Recognition of single upper limb motor imagery tasks from EEG using multi-branch fusion convolutional neural network

Rui Zhang, Yadi Chen, Zongxin Xu, Lipeng Zhang, Yuxia Hu and Mingming Chen*

Henan Key Laboratory of Brain Science and Brain-Computer Interface Technology, School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou, China

Motor imagery-based brain-computer interfaces (MI-BCI) have important application values in the field of neurorehabilitation and robot control. At present, MI-BCI mostly use bilateral upper limb motor tasks, but there are relatively few studies on single upper limb MI tasks. In this work, we conducted studies on the recognition of motor imagery EEG signals of the right upper limb and proposed a multi-branch fusion convolutional neural network (MF-CNN) for learning the features of the raw EEG signals as well as the two-dimensional time-frequency maps at the same time. The dataset used in this study contained three types of motor imagery tasks: extending the arm, rotating the wrist, and grasping the object, 25 subjects were included. In the binary classification experiment between the grasping object and the arm-extending tasks, MF-CNN achieved an average classification accuracy of 78.52% and kappa value of 0.57. When all three tasks were used for classification, the accuracy and kappa value were 57.06% and 0.36, respectively. The comparison results showed that the classification performance of MF-CNN is higher than that of single CNN branch algorithms in both binary-class and three-class classification. In conclusion, MF-CNN makes full use of the time-domain and frequency-domain features of EEG, can improve the decoding accuracy of single limb motor imagery tasks, and it contributes to the application of MI-BCI in motor function rehabilitation training after stroke.

KEYWORDS

single upper limb motor imagery, deep learning, brain-computer interface (BCI), convolutional neural network (CNN), feature fusion

1. Introduction

The brain-computer interface (BCI) establishes a channel for information exchange between the human brain and the outside world. It decodes the user's intent through reading and analyzing brain signals (Wolpaw et al., 2002), and has been linked to a wide range of devices, including the use of spellers, wheelchairs, robotic arms and robotic exoskeletons (Kaufmann and Kubler, 2014; Kwak et al., 2015; He et al., 2018; Kim et al., 2018; Penaloza and Nishio, 2018; Yu et al., 2018; Jeong et al., 2019; Yao et al., 2022). Among the various types of BCI paradigms, MI-BCI is one of the most important one because it has potential clinical application value. MI is a mental process that mimics motor intention without

actually eliciting motor behavior. It is an actively evoked EEG signal that has high application values in the field of neurorehabilitation because it can independently elicit potential activity in motor-related brain regions without external stimulation (Pfurtscheller and Neuper, 2001). The MI-BCI detects the user's motor intentions by capturing the potential changes, and the output command could be used to control functional electrical stimulation (FES), exoskeletons, or other rehabilitation assistive equipment (Biasiucci et al., 2018; Zhao et al., 2022). Thus MI-BCI is valuable in the medical rehabilitation pathway for patients with motor dysfunction through the provision of active rehabilitation training (Jeong et al., 2019; Romero-Laiseca et al., 2020). A large number of studies have shown that the addition of MI-BCI helps to promote the recovery of motor function and improve the quality of life of patients (Cervera et al., 2018; Yuan et al., 2021).

The majority of current researches on motor imagery EEG signal recognition focuses on movements of different body parts, such as the tongue, hands, and feet. These studies have produced excellent results, but it is uncommon to find studies on motor imagery EEG signal recognition of tasks that involve the same side of the limb. It is well known that limb motor dysfunction caused by stroke is often unilateral. In BCI-based rehabilitation training, motor imagery tasks using unilateral limbs are more natural and intuitive than motor imagery tasks between different body parts (Tavakolan et al., 2017; Ubeda et al., 2017). However, the classification of single limb motor imagery is more difficult and complex than that of different parts of the body, because similar brain regions are activated when performing different motor tasks for unilateral limbs (Bigdely-Shamlo et al., 2015; Jas et al., 2016; Taulu and Larson, 2021). Considering the low spatial resolution of EEG, it is not feasible to use the algorithms for multi-limb motor imagery EEG recognition to identify unilateral limb motor imagery EEG.

The issue of unilateral limb movement task recognition has begun to be focused on by some researchers in recent years. Edelman et al. (2016) reported that source space analysis can improve the classification accuracy of wrist movements, four different movements of the right hand (i.e., flexion and extension of the arm; left and right rotation of the wrist) were recognized with a classification accuracy of 81.4%. Ofner et al. (2017) encoded motor imagery tasks for the right hand into the time domain of low-frequency EEG signals to classify six different movements, including elbow flexion/extension, forearm left/right rotation, and hand opening/closing, and achieved an accuracy of 27%. A novel classification strategy using the combination of EMG and EEG signals was proposed by Li et al. (2017). They recognized a variety of upper limb movements such as hand open/close and wrist pronation/supination, and results showed that the classification performance achieved by the fusion features of EMG and EEG signals is significantly higher than that obtained by a single signal source of either EMG or EEG across all subjects. Loopez-Larraz et al. (2018) further used EMG activity as a complementary information to EEG to detect the motor intention, and also found that the fusion features achieved higher classification accuracy than EEG or EMG-based methods.

The end-to-end deep learning techniques provide a new development path for the recognition of motor imagery EEG. Inspired by the filter bank common spatial pattern (FBCSP), Schirrmeister et al. (2017) proposed three types of CNN-based

models for motor imagery classification based on the number of layers. Jeong et al. (2020b) proposed a hierarchical flow convolutional neural network model consisting of a two-stage CNN for extracting relevant features for multi-class tasks and decoding arm rotation tasks. Zhang X. et al. (2019) proposed a network model CNN-LSTM, the motor imagery EEG data were spatially filtered by the FBCSP algorithm to extract the spatial domain feature information from the original data at first, then the extracted feature were fed into the CNN, and the final classification was performed by the LSTM. Cho et al. (2021) proposed a two-stage network structure called NeuroGrasp, which used six different CNN-BLSTM networks to implicitly map EEG signals to six muscle synergy features based on EMG and generated kinematic images corresponding to the EMG signals based on the extracted features. In the second stage, the generated images and real EMG features are used together as SiamNet network input to train the model, so as to realize the classification of single upper limb motor imagery tasks.

Most of the motor imagery EEG decoding methods based on deep learning used a single type of feature, including raw EEG signals, time-frequency maps, and power spectral density features. However, a single feature input often cannot fully and effectively mine the information related to motor imagery in EEG. Inspired by multimodal classification models, we proposed a multi-branch fusion convolutional network model (MF-CNN) for solving the classification problem of a single upper limb movement imagery task, which takes the EEG signals and the corresponding time-frequency maps as inputs simultaneously to make full use of the time-domain, frequency-domain and time-frequency-domain features of the EEG signal. The original EEG signal has high-resolution temporal information, and the discriminative features can be extracted by spatio-temporal convolution, while the two-dimensional time-frequency map contains rich time-frequency domain and spatial information. In this work, we first extracted the features of the above two inputs independently using two CNNs and then performed fusion classification, and the test results on the single upper limb motor imagery dataset showed that the proposed model achieved higher classification accuracy than single-input CNN.

2. Materials and methods

2.1. Datasets

The EEG data used in this work is the "Multimodal signal dataset for 11 intuitive movement tasks from single upper extremity during multiple recording sessions" from the Giga DB dataset completed by Jeong et al. (2020a). The dataset included intuitive upper limb movement data from 25 subjects, who were required to perform three types of motor tasks in a total of 11 categories, including 6 directions of arm extension movement (up, down, left, right, front, back), 3 kinds of object grasping action (cup, card, ball) and 2 kinds of wrist-twisting action (left rotation, right rotation), each type of movement was randomly executed 50 times, corresponding to 11 movements designed to be associated with each segmental movement of the arm, hand, and wrist, rather than continuous limb movements. The dataset included

not only EEG data but also magnetoencephalography (EMG) and electrooculogram (EOG) data, which are collected simultaneously in the same experimental setting while ensuring no interference between them. The data were acquired using a 60-channel EEG, 7-channel EMG, and 4-channel EOG. In the current work, only motor imagery EEG data were used, the EEG sensors were placed according to the international 10–20 system, and the sampling rate was set as 2,500 Hz. Our goal is to classify the motor imagery EEG of the three types of actions, so we selected forward extension of the arm, grasping the cup, and rotation of the wrist to the left from the three types of actions for the following study.

2.2. Algorithm framework

The workflow of the algorithm was shown in **Figure 1**. The time-frequency maps were firstly obtained by continuous wavelet transform (CWT) method, then both the EEG signals and the corresponding time-frequency maps were sent to the MF-CNN model, which consisted of two CNN network branches. After the process of convolution and pooling, the output features from the two branches were fused and combined into a one-dimensional vector. Finally, the one-dimensional feature vector was sent into a classifier to obtain the prediction results.

2.3. EEG signal pre-processing

When subjects were preparing or performing motor tasks, event-related desynchronization (ERD) and event-related synchronization (ERS) can be observed in the sensorimotor cortex of the brain (Pfurtscheller and da Silva, 1999; McFarland et al., 2000). Therefore, we selected 20 EEG channels on the sensorimotor cortex region to analyze (including FC1-6, C1-6, CP1-6, CZ, and CPZ). The selected EEG data were band-pass filtered within 8–30 Hz (Sreeja et al., 2017) and downsampled to 250 Hz. All the 4 s of EEG data during the motor task of a single trial were intercepted for subsequent processing, thus the EEG segment of each trial could be defined as a $20 \times 1,000$ matrix, where 20 was the number of channels and 1,000 was the length of

the data. The preprocessed EEG signals were used as input for the EEG-CNN branch and the time-frequency map conversion.

In terms of time-frequency map transformation, Tabar and Halici (2017) proposed a method based on the short-time Fourier transform (STFT) to extract time-frequency features and constructed a three-channel stacked time-frequency map for subsequent classification. However, the time window of the STFT algorithm is fixed, so the time-frequency resolution is also fixed, which causes the problem of incompatibility between the time resolution and the spectral resolution. To solve this problem, wavelet transform based time-frequency analysis methods have been widely introduced to EEG signal feature extraction (Zhang et al., 2021). The wavelet transform replaced the infinite-length triangular function with a finite-length wavelet basis with attenuation, which made the window width inconsistent and thus enabled better local feature extraction. We chose Morlet wavelet as the basis function for the wavelet transform. As a single-frequency complex sinusoidal function under Gaussian envelope Morlet wavelet is the most commonly used complex-valued wavelet. Because it has a better local resolution in the time and frequency domain, it is often used in the decomposition of complex signals and time-frequency analysis (Lee and Choi, 2019). The features extracted from EEG signals through CWT include time and frequency information and are finally converted into two-dimensional time-frequency maps. **Figure 2** showed the example time-frequency maps of the three channels C3, C4, and CZ.

Since the conversion of the time-frequency map is generated for each channel individually, the 20 EEG channels we used could not all be combined into one image. And if only a few channels were selected, a lot of helpful information would be lost. To effectively utilize the information of each channel, we preprocessed the data to extract features and used CSP to filter the 20 channels of EEG signals in the spatial domain to obtain "virtual channels," and then generated the time-frequency maps. The basic principle of CSP is to find a set of optimal spatial filters for projection by diagonalizing matrices so that the difference in variance values between the two types of signals is maximized (Ramoser et al., 2000). For the three classification tasks we used the "One vs. Rest" strategy to extend the CSP to achieve multi-class CSP feature extraction (Dornhege et al., 2004). The spatially filtered EEG can be
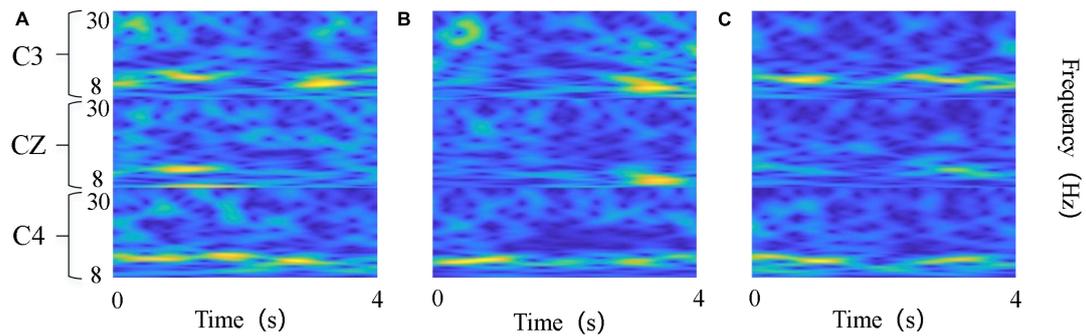

FIGURE 1
Workflow of the proposed algorithm.
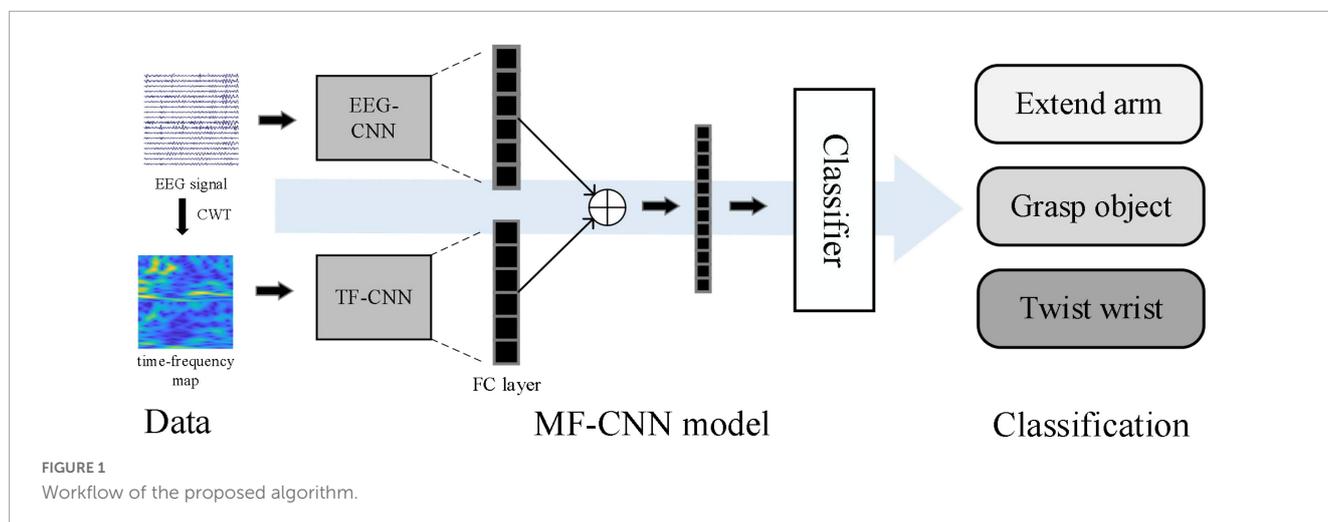
FIGURE 2
Time-frequency maps of the three kinds of tasks. **(A)** Left wrist rotation; **(B)** cup grasping; **(C)** forward arm extension. The abscissa denotes time points, and the ordinate denotes frequency bands.

calculated as:

$$Z_{M \times N} = W_{M \times M} E_{M \times N} \qquad (1)$$

where W is the projection matrix of CSP, M is the number of EEG data channels; N is the data length; E is the EEG data matrix; Z is the obtained EEG on "virtual channels."

The information of the feature matrix generated by the CSP algorithm is not equivalent, and the feature information is mainly concentrated in the head and tail of the feature matrix, while the middle feature information is not obvious and can be ignored. Therefore, the first m rows and the last m rows (2 m < M) of $Z_{M \times N}$ were usually selected. In this work, we chose $m = 1$, that is, the first and the last row of $Z_{M \times N}$ were selected to calculate the time-frequency map. The CWT was applied to the spatially filtered EEG data during the 4 s motor imagery to obtain time-frequency maps, and the maps were then saved as images with a resolution of 64 × 64. Such procedures were applied to all trials, and finally the motor imagery time-frequency map dataset was obtained. An example was shown in Figure 3.

2.4. Structure of MF-CNN

The classification of motor imagery EEG signals using deep learning networks based on CNN has proven successful and has good feature extraction capabilities (Lee and Kwon, 2016; Zhang P. et al., 2019). The common CNN models include convolutional layer, pooling layer, activation function, and fully connected layer. The convolution in the network is a local operation that can extract the deep features of the input signal through the kernel function, then the feature information can be obtained by the operation of each layer of the CNN model. In the convolution phase, the network input is convolved with the convolution kernel, and then the activation function $f(a)$ is used to output the feature maps, which can be expressed for each convolution layer as:

$$h_{ij}^k = f(a) = f((W^k * x)_{ij} + b_k) \qquad (2)$$

where x represents the input data, W_k is the weight matrix of the k_{th} convolution kernel, b_k corresponding to the deviation of the convolution kernel k, i and j denote the number of adjacent convolutional layers.

In the current work, the ReLU function was chosen as the activation function (Clevert et al., 2015), and it was defined as follows:

$$f(a) = ReLU(a) = ln(1 + e^a) \qquad (3)$$

The main purpose of the fully connected layer in a CNN is classification. To merge the features acquired from the previous side, each node in the fully connected layer is connected to full nodes in the preceding layer. After a number of prior convolutions, it can combine the local information with category differentiation, and the output of the final fully-connected layer is then sent to the classifier to output the prediction result.

Figure 4 showed the network structures of MF-CNN proposed in this study, it extracted the features of the raw EEG data and the time-frequency map simultaneously by using two CNN branches, and could obtain more comprehensive information hidden in the motor imagery EEG.

The EEG-CNN branch extracted spatial and temporal features from the raw EEG data, the dimensionality of the input EEG signal was 20 × 1,000 (channels × points). The input was successively fed through a feature extraction module made up of two convolutional layers and a maximum pooling layer in this branch. A one-dimensional convolutional kernel along the horizontal axis was used to extract the features of each channel to obtain the feature map as the output of this layer. The size of the convolution kernel was set to 3 × 1, and the step size was 1. After convolution, a feature map of the form $N_w \times N_f$ could be obtained, where N_w is the vector and N_f is the number of convolution kernels. Then, the data from the convolutional layer was downsampled using the pooling layer, which set a kernel size of 2 × 1 and a step size of 2. Subsequently, the fully connected layer flattens the features extracted through the convolutional layer.

The TF-CNN branch performed feature extraction on the input time-frequency map, and the size of the time-frequency map was 64 × 64 × 3, which represented an RGB image of size 64 × 64. VGG16 was used as the basic network framework in this branch (Zhao-Hong et al., 2019), the main feature of which was the inclusion of convolutional kernel computation and feedforward structure. It contained 16 hidden layers (13 convolutional layers and 3 fully connected layers), the convolutional part used a convolutional kernel of size 3 × 3 with a step size of 1, and a max pooling layer of size 2 × 2 with a step size of 2.
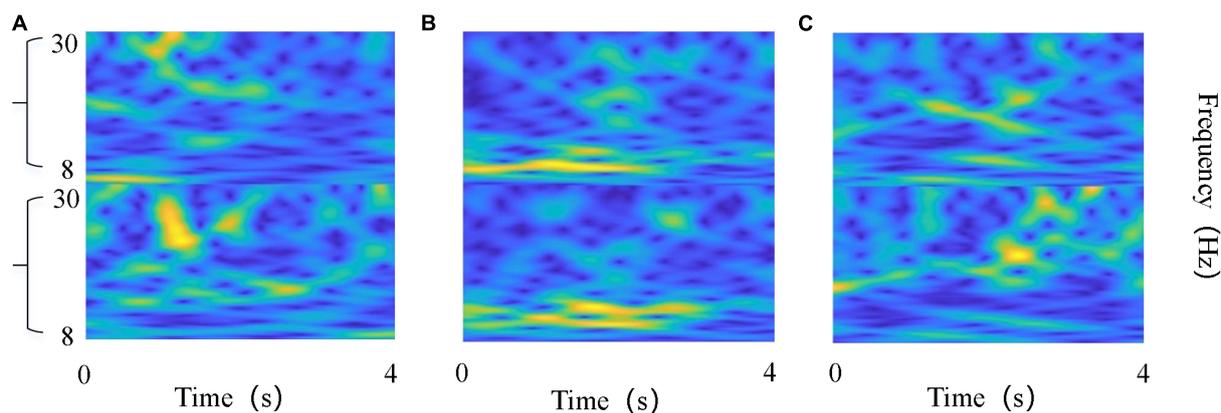
FIGURE 3
Time-frequency maps of the EEG on "virtual channel". **(A)** Left wrist rotation; **(B)** cup grasping; **(C)** forward arm extension. The abscissa denotes time points, and the ordinate denotes frequency bands.
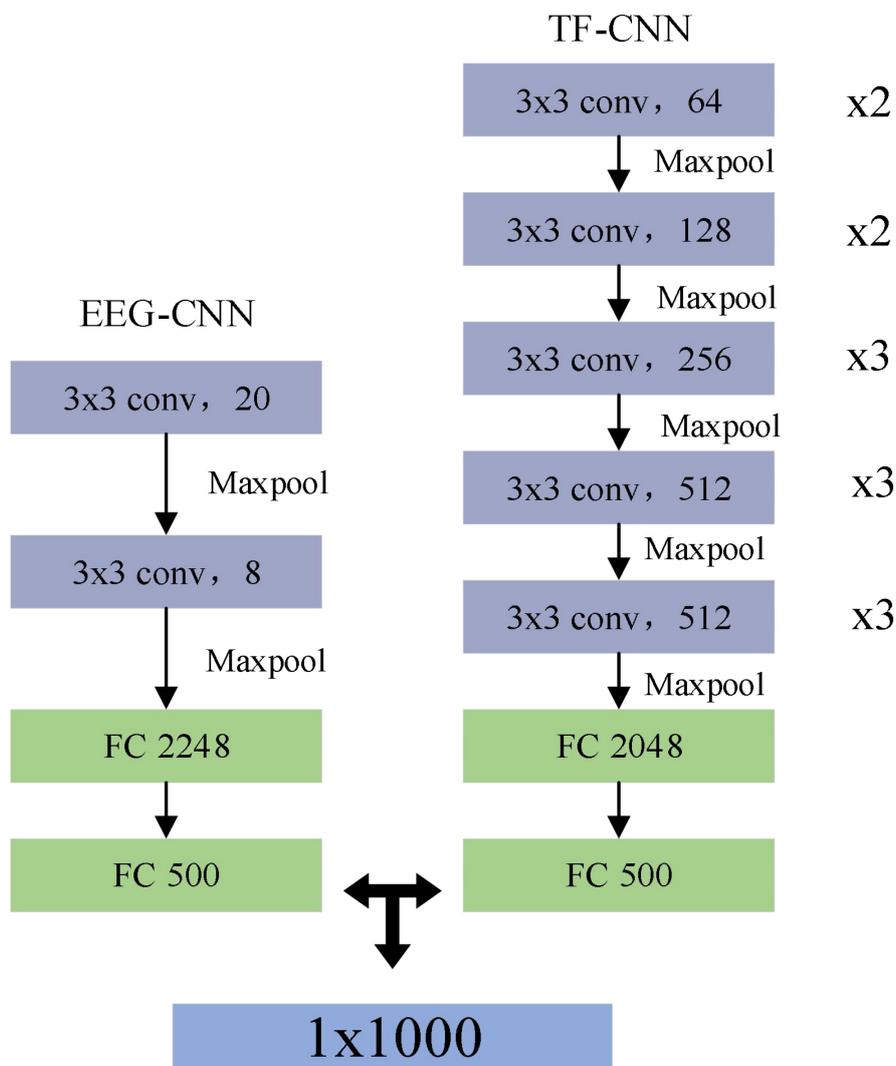
FIGURE 4
Structure of multi-branch fusion convolutional neural network (MF-CNN) model.

In the model training phase of the above two branches, the parameters of each network layer were updated using the Adam optimizer with β_1 0.9 and β_2 0.999, with an initial learning rate of 0.01.

2.5. Feature fusion method

Generally, fusion methods can be applied in two different ways: decision-level fusion and feature-level fusion. Decision-level fusion first trains different modalities with different models and then fuses the results of multiple model outputs. Feature-level fusion combines two or more feature vectors to construct a single feature vector to include more information (Zhang P. et al., 2019; Hatipoglu Yilmaz and Kose, 2021). In this study, feature-level fusion was selected. Before the feature fusion, the individual feature vector must have sufficient relevant features in order to provide a good classification model and achieve high classification performance. In CNN, the fully connected layer can integrate local information into global features for classification, which contains enough information. In addition, the output dimension of the last fully connected layer is consistent with the category of the sample, and the obtained information has been compressed, so it is not appropriate to serve as the final feature vector. Therefore, we chose to use the penultimate fully connected layer of these two branch networks as the fusion layer, and fused their outputs as the extracted features. Suppose the output feature vector of the EEG-CNN branch was $A = \{a_1, \cdots, a_m\}$, where m is the length of A, the feature vector output from the TF-CNN branch was $B = \{b_1, \cdots, b_n\}$, where n is the length of B. Then the fusion feature vector could be defined as $C = \{a_1, \cdots, a_m, b_1, \cdots, b_n\}$, and it is fed into the support vector machine (SVM) to complete the classification finally.

2.6. Performance evaluations

The classification accuracy was used as an evaluation criterion to compare the model's performance, which was calculated as follows.

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (4)$$

where T_P was the true-positives field in the confusion matrix, T_N was the true-negatives field, F_P was the false-positives field in the confusion matrix, F_N was the false-negatives field. It indicates the probability of correct prediction in all samples. In this paper, we compared the accuracy of six algorithms, including our proposed MF-CNN, the two single-branch CNNs (EEG-CNN and TF-CNN), EEGNET (Lawhern et al., 2018), ALEXNET (Iandola et al., 2016), and the classical CSP. EEG-CNN, EEGNET, and CSP used EEG signals as inputs, which are pre-processed in the same procedures as described in (section "2.3. EEG Signal pre-processing"). TF-CNN and ALEXNET used time-frequency maps as input for image classification.

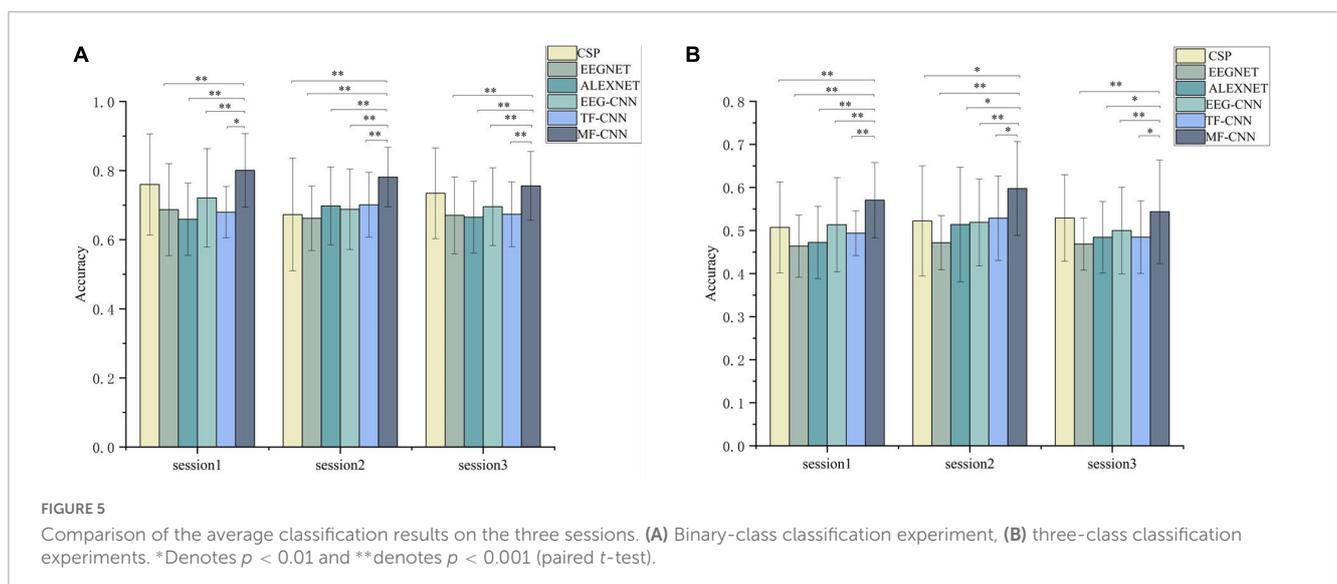In addition, we calculated kappa values (Tabar and Halici, 2017).

$$kappa = \frac{P_0 - P_e}{1 - P_e} \quad (5)$$

where p_0 represents the average classification accuracy and p_e represents the random classification accuracy for the n-class classification task.

3. Results

In this work, three-class and binary-class classification (grasp object vs. extend arm) test tasks were carried out separately to verify the performance of the proposed algorithm. The classification accuracies were calculated by using the five-fold cross-validation strategy, each subject's EEG data was divided into five equal subsets, one of which was randomly chosen as the testing dataset and the other subsets served as the training dataset. Such procedures were repeated five times, and the average accuracy was determined as the final classification accuracy. The three sessions for each subject were tested separately.

In order to verify the advantages brought by the dual-branch CNN, we compared the classification performance of MF-CNN

FIGURE 5
Comparison of the average classification results on the three sessions. **(A)** Binary-class classification experiment, **(B)** three-class classification experiments. *Denotes $p < 0.01$ and **denotes $p < 0.001$ (paired t-test).

model and single-branch CNN model. The single-branch CNN model was set up as an EEG-CNN branch for processing raw EEG signals and a TF-CNN branch for processing the time-frequency maps. The network architectures of these two single-branch CNN models were same as the EEG-CNN and TF-CNN branches in MF-CNN.

Figure 5 showed the classification results of the 25 subjects, the average classification accuracies of the single EEG-CNN branch were 70.8 and 51.08% separately for the binary-class and three-class classification experiments, while the single TF-CNN branch achieved 68.4 and 50.24%, respectively. It indicated that discriminative feature information can be extracted by the two kinds of single CNN branches. The accuracy obtained was higher than EEGNET and ALEXNET, but lower than CSP. After merging the features obtained from the two branches, MF-CNN achieved average accuracies of 78.52 and 57.06% for the two classification experiments, both of which were higher than that of the single CNN branch model, and also higher than CSP, EEGNET and ALEXNET.

The statistical analysis was further performed between the four algorithms using paired t-test. The results demonstrated that the accuracies achieved by MF-CNN were significantly higher than that of EEG-CNN and TF-CNN in all sessions. In addition, the accuracy of MF-CNN is higher than that of the deep learning algorithms EEGNET and ALEXNET used as comparisons.

The confusion matrix of the three deep learning network models were obtained. As shown in Figures 6, 7, the column represented the true label, and the row represented the predicted label. It can be seen that the probability of

correct recognition of each motor imagery task by MF-CNN is higher than that of EEG-CNN and TF-CNN, and all the true positive values are greater than the true negative and false negative values for the three deep learning network models.

Finally, we calculated the kappa coefficient for each subject, and the mean results were shown in Table 1. The binary-class classification experiments obtained higher kappa values than the three-class classification experiment for all three deep learning models, and MF-CNN outperformed EEG-CNN and TF-CNN in the two experiments.

4. Discussion

In this study, we performed feature fusion at the feature level to recognize the single upper limb motor imagery tasks by using deep learning approach. The dataset we used include three different types of movements, including forward extension of the arm, grasping the cup, and rotation of the wrist to the left. These are complex movements of the upper limb of the body and are commonly used in daily life. The accurate classification on the motor imagery of these three movements is of great significance in the application of BCI-based upper limb motor rehabilitation training. In this paper, the MF-CNN model was proposed to extract fusion features from the original EEG signal and corresponding time-frequency map. In the comparative experiment conducted on the single upper limb motor imagery dataset, MF-CNN model achieved

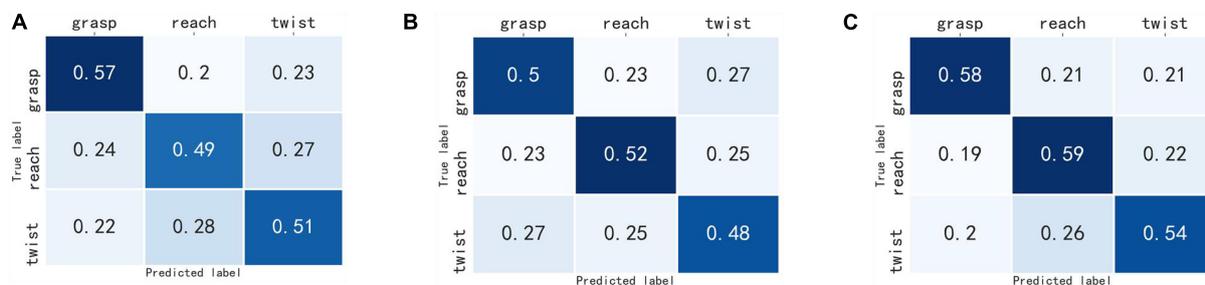

FIGURE 6
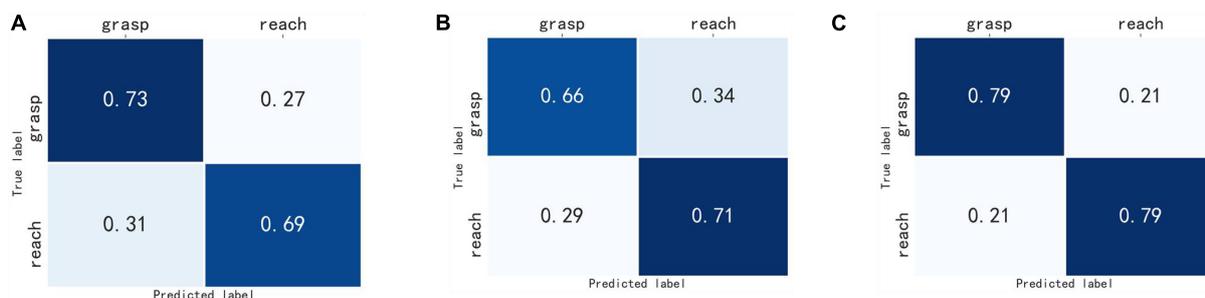Confusion matrix of the three-class classification experiment. **(A)** EEG-CNN, **(B)** TF-CNN, **(C)** MF-CNN.

FIGURE 7
Confusion matrix of the binary-class classification experiment. **(A)** EEG-CNN, **(B)** TF-CNN, **(C)** MF-CNN.

TABLE 1   Kappa values of the three deep learning models.

	EEG-CNN	TF-CNN	MF-CNN
Three-class classification	0.2662 ± 0.01	0.2536 ± 0.03	0.3559 ± 0.04
Binary-class classification	0.415 ± 0.04	0.3696 ± 0.02	0.5704 ± 0.02

better classification performance than two single CNN branches, EEGNET, ALEXNET, and CSP.

The EEG signal is non-stationary and non-linear (Yang et al., 2022). One of the most valuable methods for analyzing EEG signals is to transform them from one-dimensional time-domain signal to two-dimensional time-frequency map, which can concurrently combine the frequency feature in the time-domain and frequency-domain. The STFT and WT are the typical approaches for time-frequency analysis (Tabar and Halici, 2017; Yang et al., 2022). The STFT is obtained by adding a window on the basis of the Fourier transform. It has the ability of time-frequency analysis by using a fixed window function to analyze the signal segment. However, there are some shortcomings in the determination of the window function. If the window function is too narrow, the frequency domain analysis will be inaccurate; if it is too wide, the signal features in the time domain will be imprecise, affecting the time resolution. The WT is based on the Fourier transform but replacing the infinitely long triangular function base with a finite length and decaying wavelet base, and introduces scale and translation factors so that the resolution of the window function can change with the frequency characteristics. Compared with STFT, WT has the ability to obtain the local characteristics of the signal in both the time domain and the frequency domain (Khorrami and Moavenian, 2010). CWT offers a greater time-frequency resolution and can express the 3–5 s MI-EEG signal more precisely. Therefore, the EEG signal is transformed into a two-dimensional time-frequency map using the CWT method in the current study.

Previous studies based on deep learning usually used multi-channel stacked time-frequency maps as input to recognize motor imagery EEG (Dai et al., 2019). We have also tried this method, but could not obtain higher accuracy, only about 50% accuracy was achieved when using the time-frequency maps of C3, CZ, and C4. The reason for this may be that the aim of this study is to discriminative the motor imagery EEG of unilateral upper limbs, rather than the recognition of bilateral upper limb motor imagery in most studies. The difference between different actions in the unilateral upper limb motor imagery EEG is more minor (Ofner et al., 2017; Cho et al., 2021), thus it is challenging to obtain discriminative features with fewer channels. In order to make full use of the hidden information in the unilateral limb motor imagery EEG, we selected the EEG signals of 20 channels covering the sensorimotor cortex of the brain for analysis. However, it is not suitable to directly stack the 20-channel time-frequency maps as the input of TF-CNN. To solve this problem, we proposed to convert the time-frequency map based on the virtual channel after CSP spatial filtering. CSP

could extract the spatial distribution components of each class from the multi-channel EEG data (Ramoser et al., 2000), and the virtual channel signal generated after spatial filtering contained the discriminative information between classes. The results shown in Figure 5 validated the effectiveness of this approach.

There are many successful applications for EEG signal classification using feature fusion methods of multi-modal signals. For instance, the feature fusions of facial pictures or sound signals with EEG signals have been proven to improve the classification accuracy of emotion recognition (Wagner et al., 2011; Xing et al., 2019). In the current study, the two-dimensional time-frequency maps converted by raw EEG signals were used as a supplement to the time-domain EEG signal. Since the time-frequency maps were calculated from the original EEG signals, this did not increase the complexity of the data acquisition and was suitable for rehabilitation training scenarios. In the processing of time-frequency images, TF-CNN was carried out from the perspective of image processing, which is quite different from the time-domain EEG signals processing of EEG-CNN. The information extracted from the two CNN branches were complementary, MF-CNN fused these information to make them complement each other. The results shown in Figure 6 validated that the classification accuracy of single upper limb motor imagery EEG could be improved by such fusion strategy.

5. Conclusion

In this study, we proposed a deep learning framework named MF-CNN for classifying EEG signals associated with single upper limb motor imagery. There are two branches in MF-CNN, which can simultaneously extract features from the original EEG signal and the two-dimensional time-frequency map, and fully learn the time domain and time-frequency domain features of the EEG signal. The binary-class and three-class classification test results on the unilateral upper limb motor imagery dataset demonstrated that the proposed MF-CNN can improve the classification performance of unilateral upper limb motor imagery EEG effectively.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: http://gigadb.org/dataset/100788.

Ethics statement

This study was reviewed and approved by the Institutional Review Board at Korea University (1040548-KU-IRB-17-181-A-2). The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication

of any potentially identifiable images or data included in this article.

Author contributions

RZ and YC conceptualized the study, performed the majority of the experiments and analyses, made the figures, and wrote the first draft of the manuscript. ZX, LZ, YH, and MC performed some experiments, updated the figures, performed the statistics, and edited the manuscript. All authors approved the submitted version.

Funding

Conflict of interest

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

Publisher's note

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

References

Biasiucci, A., Leeb, R., Iturrate, I., Perdikis, S., Al-Khodairy, A., Corbet, T., et al. (2018). Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke. *Nat. Commun.* 9:2421. doi: 10.1038/s41467-018-04673-z

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K. M., and Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Front. Neuroinform.* 9:16. doi: 10.3389/fninf.2015.00016

Cervera, M. A., Soekadar, S. R., Ushiba, J., Millán, J. D. R., Liu, M., Birbaumer, N., et al. (2018). Brain-computer interfaces for post-stroke motor rehabilitation: A meta-analysis. *Ann. Clin. Transl. Neurol.* 5, 651–663.

Cho, J. H., Jeong, J. H., and Lee, S. W. (2021). NeuroGrasp: Real-Time EEG classification of high-level motor imagery tasks using a dual-stage deep learning framework. *IEEE Trans. Cybern.* 52, 13279–13292. doi: 10.1109/tcyb.2021.3122969

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

Dai, M., Zheng, D., Na, R., Wang, S., and Zhang, S. (2019). EEG classification of motor imagery using a novel deep learning framework. *Sensors* 19:551.

Dornhege, G., Blankertz, B., Curio, G., and Muller, K. R. (2004). Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. *IEEE Trans. Biomed. Eng.* 51, 993–1002. doi: 10.1109/tbme.2004.827088

Edelman, B. J., Baxter, B., and He, B. (2016). EEG source imaging enhances the decoding of complex right-hand motor imagery tasks. *IEEE Trans. Biomed. Eng.* 63, 4–14. doi: 10.1109/TBME.2015.2467312

Hatipoglu Yilmaz, B., and Kose, C. (2021). A novel signal to image transformation and feature level fusion for multimodal emotion recognition. *Biomed. Tech.* 66, 353–362. doi: 10.1515/bmt-2020-0229

He, Y., Eguren, D., Azorin, J. M., Grossman, R. G., Luu, T. P., and Contreras-Vidal, J. L. (2018). Brain-machine interfaces for controlling lower-limb powered robotic systems. *J. Neural Eng.* 15:021004. doi: 10.1088/1741-2552/aaa8c0

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size. *arXiv* [Preprint]. doi: 10.48550/arXiv.1602.07360

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., and Gramfort, A. (2016). Autoreject: Automated artifact rejection for MEG and EEG data. *Neuroimage* 159, 417–429. doi: 10.1016/j.neuroimage.2017.06.030

Jeong, J.-H., Lee, B.-H., Lee, D.-H., Yun, Y.-D., and Lee, S.-W. (2020b). EEG classification of forearm movement imagery using a hierarchical flow convolutional neural network. *IEEE Access.* 8, 66941–66950.

Jeong, J. H., Cho, J. H., Shim, K. H., Kwon, B. H., Lee, B. H., Lee, D. Y., et al. (2020a). Multimodal signal dataset for 11 intuitive movement tasks from single upper extremity during multiple recording sessions. *Gigascience* 9:giaa098. doi: 10.1093/gigascience/giaa098

Jeong, J.-H., Shim, K.-H., Kim, D.-J., and Lee, S.-W. (2019). "Trajectory decoding of arm reaching movement imageries for brain-controlled robot arm system," in *Paper Presented at the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* (Honolulu: EMBC), 2019. doi: 10.1109/EMBC.2019.8856312

Kaufmann, T., and Kubler, A. (2014). Beyond maximum speed-a novel two-stimulus paradigm for brain-computer interfaces based on event-related potentials (P300-BCI). *J. Neural Eng.* 11:056004. doi: 10.1088/1741-2560/11/5/056004

Khorrami, H., and Moavenian, M. (2010). A comparative study of DWT, CWT and DCT transformations in ECG arrhythmias classification. *Expert Syst. Appl.* 37, 5751–5757.

Kim, K.-T., Suk, H.-I., and Lee, S.-W. (2018). Commanding a brain-controlled wheelchair using steady-state somatosensory evoked potentials. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 654–665. doi: 10.1109/TNSRE.2016.2597854

Kwak, N.-S., Muller, K.-R., and Lee, S.-W. (2015). A lower limb exoskeleton control system based on steady state visual evoked potentials. *J. Neural Eng.* 12:056009. doi: 10.1088/1741-2560/12/5/056009

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., and Lance, B. J. (2018). EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. *J. Neural Eng.* 15:056013. doi: 10.1088/1741-2552/aace8c

Lee, H., and Choi, Y. S. (2019). Application of continuous wavelet transform and convolutional neural network in decoding motor imagery brain-computer interface. *Entropy* 21:1199. doi: 10.3390/e21121199

Lee, H., and Kwon, H. (2016). Going deeper with contextual CNN for hyperspectral image classification. *IEEE Trans. Image Process.* 26, 4843–4855. doi: 10.1109/TIP.2017.2725580

Li, X. X., Samuel, O. W., Zhang, X., Wang, H., Fang, P., and Li, G. L. (2017). A motion-classification strategy based on sEMG-EEG signal combination for upper-limb amputees. *J. Neuroeng. Rehabil.* 14, 1–13. doi: 10.1186/s12984-016-0212-z

Loopez-Larraz, E., Birbaumer, N., and Ramos-Murguialday, A. (2018). "A hybrid EEG-EMG BMI improves the detection of movement intention in cortical stroke patients with complete hand paralysis," in *Paper Presented at the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* (Honolulu: EMBC). doi: 10.1109/EMBC.2018.8512711

McFarland, D. J., Miner, L. A., Vaughan, T. M., and Wolpaw, J. R. (2000). Mu and beta rhythm topographies during motor imagery and actual movements. *Brain Topogr.* 12, 177–186. doi: 10.1023/a:1023437823106

Ofner, P., Schwarz, A., Pereira, J., and Muller-Putz, G. R. (2017). Upper limb movements can be decoded from the time-domain of low-frequency EEG. *PLoS One* 12:e0182578. doi: 10.1371/journal.pone.0182578

Penaloza, C. I., and Nishio, S. (2018). BMI control of a third arm for multitasking. *Sci. Robot.* 3:eaat1228. doi: 10.1126/scirobotics.aat1228

Pfurtscheller, G., and da Silva, F. H. L. (1999). Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/s1388-2457(99)00141-8

Pfurtscheller, G., and Neuper, C. (2001). Motor imagery and direct brain-computer communication. *Proc. IEEE* 89, 1123–1134. doi: 10.1109/5.939829

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

Romero-Laiseca, M. A., Delisle-Rodriguez, D., Cardoso, V., Gurve, D., Loterio, F., Posses Nascimento, J. H., et al. (2020). A low-cost lower-limb brain-machine interface triggered by pedaling motor imagery for post-stroke patients rehabilitation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 988–996. doi: 10.1109/TNSRE.2020.2974056

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., et al. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain Mapp.* 38, 5391–5420. doi: 10.1002/hbm.23730

Sreeja, S. R., Rabha, J., Samanta, D., Mitra, P., and Sarma, M. (2017). "Classification of motor imagery based EEG signals using sparsity approach," in *Paper Presented at the 9th International Conference on Intelligent Human Computer Interaction (IHCI)*, (Evry: IHCI).

Tabar, Y. R., and Halici, U. (2017). A novel deep learning approach for classification of EEG motor imagery signals. *J. Neural Eng.* 14:016003. doi: 10.1088/1741-2560/14/1/016003

Taulu, S., and Larson, E. (2021). Unified expression of the quasi-static electromagnetic field: Demonstration with MEG and EEG signals. *IEEE Trans. Biomed. Eng.* 68, 992–1004. doi: 10.1109/TBME.2020.3009053

Tavakolan, M., Frehlick, Z., Yong, X., and Menon, C. (2017). Classifying three imaginary states of the same upper extremity using time-domain features. *PLoS One* 12:e0174161. doi: 10.1371/journal.pone.0174161

Ubeda, A., Azorin, J. M., Chavarriaga, R., and Millan, J. D. (2017). Classification of upper limb center-out reaching tasks by means of EEG-based continuous decoding techniques. *J. Neuroeng. Rehabil.* 14, 1–14. doi: 10.1186/s12984-017-0219-0

Wagner, J., Andre, E., Lingenfelser, F., and Kim, J. (2011). Exploring fusion methods for multimodal emotion recognition with missing data. *IEEE Trans. Affect. Comput.* 2, 206–218.

Wolpaw, R. J., Birbaumer, N., McFarland, J. D., Pfurtscheller, G., and Vaughan, T. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791.

Xing, B., Zhang, H., Zhang, K., Zhang, L., Wu, X., Shi, X., et al. (2019). Exploiting EEG signals and audiovisual feature fusion for video emotion recognition. *IEEE Access.* 7, 59844–59861.

Yang, J., Gao, S., and Shen, T. (2022). A two-branch CNN fusing temporal and frequency features for motor imagery EEG decoding. *Entropy* 24:376. doi: 10.3390/e24030376

Yao, D., Qin, Y., and Zhang, Y. (2022). From psychosomatic medicine, brain–computer interface to brain–apparatus communication. *Brain Apparatus Commun.* 1, 66–88.

Yu, Y., Liu, Y., Jiang, J., Yin, E., Zhou, Z., and Hu, D. (2018). An asynchronous control paradigm based on sequential motor imagery and its application in wheelchair navigation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 2367–2375. doi: 10.1109/TNSRE.2018.2881215

Yuan, K., Chen, C., Wang, X., Chu, W. C.-W., and Tong, R. K.-Y. (2021). BCI training effects on chronic stroke correlate with functional reorganization in motor-related regions: A concurrent EEG and fMRI study. *Brain Sci.* 11:56. doi: 10.3390/brainsci11010056

Zhang, H., Zhao, X., Wu, Z., Sun, B., and Li, T. (2021). Motor imagery recognition with automatic EEG channel selection and deep learning. *J. Neural Eng.* 18:016004. doi: 10.1088/1741-2552/abca16

Zhang, P., Wang, X., Zhang, W. H., and Chen, J. F. (2019). Learning spatial-spectral-temporal EEG features with recurrent 3D convolutional neural networks for cross-task mental workload assessment. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 31–42. doi: 10.1109/tnsre.2018.2884641

Zhang, R., Zong, Q., Dou, L., and Zhao, X. (2019). A novel hybrid deep learning scheme for four-class motor imagery classification. *J. Neural Eng.* 16:066004. doi: 10.1088/1741-2552/ab3471

Zhang, X., Shen, J., Din, Z. U., Liu, J., Wang, G., and Hu, B. (2019). Multimodal depression detection: Fusion of electroencephalography and paralinguistic behaviors using a novel strategy for classifier ensemble. *IEEE J. Biomed. Health Inform.* 23, 2265–2275. doi: 10.1109/JBHI.2019.2938247

Zhao, C.-G., Ju, F., Sun, W., Jiang, S., Xi, X., Wang, H., et al. (2022). Effects of training with a brain–computer interface-controlled robot on rehabilitation outcome in patients with subacute stroke: A randomized controlled trial. *Neurol. Ther.* 11, 679–695. doi: 10.1007/s40120-022-00333-z

Zhao-Hong, X. U., Liu, Y., Quan, J. C., and Chen, W. U. (2019). Buildings segmentation of remote sensing images based on VGG16 pre-encoding. *Sci. Technol. Eng.* 19, 250–255.