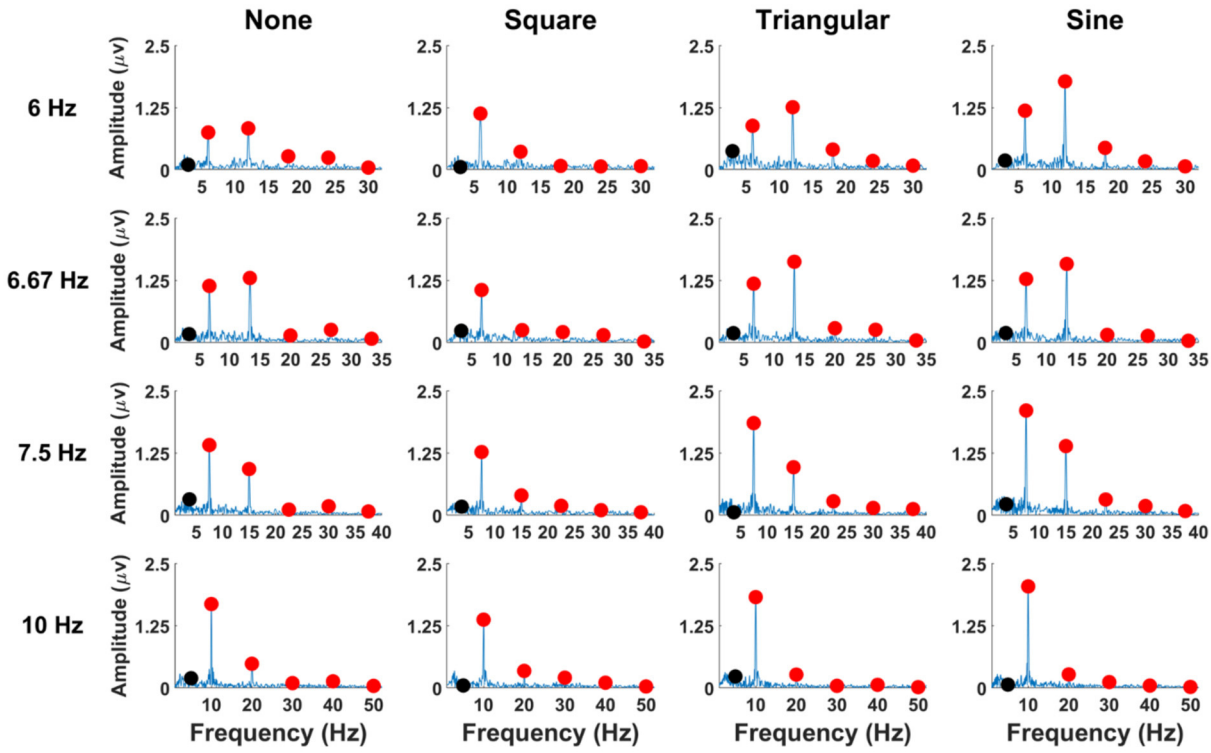



FIGURE 3
SSVEP amplitudes of the averaged EEG signals across all the participants at the Oz electrode for FS-based hybrid visual stimuli. Red circles indicate the stimulation frequencies and their harmonics, and black circles represent the subharmonic frequencies.

# Results

## FS-based hybrid visual stimuli

Figure 3 illustrates the grand mean amplitude spectra of SSVEPs averaged across all participants with respect to waveforms of periodic motions. The amplitudes of SSVEPs were obtained from the EEG signals of 5-s long recorded at the Oz electrode. Here, the first five harmonic components of the stimulation frequencies, which were used for the classification, and the subharmonic components were presented in the figure. The red circles indicate the fundamentals and harmonics of stimulation frequencies, and the black circles represent the subharmonics. For FS-based hybrid visual stimuli, clear SSVEP peaks were mainly evoked at the fundamental and second harmonic frequencies. No SSVEP peaks were observed at the subharmonic frequency. The grand average amplitudes of each SSVEP component are listed in Supplementary Table 1.

The grand mean amplitudes of SSVEP components averaged over all stimulation frequencies across all the participants are illustrated in Figure 4 as a function of waveforms of periodic motions, where the error bars represent the standard errors. The statistical analyses were performed to compare the differences in the amplitude of SSVEP components at the subharmonic, fundamental, and second harmonic frequencies among FS-based hybrid visual stimuli. Four SSVEP amplitudes at each harmonic frequency were calculated from the EEG signals averaged over each stimulation frequency recorded at the Oz electrode for each participant. Consequently, a total of 80 SSVEP amplitudes (4 stimulation frequencies $\times$ 20 participants) were statistically compared. The Friedman test indicated significant differences in the amplitudes at fundamental and second harmonic frequencies (subharmonic frequency: $\chi^2 = 0.05$, $p = 0.998$, fundamental frequency: $\chi^2 = 28.26$, $p < 0.001$, second harmonic frequency: $\chi^2 = 32.81$, $p < 0.001$). At the fundamental frequency, SSVEP amplitude elicited by FS-Sine and FS-Triangular was significantly higher than that elicited by FS-None and FS-Square (FDRs-corrected $p < 0.05$ between FS-Square vs FS-Triangular, and FDRs-corrected $p < 0.001$ for the others, Wilcoxon signed-rank test). For the second harmonic frequency, SSVEP amplitude evoked by FS-Sine was significantly higher than that evoked by FS with other waveforms ($p < 0.005$ for FS-None and $p < 0.001$ for the others, Wilcoxon signed-rank test with FDRs correction). Additionally, FS-Square elicited significantly lower

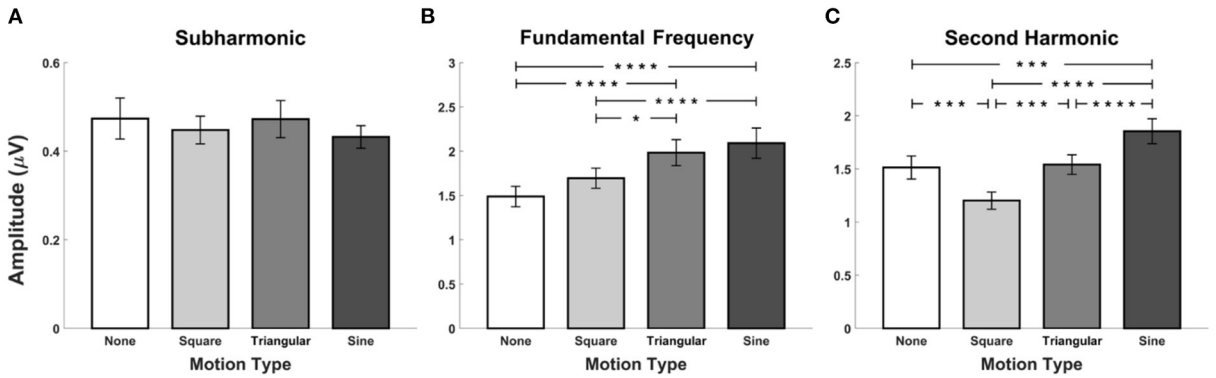



FIGURE 4
Grand mean SSVEP amplitudes for FS-based hybrid visual stimuli averaged across all the participants at **(A)** the subharmonic frequency, **(B)** the fundamental frequency, and **(C)** the second harmonic frequency. Error bars represent the standard errors. Here, the asterisks of *, ***, and **** represent FDRs-corrected $p < 0.05$, $p < 0.005$, and $p < 0.001$, respectively (Wilcoxon signed-rank test).

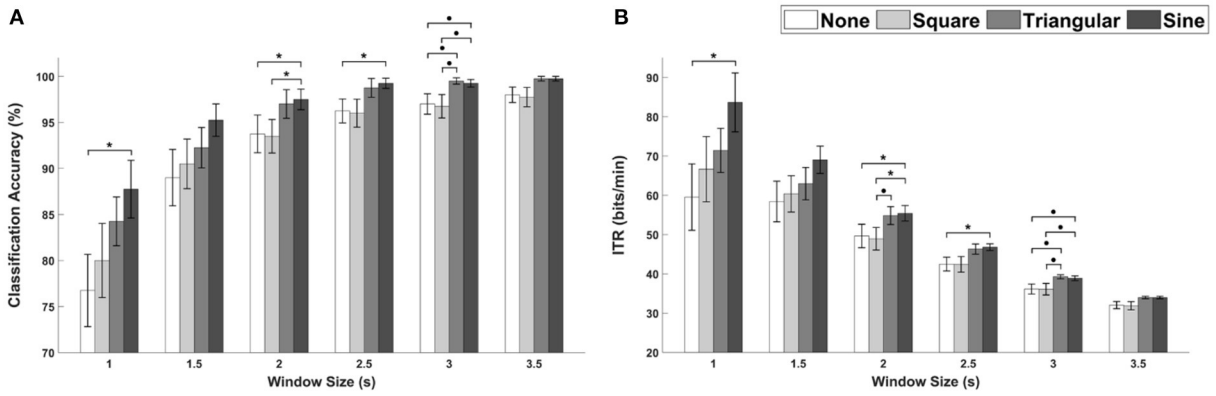



FIGURE 5
The average performance of FS-based hybrid visual stimuli in terms of **(A)** classification accuracies and **(B)** ITRs as a function of waveforms with different window sizes. Error bars represent the standard errors. Here, •$p < 0.1$ and *$p < 0.05$ are the FDRs-corrected $p$-values from Wilcoxon signed-rank test.

SSVEP amplitude than that elicited by the other waveforms (FDRs-corrected $p < 0.005$ for FS-None and FS-Triangular, and FDRs-corrected $p < 0.001$ for FS-Sine, Wilcoxon signed-rank test).

The average classification accuracies and ITRs for FS-based hybrid visual stimuli with respect to different window sizes are depicted in Figures 5A,B, respectively. The Friedman test indicated statistically significant differences for all window sizes except for 1.5 and 3.5 s (1 s, $\chi^2 = 8.10$, $p < 0.5$; 1.5 s, $\chi^2 = 6.20$, $p = 0.102$; 2 s, $\chi^2 = 9.50$, $p < 0.05$; 2.5 s, $\chi^2 = 11.38$, $p < 0.001$; 3 s, $\chi^2 = 9.88$, $p < 0.05$; 3.5 s, $\chi^2 = 7.47$, $p = 0.058$, identical to both the classification accuracy and ITR). The Wilcoxon signed-rank test with FDRs correction showed statistically significant differences in both classification accuracy and ITR between FS-None and FS-Sine for window sizes of 1, 2, and 2.5 s ($p < 0.5$ for both classification accuracies and ITRs). Additionally, the performances of FS-Sine were significantly higher than those of FS-Square for the window size of 2 s ($p < 0.5$, Wilcoxon signed-rank test with FDRs correction). As illustrated in the figure, the performances of SSVEP-based BCI could be improved by incorporating triangular- and sine-wave periodic motions into the conventional FS for all window sizes. For FS, FS-Sine exhibited the highest average performances for every window size except for 3 s, especially for short window sizes.

Figure 6A illustrates the fatigue scores of FS-based hybrid stimuli as a function of periodic motion waveforms. The gray bars represent the interquartile ranges from the first quartile to the third quartile and white circles indicate the median values. The averaged fatigue scores were 4.8 ± 1.82, 4.6 ± 1.90, 5.4 ± 2.01, and 5.3 ± 2.05 for FS-None, FS-Square, FS-Triangular, and FS-Sine, respectively. A statistically significant difference was not observed in the Friedman test ($\chi^2 = 2.55$, $p = 0.467$).

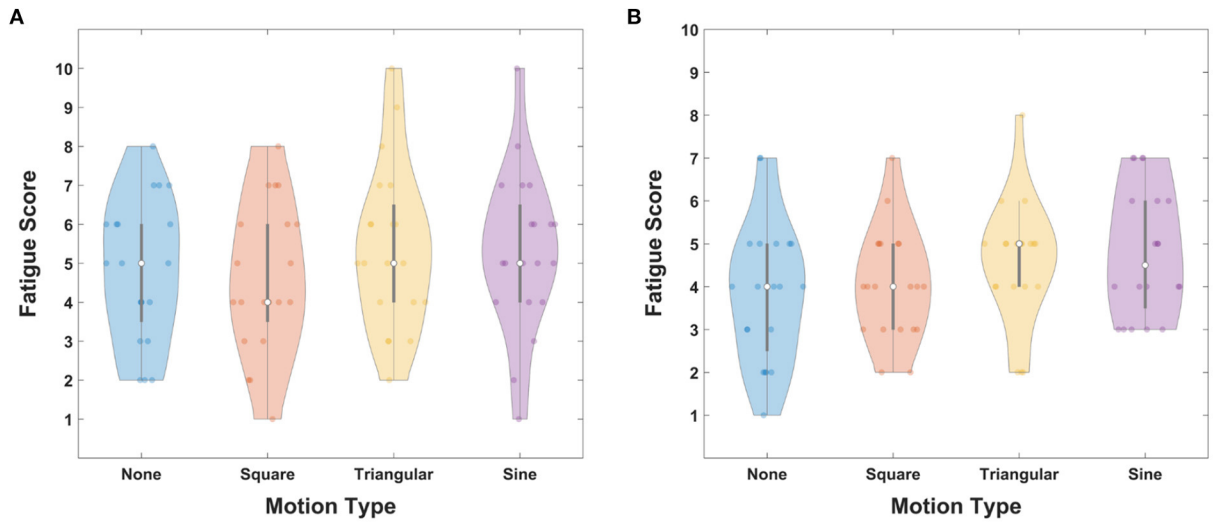



FIGURE 6
Fatigue scores for **(A)** FS-based hybrid visual stimuli and **(B)** PRS-based hybrid visual stimuli as a function of waveforms. Gray bars represent the interquartile range from 25 to 75% and white circles indicate median values.

## PRS-based hybrid visual stimuli

The grand mean amplitude spectra of SSVEPs averaged across all participants are illustrated in Figure 7 as a function of waveforms of PRS-based hybrid visual stimuli. The red circles indicate the fundamental and harmonic frequencies, and the black circles represent the subharmonic frequencies. Unlike the FS-based hybrid visual stimuli, clear SSVEP peaks were observed at the subharmonic frequency for PRS-Square, PRS-Triangular, and PRS-Sine, as expected. The grand mean amplitudes of each SSVEP component are listed in Supplementary Table 2.

The grand mean amplitudes of SSVEP components at the subharmonic, fundamental, and second harmonic frequencies averaged over all stimulation frequencies across all the participants are illustrated in Figure 8 with respect to the periodic motion waveforms incorporated with PRS. The error bars represent the standard errors. The Friedman test indicated significant differences in amplitudes at all harmonic frequencies ($\chi^2 = 110.53$, $p < 0.001$; $\chi^2 = 10.69$, $p < 0.05$; $\chi^2 = 32.81$, $p < 0.05$). PRS-None evoked the lowest SSVEP amplitudes at the subharmonic frequency ($p < 0.001$ for PRS-Square, PRS-Triangular, and PRS-Sine, Wilcoxon signed-rank test with FDRs correction). In addition, the SSVEP amplitude induced by PRS-Triangular was significantly lower than that induced by PRS-Square and Sine (FDRs-corrected $p < 0.001$, Wilcoxon signed-rank test). For the fundamental frequency, PRS-Triangular induced significantly lower SSVEP amplitudes compared to other PRS-based hybrid visual stimuli and even conventional PRS ($p < 0.05$, Wilcoxon signed-rank test with FDRs correction). The SSVEP amplitudes elicited by PRS-None and PRS-Square were significantly higher than those elicited by PRS-Triangular and PRS-Sine at the second harmonic frequency ($p < 0.001$ between PRS-None and PRS-Triangular, and $p < 0.05$ for the others, Wilcoxon signed-rank test, FDRs-corrected).

Figures 9A,B depict the average classification accuracies and ITRs, respectively, for PRS-based hybrid visual stimuli with respect to different window sizes. Here, all the performances of SSVEP-based BCIs were evaluated using FBCCA for PRS-None and sFBCCA for PRS-Square, PRS-Triangular, and PRS-Sine cases. The Friedman test indicated statistically significant differences for window sizes of 1.5 and 2 s (1 s, $\chi^2 = 5.59$, $p < 0.133$; 1.5 s, $\chi^2 = 9.30$, $p < 0.05$; 2 s, $\chi^2 = 8.06$, $p < 0.05$; 2.5 s, $\chi^2 = 6.38$, $p < 0.1$; 3 s, $\chi^2 = 7.12$, $p < 0.1$; 3.5 s, $\chi^2 = 4.63$, $p = 0.201$, identical to both the classification accuracy and ITR). For all window sizes, PRS-Square showed the highest performance in terms of both classification accuracy and ITR, although statistically significant differences were not observed.

To investigate the effect of sFBCCA, the SSVEP-based BCI performances for PRS-Square were evaluated using FBCCA and sFBCCA with respect to different window sizes. In Figure 10, the white and gray bars indicate the averaged classification accuracies and ITRs evaluated using FBCCA and sFBCCA, respectively. The error bars represent standard errors. The SSVEP-based BCI performances evaluated using sFBCCA were significantly improved compared to those evaluated using FBCCA for every window size except 3.5 s (Wilcoxon signed-rank test). The result demonstrated that the proposed sFBCCA could significantly improve the performance of SSVEP-based BCIs when PRS-based hybrid visual stimuli are employed.

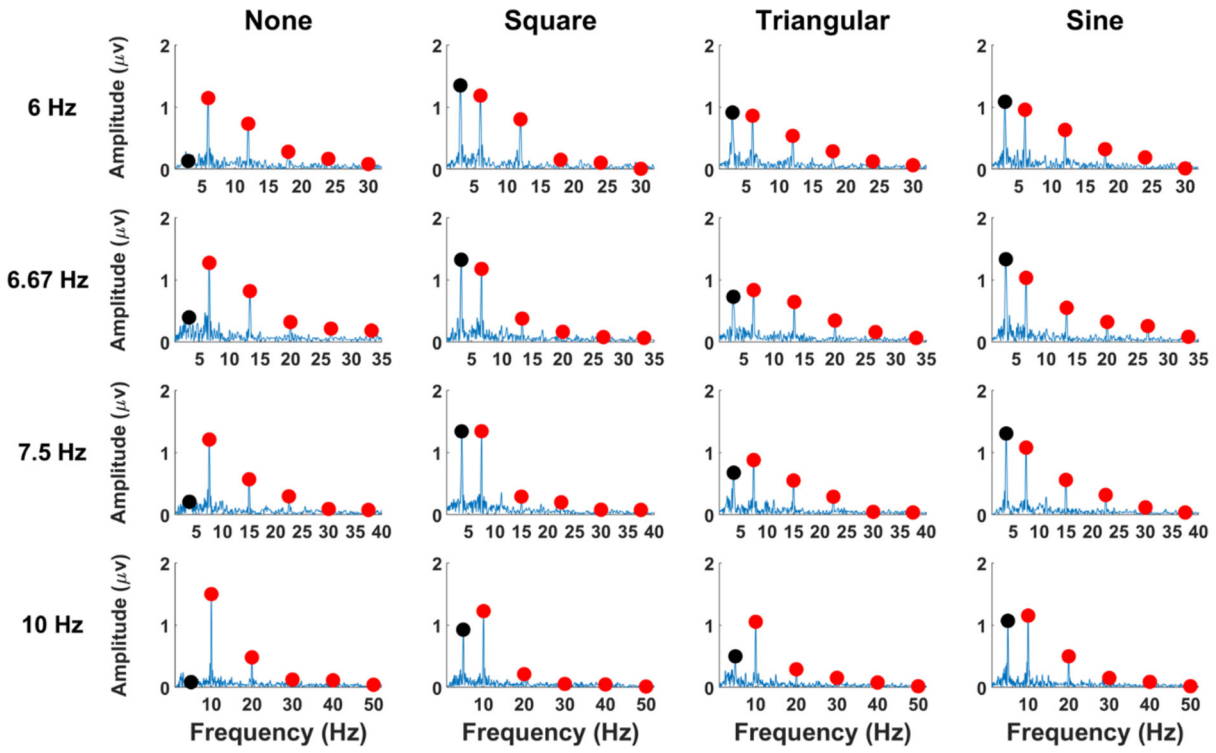



FIGURE 7
SSVEP amplitudes of averaged EEG signals over all participants for PRS-based hybrid visual stimuli at the Oz electrode. Red circles indicate the stimulation frequencies and their harmonics, and black circles represent the subharmonic frequencies.

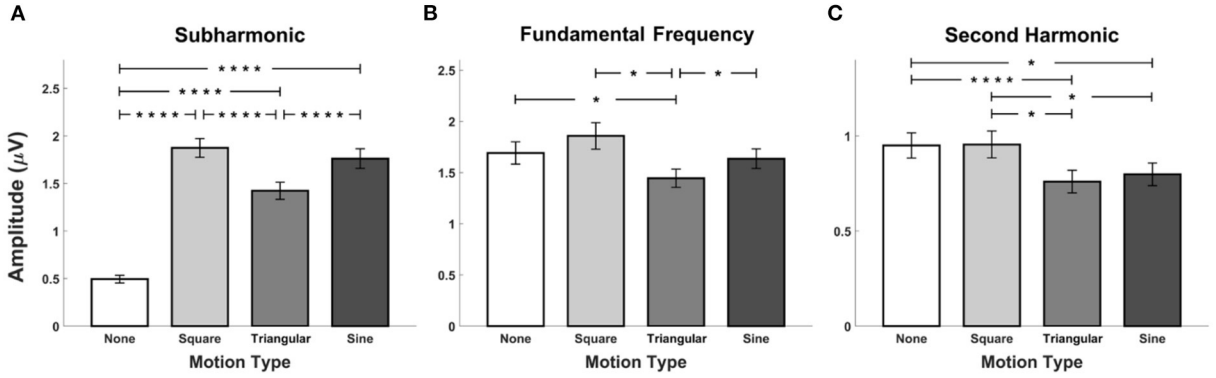



FIGURE 8
Grand mean SSVEP amplitudes for PRS-based hybrid visual stimuli averaged across all the participants at **(A)** the subharmonic frequency, **(B)** the fundamental frequency, and **(C)** the second harmonic frequency. Error bars represent the standard errors. Asterisks of * and **** represent FDRs-corrected $p < 0.05$ and $p < 0.001$, respectively (Wilcoxon signed-rank test).

The fatigue scores for PRS-based hybrid visual stimuli are illustrated in Figure 6B as a function of periodic motion waveforms. The gray bars represent the interquartile ranges from 25 to 75% and white circles indicate the median values. For PRS-None, PRS-Square, PRS-Triangular, and PRS-Sine, the averaged fatigue scores were reported as $3.85 \pm 1.63$, $4.05 \pm 1.28$, $4.55 \pm 1.43$, and $4.8 \pm 1.51$, respectively.

## Comparison between FS-sine and PRS-square

The average classification accuracies and ITRs for FS-Sine and PRS-Square, which exhibited the highest performances among FS- and PRS-based hybrid visual stimuli, are shown in Figures 11A,B, respectively. The differences in the average

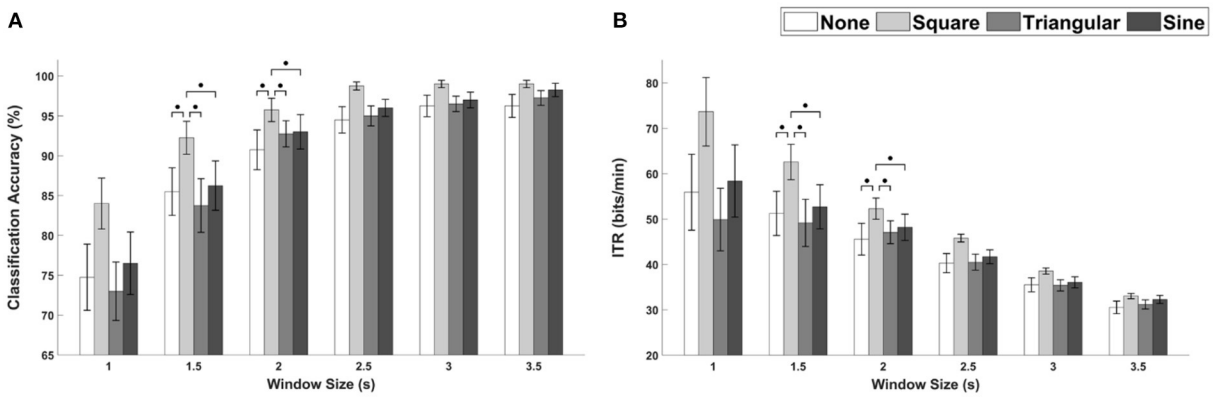



FIGURE 9
The average performance of PRS-based hybrid visual stimuli in terms of **(A)** classification accuracies and **(B)** ITRs as a function of waveforms with different window sizes. Error bars represent the standard errors. Here, $\bullet p < 0.1$ is the FDRs-corrected $p$-value from Wilcoxon signed-rank test.

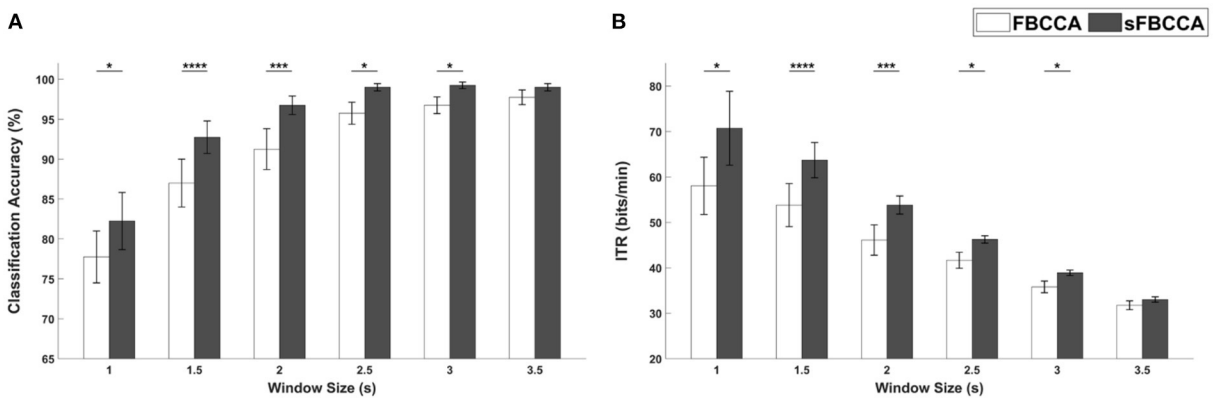



FIGURE 10
**(A)** Classification accuracies and **(B)** ITRs of SSVEP-based BCI for PRS-Square evaluated using FBCCA and sFBCCA with different window sizes. Error bars represent the standard errors. Here, *, ***, and **** represent $p < 0.05$, $p < 0.005$, and $p < 0.001$, respectively (Wilcoxon signed-rank test).

classification accuracies were reported to be 5.5, 2.5, 0.75, 0.25, 0.00, and 0.75%p for window sizes of 1, 1.5, 2, 2.5, 3, and 3.5 s, respectively. As for the ITRs, the differences were 12.93, 5.35, 1.60, 0.56, 0.00, 0.94 bits/min for the 1-, 1.5-, 2-, 2.5-, 3-, and 3.5-s window sizes. Statistically significant differences were not observed (Wilcoxon signed-rank test).

The violin plot in Figure 12 illustrates the fatigue scores for FS-Sine and PRS-Square. The distributions of fatigue scores from the first quartile to the third quartile are presented as gray bars and the median values are indicated as white circles. The average fatigue scores were reported to be 5.3 $\pm$ 2.05 and 4.05 $\pm$ 1.28 for FS-Sine and PRS-Square, respectively. A statistically significant difference in the fatigue score was observed between the FS-Sine and PRS-Square conditions ($p < 0.005$, Wilcoxon signed-rank test), implying that PRS-Square is more visually comfortable to the users than FS-Sine.

## Discussion

In this study, we proposed novel types of hybrid visual stimuli that incorporate periodic motions into conventional SSVEP visual stimuli. Periodic motions were realized by changing the size of the visual stimulus according to four different types of waveforms. We then investigated the effect of periodic motion waveforms for the hybrid visual stimuli on the performances of SSVP-based BCIs, for the first time. Our results demonstrated that the conventional SSVEP visual stimuli combined with appropriate periodic motions could increase the SSVEP amplitudes significantly, resulting in the enhancement of SSVEP-based BCI performances. For FS, the hybrid stimulus of FS-Sine elicited the highest SSVEP amplitudes at the fundamental and second harmonic frequencies, thereby resulting in the highest average performances in terms of

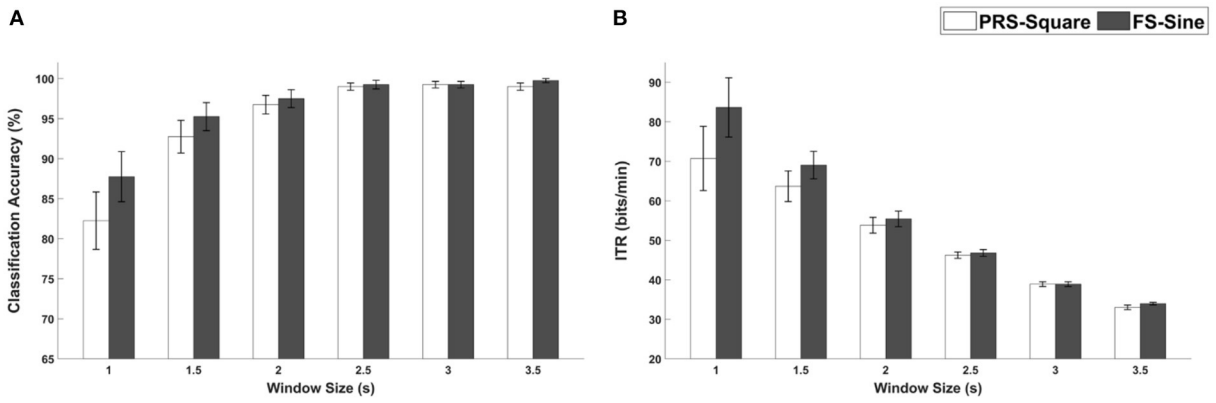



FIGURE 11
**(A)** Classification accuracies and **(B)** ITRs for PRS-Square and FS-Sine with different window sizes. Error bars represent the standard errors.

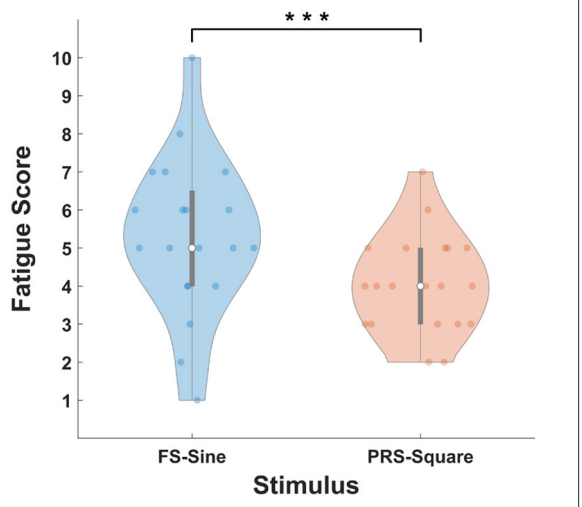



FIGURE 12
Average fatigue scores for FS-Sine and PRS-Square. Gray bars represent the interquartile range from 25 to 75% and white circles indicate the median values. Here, the asterisk of *** represents $p < 0.005$ (Wilcoxon signed-rank test).

classification accuracies and ITRs for every window size except for 3 s. As for PRS, PRS-Square evoked the highest SSVEP components, thereby exhibiting the highest performances for all window sizes. No statistically significant difference in the performances between FS-Sine and PRS-Square was observed; however, the visual fatigue score of PRS-Square was significantly lower than that of FS-Sine. Visual fatigue is one of the main obstacles to implementing practical SSVEP-based BCIs because visual fatigue generally decreases SSVEP amplitudes, yielding degradation of overall SSVEP BCI performances (Makri et al., 2015; Ajami et al., 2018). Therefore, our results suggest that the proposed PRS-Square is the most appropriate stimulus that could improve the SSVEP-based BCI performance without inducing high visual fatigue. It is believed that the use of PRS-Square stimuli has a great potential to improve the practicality of SSVEP-based BCIs, particularly for long-term use.

We hypothesized that the performances of SSVEP-based BCIs with any kind of hybrid visual stimuli could outperform those with conventional SSVEP visual stimuli because the hybrid visual stimuli could induce both SSVEP and SSMVEP. However, unlike our expectation, FS-Square and PRS-Triangular exhibited lower average classification accuracies and ITRs than the conventional visual stimuli for some window sizes. In addition, the periodic motion of the same waveform showed different effects on FS and PRS. For example, contrary to FS-Square, PRS-Square achieved the highest classification accuracies and ITRs for every window size, suggesting that it is important to combine conventional visual stimuli with periodic motions with appropriate waveforms for implementing high-performance SSVEP-based BCIs. We believe that these results might originate from different mechanisms of FS and PRS to evoke SSVEP. However, imaging modalities with higher spatial resolution such as fMRI would be necessary to further investigate the mechanisms of FS and PRS to evoke SSVEP. On the other hand, the luminance or pattern of FS and PRS was changed according to the square waveform because previous studies (Teng et al., 2011; Chen et al., 2019) demonstrated that square-wave FS achieved significantly higher classification performances than other waveform stimuli. However, as there is a possibility that the hybrid visual stimuli combining FS or PRS of other waveforms with periodic motions might improve the performance of SSVEP-based BCI, further investigations would be necessary for the future.

In contrast to the FS-based hybrid visual stimuli, the PRS-based hybrid visual stimuli were implemented by incorporating periodic motions whose stimulation frequency was half of PRS frequency. Although not mentioned in this manuscript, we also tested PRS-based hybrid visual stimuli with periodic motions of which the stimulation frequency was the same as

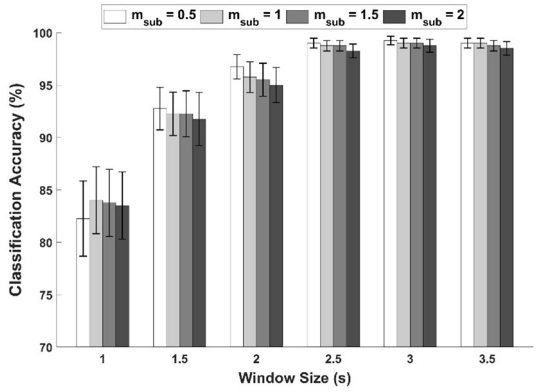



FIGURE 13
Average classification accuracies for PRS-Square as a function of $m_{sub}$ in sFBCCA with different window sizes.

that of PRS in our preliminary tests. For PRS with periodic motions of twice the frequency, most participants complained of severe visual fatigue and discomfort due to the rapid change in the stimulus size although almost the same classification accuracy as the PRS-based hybrid visual stimuli employed in this study was reported. As a result, the stimulation frequency of periodic motions was determined as half of PRS frequency. Since the use of the reduced stimulation frequency for periodic motions evoked subharmonic component, we extended the conventional FBCCA and proposed sFBCCA to fully exploit useful information contained in SSVEP evoked by the proposed PRS-based hybrid visual stimuli. The use of sFBCCA significantly enhanced the classification performances, compared to the results of FBCCA applied to PRS-Square, as shown in Figure 10. In the sFBCCA, the index of subharmonic, $m_{sub}$, was set to 0.5, which showed the highest classification accuracies for all window sizes except 1 s, as shown in Figure 13. However, there is still a possibility of further improvement of SSVEP-based BCI performances by optimizing sFBCCA parameters. Additionally, PRS-based hybrid visual stimulus has a promising possibility of increasing the number of commands limited by the refresh rate of the LCD monitor (Li et al., 2021), thanks to its characteristics of dual main stimulation frequencies induced by SSVEP and SSMVEP. This would be one of the promising topics we would like to further investigate in our future studies for implementing high-performance SSVEP-based 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 Institutional Review Board Committee of Hanyang University. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

JK designed the experiment, developed the algorithm, and analyzed the data. JH developed the visual stimulation program. HN performed experiments. C-HI supervised the study. JK and C-HI wrote the manuscript. All authors reviewed the manuscript.

## Funding

This work was supported in part by Hyundai Motor Group in 2021. 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. This work was also supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (No. NRF2019R1A2C2086593), and in part by the Institute for Information and communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (2017-0-00432).

## Conflict of interest

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

## Publisher's note

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

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fninf.2022.997068/full#supplementary-material

# References


Abiri, R., Borhani, S., Sellers, E. W., Jiang, Y., and Zhao, X. (2019). A comprehensive review of EEG-based brain–computer interface paradigms. *J. Neural Eng.* 16, 011001. doi: 10.1088/1741-2552/aaf12e

Ajami, S., Mahnam, A., and Abootalebi, V. (2018). An adaptive SSVEP-based brain-computer interface to compensate fatigue-induced decline of performance in practical application. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 2200–2209. doi: 10.1109/TNSRE.2018.2874975

Armengol-Urpi, A., and Sarma, S. E. (2018). "Sublime: a hands-free virtual reality menu navigation system using a high-frequency SSVEP-based brain-computer interface," in *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, (Tokyo, Japan: Association for Computing Machinery) 1–8. doi: 10.1145/3281505.3281514

Arpaia, P., De Benedetto, E., and Duraccio, L. (2021). Design, implementation, and metrological characterization of a wearable, integrated AR-BCI hands-free system for health 4.0 monitoring. *Measurement* 177, 109280. doi: 10.1016/j.measurement.2021.109280

Bieger, J., Molina, G. G., and Zhu, D. (2010). "Effects of stimulation properties in steady-state visual evoked potential based brain-computer interfaces," in *Proceedings of 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (Buenes Aires, Argentina: IEEE) 3345–3348.

Chen, X., Wang, Y., Gao, S., Jung, T.-P., and Gao, X. (2015). Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface. *J. Neural Eng.* 12, 046008. doi: 10.1088/1741-2560/12/4/046008

Chen, X., Wang, Y., Zhang, S., Xu, S., and Gao, X. (2019). Effects of stimulation frequency and stimulation waveform on steady-state visual evoked potentials using a computer monitor. *J. Neural Eng.* 16, 066007. doi: 10.1088/1741-2552/ab2b7d

Chen, Y., Yang, C., Ye, X., Chen, X., Wang, Y., and Gao, X. (2021). Implementing a calibration-free SSVEP-based BCI system with 160 targets. *J. Neural Eng.* 18, 046094. doi: 10.1088/1741-2552/ac0bfa

Choi, G.-Y., Han, C.-H., Jung, Y.-J., and Hwang, H.-J. (2019a). A multi-day and multi-band dataset for a steady-state visual-evoked potential–based brain-computer interface. *GigaScience* 8, giz133. doi: 10.1093/gigascience/giz133

Choi, K- M., Park, S., and Im, C.-H. (2019b). Comparison of visual stimuli for steady-state visual evoked potential-based brain-computer interfaces in virtual reality environment in terms of classification accuracy and visual comfort. *Comput. Intell. Neurosci.* 2019, 9680697. doi: 10.1155/2019/9680697

Dai, G., Zhou, J., Huang, J., and Wang, N. (2020). HS-CNN: a CNN with hybrid convolution scale for EEG motor imagery classification. *J. Neural Eng.* 17, 016025. doi: 10.1088/1741-2552/ab405f

Daly, J. J., and Wolpaw, J. R. (2008). Brain–computer interfaces in neurological rehabilitation. *Lancet Neurol.* 7, 1032–1043. doi: 10.1016/S1474-4422(08)70223-0

Duszyk, A., Bierzyńska, M., Radzikowska, Z., Milanowski, P., Kuś, R., Suffczyński, P., et al. (2014). Towards an optimization of stimulus parameters for brain-computer interfaces based on steady state visual evoked potentials. *PLOS ONE* 9, e112099. doi: 10.1371/journal.pone.0112099

Ge, S., Jiang, Y., Wang, P., Wang, H., and Zheng, W. (2019). Training-free steady-state visual evoked potential brain–computer interface based on filter bank canonical correlation analysis and spatiotemporal beamforming decoding. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 1714–1723. doi: 10.1109/TNSRE.2019.2934496

Herrmann, C. S. (2001). Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Exp. Brain Res.* 137, 346–353. doi: 10.1007/s002210100682

Jukiewicz, M., and Cysewska-Sobusiak, A. (2016). Stimuli design for SSVEP-based brain computer-interface. *Int. J. Electr. Telecomm.* 62, 109–113. doi: 10.1515/eletel-2016-0014

Kim, H., and Im, C.-H. (2021). Influence of the number of channels and classification algorithm on the performance robustness to electrode shift in steady-state visual evoked potential-based brain-computer interfaces. *Front. Neuroinform.* 15, 750839. doi: 10.3389/fninf.2021.750839

Kim, M., Kim, M.-K., Hwang, M., Kim, H.-Y., Cho, J., and Kim, S.-P. (2019). Online home appliance control using EEG-based brain–computer interfaces. *Electronics* 8, 1101. doi: 10.3390/electronics8101101

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

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

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., et al. (2018). A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *J. Neural Eng.* 15, 031005. doi: 10.1088/1741-2552/aab2f2

Makri, D., Farmaki, C., and Sakkalis, V. (2015). "Visual fatigue effects on steady state visual evoked potential-based brain computer interfaces," in *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER),* (Montpellier, France: IEEE). doi: 10.1109/NER.2015.7146562

Nakanishi, M., Wang, Y., Chen, X., Wang, Y. T., Gao, X., and Jung, T. P. (2018). 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., Wang, Y.-T., and Jung, T.-P. (2015). A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials. *PLoS ONE* 10, e0140703. doi: 10.1371/journal.pone.0140703

Park, S., Cha, H., and Im, C.-H. (2019). Development of an online home appliance control system using augmented reality and an SSVEP-based brain–computer interface. *IEEE Access* 7, 163604–163614. doi: 10.1109/ACCESS.2019.2952613

Perera, C. J., Naotunna, I., Sadaruwan, C., Gopura, R. A. R. C., and Lalitharatne, T. D. (2016). "SSVEP based BMI for a meal assistance robot," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, (Budapest, Hungary: IEEE) 002295–002300. doi: 10.1109/SMC.2016.7844580

Rabiul Islam, M., Khademul Islam Molla, M., Nakanishi, M., and Tanaka, T. (2017). Unsupervised frequency-recognition method of SSVEPs using a filter bank implementation of binary subband CCA. *J. Neural Eng.* 14, 026007. doi: 10.1088/1741-2552/aa5847

Teng, F., Chen, Y., Choong, A. M., Gustafson, S., Reichley, C., Lawhead, P., et al. (2011). Square or sine: finding a waveform with high success rate of eliciting SSVEP. *Comput. Intell. Neurosci.* 2011, 1–5. doi: 10.1155/2011/364385

Waytowich, N., Lawhern, V. J., Garcia, J. O., Cummings, J., Faller, J., Sajda, P., et al. (2018). Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials. *J. Neural Eng.* 15, 066031. doi: 10.1088/1741-2552/aae5d8

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

Xie, J., Xu, G., Wang, J., Zhang, F., and Zhang, Y. (2012). Steady-state motion visual evoked potentials produced by oscillating newton's rings: implications for brain-computer interfaces. *PLoS ONE* 7, e39707. doi: 10.1371/journal.pone.0039707

Xu, L., Xu, M., Jung, T.-P., and Ming, D. (2021). Review of brain encoding and decoding mechanisms for EEG-based brain–computer interface. *Cogn. Neurodyn.* 15, 569–584. doi: 10.1007/s11571-021-09676-z

Yan, W., Xu, G., Li, M., Xie, J., Han, C., Zhang, S., et al. (2017). Steady-state motion visual evoked potential (SSMVEP) based on equal luminance colored enhancement. *PLoS ONE* 12, e0169642. doi: 10.1371/journal.pone.0169642

Zerafa, R., Camilleri, T., Falzon, O., and Camilleri, K. P. (2018). To train or not to train? A survey on training of feature extraction methods for SSVEP-based BCIs. *J. Neural Eng.* 15, 051001. doi: 10.1088/1741-2552/aaca6e

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, Y., Xie, S. Q., Wang, H., Yu, Z., and Zhang, Z. (2020). Data analytics in steady-state visual evoked potential-based brain-computer interface: a review. *IEEE Sens. J.* 21, 1124–1138. doi: 10.1109/JSEN.2020.3017491

Zhao, J., Zhang, W., Wang, J. H., Li, W., Lei, C., Chen, G., et al. (2020). Decision-making selector (DMS) for integrating CCA-based methods to improve performance of SSVEP-based BCIs. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 1128–1137. doi: 10.1109/TNSRE.2020.2983275

Zhu, D., Bieger, J., Garcia Molina, G., and Aarts, R. M. (2010). A survey of stimulation methods used in SSVEP-based BCIs. *Comput. Intell. Neurosci.* 2010, 702357. doi: 10.1155/2010/702357

frontiers | Frontiers in Neurorobotics

TYPE Original Research
PUBLISHED 30 September 2022
DOI 10.3389/fnbot.2022.995552



OPEN ACCESS

EDITED BY
Ke Liu,
Chongqing University of Posts and Telecommunications, China

REVIEWED BY
Long Chen,
Tianjin University, China
Jiahui Pan,
South China Normal University, China

*CORRESPONDENCE
Guijun Chen
chenguijun@tyut.edu.cn
Xueying Zhang
tyzhangxy@163.com

RECEIVED 16 July 2022
ACCEPTED 06 September 2022
PUBLISHED 30 September 2022

CITATION
Chen G, Zhang X, Zhang J, Li F and Duan S (2022) A novel brain-computer interface based on audio-assisted visual evoked EEG and spatial-temporal attention CNN. *Front. Neurorobot.* 16:995552. doi: 10.3389/fnbot.2022.995552



# A novel brain-computer interface based on audio-assisted visual evoked EEG and spatial-temporal attention CNN

Guijun Chen*, Xueying Zhang*, Jing Zhang, Fenglian Li and Shufei Duan

College of Information and Computer, Taiyuan University of Technology, Taiyuan, China

**Objective:** Brain-computer interface (BCI) can translate intentions directly into instructions and greatly improve the interaction experience for disabled people or some specific interactive applications. To improve the efficiency of BCI, the objective of this study is to explore the feasibility of an audio-assisted visual BCI speller and a deep learning-based single-trial event related potentials (ERP) decoding strategy.

**Approach:** In this study, a two-stage BCI speller combining the motion-onset visual evoked potential (mVEP) and semantically congruent audio evoked ERP was designed to output the target characters. In the first stage, the different group of characters were presented in the different locations of visual field simultaneously and the stimuli were coded to the mVEP based on a new space division multiple access scheme. And then, the target character can be output based on the audio-assisted mVEP in the second stage. Meanwhile, a spatial-temporal attention-based convolutional neural network (STA-CNN) was proposed to recognize the single-trial ERP components. The CNN can learn 2-dimentional features including the spatial information of different activated channels and time dependence among ERP components. In addition, the STA mechanism can enhance the discriminative event-related features by adaptively learning probability weights.

**Main results:** The performance of the proposed two-stage audio-assisted visual BCI paradigm and STA-CNN model was evaluated using the Electroencephalogram (EEG) recorded from 10 subjects. The average classification accuracy of proposed STA-CNN can reach 59.6 and 77.7% for the first and second stages, which were always significantly higher than those of the comparison methods ($p < 0.05$).

**Significance:** The proposed two-stage audio-assisted visual paradigm showed a great potential to be used to BCI speller. Moreover, through the analysis of the attention weights from time sequence and spatial topographies, it was proved that STA-CNN could effectively extract interpretable spatiotemporal EEG features.

KEYWORDS

**brain-computer interface, audio-assisted visual evoked EEG, space division multiple access, spatial-temporal attention, convolutional neural network**

## Introduction

As an emerging human-computer interaction technique, the brain-computer interface (BCI) can realize the communication between the brain and the external devices without depending on the peripheral nervous and muscular tissues. The BCI can significantly improve the interaction experience for disabled people or some specific interactive applications including medical rehabilitation, healthcare, intelligent control, entertainment and so on (Chaudhary et al., 2016; Song et al., 2020). The scalp Electroencephalogram (EEG)-based BCI system has received more attention due to its easily used, relatively inexpensive, and high time resolution. Currently, some kinds of EEG signals with intentions modulated from the large neuronal activity are widely used in BCI systems including sensorimotor rhythm (SMR), steady-state visual evoked potential (SSVEP), and event-related potential (ERP). The SMR-based BCI usually requires a relatively long training time and even becomes ineffective after a certain amount of training for some users (Blankertz et al., 2010). The SSVEP-based BCI usually has a strong visual stimulation, which could cause the user's visual fatigue (Allison et al., 2014).

In the past few years, ERP-based BCIs have been widely investigated. One is the P300 speller, where a P300 component is elicited when the target character in a matrix is flashed with a small probability (Aloise et al., 2012). The P300 is a positive peak potential with a latency of about 300ms after the stimulus onset. To avoid flashing stimuli, the motion-onset visual evoked potential (mVEP) has been widely applied in BCI by attending to the target with a moving bar in an overt or covert way (Hong et al., 2009; Schaeff et al., 2012). The mVEP is composed of three main ERP components: P1 (P100), N2 (N200) and P2 (P200). The positive peak P1 with a latency of about 130 ms and the late negative peak N2 with a latency of 160–200 ms are considered as the main motion specific components (Zhang et al., 2015).

However, most of the ERP-based BCI must take a long time to output a target, where the stimuli must traverse all the target and nontarget with mutiple different time slices. To improve the detection speed, the dual-directional motion encoding paradigm was presented to reduce the stimuli presentation time by half (Liu et al., 2021). A new speller based on miniature asymmetric visual evoked potentials and space-code division multiple access (SDMA) scheme was developed, which can reduce stimuli time to implement an efficient BCI (Xu et al., 2018). For the SDMA scheme, the stimuli of targets and nontargets appear at different locations in the visual field simultanously, where an intended stimulus is attended to output the target quickly (Gao et al., 2014). Therefore, this study explored a new SDMA scheme to develop an efficient mVEP-based speller.

Compared with spontaneous EEG, the amplitude of single-trial ERP is so small that it is difficult to identify the target. Generally, to improve the signal-to-noise ratio (SNR) of ERP, averaging the EEG over several trials is used to obtain the discriminated ERP components. Nevertheless, it would decrease the output speed of the BCI system. An audiovisual hybrid BCI was designed to evoke stronger P100, N200, and P300 responses than the visual modality (Wang et al., 2015). The observed audiovisual integration effects can enhance the discriminability between target and nontarget brain responses. Moreover, an audiovisual P300-speller paradigm was proposed, which significantly improved the classification accuracies compared with the visual-based P300-speller (Lu et al., 2019). So, to enhance the quality of the ERP components, a semantically congruent audio-assisted mVEP paradigm was further used to output the target character in this study.

In addition, it is essential to decoding the ERP from a single-trial EEG to achieve fast and accurate target output. In some methods, the ERP components and spontaneous EEG were separated from a single-trial EEG based on a priori ERP pattern using wavelet transform (WT) (Quiroga, 2005), independent component analysis (ICA) (Lee et al., 2016) and so on. An iterative principal component analysis (PCA) method was proposed to extract single-trial ERP by reconstructing the principal components with a higher correlation with the target ERP (Mowla et al., 2016). Other methods aimed to improve the classification performance of single-trial ERP. The linear discriminant analysis (LDA) usually worked well for single-trial ERP classification. However, an accurate covariance matrix estimation was required in high-dimensional feature spaces. A shrinkage LDA was proposed to achieve excellent results for single-trial ERP classification (Blankertz et al., 2011). Meanwhile, a spatial-temporal discriminant analysis (STDA) algorithm was introduced to learn spatial and temporal projection matrices collaboratively by adopting matrix features, and the ill-conditional problem of covariance matrix can be effectively solved (Zhang et al., 2013). To enhance the SNR of ERP and classification accuracy simultaneously, current detection methods of single-trial ERP were reviewed, and the best performance of the xDAWN-based spatial filter and Bayesian LDA method was proved during a rapid serial visual presentation task (Cecotti and Ries, 2017). A data-adaptive spatiotemporal filtering method based on the clustering and array WT was proposed to improve the discriminative features of single-trial ERP (Molla et al., 2018). To adapt to the ERP diversities, the discriminative canonical pattern matching (DCPM) was proposed and obtained outperformed classification performance for the single-trial classification of EEG datasets including P300, mVEP components and so on (Xiao et al., 2020).

Recently, deep learning has been demonstrated that it can deal with EEG feature learning and classification effectively (Amin et al., 2019). A convolutional neural network (CNN) with a layer dedicated to spatial filtering was proposed to detect

the single-trial ERP (Cecotti et al., 2014). The EEGNet, using the depthwise and separable CNN, was introduced to construct an EEG-specific model, which achieved comparably high performance for within-subject and cross-subject classification (Lawhern et al., 2018). Furthermore, a novel CNN model was proposed to better use the phase-locked characteristic to extract spatiotemporal features for single-trial ERP classification (Zang et al., 2021). However, due to the inter-trial and inter-subject variability of single-trial ERP, it is still challenging to build an efficient decoding strategy for single-trial ERP. Current studies have suggested that large inter-trial and inter-subject differences exist in the amplitude and latency of ERP components. So, it becomes crucial to construct an adaptive learning model to extract the spatial-temporal features from single-trial EEG.

In sum, there are still some current challenges to the application of the EEG-based BCI, including the friendly cognitive load and EEG characteristics-guided BCI classification algorithms (Xu et al., 2021). Compared with the flashing or flickering visual BCIs, the mVEP is a convenient way to encode targets with briefly moving stimuli (Libert et al., 2022b). On single trial classification, CNN can achieve comparable performance to both the LDA and support vector machine, but slightly less stable and interpretable (Vareka, 2020). In this study, similar to the Hex-o-Spell (Treder and Blankertz, 2010), a two-stage overt attention BCI speller combining with the mVEP and semantically congruent audio evoked ERP was designed to output a target by taking advantage of audiovisual properties. The main contributions of this paper are as follows.

(1) In the first stage, the different character groups coded with mVEP were presented simultaneously in the different locations of the visual field based on a new SDMA scheme to improve the efficiency of visual stimuli presentation.
(2) The target character was selected based on the audio-assisted mVEP in the second stage, which can enhance the quality of the ERP components.
(3) The spatial-temporal attention-based CNN (STA-CNN) was proposed to deal with single-trial ERP components learning and classification. The STA-CNN can effectively extract interpretable spatiotemporal EEG features by adaptively learning probability weights.

The rest of the paper is organized as follows: materials and methods are demonstrated in Section Materials and methods. Then experiment results of our proposed BCI speller are presented in Section Experiment results.. Finally, the discussion and conclusion of this paper are provided in Section Discussion and conclusions.

# Materials and methods

## Two-stage audio-assisted visual BCI paradigm

This study implemented a two-stage audio-assisted visual copy-spelling BCI, as shown in Figure 1. The paradigm was designed by using the Psychtoolbox in the Matlab 2012b environment. The visual stimuli were presented on a 17-inch LCD monitor with a 60 Hz refresh rate and 1440 × 900 pixels resolution. The audio stimuli were played by the headphone at a sensible volume.

In the first stage, forty alphanumeric characters were divided into 8 groups with 5 characters in each group, as shown in Figure 1A, wherein the size of each character group area was 280 × 280 pixels, and the size of each motion visual stimulus (red vertical bar) was 10 × 80 pixels. The target character group was selected based on the mVEP with a new SDMA scheme. For the SDMA scheme, three sub-trial motion visual stimuli sequences constituted eight parallel spatial channels. In each group, the red vertical bar appeared on the left side and moved rightward until it reached the right side, which lasted for 0.3 s as a brief motion-onset stimulus. Specifically, the motion-onset stimulus from left to right was regarded as code "1", while no motion-onset stimulus was regarded as code "0". The interval between two successive motion-onset stimuli was 0.2 s, and a complete stimulation sequence lasted for 1.5 s. Eight groups of code sequences were allocated to different character groups, as shown in Figure 1B. Take character A as an example, and its group code is "101". That is, the 'moving bar—none—moving bar' was presented by turns in the location of the top left group. During the spelling period, the motion-onset stimuli would be presented simultaneously for all character groups with different code sequences. The spatial information is embedded in the group codes. After three sub-trials of motion-onset stimuli in the first stage, the target character group would be selected.

Upon choosing of a character group, the speller switches to the second stage, and the target character can be selected based on the audio-assisted mVEP. One example of the second stage paradigm, including motion-onset stimuli and pronunciation of character A, is shown in Figure 1C. The motion-onset stimuli and semantically congruent audio (pronunciation) for each character would be presented simultaneously. During the presentation of audiovisual stimuli, when a moving bar was presented on a character, the pronunciation of the character was played by the headphone. Each group contained 5 characters, and the audiovisual stimuli of each character were presented randomly for 0.3 s with a time interval of 0.2 s. A complete stimulation sequence lasted for 2.5 s. Specifically, the audiovisual stimulation was similar to the "oddball" paradigm, and the target stimulation produced a P300 response. In the stimulation interface, the size of each character area was 170 × 250

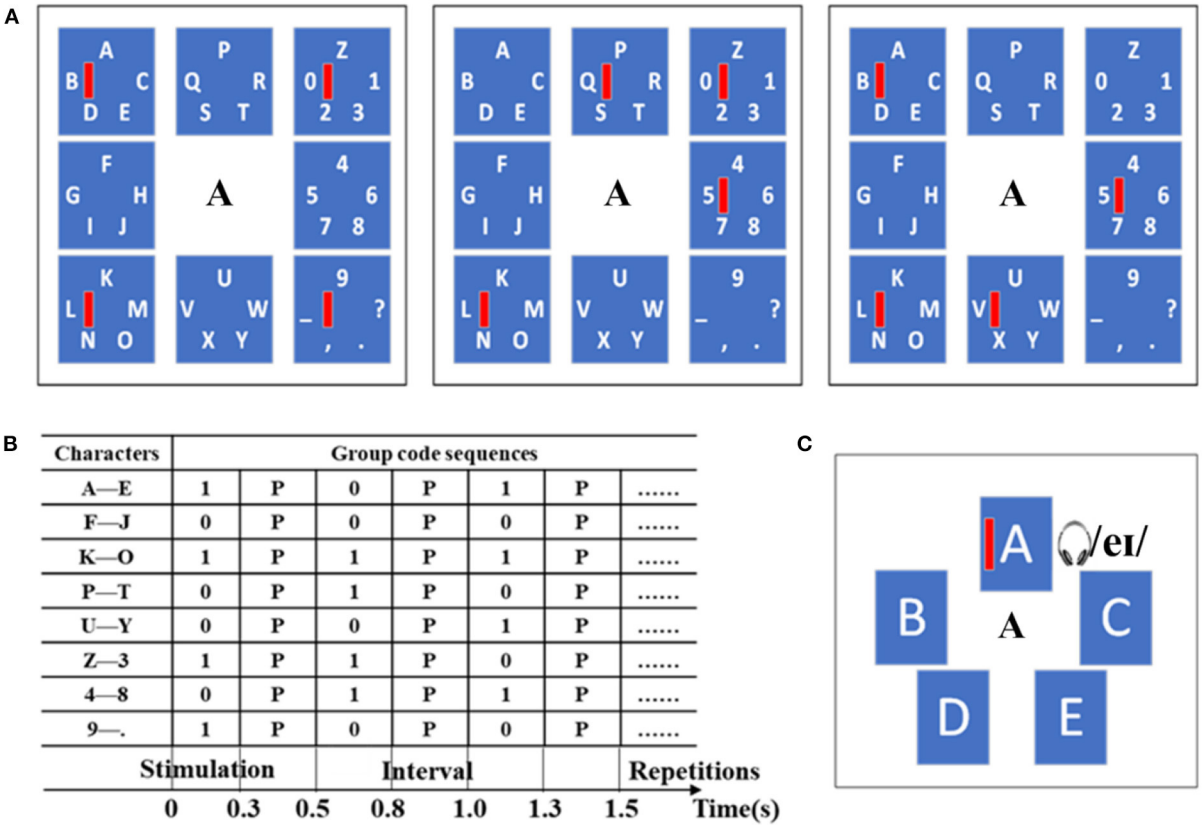



FIGURE 1
Speller paradigm example. **(A)** The first stage paradigm includes 3 sub-trials motion-onset stimuli based on the SDMA scheme. **(B)** A list of group code sequences for the first stage. **(C)** One example of the second stage paradigm includes motion-onset stimuli and pronunciation of character A simultaneously.

pixels, and the size of each motion visual stimulus was 10 × 100 pixels.

## Subjects and experimental procedure

Ten healthy volunteers (22–26 years of age, 7 males, all right-handed) with normal hearing and normal or corrected to normal vision participated in this study. The experimental procedures were performed in accordance with the Declaration of Helsinki. The written informed consent was obtained from all subjects before the experiments, and the required tasks of the study were explained. After the experiments, the subjects received money for their participation. A total of 467 characters, including 10 sentences, were spelled in the copy-spelling task for each subject, with a 2 min rest between the sentences.

During the experiment, subjects were seated 50 centimeters in front of the LCD monitor. When a target character was introduced, it was shown on the screen center. In the first stage, the subjects were asked to pay attention to the center of the character group where the target character is located. In the second stage, the subjects were asked to pay attention to the target character. During the experiment, the subjects were asked to keep their heads as still as possible and blink less. And then, EEG was recorded using the Neuroscan SynAmps2 system with 64 channels referring to the international 10–20 electrode positions (Xu et al., 2018). The reference electrode was put in the position near Cz, and the ground electrode was put in the position near Fz. The impedance between the scalp and the electrode is $<10\ k\Omega$. The recorded EEG was bandpass-filtered at 0.1–100 Hz, sampled at a rate of 1000 Hz, and then stored in a computer.

After the EEG data were acquired, the recorded EEG data were re-referenced to the average of the bilateral mastoids (M1 and M2), filtered by a band-pass filter at 1–30 Hz, and down-sampled at 200 Hz. A 0.6 s time window was used to extract event-related data frames from – 0.1 to 0.5 s after stimulus onset, and 0.1 s baseline correction was applied in the first and the second stages. The format of a single trial EEG data in both two stages was a matrix of 62 channels × 100 time samples.

## Spatial-temporal attention CNN model

To enhance the discriminative event-related features from spatial-temporal domains, the spatial-temporal attention CNN (STA-CNN) model is proposed, which consists of four modules, as shown in Table 1.

The module 1 is mainly used for temporal filtering, which contains a reshape layer, a convolutional layer (Conv2D), and a batch normalization (BN) layer. The reshape layer transforms the EEG data into the input format of the Conv2D layer. And then, we perform a convolutional step in time sequence, and a 2D convolutional filter of size (1, 50) and stride 1 is used to output 16 feature maps containing the EEG data at different band-pass frequencies. The time length of the output is still 100 due to a 2D zero-padding of size (24, 25, 0, 0). In addition, the BN layer is performed before the activation function to avoid the distribution shift (Ioffe and Szegedy, 2015), and the exponential linear unit (ELU) activation function is used.

The module 2 performs the temporal features extraction, including a temporal attention layer, a Conv2D layer, and a BN layer. In the temporal attention layer, we adopt the adaptive event-related features learning, which can assign weights to different time samples based on importance. Suppose the feature maps $T \in R^{Ns\times Nf\times Nc\times Nt}$ from the module 1, we first apply a grand average pooling (GAP) for each time sample from different channels to obtain temporal-wise statistics $\overline{T} \in R^{Ns\times Nf\times 1\times Nt}$, where $N_s$ is the batch size, $N_f$ is the number of filters, Nc is the number of channels, $N_t$ is the number of time points. The temporal attention mechanism adopts two fully-connected (FC) layers, including a dimensionality-reduction Linear layer 1 with tanh activation function and a dimensionality-increasing Linear layer 2, to reduce model complexity and improve generalizability. Thus, the temporal attention mechanism is expressed as follows.

$$A_t = softmax\left(Linear2\left(\tanh\left(Linear1\left(\overline{T}\right)\right)\right)\right) \quad (1)$$

where the softmax function transforms the importance of time points to a probability distribution. Finally, we consider probability as the weight to recode the feature maps $T$ at each time point. Thus, the attentive temporal feature can be represented as follows.

$$T_a = T \bullet A_t \quad (2)$$

The design of temporal attention on different periods utilizes relatively stable latency of event-related features for different channels. Then, we further perform a convolutional step in time sequence, and a 2D convolutional filter of size (1, 51) and stride 1 is used to output 32 feature maps. The time length of output becomes 50 to reduce the temporal dimension. And the BN layer is performed before the ELU activation function. To prevent over-fitting, we use the Dropout technique (Srivastava et al., 2014) and set the dropout probability to 0.5.

To further extract spatial information from the feature maps, the module 3 consists of a spatial attention layer, a Conv2D layer, and a BN layer. Similar to the temporal attention layer, the spatial attention layer assigns weights to different channels based on importance. Suppose the feature maps $S \in R^{Ns\times Nf\times Nc\times Nt}$ from the module 2, we apply a GAP for each channel of feature maps to obtain channel-wise statistics $S \in R^{Ns\times Nf\times Nc\times 1}$. The spatial attention mechanism also adopts two FC layers 3 and 4, which are expressed as follows.

$$A_s = softmax\left(Linear4\left(\tanh\left(Linear3\left(\overline{S}\right)\right)\right)\right) \quad (3)$$

Finally, we consider probability as the weight to recode the feature maps $S$ in each channel as follows.

$$S_a = S \bullet A_s \quad (4)$$

Compared with the traditional channel attention (Woo et al., 2018), this study only utilizes the average pooling instead of the sum of average and maximum pooling to become insensitive to the noise in EEG feature learning. Then, to learn a spatial filter, we further perform a 2D convolutional filter of size (62, 1) and stride 1 to output 4 feature maps. The BN layer is used before the ELU activation function. A maximum pooling layer of size (1, 5) and stride 5 is utilized to reduce the feature dimensions. To prevent over-fitting, we use the Dropout technique and set the dropout probability to 0.5.

In the module 4, after feature maps are flattened into vectors, a dense layer with the softmax function is used as the classifier of the model. The output size of the dense layer is set to 2, which corresponds to the target and non-target classes.

In summary, we have designed a model, as shown in Figure 2 to extract spatial-temporal features and classification from single-trial EEG data. The model was trained using the Adam optimizer and the categorical cross-entropy loss function in PyTorch. We ran 300 training iterations and performed validation stopping, saving the model weights when we got the lowest loss of validation set.

# Experiment results

## ERP components analysis

The performance of the proposed two-stage audio-assisted visual BCI paradigm and the STA-CNN model was evaluated using the EEG recorded by our experiment in Section Materials and methods. A total of 467 characters, including 10 sentences, were spelled for each subject. Hence, 714 target EEG segments and 687 nontarget EEG segments in the first stage, and 467 target EEG segments and 1868 nontarget EEG segments in the second stage were recorded for each subject.

We firstly analyzed the ERP components evoked from the audio-assisted visual BCI paradigm. The grand average of

TABLE 1 Parameters setting of STA-CNN model.

| Module | Layer | #Filters | Size | Strides | Output | Options |
|---|---|---|---|---|---|---|
| 1 | Input | | | | (62, 100) | |
| | Reshape | | | | (1, 62, 100) | |
| | Conv2D | 16 | (1, 50) | 1 | (16, 62, 100) | Padding = (24, 25) |
| | BatchNorm | | | | (16, 62, 100) | |
| | Activation | | | | (16, 62, 100) | ELU |
| 2 | TemporalAttention | | | | (16, 62, 100) | |
| | Conv2D | 32 | (1, 51) | 1 | (32, 62, 50) | |
| | BatchNorm | | | | (32, 62, 50) | |
| | Activation | | | | (32, 62, 50) | ELU |
| | Dropout | | | | (32, 62, 50) | $P = 0.5$ |
| 3 | ChannelAttention | | | | (32, 62, 50) | |
| | Conv2D | 4 | (62, 1) | 1 | (4, 1, 50) | |
| | BatchNorm | | | | (4, 1, 50) | |
| | Activation | | | | (4, 1, 50) | ELU |
| | MaxPool2d | | (1, 5) | 5 | (4, 1, 10) | |
| | Dropout | | | | (4, 1, 10) | $P = 0.5$ |
| 4 | Flatten | | | | (40) | |
| | Dense | | | | (2) | Softmax |

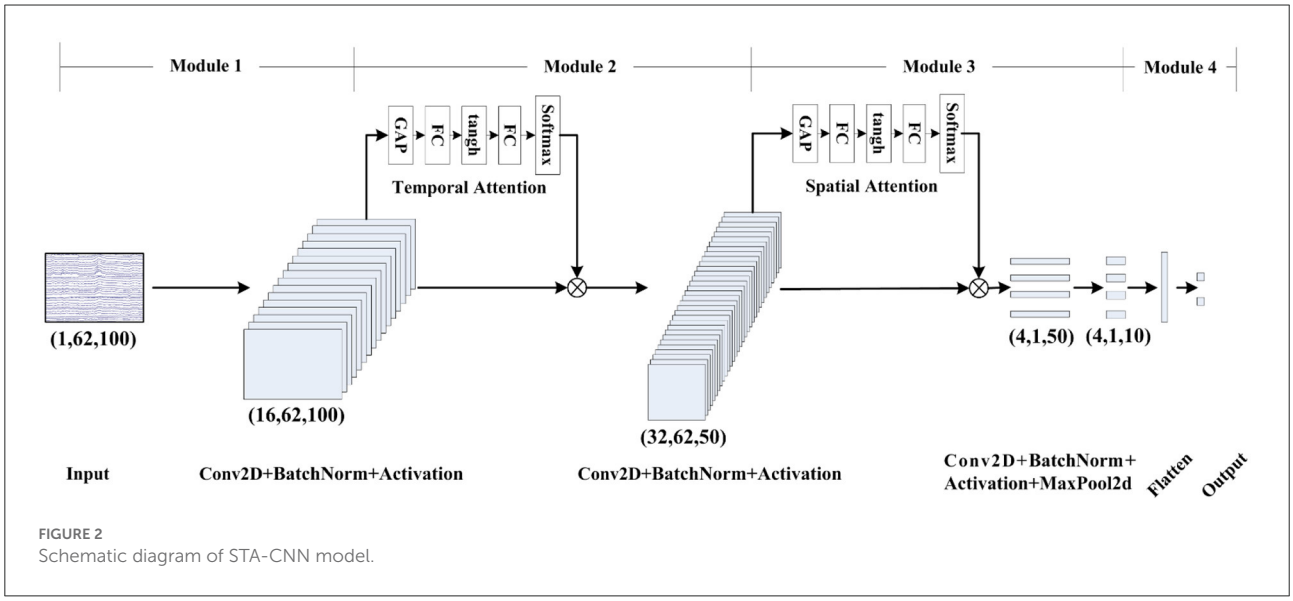



FIGURE 2
Schematic diagram of STA-CNN model.

the target and nontarget EEG epochs in the first stages and the second stages for each subject were calculated separately. Figure 3 illustrates the averaged scalp potentials on 62 channels and topographies from the target and nontarget EEG epochs for one example subject S10. The amplitude of some time periods of target-related signal in Figures 3A,B is obviously higher than that of nontarget-related signal in Figures 3C,D. The three main time periods in Figure 3A with latencies of about 90, 130, and 195 ms after the motion-onset stimulus could be categorized as the P1, N2, and P2 components of mVEP. Compared with Figure 3A, the latencies of the mVEP components in Figure 3B become shorter, which are about 50, 90, and 180 ms after the motion-onset stimulus. This may be due to the earlier picked-up target cues in the first stage and the congruent audiovisual integration effect (Hessler et al., 2013; Simon and Wallace, 2018). Besides the mVEP components, there is an obviously larger amplitude with a latency of 350 ms in Figure 3B, which could be categorized to the P300 component evoked by the audio-assisted visual oddball paradigm. The above-mentioned P1, N2, P2, and P300

components are distributed in the central, temporo-occipital, and associate parietal cortical areas and dominate in the right hemisphere, which are consistent with the previous findings about mVEP and P300 components (Kuba et al., 2007; Guo et al., 2008; Belitski et al., 2011).

Moreover, to further analyze the ERP components during the character spelling process, taking character A as an example, the grand average of target and nontarget related signals in the first stage and the second stage on channel P4 are shown in Figure 4. The reason for choosing channel P4 is based on the significant differences between target and nontarget related signals shown in Figure 3 and the previous findings (Zhang et al., 2015). In the first stage, the three epochs could be coded as "101" for the group code sequence of characters A-E. The amplitudes of the target-related mVEP components, including P1, N2, and P2 (color area) during the first and third epochs are significantly higher than that of the nontarget-related signal during the second epoch, as shown in Figure 4A, where the mVEP components from the first 500 ms and the third 500 ms epochs are coded as "1" and the nontarget signal in the middle 500 ms epoch is coded as '0'. In the second stage, besides the P1, N2, and P2 components, the audio-assisted visual target stimuli can evoke obviously P300, while the nontarget stimuli had no obvious ERP components. Therefore, the proposed two-stage audio-assisted visual stimulus paradigm shows great potential to be used for BCI speller.

## Single-trial ERP classification performance

Due to the different spatial-temporal characteristics of the ERP components in both two stages, the STA-CNN classification models were constructed based on the single trial EEG data in the first stage and the second stage, respectively. For the integrity of spelling characters, 448 target EEG segments and 452 nontarget EEG segments for 300 characters in the first stage were used as the training set for each subject, ten percent of which were used as the validation set. Similar to the first stage, 300 target EEG segments and 1200 nontarget EEG segments in the second stage were used as the training set, 10% of which were used as the validation set. For the training set, target samples were replicated 3 times to ensure the same number of samples from the two categories in the second stage, which could avoid model deviation caused by an unbalanced sample number. The remaining EEG segments for 167 characters were used as the test set in both stages. The classification performance was evaluated by the metrics: accuracy and F1-score.

In this part, the proposed STA-CNN was compared with several LDA variants and deep learning methods, including STLDA, DCPM, DeepLDA (Wu et al., 2017) and EEGNet, to validate the single-trial ERP classification performance. For the above comparison methods, the model parameters in this study were set by referring to the original literature. Tables 2, 3 present the classification results in terms of accuracy and F1-score for each subject using the above five methods. The classification accuracy of the second stage with the audio-assisted visual stimulus is higher than that of the first stage with visual stimulus. The overall classification results vary with different subjects, and subjects 2, 4, 7 and 10 could obtain higher spelling performance. The average classification accuracy of proposed STA-CNN across all subjects can reach 59.6% and 77.7% in the first and second stages, which are always higher than those of the comparison methods. The paired samples $t$-test was utilized to verify whether there were significant differences in classification performance between STA-CNN and other comparison methods. The results show that the STA-CNN can obtain significantly higher accuracy ($p < 0.01$) and F1-score (STLDA: $p = 0.03 < 0.05$, DCPM: $p = 0.041 < 0.05$, DeepLDA: $p = 0.014 < 0.05$, EEGNet: $p = 0.143$) in the first stage, while the STA-CNN can obtain significantly higher accuracy ($p < 0.01$) and F1-score (EEGNet: $p = 0.027 < 0.05$, others $p < 0.01$) in the second stage.

Furthermore, according to the trained STA-CNN models in the first and second stages, we provide the total classification accuracy of the above-mentioned test set (EEG segments for 167 characters) to evaluate the effectiveness of the paradigm and classification method. In the first stage, the group (during 3 sub-trials) with the group code corresponding classifier output was chosen, and in the second stage, the character (out of 5 characters) with the highest classifier output was chosen. The total classification accuracy of 10 subjects is listed in Table 4. Notice that the chance level is 1/40 = 2.5% for the two-stage spelling paradigm. The total classification accuracy varied with different subjects and ranged from 27.0 to 61.7%. Herein, the total classification accuracy is greatly affected by the first stage, and once the spelling error occurs in the first stage, it should be returned to the group selection during the actual spelling process.

Meanwhile, we provided the ablation study to validate the effectiveness of the spatial-temporal attention module in the STA-CNN method. The CNN is the baseline model that removes the spatial and temporal attention modules. The TA-CNN is the model that removes the spatial attention module from STA-CNN. The SA-CNN is the model that removes the temporal attention module from STA-CNN. The structures and parameters of all these three models were set according to the STA-CNN model in Table 1. As shown in Table 4, compared with CNN, TA-CNN and SA-CNN achieve better performance, which validates the effectiveness of the spatial and temporal attention modules. Combing with the spatial-temporal attention module, the STA-CNN is more effective than the TA-CNN and SA-CNN. The TA-CNN extract effective target ERP features based on the difference of special time periods between the target and nontarget EEG signals, which can obtain higher classification

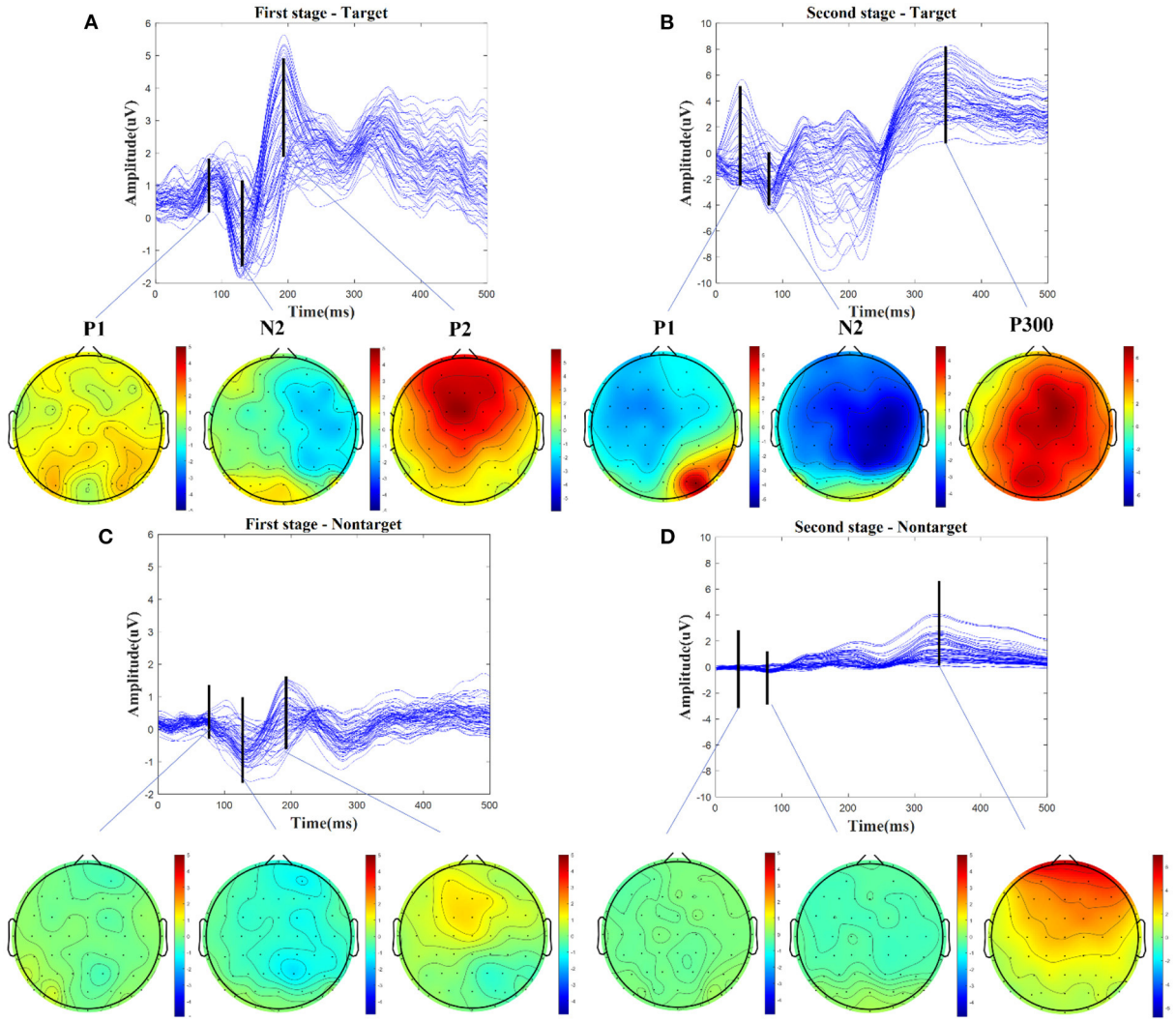



FIGURE 3
Averaged scalp potentials on 62 channels and topographies from the target and nontarget EEG epochs, for one example subject S10. **(A)** The mVEP components of P1, N2, and P2 and their topographies are in the first stage. **(B)** The audio-assisted visual evoked components of P1, N2, and P300 and their topographies are in the second stage. **(C)** The nontarget-related signal and topographies are in the first stage. **(D)** The nontarget-related signal and topographies are in the second stage.

accuracy than that of SA-CNN. And then, the paired samples *t*-test was utilized to verify whether these methods had significant differences. The results show that the STA-CNN can obtain significantly higher accuracy than other comparison methods (CNN: $p = 0.001 < 0.01$, TA-CNN: $p = 0.002 < 0.01$, SA-CNN: $p = 0.005 < 0.01$).

## Influence of spatial-temporal attention

The deep learning methods can automatically learn the EEG features, but it is difficult to determine if the spatial-temporal characteristics of ERP have been extracted efficiently. The spatial-temporal attention becomes essential to learn the individual spatial filters for particular time periods. In order to show the influence of spatial-temporal attention, Figure 5 shows the average weights of temporal and spatial attention from the test samples based on the STA-CNN model for the two stages.

In the first stage, as shown in Figure 5A, the results show that there are higher temporal weights nearby the time periods of 90, 130, and 210 ms, and higher spatial weights located at temporo-occipital cortical and parietal cortical areas. The time periods with higher temporal weights are similar to the latencies of mVEP in Figure 3A, and the higher spatial weights reflect the differences between

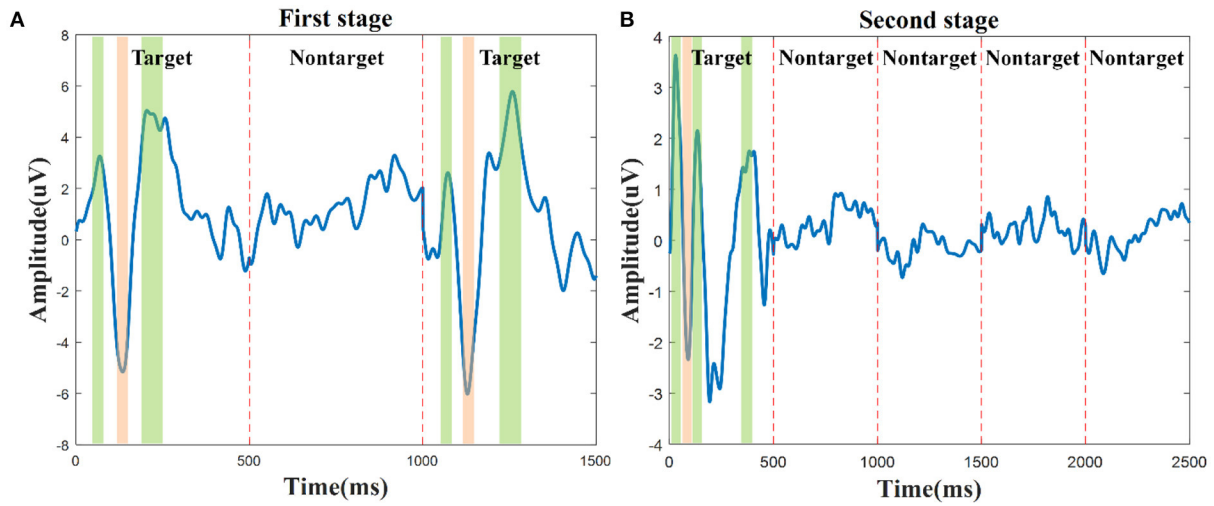



FIGURE 4
The grand average of target and nontarget signals during the character A spelling process for S10 was recorded on channel P4. **(A)** The mVEP of the group characters A-E in the first stage, P1, N2, and P2 are the three main target components. **(B)** The audio-assisted visual ERP components of character A in the second stage, P1, N2, P2, and P300 are the main target components.

TABLE 2 The classification accuracy and F1-score of 10 subjects using five methods in the first stage (%).

| Subject | Accuracy | | | | | F1-score | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | STLDA | DCPM | DeepLDA | EEGNet | STA-CNN | STLDA | DCPM | DeepLDA | EEGNet | STA-CNN |
| S1 | 56.3 | 56.7 | 54.9 | 57.9 | 59.1 | 60.7 | 60.6 | 55.2 | 61.6 | 59.7 |
| S2 | 74.9 | 74.3 | 77.1 | 78.0 | 78.2 | 76.0 | 75.1 | 78.1 | 79.1 | 78.6 |
| S3 | 50.9 | 51.5 | 53.3 | 53.3 | 54.3 | 53.2 | 54.4 | 54.5 | 55.9 | 59.9 |
| S4 | 60.3 | 60.1 | 62.1 | 62.5 | 63.7 | 58.8 | 60.6 | 59.1 | 60.2 | 62.6 |
| S5 | 50.3 | 50.5 | 51.5 | 53.3 | 56.5 | 51.0 | 49.3 | 50.3 | 52.1 | 56.4 |
| S6 | 48.7 | 50.3 | 52.3 | 50.5 | 50.5 | 52.0 | 54.3 | 52.1 | 47.5 | 50.4 |
| S7 | 58.7 | 58.3 | 60.1 | 60.5 | 62.7 | 62.8 | 61.5 | 62.3 | 64.8 | 60.3 |
| S8 | 50.5 | 50.5 | 50.5 | 52.1 | 53.7 | 52.6 | 50.4 | 52.9 | 53.3 | 55.2 |
| S9 | 50.5 | 52.1 | 51.7 | 54.1 | 54.5 | 51.9 | 54.6 | 53.8 | 56.6 | 59.0 |
| S10 | 56.9 | 54.5 | 57.3 | 61.1 | 63.3 | 59.3 | 57.0 | 57.4 | 61.4 | 64.6 |
| Mean ± Std | 55.8 ± 7.8 | 55.9 ± 7.3 | 57.1 ± 8.0 | 58.3 ± 8.1 | 59.6 ± 8.0 | 57.8 ± 7.7 | 57.8 ± 7.4 | 57.6 ± 8.0 | 59.2 ± 8.7 | 60.7 ± 7.4 |

target and non-target activated brain regions in Figures 3A,C. Similar to the first stage, as shown in Figure 5B, there are also higher temporal weights nearby the time periods of 50, 100, 180, and 310 ms, and higher spatial weights located at occipital and right temporo-parietal cortical areas in the second stage, which are similar to the latencies of ERPs and the differences between target and nontarget activated brain regions in Figures 3B,D. These results are consistent with the ERP components analysis in Section Introduction and Figure 3. The proposed STA-CNN benefits from spatial-temporal weights of attention mechanism to learn the ERP features effectively, and thus it can achieve superior performance.

## Discussion and conclusions

The mVEP-based paradigm is suitable for BCI speller application because it can encode the intentions as the identifiable target components and does not make subjects feel visual fatigue even for a long-time use (Liu et al., 2019). However, the problems restricting the practical application of mVEP-based BCI are the coding efficiency of the large command set and the decoding accuracy of the single-trial ERP due to low SNR (Lotte et al., 2018; Xu et al., 2020). Similar to telecommunication systems, the multiple targets coding strategy aims to simultaneously share the bandwidth from time, frequency, code and space with the least performance

TABLE 3 The classification accuracy and F1-score of 10 subjects using five methods in the second stage (%).

| Subject | Accuracy | | | | | F1-score | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | STLDA | DCPM | DeepLDA | EEGNet | STA-CNN | STLDA | DCPM | DeepLDA | EEGNet | STA-CNN |
| S1 | 52.7 | 54.0 | 63.5 | 67.9 | 71.4 | 28.8 | 28.1 | 36.6 | 30.6 | 34.2 |
| S2 | 74.5 | 72.1 | 72.0 | 74.7 | 79.2 | 48.9 | 42.5 | 41.2 | 49.2 | 52.5 |
| S3 | 54.3 | 58.6 | 64.6 | 60.1 | 63.0 | 30.8 | 32.7 | 33.0 | 34.8 | 34.1 |
| S4 | 83.1 | 85.4 | 87.3 | 87.7 | 88.7 | 64.1 | 67.7 | 70.1 | 71.3 | 72.8 |
| S5 | 76.5 | 74.6 | 77.8 | 80.0 | 83.0 | 49.2 | 47.3 | 48.8 | 52.2 | 55.9 |
| S6 | 61.3 | 57.0 | 68.0 | 65.5 | 69.6 | 34.2 | 34.4 | 35.7 | 41.2 | 39.5 |
| S7 | 83.2 | 84.4 | 84.4 | 86.6 | 87.9 | 65.4 | 67.2 | 64.5 | 68.0 | 71.2 |
| S8 | 59.4 | 58.6 | 56.3 | 60.5 | 65.5 | 24.5 | 32.7 | 28.0 | 25.7 | 26.8 |
| S9 | 71.7 | 73.1 | 78.4 | 78.0 | 79.9 | 43.5 | 46.3 | 51.1 | 49.7 | 52.5 |
| S10 | 78.3 | 79.5 | 87.1 | 87.7 | 89.0 | 60.2 | 61.9 | 72.2 | 71.5 | 75.5 |
| Mean ± Std | 69.5 ± 11.6 | 69.7 ± 11.8 | 73.9 ± 10.8 | 74.9 ± 10.9 | 77.7 ± 9.8 | 45.0 ± 15.1 | 46.1 ± 14.9 | 48.1 ± 16.0 | 49.4 ± 16.7 | 51.5 ± 17.6 |

TABLE 4 The total classification accuracy of 10 subjects for the ablation study (%).

| Subject | Total classification accuracy | | | |
|---|---|---|---|---|
| | CNN | TA-CNN | SA-CNN | STA-CNN |
| S1 | 25.8 | 30.5 | 29.9 | 31.7 |
| S2 | 55.1 | 58.7 | 56.3 | 61.7 |
| S3 | 24.0 | 26.4 | 24.6 | 28.1 |
| S4 | 50.3 | 55.7 | 53.3 | 56.9 |
| S5 | 46.1 | 53.9 | 47.9 | 55.7 |
| S6 | 32.9 | 30.5 | 24.6 | 29.9 |
| S7 | 52.4 | 59.0 | 57.5 | 59.9 |
| S8 | 22.8 | 25.2 | 24.6 | 27.0 |
| S9 | 35.9 | 43.7 | 40.1 | 44.3 |
| S10 | 53.9 | 55.1 | 59.3 | 56.3 |
| Mean ± Std | 39.9 ± 13.1 | 43.9 ± 14.2 | 41.8 ± 14.8 | 45.2 ± 14.5 |

degradation (Gao et al., 2014). In this paper, the SDMA method was utilized to present multiple motion-onset visual stimuli in the different locations of the visual field simultaneously, as shown in Figure 1, which can effectively improve the coding efficiency of spelling intentions. For example, to achieve the presentation of 40 characters, at least 14 times presentations are required using a determinant matrix. In contrast, this paper requires 8 presentations, including 3 parallel mVEP stimuli for determining group codes and 5 audio-assisted visual stimuli for determining character codes. The purpose is to improve the SNR of ERP components by utilizing the integration effect of audiovisual stimuli. Suppose the group and character codes are presented in parallel based on mVEP, a target character can be coded with a maximum of 6 times presentation, which can achieve a higher output speed.

On the other hand, according to the characteristics of mVEP and P300 from the audiovisual stimulus, extracting temporal and spatial information from single-trial EEG is the key to effectively decoding target ERP (Wirth et al., 2020). The traditional method is the grand average to improve the SNR of ERP. The second one is to extract ERP components from a single EEG according to the prior knowledge, such as wavelet transform, PCA, ICA and so on, but the computational complexity is high, and the result is not good. The others are using classification algorithms to identify targets and nontargets by mapping the original EEG to the separable space, such as the LDA method and its variants for optimizing key temporal segments and spatial activation positions of ERP. The development of deep learning (Li et al., 2018) has obvious advantages for decoding ERP, especially EEGNet has achieved good results. Based on EEGNet, this paper further introduces a spatial-temporal attention mechanism, which can effectively learn the key spatial-temporal features and make the deep learning method better interpretable. As seen in Figure 5,

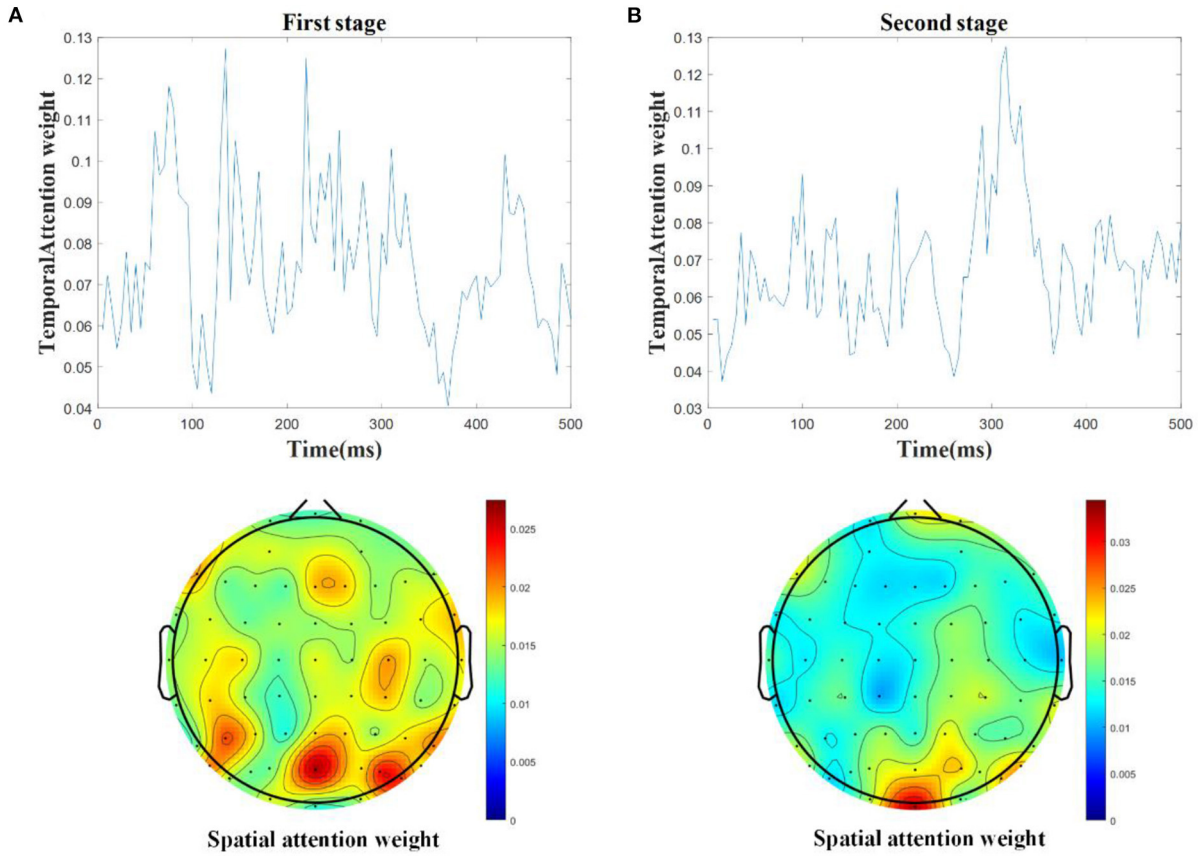



FIGURE 5
The average weights of temporal and spatial attention from the test samples based on STA-CNN using line charts and topographies in the first stage **(A)** and the second stage **(B)**.

the spatial-temporal attention mechanism can obtain larger weighted values in the time period corresponding to the active components of mVEP and P300, as well as in the spatial channels corresponding to the active brain areas of the target. Moreover, the deep learning method can realize end-to-end feature learning, thereby improving the adaptive ability between different subjects or trials.

The high reliability and robustness of audiovisual BCI should be furtherly considered for different subjects, different times, and different scenarios (Liu et al., 2020). According to the results in Table 4, the total classification accuracy of our paradigm is greatly affected by the first stage, which still needs to be improved. Due to the visual interference in the first stage, the classification accuracy of SDMA-based mVEP is not high. According to the literature (Lu et al., 2020), audiovisual integration could enhance the activation of attention-related brain areas. We tried to introduce the semantically congruent audio (pronunciation) to enhance the strength of the target ERPs in the second stage. The experimental results showed that the classification accuracy in the second stage was higher than in the first stage, which proves the audio-assisted effect's positive influence. But there are 5 characters that need to be traversed one by one in the second stage, which would lead to a decrease in presentation efficiency. To improve the efficiency of the BCI paradigm, we analyze further possible strategies, including novel paradigms to enhance the EEG features, such as the leftwards or rightwards motion-onset stimuli translating (Libert et al., 2022b) and the two-dimensional auditory stimuli with both pitch (high/medium/low) and direction (left/middle/right) (Hohne et al., 2011), and the stable classification algorithm of ERP for cross subjects or scenarios, such as the analytic beamformer transformation (Libert et al., 2022a), ternary classification method (Zhang et al., 2021) and some transfer learning methods.

This study proposed the spatial-temporal attention CNN method for decoding a novel audio-assisted mVEP-based BCI speller. A two-stage stimulation framework combined with mVEP and semantically congruent audio evoked P300 was designed based on a new SCDMA scheme to improve efficiency. Meanwhile, the STA-CNN method was proposed to deal with single-trial ERP components learning and classification.

Specifically, the spatial-temporal attention mechanism can enhance the discriminative event-related features by adaptively learning probability weights. The experiment results, obtained from a dataset including 10 subjects, showed that the classification accuracy and F1-score were significantly improved using the proposed STA-CNN compared with the LDA variant and deep learning methods. Moreover, through the analysis of the attention weights from time sequence and spatial topographies, it was proved that STA-CNN could effectively extract interpretable spatiotemporal features. It is possible to extend the proposed strategy in the mVEP-based BCI system in the online test scenario, and future studies are needed to avoid the mutual interference of different intentions in the SDMA scheme and develop a robust classification algorithm of ERP.

## 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 Taiyuan University of Technology. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

GC and XZ contributed to the conception and design of the study. GC organized the database, performed the analysis, and wrote the first draft of the manuscript. JZ, FL, and SD contributed to the manuscript revision. All authors participated to the scientific discussion. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 62201377, 62271342, 62171307, and 12004275), Research Project Supported by Shanxi Scholarship Council of China (Grant No. 2022-072), Scientific and Technological Innovation Project in Higher Education Institutions of Shanxi Province, China (Grant No. 2019L0189), and MOE of PRC Industry-University Collaborative Education Program (Grant No. 202002035019, Kingfar-CES Human Factors and Ergonomics).

## 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

Allison, Z. B., Jin, J., Zhang, Y., and Wang, X. (2014). A four-choice hybrid P300/SSVEP BCI for improved accuracy. *Brain Comput. Interfaces* 1, 17–26. doi: 10.1080/2326263X.2013.869003

Aloise, F., Arico, P., Schettini, F., Riccio, A., Salinari, S., Mattia, D., et al. (2012). A covert attention P300-based brain-computer interface: geospell. *Ergonomics* 55, 538–551. doi: 10.1080/00140139.2012.661084

Amin, S. U., Alsulaiman, M., Muhammad, G., Mekhtiche, M. A., and Hossain, M. S. (2019). Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. *Future Generation Comput. Syst. Int. J. Escience* 101, 542–554. doi: 10.1016/j.future.2019.06.027

Belitski, A., Farquhar, J., and Desain, P. (2011). P300 audio-visual speller. *J. Neural Eng.* 8, 025022. doi: 10.1088/1741-2560/8/2/025022

Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Muller, K. R. (2011). Single-trial analysis and classification of ERP components—a tutorial. *Neuroimage* 56, 814–825. doi: 10.1016/j.neuroimage.2010.06.048

Blankertz, B., Sannelli, C., Haider, S., Hammer, E. M., Kubler, A., Muller, K. R., et al. (2010). Neurophysiological predictor of SMR-based BCI performance. *Neuroimage* 51, 1303–1309. doi: 10.1016/j.neuroimage.2010.03.022

Cecotti, H., Eckstein, M. P., and Giesbrecht, B. (2014). Single-trial classification of event-related potentials in rapid serial visual presentation tasks using supervised spatial filtering. *Ieee Trans. Neural Networks Learn. Syst.* 25, 2030–2042. doi: 10.1109/TNNLS.2014.2302898

Cecotti, H., and Ries, A. J. (2017). Best practice for single-trial detection of event-related potentials: application to brain-computer interfaces. *Int. J. Psychophysiol.* 111, 156–169. doi: 10.1016/j.ijpsycho.2016.07.500

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

Gao, S. K., Wang, Y. J., Gao, X. R., and Hong, B. (2014). Visual and auditory brain-computer interfaces. *IEEE Trans. Biomed. Eng.* 61, 1436–1447. doi: 10.1109/TBME.2014.2300164

Guo, F., Hong, B., Gao, X. R., and Gao, S. K. (2008). A brain-computer interface using motion-onset visual evoked potential. *J. Neural Eng.* 5, 477–485. doi: 10.1088/1741-2560/5/4/011

Hessler, D., Jonkers, R., Stowe, L., and Bastiaanse, R. (2013). The whole is more than the sum of its parts—audiovisual processing of phonemes

investigated with ERPs. *Brain Lang.* 124, 213–224. doi: 10.1016/j.bandl.2012.12.006

Hohne, J., Schreuder, M., Blankertz, B., and Tangermann, M. (2011). A novel 9-class auditory ERP paradigm driving a predictive text entry system. *Front. Neurosci.* 5, 1–10. doi: 10.3389/fnins.2011.00099

Hong, B., Guo, F., Liu, T., Gao, X. R., and Gao, S. K. (2009). N200-speller using motion-onset visual response. *Clin. Neurophysiol.* 120, 1658–1666. doi: 10.1016/j.clinph.2009.06.026

Ioffe, S., and Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. *Int. Conf. Machine Learn.* 37, 448–456. doi: 10.5555/3045118.3045167

Kuba, M., Kubova, Z., Kremlacek, J., and Langrova, J. (2007). Motion-onset VEPs: Characteristics, methods, and diagnostic use. *Vision Res.* 47, 189–202. doi: 10.1016/j.visres.2006.09.020

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, W. L., Tan, T., Falkmer, T., and Leung, Y. H. (2016). Single-trial event-related potential extraction through one-unit ICA-with-reference. *J. Neural Eng.* 13, 066010. doi: 10.1088/1741-2560/13/6/066010

Li, J. C., Yu, Z. L., Gu, Z. H., Wu, W., Li, Y. Q., and Jin, L. W. (2018). A hybrid network for ERP detection and analysis based on restricted boltzmann machine. *IEEE Trans. Neural Syst. Rehabilitat. Eng.* 26, 563–572. doi: 10.1109/TNSRE.2018.2803066

Libert, A., van den Kerchove, A., Wittevrongel, B., and Van Hulle, M. M. (2022a). Analytic beamformer transformation for transfer learning in motion-onset visual evoked potential decoding. *J. Neural Eng.* 19, 026040. doi: 10.1088/1741-2552/ac636a

Libert, A., Wittevrongel, B., Camarrone, F., and Van Hulle, M. M. (2022b). Phase-spatial beamforming renders a visual brain computer interface capable of exploiting EEG electrode phase shifts in motion-onset target responses. *IEEE Trans. Biomed. Eng.* 69, 1802–1812. doi: 10.1109/TBME.2021.3136938

Liu, D. K., Liu, C., Chen, J. J., Zhang, D., and Hong, B. (2021). Doubling the speed of N200 speller via dual-directional motion encoding. *IEEE Trans. Biomed. Eng.* 68, 204–213. doi: 10.1109/TBME.2020.3005518

Liu, D. K., Liu, C., and Hong, B. (2019). "Bi-directional visual motion based BCI speller," in *2019 9th International Ieee/Embs Conference on Neural Engineering (Ner)*, p. 589–592. doi: 10.1109/NER.2019.8717075

Liu, S., Wang, W., Sheng, Y., Zhang, L. D., Xu, M. P., and Ming, D. (2020). Improving the cross-subject performance of the ERP-based brain-computer interface using rapid serial visual presentation and correlation analysis rank. *Front. Hum. Neurosci.* 14, 296. doi: 10.3389/fnhum.2020.00296

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., et al. (2018). A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update. *J. Neural Eng.* 15, 031005. doi: 10.1088/1741-2552/aab2f2

Lu, Z. H., Li, Q., Gao, N., and Yang, J. J. (2020). Time-varying networks of ERPs in P300-speller paradigms based on spatially and semantically congruent audiovisual bimodality. *J. Neural Eng.* 17, 046015. doi: 10.1088/1741-2552/aba07f

Lu, Z. H., Li, Q., Gao, N., Yang, J. J., and Bai, O. (2019). A novel audiovisual p300-speller paradigm based on cross-modal spatial and semantic congruence. *Front. Neurosci.* 13, 1040. doi: 10.3389/fnins.2019.01040

Molla, M. K. I., Morikawa, N., Islam, M. R., and Tanaka, T. (2018). Data-adaptive spatiotemporal ERP cleaning for single-trial BCI implementation. *IEEE Trans. Neural Syst. Rehabilitat. Eng.* 26, 1334–1344. doi: 10.1109/TNSRE.2018.2844109

Mowla, M. R., Ng, S. C., Zilany, M. S. A., and Paramesran, R. (2016). Single-trial evoked potential estimation using iterative principal component analysis. *IEEE Sens. J.* 16, 6955–6960. doi: 10.1109/JSEN.2016.2591582

Quiroga, R. Q. (2005). Single-trial event-related potentials with wavelet denoising: method and applications. *Unveiling Mystery Brain Neurophysiol Invest Brain Funct.* 1278, 429–432. doi: 10.1016/j.ics.2004.11.062

Schaeff, S., Treder, M. S., Venthur, B., and Blankertz, B. (2012). Exploring motion VEPs for gaze-independent communication. *J. Neural Eng.* 9, 045006. doi: 10.1088/1741-2560/9/4/045006

Simon, D. M., and Wallace, M. T. (2018). Integration and temporal processing of asynchronous audiovisual speech. *J. Cogn. Neurosci.* 30, 319–337. doi: 10.1162/jocn_a_01205

Song, Y. H., Cai, S. Q., Yang, L., Li, G. F., Wu, W. F., and Xie, L. H. (2020). A practical EEG-based human-machine interface to online control an upper-limb assist robot. *Front. Neurorobot.* 14, 32. doi: 10.3389/fnbot.2020.00032

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958. doi: 10.5555/2627435.2670313

Treder, M. S., and Blankertz, B. (2010). (C)overt attention and visual speller design in an ERP-based brain-computer interface. *Behav. Brain Funct.* 6, 28. doi: 10.1186/1744-9081-6-28

Vareka, L. (2020). Evaluation of convolutional neural networks using a large multi-subject P300 dataset. *Biomed. Signal Process. Control* 58, 1–7. doi: 10.1016/j.bspc.2019.101837

Wang, F., He, Y. B., Pan, J. H., Xie, Q. Y., Yu, R. H., Zhang, R., et al. (2015). A novel audiovisual brain-computer interface and its application in awareness detection. *Sci. Rep.* 5, 9962. doi: 10.1038/srep09962

Wirth, C., Dockree, P. M., Harty, S., Lacey, E., and Arvaneh, M. (2020). Towards error categorisation in BCI: single-trial EEG classification between different errors. *J. Neural Eng.* 17, 016008. doi: 10.1088/1741-2552/ab53fe

Woo, S. H., Park, J., Lee, J. Y., and Kweon, I. S. (2018). CBAM: convolutional block attention module. *Comput. Vis. Eccv 2018 Pt Vii.* 11211, 3–19. doi: 10.1007/978-3-030-01234-2_1

Wu, L., Shen, C. H., and van den Hengel, A. (2017). Deep linear discriminant analysis on fisher networks: A hybrid architecture for person re-identification. *Pattern Recognit.* 65, 238–250. doi: 10.1016/j.patcog.2016.12.022

Xiao, X. L., Xu, M. P., Jin, J., Wang, Y. J., Jung, T. P., and Ming, D. (2020). Discriminative canonical pattern matching for single-trial classification of ERP components. *IEEE Trans. Biomed. Eng.* 67, 2266–2275. doi: 10.1109/TBME.2019.2958641

Xu, M. P., Han, J., Wang, Y. J., 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. P., He, F., Jung, P., Gu, X. S., 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. P., Xiao, X. L., Wang, Y. J., Qi, H. Z., 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

Zang, B. Y., Lin, Y. F., Liu, Z. W., and Gao, X. R. (2021). A deep learning method for single-trial EEG classification in RSVP task based on spatiotemporal features of ERPs. *J. Neural Eng.* 18, 0460c0468. doi: 10.1088/1741-2552/ac1610

Zhang, C. C., Qiu, S., Wang, S. P., and He, H. G. (2021). Target detection using ternary classification during a rapid serial visual presentation task using magnetoencephalography data. *Front. Comput. Neurosci.* 15, 619508. doi: 10.3389/fncom.2021.619508

Zhang, R., Xu, P., Chen, R., Ma, T., Lv, X. L., Li, F. L., et al. (2015). An adaptive motion-onset VEP-based brain-computer interface. *IEEE Trans. Auton. Ment. Dev.* 7, 349–356. doi: 10.1109/TAMD.2015.2426176

Zhang, Y., Zhou, G. X., Zhao, Q. B., Jin, J., Wang, X. Y., and Cichocki, A. (2013). Spatial-temporal discriminant analysis for erp-based brain-computer interface. *IEEE Trans. Neural Syst. Rehabilitat. Eng.* 21, 233–243. doi: 10.1109/TNSRE.2013.2243471

frontiers | Frontiers in Neuroinformatics

TYPE Original Research
PUBLISHED 05 October 2022
DOI 10.3389/fninf.2022.952474



OPEN ACCESS

EDITED BY
Jiayang Guo,
Xiamen University, China

REVIEWED BY
Jing Jin,
East China University of Science and Technology, China
Jipeng Wu,
Shenzhen Institute of Advanced Technology (CAS), China

*CORRESPONDENCE
Moran Cerf
moran@morancerf.com

RECEIVED 25 May 2022
ACCEPTED 24 August 2022
PUBLISHED 05 October 2022

CITATION
Wang G and Cerf M (2022) Brain-Computer Interface using neural network and temporal-spectral features.
*Front. Neuroinform.* 16:952474.
doi: 10.3389/fninf.2022.952474



# Brain-Computer Interface using neural network and temporal-spectral features

Gan Wang[1] and Moran Cerf[2]*

[1]School of Mechanical and Electrical Engineering, Soochow University, Suchow, China,
[2]Interdepartmental Neuroscience Program, Northwestern University, Evanston, IL, United States

Brain-Computer Interfaces (BCIs) are increasingly useful for control. Such BCIs can be used to assist individuals who lost mobility or control over their limbs, for recreational purposes such as gaming or semi-autonomous driving, or as an interface toward man-machine integration. Thus far, the performance of algorithms used for thought decoding has been limited. We show that by extracting temporal and spectral features from electroencephalography (EEG) signals and, following, using deep learning neural network to classify those features, one can significantly improve the performance of BCIs in predicting which motor action was imagined by a subject. Our movement prediction algorithm uses Sequential Backward Selection technique to jointly choose temporal and spectral features and a radial basis function neural network for the classification. The method shows an average performance increase of 3.50% compared to state-of-the-art benchmark algorithms. Using two popular public datasets our algorithm reaches 90.08% accuracy (compared to an average benchmark of 79.99%) on the first dataset and 88.74% (average benchmark: 82.01%) on the second dataset. Given the high variability within- and across-subjects in EEG-based action decoding, we suggest that using features from multiple modalities along with neural network classification protocol is likely to increase the performance of BCIs across various tasks.

KEYWORDS

**Brain-Computer Interfaces, motor, EEG, neural networks, deep learning**

## Introduction

Brain-Computer Interfaces (BCIs) act as a link between neural activity and machine operations. The BCI extracts data from electrodes or sensors acquiring neural signals and translates those data into digital code (Bulárka and Gontean, 2016). Applications of BCI include those focused on improved health outcomes (i.e., rehabilitation of impaired motor function; Courtine et al., 2013), restoration of sensory functions (Hochberg et al., 2012), interpreting thoughts from individuals who cannot otherwise communicate them

(Cerf et al., 2010), enhanced control of devices (i.e., operating heavy machinery, flying drones, or driving; Chiuzbaian et al., 2019, Nader et al., 2021), or recreational uses (i.e., gaming; Cerf and Garcia-Garcia, 2017). Invasive BCIs, such as ones built on single-neuron recordings, have recently shown high accuracy in interpreting human/animal intentions, actions, and imagery (Cerf et al., 2010; Hochberg et al., 2012). Non-invasive tools such as ones using electroencephalography (EEG) data have demonstrated high performance in interpreting thoughts and actions. For example, interpreting imagined motor action–a commonly used task for evaluating BCIs–has shown decoding accuracies ranging between 70 and 85% in recent works (Gordleeva et al., 2017).

Notably, BCIs based on motor imagination (MI) tasks, where subjects imagine an action (i.e., clenching of the fist) and the BCI aims to identify the action imagined, have shown remarkable improvement in recent years. In a typical MI task, the BCI derives neural signatures (i.e., power changes in the alpha and beta rhythms extracted from sensory-motor regions) that accurately predict the action intent following a training period. Given that non-invasive signals generated by EEG are often contaminated by artifacts derived from eye movement or muscle movement, a typical EEG-based BCI requires larger training data and isolated trials to increase the action identification accuracy. The repeated trials enable the averaging of the event-related signals and the extraction of a synchronized clean input. Variance across individual subjects, electrode montages, experimental sessions, and trial types add difficulty to the interpretation of the signals.

Given the challenges in EEG-based BCI development using noisy inputs, numerous methods have been proposed to improve the decoders performance (Lebedev and Nicolelis, 2006; Lotte et al., 2007; Prashant et al., 2015; Abiri et al., 2019; Andersen et al., 2019). The suggested methods often focus on the isolation of temporal or spectral components in the signal. Algorithms based on spectral feature selection are more prominent in the BCI arsenal since the time courses of event-related synchronization (or de-synchronization) vary heavily among subjects during motor tasks (Hochberg et al., 2012; Andersen et al., 2019).

Within the feature selection BCIs signal toolkit arsenal, common spatial patterns (CSPs) algorithms are dominant (Bhatti et al., 2019). These algorithms seek to find an optimal spatial filter that distinguishes one brain state from another. In EEG, the performance of CSPs is highly sensitive to the choice of frequency bands, making the decision on which filter to use heavily dependent on the recording configuration. To afford some generalization, variants of CSP were proposed as ways to improve the signal processing. Those variants often use narrower frequency bands (termed: sub-band CSP; SBCSP; Novi et al., 2007) and Filter Banks (FBCSP; Ang et al., 2008) and show increased performance for action decoding, yet are still scarce.

In addition to the extended frequency bands and filters improvements, recent attempts to include temporal signals in BCIs emerged in the form of Temporally Constrained Group Spatial Pattern (TSGSP) algorithms (Zhang et al., 2018). TSGSP optimally select the CSP features by considering different temporal windows for signal extraction derived from multi-task learnings. That is, instead of collapsing all the trials within one MI class (i.e., all left-hand movement trials) various MI tasks are combined to suggest the ideal CSP for a specific individual subject. The TSGSP algorithms use Support Vector Machines (SVM) for the classification of new trials to their corresponding action class. This inclusion of temporal data was recently shown to improve the performance of CSP-based BCIs (Sakhavi et al., 2018; Zhang et al., 2018; Deng et al., 2021).

Neural network based classifiers that frequently show superiority in data-rich non-linear clustering tasks such as MI decoding were recently suggested as a potential improvement for the CSP algorithms (Bhatti et al., 2019). Specifically, the usage of Sequential Backward Floating Selection method along with a radial basis function neural network (RBFNN) for optimal CSP features selection was suggested as a potential superior algorithm for BCIs (Bhatti et al., 2019).

Here we implement and test a combination of the suggested improvements for MI decoding and show the tuning curves of key parameters driving the performance increase. Namely, we introduce a number of additions to the BCI motor classification algorithms arsenal. First, we incorporate both temporal and spectral features in the MI BCI. Second, we use sub-bands rather than typical frequency bands for the BCI inputs. Third, we combine the successful Sequential Backward Selection (SBS) method with CSP features for the temporal-spectral feature selection. Fourth, we separate the feature selection process from the following feature classification process. Finally, we incorporate the suggested RBFNN (rather than SVM) in the motor classification. We demonstrate the effectiveness of our method using popular public datasets and compare our performance to the current state-of-the-art BCI benchmark algorithms.

This work contributes to the BCI literature by showing that the combination of a SBS and temporal-spectral EEG signals with RBFNN significantly outperforms other methods. This is the first work to test the combination of all previously suggested improvements to existing algorithms in a single implementation (see Sakhavi et al., 2018; Zhang et al., 2018; Deng et al., 2021; for discussions of the improvements implemented here).

# Materials and methods

## Data

Two popular BCI datasets were used for the algorithm testing:

### Dataset 1

Brain-computer interface competition IV, dataset 2a, which contains 22-channel EEG data recorded from 9 healthy subjects (A01–A09) participating in different MI tasks. In each task, subjects were asked to imagine movement of the left hand, right hand, feet, and tongue. The experiment consisted of two sessions. In each session, there were 72 trials for each of the four classes of movement. The EEG signals were sampled at 250 Hz and bandpass filtered between 0.5 and 100 Hz with a 50 Hz notch filter. We used the data from the left- and right-hand imagery tasks alone to align with the second dataset and some of the benchmark algorithms that focused solely on those movement classes.

### Dataset 2

Brain-computer interface competition IV, dataset 2b, which contains 3-channel EEG data recorded from 9 different subjects (B01–B09) participating in two MI tasks. The experimental protocol was nearly identical to dataset 1 other than the fact that subjects only imagined movements of the left-hand and right-hand, and that instead of two sessions there were five session. For each subject, separate training and testing sets were available. The EEG signals were sampled at 250 Hz and bandpass filtered between 0.5 and 100 Hz with a 50 Hz notch filter.

See Leeb et al. (2008) for additional details on the two datasets.

## Feature extraction

### Pre-processing

Raw EEG signals were filtered between 4 and 40 Hz with fifth-order Butterworth filter. For each trial, we used samples between 500 and 4,500 ms from the trial onset in the analyses. The first 500 ms were excluded, in alignment with the instructions of the BCI IV competition winners, because of response times deviations across trials.

### Feature selection

The neural signals were divided to five overlapping 2-s windows with a step size of 500 ms. This ensured temporal generalizability within a trial. Following, the data were filtered along 17 overlapping frequency bands ranging from 4 to 40 Hz with a 2 Hz step. Finally, a common spatial filter (Bhatti et al., 2019) was identified such that it maximized the variance within a single class (i.e., across all left-hand trials) and minimized the variance across classes (i.e., between left-hand and right-hand trials).

The data for a single trial were represented as a matrix, $X \in R^{N \cdot T}$ (with $N$ reflecting the number of channels, and $T$ the time) whose normalized covariance matrix, $C$, is:

$$C = \frac{XX^T}{trace(XX^T)} \tag{1}$$

Averaging across all trials within a class yielded a matrix, $C_t$ ($t$ indicating the class type).

The spatial covariance was calculated by averaging all covariance matrices:

$$C_c = \overline{C_{left-hand}} + \overline{C_{right-hand}} \tag{2}$$

The $C_c$ matrix was white transformed:

$$C_C = U_C \lambda_C U_C^T \tag{3}$$

with $U_C$ the eigenvector matrix and $\lambda_C$ the eigenvectors.

Defining $P$ as:

$$P = \sqrt{\lambda_C^{-1}} U_C^T \tag{4}$$

the individual class matrices were transformed to:

$$S_{left-hand} = P\overline{C_{left-hand}}P^T \tag{5}$$

$$S_{right-hand} = P\overline{C_{right-hand}}P^T \tag{6}$$

such that the $S_t$ matrices have the same eigenvectors.

Given that $S_t$ could be represented as $B\lambda_t B^T$ with $B$ the eigenvectors matrix and $\lambda_t$ the eigenvalues:

$$S_t = B\lambda_t B^T \tag{7}$$

the projection matrix, $W$, was derived:

$$W = B^T P \tag{8}$$

Thus, the EEG data were projected to a matrix, $Z$:

$$Z = W^T X \tag{9}$$

where the columns of $Z$ corresponded to the data's spatial source distribution vectors. The vectors maximized the variance across classes and corresponded to the maximum eigenvalues ($\lambda_{left-hand}$ and $\lambda_{right-hand}$ ).

Finally, the classification features were represented by:

$$f_p = \log\left(\frac{var(Z_p)}{\sum_{i=1}^{n} var(Z_i)}\right) \tag{10}$$

where $Z_p$ are the CSPs ($p = 1..N$).

A subset of $Z$ (first and last $m$ rows) were used in further analyses.

An SBS (Pasyuk et al., 2019) was used to reduce the initial 85-feature set (17 frequency bands × 5 time-windows) from each individual trial. According to the SBS criteria, in every iteration of the algorithm the feature yielding the lowest accuracy was discarded. That is, if the initial performance with all 85 features was, say, 87%, the performance using 84 features was computed next, leaving one feature out in each iteration ($f_1$ = 78%, $f_2$ = 82%, $f_3$ = 77%, ...). Comparing all 85 leave-one-out iterations, the feature whose contribution to the performance was lowest (i.e., one without whom the performance drops least; $f_2$ in the particular example) was

discarded. Following, the performance of the remaining 84 features was set as the anchor performance and the evaluation was repeated with 83 features. Each run led to a drop of a single feature. The optimal performance across all 3,655 iterations (85+84+...) was regarded the network's accuracy, with the feature set yielding the highest performance being the preferred set.

## Neural network

An RBFNN was used for the classification. The network consisted of two layers: an input layer and a hidden layer. The output of the hidden layer was summed proportionally to the input features to yield the output classification. Formally, this is represented as:

$$F(x) = \sum_{i=1}^{k} w_i f_i(\boldsymbol{x}, \boldsymbol{c}_i) + b \tag{11}$$

where $w_i$ are the weights, $f_i$ the Gaussian radial basis functions, $c_i$ the center values of the Gaussian radial function, $b$ the bias, and $k$ the number of neurons in the hidden layer.

With $f_i$ formally computed as:

$$f_i(\boldsymbol{x}, \boldsymbol{c}_i) = e^{\left(\frac{-||\boldsymbol{x}-\boldsymbol{c}_i||^2}{2\sigma_i^2}\right)} \tag{12}$$

where $\sigma_i$ is the standard deviation.

In each iteration of the RBFNN implementation the extracted input features are scaled and used to train the network, followed by a testing. The network was implemented using Matlab's *newrbe* function default hyperparameters, with the spread of the radial basis functions set to 16.

## Implementation

The implementation of the method–pre-processing, feature selection, and neural network classification are available online at https://www.morancerf.com/publications.

### Analyses

We compared our algorithm's performance to that of all state-of-the-art methods which: (a) were published in the last 5 years, (b) used the same datasets as ours, and (c) were implemented on both the left- and right-hand MI data. We used one implementation of each method to avoid focusing on coding variations but rather on conceptual differences in the protocol. Altogether, 38 methods were compared to our algorithms, and 19 were not included in our analyses because they did not satisfy the inclusion criteria (namely, those algorithms used different movement classes outside of the ones we tested).

For dataset 1 we compared our performance to the following methods (see results in **Table 1**):

(1) Deep Neural Network (DNN) (Kumar et al., 2016)
(2) Kernel Principal Component Analysis using Conformal-Isometric Linearizing Kernel (KPCA-CILK) (Sadatnejad and Ghidary, 2016)
(3) Weighted Overlap Add Common Spatial Patterns (WOLA-CSP) (Belwafi et al., 2018)

TABLE 1 Performance comparison for dataset 1, sorted by accuracy.

| Method | Year | A01 | A02 | A03 | A04 | A05 | A06 | A07 | A08 | A09 | Mean ± std |
|---|---|---|---|---|---|---|---|---|---|---|---|
| SBS-FBCSP | 2022 | 80.00 | 72.76 | 83.79 | 70.42 | 73.10 | 68.97 | 75.17 | 77.93 | 77.59 | 75.52 ± 4.76 |
| DNN | 2016 | 86.81 | 66.70 | 95.83 | 76.39 | 57.64 | 68.06 | 75.00 | 93.75 | 77.08 | 77.47 ± 12.72 |
| KPCA CILK | 2016 | 88.89 | 59.03 | 90.28 | 78.47 | 62.50 | 75.00 | 72.92 | 93.06 | 87.50 | 78.63 ± 12.34 |
| WOLA-CSP | 2018 | 86.81 | 63.19 | 94.44 | 68.75 | 56.25 | 69.44 | 78.47 | 97.91 | 93.75 | 78.78 ± 15.15 |
| MEMDBF-CSP-LDA | 2019 | 90.78 | 57.75 | 97.08 | 70.69 | 61.48 | 70.37 | 72.14 | 97.76 | 94.62 | 79.19 ± 15.85 |
| JSTFD-LDA | 2020 | 86.40 | 55.90 | 96.30 | 73.10 | 89.50 | 58.20 | 76.10 | 93.80 | 86.60 | 79.54 ± 14.78 |
| nCSP-TSLR | 2019 | 89.23 | 76.15 | 90.60 | 71.38 | 59.82 | 63.26 | 91.70 | 89.18 | 85.26 | 79.62 ± 12.36 |
| W-CNN | 2019 | 76.67 | 72.00 | 90.00 | 73.33 | 83.33 | 80.00 | 82.67 | 80.00 | 80.00 | 79.78 ± 5.45 |
| SS-MEMDBF | 2018 | 91.49 | 60.56 | 94.16 | 76.16 | 58.52 | 68.52 | 78.57 | 97.01 | 93.85 | 79.87 ± 15.01 |
| CSP-Wavelet + LOG | 2020 | 93.06 | 61.81 | 95.83 | 72.92 | 58.33 | 68.06 | 81.25 | 95.14 | 93.06 | 79.94 ± 15.06 |
| SW-LSR | 2021 | 86.81 | 64.58 | 95.83 | 67.36 | 68.06 | 67.36 | 80.56 | 97.22 | 92.36 | 80.02 ± 13.45 |
| EEGnet | 2016 | 71.43 | 78.51 | 100 | 64.28 | 71.43 | 78.57 | 71.43 | 92.86 | 100 | 80.95 ± 13.37 |
| R-MDRM | 2019 | 91.61 | 63.28 | 97.20 | 72.91 | 64.08 | 69.71 | 81.25 | 96.52 | 92.30 | 80.98 ± 13.86 |
| SR-MDRM | 2019 | 90.21 | 63.28 | 96.55 | 76.38 | 65.49 | 69.01 | 81.94 | 95.14 | 93.01 | 81.22 ± 13.19 |
| TSGSP | 2018 | 87.00 | 64.70 | 93.80 | 74.30 | 90.40 | 63.90 | 91.40 | 95.80 | 81.30 | 82.51 ± 12.24 |
| DCR-MEMD | 2021 | 89.79 | 94.18 | 78.92 | 94.01 | 71.32 | 86.71 | 89.36 | 82.11 | 86.18 | 85.84 ± 7.40 |
| *Ours* | *2022* | *93.45* | *84.83* | *95.52* | *88.33* | *86.55* | *83.10* | *88.97* | *95.52* | *94.48* | *90.08* ± 4.78 |

(4) Multivariate Empirical Mode Decomposition Based Filtering-Common Spatial Pattern-Linear Discriminant Analysis (MEMDBF-CSP-LDA) (Gaur et al., 2019)
(5) Joint Spatio-temporal Filter Design Linear Discriminant Analysis (JSTFD-LDA) (Jiang et al., 2020)
(6) Normalized Common Spatial Pattern Tangent Space Logistic Regression (nCSP-TSLR) (Olias et al., 2019)
(7) Wavelet Convolutional Neural Network (W-CNN) (Xu et al., 2018)
(8) Subject Specific Multivariate Empirical Mode Decomposition Based Filtering (SS-MEMDBF) (Gaur et al., 2018)
(9) Common Spatial Pattern-Filter Bank-Log (CSP-FB-LOG) (Zhang S. et al., 2020)
(10) Sliding Window-Longest Consecutive Repetition (SW-LSR) (Gaur et al., 2021)
(11) EEG Network (EEGnet) (Lawhern et al., 2018)
(12) Regularized Minimum Distance to Riemannian Mean (R-MDRM) (Singh et al., 2019)
(13) Spatial Regularized Minimum Distance to Riemannian Mean (SR-MDRM) (Singh et al., 2019)
(14) Temporally Constrained Sparse Group Spatial Patterns (TSGSP) (Zhang et al., 2018)
(15) Dynamic Channel Relevance-Multivariate Empirical Mode Decomposition (DCR-MEMD) (Song and Sepulveda, 2018)

For dataset 2 we compared our results to the following methods (see results in **Table 2**):

(1) Robust Support Matrix Machine (RSMM) (Zheng et al., 2018)
(2) Deep Learning with Variational Autoencoder (DLVA) (Dai et al., 2019)
(3) Sparse Group Representation Model (SGRM) (Jiao et al., 2018)
(4) Unsupervised Discriminative Feature Selection (UDFS) (Al Shiam et al., 2019)
(5) Sparse Spectro-temporal Decomposition Squeeze-and-Excitation Convolutional Neural Network (SSD-SE-CNN) (Sun et al., 2020)
(6) Wavelet Spatial Filter Convolution Network (WaSF ConvNet) (Dy et al., 2019; Fang et al., 2022)
(7) Neighborhood Component analysis based Feature Selection (NCFS) (Molla et al., 2020)
(8) Common Spatial Pattern-Wavelet-Log (CSP-Wavelet-LOG) (Zhang S. et al., 2020)
(9) Multi-Attention Adaptation Network (MAAN) (Chen et al., 2021)
(10) Multilayer Temporal Pyramid Pooling EEG Network (MTPP-EEGNet) (Ha and Jeong, 2020)
(11) Dynamic Joint Domain Adaptation (DJDA) (Hong et al., 2021)
(12) SincNet-based Hybrid Neural Network (SHNN) (Liu et al., 2022)
(13) Tangent Space Linear Discriminant Analysis (TSLDA) (Ai et al., 2019; Fang et al., 2022)
(14) Deep Representation-based Domain Adaptation (DRDA) (Zhao et al., 2020)
(15) Random Forest Dynamic Frequency Feature Selection (RF-DFFS) (Luo et al., 2016)
(16) Frequential Deep Belief Network (FDBN) (Lu et al., 2016)
(17) Temporally constrained Sparse Group Spatial Patterns (TSGSP) (Zhang et al., 2018)
(18) Multi-branch Multi-scale Convolutional Neural Network (MMCNN) (Jia et al., 2020)
(19) Wavelet Package Decomposition Spatio-Temporal Discrepancy Feature (WPD-STDF) (Luo et al., 2019)
(20) Central Distance Loss Convolutional Neural Network (CD-CNN) (Yang et al., 2021)
(21) Filter Banks and Riemannian Tangent Space (FBRTS) (Fang et al., 2022)

Additionally, we implemented a version of the Sequential Backward Selection Filter Bank Common Spatial Patterns (SBS-FBCSP) algorithm, which is an adaptation of the Sub-Band Common Spatial Patterns with Sequential Backward Floating Selection (SBCSP-SBFS) proposed by Bhatti et al. (2019). The original SBCSP-SBFS algorithm did not use temporal features and was limited to 12 overlapping frequency bands (4–30 Hz). Conceptually, the SBS-FBCSP algorithm resembled our method in that it, too, used sub-bands and CSPs feature selection. SBS-FBCSP differed from our method in that it used the full trial as temporal dimension.

For dataset 1, we varied the parameter $m$ from 1 to 7 (with $2m$ options yielding up to 14 features in each trial) since the parameter selection impacts the performance. To calculate the accuracy, we used 5-fold cross validation with all the trials from the first dataset (combining the first and second sessions onto one data set).

For dataset 2 we varied $m$ from 1 to 3 yielding up to six features. To calculate the accuracy, we combined all sessions data in random order and used 80% of the trials for training and the remaining 20% for testing sessions as training set and the remaining two for testing, with five-fold cross validation (see Luo et al., 2016).

The excluded methods were:

(1) Distance Preservation to Local Mean (Davoudi et al., 2017)

(2) Neighborhood Rough Set Classifier (Udhaya Kumar and Hannah Inbarani, 2017)

(3) Channel-wise Convolution with Channel Mixing (Sakhavi et al., 2018)

TABLE 2 Performance comparison for dataset 1, sorted by accuracy.

| Method | Year | B01 | B02 | B03 | B04 | B05 | B06 | B07 | B08 | B09 | Mean ± std |
|---|---|---|---|---|---|---|---|---|---|---|---|
| SBS-FBCSP | 2022 | 70.14 | 60.29 | 62.50 | 89.19 | 82.43 | 73.61 | 66.67 | 74.34 | 79.86 | 73.23 ± 9.49 |
| RSMM | 2016 | 72.50 | 56.43 | 55.63 | 97.19 | 88.44 | 78.75 | 77.50 | 91.88 | 83.44 | 77.97 ± 14.56 |
| DLAV | 2019 | 76.10 | 67.30 | 71.80 | 95.40 | 82.30 | 82.10 | 77.50 | 75.30 | 75.90 | 78.19 ± 7.95 |
| SGRM | 2019 | 76.30 | 56.00 | 49.20 | 98.20 | 91.10 | 74.80 | 88.30 | 85.40 | 84.90 | 78.24 ± 16.26 |
| UDFS | 2019 | 76.09 | 58.64 | 53.45 | 99.38 | 83.83 | 76.96 | 83.15 | 90.66 | 83.48 | 78.40 ± 14.53 |
| SSD-SE-CNN | 2021 | 78.50 | 67.90 | 68.30 | 96.50 | 81.40 | 85.70 | 76.90 | 79.30 | 79.60 | 79.34 ± 8.65 |
| WaSF ConvNet | 2019 | 73.80 | 64.20 | 85.70 | 96.20 | 85.20 | 68.50 | 88.30 | 90.10 | 81.50 | 81.50 ± 10.58 |
| NCFS | 2020 | 79.25 | 63.48 | 56.65 | 99.28 | 88.67 | 79.96 | 88.76 | 92.66 | 84.95 | 81.52 ± 13.72 |
| CSP-FB-LOG | 2020 | 88.75 | 52.50 | 48.75 | 98.75 | 88.75 | 90.00 | 90.00 | 92.50 | 83.75 | 81.53 ± 17.98 |
| MAAN | 2021 | 82.81 | 60.36 | 59.06 | 97.50 | 91.88 | 86.38 | 84.06 | 93.44 | 86.88 | 82.49 ± 13.73 |
| MTPP-EEGNet | 2020 | 78.75 | 66.43 | 67.50 | 95.00 | 94.38 | 84.38 | 85.31 | 92.19 | 81.56 | 82.83 ± 10.61 |
| DJDA | 2021 | 83.44 | 58.57 | 59.06 | 98.13 | 96.56 | 84.38 | 86.25 | 92.81 | 87.81 | 83.00 ± 14.64 |
| SHNN | 2022 | 83.33 | 61.76 | 58.33 | 97.30 | 91.89 | 88.89 | 86.11 | 92.11 | 91.67 | 83.49 ± 13.89 |
| TSLDA | 2019 | 76.30 | 68.90 | 86.40 | 94.20 | 88.10 | 72.30 | 89.20 | 92.80 | 87.30 | 83.94 ± 9.13 |
| DRDA | 2021 | 81.37 | 62.86 | 63.63 | 95.94 | 93.56 | 88.19 | 85.00 | 95.25 | 90.00 | 83.98 ± 12.67 |
| RF-DFFS | 2016 | 73.24 | 67.48 | 63.01 | 97.40 | 95.49 | 86.66 | 84.68 | 95.93 | 92.61 | 84.06 ± 13.06 |
| FDBN | 2016 | 81.00 | 65.00 | 66.00 | 98.00 | 93.00 | 88.00 | 82.00 | 94.00 | 91.00 | 84.22 ± 11.94 |
| TSGSP | 2018 | 84.00 | 62.60 | 56.30 | 99.40 | 94.80 | 83.80 | 94.10 | 93.30 | 90.10 | 84.27 ± 15.01 |
| MMCNN | 2020 | 84.90 | 70.40 | 75.50 | 96.30 | 92.40 | 86.30 | 87.60 | 84.20 | 81.80 | 84.38 ± 7.92 |
| WPD-STDF | 2019 | 69.50 | 64.00 | 86.50 | 96.00 | 94.00 | 87.00 | 83.00 | 95.50 | 92.00 | 85.28 ± 11.47 |
| CD-CNN | 2021 | 79.69 | 60.71 | 82.19 | 96.87 | 94.37 | 89.37 | 82.19 | 93.75 | 90.00 | 85.46 ± 11.08 |
| FBRTS | 2022 | 82.40 | 75.20 | 86.90 | 95.20 | 89.70 | 80.20 | 90.50 | 91.20 | 91.10 | 86.93 ± 6.40 |
| *Ours* | *2022* | *90.28* | *75.00* | *73.61* | *100* | *97.30* | *90.28* | *84.03* | *92.11* | *95.83* | *88.72 ± 9.40* |
| Li M.-A. et al., 2019 | 2019 | – | – | – | – | – | – | – | – | – | 96.48 |
| Li M. et al., 2021 | 2021 | – | – | – | – | – | – | – | – | – | 97.03 |

Two of the algorithms that were excluded from our benchmark comparisons are shown in the table in gray. These algorithms were excluded since they could only be implemented on the second dataset. However, since they showed higher performance than ours on we listed them here to highlight their potential superiority (no individual subject data were available for these works, hence we only show the overall average performance).

(4) Gated Recurrent Unit Recurrent Neural Network Long-Short Term Memory-Recurrent Neural Network (Luo et al., 2018)

(5) Deep Recurrent Spatial-Temporal Neural Network (Ko et al., 2018)

(6) Long-Short Term Memory network (Wang et al., 2018)

(7) Dempster-Shafer Theory (Razi et al., 2019)

(8) Densely Feature Fusion convolutional neural Network (Li D. et al., 2019)

(9) Convolutional Neural Network Long-term Short-term Memory Network (Zhang R. et al., 2019)

(10) Multi-branch 3D Convolutional Neural Network (Zhao et al., 2019)

(11) Channel-Projection Mixed-scale convolutional neural Network (Li Y. et al., 2019)

(12) Convolutional Recurrent Attention Model (Zhang D. et al., 2019)

(13) Weight-based Feature Fusion Convolutional Neural Network (Amin et al., 2019)

(14) Multi-Scale Fusion Convolution Neural Network (Li D. et al., 2020)

(15) Multiple Kernel Stein Spatial Patterns (Galindo-Noreña et al., 2020)

(16) Graph-based Convolutional Recurrent Attention Model (Zhang D. et al., 2020)

(17) Temporal-Spatial Convolutional Neural Network (Chen et al., 2020)

(18) Temporal-Spectral-based Squeeze-and-Excitation Feature Fusion Network (Li Y. et al., 2021)

(19) Shallow Convolution Neural Network and Bidirectional Long-Short Term Memory (Lian et al., 2021)

(20) Temporal Convolutional Networks-Fusion (Musallam et al., 2021)

(21) EEG-Inception-Temporal Network (Salami et al., 2022)

# Results

## Performance

Our algorithm, which we term Sequential Backward Selection with Temporal Filter Bank Common Spatial Patterns (SBS-TFBCSP), significantly outperformed the average performance (79.99% ± 2.23; mean ± std) of all other algorithms. By 12.61% ($T(8) = 5.057$, $p < 0.001$; $t$-test; **Table 1**) and outperformed each of those algorithms individually. The algorithm outperformed the contender leading algorithm (DCR-MEMD) by 4.94%, yet this was not significant ($T(8) = 1.322$, $p = 0.223$, $t$-test). While conceptually

similar, the SBS-FBCSP yielded the lowest score among the methods compared.

Using the second dataset, our algorithm significantly outperformed the average (82.01% ± 3.25) of all other algorithms by 8.18% ($T(8) = 5.697$, $p < 0.001$; $t$-test) and each of those algorithms individually (Table 2). Comparing our algorithm's performance to the leading state-of-the-art contender algorithm (FBRTS), we see a non-significant 2.06% increase in performance favoring our method ($T(8) = 0.707$, $p = 0.499$, $t$-test). The SBS-FBCSP again yielded the lowest performance among the methods compared.

Noting that the performance of SBS-FBCSP is lower across datasets while the key difference between our algorithm and the SBS-FBCSP is the features selected, we suggest that the inclusion of temporal features in the CSPs is likely driving the performance increase (Figure 1). The expansion of the frequency range implemented in our algorithm increases the feature selection granularity, and in turn the performance. As an intuition for the advantage of the method with respect to the feature selection, we show examples (subjects A01, A02; chosen arbitrarily; Figure 1) where the feature-subsets selected by the algorithms are highlighted. In both subjects, a larger proportion of the selected features were drawn from the last 2 s (which SBS-FBCSP would ignore since it averages across the entire 4-s window). Additionally, a number of the selected features were drawn from frequency bands above 30 Hz which would be excluded in the standard SBCSP-SBFS implementations (Bhatti et al., 2019) because they correspond to frequencies not typically associated with MI.

To further investigate the difference between our work and similar methods we highlight two additional algorithms that share various features with ours. The Sparse Filter Bank Common Spatial Pattern (SFBCSP) and the Multiple Windows SFBCSP (SFBCSP-MT) both used a feature selection to choose the CSP features from multiple filter banks (SFBCSP) and 5 (dataset 1) or 6 (dataset 2) time windows (SFBCSP-MT). However, the contender algorithms did not use SBS. Our method outperformed the SFBCSP contender algorithm among eight of the nine subjects using both dataset 1 and 2, and among seven (dataset 1) and eight (dataset 2) of the nine subjects with the SFBCSP-MT contender algorithm (Tables 3, 4).

Taken together, our results suggest that the performance increase is driven by broader choice of inputs, and the feature selection process.

## Results of other performance measure

To further evaluate the performance of our algorithm we used additional standard accuracy metrics. Namely, we estimated the Positive Precision Value (PPV, $^{TP}/_{TP+FP}$), Negative Precision Value (NPV, $^{TN}/_{TN+FN}$), sensitivity (True Positive Rate, $^{TP}/_{TP+FN}$), specificity (True Negative Rate, $^{TN}/_{TN+FP}$), and Kappa value ($^{P_o - P_e}/_{1 - P_e}$), where TP represents the number of testing samples whose real value aligned with the model prediction (True Positive), TN represents the number of testing samples whose real value and model predicted values were both negative (True Negative), FP represents the number of testing samples whose real value is negative while their model predicted value is positive (False Positive), and FN represents the number of testing samples whose real value is positive while their model predicted value is negative (False Negative). $P_o$ is the proportion of observed agreement, and $P_e$ probability that the agreement is at chance. In both dataset 1 (Table 5) and dataset 2 (Table 6) our algorithm proved superior compared to the SBS-FBCSP using those metrics.

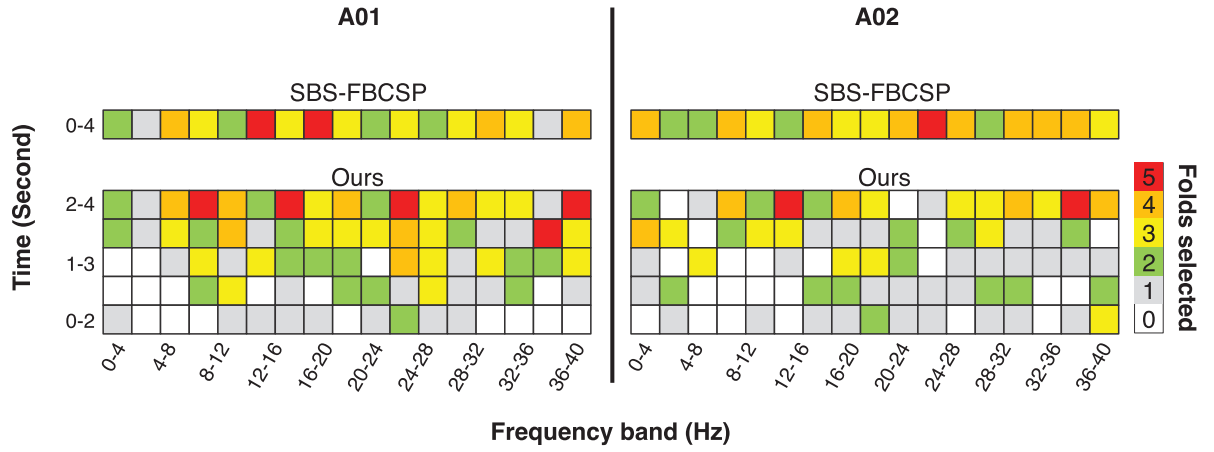



FIGURE 1
Illustration of the results of the 5-fold feature selection comparing a contender algorithm (SBS-FBCSP) and our algorithm, which namely differ in the breakdown of temporal features. The colors (taken from subject A01, A02; chosen arbitrarily) denote the number of times a feature-subset was selected during the 5-fold cross validation.

TABLE 3 Performance comparisons of our method and two similar ones, for dataset 1.

| Measurement | A01 | A02 | A03 | A04 | A05 | A06 | A07 | A08 | A09 | Mean ± std |
|---|---|---|---|---|---|---|---|---|---|---|
| SFBCSP | 82.80 | 59.80 | 92.00 | 68.00 | 81.03 | 59.30 | 89.10 | 92.80 | 70.20 | 77.30 ± 13.30 |
| SFBCSP (MT) | 84.10 | 62.90 | 92.90 | 71.60 | 86.90 | 61.20 | 89.80 | 94.30 | 80.90 | 80.50 ± 12.50 |
| *Ours* | *93.45* | *84.83* | *95.52* | *88.33* | *86.55* | *83.10* | *88.97* | *95.52* | *94.48* | *90.08 ± 4.78* |

Performance metrics for the algorithms were taken from Zhang et al. (2018).

TABLE 4 Performance comparisons of our method and two similar ones, for dataset 2.

| Measurement | B01 | B02 | B03 | B04 | B05 | B06 | B07 | B08 | B09 | Mean ± std |
|---|---|---|---|---|---|---|---|---|---|---|
| SFBCSP | 79.10 | 59.00 | 53.10 | 98.90 | 91.50 | 81.30 | 90.80 | 88.90 | 85.40 | 80.90 ± 15.30 |
| SFBCSP (MT) | 81.80 | 60.30 | 54.00 | 99.10 | 92.60 | 82.00 | 91.80 | 91.10 | 87.30 | 82.20 ± 15.30 |
| *Ours* | *90.28* | *75.00* | *73.61* | *100* | *97.30* | *90.28* | *84.03* | *92.11* | *95.83* | *88.72 ± 9.40* |

Performance metrics for the algorithms were taken from Zhang et al. (2018).

TABLE 5 Performance comparison between SBS-FBCSP and our method, using dataset 1.

| Measurement | A01 | A02 | A03 | A04 | A05 | A06 | A07 | A08 | A09 | Mean ± std |
|---|---|---|---|---|---|---|---|---|---|---|
| SBS-FBCSP | | | | | | | | | | |
| PPV | 81.10 | 74.57 | 85.91 | 70.78 | 75.19 | 72.63 | 75.00 | 80.28 | 78.65 | 77.12 ± 4.75 |
| NPV | 80.04 | 72.64 | 82.97 | 71.93 | 71.46 | 67.41 | 76.85 | 77.52 | 77.41 | 75.36 ± 4.85 |
| TPR | 79.31 | 71.03 | 82.07 | 71.67 | 68.97 | 62.07 | 77.24 | 77.24 | 76.55 | 74.02 ± 6.15 |
| TNR | 80.69 | 75.17 | 85.52 | 69.17 | 77.24 | 75.86 | 73.10 | 78.62 | 78.62 | 77.11 ± 4.64 |
| Kappa | 60.00 | 46.21 | 67.59 | 40.83 | 46.21 | 37.93 | 50.34 | 55.86 | 55.17 | 51.13 ± 9.47 |
| *Ours* | | | | | | | | | | |
| PPV | 94.58 | 85.77 | 94.57 | 88.65 | 88.48 | 82.28 | 90.51 | 94.75 | 92.40 | 90.22 ± 4.35 |
| NPV | 92.89 | 85.48 | 96.62 | 88.39 | 87.62 | 86.56 | 88.10 | 96.56 | 97.35 | 91.06 ± 4.79 |
| TPR | 92.41 | 84.14 | 96.55 | 88.33 | 85.52 | 86.21 | 87.59 | 96.55 | 97.24 | 90.50 ± 5.23 |
| TNR | 94.48 | 85.52 | 94.48 | 88.33 | 87.59 | 80.00 | 90.34 | 94.48 | 91.72 | 89.66 ± 4.89 |
| Kappa | 86.90 | 69.66 | 91.03 | 76.67 | 73.10 | 66.21 | 77.93 | 91.03 | 88.97 | 80.17 ± 9.56 |

TABLE 6 Performance comparison between SBS-FBCSP and our method, using dataset 2.

| Measurement | B01 | B02 | B03 | B04 | B05 | B06 | B07 | B08 | B09 | Mean ± std |
|---|---|---|---|---|---|---|---|---|---|---|
| SBS-FBCSP | | | | | | | | | | |
| PPV | 65.93 | 61.67 | 62.50 | 87.18 | 83.33 | 72.97 | 64.29 | 74.03 | 77.22 | 72.12 ± 9.25 |
| NPV | 77.36 | 59.21 | 62.50 | 91.43 | 81.58 | 74.29 | 70.00 | 74.67 | 83.08 | 74.90 ± 10.09 |
| TPR | 83.33 | 54.41 | 62.50 | 91.89 | 81.08 | 75.00 | 75.00 | 75.00 | 84.72 | 75.88 ± 11.51 |
| TNR | 56.94 | 66.18 | 62.50 | 86.49 | 83.78 | 72.22 | 58.33 | 73.68 | 75.00 | 70.57 ± 10.49 |
| Kappa | 40.28 | 20.59 | 25.00 | 78.38 | 64.86 | 47.22 | 33.33 | 48.68 | 59.72 | 46.45 ± 18.97 |
| *Ours* | | | | | | | | | | |
| PPV | 86.25 | 73.61 | 75.00 | 100 | 97.30 | 92.65 | 78.82 | 92.00 | 95.83 | 87.94 ± 9.95 |
| NPV | 95.31 | 76.56 | 72.37 | 100 | 97.30 | 88.16 | 91.53 | 90.91 | 95.83 | 89.77 ± 9.44 |
| TPR | 95.83 | 77.94 | 70.83 | 100 | 97.30 | 87.50 | 93.06 | 90.79 | 95.83 | 89.90 ± 9.68 |
| TNR | 84.72 | 72.06 | 76.39 | 100 | 97.30 | 93.06 | 75.00 | 92.11 | 95.83 | 87.39 ± 10.60 |
| Kappa | 80.56 | 50.00 | 47.22 | 100 | 94.59 | 80.56 | 68.06 | 82.89 | 91.67 | 77.28 ± 18.74 |

Specifically, with respect to dataset 1, our algorithm significantly outperformed the PPV of the SBS-FBSP (77.12% ± 4.75) by 16.99% ($T(8) = 13.653$, $p < 10^{-7}$; $t$-test), the NPV of the SBS-FBSP (75.36% ± 4.85) by 20.83% ($T(8) = 14.704$, $p < 10^{-7}$; $t$-test), the TPR of the SBS-FBSP (74.02% ± 6.15) by 22.26% ($T(8) = 11.469$, $p < 10^{-6}$; $t$-test), the TNR of the SBS-FBSP (77.11% ± 4.64) by 16.28% ($T(8) = 8.125$, $p < 10^{-5}$; $t$-test), and the Kappa of the

SBS-FBSP (51.13% ± 9.47) by 56.80% ($T(8) = 18.318$, $p < 10^{-8}$; $t$-test).

In dataset 2, our algorithm again significantly outperformed the PPV of the SBS-FBSP (72.12% ± 9.25) by 21.94% ($T(8) = 14.33$, $p < 10^{-7}$; $t$-test), the NPV of the SBS-FBSP (74.90% ± 10.09) by 19.85% ($T(8) = 10.952$, $p < 10^{-6}$; $t$-test), the TPR of the SBS-FBSP (75.88% ± 11.51) by 18.48% ($T(8) = 8.514$, $p < 10^{-5}$; $t$-test), the TNR of the SBS-FBSP (70.57% ± 10.49) by 23.83% ($T(8) = 8.169$, $p < 10^{-5}$; $t$-test), the Kappa of the SBS-FBSP (46.45% ± 18.97) by 66.37% ($T(8) = 15.479$, $p < 10^{-7}$; $t$-test).

## Parameters sensitivity

Given that the performance of our proposed method heavily depends on the selection of the $m$ parameter we tested the robustness of our results by enumerating over all $m$ values possible in dataset 1 (Table 7) and dataset 2 (Table 8). While, indeed, the choice of $m$ impacts the algorithm performance across subjects, the average difference in performance for dataset 1 was 2.23% ± 0.85 and average difference in performance for dataset 2 of 0.99% ± 0.67 (with the highest drop in performance yielding 83.94% accuracy). The lowest performance was aligned with the accuracy of the DCR-MEMD algorithm, but better than all other methods. The highest performance drop yielded an accuracy of 86.79%. which was on par with the FBRTS method but better than all other methods. Combined, these results suggest that the method is robust to perturbations of its single free parameter and maintains its efficiency irrespective of the parameter choice.

Additionally, as our algorithm used temporal windows similar to those suggested in previous work (Zhang et al., 2018), yet the selection of number of windows in both ours and the previous work was arbitrary, we estimated the sensitivity of the algorithm to the selection of window sizes. We altered the number of temporal windows used from 4 to 6 to see the impact of this change on the accuracy. We used this range under the assumption that keeping the number of windows proportional to the number of frequency bands would align with existing works and the theoretical reasoning that they suggest for the bin sizes (Zhang et al., 2018). Testing the algorithm with varying window sizes shows that the range of perturbations yields a performance change of ±1.74%, proportional to the number of windows used. While manipulating the window size impacted the performance, the change was not significant. That is, the impact of ±1 window size usage had a marginal difference in performance (±1.29% on average for dataset 1, and ±0.87% for dataset 2). This non-significant change in performance along with the fact that a change from a single window (SBS-FBCSP) to 5 bins (ours) yields a notable difference suggest that there is a plateau in the performance increment after four bins.

## Ablation study

To further investigate the validity of the proposed algorithm we conducted a series of tests where we hindered the algorithm's inputs and evaluated the performance change. As one key difference between our algorithm and existing ones is the inclusion of both temporal and spectral bands, we varied both input features. In a series of ablation studies, we decreased the range of spectral features from 17 (our algorithm) to

TABLE 7 Performance comparison of different values of $m$ for our method, using dataset 1.

| $m$ | A01 | A02 | A03 | A04 | A05 | A06 | A07 | A08 | A09 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 84.48 | 80.69 | 88.28 | 83.75 | 82.41 | 77.24 | 82.41 | 86.56 | 89.66 |
| 2 | 88.97 | 81.38 | 92.41 | 84.58 | 84.83 | 80.69 | 85.52 | 86.90 | 93.10 |
| 3 | 89.31 | 83.79 | 94.83 | 84.58 | 84.83 | 82.76 | 85.52 | 90.69 | 93.79 |
| 4 | 91.38 | 84.83 | 94.83 | 85.00 | 85.52 | 83.10 | 85.17 | 93.10 | 94.14 |
| 5 | 93.45 | 81.38 | 95.52 | 88.33 | 84.14 | 81.38 | 87.24 | 94.83 | 93.79 |
| 6 | 92.41 | 83.45 | 95.52 | 86.25 | 86.55 | 82.41 | 88.62 | 95.17 | 93.45 |
| 7 | 93.45 | 82.41 | 93.79 | 87.50 | 83.79 | 82.41 | 88.97 | 95.52 | 94.48 |
| STD | 3.21 | 1.51 | 2.59 | 1.70 | 1.32 | 2.02 | 2.27 | 3.85 | 1.62 |

TABLE 8 Performance comparison of different values of $m$ for our method, using dataset 2.

| $m$ | B01 | B02 | B03 | B04 | B05 | B06 | B07 | B08 | B09 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 90.28 | 75.00 | 72.92 | 100 | 93.92 | 90.28 | 84.03 | 91.47 | 93.75 |
| 2 | 88.89 | 72.79 | 73.61 | 100 | 97.30 | 90.28 | 84.03 | 92.11 | 95.83 |
| 3 | 89.58 | 70.69 | 72.22 | 100 | 95.95 | 88.89 | 83.33 | 90.13 | 93.06 |
| STD | 0.70 | 2.16 | 0.70 | 0 | 1.70 | 0.80 | 0.40 | 1.01 | 1.44 |

12 (as is done in contender algorithms) and the range of temporal features from five (our algorithm) to one (as is done in contender algorithms, namely SBS-FBCSP). Across all tests, the feature selection (Sequential Backward Selection) and the classifications parameters were held constant. Across all ablation tests, the performance drop ranged from −4.62% to −2.05% for dataset 1, and −7.33% to −1.61% for dataset 2. Our algorithm remained on par with the state-of-the-art benchmarks despite the drop in performance. The algorithm maintained its superiority for dataset 1, and ranked 15th (out of 21) for dataset 2 at its most hindered state, when the number of frequency bands used was lowest. That is, the selection of time windows and frequency bands that led to our algorithm's performance seem to be mostly sensitive to the number of frequency bands used as inputs (**Figure 2**). Importantly, the drop in frequency bands to a lower number puts our algorithm in line with the contender ones, suggesting that some of the improvement is contingent on this input feature broadening.

Additionally, we replicated the accuracy metrics tests with the ablated inputs to evaluate the impact of the input on performance in an additional manner (**Table 9**). We attempted various implementations of the model with input features ranging from 12 to 17 frequency bands and 3–5 temporal windows. Our algorithm significantly outperformed a variety of contender algorithms with ablated input. Highlighting three of the ablation studies ("Ablation 1" with 60 input features, "Ablation 2" with 51 input features, and "Ablation 3" with 36 input features), our algorithm maintained its performance improvement. Specifically, for ablation test "1" our algorithm outperformed the non-ablated input by over 2% ($T(8) = 4.143$, $p = 0.003$; $t$-test) using dataset 1, and by over 1.5% using dataset 2 ($T(8) = 2.024$, $p = 0.078$; $t$-test). Similarly, in ablation test "2," dataset 1 ($T(8) = 3.869$, $p = 0.005$; $t$-test), dataset 2 ($T(8) = 2.883$, $p = 0.020$; $t$-test) as well as in ablation test "3," dataset 1 ($T(8) = 7.051$, $p < 10^{-4}$; $t$-test) and dataset 2 ($T(8) = 6.553$, $p < 10^{-4}$; $t$-test) the performance was consistency significantly higher for the non-ablated implementation.

## Comparison of computational time

Finally, to demonstrate that the new method is useful for BCI applications, we tested its computational efficiency. As BCIs require not only high decoding accuracy but also relatively fast parsing of the intended motion, a speedy classification is important. We used a 2.67 GHz i5-M480 processor with 4 Gb RAM to analyze the classification speed.

Runtime profiling of the algorithm took 366.91 ± 51.29 s for the entire assessment. While this is nearly 2.8 orders of magnitude longer than the similar contender algorithm (SBS-FBCSP) which took only 8.05 ± 3.02 s) this test compared both the feature selection/validation and classification. As the feature selection is only required for the model training, a comparison of the online classification alone showed that our algorithm is on par with competing algorithms that report their computational efficiency (Zhang et al., 2018). Namely, it is within 3 s from the SBS-FBCSP algorithm (n.s.). Together with the improved classification accuracy, we argue, the sacrifice in computational efficiency still renders our method ideal for BCI applications, and comparable to leading benchmark algorithms.

# Discussion

We evaluated the performance of a novel neural decoding algorithm, which used both temporal and spectral EEG signals, in predicting a motor action planned by subjects. Our algorithm showed increased accuracy of 2.06–4.94% above benchmark algorithms using two different standard dataset (**Tables 1**, **2**).

The main differences between our method and the state-of-the-art algorithms tested were the inclusion of both temporal and spectral signals as inputs, and the extended features selection process. We suggest that these changes are key drivers of the performance improvement. Namely, we propose that the combined feature sets capture information that amplifies the variance within trials of a single individual and therefore increase the performance. To explore this hypothesis, we performed an ablation study where we hindered the inputs by altering the set of features included in the analyses and showed that the decoding accuracy decreased by an average of 3.88%. Even with the drop in accuracy, our algorithm was on par with state-of-the-art algorithms. As a sanity check, our results show that a decrease of the number of temporal features to a single feature yielded performance that was parallel to that of contender methods which only used spectral features.

Given that our method relies on the choice of a free parameter, $m$, we also tested the algorithm's robustness to the parameter selection and showed that the results remain consistent (**Tables 7**, **8**). Further, given that the choice of temporal window size was done arbitrarily in previous works, we tested a range of windows as well as a numbers of frequency bands permutations and showed that the results remain within ±1.49% for dataset 1 and ±1.74% for dataset 2, indicating that the decision is valid and reasonable.

In dataset 2, two algorithms outperformed our implementation. Both algorithms used an approach that deviated from traditional feature extraction methods. One algorithm used multi-scale CNN as a mechanism for the feature selection and the other used montage irregularities. These algorithms' average performance increase was 8.03% (0.85 standard deviations) above our method. Given that both our algorithm and the contender ones show an effective deviation from traditional feature extraction methods, we suggest that a focus on improving this part of the MI classification process may be key to the success of novel methods.

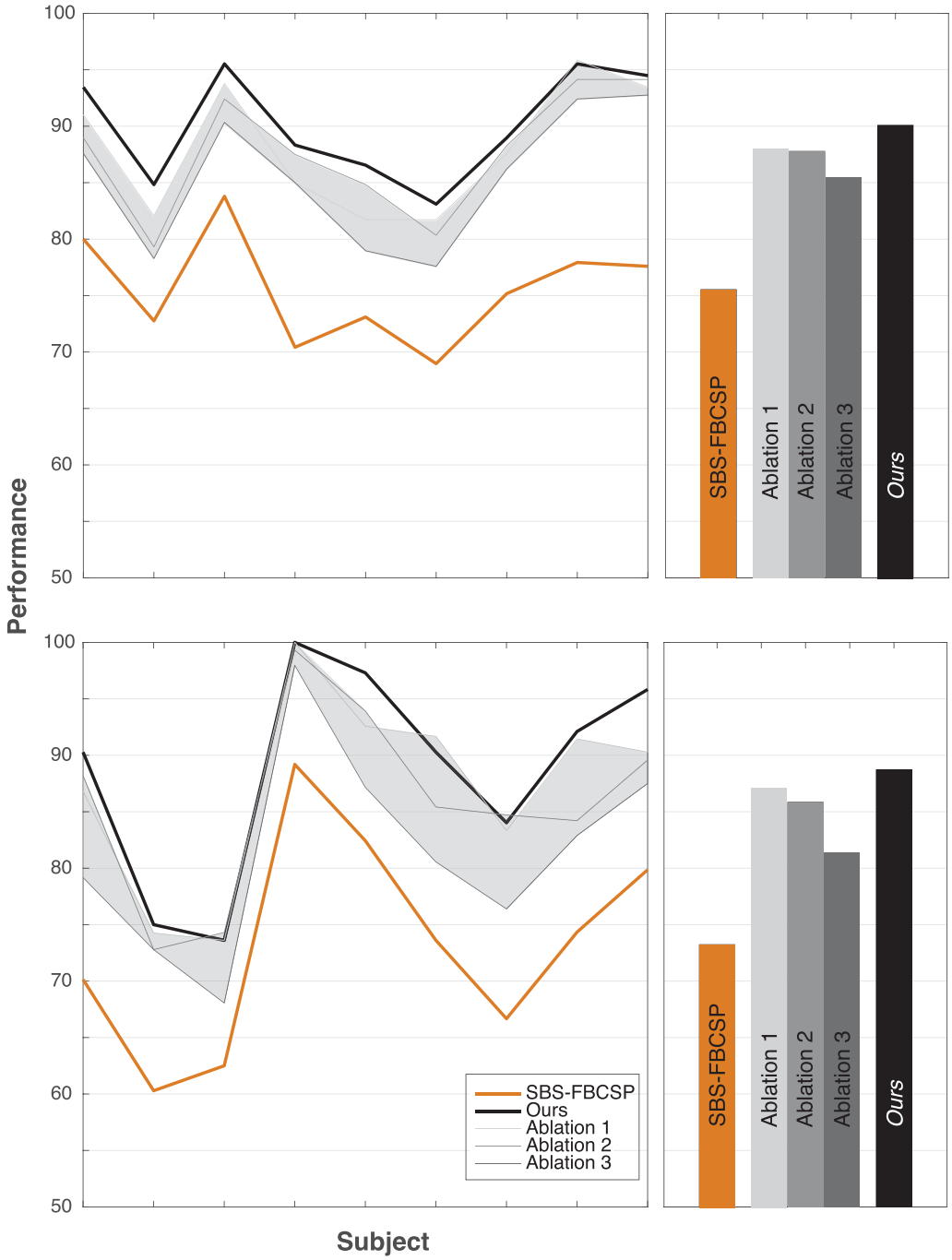



FIGURE 2
Ablation tests. Reducing the number of features as ablation tests for dataset 1 **(top row)** and dataset 2 **(bottom row)**. The shaded areas depict the range of performance for all nine subjects across all ablation tests, with the three tests showing the extremal performance highlighted individually as "Ablation #." Right panels show the average performance across all subjects. "Ablation 1" corresponds to a test that included all five time windows (500–4,500 ms range, with 2 s windows size, and 500 ms step size) and 12 frequency bands (4–30 Hz range, with 4 Hz window size, and 2 Hz step size) for a total of 60 input features (12 bands × 5 time windows) reduced gradually to 10 features through the selection. "Ablation 2" corresponds to a test with three time windows (500–3,500 ms range, with 2 s windows size, and 500 ms step size) and 17 frequency bands (4–40 Hz range, with 4 Hz window size, and 2 Hz step size) for a total of 51 input features (17 bands × 3 time windows) reduced gradually to 10 features through the selection. "Ablation 3" corresponds to a test with three time windows (500–3,500 ms range, with 2 s windows size, and 500 ms step size) and 12 frequency bands (4–30 Hz range, with 4 Hz window size, and 2 Hz step size) for a total of 36 input features (12 bands × 3 time windows) reduced gradually to 10 features through the selection.

TABLE 9 Details of three ablation tests using our algorithm in dataset 1 and 2.

| Measurement | A01 | A02 | A03 | A04 | A05 | A06 | A07 | A08 | A09 | Mean ± std | Significance [$T(8) = (T,p,$ $t$-test)] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Ablation 1 | | | | | | | | | | | |
| PPV | 89.94 | 84.06 | 92.35 | 83.33 | 84.94 | 83.32 | 89.52 | 97.23 | 91.75 | 88.49 ± 4.88 | 2.033, 0.076 |
| NPV | 92.38 | 82.36 | 95.77 | 87.50 | 79.92 | 81.20 | 86.39 | 94.83 | 95.76 | 88.46 ± 6.43 | 3.232, **0.012** |
| TPR | 92.41 | 80.69 | 95.86 | 88.33 | 77.93 | 80.00 | 85.52 | 94.48 | 95.86 | 87.90 ± 7.16 | 2.908, **0.020** |
| TNR | 89.66 | 83.45 | 91.72 | 81.67 | 85.52 | 83.45 | 89.66 | 97.24 | 91.03 | 88.16 ± 5.02 | 1.394, 0.201 |
| Kappa | 82.07 | 64.14 | 87.59 | 70.00 | 63.45 | 63.45 | 75.17 | 91.72 | 86.90 | 76.05 ± 11.34 | 4.140, **0.003** |
| Ablation 2 | | | | | | | | | | | |
| PPV | 88.54 | 79.32 | 94.15 | 89.93 | 84.94 | 76.19 | 90.06 | 94.69 | 92.96 | 87.86 ± 6.51 | 2.235, 0.056 |
| NPV | 89.90 | 80.00 | 91.81 | 86.01 | 84.97 | 86.98 | 87.42 | 93.89 | 96.04 | 88.56 ± 4.91 | 4.034, **0.004** |
| TPR | 89.66 | 80.00 | 91.03 | 85.00 | 84.83 | 88.97 | 86.21 | 93.79 | 95.86 | 88.37 ± 4.91 | 2.701, **0.027** |
| TNR | 88.28 | 78.62 | 93.79 | 90.00 | 84.83 | 71.72 | 90.34 | 94.48 | 92.41 | 87.16 ± 7.59 | 2.021, 0.078 |
| Kappa | 77.93 | 58.62 | 84.83 | 75.00 | 69.66 | 60.69 | 76.55 | 88.28 | 88.28 | 75.54 ± 10.95 | 3.871, **0.005** |
| Ablation 3 | | | | | | | | | | | |
| PPV | 87.23 | 80.10 | 88.30 | 86.12 | 79.23 | 78.15 | 85.33 | 91.70 | 93.07 | 85.47 ± 5.36 | 4.870, **0.001** |
| NPV | 88.94 | 77.16 | 92.92 | 84.55 | 80.72 | 78.15 | 88.37 | 93.82 | 92.59 | 86.36 ± 6.47 | 5.048, $\mathbf{10^{-4}}$ |
| TPR | 88.28 | 75.86 | 93.10 | 84.17 | 78.62 | 77.24 | 88.28 | 93.79 | 92.41 | 85.75 ± 7.07 | 4.792, **0.001** |
| TNR | 86.90 | 80.69 | 87.59 | 85.83 | 79.31 | 77.93 | 84.14 | 91.03 | 93.10 | 85.17 ± 5.18 | 4.313, **0.003** |
| Kappa | 75.17 | 56.55 | 80.69 | 70.00 | 57.93 | 55.17 | 72.41 | 84.83 | 85.52 | 70.92 ± 11.97 | 7.047, $\mathbf{10^{-4}}$ |
| Ablation 1 | | | | | | | | | | | |
| PPV | 82.72 | 75.38 | 75.00 | 100 | 89.87 | 95.45 | 82.43 | 94.37 | 87.18 | 86.93 ± 8.81 | 0.669, 0.552 |
| NPV | 92.98 | 73.24 | 72.37 | 100 | 95.65 | 88.46 | 84.29 | 88.89 | 93.94 | 87.76 ± 9.61 | 2.618, **0.031** |
| TPR | 93.06 | 72.06 | 70.83 | 100 | 95.95 | 87.50 | 84.72 | 88.16 | 94.44 | 87.41 ± 10.19 | 2.577, **0.033** |
| TNR | 80.56 | 76.47 | 76.39 | 100 | 89.19 | 95.83 | 81.94 | 94.74 | 86.11 | 86.80 ± 8.68 | 0.308, 0.766 |
| Kappa | 73.61 | 48.53 | 47.22 | 100 | 85.14 | 83.33 | 66.67 | 82.89 | 80.56 | 74.22 ± 17.42 | 1.897, 0.094 |
| Ablation 2 | | | | | | | | | | | |
| PPV | 83.13 | 69.62 | 75.36 | 100 | 90.12 | 89.23 | 83.78 | 88.24 | 91.30 | 85.64 ± 9.01 | 1.942, 0.088 |
| NPV | 95.08 | 77.19 | 73.33 | 98.67 | 98.51 | 82.28 | 85.71 | 80.95 | 88.00 | 86.64 ± 9.20 | 2.213, 0.058 |
| TPR | 95.83 | 80.88 | 72.22 | 98.65 | 98.65 | 80.56 | 86.11 | 78.95 | 87.50 | 86.59 ± 9.43 | 1.881, 0.097 |
| TNR | 80.56 | 64.71 | 76.39 | 100 | 89.19 | 90.28 | 83.33 | 89.47 | 91.67 | 85.07 ± 10.27 | 1.427, 0.191 |
| Kappa | 76.39 | 45.59 | 48.61 | 98.65 | 87.84 | 70.83 | 69.44 | 68.42 | 79.17 | 71.66 ± 16.94 | 2.938, **0.019** |
| Ablation 3 | | | | | | | | | | | |
| PPV | 74.42 | 69.14 | 66.25 | 100 | 86.67 | 85.48 | 74.36 | 83.78 | 86.49 | 80.73 ± 10.58 | 5.884, $\mathbf{10^{-4}}$ |
| NPV | 86.21 | 78.18 | 70.31 | 96.10 | 87.67 | 76.83 | 78.79 | 82.05 | 88.57 | 82.75 ± 7.71 | 4.510, **0.002** |
| TPR | 88.89 | 82.35 | 73.61 | 95.95 | 87.84 | 73.61 | 80.56 | 81.58 | 88.89 | 83.70 ± 7.42 | 2.953, **0.018** |
| TNR | 69.44 | 63.24 | 62.50 | 100 | 86.49 | 87.50 | 72.22 | 84.21 | 86.11 | 79.08 ± 12.77 | 5.040, **0.001** |
| Kappa | 58.33 | 45.59 | 36.11 | 95.95 | 74.32 | 61.11 | 52.78 | 65.79 | 75.00 | 62.78 ± 17.75 | 6.560, $\mathbf{10^{-4}}$ |

Bold significance values indicate $p$-values below 0.05.

## Contribution

In addition to proposing a new algorithm that implements various suggestions from a large corpus of prior works and yielding an improved performance, we also demonstrate the robustness of the method in multiple ways. We estimate the algorithm on two different datasets (allowing for generalizability of the implementation) and identify dominant parameters driving the performance. We situate the work in the context of existing algorithms and suggest that the process of feature extraction followed by independent classification maximizes the performance yield. Using inputs that are not traditionally considered for MI the expansion of classification set affords the algorithm a richer idiosyncratic noise minimization and tuning option. We show that the algorithm is offering an improvement without considerable hyperparameters tuning. Finally, we show that expanding the input set and the processing steps does not come at a significant cost with respect to decoding speed. The proposed algorithm can show generalized improvement in near real-time on consumer-grade computation tools, making

it a viable method for future implementations by practitioners (Massaro et al., 2020) as well as academics (Cerf et al., 2007).

## Prior works

Our method is not the first to consider the multi-modal structure of EEG signals along with a dedicated classification tool during MI. For example, previous work (Deng et al., 2021) has combined temporally constrained group LASSO with CNN to interpret the underlying mechanisms driving the successful EEGNet decoding (Lawhern et al., 2018). Similarly, a framework for time frequency CSP smoothing was recently implemented to improve EEG decoding performance through ensemble learning (Miao et al., 2021). Both those methods focused on selecting CSP features by ranked weight. Conversely, our method incorporated the temporal features selection using a neural network. The neural networks classifiers were previously suggested as an extension of the establish body of works for MI tasks (Bhatti et al., 2019), yet were not implemented. Our work suggest that the non-linear feature selection provided by the network yields notable performance increase.

Focusing on the neural network implementation, it is noteworthy that a number of classifiers were proposed as variations on the method we used. Due to the recent developments in deep learning algorithms a majority of the methods proposed focused on CNN for the motor classification (Lawhern et al., 2018; Xu et al., 2018; Amin et al., 2019; Dy et al., 2019; Zhang D. et al., 2019; Zhang R. et al., 2019; Zhao et al., 2019; Chen et al., 2020; Ha and Jeong, 2020; Jia et al., 2020; Lian et al., 2021; Musallam et al., 2021). Specifically, Sakhavi et al. (2018) utilized CNN with temporal data, spectral data, and combination of these data to show a notable improvement in the classification performance compared to benchmark methods. Similarly, Dai et al. (2019) and Sun et al. (2020) showed that adoption of Squeeze-and-Excitation networks (Hu et al., 2018) in the CNN architecture improved the classification further because they accounted for the inter-dependencies among the EEG channels in the calibration of the spectral responses. In parallel, Zhang D. et al. (2019, 2020) and Chen et al. (2021) have implemented attentional mechanisms within the neural network architecture to benefit from the temporal dynamics of subject-specific signal properties. In line with these methods, Zhang D. et al. (2019) and Jia et al. (2020) deployed a multi-branch strategy that benefited from the idiosyncratic temporal-properties of different subjects by utilizing complementary networks. Applying the same logic to spatial-temporal signals, Li Y. et al. (2019) used CNN to capture mixed-scale temporal information and improve the decoding accuracy. In addition to improving the input signal features selection, novel methods have focused on bettering the feature discrimination and selection strategies (Yang et al., 2021) and the data augmentation tools (Li Y. et al., 2019; Yang et al., 2021). Specifically, investigating the input features further, Jin et al. (2021) have introduced time filter to a task-related component analyses method that enhanced the signal detection. The works used singular value decomposition to suppresses the general noise and increase the classification accuracy. The method was implemented on steady-state visual evoked potential based BCIs which are different than our data, but it is likely that the method will be useful for our data as well because of the similarity in decoding performance. Beyond similarity in noise reduction, previous works have also improved the feature selection optimization as we did. Jin et al. (2020) implemented feature selection based on the Dempster-Shafer theory which considers the distribution of the features and found the optimal combination of CSPs that minimized the influence of non-stationarity in the signal. Similar to our implementation, this method took into account the inherent defects of CSPs. Further, the work proposed an investigation of the temporal-spectral feature binning for the BCIs similar to the way bins were integrated into the sequential backward feature selection process in our work. Additionally, Jin et al. (2019) have proposed a correlation-based channel selection combined with regularized CSP (RCSP) as a way to improve the classification accuracy. The method seems to align with ours in its performance despite the fact that the RCSP does not consider both the temporal and spectral properties of the MI. The inclusion of both temporal and spectral feature types is suggested in the work as a future endeavor to be investigated. Completing the previous work, Li et al. (2018) have reported that using multiple modality inputs (in their work, both audio and visual signals) to enhance the representation of incoming signals yield increased accuracy in action decoding task (in their case: decoding "crying" vs. "laughing"). The work suggests that in addition to richer signal, the multi-modal inputs afford comprehensive data that benefits from the internal correlation among features. Besides being analogous to our work in their approach to the decoding, these works suggest that rich (or even superfluous or noisy) data inputs can prove useful in classification improvement. While the data in some of the listed works are different (i.e., fMRI data, or different tasks data) we intuit that the methods could be used to improve our work toward an even greater accuracy in the MI decoding. Finally, the network architecture itself was optimized in several works. For example, LSTM and RNN were incorporated in the CNN with the intent to capture additional properties of the EEG signal segments (Ko et al., 2018; Zhang D. et al., 2019; Zhang R. et al., 2019; Lian et al., 2021). Together, all these methods have demonstrated the benefit of incorporating subject-specific temporal properties in the neural network and the advantages these data have in improving the decoding performance.

Since our method implements feature selection and subject-idiosyncratic inputs in the training, as well as

further granular features breakdown along the one discussed here, we suggest that our proposed method benefits from the collection of previous advantages. Namely, as our method separated the feature selection process from the following classification task, we suggest that the two-stage process, which enabled the reduction of features number, is one of the significant drivers of the performance increase.

## Comparison to leading contender algorithms

Comparing our method to an algorithm that uses similar routines (SBS-FBCSP) showed average increased performance of 20.22% (19.28% for dataset 1, and 21.15% for dataset 2). Similarly, a comparison to two other algorithms that share key characteristics with ours, albeit with less direct alignment in the protocol (SFBCSP and SFBCSP-MT) showed an incremental performance increase for our proposed method.

While the SBS-FBCSP was the algorithm conceptually closest to ours and, therefore, the subject of the main comparison, it is useful to highlight some of the similarities and differences between our method and other popular classification protocols.

Examining the notable similarities and differences between our method and 15 methods tested with dataset 1 (Table 10) or 21 tested with dataset 2 (Table 11) we note the main difference being the type of features selected as inputs, and the separation of the feature selection and generation steps.

## Limitations

The proposed decoding method suffers from a number of limitations that are driven by the extension of the temporal components. First, the method requires a priori intuition about the data in order to accurately choose the temporal segments. To prove the method's superiority in datasets where no prior knowledge is available it would be useful to test either arbitrary datasets, or randomly selected temporal windows. If

TABLE 10 Comparison of contender algorithm implemented with dataset 1.

| | Method | Similarity | Difference |
|---|---|---|---|
| 1 | DNN | Combined CSP with neural network | Did not use both temporal and spectral information |
| 2 | KPCA-CILK | Applied a conformal transformation to decrease the non-Euclidian characteristics of the signal while preserving the geometry | Did not use both temporal and spectral information |
| 3 | WOLA-CSP | Performed dynamic filtering of the EEG signal | Implemented the BCI on embedded platform, and did not use both temporal and spectral information |
| 4 | MEMDBF-CSP-LDA | Adopted common spatial pattern on reconstructed data from the multivariate empirical mode decomposition | Did not use both temporal and spectral information |
| 5 | JSTFD-LDA | Considered both temporal and spatial features by extending the CSPs | Did not use spectral information |
| 6 | nCSP-TSLR | Normalized and regularized the CSP to improve performance | Did not use temporal and spectral information |
| 7 | W-CNN | Took the wavelet time-frequency image of the EEG as input for the CNN | Both feature generation and selection done by the neural network |
| 8 | SS-MEMDBF | Utilized the MEMD to extract cross channel information as well as localize the specific frequency information | Did not use temporal information |
| 9 | CSP-FB-LOG | Adopted ensemble learning for feature selection from newly proposed feature extraction based on CSPs | Did not use temporal information |
| 10 | SW-LSR | Introduced the sliding windows techniques into the CSP | Did not use spectral information |
| 11 | EEGnet | Utilized depth-wise and separable convolutions in the CNN | Both feature generation and selection done by the neural network |
| 12 | R-MDRM | Regularized the covariance matrices using data from prior analyses of the EEG channels in small sample settings to reduce calibration time | Did not use both temporal and spectral information |
| 13 | SR-MDRM | Regularized the covariance matrices using data from other subjects in small sample settings to reduce calibration time | Did not use both temporal and spectral information |
| 14 | TSGSP | Adopted Group Lasso selecting the temporal-spectral common spatial pattern features in a multi-task learning manner. The selection of filter banks as well as temporal windows was similar to ours | Used SVM. Tuned three parameters rather than one |
| 15 | DCR-MEMD | Utilized the Gini and Maximum Information Coefficient for optimal channel selection as well as Multivariate Empirical Mode Decomposition for feature extraction | Did not use both temporal and spectral information |

TABLE 11 Comparison of contender algorithm implemented with dataset 2.

| | Method | Similarity | Difference |
|---|---|---|---|
| 1 | RSMM | Adopted a novel unified framework of robust matrix classifier as well as decomposition of EEG solved by alternating direction method of multipliers | Eliminated the effect of outlier and noise. Did not use temporal and spectral information |
| 2 | DLVA | Simultaneously incorporated temporal, spatial and spectral information using a CNN-variational autoencoder | Both feature generation and selection done by the neural network, and had no independent feature selection process |
| 3 | SGRM | Reduced the required training samples from target subject using auxiliary data from other subjects | Did not use temporal and spectral information |
| 4 | UDFS | Utilized an unsupervised feature selection strategy to select the potential CSP feature from multiple frequency bands | Did not use temporal information |
| 5 | SSD-SE-CNN | Proposed squeeze-and-excitation blocks embedded CNN to explore time-frequency features and classification | No independent feature selection process |
| 6 | WaSF ConvNet | End-to-end CNN using time-frequency and spatial information and using wavelet-like kernels to reduce the number of parameters | No independent feature selection process |
| 7 | CSP-Wavelet-LOG | Ensemble learning for feature selection using CSP-based feature extraction | Did not use temporal and spectral information |
| 8 | MAAN | CNN with multi-attention layers to capture the spatial property of the signal as well a domain discriminator to reduce the difference between sessions | Did not use spectral information |
| 9 | MTPP-EEGNet | Multi-layer temporal pyramid pooling approach incorporated into the CNN | Did not use spectral information |
| 10 | UDFS | Utilized an unsupervised feature selection strategy to select the potential CSP features from specific multiple frequency bands | Did not use temporal information |
| 11 | DJDA | Novel dynamic joint domain adaptation network based on adversarial learning strategy to learn domain-invariant feature representation using information from the source session | Did not use temporal and spectral information |
| 12 | SHNN | Consider both temporal and spectral information by segmenting the raw EEG into different windows and band-pass filtering the signal | Both feature generation and selection done by the neural network |
| 13 | TSLDA | Linear discriminant analysis included in covariance matrix | Covariance matrix did not use temporal and spectral information |
| 14 | DRDA | Deep representation-based domain adaptation to improve the classification performance on a single subject using information from multiple subject sources | CNN did not use temporal and spectral information for both the source and target domains |
| 15 | RF-DFFS | Dynamic feature selection strategy where EEG frequency domain features are selected one by one in a boosting protocol | Did not use temporal information |
| 16 | FDBN | Deep Belief Network classifier using the FFT features | Did not use temporal information |
| 17 | TSGSP | Group LASSO selecting the temporal-spectral CSPs in a multi-task learning manner. Selection of filter banks as well as temporal windows | Used SVM. Tuned three parameters rather than one |
| 18 | MMCNN | Multi-scale, multi-branch CNN to overcome the variation between time and subjects using convolution kernel in different sizes | Did not use both temporal and spectral information |
| 19 | WPD-STDF | New spatio-temporal discrepancy feature combined with frequency information | No feature selection strategy |
| 20 | CD-CNN | EEG data augmented using a circular translation strategy, followed by a central vector shift strategy to strengthen the discriminative power of the CNN | Did not use both temporal and spectral information |
| 21 | FBRTS | Fusing the features extracted from CSP as well as multiple time windows | Fusion strategy did not have feature selection strategy |

the method proves superior even with those selections, it will be regarded more robust.

Second, the algorithm has additional degrees of freedom that could be optimized with regards to the selection of hyperparameters. We elected to use the default ones operationalized by the Matlab implementation (Matlab version 2018a) without any additional tuning, but recognize that future work could focus solely on tuning those hyperparameters. Given the lack of theoretical justification for any alternative choice and given that the contender algorithms also used the default hyperparameters without additional emphasis on tuning, we did not deviate from this norm. Ideally, future work will yield theoretical reasoning for some of the tuning alternatives and thereby improve the algorithm's performance as well as its usefulness for varying test cases outside of the MI implemented here.

Third, our method could be orders of magnitude slower in its initial computation training time than other methods. This means that usage of the method for BCIs that continuously update the feature set would either be challenging or require extensive computational resources. To overcome this challenge, one should investigate whether smaller time-window sizes (presumably yielding faster processing) could produce higher performance. Shorter time-window that maintain the high performance would elevate the usefulness of the algorithm.

Fourth, it is not clear whether the method would easily generalize to BCI tasks outside of MI. Specifically, because MI tasks are less likely to show the types of noise that pollutes active motor actions, the fact that our method shows superiority in one domain does not guarantee its success in others. We focused on implementing the method on MI tasks as these are the ones mostly implemented thus far and because of their ecological validity in the context of therapy and rehabilitation (Sokol et al., 2019). Implementing the method in other domains (i.e., language decoding) would validate the usefulness of the method further, or highlight the differences in the BCI uses.

### Future directions

Two research venues that directly extend our work are: i) the enhancement of the features selection granularity (while attempting to maintain the feature-classification performance), and ii) the generalization of the temporal features classification process. Specifically, as EEG and other biological signals are heavily dependent on combined temporal and spectral dynamics, usage of feature selection process with tools such as the recently proposed attention guided neural networks (Sun et al., 2019) may improve the ability to extract the appropriate features without a priori knowledge on the data. This would make the algorithm generalizable to other BCI inputs.

Further, as the majority of the benchmark algorithms we compared use neural networks for the full classification process (thereby effectively using all the available features without pre-selection) we suggest that amending the benchmark algorithms to focusing on the deep learning ones incorporating the two-step selection-classification process may increase the performance of all the benchmark methods.

It has not escaped our notice that as SVMs were previously shown to be superior with respect to feature classification (whereas deep learning networks were shown to be superior in BCI feature selection; Li Y. et al., 2019; Deng et al., 2021; Tiwari and Chaturvedi, 2021) a combination of both methods might improve our algorithm further and allow it to generalize to tasks outside of MI or motor control (i.e., non-verbal communication, language decoding, or parsing of thoughts).

## Conclusion

In this work we have shown that an algorithm which incorporates both temporal and spectral EEG inputs can yield high performance in recognizing which action was imagined by a subject. The algorithm uses SBS technique to reduce the number of inputs and to identify which inputs are less likely to be idiosyncratic across subjects. Once the input features are selected, a RBFNN is used for the classification of the action. We suggest that the method yields performance improvements compared to existing protocols primarily because the inclusion of the large subset of features reduces the individual noise idiosyncrasies within subjects. The suggested algorithm incorporates many of the benefits of the current corpus of state-of-the art BCI protocols and implements the improvements suggestion offered by numerous prior works. In line with these prior suggestion, the method could be applicable for other neural classification problems, modalities, and domains outside of the ones tested herein.

## Data availability statement

Publicly available datasets were analyzed in this study. These datasets can be found here: Leeb et al. (2008).

## Author contributions

GW conducted the analyses. Both authors wrote the manuscript and generated the final output.

## 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

Abiri, R., Borhani, S., Sellers, E. W., Jiang, Y., and Zhao, X. (2019). A comprehensive review of EEG-based brain–computer interface paradigms. *J. Neural Eng.* 16:11001. doi: 10.1088/1741-2552/aaf12e

Ai, Q., Chen, A., Chen, K., Liu, Q., Zhou, T., Xin, S., et al. (2019). Feature extraction of four-class motor imagery EEG signals based on functional brain network. *J. Neural Eng.* 16:026032. doi: 10.1088/1741-2552/ab0328

Al Shiam, A., Islam, M. R., Tanaka, T., and Molla, M. K. I. (2019). "Electroencephalography based motor imagery classification using unsupervised feature selection," in *Proceedings of the 2019 international conference on cyberworlds (CW)*, Kyoto, 239–246. doi: 10.1109/CW.2019.00047

Amin, S. U., Alsulaiman, M., Muhammad, G., Bencherif, M. A., and Hossain, M. S. (2019). Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification. *IEEE Access* 7, 18940–18950. doi: 10.1109/ACCESS.2019.2895688

Andersen, R. A., Aflalo, T., and Kellis, S. (2019). From thought to action: The brain–machine interface in posterior parietal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 116, 26274–26279. doi: 10.1073/pnas.1902276116

Ang, K. K., Chin, Z. Y., Zhang, H., and Guan, C. (2008). "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proceedings of the 2008 IEEE international joint conference on neural networks (IEEE World congress on computational intelligence)*, (Hong Kong: IEEE), 2390–2397.

Belwafi, K., Romain, O., Gannouni, S., Ghaffari, F., Djemal, R., and Ouni, B. (2018). An embedded implementation based on adaptive filter bank for brain–computer interface systems. *J. Neurosci. Methods* 305, 1–16. doi: 10.1016/j.jneumeth.2018.04.013

Bhatti, M. H., Khan, J., Khan, M. U. G., Iqbal, R., Aloqaily, M., Jararweh, Y., et al. (2019). Soft computing-based EEG classification by optimal feature selection and neural networks. *IEEE Trans. Industr. Inform.* 15, 5747–5754. doi: 10.1109/TII.2019.2925624

Bulárka, S., and Gontean, A. (2016). "Brain-computer interface review," in *Proceedings of the 2016 12th IEEE international symposium on electronics and telecommunications (ISETC)*, (Timisoara: IEEE), 219–222. doi: 10.1109/ISETC.2016.7781096

Cerf, M., and Garcia-Garcia, M. (2017) *Consumer neuroscience*. Cambridge, MA: MIT Press.

Cerf, M., Cleary, D. R., Peters, R. J., Einhäuser, W., and Koch, C. (2007). Observers are consistent when rating image conspicuity. *Vision Res.* 47, 3052–3060. doi: 10.1016/j.visres.2007.06.025

Cerf, M., Thiruvengadam, N., Mormann, F., Kraskov, A., Quiroga, R. Q., Koch, C., et al. (2010). On-line, voluntary control of human temporal lobe neurons. *Nature* 467, 1104–1108. doi: 10.1038/nature09510

Chen, J., Yu, Z., Gu, Z., and Li, Y. (2020). Deep temporal-spatial feature learning for motor imagery-based brain–computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 2356–2366. doi: 10.1109/TNSRE.2020.3023417

Chen, P., Gao, Z., Yin, M., Wu, J., Ma, K., and Grebogi, C. (2021). Multiattention adaptation network for motor imagery recognition. *IEEE Trans. Syst. Man Cybernet. Syst.* 52, 5127–5139. doi: 10.1109/TSMC.2021.3114145

Chiuzbaian, A., Jakobsen, J., and Puthusserypady, S. (2019). "Mind controlled drone: An innovative multiclass SSVEP based brain computer interface," in *Proceedings of the 2019 7th international winter conference on brain-computer interface (BCI)*, Gangwon, 1–5. doi: 10.1109/IWW-BCI.2019.8737327

Courtine, G., Micera, S., DiGiovanna, J., del, R., and Millán, J. (2013). Brain–machine interface: Closer to therapeutic reality? *Lancet* 381, 515–517. doi: 10.1016/S0140-6736(12)62164-3

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. doi: 10.3390/s19030551

Davoudi, A., Ghidary, S. S., and Sadatnejad, K. (2017). Dimensionality reduction based on distance preservation to local mean for symmetric positive definite matrices and its application in brain–computer interfaces. *J. Neural Eng.* 14:036019. doi: 10.1088/1741-2552/aa61bb

Deng, X., Zhang, B., Yu, N., Liu, K., and Sun, K. (2021). Advanced TSGL-EEGNet for motor imagery EEG-based brain-computer interfaces. *IEEE Access* 9, 25118–25130. doi: 10.1109/ACCESS.2021.3056088

Dy, Z., Fz, T., and Bl, S. (2019). Learning joint space–time–frequency features for EEG decoding on small labeled data. *Neural Netw.* 114, 67–77. doi: 10.1016/j.neunet.2019.02.009

Fang, H., Jin, J., Daly, I., and Wang, X. (2022). Feature extraction method based on filter banks and Riemannian tangent space in motor-imagery BCI. *IEEE J. Biomed. Health Inform.* 26, 2504–2514. doi: 10.1109/JBHI.2022.3146274

Galindo-Noreña, S., Cárdenas-Peña, D., and Orozco-Gutierrez, Á (2020). Multiple kernel stein spatial patterns for the multiclass discrimination of motor imagery tasks. *Appl. Sci.* 10:8628. doi: 10.3390/app10238628

Gaur, P., Gupta, H., Chowdhury, A., McCreadie, K., Pachori, R. B., and Wang, H. (2021). A sliding window common spatial pattern for enhancing motor imagery classification in EEG-BCI. *IEEE Trans. Instrum. Meas.* 70, 1–9. doi: 10.1109/TIM.2021.3051996

Gaur, P., Pachori, R. B., Wang, H., and Prasad, G. (2018). A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry. *Expert Syst. Appl.* 95, 201–211. doi: 10.1016/j.eswa.2017.11.007

Gaur, P., Pachori, R. B., Wang, H., and Prasad, G. (2019). An automatic subject specific intrinsic mode function selection for enhancing two-class EEG-based motor imagery-brain computer interface. *IEEE Sens. J.* 19, 6938–6947. doi: 10.1109/JSEN.2019.2912790

Gordleeva, S. Y., Lukoyanov, M. V., Mineev, S. A., Khoruzhko, M. A., Mironov, V. I., Kaplan, A. Y., et al. (2017). Exoskeleton control system based on motor-imaginary brain–computer interface. *Mod. Technol. Med.* 9:31. doi: 10.17691/stm2017.9.3.04

Ha, K.-W., and Jeong, J.-W. (2020). Temporal pyramid pooling for decoding motor-imagery EEG signals. *IEEE Access* 9, 3112–3125. doi: 10.1109/ACCESS.2020.3047678

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

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

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, (Salt Lake City, UT: IEEE), 7132–7141. doi: 10.1109/CVPR.2018.00745

Jia, Z., Lin, Y., Wang, J., Yang, K., Liu, T., and Zhang, X. (2020). "MMCNN: A multi-branch multi-scale convolutional neural network for motor imagery classification," in *Proceedings of the joint European conference on machine learning and knowledge discovery in databases*, eds F. Hutter, K. Kersting, J. Lijffijt, and I. Valera (Cham: Springer), 736–751. doi: 10.1007/978-3-030-67664-3_44

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

Jiao, Y., Zhang, Y., Chen, X., Yin, E., Jin, J., Wang, X., et al. (2018). Sparse group representation model for motor imagery EEG classification. *IEEE J. Biomed. Health Inform.* 23, 631–641. doi: 10.1109/JBHI.2018.2832538

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

Jin, J., Wang, Z., Xu, R., Liu, C., Wang, X., and Cichocki, A. (2021). Robust similarity measurement based on a novel time filter for SSVEPs detection. *IEEE Trans. Neural Netw. Learn. Syst.* doi: 10.1109/TNNLS.2021.3118468 [Epub ahead of print].

Jin, J., Xiao, R., Daly, I., Miao, Y., Wang, X., and Cichocki, A. (2020). Internal feature selection method of CSP based on L1-norm and Dempster–Shafer theory. *IEEE Trans. Neural Netw. Learn. Syst.* 32, 4814–4825. doi: 10.1109/TNNLS.2020.3015505

Ko, W., Yoon, J., Kang, E., Jun, E., Choi, J.-S., and Suk, H.-I. (2018). "Deep recurrent spatio-temporal neural network for motor imagery based BCI," in *Proceedings of the 2018 6th international conference on brain-computer interface (BCI)*, (Gangwon: IEEE), 1–3. doi: 10.1109/IWW-BCI.2018.8311535

Kumar, S., Sharma, A., Mamun, K., and Tsunoda, T. (2016). "A deep learning approach for motor imagery EEG signal classification," in *Proceedings of the 2016 3rd Asia-Pacific world congress on computer science and engineering (APWC on CSE)*, (Nadi: IEEE), 34–39. doi: 10.1109/APWC-on-CSE.2016.017

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.* 15:56013. doi: 10.1088/1741-2552/aace8c

Lebedev, M. A., and Nicolelis, M. A. L. (2006). Brain–machine interfaces: Past, present and future. *Trends Neurosci.* 29, 536–546. doi: 10.1016/j.tins.2006.07.004

Leeb, R., Brunner, C., Müller-Putz, G., Schlögl, A., and Pfurtscheller, G. (2008). *BCI competition 2008–Graz data set B*. Graz: Graz University of Technology, 1–6.

Li, D., Wang, J., Xu, J., and Fang, X. (2019). Densely feature fusion based on convolutional neural networks for motor imagery EEG classification. *IEEE Access* 7, 132720–132730. doi: 10.1109/ACCESS.2019.2941867

Li, D., Xu, J., Wang, J., Fang, X., and Ji, Y. (2020). A multi-scale fusion convolutional neural network based on attention mechanism for the visualization analysis of EEG signals decoding. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 2615–2626. doi: 10.1109/TNSRE.2020.3037326

Li, M., Han, J., and Yang, J. (2021). Automatic feature extraction and fusion recognition of motor imagery EEG using multilevel multiscale CNN. *Med. Biol. Eng. Comput.* 59, 2037–2050. doi: 10.1007/s11517-021-02396-w

Li, M.-A., Han, J.-F., and Duan, L.-J. (2019). A novel MI-EEG imaging with the location information of electrodes. *IEEE Access* 8, 3197–3211. doi: 10.1109/ACCESS.2019.2962740

Li, Y., Guo, L., Liu, Y., Liu, J., and Meng, F. (2021). A temporal-spectral-based squeeze-and-excitation feature fusion network for motor imagery EEG decoding. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 1534–1545. doi: 10.1109/TNSRE.2021.3099908

Li, Y., Wang, F., Chen, Y., Cichocki, A., and Sejnowski, T. (2018). The effects of audiovisual inputs on solving the cocktail party problem in the human brain: An fMRI study. *Cereb. Cortex* 28, 3623–3637. doi: 10.1093/cercor/bhx235

Li, Y., Zhang, X.-R., Zhang, B., Lei, M.-Y., Cui, W.-G., and Guo, Y.-Z. (2019). A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 1170–1180. doi: 10.1109/TNSRE.2019.2915621

Lian, S., Xu, J., Zuo, G., Wei, X., and Zhou, H. (2021). A novel time-incremental end-to-end shared neural network with attention-based feature fusion for multiclass motor imagery recognition. *Comput. Intell. Neurosci.* 2021:6613105. doi: 10.1155/2021/6613105

Liu, C., Jin, J., Daly, I., Li, S., Sun, H., Huang, Y., et al. (2022). SincNet-based hybrid neural network for motor imagery EEG decoding. *IEEE Trans. Neural Syst. Rehabil. Eng.* 30, 540–549. doi: 10.1109/TNSRE.2022.3156076

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. *J. Neural Eng.* 4, R1–R13. doi: 10.1088/1741-2560/4/2/R01

Lu, N., Li, T., Ren, X., and Miao, H. (2016). A deep learning scheme for motor imagery classification based on restricted Boltzmann machines. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 566–576. doi: 10.1109/TNSRE.2016.2601240

Luo, J., Feng, Z., and Lu, N. (2019). Spatio-temporal discrepancy feature for classification of motor imageries. *Biomed. Signal Process. Control* 47, 137–144. doi: 10.1016/j.bspc.2018.07.003

Luo, J., Feng, Z., Zhang, J., and Lu, N. (2016). Dynamic frequency feature selection based approach for classification of motor imageries. *Comput. Biol. Med.* 75, 45–53. doi: 10.1016/j.compbiomed.2016.03.004

Luo, T., Zhou, C., and Chao, F. (2018). Exploring spatial-frequency-sequential relationships for motor imagery classification with recurrent neural network. *BMC Bioinformatics* 19:344. doi: 10.1186/s12859-018-2365-1

Massaro, S., Drover, W., Cerf, M., and Hmieleski, K. M. (2020). Using functional neuroimaging to advance entrepreneurial cognition research. *J. Small Bus. Manag.* 1–29. doi: 10.1080/00472778.2020.1824527

Miao, Y., Jin, J., Daly, I., Zuo, C., Wang, X., Cichocki, A., et al. (2021). Learning common time-frequency-spatial patterns for motor imagery classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 699–707. doi: 10.1109/TNSRE.2021.3071140

Molla, M. K. I., Al Shiam, A., Islam, M. R., and Tanaka, T. (2020). Discriminative feature selection-based motor imagery classification using EEG signal. *IEEE Access* 8, 98255–98265. doi: 10.1109/ACCESS.2020.2996685

Musallam, Y. K., AlFassam, N. I., Muhammad, G., Amin, S. U., Alsulaiman, M., Abdul, W., et al. (2021). Electroencephalography-based motor imagery classification using temporal convolutional network fusion. *Biomed. Signal Process. Control* 69:102826. doi: 10.1016/j.bspc.2021.102826

Nader, M., Jacyna-Golda, I., Nader, S., and Nehring, K. (2021). Using BCI and EEG to process and analyze driver's brain activity signals during VR simulation. *Transport* 60, 137–153. doi: 10.5604/01.3001.0015.6305

Novi, Q., Guan, C., Dat, T. H., and Xue, P. (2007). "Sub-band common spatial pattern (SBCSP) for brain-computer interface," in *Proceedings of the 2007 3rd international IEEE/EMBS conference on neural engineering*, (Kohala Coast, HI: IEEE), 204–207. doi: 10.1109/CNE.2007.369647

Olias, J., Martín-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.* 27, 895–904. doi: 10.1109/TNSRE.2019.2905894

Pasyuk, A., Semenov, E., and Tyuhtyaev, D. (2019). "Feature selection in the classification of network traffic flows," in *Proceedings of the 2019 international multi-conference on industrial engineering and modern technologies (FarEastCon)*, (Vladivostok: IEEE), 1–5. doi: 10.1109/FarEastCon.2019.8934169

Prashant, P., Joshi, A., and Gandhi, V. (2015). "Brain computer interface: A review," in *Proceedings of the 2015 5th Nirma University international conference on engineering (NUiCONE)*, (Ahmedabad: IEEE), 1–6. doi: 10.1109/NUICONE.2015.7449615

Razi, S., Mollaei, M. R. K., and Ghasemi, J. (2019). A novel method for classification of BCI multi-class motor imagery task based on Dempster–Shafer theory. *Inform. Sci.* 484, 14–26. doi: 10.1016/j.ins.2019.01.053

Sadatnejad, K., and Ghidary, S. S. (2016). Kernel learning over the manifold of symmetric positive definite matrices for dimensionality reduction in a BCI application. *Neurocomputing* 179, 152–160. doi: 10.1016/j.neucom.2015.11.065

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

Salami, A., Andreu-Perez, J., and Gillmeister, H. (2022). EEG-ITNet: An explainable inception temporal convolutional network for motor imagery classification. *IEEE Access* 10, 36672–36685. doi: 10.1109/ACCESS.2022.3161489

Singh, A., Lal, S., and Guesgen, H. W. (2019). Small sample motor imaginery classification using regularized riemannian features. *IEEE Access* 7, 46858–46869.

Sokol, L. L., Young, M. J., Paparian, J., Kluger, B. M., Lum, H. D., Besbris, J., et al. (2019). Advance care planning in Parkinson's disease: Ethical challenges and future directions. *NPJ Parkinsons Dis.* 5:24. doi: 10.1038/s41531-019-0098-0

Song, Y., and Sepulveda, F. (2018). A novel technique for selecting EMG-contaminated EEG channels in self-paced brain–computer Interface task onset. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 1353–1362. doi: 10.1109/TNSRE.2018.2847316

Sun, B., Zhao, X., Zhang, H., Bai, R., and Li, T. (2020). EEG motor imagery classification with sparse spectrotemporal decomposition and deep learning. *IEEE Trans. Autom. Sci. Eng.* 18, 541–551. doi: 10.1109/TASE.2020.3021456

Sun, L., Shao, W., Zhang, D., and Liu, M. (2019). Anatomical attention guided deep networks for ROI segmentation of brain MR images. *IEEE Trans. Med. Imaging* 39, 2000–2012. doi: 10.1109/TMI.2019.2962792

Tiwari, A., and Chaturvedi, A. (2021). A novel channel selection method for BCI classification using dynamic channel relevance. *IEEE Access* 9, 126698–126716. doi: 10.1109/ACCESS.2021.3110882

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.* 28, 3239–3258. doi: 10.1007/s00521-016-2236-5

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

Xu, B., Zhang, L., Song, A., Wu, C., Li, W., Zhang, D., et al. (2018). Wavelet transform time-frequency image and convolutional network-based motor imagery EEG classification. *IEEE Access* 7, 6084–6093. doi: 10.1109/ACCESS.2018.2889093

Yang, L., Song, Y., Ma, K., and Xie, L. (2021). Motor imagery EEG decoding method based on a discriminative feature learning strategy. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 368–379. doi: 10.1109/TNSRE.2021.3051958

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 Inform.* 24, 2570–2579. doi: 10.1109/JBHI.2020.2967128

Zhang, D., Yao, L., Chen, K., and Monaghan, J. (2019). A convolutional recurrent attention model for subject-independent EEG signal analysis. *IEEE Signal Process. Lett.* 26, 715–719. doi: 10.1109/LSP.2019.2906824

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, S., Zhu, Z., Zhang, B., Feng, B., Yu, T., and Li, Z. (2020). The CSP-based new features plus non-convex log sparse feature selection for motor imagery EEG classification. *Sensors* 20:4749.

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

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

Zhao, X., Zhang, H., Zhu, G., You, F., Kuang, S., and Sun, L. (2019). A multi-branch 3D convolutional neural network for EEG-based motor imagery classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 2164–2177. doi: 10.1109/TNSRE.2019.2938295

Zheng, Q., Zhu, F., and Heng, P.-A. (2018). Robust support matrix machine for single trial EEG classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 551–562. doi: 10.1109/TNSRE.2018.2794534

frontiers | Frontiers in Neuroinformatics

TYPE Review
PUBLISHED 25 October 2022
DOI 10.3389/fninf.2022.997282



OPEN ACCESS

EDITED BY
Duo Chen,
Nanjing University of Chinese Medicine, China

REVIEWED BY
Wentao Wang,
Oracle, United States
Rj Wang,
PCL, China

*CORRESPONDENCE
Jipeng Wu
jp.wu@siat.ac.cn
Zhiyu You
youzhiyu022@163.com
Jianbing Xiahou
jbxiahou@qztc.edu.cn
Menglin Ouyang
ouyangmenglin@163.com

†These authors have contributed equally to this work

RECEIVED 18 July 2022
ACCEPTED 11 October 2022
PUBLISHED 25 October 2022

CITATION
Cui X, Wu Y, Wu J, You Z, Xiahou J and Ouyang M (2022) A review: Music-emotion recognition and analysis based on EEG signals. *Front. Neuroinform.* 16:997282. doi: 10.3389/fninf.2022.997282



# A review: Music-emotion recognition and analysis based on EEG signals

Xu Cui[1†], Yongrong Wu[2†], Jipeng Wu[3*], Zhiyu You[4*], Jianbing Xiahou[5*] and Menglin Ouyang[6*]

[1]The Art School, Xiamen University, Xiamen, China, [2]Department of Software Engineering, Xiamen University, Xiamen, China, [3]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [4]National Institute for Data Science in Health and Medicine, School of Medicine, Xiamen University, Xiamen, China, [5]The Mathematics and Computer School, Quanzhou Normal University, Quanzhou, China, [6]The Affiliated Hospital of Medical School, Ningbo University, Ningbo, China

Music plays an essential role in human life and can act as an expression to evoke human emotions. The diversity of music makes the listener's experience of music appear diverse. Different music can induce various emotions, and the same theme can also generate other feelings related to the listener's current psychological state. Music emotion recognition (MER) has recently attracted widespread attention in academics and industry. With the development of brain science, MER has been widely used in different fields, e.g., recommendation systems, automatic music composing, psychotherapy, and music visualization. Especially with the rapid development of artificial intelligence, deep learning-based music emotion recognition is gradually becoming mainstream. Besides, electroencephalography (EEG) enables external devices to sense neurophysiological signals in the brain without surgery. This non-invasive brain-computer signal has been used to explore emotions. This paper surveys EEG music emotional analysis, involving the analysis process focused on the music emotion analysis method, e.g., data processing, emotion model, and feature extraction. Then, challenging problems and development trends of EEG-based music emotion recognition is proposed. Finally, the whole paper is summarized.

KEYWORDS

**EEG, emotions, music, recognition, song**

## 1. Introduction

Music is a product created by musicians in the brain and mapped to reality through playing and is also a meaningful way to express emotions and generate emotional resonance. Music is a sound symbol representing people's thoughts and has specific acoustic and structural characteristics. The structure and features of music, such as pitch, tonality, rhythm, and volume, play an essential role in emotional expression. For example, fast rhythms (dense rhythms) typically indicate high arousal levels, while slow rhythms (sparse rhythms) indicate low arousal emotions. Music has many elements, including loudness, pitch, beat, tempo, rhythm, melody, harmony, texture,

style, timbre, dynamics, structure, and more. Generally speaking, music has three essential characteristics: loudness, pitch, and timbre (Lin et al., 2010). Music is a type of sound caused by objects vibrating the air regularly. It enters the brain through the human ear and acts on the cranial nerves. Besides, music is one of the ways of expressing emotions, which can evoke strong emotional responses from listeners with good cross-cultural, consistent consistency. Listening to music is an easy and effective way to change your mood or reduce stress. Neuroscientific research has found that the melody and rhythm of music directly affect the human ear, and the listener can perceive the emotion expressed by music in the subconscious state. Studies have shown that music can induce various emotions in humans, including sadness, nostalgia, tension, happiness, relaxation, calmness, and joy. Listening to music causes emotional mechanisms that are very complex, which often involves a series of related interdisciplinary knowledge such as psychology, physiology, and neuroscience. It isn't easy to find an intuitive quantitative method to analyze. When musical stimulation acts on a person's auditory system, the brain will temporarily increase the activity of specific nerve cells, thereby affecting the expression of people's emotions. Music signals and EEG are closely related in the root (Lin et al., 2006). The EEG often used to verify the influence of music on human brain activity. Studying the relationship between EEG signals and music can help discover how music activates corresponding neurons in the brain. And it can also affect the expression of human emotions. Therefore, studying the correlation between music and EEG signals is of great significance.

Human emotion is a complex phenomenon closely related to human mental state and emotion, including psychological, physiological, and social aspects (Cacioppo et al., 2021). Emotions can be described as changes in the form of the human brain in response to specific events. Emotions play a crucial role in numerous daily experiences of human beings and exert significant influence on their life, such as cognition, perception, and rational decision-making. Given the correlation between the two, emotion has been studied in psychology, philosophy, and neurobiology, which is consistent with the basis of emotion neuroscience (Panksepp, 2004). Although emotions are essential to human communication and interaction, until now, there has been no transparent, automated system for emotion recognition in human society. With the continuous development of artificial intelligence, the demand for emotional intelligence in the human-machine interface is increasingly amplified (Picard, 2003). Based on this, developing and implementing a new automatic emotional recognition algorithm becomes crucial. In this way, it is possible to improve and humanize the interaction between humans and machines. Many studies (Cloitre et al., 2019; Dickerson et al., 2020) have focused on detecting emotions by analyzing physical responses to emotional conditions. To this end, many studies have assessed the effects of emotion on different physiological variables, with particular efforts being made to evaluate EEG recordings (Alarcao and Fonseca, 2017). In recent decades, EEG-based emotion recognition research has gained popularity in many disciplines. However, available scientific data on emotional states and their structure are still limited. Researchers have verified that certain music can change neural activity and stimulate people's potential. Activity in the EEG correlates strongly with music-induced emotions. Vuilleumier and Trost (2015) shows that emotion recognition in music is a relatively rapid process.

Nordström and Laukka (2019) found that music-evoked emotions share some characteristics with basic emotions that map to valences observed in other emotional contexts (Vuilleumier and Trost, 2015). These findings suggest that emotional responses are multidimensional. Using functional magnetic resonance imaging (fMRI), Bodner and Shaw compared the changes in electrical signals in the brain as participants listened to Mozart's music and other music (Bodner et al., 2001). The results showed that in addition to the expected temporal lobe activation, the subjects listening to Mozart also activated the brain's frontal lobe, with significant $\alpha$-wave changes. This phenomenon may be because Mozart's music is highly organized, and the regular repetition of melody is similar to the rhythmic cycle of brain electricity, which affects the human body. In addition, most previous studies on emotion recognition are based on body movement, posture, voice, and expression. However, tone and facial expressions can be deliberately hidden, so the corresponding credit is inaccurate. In contrast, EEG captures the electrical signals produced by neurons, and humans cannot control physiological signals on purpose, so EEG-based emotion recognition is more reliable.

Brain functional network studies include magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), and functional near-infrared spectroscopy (fNIRS). MRI is a very complex imaging examination. MRI provides in a particular brain, such as a tumor patient's. MRI shows only a still image of the brain, which is an anatomical image, with no actual brain activdes a map of the brain at a given moment. This structural information can be used to determine the size of a person's brain or whether there are abnormalitity. To get images of brain activity, fMRI is needed. FMRI can observe the specific brain activity clearly under test in the form of pictures, which has incomparable advantages. However, fMRI equipment will produce huge background noise when working, which seriously affects the effect of music appreciation. Meanwhile, fMRI has low time resolution, and the brain imaging obtained by fMRI is the average of brain activity within a few seconds. However, the brain responds effectively to stimuli within 30 ms, and the temporal resolution of fMRI makes it impossible to accurately analyze brain activity. fNIRS is the near-infrared spectroscopy (NIRS) used in functional neuroimaging. In application, the brain's response can be detected according to the behavior of neurons. It is a neuroimaging technique that uses spectroscopy

to measure the level of brain activity. Functional near-infrared spectroscopy (FNIR) is a non-invasive functional brain imaging technique developed in recent years. But all of these methods are more expensive than EEG, require more training, and, most importantly, are less portable. And in the experiment, sometimes as long as you wear headphones, you can detect the quality of the data, so EEG is more widely used in the test.

The connection between music and the brain is interconnected. People with hearing impairments can't perceive pitch and rhythm correctly, but their brains are unaffected. Different music can trigger different emotions, and the same music can trigger other emotions, which has a lot to do with the current state of mind of the audience. In recent years, with the development of brain science, music, mood, and the correlation with EEG research have gradually become cutting-edge research in this area. With the help of EEG, research content can use machine learning algorithms to capture and study brain activity signals to predict emotional recognition patterns and translate them into commands. The development trend of EEG recognition of music arousing emotion is proposed. Based on this, this paper makes some conclusions about the brain mapping of music-induced emotion.

The main contributions of this paper are as follows:

- We give a detailed introduction to the music-based emotion recognition method and analyze the emotion analysis of EEG signals in detail.
- We introduced EEG-based music emotion recognition methods in detail, including data collection and processing, feature extraction, etc
- We discuss current challenges and future research directions for EEG-based music emotion.

The rest of the paper mainly includes some preliminary knowledge of music emotion recognition is introduced in Section 2. Then Section 3 reviews and summarizes the framework of EEG-based music emotion recognition. Section 4 discusses the current challenges and future work. Finally, we conclude this paper in Section 5.

## 2. Preliminary knowledge

With the rapid development of information science, various digital technologies are flourishing. Audio digitization has become an intermediate force in the digital trend. Audio is the result of digitizing and preserving sound, and music is an essential part of the audio. In recent years, digital music has attracted many scholars to research it. Compared with traditional music, digital music reflects its advantages in production, storage, dissemination, and retrieval, e.g., digital signal processing can reduce the cost of music storage and the application. At the same time, the popularization of the Internet has promoted the spread of digital music. With the development of audio retrieval technology, text content is unable to meet users' needs. In this context, digital audio content has become a hot spot of digital audio technology.

However, music's essential feature is based on emotion, and more music studies point out that emotion is a necessary criterion for people in music retrieval. The core of emotion-based music retrieval is music emotion recognition, which is the essential research direction of digital music. Music emotion recognition research covers musicology, psychology, music acoustics, audio signal processing, natural language processing, machine learning, and other fields. More than 100 related studies have shown that different listeners are generally consistent in judging the emotions expressed by music. Therefore, it is possible to perform emotion recognition with high accuracy.

Music emotion recognition began in the 1960s and has been around for decades. At the beginning of artificial intelligence, some people proposed the relationship between music content and emotion. Since then, universities and research institutes have conducted related research. In recent years, with the development of artificial intelligence, music emotion recognition research has progressed rapidly and has been successfully applied in various fields, e.g., music emotion retrieval, music art performance, intelligent space design, etc. Music emotion recognition technology is the research field of music visualization. It has laid the foundation with broad research prospects and essential application value.

In recent years, music emotion recognition has progressed rapidly and become one of the important research directions in digital music. In the current music emotion research, there are many problems need to be resolved, such as the scarcity of music emotion data sets, the difficulty of emotion quantification, and the limited accuracy of emotion recognition, with the test. This section will introduce the research status of music emotion recognition.

### 2.1. Music emotion recognition

Music emotion recognition is an interdisciplinary work that requires deep analysis and understanding of music, which not only involves signal processing and artificial intelligence but also needs to understand many fields such as auditory perception, psychology, and music theory. People's judgment of music emotion is subjective, and different experiences or ideas will influence the determination of music emotion. The characteristics of music, e.g., timbre, rhythm, and lyrics, also affect people's perception and judgment of musical emotion. Music expresses emotion with essential elements such as pitch, length, strength, and timbre, and how quantifying music's emotional characteristics is the key to solving the music emotion recognition problem.

Xia et al. (2012) used a continuous emotion mental model and a regression prediction model to generate robot dance movements based on constructing emotional changes and music beat sequences. Schmidt et al. (2010) chose a continuous emotion model to link the dynamic content of music with the acoustic feature model. Then the established regression model can study the emotional changes in music over time. Bresin and Friberg (2011) invited 20 music experts to express emotions such as happiness, sadness, fear, and calm by the numerical combination of 7 characteristic quantities in the device and obtained the relationship between usual quantities and musical emotions. Yang et al. (2006) used a continuous emotion mental model with regression modeling to predict the emotional value of music. Then, two fuzzy classifiers were used to measure the emotional intensity to identify the emotional content of the music. Sarkar and Saha (2015) using convolutional neural networks to identify music models and compared them with commonly classifiers, e.g., BP neural networks.

In recent years, music emotion recognition has received extensive attention. With the combination of deep learning methods, emotion recognition has been dramatically improved. However, music emotion recognition is a long-term task that still needs continuous innovation and improvement. As the in-depth application in various fields, music emotion recognition has created incomparable value, promoting the development of other areas. Music emotion recognition can accurately carry out personalized music recommendations. Adjusting the music according to the emotional needs of individuals can well solve the problem of individual differences and make music retrieval methods more diversified. With the further application of musical emotion recognition in the medical field, psychological therapy with music has become an effective treatment method in recent years. At the same time, music emotion recognition also plays an essential role in dealing with brain nerve problems. Therefore, the research on musical emotion recognition is of great significance to the in-depth development of various fields.

Music emotion recognition can accurately carry out personalized music recommendations. Adjusting the music according to the emotional needs of individuals can well solve the problem of individual differences and make music retrieval methods more diversified. With the further application of musical emotion recognition in the medical field, psychological therapy with music has become an effective treatment method in recent years. At the same time, music emotion recognition also plays an essential role in dealing with brain nerve problems. Therefore, the research on musical emotion recognition is of great significance to the in-depth development of various fields.

## 2.2. The processing of MER

The existing MER algorithms are almost based on supervised learning. Therefore, establishing a learning library is necessary, i.e., music and related data, which is shown in Figure 1. Then, the music features are extracted by an emotional model, which forms feature vectors using dimensional reduction. This way, the music emotion recognition model is calculated by training the feature vectors and emotion labels. Finally, the performance feature is extracted from the unknown music, i.e., emotional test, and the classification result can be conducted by the trained recognition model.

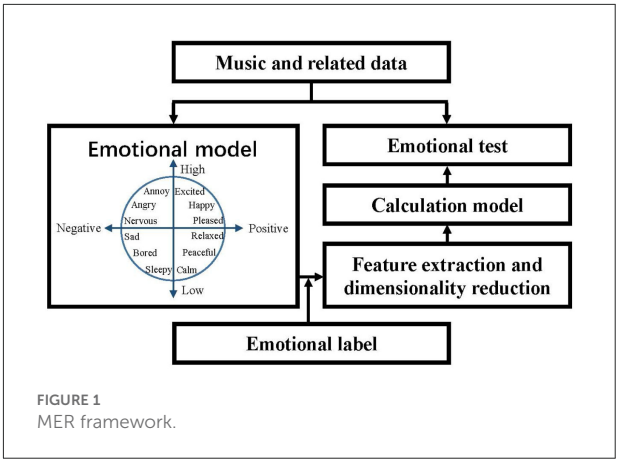



FIGURE 1
MER framework.

## 2.3. MER models

The analysis and identification of music emotion require using a music emotion model. The music emotion model can effectively solve the problem that emotion is difficult to quantify. The basis for analyzing and identifying music emotion is selecting a suitable model according to the characteristics. The music sentiment analysis model can generally be divided into three parts, namely the music feature model, music emotion model, and cognitive classification model. Among them, the music emotion model is the basis for the final classification, and its selection is essential. Classical music emotion models include Hevner model (Hevner, 1936), Thayer model (Thayer, 1990), TWC model (Tellegen et al., 1999), and PAD model (Russell, 1980).

In computer music emotion analysis, the Hevner model is a commonly used psychological model of music emotion. As in Figure 2A, Hevner first proposed it in 1936. Hevner categorizes emotional adjectives into eight categories: solemn, sad, dreamy, quiet, graceful, happy, excited, and powerful. Each type of adjective is further subdivided into several more detailed and extensive emotional adjectives, totaling 67 words. The model combines musicology and psychology, is richer in selecting emotional keywords, and has a good effect on the emotional identification of musical works. Based on the Hevner model, Farnsworth (1954) found that several Hevner adjective clusters were not accurate enough to describe emotions and then

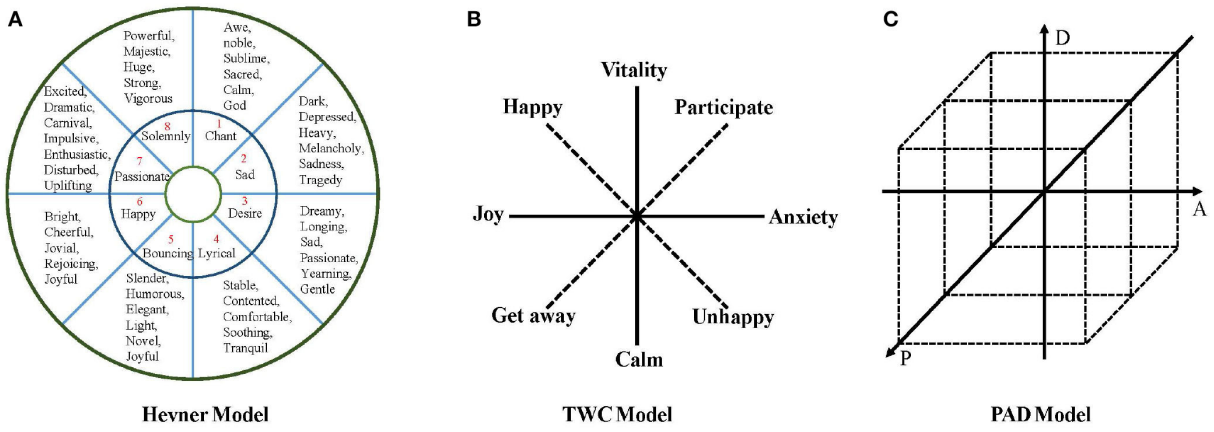



FIGURE 2
The common MER model. **(A)** Hevner model. **(B)** TWC model. **(C)** PAD model.

effectively updated the model using 50 Hevner adjectives. The internal consistency was higher than in the original model, and the category distinction was better. In 2003, Schubert (2003) updated the Hevner model emotional adjective table, and the final list contained 46 words, which were divided into nine groups in the emotional space.

Thayer (1990) believes that the two underlying dimensions of emotion are two independent dimensions of arousal: energy awakening and tension awakening. Thayer's emotional model is two-dimensional, with pressure as the abscissa and energy as the ordinate. According to the power from calm to vitality and the force from happiness to anxiety, the plane is divided into four extreme areas, i.e., vitality, anxiety, contentment, and depression. The Thayer model is proposed from the perspective of psychology. It is described by dimensional thinking and can establish a good relationship with acoustic characteristics. Therefore, it is suitable for emotional recognition of audio music such as MP3 and WAV. However, it is still necessary to map to discrete emotion representations in the specific emotion recognition process. As shown in Figure 2B, Tellegen et al. (1999) revised and supplemented the Thayer model and proposed a TWC model, which used 38 adjectives to describe emotions, and added a set of coordinate systems based on the original two-dimensional coordinates. The horizontal and vertical coordinates are Happy and involved. The TWC model not only retains the Thayer model's natural and smooth emotional transition but also dramatically enriches the descriptions of musical emotions.

The PAD (Russell, 1980) three-dimensional effective model is a dimensional measurement model first proposed by Mehrabian and Russell and is widely used in psychology, marketing, and product satisfaction research, which is shown in Figure 2C. The model divides emotions into three dimensions: pleasure, activation, and dominance, where P represents pleasure, representing the positive and negative characteristics of an individual's emotional state, A means the degree of activation, representing the level of individual neurophysiological activation, and D represents dominance, representing the individual Control over the situation and others. PAD model regards the degree of emotion as a spatial angle coordinate system and divides feeling into three dimensions, pleasure, activation, and dominance. Among them, pleasure represents the trend of an individual's emotional state, activation degree represents the neurophysiological activation level of the individual, and dominance refers to the degree of domination of the individual relative to the situation and others. The model describes subtle emotional intensity through different coordinate points. Compared with traditional emotion description methods, the characteristics and advantages of the PAD model are that it can distinguish emotional states such as anger and fear. The PAD model can describe a state of emotional intensity according to specific spatial coordinates and define emotional words continuously and subtly. The Thayer model has only two dimensions and lacks the richness of emotions, while the Hevner emotional ring model classifies and describes emotions in too much detail. Except for the 8-dimensional emotional ring, each category is subdivided into several sub-categories. This classification is too fine-tuned, and complex semantic models are unnecessary for simple music with a single emotion.

In addition to the emotion model listed in Figure 2, some studies utilize probability distributions (Yang and Chen, 2011), rankings (Fan et al., 2017), and antonym pairs (Liu et al., 2019) to express musical emotion. Probability distributions represent the emotion of a song as a probability distribution in the emotion space, which can alleviate the problem of emotional subjectivity. Ranking and recommending music works according to emotional intensity can reduce the

cognitive burden caused by continuous dynamic polarity and the inaccuracy caused by emotional subjectivity. Antonym pairs can make sentiment labels more objective. Figure 2 mainly introduces several classical music emotion models. The discrete model and multidimensional model are extensions of these methods.

### 2.3.1. Discrete model

In the discrete emotion model research, experts, and scholars assume that all types of emotions can be described by a specific subset of primary emotional states. For example, Ekman's team believes there are only six basic emotions (Ekman et al., 1987), including anger, disgust, fear, happiness, sadness, and surprise. The number of basic emotions is fixed and limited, and there is no appropriate way to reflect the complexity and variability of emotions. In the research process, scholars found that the discrete emotion model is used to quantify the emotion type and intensity. However, it has to face many limitations when solving practical application problems. For example, Gelbrich (2010) found that negative emotions of anger were associated with intentions to complain and negative word of mouth, whereas frustration and helplessness were not.

### 2.3.2. Multidimensional model

Another emotion model (Verma and Tiwary, 2017; Shu et al., 2018) provides a more effective way to quantify the types and intensity of emotions (Russell, 1980). The dimensional approach is based on the principle that human emotions can be represented as any point in two or three consecutive dimensions. One of the most prominent models is Russell's arousal valence emotion model. In this bipolar model, valence is represented by a positive or negative horizontal axis representing emotion. The vertical axis represents arousal, which describes the level of emotional activation. In this study, we used the arousal valence emotion model, which has been proven effective and reliable for identifying emotions during music listening, to convey human emotions (Salzman and Fusi, 2010).

## 2.4. Emotion feature

Music emotion classification is one of the tasks in the field of audio signal processing. The research work initially analyzed the emotional trend of audio from the perspective of audio signals, and solved the problem of music emotion classification by using common features such as timbre, dynamics and rhythm. Most of the methods used were some classic machine learning methods, which greatly improved the classification accuracy and efficiency. At present, there are two main methods for feature extraction. One is to train traditional machine learning (ML) models to predict emotions by extracting and using manual features. The other is doing it together through deep learning (DL) models.

- **Machine learning features** MER methods based on traditional machine learning can be divided into three categories, namely song-level classification MER, song-level dimension MER and dimension MEVD.
- **Song-level classification MER** Most song level MER classifications are based on mathematical-statistical models, of which K-nearest neighbors (KNN) and Support Vector Machine (SVM) are more common. In addition, decision tree (DT), random forest (RF), and Naive Bayes (NB) are also used for MER classification. MARSYAS (Li and Ogihara, 2003) extracted timbre, rhythm, and pitch features and input them into SVM for classification. However, the different emotion categories varied greatly. Yang et al. (2008a) evaluated various feature and fusion methods to improve classification accuracy through the late fusion of subtask merging. Liu et al. (2015) proposed Multi-emotional Similarity Preserving Embedding (ME-SPE), which combines ME-SPE with Calibration Label Ranking (CLR) to identify the emotion of music.
- **Song-level dimension MER** Commonly used MER regression algorithms include support vector regression (SVR), linear regression (LR), multiple linear regression (MLR), Gaussian process regression (GPR), and acoustic affective Gaussian (AEG). Yang et al. (2008b) take MER as a regression problem, extracting 114 audio features from existing toolkits and entering them into SVR. Malheiro et al. (2016) proposes three novel features of lyrics: slang presence, structural analysis, and semantic features. Wang et al. (2012) proposed a new generation model named Acoustic Emotional Gaussian (AEG) to identify emotional music, which shows higher accuracy than SVR and MLR models. Chen et al. (2017) discussed how to adjust AEG model to a personalized MER model with minimum user load. Chen et al. (2014a) explored the personalized MER using LR-based models. Fukayama and Goto (2016) input new acoustic signal into account using GPR. Soleymani et al. (2014) try static and dynamic methods by ML and DL methods.
- **Dimension MEVD** The idea of musical mood tracking was first proposed in Lu et al. (2005). Lu et al. (2005) divides the whole song into several independent segments and assigns an emotional label to each segment. The regression method can also be applied to MEVD but lacks time information. Schmidt et al. (2010) shows superior results in regression tasks of the spectral contrast features using multiple audio features. Xianyu et al. (2016) proposed DS-SVR method using two independent SVR models. One identifies mood changes between different songs, the other detects mood changes within a song, and then merges the results of the two SVRS into the final result.

## 2.5. Emotion evoking methods

In a broad sense, some stimuli can be classified as related to the organism's survival, while others are related to its growth. In this case, "emotional" stimuli are one such category, which creates a perception of understanding the situation and expresses the desired behavioral response accordingly. For example, galvanic skin response (GSR) evokes emotions by observing changes in electrical properties and autonomic nervous system activity (Das et al., 2016). In the current research on emotion recognition, various methods of emotion arousal (Widmann et al., 2018) have been developed. When we choose the proper arousal, it improves the accuracy of the data. Therefore, how to choose suitable and effective media to induce emotion experiments becomes the critical point of emotion recognition. By analyzing the source of emotional arousal materials, emotional arousal methods are usually divided into internal stimuli and external stimuli (Etkin et al., 2015; Alarcao and Fonseca, 2017). In the process of studying music to arouse emotions, experts and scholars also have their different views. Becker claimed that "music-induced emotional responses are not spontaneous" (Becker, 2001) and on the other hand, contrary to him, Peretz believed that "emotions are spontaneous and difficult to conceal" (Peretz, 2001). Noy's opinion is that "the emotions evoked by music are different from those generated in daily life and interpersonal communication" (Hunter and Schellenberg, 2010), and Peretz believes that "there is no theory that postulates this specificity of musical emotions" (Peretz, 2001).

## 2.6. Emotional information acquisition

In the experiments, it is not difficult to find that the results obtained by using multimedia emotional labels cannot be generalized to more interactive situations or everyday environments. To address this real-world problem, new research that uses interactive emotional stimuli to ensure the generality of Brain Computer Interfaces (BCI) results would be welcome (Pallavicini et al., 2018). Many experiments elicit emotions in different environments but don't use electroencephalography devices to record changes. Instead, other physiological indicators such as heart rate, electrical skin changes, and respiratory rate are collected. Conceptually, these paradigms could be helpful if replicated for EEG signal acquisition (Iacoviello et al., 2015).

Computer technology and emotion recognition algorithms are usually used when developing a human-computer emotional interaction system. Spontaneous physiological signals can be detected from a physiological perspective, allowing us to objectively and effectively analyze emotional states (Healey and Picard, 2005; Kreibig, 2010). When we want to know the activity of the cardiac autonomic nervous system corresponding to different emotional states, we can use the ECG to complete it (Agrafioti et al., 2011). Respiratory signals (RSP) can reveal a lot of information about emotions. For example, the frequency or depth of breathing is closely related to emotional changes (Zhang et al., 2017). However, during the experiment, the spontaneous physiological signals in the case will encounter problems such as no quantification standard and low classification accuracy when quantifying emotions (Fairclough, 2009; Nie et al., 2011). In processing automatic signals, EEG signals can provide a direct and comprehensive method for emotion recognition (Mauss and Robinson, 2009; Waugh et al., 2015).

## 2.7. EEG

EEG is a process of monitoring and recording information about the electrical activity of the human brain by generating spontaneous, rhythmic impulses from neurons in the brain. In neuroscience and psychology, EEG signals can describe the emotional state of the brain and human behavior, which can pick up subtle fluctuations in a person's emotional state. However, EEG signals are too weak to be recorded and are easily interfered by other physiological factors. In addition, to make the EEG signal non-linear, the original EEG signal needs to be preprocessed.

EEG has been widely used in clinical practice, and its characteristics can be summarized as follows:

- The EEG signal is feeble, with an amplitude around microvolts. Collecting EEG signals is easy to be interfered with by other signals. Therefore, the original EEG signal can be used for subsequent research after filtering, noise reduction, and further processing.
- The advantages of EEG signals are that they can directly reflect brain activity and change quickly. People's thinking, emotional changes, and other physiological factors can cause changes in EEG. An electroencephalogram differs from a typical electrocardiogram, a non-stationary and non-linear time series.

## 2.8. EEG datasets

With the rapid development of emotion recognition, a series of standardized emotion trigger databases have been established with corresponding emotion labels provided by psychologists. However, current research using music-based methods of emotional arousal is still limited without generally accepted standards. The researchers developed a number of mood databases on EEG. The benefit of open databases is that more standard methods can be used to verify the performance of emotional classification when different algorithms or

TABLE 1 A summary of datasets.

| Datasets | Conceptualization | Number of songs | Data type | Genres |
|---|---|---|---|---|
| MediaEval | Dimensional | 744 | MP3 | Rock, Pop, Soul, Blues |
| CAL500 | Categorical | 500 | MP3 | - |
| CAL500exp | Categorical | 3,223 | MP3 | - |
| AMG 1608 | Dimensional | 1,608 | MAV | Rock, Metal, Country, Jazz |
| DEAM | Dimensional | 1,802 | MP3 | Rock, Pop, Electronic |
| Soundtracks | Both | 360 | MP3 | Rap, R&B, Electronic |
| Emotify | Categorical | 400 | MP3 | Rock, Classical, Pop, Electronic |
| DEAP | Dimensional | 40 | CSV | - |

classification models are used in experiments. Due to music copyright restrictions, some MER researchers use self-built and unpublished data sets. Table 1 lists some common data sets that are commonly used.

- **MediaEval**[1] The data set contains randomly extracted 45-s musical snippets of complete songs. The 45 s excerpt (clip) is also annotated for the full length clip using a 9 point arousal and potency level. A set of features extracted by openSMILE are also available for data. The dataset initially had 1,000 creative general License songs annotated continuously (dynamically) on the wake and potency dimensions. We found some redundant songs and fixed some problems, which reduced the number of songs to 744.
- **CAL500**[2] is a data set designed to evaluate music information retrieval systems. It consists of 502 songs from western pop music. Audio is represented as a time series of the first 13 Meir frequency cepstrum coefficients (and their first and second derivatives), extracted by sliding a 12 ms semi-overlapping short time window across the waveform of each song. Each song is annotated by at least 3 people and contains 135 music-related concepts covering six semantic categories.
- **CAL500exp**[3] is a rich version of the CAL500 Music Information Retrieval dataset. CAL500exp is designed to facilitate automatic tagging of music on smaller time scales. The dataset consisted of the same songs, divided into 3,223 acoustically homogenous segments ranging in length from 3 to 16 s. Tags are annotated at the segment level rather than the track level. Notes are obtained from annotators with a strong musical background.
- **AMG 1608**[4] It's a data set for emotional analysis of music. It contains frame-level acoustic features extracted from 1,608 30-s music clips and corresponding valency awakening (VA) notes provided by 665 subjects. The dataset consists of two parts: Campus subset. It is a subset of 240 songs annotated by 22 subjects from National Taiwan University and Academia Sinica. Amazon Mechanical Turk (AMT) subset. This subset contained 643 subjects using annotations for all 1,608 songs provided by the AMT each song received a total of 15 emotional annotations from each subject in this subset.
- **DEAM**[5] The DEAM dataset consists of 1,802 fragments and complete songs annotated consecutively (per second) and for the entire song with valence and wake values. This dataset is a collection of data sets for the "Emotions in Music" task from the 2013–2015 MediaEval benchmark campaign, in addition to the original tags. The purpose of the emotional annotation collection is to detect the emotions expressed by music and musician composition content.
- **Soundtracks**[6] The dataset provides complete audio files, documents, and behavioral ratings for the sharing of stimulating material for academic research. Audio files have been compressed into MP3 files using medium quality compression.
- **Emotify**[7] The dataset contains extracts (1 min long) from 400 songs in 4 genres (rock, classical, pop, electronic). Notes were collected using the GEMS Scale (Geneva Emotional Music Scale). Each participant could choose up to three items from the scale (the strong emotions he felt while listening to the song).
- **DEAP**[8] The most commonly used database for EEG emotion recognition is DEAP. The database was created

1 http://www.multimediaeval.org/mediaeval2015/emotioninmusic2015/

2 https://paperswithcode.com/dataset/cal500

3 https://paperswithcode.com/dataset/cal500exp

4 http://amg1608.blogspot.com/

5 https://cvml.unige.ch/databases/DEAM/

6 https://osf.io/p6vkg/

7 http://www2.projects.science.uu.nl/memotion/emotifydata/

8 http://www.eecs.qmul.ac.uk/mmv/datasets/deap/

by a consortium of four universities in the UK, the Netherlands and Sweden. The researchers collected physiological signals from 32 participants as they watched a 1-min music video and recorded the participants' SAM scale before and after the experiment (Koelstra et al., 2011). SAM scale was used to collect self-assessment of emotion in valence, arousal, dominance, and liking dimensions.

## 3. EEG-based music emotion recognition

With the emergence of advanced analytical tools, including mathematical models and brain-machine interface apparatus, the connectivity patterns between music and mood have attracted ever-lasting attention in the last decades. EEG plays a pivotal role in exploring the human body. EEG-based music emotion recognition combines the methods commonly used in processing EEG data and application in music emotion recognition. Other detection methods often describe the mental and physical changes of the human body by detecting the physiological response of the human body. An electrocardiogram (ECG) records changes in the heart's electrical activity during each cardiac cycle. Electromyogram (EMG) uses sensing electrodes to capture electrical signals from muscle impulses beneath the skin. Galvanic skin response (GSR) describes the mental state of the human body through the secretion of sweat glands. Electrooculography (EOG) records the electrical potential between the cornea and the retina of human eyes. Other detections such as speech, facial expression, heart rate, and respiratory rate are commonly used to profile emotional states. Of all the methods, EEG is the most used approach to physiological signals. EEG works by placing probe electrodes on a particular scalp and monitors the most complex response center of the human body indirectly, the brain. The most complex signals can be obtained, meaning that EEG signals contain valuable information. EEG-based music-induced emotion recognition is a method of great essence.

### 3.1. Data collection

Many factors have to be taken into consideration when acquiring MER data. This section presents an overview of common restrictions and unwritten rules followed by a mass of investigators. First of all, most subjects were between 20 and 30 years of age, and the ratio of men to women remained balanced. The subjects were also required to be free of brain diseases and mental illness. In music-related emotion studies, it is also necessary to exclude whether the subjects have received any professional training in music. Moreover, liking or familiarity with the music must also be considered in particular research tasks (Hadjidimitriou and Hadjileontiadis, 2012; Thammasan et al., 2017b; Naser and Saha, 2021). In addition, subjects were usually required to rest before data acquisition. Data acquisition is generally carried out in a dark and quiet environment, excluding interference from the external environment. A certain amount of participants, mainly from 9 to 27 in total, was usually allowed in a study. In another aspect, The choice of music was relatively personalized and varied considerably in different studies. An endeavor (Sangnark et al., 2021) have been made toward the perceptual distinctions from listening to two music types: music without lyrics (Melody) and music with lyrics (Song), which denoted that music of different forms could bring different emotional information within. Increasing numbers of researchers have been willing to using standard music databases for study as MER continues to develop, which eliminates the subjective operation of music selection and reduces the inconsistency of outcomes of different studies. Moreover, Some databases contain EEG data from subjects, which saves the experimenters from having to collect EEG data.

There are two main annotation methods for music emotion label annotation. The first one is Expert-based annotation. A given piece of music has been emotionally labeled, and the people's feelings are unique and subjective. The same song can bring different feelings to different people, which triggers the appearance of another criterion called Subject-based annotation. Subjects were asked to rate the degree of musical valence, arousal, and dominance. Target music selected from a set of music candidates shares a standard emotional label across all subjects, and the difference between different songs is significant enough simultaneously.

In different studies, the classification criteria of emotions vary according to the various tasks that researchers are interested in or the difficulty of the tasks. The most adopted is the two-dimensional emotion model. Self-evaluation reports, such as the Self-Assessment Manikin (SAM), are commonly used for evaluating a person's mental state in bipolar dimensions. The subjects were required to rate valence and arousal on a scale of 1 to 9. The higher the score, the stronger it is. There are usually four types: HVHA, HVLA, LVHA, and LVLA, namely happy, angry, relaxed, and sad. Some studies divide a single dimension into three categories: a score below three is considered low, and a score over six is considered high. In classification tasks, accuracy and F-score are generally used as model indicators. Unlike classification tasks, Sam Score can also be predicted by regression and other methods, while Root Mean Square Error (RMSE) is used as an indicator to evaluate the quality of the model.

### 3.2. Data preprocessing

EEG signal feature extraction has always been the key to music emotion recognition. Data preprocessing can transform

TABLE 2 A summary of EEG data.

| Frequency band | Frequency range (Hz) | Describe |
|---|---|---|
| $\alpha$ | 8–13 | Appear in a relaxed state |
| $\beta$ | 14–30 | Appear in a state of tension |
| $\delta$ | 1–3 | Appear in a state of extreme exertion |
| $\gamma$ | 31–50 | Appear in a state of concentration |
| $\theta$ | 4–7 | Occurs in people with mental illness |

complex EEG signals into data structures that are easier to understand and use. Different experiments used different pretreatment processes, which would affect the subsequent results. In general, according to the purpose of the pretreatment, the processing process can be divided into noise filtering and signal feature extraction.

Raw EEGs typically incorporate noise generated by eye movements, facial muscle movements, heart rate, and breathing. ICA (Independent Component Analysis) and Blind Source Separation commonly remove the above noise before processing EEG data. Another option is to filter out EEG bands of no interest. In the past few decades, EEGs have generally been classified into the following bands with different frequencies, i.e., $\delta$ (1–3 Hz), $\theta$ (4–7 Hz), $\alpha$ (8–13 Hz), $\beta$ (14–30 Hz), and $\gamma$ (31–50 Hz), as shown in Table 2. $\delta$ band is related to Deep sleep. $\beta$ and $\gamma$ bands are related to Relaxed. In many studies, $\delta$ and $\gamma$ bands are usually excluded to reduce EMG and power line artifacts. EEG sequence is a typical time-varying signal, and a range of signal analysis techniques is proven to extract information from that effectively. The signal feature extraction method varies in different studies. Based on the above partial noise filtering methods, signal feature extraction can be divided into three parts: time, frequency, and time-frequency domain.

- **Time domain** Independent Component Analysis attempts to decompose a multivariate signal into independent non-Gaussian signals by maximizing the statistical independence of the estimated components. As mentioned above, it is applied to lessen the noise caused mainly by eye movement. Fractal dimensions, a widely used measure of complexity and irregularity, characterize a broad spectrum of objects, especially human physiology. The Higuchi algorithm is a selectable approach to calculating the FD value of EEG data and could give an outcome close to the theoretical value. Fluctuation Analysis is a universal analytic method of various fields for time-varying signals.

  Time series can generally decompose into trend, periodic, and random terms. Fluctuation analysis is a method used to judge whether the noise items in the time series have positive or negative self-correlation and whether the self-correlation is a long-range correlation. The detrended Fluctuation analysis is an improvement to the detrended analysis. The purpose is to eliminate the influence of the trend item on the detrended analysis.

  The entropy of the signal is a dimensionless indicator used to characterize the complexity of the signal sequence. The larger the entropy value, the greater the signal complexity. It is a big family containing approximate entropy, sample entropy, multiscale entropy, etc. Sample entropy is widely used in EEG signal processing because the calculation of sample entropy does not depend on data length and has a better consistency.
- **Frequency domain** The Frequency Domain refers to the analytic space in which mathematical functions or signals are conveyed in terms of frequency. Fourier transform and its variant, such as the short-time Fourier transform, are the most common paradigms to convert the time function into a set of sine waves representing different frequencies. Power spectral density has been adopted based on a fast Fourier transform with a window to obtain EEG signal frequency information.
- **Time-Frequency domain** Wavelet transform is a feature extraction algorithm combining the time and frequency domains. The wavelet decomposes the signal into different approximation and detail levels according to a specific frequency range while preserving the time information of the signal. The discrete wavelet transforms the signal into coarse approximations and details associated with low-pass and high-pass filters.

  Empirical Mode Decomposition decomposes the signal according to the time scale characteristics of the data itself. It does not need to set any basis function in advance, which is different from the Fourier decomposition methods based on a priori harmonic and wavelet basis functions. The EMD method can theoretically be applied to any signal decomposition, so it has obvious advantages in processing the non-stationary and non-linear signals with a high signal-to-noise ratio.

## 3.3. Model collection

Sequentially, feature extraction is closely followed by model training. In this review, various relevant algorithms have been collected according to the taxonomy of machine learning. In most cases, MER is a label-available task. Commonly used classical classifiers or regressors, including General Linear Regression(GLR), Support Vector Machine(SVM), and Random Forest(RF), had yielded precise accuracy. SVM was the most used tool among all traditional algorithms and achieved remarkable prediction performance in different studies (Sourina et al., 2012; Thammasan et al., 2016c; Bo et al., 2019).

Radial basis function (RBF) kernel-based SVM can non-linearly map features onto a higher dimension space. It has

been widely approved that kernel-based SVM has ensured better representation for samples and robustness. Thus, kernel-based SVM has been popular in MER (Thammasan et al., 2017a; Avramidis et al., 2021; Luo et al., 2022). Other machine learning methods like random forest (Pisipati and Nandy, 2021) and KNN (Bhatti et al., 2016) were also developed for MER. Fatemeh introduced a novel approach named Fuzzy Parallel Cascades outperformed the CNN-LSTM model.

Recently, with the trending usage of neural networks, deep learning-based algorithms, e.g., Multi-Layer Perception (MLP), Long Short-Term Memory (LSTM), have been referred to as better substitutes for traditional machine learning methods. In Rahman et al. (2020), MLP selected handcraft features as input. Convolution Neuron Network (CNN) is well-suited to processing image data with the bias of transitional invariance. Thus, the power spectrogram generated by the EEG frequency signal was adopted reasonably (Er et al., 2021; Liu et al., 2022). LSTM was born for time series data since it can keep track of arbitrary long-term dependencies in the input sequences. Luo et al. (2022) utilizes LSTM for sequence generation in his work. Additionally, deep learning algorithms overcome a troublesome and controversial problem, namely feature extraction, which liberates researchers from handcrafted feature selection to a certain extent. Panayu (Keelawat et al., 2019) used almost raw EEG signal as input into a 5-layer CNN without feature extraction and revealed supremacy compared to SVM. Sheykhivand et al. (2020) uses a fusion network of CNN and LSTM had been developed. Deep Belief Networks (Thammasan et al., 2016a) and Stacked Sparse Auto-Encoder (Li and Zheng, 2021) invented another path to solving MER.

## 3.4. EEG emotion recognition method

EEG is a physiological detection method mainly used to reflect brain activity. It has rich information on mental activity and is widely used by doctors and scientists to study brain function and diagnose neurological diseases. However, EEG signals generate a huge amount of data that can be difficult to analyze by observation during a study. Therefore, how to efficiently extract the required information from EEG signals has become the most urgent problem to be solved.

### 3.4.1. EGG feature selection

High-dimensional EEG features may contain a large number of unnecessary features. In order to improve the classification performance, it is necessary to filter out the EEG features related to emotion before modeling the emotion classifier. Feature selection can be simply divided into supervised and unsupervised feature selection according to whether label information is used or not.

### 3.4.2. Supervised feature selection

In supervised learning, Linear Discriminant Analysis (LDA) is a dimensionality reduction method for feature selection. The main idea of LDA is to find the appropriate projection direction according to the discriminative information of the class (Koelstra et al., 2011). The projection direction can be determined when the minimum within-class variance and the maximum within-class variance are simultaneously achieved. On the other hand, Maximum Relevance Minimum Redundancy (mRMR) is also an algorithm commonly used by supervised learning to obtain EEG features. It uses MI algorithm to select EEG features that can satisfy both maximum correlation and minimum redundancy (Atkinson and Campos, 2016).

### 3.4.3. Unsupervised feature selection

Principal Component Analysis (PCA) is a commonly used method for unsupervised feature selection. By projecting the samples into a low-dimensional space, a series of linearly independent principal components are obtained. Principal component analysis preserves as much data information as possible by minimizing reconstruction errors during feature selection. K Nearest Neighbor (KNN) is also a non-parametric statistical method for classification and regression. Its core strategy is to identify k samples closest to the unknown sample point, and determine the classification information of the unknown samples from the majority of k samples (Hwang and Wen, 1998; Zhang et al., 2006). The selection of k value generally depends on the data used. Increasing k value in the classification task can suppress the interference of noise, but it will also blur the boundary between categories. Its advantages are easy to implement, small sample size and low computational complexity. It is relatively simple to implement and robust to noisy training data.

### 3.4.4. EEG signal preprocessing and analysis

Since the architecture of the EEG headset is not exactly the same, and the cost of the equipment is also different, it is necessary to set different EEG equipment during the experiment. The main difference between these devices is the duration of time when the brainwave signals are collected, and the type of electrodes is also a factor (Chen et al., 2014b; Casson, 2019). Due to the sensitivity of the electrodes, the experiment required participants to remain still as they began collecting brainwave signals.

Over the past few years, there has been a lot of interest in using EEG signals to observe mood changes. In order to efficiently use EEG signals for emotion identification, the following steps are performed:

- Participants had to be tested for musical stimulation.

- The participants' brain voltage changes were observed and recorded throughout the experiment.
- Remove noise, shadow, and other interference items.
- The experimental results were analyzed and the eigenvalues were extracted.
- Training data, analyzing, and interpreting raw signals.

EEG signals have better temporal resolution than spatial resolution. In the process of music arousing emotion, the time change of electroencephalogram can be observed, including its amplitude and the change with time. In the process of EEG signal acquisition, due to the influence of environment and equipment, many noises are usually introduced, such as power frequency noise, ECG, EOG, and EMG caused by physiological signals of human body. In order to obtain relatively pure EEG data, it is necessary to preprocess the original EEG signal. Power frequency noise is mainly caused by the power supply of the device itself, and its frequency is 50 Hz. In the experimental process, the usual way is to use a 50Hz notch filter to work at 50 Hz frequency to remove the power frequency noise. In addition, the ECG is generated by the rhythmic operation of the heart, and the amplitude is large. However, the heart is far away from the electrode position, and the ECG signal is greatly weakened when it reaches the scalp. Therefore, this part is often ignored when we preprocess EEG signals.

The most widely used EEG analysis methods include four categories: time domain, frequency domain, time frequency domain, and non-linear method (Babloyantz et al., 1985; Phneah and Nisar, 2017).

(1) There are two main methods of time-domain EEG analysis: linear prediction and component analysis. Linear prediction is a linear combination of past output values with current and past input values. Component analysis is the mapping of datasets to feature sets for unsupervised learning.

(2) Spectrum analysis is to obtain frequency domain information in the EEG waveform through statistics and Fourier transform. Among them, power spectrum analysis is the most commonly used method.

## 3.5. Results and analysis

Works in MER were extremely hard to compare with each other owing to the processes of data acquisition and feature elicitation. In this part, we present the contribution of relevant research and shed light on the distinction between closely related jobs. Most studies focus on the improvement of prediction. Some others pay additional attention to feature selection for saving computational complexity or seek profound mechanisms inside. Tables 3, 4 present an overview of the scope and contributions of relevant works.

To summarize the works mentioned above, this article divides them into four categories regarding their focus and contributions.

### 3.5.1. Performance comparison

Many works have listed a series of algorithms. Therefore, differences can be quickly concluded by their achievements in prediction results. Diverse versions of SVMs have been investigated in Lin et al. (2009). In Bhatti et al. (2016), MLP had been proven to outperform KNN and SVM in MER. A comparison to random forest and multilayer perceptron shows slight superiority of hyper pipes (Zainab and Majid, 2021). In addition, model selection is time-dependent. In the early years of MER, shallow machine learning methods, e.g., SVM, kernel SVM, KNN, and RF, were widely adopted. As time went by, algorithms went more profound and more complex, like CNN (Salama et al., 2018; Keelawat et al., 2019; Er et al., 2021), LSTM (Sheykhivand et al., 2020), and SSAE (Li and Zheng, 2021), and promising performances have been witnessed.

### 3.5.2. Features selection preference

Generally, linear features like Mean, Standard Deviation are considered insufficient for precise prediction. Therefore, non-linear hand-crafted features showed up in the traditional machine learning methods. Several taxonomies of feature elicitation patterns have been broadly endorsed. Most works set a high value on FD in the time domain. The entropy of many forms of EEG signal was also available in predicting emotion. Undoubtedly, PSD is the most used method in the frequency domain and Wavelet Transform is pivotal in the time-frequency domain. However, earthshaking change have taken place as deep learning based algorithms began to flourish in MER, which diminished the importance of feature selection largely.

### 3.5.3. Accessory effective factors

Many works also took liking or familiarity into consideration because whether a person enjoys or gets familiar with a specific song would have an impact on the correctness of prediction to a great extent. Such investigation has been carried out by Naser (Naser and Saha, 2021), who found that testing in low-liking music could perform better than high-like. In Thammasan et al. (2017b), the same phenomenon can be observed. Familiarity effects in MER have been proven to impair the capacity of the model to predict. Both observations indicate that liking or familiarity would impact emotion recognition.

### 3.5.4. Multi-modal fusion

Despite a series of efforts have been exerted in algorithm melioration, another promising path had been opened up for

TABLE 3 Details of music emotion recognition algorithms based traditional method.

| Method | Dataset | Features | Classifier/ regressor | Performance |
|---|---|---|---|---|
| Avramidis et al. (2021) | DEAP | PSD, HFD, MFD, MADFA | RBF-SVM | Accuracy of 67% in Binary classification of arousal. |
| Hasanzadeh et al. (2021) | 15 recruited listened to 7 songs | Spectrograms from Morlet wavelet transform | Fuzzy Parallel Cascades | 2 types regression of Valence with RMSE 0.089. |
| Thammasan et al. (2016c) | 15 recruited listened to 16 songs selected from MIDI | HFD | SVM | 3% performance increase over the non-filtered. |
| Zainab and Majid (2021) | 27 recruited listened to bilingual audio music of five genres | PSD, HFD, Hjorth Parameters. A series of linear measures of time domain | Hyper Pipes | Accuracy of 83.95% in quaternary classification.. |
| Thammasan et al. (2016b) | 12 recruited listened to 16 songs selected from MIDI | FD for EEG and handcraft feature for music | SVM | MCC of 84.17 and 90.25% in binary classification of arousal and valence, respectively. |
| Naser and Saha (2021) | DEAP | Wavelet transform, functional connectivity, graph-theory based features | RBF-SVM | Accuracy of arousal, valence, and dominance were 22.50, 14.87, and 19.44% above the empirical chance-level, respectively. |
| Thammasan et al. (2017b) | DEAP | PSD, HFD | Kernel SVM, MLP, Decision Tree | An average of 5% classification improvement of Unfamiliar set above familiar set in three methods. |
| Shahabi and Moghimi (2016) | 19 recruited listened to six classical music excerpts | Connectivity matrices | SVM | Joyful vs. neutral, joyful vs. melancholic and familiar vs. unfamiliar trials reach accuracy of 93.7, 80.43, and 83.04%, respectively. |
| Lin et al. (2009) | 26 recruited | PSD | One-against-one scheme SVM | Accuracy of 92.57% in quaternary classification. |
| Bhatti et al. (2016) | 30 recruited listened to 4 genres of music | Latency to Amplitude Ratio, PSD, Wavelet transform | MLP, KNN, SVM | Accuracy of 78.11% (MLP) in quaternary classification. |

better performance, that is multi-modal fusion. Multi-modal is a trend in future works owing to the shortness of EEG signals, which is too complex to understand for a machine, speaking of only a few data available for training in a very single experiment. Data from other pathways are welcome for affective recognition that would improve performance. Like Nattapong (Thammasan et al., 2016b) did in his work, he fused features from both EEG and music and reached higher performance.

### 3.5.5. Applications

Music emotional recognition is an interdisciplinary field that spans medical psychology and computer science. Human's affection could be precisely detected by machine. By this way, it's certified that the inner alteration in mood could be effected by the external music. The quantified emotional score could be viewed as a criterion in music therapy. A bulk of jobs (Shahabi and Moghimi, 2016; Thammasan et al., 2017b) centered on the brain communication by analyzing connectivity among disparate EEG channels when exposed to music stimulation, which attempted to uncover the function of areas of brain. Interestingly, A reverse research have been conducted that generated emotion-related music by feeding neuron network with EEG signal (Li and Zheng, 2021).

## 4. Discussion and future work

Music emotion recognition is an interdisciplinary subject with a wide range of applications. Many researchers have carried on the in-depth discussion on it and have made some achievements. However, music emotion recognition is still in the process of development and should be further explored. The subjectivity of music emotion has a significant influence on the expression of emotion. In analyzing music emotion, studying its characteristics is a basic premise for emotional calculation. It is not trivial for researchers to obtain objective and accurate emotional expression and establish effective and high-quality teaching resources because of the subjective nature of the expression and emotion inspired by music. Recently, music emotion recognition-based EEG is an emerging topic in affective computing and social signal processing, gaining more attention.

TABLE 4 Details of music emotion recognition algorithms based deep learning method.

| Method | Dataset | Features | Classifier/ regressor | Performance |
|---|---|---|---|---|
| Keelawat et al. (2019) | 12 recruited listened to 16 songs selected from MIDI | Segmented EEG | CNN | Accuracy of 78.36 and 83.67% in binary classification of arousal and valence, respectively. |
| Er et al. (2021) | Nine recruited listened to 16 audio tracks | Power spectrogram | Pretrained VGG16 | Accuracy of 73.28% in quaternary classification. |
| Thammasan et al. (2016a) | 15 recruited listened to 16 songs selected from MIDI | HFD, PSD, Discrete Wavelet Transform | Deep Belief Networks | Accuracy of 81.98% in binary classification of arousal and valence. |
| Rahman et al. (2020) | 24 recruited listened to Twelve songs | DFA, Approximate Entropy, Fuzzy Entropy, Shannon's Entropy, Permutation Entropy, Hjorth Parameters, Hurst Exponent | Neuron Network | 3 emotion scales (Depressing vs. Exciting and Sad vs. Happy and Irritating vs. Soothing). |
| Liu et al. (2022) | 15 recruited listened to 13 music excerpts | Power spectrogram | Xception | Accuracy of 76.84% in HVHA vs. LVLA |
| Luo et al. (2022) | DEAP | PSD | RBF-SVM, LSTM | A SAM score of 6.17(high) and 4.76(low) in continuous valance scale, that is close to 6.98 and 4.36 evaluated in music database. |
| Hsu et al. (2018) | IADS | Segmented EEG | Neuron Network | MSE of 1.865 in 2D continuous SAM score. |
| Sheykhivand et al. (2020) | 16 recruited listened to ten music excerpts | Segmented EEG | CNN, LSTM | Accuracy of 76.84% in HVHA vs. LVLA. |
| Li and Zheng (2021) | 21 recruited listened to 15 music excerpts | Segmented EEG | Stacked Sparse Auto-Encoder | Accuracy of 59.5% and 66.8% in binary classification of arousal and valence, respectively. |
| Salama et al. (2018) | DEAP | Segmented EEG | 3D CNN | Accuracy of 88.49% and 87.44% in binary classification of arousal and valence, respectively. |

Development in this field is to meet the demand of People's Daily lives, such as in human-computer interaction, the machine can communicate and understand humans.

In the future, multimodal-signals-based MER can be studied. For example, music emotion analysis based on physiological and EEG signals can obtain more comprehensive and practical information by analyzing the correlation between EEG signals and physiological signals. At the same time, signal conversion based on music sentiment analysis is also a promising research direction. The mutual conversion between EEG signals and physiological signals is realized through music emotion analysis.

## 5. Conclusion

Music plays an essential role in People's Daily life. It is necessary to relax their emotions and regulate their physical and mental health. It can affect people's emotions, intelligence, and psychology. As the research progressed, the researchers began to explore what happens to the body during listening to music and the association between different music and induced emotions. EEG is one of the physiological signals of the human body, which contains rich physiological and disease information and is widely used in clinical practice.

## Author contributions

YW, JW, and ZY performed the literature search. XC provided facilities and support. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the Ningbo Public Welfare Technology Plan Project (2019C50081), the Shenzhen Governmental Science and Technology for Basic Research Program Grant (JCYJ20180507182446643), the Shenzhen Governmental Science and Technology for Shenzhen-Hong Kong Joint Program Grant (SGDX20201103095406023), the Guangdong Province Science and Technology Program

Grant (2018A050501010), the Natural Science Foundation of Fujian Province of China (No. 2019J01573), the Shenzhen Science and Technology for International Cooperation Research Grant (GJHZ20190821155201661), and SIAT Innovation Program for Excellent Young Researchers (E1G062, E2G037).

## 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

Agrafioti, F., Hatzinakos, D., and Anderson, A. K. (2011). ECG pattern analysis for emotion detection. *IEEE Trans. Affect. Comput.* 3, 102–115. doi: 10.1109/T-AFFC.2011.28

Alarcao, S. M., and Fonseca, M. J. (2017). Emotions recognition using EEG signals: a survey. *IEEE Trans. Affect. Comput.* 10, 374–393. doi: 10.1109/TAFFC.2017.2714671

Atkinson, J., and Campos, D. (2016). Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Expert Syst. Appl.* 47, 35–41. doi: 10.1016/j.eswa.2015.10.049

Avramidis, K., Zlatintsi, A., Garoufis, C., and Maragos, P. (2021). "Multiscale fractal analysis on EEG signals for music-induced emotion recognition," in *2021 29th European Signal Processing Conference (EUSIPCO)* (Dublin: IEEE), 1316–1320. doi: 10.23919/EUSIPCO54536.2021.9616140

Babloyantz, A., Salazar, J., and Nicolis, C. (1985). Evidence of chaotic dynamics of brain activity during the sleep cycle. *Phys. Lett. A* 111, 152–156. doi: 10.1016/0375-9601(85)90444-X

Becker J. (2001). "Anthropological perspectives on music and emotion," in *Music and Emotion: Theory and Research*, eds P. N. Juslin and J. A. Sloboda (New York, NY: Oxford University Press), 135–160.

Bhatti, A. M., Majid, M., Anwar, S. M., and Khan, B. (2016). Human emotion recognition and analysis in response to audio music using brain signals. *Comput. Hum. Behav.* 65, 267–275. doi: 10.1016/j.chb.2016.08.029

Bo, H., Ma, L., Liu, Q., Xu, R., and Li, H. (2019). Music-evoked emotion recognition based on cognitive principles inspired EEG temporal and spectral features. *Int. J. Mach. Learn. Cybernet.* 10, 2439–2448. doi: 10.1007/s13042-018-0880-z

Bodner, M., Muftuler, L. T., Nalcioglu, O., and Shaw, G. L. (2001). FMRI study relevant to the mozart effect: brain areas involved in spatial-temporal reasoning. *Neurol. Res.* 23, 683–690. doi: 10.1179/016164101101199108

Bresin, R., and Friberg, A. (2011). Emotion rendering in music: range and characteristic values of seven musical variables. *Cortex* 47, 1068–1081. doi: 10.1016/j.cortex.2011.05.009

Cacioppo, J. T., Freberg, L., and Cacioppo, S. (2021). *Discovering Psychology: The Science of Mind*. Boston, MA: Cengage Learning.

Casson, A. J. (2019). Wearable EEG and beyond. *Biomed. Eng. Lett.* 9, 53–71. doi: 10.1007/s13534-018-00093-6

Chen, Y.-A., Wang, J.-C., Yang, Y.-H., and Chen, H. (2014a). "Linear regression-based adaptation of music emotion recognition models for personalization," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Florence: IEEE), 2149–2153. doi: 10.1109/ICASSP.2014.6853979

Chen, Y.-A., Wang, J.-C., Yang, Y.-H., and Chen, H. H. (2017). Component tying for mixture model adaptation in personalization of music emotion recognition. *IEEE/ACM Trans. Audio Speech Lang. Process.* 25, 1409–1420. doi: 10.1109/TASLP.2017.2693565

Chen, Y.-H., De Beeck, M. O., Vanderheyden, L., Carrette, E., Mihajlović, V., Vanstreels, K., et al. (2014b). Soft, comfortable polymer dry electrodes for high quality ECG and EEG recording. *Sensors* 14, 23758–23780. doi: 10.3390/s141223758

Cloitre, M., Khan, C., Mackintosh, M.-A., Garvert, D. W., Henn-Haase, C. M., Falvey, E. C., et al. (2019). Emotion regulation mediates the relationship between aces and physical and mental health. *Psychol. Trauma Theory Res. Pract. Policy* 11, 82. doi: 10.1037/tra0000374

Das, P., Khasnobish, A., and Tibarewala, D. (2016). "Emotion recognition employing ECG and GSR signals as markers of ANS," in *2016 Conference on Advances in Signal Processing (CASP)* (Pune: IEEE), 37–42. doi: 10.1109/CASP.2016.7746134

Dickerson, K. L., Skeem, J. L., Montoya, L., and Quas, J. A. (2020). Using positive emotion training with maltreated youths to reduce anger bias and physical aggression. *Clin. Psychol. Sci.* 8, 773–787. doi: 10.1177/2167702620902118

Ekman, P., Friesen, W. V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *J. Pers. Soc. Psychol.* 53, 712. doi: 10.1037/0022-3514.53.4.712

Er, M. B., Çiğ, H., and Aydilek, İ. B. (2021). A new approach to recognition of human emotions using brain signals and music stimuli. *Appl. Acoust.* 175, 107840. doi: 10.1016/j.apacoust.2020.107840

Etkin, A., Büchel, C., and Gross, J. J. (2015). The neural bases of emotion regulation. *Nat. Rev. Neurosci.* 16, 693–700. doi: 10.1038/nrn4044

Fairclough, S. H. (2009). Fundamentals of physiological computing. *Interact. Comput.* 21, 133–145. doi: 10.1016/j.intcom.2008.10.011

Fan, J., Tatar, K., Thorogood, M., and Pasquier, P. (2017). "Ranking-based emotion recognition for experimental music," in *ISMIR* (Suzhou: International Society for Music Information Retrieval), 368–375.

Farnsworth, P. R. (1954). A study of the Hevner adjective list. *J. Aesthet. Art Crit.* 13, 97–103. doi: 10.1111/1540_6245.jaac13.1.0097

Fukayama, S., and Goto, M. (2016). "Music emotion recognition with adaptive aggregation of Gaussian process regressors," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Shanghai: IEEE), 71–75. doi: 10.1109/ICASSP.2016.7471639

Gelbrich, K. (2010). Anger, frustration, and helplessness after service failure: coping strategies and effective informational support. *J. Acad. Market. Sci.* 38, 567–585. doi: 10.1007/s11747-009-0169-6

Hadjidimitriou, S. K., and Hadjileontiadis, L. J. (2012). Toward an EEG-based recognition of music liking using time-frequency analysis. *IEEE Trans. Biomed. Eng.* 59, 3498–3510. doi: 10.1109/TBME.2012.2217495

Hasanzadeh, F., Annabestani, M., and Moghimi, S. (2021). Continuous emotion recognition during music listening using EEG signals: a fuzzy parallel cascades model. *Appl. Soft Comput.* 101, 107028. doi: 10.1016/j.asoc.2020.107028

Healey, J. A., and Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transport Syst.* 6, 156–166. doi: 10.1109/TITS.2005.848368

Hevner, K. (1936). Experimental studies of the elements of expression in music. *Am. J. Psychol.* 48, 246–268. doi: 10.2307/1415746

Hsu, J.-L., Zhen, Y.-L., Lin, T.-C., and Chiu, Y.-S. (2018). Affective content analysis of music emotion through EEG. *Multimedia Syst.* 24, 195–210. doi: 10.1007/s00530-017-0542-0

Hunter, P. G., and Schellenberg, E. G. (2010). "Music and emotion," in *Music Perception* (New York, NY: Springer), 129–164. doi: 10.1007/978-1-4419-6114-3_5

Hwang, W. J., and Wen, K. W. (1998). Fast KNN classification algorithm based on partial distance search. *Electron. Lett*. 34. 2062–2063. doi: 10.1049/el:19981427

Iacoviello, D., Petracca, A., Spezialetti, M., and Placidi, G. (2015). A real-time classification algorithm for EEG-based BCI driven by self-induced emotions. *Comput. Methods Prog. Biomed*. 122, 293–303. doi: 10.1016/j.cmpb.2015.08.011

Keelawat, P., Thammasan, N., Numao, M., and Kijsirikul, B. (2019). Spatiotemporal emotion recognition using deep CNN based on EEG during music listening. *arXiv preprint arXiv:1910.09719*.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., et al. (2011). DEAP: a database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput*. 3, 18–31. doi: 10.1109/T-AFFC.2011.15

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: a review. *Biol. Psychol*. 84, 394–421. doi: 10.1016/j.biopsycho.2010.03.010

Li, T., and Ogihara, M. (2003). "Detecting emotion in music," in *4th International Conference on Music Information Retrieval* (Baltimore, MD).

Li, Y., and Zheng, W. (2021). Emotion recognition and regulation based on stacked sparse auto-encoder network and personalized reconfigurable music. *Mathematics* 9, 593. doi: 10.3390/math9060593

Lin, W.-C., Chiu, H.-W., and Hsu, C.-Y. (2006). "Discovering EEG signals response to musical signal stimuli by time-frequency analysis and independent component analysis," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference* (Shanghai: IEEE), 2765–2768. doi: 10.1109/IEMBS.2005.1617045

Lin, Y.-P., Wang, C.-H., Jung, T.-P., Wu, T.-L., Jeng, S.-K., Duann, J.-R., et al. (2010). EEG-based emotion recognition in music listening. *IEEE Trans. Biomed. Eng*. 57, 1798–1806. doi: 10.1109/TBME.2010.2048568

Lin, Y.-P., Wang, C.-H., Wu, T.-L., Jeng, S.-K., and Chen, J.-H. (2009). "EEG-based emotion recognition in music listening: a comparison of schemes for multiclass support vector machine," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing* (Taipei: IEEE), 489–492. doi: 10.1109/ICASSP.2009.4959627

Liu, H., Fang, Y., and Huang, Q. (2019). "Music emotion recognition using a variant of recurrent neural network," in *2018 International Conference on Mathematics, Modeling, Simulation and Statistics Application (MMSSA 2018)* (Shanghai: Atlantis Press), 15–18. doi: 10.2991/mmssa-18.2019.4

Liu, J., Sun, L., Huang, M., Xu, Y., and Li, R. (2022). Enhancing emotion recognition using region-specific electroencephalogram data and dynamic functional connectivity. *Front. Neurosci*. 16, 884475. doi: 10.3389/fnins.2022.884475

Liu, Y., Liu, Y., Zhao, Y., and Hua, K. A. (2015). What strikes the strings of your heart?–Feature mining for music emotion analysis. *IEEE Trans. Affect. Comput*. 6, 247–260. doi: 10.1109/TAFFC.2015.2396151

Lu, L., Liu, D., and Zhang, H.-J. (2005). Automatic mood detection and tracking of music audio signals. *IEEE Trans. Audio Speech Lang. Process*. 14, 5–18. doi: 10.1109/TSA.2005.860344

Luo, G., Chen, H., Li, Z., and Wang, M. (2022). "Music generation based on emotional EEG," in *2022 the 6th International Conference on Innovation in Artificial Intelligence (ICIAI)* (Harbin), 143–147. doi: 10.1145/3529466.3529492

Malheiro, R., Panda, R., Gomes, P., and Paiva, R. P. (2016). Emotionally-relevant features for classification and regression of music lyrics. *IEEE Trans. Affect. Comput*. 9, 240–254. doi: 10.1109/TAFFC.2016.2598569

Mauss, I. B., and Robinson, M. D. (2009). Measures of emotion: a review. *Cogn. Emot*. 23, 209–237. doi: 10.1080/02699930802204677

Naser, D. S., and Saha, G. (2021). Influence of music liking on EEG based emotion recognition. *Biomed. Signal Process. Control* 64, 102251. doi: 10.1016/j.bspc.2020.102251

Nie, D., Wang, X.-W., Shi, L.-C., and Lu, B.-L. (2011). "EEG-based emotion recognition during watching movies," in *2011 5th International IEEE/EMBS Conference on Neural Engineering* (Boston, MA: IEEE), 667–670. doi: 10.1109/NER.2011.5910636

Nordström, H., and Laukka, P. (2019). The time course of emotion recognition in speech and music. *J. Acoust. Soc. Am*. 145, 3058–3074. doi: 10.1121/1.5108601

Pallavicini, F., Ferrari, A., Pepe, A., Garcea, G., Zanacchi, A., and Mantovani, F. (2018). "Effectiveness of virtual reality survival horror games for the emotional elicitation: preliminary insights using resident evil 7: biohazard," in *International Conference on Universal Access in Human-Computer Interaction* (Las Vegas, NV: Springer), 87–101. doi: 10.1007/978-3-319-92052-8_8

Panksepp, J. (2004). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. New York, NY: Oxford University Press. doi: 10.1176/appi.ajp.159.10.1805

Peretz, I. (2001). "Listen to the brain. A biological perspective on musical emotions," in *Music and Emotion: Theory and Research*, eds P. N. Juslin and J. A. Sloboda (New York, NY: Oxford University Press), 105–134.

Phneah, S. W., and Nisar, H. (2017). EEG-based alpha neurofeedback training for mood enhancement. *Austral. Phys. Eng. Sci. Med*. 40, 325–336. doi: 10.1007/s13246-017-0538-2

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

Pisipati, M., and Nandy, A. (2021). "Human emotion recognition using EEG signal in music listening," in *2021 IEEE 18th India Council International Conference (INDICON)* (Guwahati: IEEE), 1–6. doi: 10.1109/INDICON52576.2021.9691724

Rahman, J. S., Gedeon, T., Caldwell, S., and Jones, R. (2020). "Brain melody informatics: analysing effects of music on brainwave patterns," in *2020 International Joint Conference on Neural Networks (IJCNN)* (IEEE), 1–8. doi: 10.1109/IJCNN48605.2020.9207392

Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol*. 39, 1161. doi: 10.1037/h0077714

Salama, E. S., El-Khoribi, R. A., Shoman, M. E., and Shalaby, M. A. W. (2018). EEG-based emotion recognition using 3D convolutional neural networks. *Int. J. Adv. Comput. Sci. Appl*. 9, 329–337. doi: 10.14569/IJACSA.2018.090843

Salzman, C. D., and Fusi, S. (2010). Emotion, cognition, and mental state representation in amygdala and prefrontal cortex. *Annu. Rev. Neurosci*. 33, 173. doi: 10.1146/annurev.neuro.051508.135256

Sangnark, S., Autthasan, P., Ponglertnapakorn, P., Chalekarn, P., Sudhawiyangkul, T., Trakulruangroj, M., et al. (2021). Revealing preference in popular music through familiarity and brain response. *IEEE Sens. J*. 21, 14931–14940. doi: 10.1109/JSEN.2021.3073040

Sarkar, R., and Saha, S. K. (2015). "Music genre classification using emd and pitch based feature," in *2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR)* (Kolkata: IEEE), 1–6. doi: 10.1109/ICAPR.2015.7050714

Schmidt, E. M., Turnbull, D., and Kim, Y. E. (2010). "Feature selection for content-based, time-varying musical emotion regression," in *Proceedings of the International Conference on Multimedia Information Retrieval*, 267–274. doi: 10.1145/1743384.1743431

Schubert, E. (2003). Update of the hevner adjective checklist. *Percept. Motor Skills*, 96(3_Suppl), 1117–1122. doi: 10.2466/pms.2003.96.3c.1117

Shahabi, H., and Moghimi, S. (2016). Toward automatic detection of brain responses to emotional music through analysis of EEG effective connectivity. *Comput. Hum. Behav*. 58, 231–239. doi: 10.1016/j.chb.2016.01.005

Sheykhivand, S., Mousavi, Z., Rezaii, T. Y., and Farzamnia, A. (2020). Recognizing emotions evoked by music using cnn-lstm networks on EEG signals. *IEEE Access* 8, 139332–139345. doi: 10.1109/ACCESS.2020.3011882

Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., et al. (2018). A review of emotion recognition using physiological signals. *Sensors* 18, 2074. doi: 10.3390/s18072074

Soleymani, M., Aljanaki, A., Yang, Y.-H., Caro, M. N., Eyben, F., Markov, K., et al. (2014). "Emotional analysis of music: a comparison of methods," in *Proceedings of the 22nd ACM International Conference on Multimedia* (Orlando, FL), 1161–1164. doi: 10.1145/2647868.2655019

Sourina, O., Liu, Y., and Nguyen, M. K. (2012). Real-time EEG-based emotion recognition for music therapy. *J. Multimodal User Interfaces* 5, 27–35. doi: 10.1007/s12193-011-0080-6

Tellegen, A., Watson, D., and Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychol. Sci*. 10, 297–303. doi: 10.1111/1467-9280.00157

Thammasan, N., Fukui, K.-I., and Numao, M. (2016a). "Application of deep belief networks in EEG-based dynamic music-emotion recognition," in *2016 International Joint Conference on Neural Networks (IJCNN)* (Vancouver, BC: IEEE), 881–888. doi: 10.1109/IJCNN.2016.7727292

Thammasan, N., Fukui, K.-I., and Numao, M. (2016b). Fusion of EEG and musical features in continuous music-emotion recognition. *arXiv preprint arXiv:1611.10120*.

Thammasan, N., Fukui, K.-I., and Numao, M. (2016c). "An investigation of annotation smoothing for EEG-based continuous music-emotion recognition," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 3323–3328. Budapest: IEEE. doi: 10.1109/SMC.2016.7844747

Thammasan, N., Fukui, K.-I., and Numao, M. (2017a). "Multimodal fusion of EEG and musical features in music-emotion recognition," in *Proceedings*

*of the AAAI Conference on Artificial Intelligence* (San Francisco, CA). doi: 10.1609/aaai.v31i1.11112

Thammasan, N., Moriyama, K., Fukui, K.-I., and Numao, M. (2017b). Familiarity effects in EEG-based emotion recognition. *Brain informatics* 4, 39–50. doi: 10.1007/s40708-016-0051-5

Thayer, R. E. (1990). *The Biopsychology of Mood and Arousal*. Chicago: Oxford University Press. doi: 10.1086/417761

Verma, G. K., and Tiwary, U. S. (2017). Affect representation and recognition in 3d continuous valence-arousal-dominance space. *Multimedia Tools Appl*. 76, 2159–2183. doi: 10.1007/s11042-015-3119-y

Vuilleumier, P., and Trost, W. (2015). Music and emotions: from enchantment to entrainment. *Ann. N. Y. Acad. Sci*. 1337, 212–222. doi: 10.1111/nyas.12676

Wang, J.-C., Yang, Y.-H., Wang, H.-M., and Jeng, S.-K. (2012). "The acoustic emotion gaussians model for emotion-based music annotation and retrieval," in *Proceedings of the 20th ACM International Conference on Multimedia* (Nara), 89–98. doi: 10.1145/2393347.2396494

Waugh, C., Shing, E., and Avery, B. (2015). Temporal dynamics of emotional processing in the brain. *Emot. Rev*. 7, 323–329. doi: 10.1177/1754073915590615

Widmann, A., Schröger, E., and Wetzel, N. (2018). Emotion lies in the eye of the listener: emotional arousal to novel sounds is reflected in the sympathetic contribution to the pupil dilation response and the P3. *Biol. Psychol*. 133, 10–17. doi: 10.1016/j.biopsycho.2018.01.010

Xia, G., Tay, J., Dannenberg, R., and Veloso, M. (2012). "Autonomous robot dancing driven by beats and emotions of music," in *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems* (Richland, SC), 205–212.

Xianyu, H., Li, X., Chen, W., Meng, F., Tian, J., Xu, M., et al. (2016). "SVR based double-scale regression for dynamic emotion prediction in music," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Shanghai: IEEE), 549–553. doi: 10.1109/ICASSP.2016.7471735

Yang, Y.-H., and Chen, H. H. (2011). Prediction of the distribution of perceived music emotions using discrete samples. *IEEE Trans. Audio Speech Lang. Process*. 19, 2184–2196. doi: 10.1109/TASL.2011.2118752

Yang, Y.-H., Lin, Y.-C., Cheng, H.-T., Liao, I.-B., Ho, Y.-C., and Chen, H. H. (2008a). "Toward multi-modal music emotion classification," in *Pacific-Rim Conference on Multimedia* (Tainan: Springer), 70–79. doi: 10.1007/978-3-540-89796-5_8

Yang, Y.-H., Lin, Y.-C., Su, Y.-F., and Chen, H. H. (2008b). A regression approach to music emotion recognition. *IEEE Trans. Audio Speech Lang. Process*. 16, 448–457. doi: 10.1109/TASL.2007.911513

Yang, Y.-H., Liu, C.-C., and Chen, H. H. (2006). "Music emotion classification: a fuzzy approach," in *Proceedings of the 14th ACM International Conference on Multimedia* (New York, NY), 81–84. doi: 10.1145/1180639.1180665

Zainab, R., and Majid, M. (2021). Emotion recognition based on EEG signals in response to bilingual music tracks. *Int. Arab J. Inf. Technol*. 18, 286–296. doi: 10.34028/iajit/18/3/4

Zhang, H., Berg, A. C., Maire, M., and Malik, J. (2006). "SVM-KNN: discriminative nearest neighbor classification for visual category recognition," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (New York, NY: IEEE), 2126–2136. doi: 10.1109/CVPR.2006.301

Zhang, Q., Chen, X., Zhan, Q., Yang, T., and Xia, S. (2017). Respiration-based emotion recognition with deep learning. *Comput. Indus*. 92, 84–90. doi: 10.1016/j.compind.2017.04.005

# Frontiers in Neuroinformatics

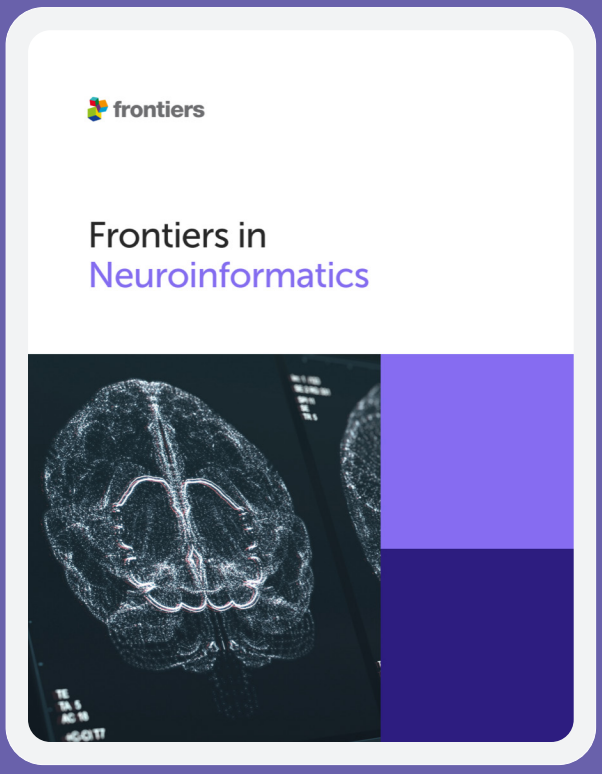
**Leading journal supporting neuroscience in the information age**

Part of the most cited neuroscience journal series, developing computational models and analytical tools used to share, integrate and analyze experimental data about the nervous system functions.

---

## 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